# Identifying Common Genetic Variants Associated With Disease Risk And Clinical Outcome In Epithelial Ovarian Cancer

*By*

Lydia Quaye

Thesis submitted to University College London in fulfilment of the requirement for the degree of Doctor of Philosophy

# Declaration

I, Lydia Quaye, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

# Abstract

Combinations of common germline low-moderate susceptibility alleles may be responsible for some of the 90% of ovarian cancer (OC) cases not explained by known risk genes. These alleles may also affect survival of OC patients.

The effects of 34 tagging single nucleotide polymorphisms (tSNPs) from candidate oncogenes (*BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA*) and 63 tSNPs from "functionally" relevant genes (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) on the risk and survival of OC sufferers were evaluated with ~1,800 cases and 3,045 controls. Associations were found between disease risk and *NMI* rs11683487 ($P\text{-}_{dominant}$=0.004) and *RUVBL1* rs13063604 ($P\text{-}_{trend}$=0.0192). These associations were not independently validated with additional samples, however, they remained significant when the results from both stages of genotyping were combined (P<0.05). Global tests of association with OC risk were significant for *BRAF*, *ERBB2*, *CASP5* and *RUVBL1* ($P\text{-}_{global}$<0.05). However, there was no evidence of an excess of significant associations from 340 SNPs investigated with the admixture maximum likelihood test ($P\text{-}_{trend}$=0.068).

*BRAF*, *FILIP1L*, *KRAS*, *RBBP8* and *RUVBL1* were also associated with the survival of all OC cases (P<0.05). When analysis was restricted to the 4 main histological subtypes of OC, additional associations were identified. Although these results are of particular interest, they were based on relatively small numbers of samples and have not been corrected for multiple testing, therefore they should be treated with

caution. The results from the secondary objective of the project, to evaluate whole genome amplification (WGA) of DNA and SNP multiplex platforms, are also described.

To conclude, associations were identified between candidate oncogenes and functionally relevant genes on the survival and susceptibility of ovarian cancer. The performance of WGA DNA on SNP multiplex genotyping platforms highlighted the importance of comparing WGA DNA with corresponding gDNA in order to ascertain quality of genotyping on the platform.

# Contents

# List of Figures

# List of Tables

# List of Appendices

# Dedication

For my parents, Benedicta (née Maison) and Brian Barrett-Stone, and my grandmother, Angela Maison, to whom I owe everything. My endless gratitude for your love, support and guidance. Thank you for being with me every step of the way.

# Acknowledgements

I would like to thank Dr Simon Gayther and Dr Susan Ramus for their supervision, help and guidance, without which this project would not have been possible. I would also like to thank my colleagues, the members of the Ovarian Cancer Association Consortium. Special thanks to Dr Honglin Song, who has been an excellent person to collaborate with over the past few years. Thank you very much, Honglin, for all your help and patience for far too many things to mention here. Thanks also to Leigh Pearce, Jonathan Tyrer and Paul Pharaoh for help with the statistics and running STATA; Susanne Krüger Kjaer and Estrid Hogdall for their encouragement and for allowing me to use their samples for my research.

Special thanks also to the people who fall under the friend and colleague category: Mark Cox, James Morris, Maria Notaridou, Dimitra Dafou, Chris Jones, Jeremy Ford, Ken Choi and everyone I have worked with at the Gynae Oncology Research lab. Thanks to those at Maple House and the Proteomic team for always making me feel welcome. Those at the Molecular Psychiatry lab: Andrew McQuillin, Jonathan Pimms *et al* for the scientific discussions.

I am indebted to Dr Lalaruque Khalique and Miss Eva Wozniak for proofreading this thesis, contributing to some of the work and being fantastic friends. Last but not least, many thanks to my family and friends for all the support and encouragement: Julie, Karen, Kirsty, Hazel, Daniel, the Bradley twins, Anila, Penny Allen, (Sp)Anna, Uncle William, Patricia, Chrystal, Grandma, Grandpa and everyone else who falls into this category.

# Publications from this thesis

M Notaridou*, L Quaye*, D Dafou, C Jones, H Song, E Høgdall, SK Kjaer, L Christensen, C Høgdall, J Blaakaer, V McGuire, AH Wu, DJ Van Den Berg, MC Pike, A Gentry-Maharaj, E Wozniak, T Sher, IJ Jacobs, J Tyrer, JM Schildkraut, PG Moorman, ES Iversen, A Jakubowska, K Medrek, J Lubiński, RB Ness, KB Moysich, G Lurie, LR Wilkens, ME Carney, S Wang-Gohrke, JA Doherty, MA Rossing, MW Beckmann, FC Thiel, AB Ekici, X Chen, J Beesley, J Gronwald, PA Fasching, J Chang-Claude, MT Goodman, G Chenevix-Trench, A Berchuck, CL Pearce, AS Whittemore, U Menon, PD Pharoah, SA Gayther, SJ Ramus; The Australian Ovarian Cancer Study Group/Australian Cancer Study (Ovarian Cancer); on behalf of the Ovarian Cancer Association Consortium (2010): Common alleles in candidate susceptibility genes associated with risk and development of epithelial ovarian cancer. *International Journal of Cancer* [in press]

L Quaye*, J Tyrer*, SJ Ramus, H Song, E Wozniak, RA DiCioccio, V McGuire, E Høgdall, C Høgdall, J Blaakaer, EL Goode, JM Schildkraut, DF Easton, S Krüger-Kjaer, AS Whittemore, SA Gayther, PD Pharoah (2009): Association between Common Germline Genetic Variation in 94 Candidate Genes or Regions and Risks of Invasive Epithelial Ovarian Cancer. *PLoS One* **4**(6):e5983.

L Quaye*, D Dafou*, SJ Ramus*, H Song, A Gentry-Maharaj, M Notaridou, E Høgdall, S Krüger-Kjaer, L Christensen, C Høgdall, DF Easton, I Jacobs, U Menon, PD Pharoah, SA Gayther (2009): Functional Complementation Studies Identify Candidate Genes and Common Genetic Variants Associated with Ovarian Cancer Survival. *Human Molecular Genetics* **18**(10):1869-78.

L Quaye, H Song, SJ Ramus, A Gentry-Maharaj, E Høgdall, RA DiCioccio, V McGuire, AH Wu, DJ Van Den Berg, MC Pike, E Wozniak, JA Doherty, MA Rossing, RB Ness, KB Moysich, C Høgdall, J Blaakaer; Ovarian Cancer Association Consortium, DF Easton, BA Ponder, IJ Jacobs, U Menon, AS Whittemore, S Krüger-Kjaer, CL Pearce, PD Pharoah, SA Gayther (2009): Tagging single-nucleotide polymorphisms in candidate oncogenes and susceptibility to ovarian cancer. *British Journal of Cancer* **100**(6):993-1001.

L Quaye, SA Gayther, SJ Ramus, RA Di Cioccio, V McGuire, E Høgdall, C Høgdall, J Blaakaer, DF Easton, BAJ Ponder, I Jacobs, S Krüger-Kjaer, AS Whittemore, CL Pearce, PD Pharoah, H Song (2008): The effects of common genetic variants in oncogenes on ovarian cancer survival. *Clinical Cancer Research* **14**(18):5833-9.

* Joint first authors

# Chapter 1: Introduction

## 1.1: Background

Ovarian cancer is the sixth most common cancer in women worldwide, but the initiation, progression and metastasis of the disease is still poorly understood (Parkin *et al*. 2005). The global incidence of ovarian cancer is approximately 205,000 per year and the death rate is 125,000 a year. Ovarian cancer is more common in Northern Europe, North America and other developed areas of the world (Parkin *et al*. 2005, Sankaranarayanan and Ferlay 2006). Africa and Asia have the lowest incidence of the disease (Parkin *et al*. 2005). The cumulative lifetime risk of ovarian cancer in the general population is 1% by the age of 70 years, but the risk is higher in individuals with a family history of breast or ovarian cancer (Sharma *et al*. 2001). Individuals with an affected first degree relative have a 3.1% chance of developing ovarian cancer (Stratton *et al*. 1998). Aside from age, family history is the strongest known risk factor for ovarian cancer (Ramus *et al*. 2007).

## 1.2: Symptoms and diagnosis of ovarian cancer

Although there are some symptoms associated with ovarian cancer, these symptoms are usually vague and non-distinct from other conditions such as irritable bowel syndrome. This can lead to a delay in diagnosing and treating ovarian cancer. Symptoms of ovarian cancer include a conspicuous abdominal mass, vaginal bleeding unrelated with menstruation, distended and hard abdomen, and abdominal pain (Lurie *et al*. 2009).

Recently, the "risk of malignancy index" (RMI) has been developed in order to pre-clinically assess an individual's risk of cancer, before they are referred to a gynaecological oncology clinician, if necessary. In ovarian cancer, the combined results of blood CA-125 levels (the molecular biomarker of ovarian and other cancers), menopausal status and transvaginal ultrasound results are used to estimate a woman's risk of disease before referral to a gynaecological oncologist (http://info.cancerresearchuk.org/cancerstats/types/ovary/symptoms/?a=5441). Definitive diagnosis of ovarian cancer is through histological examination of a suspected tumour.

## 1.3: Histological pathology of ovarian cancer

Ovarian cancer can occur in the ovarian surface epithelium, germ cells or stroma. Approximately 90% of ovarian cancer cases are of epithelial origin and of these 90% are malignant carcinomas (Auersperg *et al*. 2001, Weiss *et al*. 1977). There are three categories of epithelial ovarian cancer: benign cystadenomas, borderline epithelial ovarian cancer and invasive carcinomas (Scully 1999). Epithelial neoplasms are believed to arise from the ovarian surface epithelium, benign epithelial inclusion cysts and cyst-adenomas (Cheng *et al*. 2004), or in rare cases, from ovarian endometrial foci. There is also a theory proposing that some cases of ovarian cancer initiate from the fallopian tube (Dubeau 2008).

### 1.3.1: Histological subtypes

There are several histological subtypes of invasive epithelial ovarian cancer. The histological classification of ovarian cancer is reviewed in Kaku *et al*. (2003) and

Christie and Oehler (2006). These include serous adenocarcinoma, mucinous adenocarcinoma, endometrioid adenocarcinoma and clear cell carcinoma (McCluggage 2008). The serous histological subtype of ovarian cancer is the most common subtype, occurring in approximately 50% of malignant cases (Koonings *et al.* 1989; Seidman *et al.* 2004). Kooning *et al* found that there were differences in the age distributions of histological subtypes. There is conflicting data on whether the mucinous or the endometrioid is the second most common histological subtype. Seidman *et al* found that the frequencies of endometrioid, mucinous and clear cell ovarian cancer were 6.8%, 6% and 10%, respectively, from a sample of 220 cases (Seidman *et al*. 2004). However, Kooning *et al* established from 180 patients that the frequencies of endometrioid, mucinous and clear cell histological subtypes were 11%, 9% and 4%, respectively. Each histological subtype has a different underlying pathogenesis and "natural behaviour" in terms of disease progression. However, it has been shown that there is an element of subjectivity in the pathological typing of some samples (McCluggage 2008).

The differences in the underlying pathogenesis and behaviour of the tumours have led to the suggestion that the histological subtypes of ovarian cancer are different diseases, rather than different forms of the same disease. This suggestion is supported by the distinct molecular changes found in the different histological subtypes of ovarian cancer. *KRAS* mutations are predominantly found in mucinous tumours. However, alterations of *PTEN* and *CTNNB1* are found in low grade endometrioid carcinomas; and *BRCA1*, *BRCA2*, *TGFBR2* and *HNF1B* are associated with the clear cell subtype. The serous subtype can be separated into low and high grade carcinomas, which correlate with different molecular changes. Whilst

mutations in *TP53*, *BRCA1* and *BRCA2* are associated with high grade serous carcinomas, mutations in *BRAF* and *KRAS* are found in low grade serous tumours (Christie and Oehler 2006).

### 1.3.2: Stages of ovarian cancer

Staging of ovarian tumours is the definitive method of confirming ovarian cancer diagnosis. Staging ovarian cancer involves a pathological examination of the size of the tumour and whether the tumour has spread. The Fédération Internationale de Gynécologie et d'Obstétrique **(**FIGO) staging system is commonly used for this purpose. There are four major stages of ovarian cancer, and within each stage there are 3 sub-groups, except stage IV. In stage I (early stage) ovarian cancer the tumour is confined to either or both ovaries. Stage II comprises of tumours in one or both ovaries with pelvic extension. Stage III involves ovarian tumour(s) with microscopically confirmed peritoneal metastases outside of the pelvis and or regional lymph node metastases. Stage IV, the most advanced, involves distant metastases (http://info.cancerresearchuk.org/cancerstats/types/ovary/symptoms/?a=5441).

### 1.3.3: Grading of ovarian cancer

The grade of an ovarian tumour is based on the appearance of the cells under a microscope. There are 3 grades given to tumours, grade 1 (low-grade) contains well-differentiated cells, which look similar to normal cells. Grade 1 cells are slow-growing and are unlikely to spread. Grade 2 cells are moderately differentiated and appear more abnormal than the low grade cells. Grade 3 cells are poorly differentiated and fast-growing, with a high likelihood of spreading.

The stage, grade and histological subtype of the tumour are used for diagnostic purposes and to provide a prognosis. The pathological information is also used to evaluate the most appropriate treatment to use.

## 1.4: Treatment and survival of ovarian cancer patients

The primary treatment for the vast majority of ovarian cancer is surgical removal of the tumour. However, the full course of treatment is based on the type of ovarian cancer and also the stage of the tumour. Patients with borderline or low grade stage I tumours are likely to only require surgery, while those with stages II or III usually have surgery, followed by adjuvant, platinum-based combination chemotherapy(Cancer Research UK http://www.cancerhelp.org.uk/help/default.asp?page=3084). Patients with stage IV ovarian tumours tend to require more aggressive treatment, if the patient is well enough. Treatment of stage IV tumours typically involves shrinking the tumour with chemotherapy before and after debulking surgery. Radiotherapy is also used, in some cases, to relieve symptoms (http://www.cancerhelp.org.uk/help/default.asp?page=3084).

The five-year survival rate for ovarian cancer is between 20-30%, which is not different from around 30 years ago, and overall, 60% of ovarian cancer sufferers die from their disease (Vanderhyden *et al*. 2003). The relatively unvarying mortality rate for ovarian cancer over the past 30 years is a sharp contrast to the mortality rates of breast and cervical cancer over the same time period (see Figure 1.1). Thus, although the mortality rates of breast and cervical cancers were greater than that of

ovarian cancer in 1971, there have been consistent and significant reductions in the mortality rates for breast and cervical cancers since 1990 and 1976, respectively (Figure 1.1). As a consequence of these reductions, the mortality rate of cervical cancer has been less than ovarian cancer since 1988, and in the year 2003 the rate for cervical cancer was approximately 5 deaths per 100,000 patients compared with 12 per 100,000 for ovarian cancer.

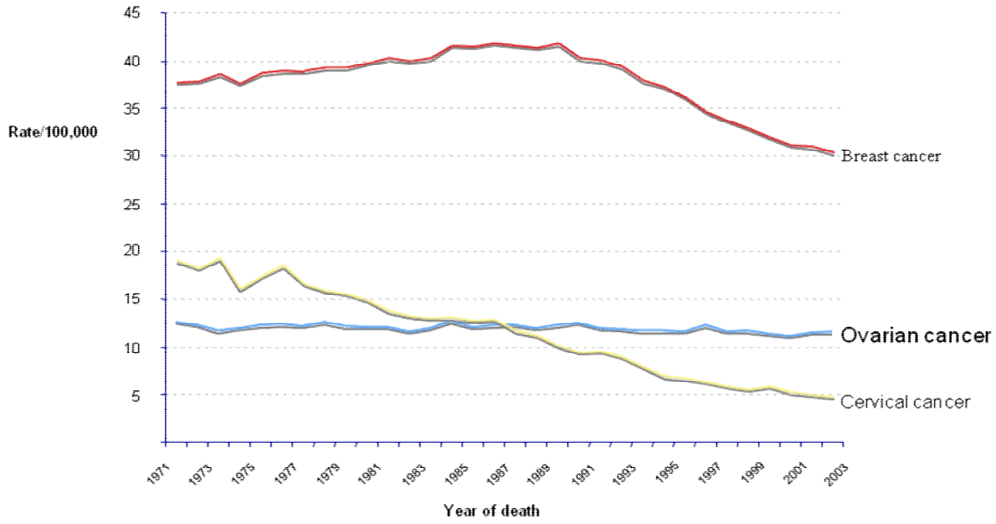



Figure 1.1: Trend of mortality rates for ovarian, breast and cervical cancer (1971-2003)

Engel and colleagues found 10-year survival rates between 32 and 34% for their ovarian cancer study participants, who were Caucasians from Germany. They also reported that despite improvements in the treatment of the disease and better survival of patients with FIGO stages I and II; the same did not apply to those with FIGO stages III-IV, and overall, there was no significant increase in the survival rate of ovarian cancer sufferers (Engel *et al*. 2002).

The pathological stage of ovarian tumours at diagnosis has the strongest effect on survival. Patients with late stage (FIGO III or IV) tumours have lower survival probabilities and thus worse prognosis, than those with early stage disease. The 5-year survival rate for ovarian cancer diagnosed in the early stages is greater than 70%. However, only 20% of ovarian cancer sufferers are diagnosed with early stage disease. In contrast, the 5-year survival rate reduces to approximately 15% for late stage disease with distant metastases, which affects a third of patients diagnosed (Cancer Research UK http://info.cancerresearchuk.org/cancerstats/types/ovary/survival/). Ovarian cancer is normally diagnosed when the disease is at an advanced stage, at which point, the prognosis is poor. This contributes to the high mortality from the disease. The age of the patient at diagnosis is also a determinant of survival. Older patients have poorer prognosis, compared with younger patients. However, this could be due to younger patients, despite their illness, being generally healthier than older patients.

## 1.5: Risk and protective factors of epithelial ovarian cancer

Aside from age, family history is the strongest risk factor for ovarian cancer (Amos and Struewing 1993). In families with affected individuals, the risk is conferred by the inheritance of a germline mutation. Other risk factors for ovarian cancer include early menarche, late menopause, infertility, nulliparity and low parity (Hildreth *et al.* 1981; Mori *et al.* 1988; DePasquale *et al.* 1998). Age has the strongest impact on the risk of ovarian cancer; as age increases, so does the risk of the disease. Greater than 80% (5,506 out of 6,596) of new ovarian cancer cases diagnosed in the UK in 2006 were at least 50 years old (Cancer Research UK). The age distribution of new

ovarian cancer cases in the UK are shown in Figure 1.2. Many of these other risk factors are a result of continuous ovulation, which encompasses early menarche, late menopause, nulliparity and low parity.

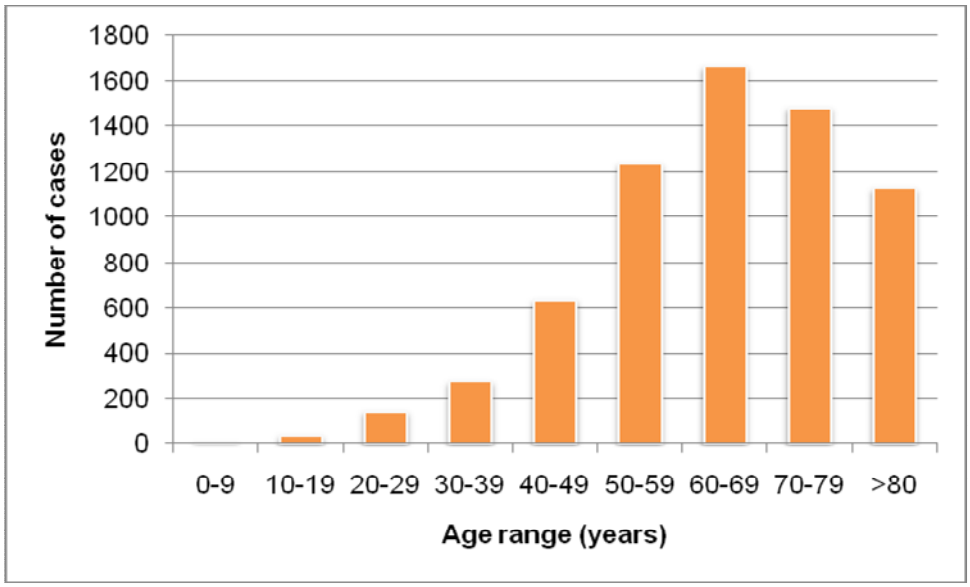



Figure 1.2: Age-distribution of new ovarian cancer cases in 2006 (UK)

N=6,596

There is conflicting evidence for the association between talcum powder use and ovarian cancer risk. Some studies have found that talcum powder use is associated with a moderate increase in ovarian cancer risk (Cramer *et al*. 1999, Gertig *et al*. 2000, Mills *et al*. 2004), but meta-analyses have not found a statistically significant association (Gross and Berg 1995, Huncharek *et al*. 2003). Studies of menopausal women on hormone replacement therapy have found that there is an increased risk of ovarian cancer in women who use oestrogen only treatment for more than 10 years, but the mechanisms through which the tumours arise are unknown (Folsom *et al*. 2004, Lacey *et al*. 2002, Rodriguez *et al*. 2001). Furthermore, there is also

conflicting data on whether factors such as consumption of alcohol, coffee, calcium, lactose, fibre or smoking increase susceptibility to ovarian cancer (Mori *et al.* 1988; Whittemore *et al.* 1988; Cramer 1989).

Use of the oral contraceptive pill is known to be a protective factor against the development of ovarian cancer (Casagrande *et al.* 1979; Franceschi *et al.* 1991). The oral contraceptive pill prevents ovulation by mimicking the levels of hormones normally present during pregnancy. Other factors such as high parity, increased duration of breast feeding, hysterectomy and tubal ligation have also been associated with reduced risk of the disease (Hildreth *et al.* 1981; Cramer *et al.* 1983; Whittemore *et al.* 1992; Hankinson *et al.* 1993).

## 1.6: Incessant ovulation and ovarian cancer

The ovarian surface epithelium is a monolayer of cells which covers the outside of the ovary (Vanderhyden *et al*. 2003) and ovarian cancer is believed to arise as a result of the continuous rupturing and mitotic repairing of the ovarian surface epithelium throughout a woman's life (Auersperg *et al*. 2001). Humans and chickens are two of very few animals known to spontaneously develop ovarian cancer. Fathalla was the first person to suggest a connection between incessant ovulation and ovarian cancer in women and hens (Fathalla 1971).

Every time mitosis occurs, there is a risk of DNA mutation. During ovulation, a follicle ruptures, releasing an ovum from the ovary. The ruptured follicle is subsequently a wound, which must be repaired by mitosis. This knowledge

combined with the fact that every time mitosis occurs there is a risk of DNA mutation, suggests a mechanism for the development of neoplasm in the ovaries. It has been proposed that the constant rupturing and repairing of the wounds on the ovarian surface throughout a woman's reproductive life contributes to the lifetime risk of ovarian cancer.

Ovarian cancer is primarily seen in perimenopausal and postmenopausal women. Greater than 90% of ovarian cancer cases are seen in women who are over 40 years old and the average age at which women in the general population are diagnosed is 60 years, and 50 years for familial cases (Holschneider and Berek 2000). This and other ovarian cancer risk factors, such as high parity, oral contraception use, support the incessant ovulation theory for the mechanism through which the disease arises (Casagrande *et al*. 1979).

## 1.7: Animal models of ovarian cancer

Little is known about the initiation, progression and metastasis of ovarian cancer despite research using ascites, primary ovarian tumour cell lines and animal models. Two varieties of animal models are used: those which spontaneously develop ovarian cancer (such as hens, some strains of mice, Wistar and Sprague-Dawley rats) and those which can be induced to develop ovarian cancer (sheep, guinea pig, rabbits) (Vanderhyden *et al*. 2003). Animal models used in ovarian cancer research are reviewed by Vanderhyden *et al*. (2003). Animal models have been useful in elucidating; the mechanism through which ovulation occurs, how inclusion cysts develop and the affects of steroids *in vivo*, but some of the results are conflicting,

and the disease is still poorly understood.

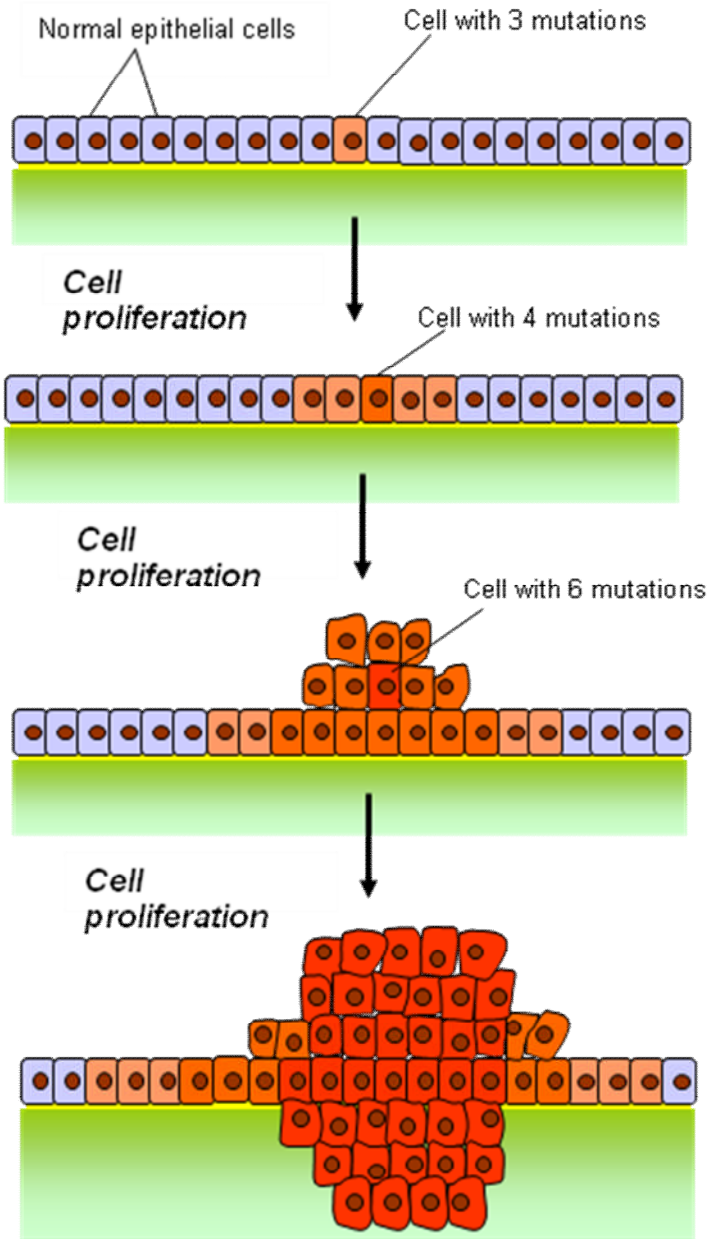



Figure 1.3: Accumulation of mutations leading cancer development

(Figure adapted from (Alberts 1994))

## 1.8: Genetics of epithelial ovarian cancer

Although the genetics of the initiation, progression and metastasis of ovarian cancer are poorly characterised, the general development of cancer is better understood. Normal cells are believed to transform into neoplastic cells after the acquisition of several mutations. Figure 1.3 shows the schematic progression of a tumour from a single mutated cell to a clump of mutated cells in a process known as clonal evolution (Alberts 1994). The mutation of some genes such as *BRCA1*, *BRCA2* and *TP53* have been proposed to lead to genomic instability, where the rate of gene mutation is accelerated due to the loss of genomic integrity and also the loss of a cell's ability to regulate normal cellular processes. The single mutated cell must have a mutation that gives it a growth advantage over the surrounding cells. The successive proliferation of the mutated cell and its daughter cells tend to lead to additional mutations. The clones with mutations for a growth advantage are continuously selected for and may become malignant.

Although the acquisition of mutations is essential for tumour development, the accumulation of mutations is not enough to cause cancer. A cell with the prerequisite genetic changes for cancer must be able to: evade apoptosis and the host's immune system; either have an increased rate of cell proliferation, or a decreased rate of cell death; become insensitive to internal and external inhibitory signals (i.e. cell-to-cell contact inhibition, anti-growth signals); become self sufficient in growth signals and either prevent cell differentiation; or promote cell de-differentiation (Boon 1993, Hanahan and Weinberg 2000). Angiogenesis must also occur in order for a tumour mass to get sufficient nutrients to grow beyond a

critical size. Ovarian cancer and other malignancies occur as a result of the accumulation of genetic alterations and favouring environment for tumour growth. Two groups of genes which are implicated in ovarian cancer development are oncogenes and tumour suppressor genes.

### 1.8.1: Oncogenes

Proto-oncogenes are essential in the normal functioning of cells, particularly in the regulation of cell division, proliferation, survival, motility and apoptosis. In adults, proto-oncogenes respond to stimuli from wound sites to repair the damage by stimulating growth factors. Oncogenes are mutated forms of proto-oncogenes. Some activating mutations can be within coding regions or regulatory elements. Proto-oncogenes can also be transformed by amplification of the region. Chemical carcinogens, ionising radiation, errors in DNA replication and faulty DNA damage repair can also cause the activating mutations (Balmain *et al*. 2003). Mutated oncogenes may still be able to elevate growth factor production and stimulate cell mitosis, but the activity may be poorly regulated, and this lack of regulation can lead to the transformation of normal cells into tumour cells (Hogdall *et al*. 2003a, Rhim 1988). Proto-oncogenes primarily have a dominant effect on cells, therefore the mutation of a single copy of the gene is sufficient for the gene to become an oncogene (Aunoble *et al*. 2000).

A number of oncogenes have been implicated in the development of ovarian cancer. These oncogenes include *AKT2, BCL2*, *BRAF*, *CDKN2A, MYC, CSF1R, CTNNB1*, *EGFR, ERBB2, FGF3*, *HRAS*, *KRAS, MDM2* and *PIK3CA*.

*BCL2*, which is located on chromosome 18q21.3, is involved in inhibiting apoptosis (White and Gilmore 1996). *BCL2* has been found to be over-expressed in 39% of ovarian tumours (Baekelandt *et al*. 1999).

The v-raf murine sarcoma viral oncogene homologue B1 (*BRAF*) is a proto-oncogene located on chromosome 7q34. The gene encodes a 84.4kDa protein, which acts as an effecter downstream of *KRAS* in the RAS-RAF-mitogen/extracellular signal-regulated kinase (MEK)-extracellular signal regulated kinase (ERK), and mitogen-activated protein kinase (MAPK) pathway. This pathway is critical in the transduction of cell growth signals from the cytoplasm into the nucleus (Russell and McCluggage 2004). Over-expression of *BRAF* has been found in ovarian, as well as a variety of other cancers, including melanomas, colorectal and thyroid cancer (Sieben *et al*. 2004). Mutations in *BRAF* in ovarian cancer have been reported to be as high as 36% (Sieben *et al*. 2004). However, according to the Catalogue Of Somatic Mutations In Cancer (COSMIC), a database from the Sanger Institute which catalogues mutations reported in ovarian and other malignancies, *BRAF* is one of the most mutated genes in ovarian cancer, with a frequency of 12% (www.sanger.ac.uk/genetics/CGP/cosmic/). Activating mutations of *BRAF* are more common in early stage ovarian cancer. *BRAF* mutations are predominantly found in tumours of the low grade serous histological subtype (Ho *et al*. 2001, Sieben *et al*. 2004). Mutations in *BRAF* have previously been shown to be associated with poor survival in patients diagnosed with papillary thyroid cancer and colon cancer (Abubaker *et al*. 2007, Samowitz *et al*. 2005).

The v-Ki-ras2 Kirsten rat sarcoma viral oncogene homologue (*KRAS*) is located on chromosome 12p12.1. The proto-oncogene encodes a 21.6 kDa protein, which is upstream of *BRAF* in the RAS-RAF-MEK-ERK-MAP kinase pathway (Russell and McCluggage 2004). Mutations in one of three RAS proto-oncogenes (*KRAS*, *HRAS* or *NRAS*), which result in the genes becoming activated are found in approximately 25% of human cancers (Gemignani *et al*. 2003). *KRAS*, like *BRAF*, is one of the most mutated genes found in ovarian cancer tumours and cell lines, with a frequency of 15% (Forbes *et al*. 2006). Codons 12 and 13 of the oncogene appear to be mutation "hotspots". Mutations in *KRAS* are predominantly found in mucinous histological subtype of ovarian cancer (50-68%), however mutations in codons 12 and 13 have also been observed in some non-mucinous ovarian cancers (Cuatrecasas *et al*. 1997, Cuatrecasas *et al*. 1998, Gemignani *et al*. 2003). Like *BRAF*, *KRAS* mutations tend to be detected in stage 1 tumours (Gemignani *et al*. 2003, Ho *et al*. 2001, Sieben *et al*. 2004). Furthermore, somatic alterations in *KRAS* have been associated with poor survival in patients with colorectal, lung and pancreatic cancers (De Roock *et al*. 2007, Kim *et al*. 2008, Lievre *et al*. 2006).

*CDKN2A* is a cell cycle control gene on chromosome 9p21.3. *CDKN2A* induces cell cycle arrest at G1 and G2/M checkpoints. *CDKN2A* is mutated in 10% of ovarian tumours and cell lines (www.sanger.ac.uk/genetics/CGP/cosmic/).

*CTNNB1* is located on chromosome 3p22-p21.3 and is involved in cell proliferation. This gene encodes β-catenin, which is a member of the Wnt signal transduction pathway. Approximately 30% of endometrioid ovarian carcinomas have *CTNNB1* mutations. It has been demonstrated that the β-catenin is normally degraded by

*APC*. However, mutant forms are resistant to the degradation, and thus accumulate in the cytoplasm. β-catenin may form complexes with transcription factors such as TCF/Lef-1, which translocates into the nucleus and activates transcription of genes, such as *MYC*, *CCND1* (also known as cyclin D1), *C-JUN* and *FRA-1* (Christie and Oehler 2006, Schlosshauer *et al*. 2002).

The v-erb-b2 erythroblastic leukaemia viral oncogene homologue 2 (*ERBB2*), also known as human epidermal growth factor receptor-2 (HER-2) and neuro/glioblastoma derived oncogene homologue (NEU) is a proto-oncogene located on chromosome 17q21.1. The *ERBB2* proto-oncogene encodes a transmembrane protein, which acts as a growth factor receptor and is involved in cell proliferation and cell differentiation (Wu *et al*. 2004). The over-expression of *ERBB2* is believed to cause the transcriptional activation of genes involved in cell proliferation (Aunoble *et al*. 2000). Ovarian, breast, prostate, lung, gastrointestinal, kidney, liver and bladder cancers have been shown to over-express *ERBB2* (Wu *et al*. 2004). For ovarian cancer, between 20-30% of stage III and IV tumours, primary tumour cells and cell lines over-express *ERBB2* (Hellstrom *et al*. 2001). Protein expression using antibody staining on a subset of ovarian tumours from the MALOVA study showed that 39% of the carcinomas over-expressed *ERBB2* (Hogdall *et al*. 2003). These findings are indicative of a tumour growth advantage when ERBB2 is over-expressed (Hellstrom *et al*. 2001). The variations in ERBB2 expression in the MALOVA study correlated with survival; where over-expression of ERBB2 was associated with poor clinical outcome (Hogdall *et al*. 2003). It has also been found that ovarian cancer cases homozygous for a polymorphism in *ERBB2*, I655V, which results in the production of the valine amino acid instead of isoleucine, have a

shorter survival period compared with the common homozygotes (who produce the isoleucine amino acid) (Pinto *et al*. 2005). Associations between ERBB2 over-expression in tumours and survival have also been reported for breast and colon cancers (Fritz *et al*. 2005).

The catalytic Class IA p110-alpha subunit of phosphatidylinositol 3-kinase (PIK3), which is known as *PIK3CA* is located on chromosome 3q26.3. This oncogene is a lipid kinase, which is involved in the regulation of cell proliferation, adhesion transformation, survival, apoptosis, and motility (Cantley 2002, Fruman *et al*. 1998, Volinia *et al*. 1994). There are two "hot spots" in which mutations in *PIK3CA* cluster – exons 9 and 20. Exon 9 contains the sequence for the helical domain, and exon 20 encodes the kinase domain. Mutations in these "hot spots" of the gene have been found in primary tumours and cell lines of cancers such as ovary, breast, lung, brain, colon and stomach (Muller *et al*. 2007). Shayesteh *et al*. initially identified the over-expression of *PIK3CA* in 7 out of 9 ovarian carcinoma cell lines. This over-expression correlated with fluorescent *in situ* hybridisation (FISH) data, which showed that *PIK3CA* was amplified in approximately 58% of the primary ovarian tumours (Shayesteh *et al*. 1999). The role of *PIK3CA* in tumour progression is reviewed in (Roymans and Slegers 2001, Samuels and Ericson 2006). In addition, mutations in the gene or over-expression of the gene may be correlated with worse clinical outcome in patients with ovarian, breast, thyroid, lung and colon cancer (Abubaker *et al*. 2007, Kato *et al*. 2007, Li *et al*. 2006, Woenckhaus *et al*. 2007).

*MYC* is a transcription factor which has a major role in neoplastic transformation. *MYC* over-expression caused by gene amplification induces uncontrolled hyper-

proliferation and occurs in approximately 30% of epithelial ovarian cancers (Aunoble *et al*. 2000). Some of the oncogenes and tumour suppressor genes implicated in ovarian cancer are shown in Table 1.1.

### 1.8.2: Tumour suppressor genes

Tumour suppressor genes (TSGs) are responsible for the inhibition of cell proliferation. The inactivation of a TSG results in a decrease in the expression of the TSG, which may lead to neoplastic growth. It has been proposed that there are two categories of TSGs: gatekeepers and caretakers. Gatekeepers are genes which act directly to regulate cell proliferation (Levitt and Hickson 2002). The retinoblastoma (*RB1*) and adenomatous polyposis coli (*APC*) genes are gatekeepers. The normal *RB1* protein represses cell proliferation and also regulates transcription (Classon and Harlow 2002). Loss of function mutations in both copies of *RB1* may result in a mutated form of the protein being produced, which is incapable of performing its normal function. Mutations in the *RB1* gene can lead to retinoblastoma, osteosarcoma and small-cell lung cancer (Taya 1997). *APC* is believed to inhibit the β-catenin protein, which is involved in the regulation of cell signal transduction, growth and adhesion (Fearnhead *et al*. 2002). The loss of these functions can lead to cells developing the anchorage independent characteristic of cancer cells and unregulated cell proliferation. Mutations in *APC* may lead to familial adenomatous polyposis coli and sporadic colon cancer (Seitz *et al*. 2003).

Table 1.1: Oncogenes and tumour suppressor genes involved in ovarian cancer development

| Gene | Chromosome | Function | Mutations (%) |
|---|---|---|---|
| **Oncogenes** | | | |
| *AKT2* | 19q13.2 | Regulation of cell proliferation. AKT is a major mediator of survival signals that protect cells from undergoing apoptosis. | 17* |
| *BCL2* | 12q15 | Acts as an ubiquitin ligase promoting proteasome dependent degradation of p53. Transcriptional target of p53. | 39* |
| *BRAF* | 7q34 | Involved in the transduction of mitogenic signals from the cell membrane to the nucleus. | 12 |
| *CDKN2A* | 9p21.3 | Induces cell cycle arrest at G1 and G2/M checkpoints, blocking them from phosphorylating RB1 and preventing exit from G1 phase of the cell cycle. P16-INK4a could act as a negative regulator of normal cells proliferation. | 10 |
| *MYC* | 8q24.21 | Transcription factor. Involved in regulation of gene expression. | 30* |
| *CSF1R* | 5q32 | Receptor. CSF1R activation by CSF1 results in increased growth, proliferation and differentiation. | 4 |
| *ERBB2* | 17q11.2-q12, 17q21.1 | Receptor tyrosine kinase. Transmembrane receptor. | 1 |
| *KRAS* | 12p12.1 | Involved in the transduction of mitogenic signals from the cell membrane to the nucleus. | 15 |
| *PIK3CA* | 3q26.3 | Signal transduction. Activated by growth factors. | 8 |
| **Tumour suppressor genes** | | | |
| *BRCA1* | 17q21 | Transcription factor. Plays essential role in DNA repair. Needed for cell arrest after DNA damage. | 3 |
| *BRCA2* | 13q12.3 | Transcription factor. Involved in DNA double strand break repair and homologous recombination. | 2 |
| *PTEN* | 10q23.31 | A phosphatase that negatively regulates the AKT/PKB pathway. Involved in cell cycle progression and cell survival. | 8 |
| *TP53* | 17p13.1 | Transcription factor. Induces cell growth arrest/apoptosis. | 28 |
| *APC* | 5q21-q22 | Antagonist of the Wnt signalling pathway. Involved in cell migration, cell adhesion, transcriptional activation, and apoptosis. | 9 |
| *RB1* | 13q14.2 | Negative regulator of the cell cycle. Regulates transcription. | 10 |

* over-expressed/amplified.

Caretaker tumour suppressor genes, such as *BRCA1*, *BRCA2* and *TP53*, encode proteins which are involved in the regulation of DNA replication, gene transcription, DNA repair or cell cycle checkpoints. All of these processes help maintain the integrity of the genome (Levitt and Hickson 2002). The *BRCA1*, *BRCA2* and *TP53* tumour suppressors are important in ovarian and breast cancers. *BRCA1* and *BRCA2* have many functions within the cell, which include DNA damage repair, DNA recombination, transcription and cell cycle checkpoint regulation (Venkitaraman 2002). Mutations in *BRCA1* and *BRCA2* can lead to the accumulation of mutations within a cell because of the loss of the appropriate DNA repair mechanism. This allows cells containing mutations to progress through cell cycle checkpoints. The *BRCA1* and *BRCA2* genes are very important in familial cases of breast and ovarian cancer, and their expression is reduced in some sporadic cancers, however mutations in these genes are relatively low (3% and 2%, respectively) when all ovarian cancers are considered (www.sanger.ac.uk/genetics/CGP/cosmic/).

*PTEN* is mutated in 8% of ovarian tumours (www.sanger.ac.uk/genetics/CGP/cosmic/). The TSG is located on chromosome 10q23.3, which encodes a phosphatase protein that inhibits the AKT/PKB signal transduction pathway. The protein is involved in cell cycle progression and cell survival. The expression of the gene can lead to cell cycle arrest, apoptosis and a reduction of cell motility (Christie and Oehler 2006).

The *TP53* protein is crucial for transcription, DNA repair, cell cycle control and apoptosis (French *et al*. 2001, Hulla *et al*. 2001). *TP53* is one of the most often mutated genes in human cancer – over 50% of sporadic tumours have an alteration in

the *TP53* gene (www-p53.iarc.fr). Twenty-eight per cent of ovarian cancers contain *TP53* mutations (www.sanger.ac.uk/genetics/CGP/cosmic/). *TP53* is also the causal gene of Li-Fraumeni syndrome, which is characterised by an increased susceptibility to cancers. Mutations in *TP53* are predominantly found in the DNA binding domain of the gene, which disrupts the ability of the protein to bind DNA and activate transcription (Iwakuma *et al*. 2005).

### 1.8.3: Epithelial ovarian cancer and inheritance

Meta-analyses of case-control and cohort studies has demonstrated that an individual with an affected first degree relative has a 3% risk of developing ovarian cancer (Stratton *et al*. 1998). This value is greater than the risk for a woman in the general population developing ovarian cancer (1%). Since twins, both monozygotic and dizygotic, generally share the same environment *in utero* and after birth, twin studies enable the estimation of the overall contribution of inherited genes to the development of cancers. Monozygotic twins are genetically identical, and dizygotic twins share approximately 50% of their segregating genes.

Twins who are concordant for a cancer have a tumour of the same anatomical site. It can be said that genetics plays an important role in the development of cancer if the proportion of monozygotic twins concordant for a cancer is greater than that of dizygotic twins. A twin study published in 2000 compared the concordance of cancer in monozygotic and dizygotic twins. They found that genetic factors made a major contribution to susceptibility of cancers such as breast, stomach, lung, colorectal and prostate as well as ovarian cancer. From Lichtenstein's study, the heritability of ovarian cancer was estimated to be 22% (Lichtenstein *et al*. 2000).

Approximately 5-10% of ovarian cancer cases are inherited. Familial ovarian cancer is subdivided into three categories: (i) site specific ovarian cancer, (ii) breast and ovarian cancer syndrome, and (iii) hereditary non-polyposis colorectal cancer (HNPCC, also known as Lynch II syndrome) (Prat *et al.* 2005). Site-specific ovarian cancer and inherited breast and ovarian cancer syndrome are deemed to be part of the same disease syndrome spectrum because they are associated with germ-line *BRCA1* and *BRCA2* mutations (Prat *et al.* 2005).

Linkage analysis of breast and ovarian cancer families have shown that ovarian cancer is caused by *BRCA1* and *BRCA2* in the majority (> 90%) of breast and ovarian cancer syndrome families with more than 3 affected individuals. It has been demonstrated that mutations in *BRCA1* or *BRCA2* co-segregate with the disease within families. Mutations in the DNA mismatch repair genes, *MSH2*, *MLH1*, *PMS1*, *PMS2* and *MSH6*/*GTBP* inherited from HNPCC families account for approximately 10% of familial cases of ovarian cancer (Sharma *et al*. 2001).

### 1.8.4: High risk/high penetrance genes

Mutations in some genes cause a very high risk of developing a cancer. These genes are known as high risk susceptibility genes, and in cancer, most appear to have a dominant effect on the development of the cancer. The inheritance of a mutated form of the high risk gene results in a greater chance of developing the disease. Normally, the Mendelian dominant mode of inheritance means that the inheritance of one mutated copy of the causal gene is sufficient to cause the disease in the offspring. In hereditary cancer, an affected individual usually inherits a mutated

copy of the gene (such as *BRCA1* or *BRCA2*), this is known as the first hit. The second copy of the gene is lost by another mechanism, such as somatic mutation, loss of heterozygosity or methylation (the second hit) (Knudson 1971). These form the basis of the two-hit hypothesis, which was proposed as a possible explanation of the development of cancer.

Familial cancers appear at an earlier onset because the affected individuals already have a mutated gene (first hit), therefore it is assumed that it requires less time to acquire the second hit compared with sporadic cases, who need to attain both hits through somatic mutation. The dominant effects of genes on cancer development are demonstrated by the adenomatous polyposis coli (*APC*) gene and familial adenomatous polyposis (FAP) syndrome. FAP is characterised by the presence of hundreds to thousands of polyps in the colon or rectum before 40 years of age. FAP is caused by mutations in the *APC* gene. The children of FAP patients have a 50% chance of inheriting the mutated gene. Colorectal tumours from FAP patients show that in addition to the germline mutated copy of the *APC* gene, somatic mutation results in the inactivation of the normal gene copy (Fearnhead *et al*. 2002).

### 1.8.5: Ovarian cancer and high susceptibility genes

A gene which confers increased susceptibility to ovarian cancer alone has not been isolated. *BRCA1* and *BRCA2* are large genes which co-segregate with the majority of hereditary breast and ovarian cancer. *BRCA1* and *BRCA2* were both mapped through linkage analysis of families with breast and ovarian cancer syndrome. *BRCA1* is an 81.09 kb tumour suppressor gene which was mapped to chromosome 17q12-21 in 1994 (Miki *et al*. 1994). *BRCA2* (84.19kb) was mapped to chromosome

13q12-13 in 1995 and the gene was identified on chromosome 13q12.3 in 1996 (Wooster *et al*. 1995). *BRCA1* and *BRCA2* consist of 24 and 28 exons, respectively, and exon 11 from both genes constitute 60% of their coding DNA sequences (Kote-Jarai and Eeles 1999).

These two genes account for approximately 45% of epithelial ovarian cancer familial cases (Ramus *et al*. 2007). Most cases (approximately 90%) with greater than 3 first-degree relatives with ovarian cancer and breast cancer are due to *BRCA1* and *BRCA2* mutations (Pharoah and Ponder 2002). However, there are some large ovarian cancer families, which are not linked to *BRCA1* or *BRCA2* (Ramus *et al*. 2007). Some of the remaining high penetrance familial cases are linked to mutations in mismatch repair genes in hereditary non-polyposis colorectal cancer (HNPCC) cases (Lakhani and Flanagan 2002). The mutations and linkage analysis studies are reviewed in (Pharoah and Ponder 2002, Prat *et al*. 2005).

Although *BRCA1* and *BRCA2* germline mutations in breast cancer occur with equal frequency, *BRCA1* mutations are approximately four times more common than *BRCA2* mutations in ovarian cancer (Gayther *et al*. 1999). *BRCA1* mutation carriers from breast and ovarian cancer families have a greater than 40% lifetime risk of ovarian cancer, and *BRCA2* mutation carriers have a 10% risk of ovarian cancer. The fact that not all individuals with a *BRCA1* or *BRCA2* mutation develop ovarian or breast cancer suggests that the genes are not fully penetrant – a mutation does not correlate to the development of a malignancy in all mutation-carriers. The incomplete penetrance of *BRCA1* and *BRCA2* may be explained by the position of the mutation within the genes, modifying genes and environmental factors, which

affect the chances of a *BRCA1* and *BRCA2* mutation carrier developing ovarian cancer (Thompson and Easton 2002). It has been demonstrated that *BRCA1* and *BRCA2* mutation carriers on long term oral contraceptive pills have a reduced risk of ovarian cancer (Whittemore *et al*. 2004).

The diagnosis of cancer is at a younger age for mutation carriers when compared with non-carriers (Laplace-Marieze *et al*. 1999, Pharoah and Ponder 2002). There is also evidence suggesting that *BRCA1* mutation carriers are more likely to have serous adenocarcinoma histological subtype tumours than non-familial cases (Lakhani and Flanagan 2002, Rubin *et al*. 1996). Furthermore, there is inconclusive data for survival in *BRCA1* and *BRCA2* mutation carriers. Some studies have reported that *BRCA1* and *BRCA2* carriers with ovarian cancer have better long-term survival compared with non-carriers (Boyd *et al.* 2000; Chetrit *et al.* 2008). However, others have demonstrated survival advantage, which were not statistically significant between *BRCA1* and *BRCA2* mutation carriers and non-carriers (Pharoah *et al*. 1999, Ramus *et al*. 2001).

To conclude, these high risk susceptibility genes account for approximately 10% of all ovarian cancer cases (see Figure 1.4). This poses the important question – "is a proportion of the remaining ovarian cancer cases attributable to moderate or low penetrance genes?"

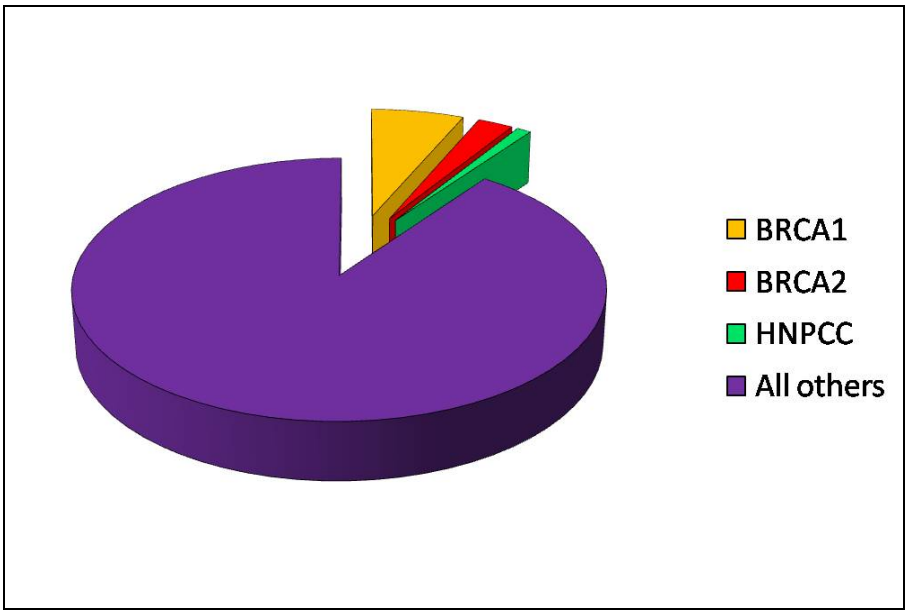



Figure 1.4: Contribution of high-risk susceptibility genes to epithelial ovarian cancer

HNPCC – genes associated with hereditary non-polyposis colon cancer

### 1.8.6: Moderate/low penetrance risk susceptibility

Statistical modelling using data from high-risk families and population-based ovarian cancer cases, have suggested that a dominant or recessive high susceptibility gene predisposing to ovarian cancer is unlikely (Antoniou *et al*. 2000). There was no significant difference between simulation of a hypothetical high risk gene with *BRCA1* and *BRCA2* and simulation without the hypothetical gene. The modelling also showed that common genes with low penetrance or rare alleles with higher risks were compatible with the observed data. However, the results from simulations of a model of relatively common alleles with moderate penetrance were inconsistent with the observed data. These suggest that some of the familial risks could be due to environmental, or modifying genetic factors (Antoniou *et al*. 2000).

### 1.8.7: Polygenic model of ovarian cancer

Linkage and segregation analysis of non-*BRCA1* and -*BRCA2* families and epidemiological modelling have suggested that ovarian cancer may be a polygenic disease. This polygenic theory is attractive because ovarian cancer, as with other complex diseases, is likely to be influenced by many genes, as well as environmental factors. The common variant: common disease hypothesis is correlated with the polygenic model (Risch, N. and Merikangas 1996; Chakravarti 1999). The hypothesis proposes that some genetic variants, with moderate effects, become common over time. These variants may predispose to common diseases and the combinations of the variants may affect differences in disease susceptibility (Pharoah *et al*. 2004).

In light that it is unlikely that there is another high-risk ovarian cancer susceptibility gene, and twin studies have suggested that genes are more important than shared environment in ovarian cancer development, it is feasible that polymorphisms of candidate genes may confer moderate- or low-penetrance susceptibility. This project aims to evaluate the risks of ovarian cancer associated with common genetic polymorphisms of candidate genes.

## 1.9: Linkage and case-control studies

Linkage analysis led to the discovery of mutations in *BRCA1* and *BRCA2* in breast and ovarian cancer syndrome families. However, linkage analysis and segregation analyses have not been successful in identifying other high-risk ovarian cancer susceptibility genes. Genetic susceptibility association studies involve the

comparison of the frequencies of candidate susceptibility variants in ovarian cancer cases with matched (by age, ethnicity, area of residence, sometimes socio-ecomonic status, parity, oral contraceptive use and other epidemiological factors depending on the number of study participants), unaffected controls to ascertain whether there are significant differences between cases and controls. Association studies can also be used to identify genetic factors which may influence response to treatment or overall survival from the disease. In survival association studies, comparisons are made between the frequencies of the genetic variables of individuals still alive, and those who have died, within a specified period of time.

As ovarian cancer has late onset and poor survival, there are insufficient numbers of older members of pedigrees to perform associations with families. Thus, case-controls association studies have greater statistical power than familial association studies to detect ovarian cancer susceptibility variants with moderate effects. Statistical power refers to the probability of rejecting a false null hypothesis. As statistical power increases, the likelihood of obtaining a false negative result (type II error) decreases, therefore increasing the chance of finding a true association. Type I error is the rejection of the null hypothesis due to chance findings. For example, the 5% significance level suggests that there is a 5 in 100 probability of obtaining a positive result by chance.

### 1.9.1: Single nucleotide polymorphisms

A single nucleotide polymorphism (SNP) is variation at a single base in a DNA sequence, which occurs with a frequency of $\geq 1\%$ in the population. SNPs with allele frequencies greater than 5% are called common polymorphisms. SNPs are the

most abundant polymorphisms in humans, with approximately 10 million variants in the human genome (Sobrino *et al*. 2005). The vast majority of SNPs are bi-allelic, which means there are two variants for the particular SNP, with one copy inherited from each parent (Doris 2002), see Figure 1.5.

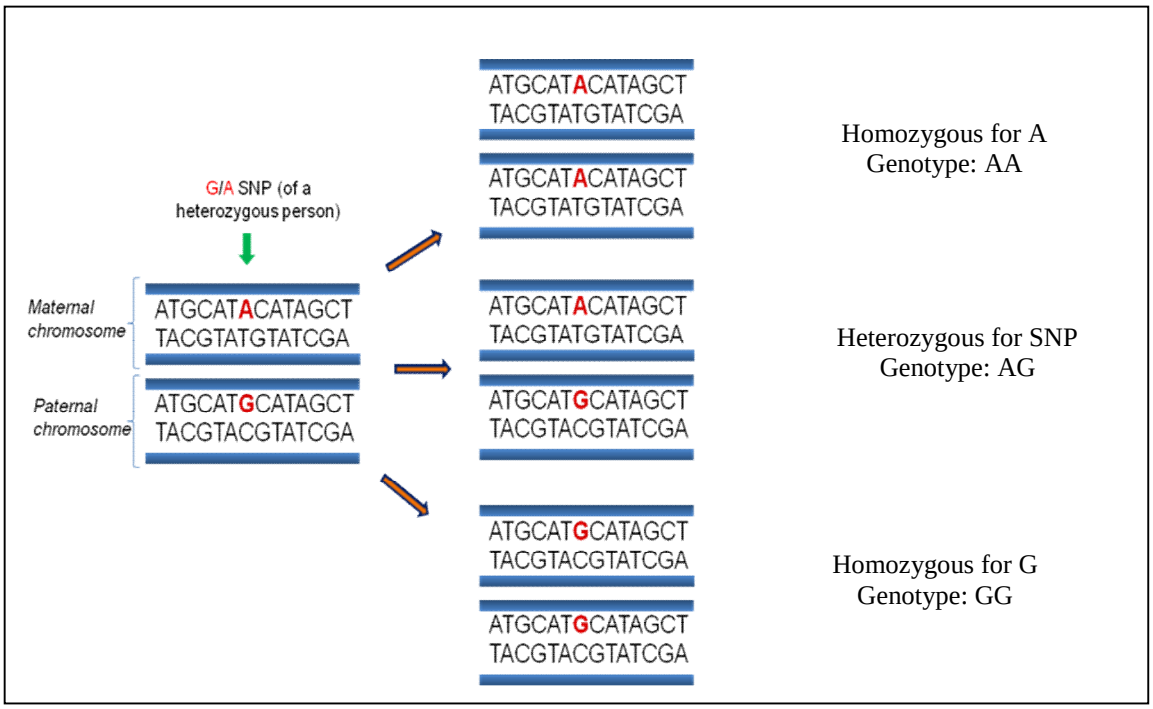



Figure 1.5: A single nucleotide polymorphism and it's possible genotypes

Between 3-5% of the human genome encode proteins, therefore the majority of SNPs are in non-coding regions of the genome, such as the introns of genes, between genes or in regions without open reading frames. SNPs within coding regions of genes are of particular interest because there is a greater chance that they may result in a variation of the biological function of the protein either by altering the folding of the protein, or the binding of the protein. A SNP can be coding or non-coding. Coding SNPs are located in the exons of genes and can be transcribed into amino

acids. Coding SNPs can be synonymous or non-synonymous. The alleles of a synonymous SNP result in the same amino acid being produced, due to the redundancy of amino acid codons. However, non-synonymous SNPs result in different amino acids being produced. These are known as missense SNPs. The amino acids translated from missense SNPs may have different charges, which may affect protein folding and binding, and subsequently the function of the protein. Another type of non-synonymous SNP exists, nonsense SNPs – these result in one of the alleles encoding a STOP codon, which may lead to a truncated protein being produced, if at all. However, most nonsense SNPs are mutations rather than polymorphisms.

SNPs in the untranslated regions (UTR) 5' or 3' of genes are also of interest. SNPs in the 5' UTR may contain sequences involved in promoting translation initiation. 5' UTRs often contain binding sites for proteins which may influence mRNA stability or translation. SNPs in 3' UTR may also be part of sequences for binding sites for proteins involved in mRNA stability or location of proteins within the cell.

Within a population, the frequencies of the alleles may be different; however the proportions of the genotypes add up to one in a population in Hardy-Weinberg equilibrium (HWE) (this is discussed below). The more frequent allele is known as the common or major allele, and the less frequent is the rare/minor allele. As shown in Figure 1.6, the allele frequencies of a SNP may be different within different populations. These differences in the allele frequencies between different

populations highlight the importance of ascertaining the ethnicity of study participants and stratifying populations during analysis.

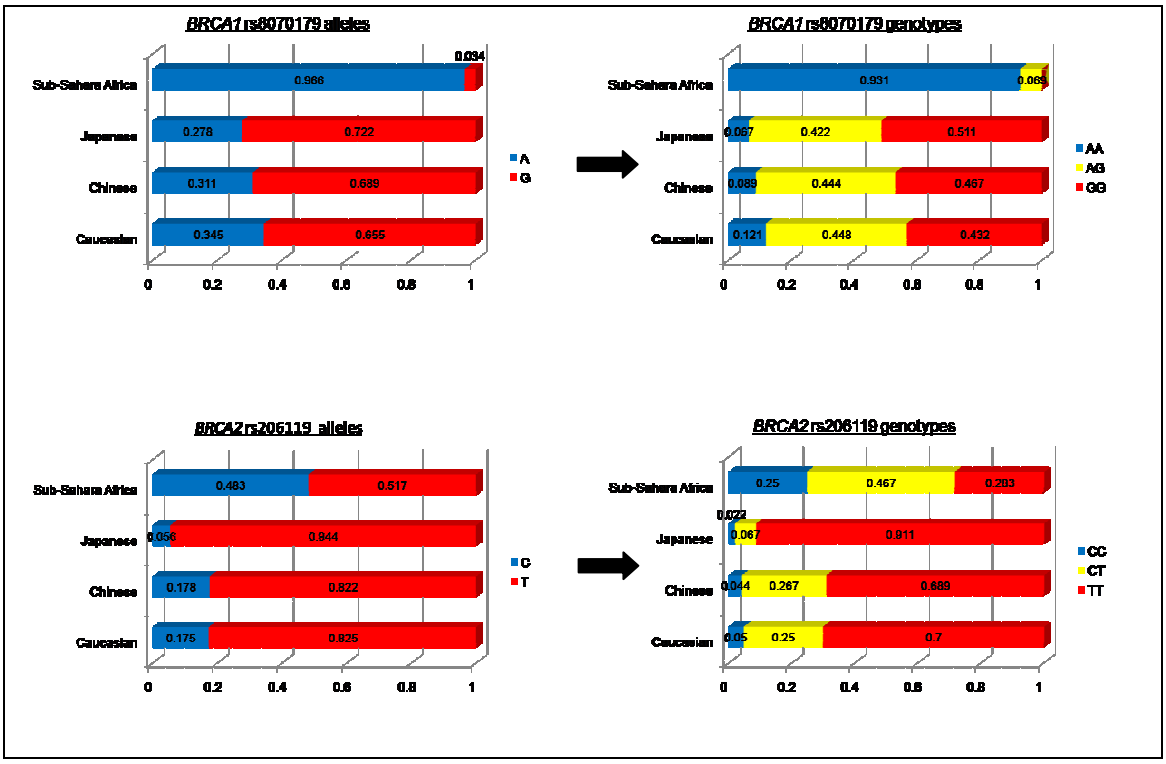



Figure 1.6: Different allele and genotype frequencies in different populations

## 1.9.2: Hardy-Weinberg equilibrium

The Hardy-Weinberg principle states that allele and genotype frequencies at an autosomal locus within an infinitely large population will reach equilibrium in a single generation where there is random mating, and there are no selective pressures, mutations, migration/emigration or random genetic drift or flow. The Hardy-Weinberg method has been demonstrated to be robust when estimating the allele frequencies of SNPs which are not "physiologically meaningful", such as the polymorphisms which encode the ABO blood groups, enzymes and DNA markers (Elston *et al.* 2002). The term "physiologically meaningful" refers to the fact that,

for example, blood performs the same function, regardless of the ABO group. It must be taken into consideration that despite the same function being performed, there are differences which prevent the transfer of blood from an individual with blood group A or B to a person whose blood group is O. The Hardy-Weinberg method is robust despite the fact that the chances of any population being able to meet all the conditions of the Hardy-Weinberg principle at any one time are very small. HWE is extensively used because the statistical power of detecting deviation from the HWE within large populations is also very small as a result of minute deviations (Chakraborty and Rao 1972; Elston *et al.* 2002).

If a bi-allelic SNP is considered, where the common allele is denoted by "*A*", the rare allele by "*a*", with allele frequencies *p* and *q*, respectively; when a population is in HWE, the frequency of "*A*" is *p*; the frequency of "a" is *q*, and $p+q=1$. The Punnett square below shows how genotypes can be derived from parents heterozygous at a SNP.

Punnett square for Hardy-Weinberg equilibrium (HWE)

| | | Female gametes | |
|---|---|---|---|
| | | A ($p$) | a ($q$) |
| Male gametes | A ($p$) | AA ($p^2$) | Aa ($pq$) |
| | a ($q$) | Aa ($pq$) | aa ($q^2$) |

Therefore, if a population is in equilibrium, the frequencies the genotypes would be:

AA (common homozygotes) = $p^2$;
Aa (rare homozygotes) = $q^2$;
Aa (heterozygotes) = $2pq$.

### 1.9.3: Linkage disequilibrium

Linkage disequilibrium (LD) is the non-random association of alleles at two or more loci. Neighbouring SNPs tend to be in LD – the SNPs are correlated with each other. The term "tag" is sometimes used to describe the correlation between SNPs with the same or similar minor allele frequencies (MAF). The correlation between neighbouring SNPs makes it unnecessary to genotype all the SNPs within a gene or chromosomal locus in order to test for association with disease. SNP tagging is described in detail in section 1.9.8 on page 60. There are two main ways of measuring LD between SNPs: disequilibrium coefficient ($r^2$), and normalised measure of Lewontin (D'). $r^2$ is a measure of the statistical correlation between two loci. For example, at two bi-allelic SNP loci on the same chromosome, if the common and rare alleles of the first locus are denoted as *A* and *a*, respectively, and the alleles of the second locus is *B* and *b*. When $r^2$ is used to calculate the LD between the alleles, the allele frequencies for *A*, *a*, *B* and *b*, are written as $\pi_A$, $\pi_a$, $\pi_B$ and $\pi_b$, respectively, and the frequencies of the haplotypes (the combinations of the alleles of the two loci) are $\pi_{AB}$, $\pi_{Ab}$, $\pi_{aB}$ and $\pi_{ab}$. Then

$$r^2 \equiv \frac{(\pi_{AB} - \pi_A \pi_B)^2}{\pi_A \pi_a \pi_B \pi_b}$$

(Pritchard and Przeworski 2001)

D' is derived from D, which measures the deviation of the frequencies of alleles or haplotypes from the equilibrium state. Therefore, D is calculated by subtraction the expected allele frequency from the observed frequency. For haplotype frequencies

$D=\pi_{AB}\pi_{ab}-\pi_{Ab}\pi_{aB}$

D is significantly greater than 0 when there is LD between alleles. D' is the absolute ratio of D compared with its minimum value, when D<0, or its maximum value, when D≥0. D' is calculated by:

$$D' = \frac{D}{\sqrt{\pi_{+A}\pi_{B+}}}$$

(Devlin and Risch 1995). When two loci are in complete LD, $r^2$=1 and D'=1, and both $r^2$ and D' tends towards 0 as the degree of correlation decreases; "0" corresponds to no LD/correlation. $r^2$ and D' can be calculated in terms of each other and allele frequencies, and $r^2$ can be calculated from D by the equation:

$$r^2 = \frac{D^2}{\pi_A \pi_a \pi_B \pi_b}$$

(Hedrick and Kumar 2001). $r^2$ is more commonly used in genetic association studies because it is inversely correlated to the sample size needed, given a fixed genetic effect. Therefore, the genotypes of a SNP can be predicted from a genotyped SNP with an $r^2 \geq 0.8$ correlation. An $r^2 \geq 0.8$ suggests a ≥80% correlation between the SNPs.

### 1.9.4: The International HapMap Project

The International HapMap Project records genetic variants, genotypes and sequences of 30 sets of (2 parents and an adult child) trios of Yoruba people from Ibadan, Nigeria; 30 trios of north and west European descent – from the Centre d'Etude du Polymorphisme Humain (CEPH) research in the United States of America; 45 unrelated individuals from Beijing, China; and 45 unrelated individuals from Tokyo, Japan. The results are freely available to researchers and they may be

used as a reference for genetic association studies. The project aims to identify and record all differences and similarities within the subjects in the project (www.hapmap.org).

## 1.10: Association study approaches

The vast majority of association studies in ovarian cancer have been conducted on candidate genes from pathways which have been implicated in neoplastic transformation, such as mismatch repair, cell cycle control and oestrogen pathways (Gayther *et al*. 2007, Goodman J. E. *et al*. 2000, Goodman M. T. *et al*. 2001b, Song *et al*. 2006a, Song *et al*. 2006b, Spurdle *et al*. 2000). Table 1.2 shows some significant genetic association studies in ovarian cancer.

### 1.10.1: Functional SNP, candidate gene approach

The first association studies in ovarian cancer were conducted on single nucleotide polymorphisms (SNPs), with variants which result in different amino acids. It was believed that association studies of these functional SNPs would be successful in identifying ovarian cancer susceptibility genes. However, the approach was not as successful as expected and resulted in the identification of an association with a variant (I31 allele of F31I) of *STK-15*, a putative oncogene (Dicioccio *et al*. 2004). This approach was also used in a study which found that the V108M polymorphism of the *catechol*-O-*methyltransferase* (*COMT*) gene was not associated with ovarian cancer risk (Goodman, J. E. *et al.* 2000).

Table 1.2: Published susceptibility association studies on ovarian cancer (positive results)

| Gene | SNP | No. cases | No. controls | OR (95% CI) | P-value | Study approach | Population | Reference |
|---|---|---|---|---|---|---|---|---|
| *P53* | A72Arg | 51 | 30 | 4.16 | 0.0058 | *a* | Greek | (Agorastos *et al*. 2004) |
| *CYP1A1* | Ile CYP1A1*3 Val | 117 | 202 | 6.08 (3.73–10.95) | $<1x10^{-3}$ | *a* | Turkey | (Aktas *et al*. 2002) |
| *BRCA2* | N372H | 1121 | 2643 | 1.36 (1.04–1.77) | 0.03 | *a* | UK, Australia | (Auranen *et al*. 2003) |
| *XRCC2* | R188H | 1600 | 4241 | 0.3 (0.1-0.9) | 0.003 | *a* | Caucasian (Denmark, UK, USA) | (Auranen *et al*. 2005) |
| *XRCC3* | rs1799796 | 1600 | 4241 | 0.08 (0.7–0.9) | 0.049 | *a* | Caucasian (Denmark, UK, USA) | (Auranen *et al*. 2005) |
| *GST* | GSTM1 null | 293 | 219 | 1.54 (1.06–2.14) | 0.025 | *a* | UK | (Baxter *et al*. 2001) |
| *PGR* | +331G/A | 973 | 802 | 0.46 (0.09-0.97) | - | *a* | White American, Australian | (Berchuck *et al*. 2004) |
| *STK15* | F31I | 1821 | 2467 | 1.17 (1.02-1.35) | 0.03 | *a* | Caucasian (Denmark, UK, USA) | (Dicioccio *et al*. 2004) |
| *MLH1* | G>A nt-93 | 899 | 931 | 1.5 (1.3–1.9) | $5x10^{-5}$ | *a* | Canadian mixed | (Harley *et al*. 2008) |
| *TGFBR1* | TGFBR1*6 A | 1155 | 983 | 1.53 (1.07-2.17) | 0.017 | *a* | Italy, Jamaica, UK, USA | (Kaklamani *et al*. 2003) |
| *PgR* | +331G/A (rs10895068 | 490 | 534 | 1.68 (1.09–2.59) | - | *a* | USA - mixed | (Risch, H. A. *et al*. 2006) |

| Gene | SNP | No. cases | No. controls | OR (95% CI) | P-value | Study approach | Population | Reference |
|---|---|---|---|---|---|---|---|---|
| *EPHX* | Tyr113His | 545 | 287 | 0.38 (0.17–0.87) | - | *a* | Australia | (Spurdle *et al*. 2001) |
| *PGR* | V660L | 987 | 1034 | 0.70 (0.57-0.85) | - | *a* | White USA | (Terry *et al*. 2005) |
| *FSHR* | Thr307Ala | 202 | 266 | 2.60 (1.56–4.34) | <0.0005 | *a* | China | (Yang *et al*. 2006) |
| *FSHR* | Asn680Ser | 202 | 266 | 2.89 (1.73–4.84) | <0.0005 | *a* | China | (Yang *et al*. 2006) |
| *BRCA1* | Q356R | 312 | 401 | 0.9 (0.5-1.4) | 0.64 | *a* | Caucasian, African American | (Wenham *et al*. 2003) |
| *CYP17* | A2 | 200 | 241 | 1.86 (1.26-2.75)* | 0.002 | *a/b* | Caucasian | (Garner *et al*. 2002) |
| *CYP1B1* | V432L | 129 | 144 | 3.8 (1.2-11.4) | 0.005 | *a/b* | White, Asian, Hawaiian | (Goodman, M. T. *et al.* 2001a) |
| *XRCC2* | R188H | 1600 | 4241 | 0.3 (0.1-0.9) | - | *b* | Caucasian (Denmark, UK, USA) | (Auranen *et al*. 2005) |
| *VDR* | rs7975232 | 72 | 148 | 2.8 (1.2–7.0) | 0.02 | *b* | USA Caucasian | (Lurie *et al*. 2007) |
| *VDR* | rs10735810 | 72 | 148 | 2.5 (1.3–4.8) | 0.04 | *b* | USA Caucasian | (Lurie *et al*. 2007) |
| *SOD2* | Val-9ala | 125 | 193 | 2.1 (1.1–4.0) | 0.04 | *b* | USA - mixed | (Olson *et al*. 2004) |
| *VDR* | rs11568820 | 94 | 173 | 0.5 (0.3–0.9) | 0.03 | *b* | USA Japanese | Lurie *et al*. 2007) |

| Gene | SNP | No. cases | No. controls | OR (95% CI) | P-value | Study approach | Population | Reference |
|---|---|---|---|---|---|---|---|---|
| *RB1* | rs2854344 | 1514 | 2415 | 0.73 (0.61-0.89) | 0.0009 | *b/c* | Caucasian (Denmark, UK, USA) | (Song *et al.* 2006b) |
| *RB1* | rs4151620 | 1514 | 2415 | 0.19 (0.07-0.53 | 0.00005 | *b/c* | Caucasian (Denmark, UK, USA) | (Song *et al.* 2006b) |
| *SHMT1* | rs9909104 | 829 | 941 | 1.2 (1.0–1.4) | 0.02 | *c* | USA Caucasian | (Kelemen *et al*. 2008) |
| *PMS2* | rs7797466 | 1531 | 2570 | 1.17 (1.03–1.33) | 0.013 | *c* | Caucasian (Denmark, UK, USA) | (Song *et al*. 2006a) |
| *CDKN1B* | rs2066827 | 4526 | 6913 | 0.93 (0.87-0.995) | 0.036 | *d* | Caucasian, African American, Asian, Hawaiian | (Gayther *et al*. 2007) |
| *CDKN2A* | rs3731257 | 4526 | 6913 | 0.91 (0.85-0.98) | 0.008 | *d* | Caucasian, African American, Asian, Hawaiian | (Gayther *et al*. 2007) |
| *PgR* | rs1042838 | 7614 | 651[†] | 1.17 (1.01–1.36) | 0.036 | *d* | USA – mixed, UK, Denmark | (Pearce *et al*. 2008) |
| *AURKA* | rs2273535 | 4624 | 8113 | 1.12 (1.01–1.24) | 0.03 | *d* | Caucasian – Denmark, UK, USA | (Ramus *et al.* 2008a) |
| *RB1* | rs2854344 | 4624 | 8113 | 0.87 (0.76–0.98) | 0.025 | *d* | Caucasian – Denmark, UK, USA | (Ramus *et al.* 2008a) |
| 9p22 | rs3814113 | 4487 | 7021 | 0.82 (0.79-0.86) | $5.1 \times 10^{-19}$ | *e* | Caucasian – Denmark, UK, USA, Australia | (Song *et al.* 2009b) |

-; P-value not given; *a*: functional SNP, candidate gene; *b*: functional SNP, candidate pathway; *c*: tSNPs, candidate pathway; *d*: consortium; *e*: genome-wide, consortium;

### 1.10.2: Functional SNP, candidate pathways approach

The functional SNP approach was then used on candidate genes from molecular pathways which were believed to be involved in ovarian cancer development. Auranen *et al* (2005) conducted an association study on genes (*BRCA1*, *NBS1*, *RAD51*, *RAD52*, *XRCC2* and *XBCC3*) involved in the DNA double strand break repair pathways. They found evidence for a decrease in ovarian cancer risk with the rare variants in *XRCC2* and *XRCC3* (R188H and rs1799796, respectively) (Auranen *et al*. 2005). Associations between ovarian cancer risk and genes involved in steroid hormone metabolism and catecholestrogen formation have also been investigated. Individuals who carried the leucine allele for the V432L polymorphism in *CYP1B1* had an increased risk of ovarian cancer (Goodman, M. T. *et al.* 2001b).

### 1.10.3: Tagging SNPs, candidate pathways approach

The next SNP association study approach involved the use of tagging SNPs (tSNPs) from candidate genes within a pathway. The tagging SNP approach takes advantage of the LD between neighbouring SNPs. SNPs in complete LD ($r^2$=1) are said to tag each other. Therefore, the genotype of a SNP which is tagged by another can be determined from the genotype of the tagging SNP (if they have the same minor allele frequencies [MAF]). SNPs which are in strong LD are inherited together, but their polymorphisms may have different MAF. The pairwise correlation coeffiecient ($r^2_p$) is the best way to measure how well a SNP tags another SNP. $r^2_p$ takes into account the loss of power incurred by using the tSNP as a marker, rather than as the causal SNP. $r^2_s$ is a measure of how well a haplotype of tSNPs tags a single SNP that is inefficiently tagged by single SNPs (Song *et al*. 2006a). The SNP which is

genotyped is known as the tagging SNP (tSNP). Figure 1.7 shows the principle of tagging SNPs.

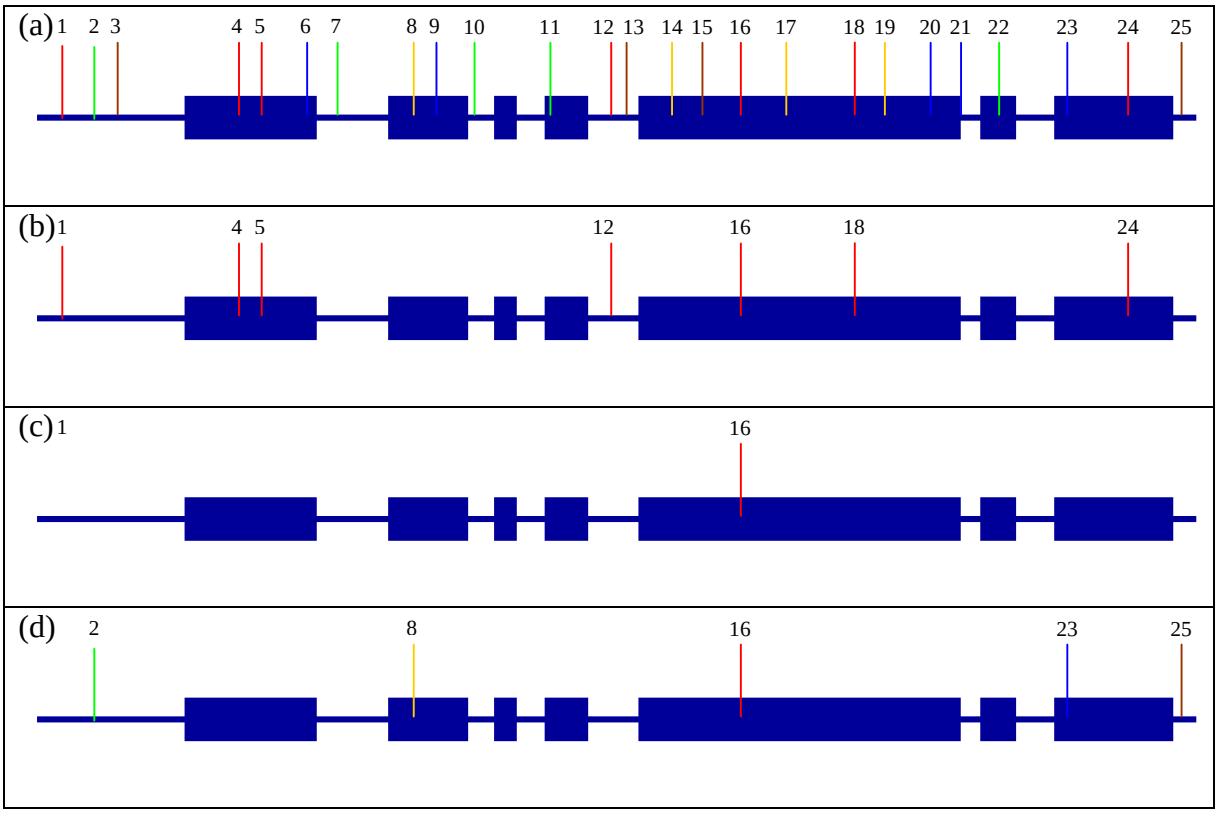



Figure 1.7: Principles of tagging SNPs

(a) This hypothetical gene contains 25 SNPs. (b) SNPs of the same colour are correlated, thus tag each other, (c) therefore only 1 SNP needs to be genotyped to gain information about all the SNPs it tags. (d) Therefore only 5 SNPs in the gene need to be genotyped in order to acquire information about all 25 SNPs in the gene.

The correlation between two SNPs is measured by $r^2$. Normally an $r^2$ of 0.8 is chosen, which means there is at least 80% correlation between the tSNP and all the SNPs it tags. This approach ensures that not all 10 million SNPs need to be genotyped in order to ascertain the genotypes of each SNP. Only approximately 500,000 tSNPs need to be genotyped to gain information about the remaining 9.5 million variants. The tSNP may be a marker of the causative SNP. This approach

was used to evaluate associations between *MLH1*, *MLH3*, *MSH2*, *MSH6*, *PMS1* and *PMS2* from the mismatch repair pathway and ovarian cancer. The rare alleles of *MSH6* (rs3136245) and *MSH3* (rs6151662) were associated with a decrease in ovarian cancer risk, and *PMS2* (rs7797466) was associated with an increase in ovarian cancer risk (Song *et al*. 2006a). This approach has also been used to identify a positive association between a variant in *SHMT1*, a member of the one-carbon transfer pathway, and an increase in ovarian cancer susceptibility (Kelemen *et al*. 2008).

### 1.10.4: Consortium approach

The consortia approach is currently the most popular strategy for genetic association studies in ovarian cancer research. This approach allows staged genotyping designs in a multi-centre collaboration. The initial stage is the genotyping of the tSNPs from the candidate genes by a group or a small number of groups within a consortium. Positive associations are genotyped by the remaining groups within the consortium to validate or refute the findings of the initial stage of research. This approach gives more statistical power to a study, reducing type I error.

The Ovarian Cancer Associations Consortium (OCAC) is a multinational consortium which co-ordinates ovarian cancer research. The group ensures that research is not duplicated unnecessarily and allows easy sharing of data. At present, the OCAC consists of 20 groups, which have published a two-stage study on candidate genes from the cell cycle control pathway (Gayther *et al*. 2007). In the first stage of the study, 88 tSNPs in 13 genes were genotyped in three study populations (MALOVA, SEARCH and GEOCS [previously FROC], also known as Stanford) consisting of

approximately 1,500 cases and 2,500 controls. There were 13 statistically significant associations found between the variants and ovarian cancer. Approximately 50% of the significant SNPs conferred an increased risk of ovarian cancer. Stage 2 of the study involved the genotyping of the five most significant tSNPs from stage 1 on approximately 2,000 cases and 3,200 controls by the remaining groups within the OCAC. The five most significant tSNPs from stage 1 were not significant with the stage 2 samples alone. There were only significant associations between ovarian cancer risk and the rare variants from *CDKN2A* (rs3731257) and *CDKN1B* (rs2066827) SNPs when the data from stages 1 and 2 were pooled (Gayther *et al*. 2007).

The consortium approach has also been used in other studies: seven best candidates from publication (Ramus *et al*. 2008), progesterone receptor (Pearce *et al*. 2008). These studies have highlighted the importance of validating results in larger studies. The lack of replication of statistically significant associations independently in later stages/studies suggests that the initial associations may have been chance findings. The larger sample sizes from consortia also allow stratification of samples by histology, race, grade, stage etc. for further analysis. Another advantage of this approach is that negative results can also be confirmed with additional samples.

### 1.10.5: Genome-wide, consortium approach

Genome-wide association studies have proved to be a success in the identification of genes which may be associated with ovarian, breast, colon and prostate cancer risks (Easton *et al*. 2007, Song *et al*. 2009, Thomas *et al*. 2008, Yeager *et al*. 2007, Zanke *et al*. 2007). In genome-wide association studies, the tagging SNP approach is used

to genotype evenly distributed SNPs within the genome and evaluate the effect of the genotyped tSNPs on disease risk. These studies tend to use a staged-design, whereby highly significant associations are further assessed in succeeding stages with additional studies. The data from the different stages are combined to increase the statistical power of detecting associations.

Genome-wide association studies have also been used in the investigation of susceptibility genes in other complex disorders such as diabetes and heart disease (Cupples *et al*. 2007, Sladek *et al*. 2007). Some of the results from these studies have been highly significant. In the breast cancer genome-wide association study, an association was found between a variant in *FGFR2* (a fibroblast growth factor receptor) and an increase in breast cancer risk, $P=2x10^{-76}$ (Easton *et al*. 2007). Genome-wide association studies involve the genotyping of thousands of SNPs throughout the human genome and performing association analyses on the SNPs genotyped.

Recently, the OCAC has published results from genome-wide association studies of ovarian cancer. The study also used the consortia approach in a 3-stage design. In the first stage 507,094 SNPs were genotyped in 1,817 invasive epithelial ovarian cancer cases and 2,353 unaffected controls. The 22,790 top ranked significantly associated polymorphisms were genotyped in an additional 4,274 ovarian cancer cases and 4,809 controls. Moreover, stage 3 involved genotyping the most significant SNPs from stage 2 in a further 2,670 cases and 4,668 controls. The data from all three stages were combined to increase the power of the study. The rare allele of the most significant SNP, rs3814113, was associated with a reduced risk of

ovarian cancer (combined stage 1-3 OR=0.82 (0.79-0.86), $P_{trend} = 5.1 \times 10^{-19}$) (Song *et al.* 2009c).

## 1.11: Survival analysis

There is substantial evidence showing that cancer patients have different responses to the same treatment (McGuire *et al.* 1996; Piccart *et al.* 2000), and chemotherapy resistance remains a very important issue; chemotherapy resistance is reviewed by (Lage and Denkert 2007). There have also been reports suggesting that chemotherapy resistance may be affected by germline genetic variation (Marsh 2005, Villafranca *et al*. 2001). These findings indicate that it is feasible that genetic polymorphisms may influence a patient's response to treatment, and thus survival from the disease. The effects may be attributed to polymorphisms in genes encoding drug targets, drug-metabolising enzymes and/or drug transporters (Pinto *et al*. 2005). Molecular markers such as *ERBB2* and *TYMS* have been identified for predicting overall survival after diagnosis of cancer ERBB2-positive breast cancer, and serous ovarian cancer (Hsu *et al.* 2004; Piccart-Gebhart *et al.* 2005; Romond *et al.* 2005). Although there have been improvements in the response to adjuvant chemotherapy, the majority of ovarian cancer patients go into remission, developing recurrent disease. Some of these recurrent cases are drug-resistant (Bristow *et al*. 2002).

Differences in survival of ovarian cancer patients have been found between *BRCA1* and *BRCA2* mutation carriers and non-mutation carriers (Chetrit *et al*. 2008, Tan *et al*. 2008). There are also publications, including findings from this project, on the effects of common genetic polymorphisms from candidate genes in mismatch repair

and cell cycle control pathways; and combinations of variants in the vascular endothelial growth factor (*EGF*) gene (Hefler *et al*. 2007, Mann *et al*. 2008, Nagle *et al*. 2007, Quaye *et al*. 2009, Quaye *et al*. 2008, Song *et al*. 2008). All of these results suggest that it is feasible that common genetic variants may affect survival from ovarian cancer.

As well as establishing associations between genetic polymorphisms and susceptibility to developing a disease, association studies can also be used to identify genetic variants that may influence survival from the disease. To do this, follow-up data is required to ascertain the vital statistics of the patients recruited into a study over a period of time (usually over 10 years). Although participating patients may die from other causes, such as heart failure, stroke etc, many are likely to die from ovarian cancer. It is also likely that the sufferers would eventually die from the disease, particularly because many cases are diagnosed in the advanced stages of disease.

In survival analyses, the frequencies of genotypes are compared between the patients which have, unfortunately, died and those still alive over a time period. Variants of SNPs are said to be associated with survival if a statistically significant difference is observed between the frequencies of the genotypes/alleles within the groups of survivors and those who die over the time period. Survival is measured by the hazard ratio (HR), which essentially, is a measure of the risk of death, based on the individual's genotype. Survival analyses using genetic polymorphisms as variables have the potential of identifying genotypes which may predict a patient's survival

over a period of time. This has the potential of becoming a prognostic tool and may also be used for identifying suitable individuals for targeted therapy.

Survival analysis may also be used to establish response to therapy or overall clinical outcome. The former could potentially be used for targeted treatment, and the latter for prognostic purposes. There are studies which have investigated and, in some instances, identified associations between SNPs and response to treatment, progression-free survival and overall clinical outcome.

Associations have been found between variants of genes such as *ABCB1*, *ERCC1* and *IL8* and response to treatment. *ABCB1* is a transporter protein, which is involved in multi-drug resistance. Associations have been found between variants of *ABCB1* in the tumour DNA of ovarian cancer cases and response to paclitaxel and carboplatin (Green *et al*. 2008). Associations between progression-free survival and ovarian cancer patients and polymorphisms of *ABCB1* in germline DNA have also been reported (Johnatty *et al*. 2008). Similar associations have been observed between *ECCR1* and response to platinum-based treatment (with tumour and germline DNA) and progression-free survival (Krivak *et al*. 2008). *ECCR1* is a component of the nucleotide excision repair pathway. The gene is involved in the repair of DNA lesions, such as those caused by ultraviolet light and electrophilic compounds. Cisplatin, a platinum-based chemotherapy agent which is used to treat many different types of cancer, including ovarian cancer, is an electrophilic compound. The drug results in the cross-linking of DNA, which consequently triggers the apoptosis pathway.

Although the findings of these publications are of interest, the results should be treated with caution since many of the findings are based on small numbers of samples (<200) (Green *et al.* 2006; Kang *et al.* 2006; Saldivar *et al.* 2007; Green *et al.* 2008; Schultheis *et al.* 2008; Steffensen *et al.* 2008). Some of the publications reporting significant findings between a common polymorphism and response to treatment or survival are listed in Table 1.3 (page 69).

Survival association studies have also been conducted on the effect of common polymorphism on overall survival from ovarian cancer (Dhar *et al.* 1999; Spurdle *et al.* 2001; Hefler *et al.* 2003; Hogdall *et al.* 2003; Li *et al.* 2005; Pinto *et al.* 2005; Beeghly *et al.* 2006; Gadducci *et al.* 2006; Green *et al.* 2006; Higashi *et al.* 2006; Kang *et al.* 2006; Obata *et al.* 2006; Santos *et al.* 2006; Six *et al.* 2006; Hefler *et al.* 2007; Nagle *et al.* 2007a; Nagle *et al.* 2007b; Mann *et al.* 2008; Song *et al.* 2008), with some statistically significant results (Dhar *et al.* 1999; Hefler *et al.* 2003; Hogdall *et al.* 2003; Li *et al.* 2005; Pinto *et al.* 2005; Beeghly *et al.* 2006; Green *et al.* 2006; Higashi *et al.* 2006; Obata *et al.* 2006; Santos *et al.* 2006; Six *et al.* 2006; Nagle *et al.* 2007a; Nagle *et al.* 2007b; Mann *et al.* 2008; Song *et al.* 2008). Some of these are shown in Table 1.3.

Table 1.3: Response to treatment and clinical outcome publications (significant SNPs)

| Gene | SNP | No. cases | HR/response rates | P-value | Treatment/ | Population | Reference/ source of DNA |
|---|---|---|---|---|---|---|---|
| **Response to treatment** | | | | | | | |
| *IL-8* | T251A | 53 | AA* (19%); AT* (0%); vs TT* (50%) | 0.006 | Cyclophosphamide & bevacizumab | USA: 45 Caucasian, 8 other | (Schultheis *et al*. 2008) |
| *ERCC1* | Codon 118 SNP | 159 | TT* (44%), CT* (41%), CC* (15%) | 0.045 | Platinum-based | Danish | (Steffensen *et al*. 2008) |
| *ERCC1* | Asn118Asn | 60 | OR=0.17 (0.04-0.74) | 0.018 | Platinum-taxane | Korean | (Kang *et al*. 2006)[§] |
| **Progression-free survival** | | | | | | | |
| *CXCR2* | C+785T | 53 | (CC*, CT*) – 7.4months vs (TT*)-3.7 months | 0.026 | Cyclophosphamide & bevacizumab | USA: 45 Caucasian, 8 other | (Schultheis *et al*. 2008) |
| *ABCB1* | 2677 G>T/A | 914 | 0.7 (0.46-1.04) | 0.039 | Paclitaxel & carboplatin, docetaxel | Australia | (Johnatty *et al*. 2008) |
| *ERCC1* | C8092A | 233 | 1.44 (1.06-1.94) | 0.018 | Paclitaxel & cisplatin | USA: 214 Caucasian; 19 other | (Krivak *et al*. 2008) |
| *XP* | *XPG* | 146 | (GG*) 8.3 months vs 24.6 months | 0.006 | Carboplatin | USA: 135 Caucasian, 21 other | (Saldivar *et al*. 2007) |
| **Overall survival** | | | | | | | |
| *ERCC1* | C8092A | 233 | 1.5 (1.07-2.09) | 0.018 | Cisplatin & paclitaxel | USA: 214 Caucasian; 19 other | (Krivak *et al*. 2008) |
| *VEFG* | Haplotype of _634C/C, _1154G/G, _2578C/C) | 563 | 2.1 (1.1-3.9) | 0.02 | Platinum-based | Austria, Germany | (Hefler *et al*. 2007) |
| *TP53* | codon 72 | 114 | HR>1 | 0.011 | Cisplatinum & paclitaxel | Portugal | (Santos *et al*. 2006) |
| *PMS2* | rs2228006 | 1473 | 0.84 (0.71–0.99) | 0.04[±] | Unknown | Caucasian UK, USA, Denmark | (Mann *et al*. 2008) |

| Gene | SNP | No. cases | HR/response rates | P-value | Treatment/ | Population | Reference/ source of DNA |
|---|---|---|---|---|---|---|---|
| *CCND2* | rs3217933;<br>rs3217901;<br>rs3217862 | 1,488<br>1,489<br>1,480 | 1.16 (1.03-1.31)<br>1.14 (1.02-1.27)<br>0.85 (0.73-1.00) | 0.02<br>0.024<br>0.043 | Unknown | Caucasian UK, USA, Denmark | (Song *et al*. 2008) |
| *CCNE1* | rs3218038 | 1,489 | 1.39 (1.04-1.85) | 0.033 | Unknown | Caucasian UK, USA, Denmark | (Song *et al*. 2008) |
| *CYP17* | 5' UTR C allele | 454 | 1.30 (1.02– 1.68) | 0.04 | Platinum based | Australian | (Nagle *et al*. 2007a) |
| *GSTP1* | Ile105Val | 448 | 0.77 (0.61–0.99) | 0.04 | Platinum based | Australian | (Nagle *et al.* 2007a) |
| *VDR* | FokI | 101 | 0.18 (0.005-0.61) | 0.006 | Paclitaxel & carboplatin, | Japan | (Tamez *et al*. 2009)[§] |

HR – hazard ratio; mo – months; § - based on tumour DNA; *Genotype; [±] No longer significant after adjusting for prognostic factors, however the effect became more pronounced.

## 1.12: The admixture maximum likelihood test

To date, there has been limited success in identifying germline variants associated with ovarian cancer predisposition. Many of the statistically significant associations are based on relatively small numbers of samples, where the statistical power to detect true positives is reduced. Furthermore, very few of the published results are corrected for multiple testing. One possible reason for the lack of multiple testing correction is that there is a lack of agreement on the most suitable test to use, because of the correlation between many of the polymorphisms evaluated within a project. This increasingly important issue has resulted in much discussion and investigation in the most appropriate method for assessing and correcting for this "experiment-wise" type I error.

The need for correction for experiment-wise type I error has led to a proposal of a global null hypothesis of no associations between any of the genetic variants from a project, and an alternative hypothesis that there are true positive significant associations between the SNPs and disease risk. A number of methods have been proposed with the aim of testing whether the null hypothesis can be rejected. The simple Bonferroni correction for multiple testing only performed best when there were only three statistically significant SNPs or 5% of the total number of SNPs tested, whichever is smaller (Pharoah *et al*. 2007, Tyrer *et al*. 2006).

Some of the proposed methods for testing the global significance of association studies include those described in (Hoh *et al.* 2001; Schaid *et al.* 2005), as well as the improved Bonferroni procedure (Simes 1986), truncated product (Zaykin *et al*. 2002), ranked truncated product of P-values (Dudbridge and Koeleman 2003). The

admixture likelihood (AML) test is a method that was created in order to assess whether there are statistically significant differences between the proportion of significant SNPs from a group/selection genotyped, and that which would be expected by chance (Tyrer *et al*. 2006).

The AML method has been tested against many of the tests currently available for controlling for multiple testing over a variety of scenarios for the alternative hypothesis, and it was found to have the same or improved statistical power than all the other methods tested (rank truncated product, unrestricted maximum likelihood, restricted space maximum likelihood, most significant SNP, Global $\chi^2$, Best subset $\chi^2$) ((Tyrer *et al.* 2006), (Pharoah *et al.* 2007)). The AML test has already been used on genotyping data from breast cancer association studies. In the study, the 710 common polymorphisms of 117 candidate genes were evaluated with AML to establish the global association between the variants and susceptibility to breast cancer (Pharoah *et al*. 2007). The test found that although the effects of individual SNPs are likely to be small, there were some variants which are associated with risk of breast cancer (Pharoah *et al*. 2007). The AML method will be used to evaluate whether a statistically significant proportion of SNPs were found to be associated with ovarian cancer risk from genotyping data spanning the past few years, and the effect size of these associations.

## 1.13: DNA amplification and genotyping platforms

Along with the evolution of the approaches used in genetic association studies of ovarian cancer, there have been developments in genotyping platforms. Genotyping platforms are used to ascertain the genotype of an individual. Although TaqMan®

and other single SNP genotyping platforms are still popular with research groups, there is an increasing need for multiplex genotyping platforms to be used. Multiplex platforms enable the genotyping of more than one SNP in a single reaction. The multiplex levels currently available differ widely from 12-plex (up to 12 SNPs genotyped in a single reaction) up to 96-plexes and more. The advent of the chip genotyping technology also allows thousands of SNPs to be genotyped in a single reaction. However, chip genotyping technology is only ever likely to be used for genome-wide association studies or evaluation of whole chromosomes due to the number of SNPs which can be analysed from a single reaction. The SNP multiplex genotyping technique offers the potential of reducing the time, amount of reagents and money spent on genotyping, and in some instances, the quantity of DNA used.

### 1.13.1: Whole genome amplification

The number of SNPs genotyped has increased exponentially as new approaches are designed. This has highlighted the importance of addressing the issue of limited amount of DNA from study individuals and the increasing number of SNPs from candidate genes which need to be genotyped. Whole genome amplification of DNA samples and SNP multiplex genotyping platforms are possible solutions for these problems. Whole genome amplification (WGA) methods are used to replicate the genome of an individual by varying magnitudes, depending on the method used. The WGA technique offers the potential of producing limitless quantities of DNA from research participants – if the re-amplification claims of some WGA products are to be believed. However there have been conflicting reports of the accuracy of the replication of some WGA methods.

Many research groups have reported complete, or near complete (>99%) concordance between non-amplified genomic material and the corresponding, whole genome-amplified DNA (Jasmine *et al*. 2008, Pan *et al*. 2008, Sorensen *et al*. 2007). However, others have found discordances between the non-amplified and amplified DNA (Pinard *et al*. 2006, Talseth-Palmer *et al*. 2008). These discordances have predominantly been a result of preferential amplification of some alleles at heterozygous loci. The fidelity of the replication of the DNA needs to be assessed due to the small effects expected in low-moderate risk models. There are two major types of WGA techniques, polymerase chain reaction (PCR)-based, and multiple strand displacement.

#### 1.13.1.1. PCR-based whole genome amplification

PCR-based WGA involves the amplification of the genome, using the PCR process, with primers which will result in the amplification of the whole genome, rather than small regions. There are several PCR-based WGA methods commercially available; Primer Extension Preamplification (PEP, (Zhang *et al*. 1992)), GenomePlex (Sigma-Aldrich®) and Degenerate Oligonucleotide PCR (DOP-PCR, (Telenius *et al*. 1992)) are such methods. PEP uses 15-mer random primers and the Taq polymerase, at low annealing temperature (to ensure low stringency binding to genomic sites). DOP-PCR is fairly similar to PEP, however there are some essential differences. Semi-degenerate oligonucleotide primers (for example ACG TGC GAG NNN NNN NNN GCT CAT) and a higher PCR annealing temperature is utilised in the DOP-PCR process. The Taq polymerase is also used in DOP-PCR. Taq is known to produce short fragments of amplified material (approximately 3 kilobases [kb]), which is suitable for SNP genotyping studies, however, not for all DNA analysis protocols.

GenomePlex® is also a PCR-based WGA method, however, the genomic DNA is converted to an OmniPlex® Library. The OmniPlex Library consists of fragmented DNA, whose flanking regions have been converted to PCR-amplifiable units. The library is amplified with universal primers. The method is said to generate 5-10ug of amplified DNA from nanogram quantities of template DNA (Sigma-Aldrich). The GenomePlex process is illustrated in Figure 1.8.

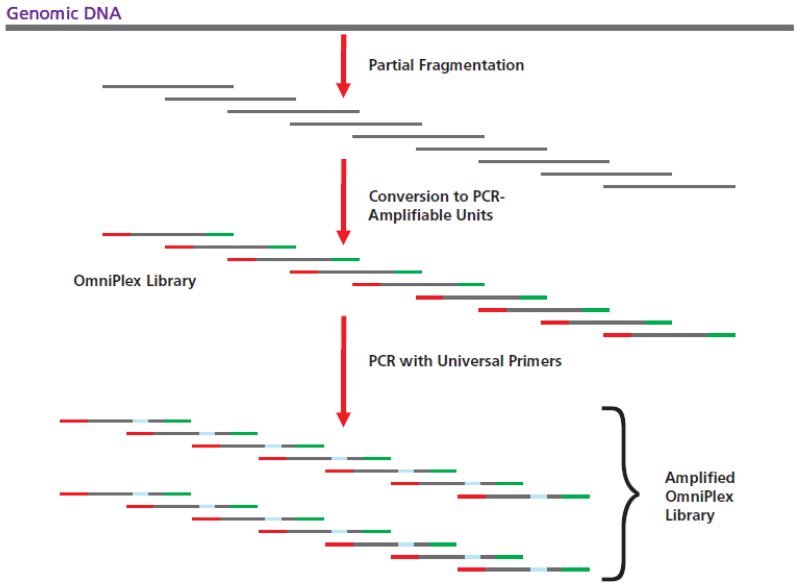



Figure 1.8: Schematic diagram of whole genome amplification with GenomePlex

### 1.13.1.2. Multiple displacement amplification

Whole genome amplification methods such as GenomiPhi™ (GE Healthcare, UK) and REPLI-g™ (Qiagen, UK) come under the multiple displacement amplification category of WGA. Both methods are based on the process illustrated in Figure 1.9.

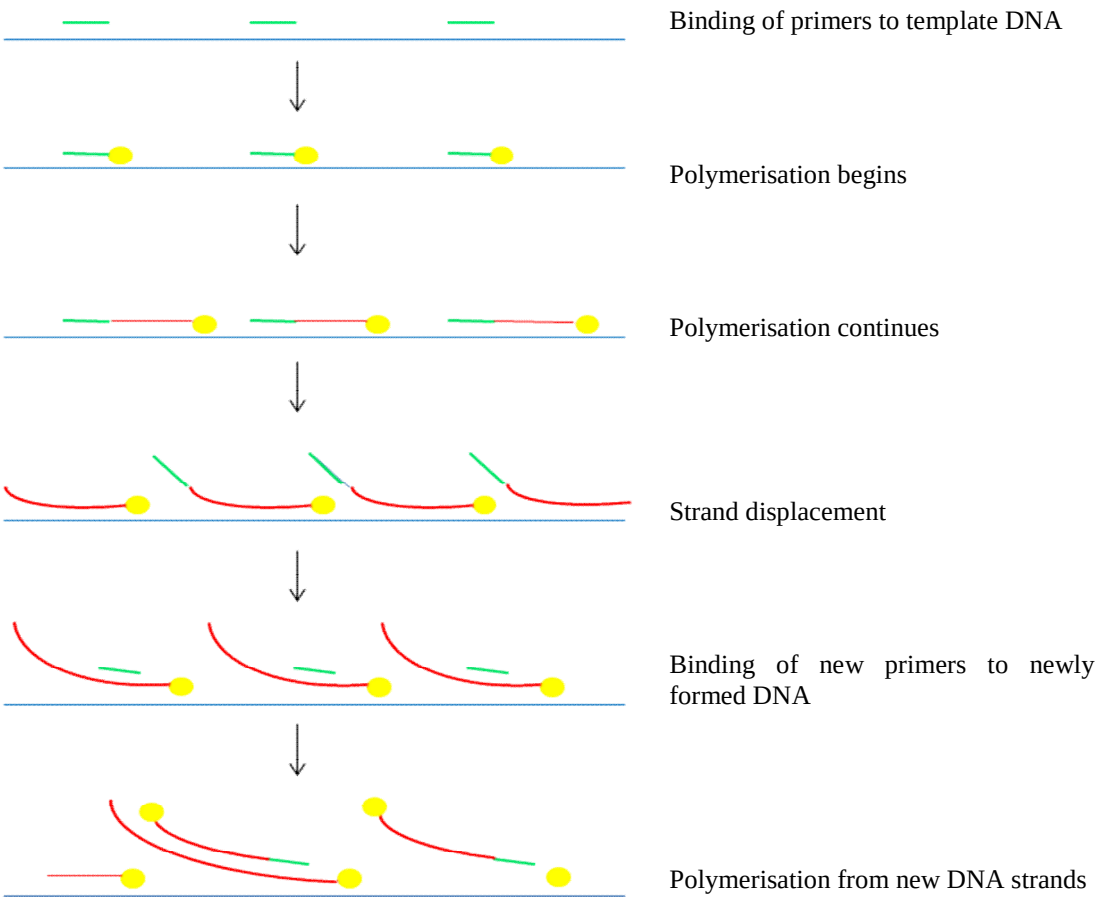



Figure 1.9: Schematic diagram of multiple strand displacement

GenomiPhi and REPLI-g use random hexamers and the bacteriophage Phi29 (φ29) DNA polymerase, which has 3' to 5' exonuclease proofreading activity. The φ29 polymerase does not detach from the template during the amplification process, and is thus, capable of producing amplified DNA that is up to 100kb in length. The major differences between GenomiPhi and REPLI-g are that the former uses heat to denature the template DNA, and the latter uses alkaline denaturation. The methods also differ in the quantities of amplified material produced. REPLI-g, apparently

generates up to 45μg of amplified DNA material, however GenomiPhi generates between 4-7μg of product.

### 1.13.2: SNP multiplex genotyping platforms

SNP multiplex genotyping methods allow the use of relatively low concentrations of DNA for the genotyping of more than one polymorphism, usually greater than 12 SNPs, in single reactions. The use of SNP multiplex genotyping platforms should drastically reduce the time required for laboratory work and the amount of DNA used for the number of SNPs per reaction. There are various ways in which the genotypes of multiple polymorphisms can be ascertained from a single reaction. These include fluorescence, mass and micro-arrays.

There are increasing numbers of multiplex genotyping platforms, such as SNPstream (microarray), SNPlex (fluorescence and mass, see Figure 1.10), OpenArray (microarray), iPLEX (mass), Illumina GoldenPath (microarray) and Fluidigm (microarray). Figure 1.10 shows the binding of a PCR product of a fluorescently tagged allele and a mass modifier which will enable the distinction of different SNPs in a SNPlex reaction. The ZipCode sequence ensures the binding of the PCR product of interest to the complementary ZipChute sequence on the hybridisation plate before the genotype is determined. The performances of the multiplex genotyping platforms need to be assessed to determine the suitability of the available DNA. Therefore, the performance of DNA amplified with four WGA methods: Genomeplex, GenomiPhi, primer extension PCR (PEP) and REPLI-g will be investigated on TaqMan and SNP multiplex genotyping platforms (iPLEX, OpenArray and SNPlex).

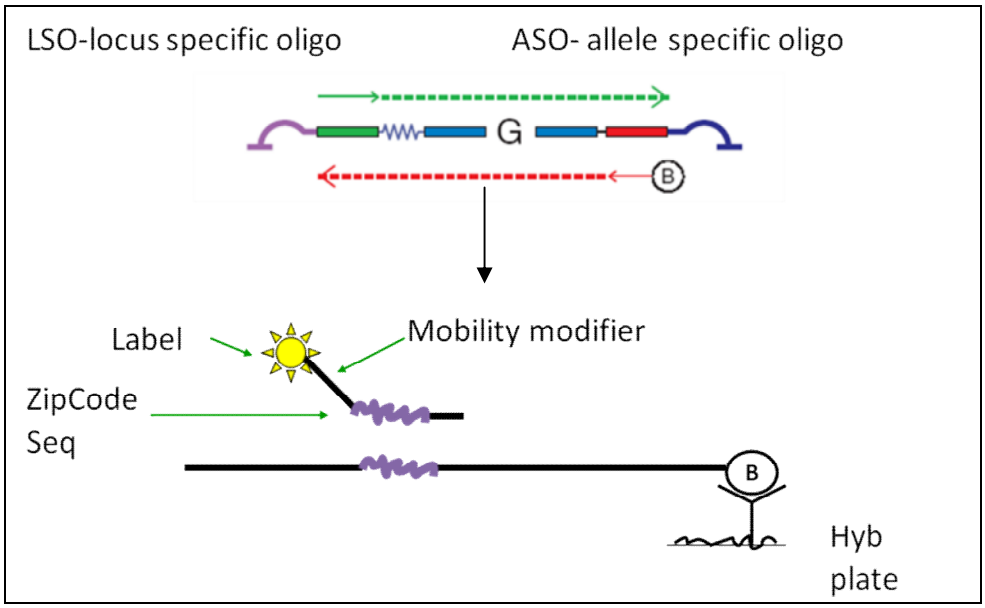



Figure 1.10: The binding of biotinylated amplicons to streptavidin-coated SNPlex hybridisation plate

Probes and linkers are linked together by phosphorylation. Blue- genome equivalent regions, red- universal reverse priming site, green-universal PCR priming site. After this step, the unbound probes, and bottom parts of linkers are enzymatically digested. The ligated probes and linkers are PCR amplified with biotin tagged universal primers. The products from this are denatured, the supernatant, containing the linker-probe, are removed. Leaving the biotinylated amplicons to bind with streptavidin-coated plates.

## 1.14: Project aims

The aims of this project are:

1. To determine if there is an effect of common variants and haplotypes of candidate oncogenes on the risk of invasive epithelial ovarian cancer.

2. To determine if there is an effect of common variants and haplotypes of functional candidate genes (associated with neoplastic suppression of ovarian cancer cell lines) on predisposition to ovarian cancer.

3. To use the admixture maximum likelihood test to assess if a significant number of associations have been found from ovarian cancer association studies.

4. To evaluate the effect of tSNPs and haplotypes from candidate oncogenes on all-cause mortality of ovarian cancer patients.

5. To investigate the effect of tSNPs and haplotypes in a series of "functional" candidates identified from *in vitro* studies on all-cause survival of ovarian cancer patients.

6. To evaluate the ease of use and quality of whole genome amplification methods.

7. To evaluate the performance of non-amplified and whole amplified DNA on multiplex SNP genotyping platforms.

# Chapter 2: Materials and Methods

## 2.1: Introduction

This chapter will describe the materials and methods used in this research. All the samples analysed were Caucasians who were either healthy, unaffected controls or individuals diagnosed with invasive epithelial ovarian cancer.

### 2.1.1: Ethics Statement

The collection and genetic analysis of all samples was approved by local review boards and ethics committees, and informed written consent was obtained from all study participants.

## 2.2: Study individuals

Thirteen population-based case control ovarian cancer sample series were used in the analyses, totalling 6,245 cases and 8,787 controls. These studies comprised of residents of the United Kingdom (SEARCH and UKOPS); Australia (AUS); Denmark (MALOVA); Germany (GER, BAVARIA), Poland (POCS, also known as JAC), and the United States of America (GEOCS, USC, DOVE, HOPE, NCOCS and HAWAII). Although many of the sample sets included non-Caucasian individuals, only the genotypes of non-Hispanic Caucasian samples of North European descent were analysed. This decision was taken because there are some significant differences in the allele frequencies of some SNPs within different ethnicities, and some variants are polymorphic in some ethnicities, but not others. The analysis of genotypes of only Caucasian minimises population stratification. This was discussed

in chapter 1. Follow-up data was only available for the GEOCS, MALOVA, SEARCH and UKOPS studies. The DOVE, HOPE, AUSTRALIA, JAC, BAV, GER, HAW and NCOCS studies were only used in the validation of statistically significant findings from stage 1 genotyping results.

The Genetic Epidemiology Ovarian Cancer Study (GEOCS, formerly known as FROC and Stanford) comprised of 327 cases and 429 controls. The cases were invasive epithelial ovarian cancer patients aged between 20 and 64 years, who were diagnosed with the disease between 1997-2002, from Alameda, Contra Costa, Marin, San Francisco, San Mateo and Santa Clara of the Greater Bay Area of San Francisco, USA. The affected individuals were all prevalent cases, therefore, they were recruited into the study after the cancer was diagnosed. The controls were recruited into the GEOCS study through random-digit dial identification from the same towns and cities of the Greater Bay Area Cancer Registry San Francisco as the cases. The controls were age (5-year categories) and ethnicity matched with the cases. The DNA from the study participants was extracted from blood samples and exfoliated buccal cells from mouthwash rinses with the Puregene Kit (Gentra Systems, Minneapolis, MN), (Lum and Le Marchand 1998). The vital status information of the GEOCS cases was obtained from the Greater Bay Cancer Registry, San Francisco twice during the study. The most current follow-up occurred in 2004. Computerised hospital tumour registry data or medical records were used for updated vital status by cancer registry staff. The state's death index was also used to follow the vital status of patients. There was a lag time of approximately 18 months with the state's death index. 147 deaths have occurred to date (45%). The majority

of the genotyping was on DNA that had been whole genome amplified with primer extension pre-amplification (PEP).

The Malignant Ovarian Cancer prediction study (MALOVA) contained 1221 controls and 446 cases from Denmark. The criteria for cases were women aged between 30 and 80 years, who were diagnosed with an invasive epithelial ovarian tumour over the December 1994 and May 1999 time period. Cases were recruited from 18 hospitals from the municipalities of Copenhagen, Frederiksborg, and counties within Copenhagen, Frederiksborg, Roskilde, Western Sealand, Storstrøm, Funen, Southern Jutland and Northern Jutland. All the cases were recruited into the study at surgery before diagnosis of the disease, therefore they are said to be incident cases. Follow-up to establish the patients' vital statistics occurred until 2003. Individuals living in Denmark have a unique personal identification number which was used to identify patients who were alive, as well as those who had died or emigrated. The cause of death of those who died during follow-up was determined by matching medical records with a Danish Hospital Reference System. Currently, there have been 301 (67%) deaths. Unaffected controls were obtained from the general female population within the same areas as the cases and the age range was also 30-80 years. Genomic DNA from both cases and controls were extracted from pre-operative blood samples by Whatman International Ltd with chloroform protocol (Ely, UK).

The UK SEARCH ovarian cancer study **(**SEARCH**)**, consisted of 1,215cases of ovarian cancer and 1,229 controls from an ongoing, population-based ovarian cancer case-control study covering the regions served by the East Anglia and West

Midlands cancer registries in the UK. The cases were younger than 70 years from East Anglia, West Midlands and Trent regions of England. Prevalent cases diagnosed between 1991 and 1998, of which there were 284 participants, were recruited for the study. The incident cases were recruited from 1998 onwards. Active follow-up was conducted at 3 and 5 years after diagnosis, and then at 5-year intervals by the Eastern and West Midlands cancer registries. The latest update was on 31$^{th}$ August 2007. Follow-up involved searching hospital information systems for recent visits and contacting general practitioners for the patient's vital status if a recent visit had not occurred. There were 230 (27%) deaths at the time of analysis. Healthy individuals, aged between 45 and 74 years, from the EPIC-Norfolk constituent of the European Prospective Investigation of Cancer (EPIC) cohort of 25,000 people were recruited as controls. The controls were from the same geographical region as the cases. The blood DNA of study participants was extracted by Whatman International Ltd.

The participants of the United Kingdom Ovarian Population Study (UKOPS) were recruited from the UK. There were 691 cases and 1,051 controls. The cases were recruited from 10 major Gynaecological Oncology National Health Service centres in England (University College London Hospital, East Kent, Gateshead, Southend, Bristol, Middlesbrough, Manchester and Portsmouth), Wales (North Wales) and Northern Ireland (Belfast), from 2006 onwards. The UKOPS controls, aged 50-76 years from the general population, were apparently healthy postmenopausal females who were recruited into the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) study. However, 75 of the cases were identified through the UKCTOCS study. DNA was extracted with the chloroform extraction method

(Sambrook and Russell 2001). The control women were followed up for cancers through the Office of National Statistics. The most current follow-up for the UKCTOCS samples was June 2008, and in August 2008 for the remaining samples. Survival data was available for 401 of the cases at the time of analysis. Of these cases, 148 were diagnosed with ovarian cancer after recruitment into the study, and the remaining were prevalent cases. At the time of analysis, there were 83 deaths (21%) out of the 391 cases with complete follow-up data.

The University of Southern California/Los Angeles County case-control studies of ovarian cancer (USC), from the USA, consisted of 434 ovarian cancer cases and 584 healthy controls, aged between 18-84 years. Recruitment began in 1993 and is ongoing as part of a larger study, known as the Los Angeles County Case-Control Studies of Ovarian Cancer (LAC-CCOC). The cases were identified from the Surveillance, Epidemiology, and End Results (SEER) registry. Unaffected study participants (controls) were matched with the cases in terms of age, race, socio-economic status, parity, oral contraceptive use, geographical residence and other ovarian cancer risk factors, however only non-Hispanic Whites were analysed in this study. DNA was extracted from blood lymphocytes with the chloroform extraction (Sambrook and Russell 2001) process or the Qiagen Blood Kit (Qiagen, Chatsworth, CA, USA). The DNA samples were sent to Molecular Staging, Inc. (New Haven, CT, USA) for whole genome amplification with RepliG™. There was no follow-up for this study.

Table 2.1: Ovarian cancer case-control populations used in study

| Population* | Cases | | | | Controls | | | |
|---|---|---|---|---|---|---|---|---|
| | Total | Age (years) | Part.~n rate (%) | Ascertainment | Total | Age (years) | Part.~n rate (%) | Ascertainment |
| MALOVA (Denmark) | 446 | 35-79 | 79 | Incident cases diagnosed 1994 -1999 from municipalities of Copenhagen & Frederiksberg & surrounding counties. | 1,221 | 35-79 | 67 | Random selection of females from the computerized Central Population Register. |
| SEARCH (UK) | 847 (368)§ | 21-74 | 69 | Cases from East Anglian, West Midlands & Trent regions of England. Prevalent cases diagnosed 1991-1998; incident cases diagnosed 1998 onwards. | 1,229 | 39-77 | 84 | Selected from the EPIC-Norfolk cohort of 25000 individuals based in the same geographical regions as the cases. |
| GEOCS (USA) | 327 | 23-64 | 75 | Consecutive cases diagnosed from 1997-2002 in Greater Bay Area Cancer Registry San Francisco. | 429 | 19-66 | 75 | Random-digit dial identification from study area. Frequency matched to cases for race/ethnicity & 5 year age group. |
| USC (USA) | 197 | 18-84 | 73 | Rapid case ascertainment through Los Angeles Cancer Surveillance program from 1999-2004. | 224 | 21-78 | 73 | Neighbourhood recruited controls, frequency matched to cases for age & ethnicity from 1993-2004. |
| UKOPS (UK) | 506 (185)§ | 35-86 | 86 | Cases from 10 gynaecological oncology National Health Service centres throughout the UK, from January 2006 onwards. | 595 (467)§ | 50-76 | 97 | Apparently healthy postmenopausal women from the general population participating in the UKCTOCS. Followed up for cancers through the Office of National Statistics. |
| DOVE (USA) | 584 | 35-74 | 75 | Cases diagnosed with primary invasive ovarian cancer between 2002-2005 from a 13-county area of Western Washington state. | 716 | 35-74 | 82 | Random-digit dial identification from study area. Frequency matched to cases for race/ethnicity & 5 year age group. |
| BAV (Germany) | 228 | 25-81 | | Hospital based study from Erlangen, Northern Bavaria, Germany, Recruitment from May 2002 to August 2008. | 234 | 24-86 | | Random selected woman from Erlangen, Northern Bavaria, Germany Recruitment from May 2002 to August 2008. |

| Population* | Cases | | | | Controls | | | |
|---|---|---|---|---|---|---|---|---|
| | **Total** | **Age (years)** | **Part.~n rate (%)** | **Ascertainment** | **Total** | **Age (years)** | **Part.~n rate (%)** | **Ascertainment** |
| GER (Germany) | 218 | 21-74 | 58 | Incident cases diagnosed 1993 -1996 from two study areas in southern Germany and identified through frequent monitoring of hospitals serving the study areas. | 416 | 23-75 | 51 | Two controls per case matched by age and recruitment area were selected from a random sample of the general female population in study area selected using population registries |
| POCS (Poland) | 603 | 23-82 | 80 | Cases diagnosed with epithelial ovarian cancer in five gynaecological oncology centres in Poland; between 1998 and 2006. | 593 | 24-74 | 90 | Healthy women from the general population were randomly selected and matched to cases with the same year of birth and geographical region. |
| NCOCS (USA) | 622 | 20-74 | 70 | Identified from 48 counties within Northern California | 747 | 22-75 | 63 | Controls identified from same region. Frequency matched to cases for age and race. |
| HAWAII (USA) | 70 | 18-84 | 66 | Rapid case ascertainment through Hawaii Tumour Registry. | 158 | 27-86 | 69 | Randomly selected from Hawaii Department of Health Annual Survey of the representatives households. |
| AUS (Australia) | 768 | 19-79 | 84 | Comprised of Cancer registries of New South Wales and Victoria. Recruited through surgical treatment centres throughout Australia. | 1,122 | 19-79 | 47 | Randomly selected from Commonwealth electoral roll. Frequency matched for age and geographical region. |
| HOPE (USA) | 276 | 25-80 | 69 | Variable source including physician offices cancer registries & pathology databases from counties of Western PA Eastern OH & Western NY. | 636 | 25-80 | 81 | Identified in same regions as cases. Frequency matched for age & ethnicity. All participants undergo home interviews. |
| **Total** | **6,245** | | | | **8,787** | | | |

§ - additional samples used in validation of functional candidates results. Part.~n - participation.

Diseases of the Ovary and their Evaluation (DOVE) study, also from the USA, is part of the SEER registry. There were 584 cases aged between 35-74 years, diagnosed with primary invasive ovarian cancer between 2002 and 2005 from a thirteen-county area of western Washington State. Controls were selected through random digit dialling. There were 716 controls who were matched to the age groups, race/ethnicity and area of residence as the cases. DNA of the cases and controls was isolated from blood or buccal cell samples.

The Hormones and ovarian cancer prediction study (HOPE) from the USA, recruited study participants from the counties of western Pennsylvania, eastern Ohio and western New York. The 276 invasive epithelial ovarian cancer cases were identified from a variety of sources including physicians' offices, cancer registries and pathology databases from the study region. Both cases and controls were individuals aged between 25 and 80 years. The 636 controls were recruited from the same cities as the cases. The controls were frequency matched for age. The case-control population sets are summarised in Table 2.1.

The Australian (AUS) case-control samples series comprised of sample collections from the Australian Cancer Study (ACS) and the Australian Ovarian Cancer Study (AOCS). There were a total of 768 cases and 1,122 controls when the two studies were amalgamated. The controls of both studies were randomly selected from Commonwealth electoral roll, and age- and geographical region-matched to the cases. The controls were aged between 19 and 81 at the time of recruitment. Participant recruitment occurred between 2002 and 2005 for ACS, and 2002-2006 for AOCS.

The cases from the ACS study were recruited from Cancer registries of New South Wales and Victoria; and the cases from the AOCS study were recruited from surgical treatment centres in Australia, and also cancer registries of Queensland, and South and West Australia. Age range of cases was 23-80 years.

The 603 invasive epithelial ovarian cancer cases from the Poland Ovarian Cancer Study (POCS), previously known as JAC, were recruited between 1998 and 2006, from five gynaecological oncology centres from four cities (Szczecin, Opole, Poznan and Rzeszów) in Poland. There was a participation rate of 80% among the ovarian cancer sufferers approached. The controls (593 individuals) of the study comprised of randomly selected healthy women from the general population. There was a 90% participation rate among the controls. The controls were matched to the cases by geographical region of residence and the year of birth.

The BAVARIA study consisted of 234 unaffected controls and 228 women diagnosed with invasive epithelial ovarian cancer. The patients were recruited from May 2002 to August 2008 from hospitals within Erlangen, of Northern Bavaria, Germany. The apparently healthy controls, aged between 24 and 86 years, were randomly selected from the same geographical area as the cases in the same time period.

German Ovarian cancer study (GER) consisted of 416 healthy controls (58% participation rate), and 218 individuals with ovarian cancer. Incident cases of ovarian cancer, diagnosed in individuals aged between 20 and 75 years were

recruited from two study areas in southern Germany. The cases were recruited from 1993 to 1996 through frequent monitoring of hospitals within the study areas. Controls from the general population were matched, 2:1 with the cases, for age and area of residence. The controls were randomly selected from population registries.

Confirmed cases of primary epithelial ovarian cancer from the Hawaii Ovarian Cancer Study (HAWAII) were recruited from residents of Oahu. The 70 affected participants were diagnosed between 1 June 1993 and 30 June 1999 in the major hospitals of Oahu. The 158 controls of the study comprised of unaffected women from the general population of Oahu. Controls were selected from lists of women who had been interviewed by the Health Surveillance Program of the Hawaii Department of Health. Participants of the Health Care Financing Administration of Oahu aged 65 years or older were randomly selected as potential controls. The controls were ethnicity and 5-year age matched with the cases in order to help minimise selection/ascertainment bias. DNA was extracted from peripheral blood leukocytes by SDS/proteinase K treatment and phenol/chloroform extraction. All the participants analysed in the study were Caucasian. There were 70 cases and 158 controls. These samples were only used in the stage 2, validation studies.

The North Carolina Ovarian Cancer Study (NCOCS) samples used comprised of 622 cases and 747 controls, all of whom were Caucasian. Eligible cases were recruited from a 48-county area of North Carolina. Rapid case ascertainment was used to identify potential study participants from the North Carolina Central Cancer Registry. This registry contains information on cancer sufferers from the general population of the state. Patients with primary ovarian cancer aged between 20 and

74 years from the 48 counties within North Carolina fulfilled the study entry criteria. List assisted random dialling was used to identify population-based controls from the same 48-county region as the cases. The controls were also ethnicity and 5-year age matched with the cases. The DNA was extracted using the PureGene DNA isolation protocol.

## 2.3: Gene and tagging SNP selection of candidate oncogenes

Various oncogenes have been implicated in the development of ovarian cancer; these include *AKT2, BCL2*, *BRAF*, *CMYC*, *CTNNB1*, *ERBB2*, *KIT*, *KRAS*, *MUC1*, *MUC2* and *PIK3CA*. SNP genotyping data on the Centre d'Etude du Polymorphisme Humain (CEPH) population for *AKT2, BCL2*, *BRAF*, CMYC, *CTNNB1*, *ERBB2*, *KIT*, *KRAS*, *MUC1*, *MUC2*, *NMI* and *PIK3CA* were downloaded from The International HapMap Project, Data Release 20/phase II Jan06. *NMI* is not an oncogene, but was accidentally selected as a result of the information being displayed when *CMYC*, an alias of *MYC*, was entered as a search term in HapMap. There was genotyping data available for only two common SNP (minor allele frequency ≥0.05) for *MYC*, so this oncogene was excluded from further evaluation. The genotyping data downloaded for these genes was CAU, which is a reference for Caucasians of north European populations. The gene selection process is discussed in chapter 3.

The reference genotyping data from HapMap was used to select a group of SNPs from each gene (with minor allele frequency [MAF] of at least 5%), which could subsequently be tagged. To do this, the genotype information was imported into Haploview version 3.32 (Barrett *et al*. 2005) and Tagger (de Bakker *et al*. 2005).

Haploview is a programme which was designed primarily for haplotype analysis. The programme can be used to perform: linkage disequilibrim (LD) and haplotype block analyses; haplotype population frequency estimation; single SNP and haplotype association tests; and permutation testing for association significance. Tagger contains an algorithm which performs tagging SNP (tSNP) selection. Tagger is able to produce a list of tSNPs by initially selecting a set of SNPs which are to be captured through the tagging approach.

The LD between a pair of SNPs is established so that SNPs which are in strong LD capture each other, and therefore only one tSNP needs to be genotyped. This tagging approach is known as the pair-wise method for SNP selection. It is possible that a SNP is in strong LD with several SNPs. This SNP is selected to be the tSNP and it is said to capture all the SNPs it tags.

Aggressive tagging is another method of tSNP selection. The initial stage of aggressive tagging is the same as that of pair-wise tagging. The additional steps include using multi-marker/SNP tests to try to capture SNPs which could not be tagged by other SNPs with the pair-wise approach. Multi-marker tests are used because in some instances, a combination of markers is in stronger LD with a SNP than another single SNP. The software then “peels back” the tSNP list by replacing some of the tSNPs with multi-marker SNPs. Haploview and Tagger have several options which can be changed by the user, and thus a user is able to select tSNPs based on a criterion.

Haploview and Tagger were used to select tSNPs that capture common genetic variation (minor allele frequency ≥ 5%) from the candidate genes, and putative regulatory regions up and down stream of the gene (within 5kb), with a minimum squared correlation of 0.8 ($r^2 \geq 0.8$). $r^2 \geq 0.8$ means that there is at least 80% correlation between the genotype tSNP and the SNPs that it tags. The quality of the HapMap data was also ascertained and only SNPs with sufficiently good quality data were selected for tagging. The selection criteria for good quality data was based on ≥80% genotyping data of the CEPH participants for each common polymorphism. The other criteria for SNP selection were for the Hardy-Weinberg equilibrium p-value to be greater than 0.01, the minimum percentage of non-missing genotypes for each SNP (of the HapMap data) to be ≥ 80% and the maximum number of Mendelian inheritance errors in the HapMap CEPH trios to be no greater than 1. The 2-3 multi-marker (aggressive) tagging option of Tagger was used to select tSNPs.

If a selected tSNP failed assay design or genotyping, an alternative tSNP was chosen where possible. The sequences for the SNPs were obtained from the National Center for Biotechnology Information (NCBI) SNP database, dbSNP, (http://www.ncbi.nlm.nih.gov/SNP/) and were used for SNP pooling and primer design.

## 2.4: Microcell-mediated chromosome transfer of chromosome 18

The microcell-mediated chromosome transfer of chromosome 18 (MMCT-18) *in vitro* and *in vivo* experiments were performed by Dr Dimitra Dafou. Details of the experimental procedure of the MMCT of chromosome 18 can be found in (Dafou *et*

*al*. 2008). Briefly, donor mouse A9 cells containing normal human chromosome 18 were micronucleated with 48 hours of colcemid. The human chromosome was tagged with selectable fusion gene marker, hygromycin phosphotransferase. Polyethylene glycol was used to fuse the donor cells to the endometrioid TOV21G, and the clear cell TOV112D, ovarian cancer cell lines. This procedure is summarised in Figure 2.1. Microcells containing the human chromosome were selected with hygromycin B. TOV21G and TOV112D hybrid clones were isolated and expanded after 2-3 weeks of culture.

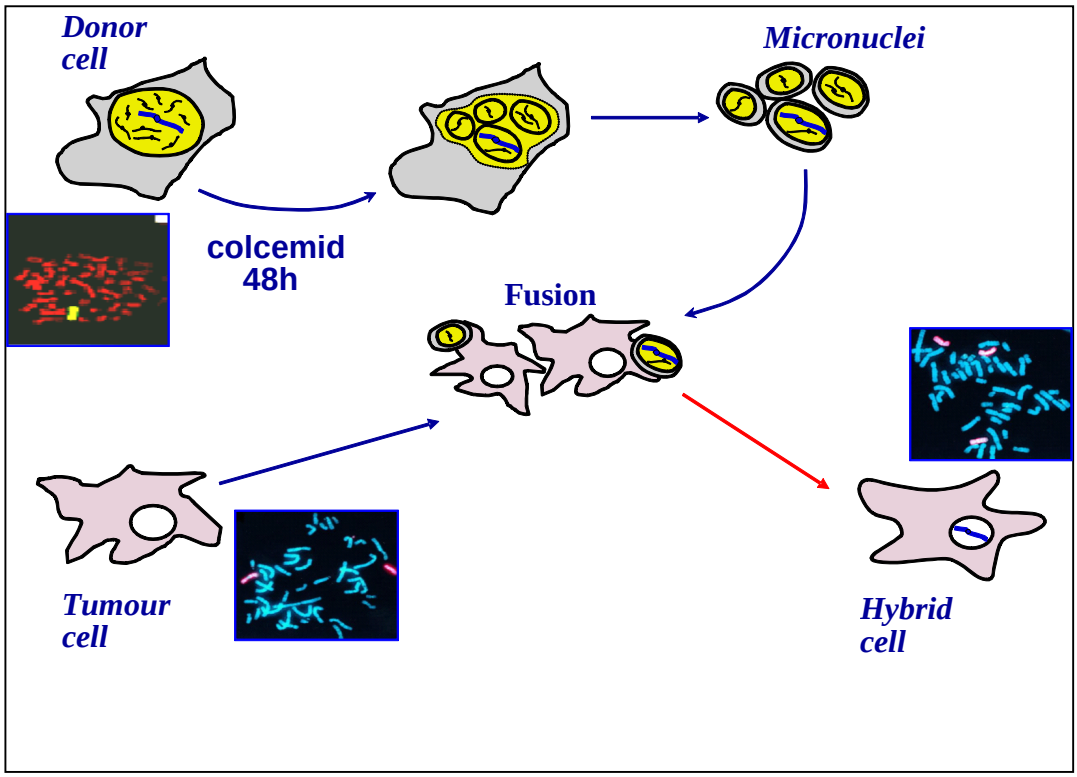



Figure 2.1: Schematic diagram of microcell-mediated chromosome transfer (MMCT)

(printed with permission from Dr Dimitra Dafou). **Hybrid cells contained an extra copy or a fragment of normal human chromosome 18. The clones showed* in vivo *and* in vitro *characteristics suggesting their phenotype was reverting back to that of non-neoplastic cells.*

The tumourigenicity of the resulting recipient: donor hybrid clones were assessed using *in vitro* and *in vivo* assays. Fluorescence *in situ* hybridisation (FISH) was also used to visualise the incorporated chromosomes in the hybrid clones.

Microsatellite analysis and array comparative genomic hybridisation were used to evaluate the regions of chromosome 18 which were transferred into the hybrid clones. The whole chromosome was transferred into the TOV21G cell line hybrids, however only the chromosomal region 18p11.21 – 18q11.2 was transferred into the TOV112D hybrids. Two hybrid clones from each cell line (18G1 and 18G5 from TOV21G, and clones18D22 and 18D23 from TOV112D) were selected based on their phenotypic characteristics - *in vivo* and *in vitro* tumour suppression.

The Applied Biosystems 32K gene expression array platform (Applied Biosystems) was used to evaluate the global gene expression levels of each parental ovarian cancer cell line and their corresponding "reverted" hybrids in triplicate. The Spotfire DecisionSite$^{TM}$ software for functional genomics (Spotfire AB, Goteborg, Sweden) and R version 1.9.1. were used to assess the fold change in gene expression between each hybrid and their parental cell line in the 32,878 probes of 29,098 genes. The corresponding P-values for the fold changes in gene expression were evaluated with an analysis of variance (ANOVA) test. The pooled data from both hybrids of each of the cell lines were also analysed using Spotfire DecisionSite$^{(TM)}$ and R software.

## 2.5: Gene and tagging SNP selection of "functional" candidate genes

Candidate genes were selected based on significant differential expression between MMCT-18 hybrids and the parental ovarian cancer cell lines. Genes with

concordant and consistent expression fold changes (up or down regulation following the insertion of the normal human chromosome 18), within the two hybrid clones, were selected. Genes with consistent expression changes between the TOV21G and TOV112D were also selected for a master-list. The master-list of the candidate genes selected based on consistent fold changes in expression between parental and the hybrids, and statistically significant P-values for the fold changes, is given in Appendix I. The candidate gene selection process is described in detail in Chapter 3.

## 2.6: Selection of genes tagging SNPs analysed with admixture maximum likelihood test

The admixture maximum likelihood (AML) test involved the evaluation of genotyping data of 3 population-based studies (GEOCS [327 cases, 429 controls], MALOVA [446 cases, 1,221 controls] and SEARCH [847 cases, 1,229 controls]). The results of the associations from the genotyping data had previously been reported in (Dicioccio *et al.* 2004; Auranen *et al.* 2005; Song *et al.* 2006a; Song *et al.* 2006b; Gayther *et al.* 2007; Song *et al.* 2007; Ghoussaini *et al.* 2008; Quaye *et al.* 2009). Over the course of the last 6 years, there have been developments in SNP association studies. The rationale and approaches used in SNP selection have also changed. The limited success in finding strongly associated genes with ovarian cancer development has also lead to new approaches being used to identify candidate genes.

Candidate gene selection for ovarian cancer associations studies have predominantly been based on biological pathways that are predicted to be involved in ovarian carcinogenesis. These pathways include DNA double strand break repair, DNA

mismatch repair and cell cycle control. Coding SNPs from the DNA double strand repair (*BRCA1, NBS1, RAD51, RAD52, XRCC2* and *XRCC3* genes) and cell cycle control (*STK15* gene) pathways were selected in the earliest studies. The alleles of these functional SNPs resulted in different amino acids being produced. These variants were selected because it was biologically plausible that they would be directly involved in altering protein function through the folding and binding of the protein. These changes would be expected to affect ovarian cancer development or susceptibility if there was a significant association.

The LD between SNPs was advantageously used in the tagging SNP approach of candidate genes from the DNA mismatch repair pathway, and all subsequent candidate genes (from cell cycle and oncogene pathways and MMCT-18). The tagging approach enabled the genotyping of smaller numbers of SNPs from genes, which would provide genotyping data for a greater number of SNPs overall. Oncogenes and tumour suppressor genes which were known or predicted to be involved in ovarian cancer development were also selected for associations studies.

Candidate genes were also selected based on differential expression of cancer parental cell lines and their suppressed, non-neoplastic normal chromosome 18 hybrids (from the functional MMCT-18 study). The genes selected from the MMCT-18 study were the only genes chosen based on putative, functionally relevant candidate genes for ovarian cancer aetiology through *in vitro* and *in vivo* assays. Furthermore, candidate SNPs validated by OCAC were selected because they had been found to be significantly associated with ovarian cancer in other population-based studies (from other members of the Ovarian Cancer Association Consortium

[OCAC]) or associated with breast cancer (identified by the Breast Cancer Association Consortium- [BCAC]).

## 2.7: Laboratory work

The vast majority of the experiments were conducted in the laboratories at University College London. Due to the collaborative nature of the studies, some of the genotyping was also conducted at Strangeways Research Laboratory at the University of Cambridge (GEOCS and SEARCH for all SNPs except MMCT and some oncogenes), University of Southern California (HOPE, DOVE, NCOCS) and Australia (AUS).

The MALOVA samples were normalised to 50ng/ul with distilled water into deep-well plates with the 8-span liquid handling (LiHa) arm of the Tecan Freedom EVO® workstation (Tecan, Reading, UK). LiHa accurately distributes low volumes with the aid of pinch valves. Filter tips were used to minimise contamination. The LiHa and TeMO®, a 96-head multi-channel pipette, were used to dilute some of the 50ng/ul DNA to 2ng/ul. The TeMO was also used to dispense 5ul of the 2ng/ul DNA to 384-well, barcoded PCR plates. Barcoded plates were used for ease of tracking the DNA plates (and sample).

## 2.8: Whole genome amplification methods

Ninety-five MALOVA control samples and one non-template test control (NTC) were whole genome amplified with GenomePlex™, GenomiPhi™, primer extension pre-amplification (PEP) and REPLI-g™. The starting concentrations of DNA amplified with GenomePlex™, GenomiPhi™ and REPLI-g was 100ng of DNA.

20ng of DNA was used for PEP amplification. All amplification reactions were performed manually in 96-well PCR plates.

### 2.8.1: Whole genome amplification with GenomePlex

100ng of 95 MALOVA samples were amplified with the GenomePlex® Whole Genome Amplification (WGA2) kit 2 (Sigma, Poole, Dorset, UK). 2µL of 50ng/µL DNA was diluted with 8µL distilled water and fragmented at 95$^{o}$C for 4 minutes. An OmniPlex library mix, containing 2µL of 1x Library Preparation Buffer and 1µL of the Library Stabilization Solution were added to each sample. The mixture was subsequently incubated for 2 minutes at 95$^{o}$C. The mixture was cooled on ice and 1µL of Library Preparation Enzyme was added. The DNA-library solution was incubated for the following conditions: 16$^{o}$C for 20 minutes, 24$^{o}$C for 20 minutes, 37$^{o}$C for 20 minutes, 75$^{o}$C for 5 minutes and cooled to 4$^{o}$C. A mixture containing 7.5µL of 10x Amplification Master Mix, 47.5µL of nuclease-free water and 5µL of WGA DNA polymerase was added to each sample. The mixture was thermocycled for 95$^{o}$C for 3 mins, 14 cycles (of 94$^{o}$C for 15 secs and 65$^{o}$C of 5 mins); and cooled to 4$^{o}$C.

A working stock of the amplified material at the concentration of 2ng/µl was stored at 4$^{o}$C, and the original and 20ng/µl stock were stored at -20$^{o}$C.

### 2.8.2: Whole genome amplification with GenomiPhi

The GenomiPhi DNA amplification kit (GE Healthcare, Bucks, UK) was also used to amplify the 95 MALOVA DNA samples. The DNA (2µL of 50ng/µL) was denatured at 95$^{o}$C for 3 minutes. An amplification mix containing 9µL of

GenomiPhi reaction buffer and 1µL of the GenomiPhi φ29 enzyme was added to each sample. The DNA-amplification mixture was incubated at 30$^{o}$C for 16 hours. The reaction was heat inactivated at 65$^{o}$C for 10 minutes and cooled to 4$^{o}$C.

### 2.8.3: Whole genome amplification with PEP

For each primer extension pre-amplification (PEP) reaction, 10ul of 2ng/µL of the 95 MALOVA samples were amplified in 50µL final volume reactions. The 40µL PCR reaction for each sample consisted of 22.75µL water, 5µL of 10x PEP buffer (1.5nM Mg), 2µL of 25nM magnesium, 5µL of 2nM dNTP, 5µL of 2000µM PEP N15-mer (5' NNN NNN NNN NNN NNN 3'), and 0.25µL Qiagen Taq polymerase. The PCR master-mix containing the PCR components were added to the DNA samples, and subsequently thermocycled for the following conditions: activation of the enzyme for 3 minutes at 94$^{o}$C, 50 cycles of (94$^{o}$C for 3 mins, 37$^{o}$C for 2 mins, 37$^{o}$C to 55$^{o}$C [RAMP at 10 seconds per $^{o}$C], 55$^{o}$C for 4 mins), incubated at 72$^{o}$C for 5 mins and cooled to 4$^{o}$C.

### 2.8.4: Whole genome amplification with REPLI-g

100ng of 95 MALOVA samples were amplified with REPLI-g Midi Kit (Qiagen, West Sussex, UK). 500µL of Solution A was prepared with 40µL 5M potassium hydroxide (KOH) and 10µL of 0.5M ethylenediamine tetra-acetic acid (EDTA) (pH8) which had been diluted in 450µL deionised water. 280µL of denaturing buffer was made up with 35µL Solution A and 245µL nuclease-free water. 560µL of neutralisation buffer was prepared with 56µL Solution B and 504µL nuclease-free water.

0.5µL of tris-EDTA (TE) was added to 2µL of 50ng/µL DNA sample. 3µL of denaturation buffer was added to the DNA. This denaturing mixture was mixed, spun and incubated at room temperature for 3 minutes. 5µL of the neutralisation buffer was mixed into the samples to neutralise the denaturing reaction. A 40µL master-mix, containing 32.4µL nuclease-free water, 15µL 4x REPLI-g buffer and 0.6µL REPLI-g DNA polymerase, was added to each 10µL denatured and neutralised DNA solution. These solutions were mixed, pulse centrifuged and incubated at 30$^{o}$C for 16 hours. The amplification reaction was inactivated by incubating the plate at 65$^{o}$C for 3 minutes. The amplified DNA was cooled to 4$^{o}$C.

The DNA of an additional 95 samples were amplified with REPLI-g by a colleague, Mr Mark Cox, to further investigate concordance of the amplified DNA.

## 2.9: DNA quantification with PicoGreen

The whole genome amplified DNA were quantified with Quant-iT™ PicoGreen® dsDNA assay (Molecular Probes, Invitrogen, Paisley, UK). A 200-fold dilution of the concentrated DMSO solution (from the PicoGreen kit) was made with TE in a plastic container in a dark room. Calf thymus DNA (Sigma) was diluted to generate a high-range size standard which would be used to extrapolate the concentration of the sample DNA. 100µL of the calf thymus stock (1µg/mL) was diluted with 900µL of TE, to make a concentration of 100µg/mL. 84µL of TE was added to 16µL of the 100µg/mL DNA to make a concentration of 16µg/mL. A 1:5 dilution of the 16µg/mL DNA was made with 100µL of the 16µg/mL DNA and 400µL of TE (to make a DNA concentration of 3.2µg/mL. Serial dilutions of the DNA were made using 200µL DNA and 200µL TE as indicated below, starting with 3.2µ/mL:

| Volume of TE (µL) | Volume of calf DNA (µL) | Calf DNA concentration (µg/mL) |
|---|---|---|
| 0 | 200 | 3.2 |
| 200 | 200 | 1.6 |
| 200 | 200 | 0.8 |
| 200 | 200 | 0.4 |
| 200 | 200 | 0.2 |
| 200 | 200 | 0.1 |
| 200 | 200 | 0.05 |
| 200 | 0 | 0 |

(The 1.6µg /mL DNA was made by mixing 200µL of the 3.2µg/mL DNA with 200µL of TE. The 0.8µg/mL DNA concentration was made by mixing 200ul of the 1.6µg/mL DNA with 200ul TE. Etcetera…).

50µL of the diluted PicoGreen was added to 50µl of the diluted calf thymus DNA (standard) in triplicate in a black plate. 5ul of each WGA DNA sample was diluted with 45ul TE and 50ul of diluted PicoGreen was added the black plate. The Tecan Genios plate reader was used to measure the DNA concentration and the data was analysed with the Magellan software (Tecan, Dorset, UK).

## 2.10: Genotyping platforms

The ninety-five MALOVA samples which were whole genome amplified with GenomePlex, GenomiPhi, PEP and REPLI-g, and their corresponding non-amplified genomic DNA were genotyped with TaqMan®, iPLEX®, SNPlex® and TaqMan® OpenArray.

### 2.10.1: TaqMan® genotyping

For each 5µL TaqMan (Applied Biosystems, Warrington, UK) genotyping reaction, a master-mix containing 2.44µL distilled water, 2.5µl Applied Biosystem's SNP

genotyping master-mix, and 0.06μL 80x Custom Assay-by-Design TaqMan probe (Applied Biosystems), was added to 10ng of DNA. TaqMan genotyping reactions for oncogene and MMCT-18 tSNPs were conducted at half volume (2.5μL); appropriate adjustments were made to the volumes of reaction components used. Normally a large mastermix was made which contained enough reaction mix for the samples being genotyped. The 5μL or 2.5μL reaction mix was dispensed into the appropriate wells of the dried DNA plates with the liquid handling (LiHa) arm of the Tecan Evo 200 robot. The DNA – master-mix solution was thermocyled for the following conditions: activation at 95$^{o}$C for 10 mins, annealed/extended for 40 cycles of (95$^{o}$C for 15 seconds, 60$^{o}$C for 1 min), and cooled to 4$^{o}$C. All thermocycling was performed on Auto-Lid Dual 384-well GeneAmp® PCR System 9700 instruments and end reaction products were read on the 7900HT Fast Real-Time PCR System using the Sequence Detection Software. Although the vast majority of Applied Biosystems TaqMan assay clustered well with the annealing temperature at 60$^{o}$C, some required different temperatures. Therefore, each TaqMan assay was tested (with 95 DNA samples and an NTC) with the annealing temperature at 60$^{o}$C, before whole population sets were genotyped. An additional 5-10 anneal extend PCR cycles were performed if the clusters were sub-optimal. If this failed, or the assay did not produce distinct clusters for the genotypes, the probe test was repeated with annealing temperature of 54$^{o}$C.

### 2.10.2: iPLEX genotyping

The MassARRAY iPLEX SNP multiplex genotyping platform was used to genotype whole genome amplified samples, and genomic GEOCS, MALOVA and SEARCH samples for the oncogene study, and iPLEX Gold was used to genotype MALOVA,

SEARCH and UKOPS for the MMCT-18 study. iPLEX Gold is an upgrade to the iPLEX system. The only real difference between the platforms are the multiplex levels. While up to 29 SNPs can be genotyped with the iPLEX assay in a single reaction, up to 40 SNPs can be assayed with the upgrade. iPLEX Gold has a wider mass range from which alleles/SNPs can be detected. The GEOCS and SEARCH samples were air-dried genomic samples, which had been plated a year prior to the lab work. Wet MALOVA DNA was used in the iPLEX runs. Desalted forward, reverse and extend primers for the iPLEX panels were manufactured by Metabion (Martinsried, Germany). The PCR with the forward and reverse primers were performed at UCL with the Tecan robot, and all post-PCR processing was conducted at Sequenom Europe in Hamburg, Germany, by the author.

A primer mix, comprising of 120μL of 500nM of each of the forward and reverse primers of all the SNPs were combined, resulting in a final concentration of 100nM in each 5μL reaction. A dNTP mix was also prepared with equal amounts (400μL) of 100nM dATP, dCTP, dGTP and dTTP. A PCR cocktail containing the following for each sample: (1.85μL distilled Milli-Q water, 0.625μL of PCR buffer with 10X magnesium chloride [$MgCl_2$], 0.325μL of 25mM $MgCl_2$, 0.1μL of 25mM dNTP mix, 1μL of primer mix [500nM of each primer], and 0.1μL of 5U/μL Hotstar Taq® DNA polymerase enzyme), was prepared and added to 10ng of DNA.

In a thermocycler, the reaction mixture was activated at 94$^{o}$C for 15 mins, cycled 45 times (at 94$^{o}$C for 20 secs, 56$^{o}$C for 30 secs, 72$^{o}$C for 60 secs), and inactivated at 72$^{o}$C for 60 secs. Unincorporated dNTPs in the PCR amplification mixture were dephosphorylated with a shrimp alkaline phospatase (SAP) cleaning step. This

involved incubating the PCR-amplified mixtures with 2μL of SAP mix (which comprised of 1.53μL distilled Milli-Q water, 0.17μL of 10x SAP buffer and 0.3μL of 1U/μL SAP enzyme for each sample). The incubation steps were 37$^{o}$C for 20 minutes and 85$^{o}$C for 5 minutes. The dephosphorylated mixture was then cooled to 4$^{o}$C. The SAP cleaning was necessary in order to prevent the remaining dNTPs being incorporated in the primer extension reactions, which could subsequently result in contamination peaks being present in the results.

The extend primers were pooled into four groups according to the mass of the extend primers. The signal-to-noise ratios of the extend primers decrease with increasing extend primer mass, therefore these adjustments in extend primer concentrations were required in order to equilibrate the signal-to-noise ratios of the extend primers of different masses. The extend primers were arranged into increasing masses, and the primers were split into 4 groups. Therefore, lower mass primers were grouped with other low mass primers and high mass primers were grouped with other high mass primers. The final concentration of the lowest mass primers was half of those in the highest mass group. Thus an extend primer mix was prepared whereby the final concentrations of the group 1 extend primers (lowest mass) was 0.625μM, group 2 was 0.833μM, group 3 was 1.042μM, and group 4 (highest mass) was 1.25μM.

An iPLEX reaction mix was made up of (for each sample): 0.755μL distilled water, 0.2μL of 10X iPLEX buffer, 0.2μL of iPLEX termination mix, 0.804μl of the extend primer mix and 0.041μL of the iPLEX enzyme. 2μL of this cocktail was added to

each sample. The mixture was mixed and covered with adhesive seal. This was subsequently cycled for the following PCR conditions:

| HOLD | 40 CYCLES | | | HOLD | HOLD |
|---|---|---|---|---|---|
| | Hold | 5 cycles | | | |
| 94°C<br>30 secs | 94°C<br>5 seconds | 52°C<br>5 secs | 80°C<br>5 secs | 72°C<br>3 mins | 4°C<br>15 mins |

The iPLEX reaction products were desalted by adding 25µL of water and 6mg of Clean resin (using a dimple plate). A nano-dispenser was used to dispense the iPLEX reaction products onto a 384-element SpectroCHIP bioarray. The SpectroCHIPs were read on Bruker™ Autoflex, a matrix-assisted laser desorption/ionization time of flight **(**MALDI-TOF) mass spectrometer.

### 2.10.3: SNPlex genotyping

SNPlex is a 48-plex SNP genotyping method from Applied Biosystems. The genomic and amplified DNA were genotyped on two separate occasions on the SNPlex platform. The first run was manually performed at UCL by the author and Applied Biosystems' SNPlex technical expert. The SNP pass rates from this run were very poor, therefore the experiments were repeated by another Applied Biosystems' SNPlex technical expert at their laboratory in Warrington, UK. The results of the second run are discussed in this work.

50ng of the genomic MALOVA DNA and 100ng of the corresponding whole genome amplified products were genotyped with the SNPlex platform. To fragment the genomic DNA, 2.5µL of the 20ng/µL stock plate was dispensed into a 384 well

plate and incubated at 95$^{o}$C for 5 minutes. The samples were dried down after DNA fragmentation. The whole genome amplified products, which did not require fragmentation, were also air-dried. The fluorescent probes and linkers were phosporylated with the following mixture: for each samples, 0.1µL pooled SNPlex ligation probes, 0.05µL of 48-plex SNPlex universal linkers, 0.125µL of nuclease-free water, 0.05µL of 10x SNPlex kinase buffer, 0.025µL of SNPlex kinase, 0.1µL of 5x SNPlex enhancer, and 0.05µL 10x dATP. This mixture, known as the SNPlex Ligation probe pool, was incubated at 37$^{o}$C for 1 hour. The activated probe pool was diluted 1:1 with 0.1xTE of pH8.

The ligation reaction was performed with the oligonucleotide ligation assay (OLA) and the activated SNPlex ligation probe pool. The OLA was prepare on ice, and consisted, for each sample, of 3.422µL of nuclease-free water, 0.5µL of SNPlex ligation buffer, 0.025µL of SNPlex ligase, 0.053µL AmpErase® UNG. The 4µL OLA master-mix and 1µL of the activated SNPlex ligation probe pool was added to each DNA sample. The DNA plate containing the ligation reaction mixture was covered with an adhesive cover and incubated at 4$^{o}$C for 10 mins, placed on a thermocycler which was at 90$^{o}$C. The plate was thermocycled for the following conditions: 3 mins at 90$^{o}$C; 30 cycles of the 3 step (15 secs at 90$^{o}$C, 30 secs at 60$^{o}$C, 30 secs at 51$^{o}$C with 2% RAMP); incubated for 10 mins at 99$^{o}$C, and cooled to 4$^{o}$C.

The ligation product was purified by exonuclease digestion. To do this, a 2x exonuclease master-mix was prepared on ice. For each sample, the master-mix contained 4.2µL nuclease-free water, 0.5µL of 10x SNPlex exonuclease buffer, 0.2µL of SNPlex lambda exonuclease, 0.1µL of SNPlex exonuclease I. 5µL of the

master-mix was added to each sample, the sample plate was covered, vortexed, and pulse spun. The plate was transferred to a thermocycler at 37°C, and incubated at 37°C for 90 mins, 80°C for 10 min and cooled to 4°C. The exonuclease reaction product was diluted with 15µL nuclease-free water.

The PCR master-mix was prepared with 2.42µL of nuclease-free water, 5µL of 2x SNPlex amplification master-mix and 0.5µL of 20x SNPlex amplification primers. 7.92µL of the PCR master-mix was added to each well of a new 384-well plate, and 2.08µL of the diluted exonuclease reaction product was also added to the plate. The plate was covered, pulse spun and thermocycled for: 95°C for 10 mins, 30 cycles of 95°C for 15 and 63°C for 1min, and then cooled to 4°C.

A 1:10 dilution of the Wash Buffer was made with deionised water. The wells of the Hybridization Plate were washed three times with 100µL of diluted Wash Buffer. 17.491µL of the SNPlex Hybridisation binding buffer was diluted with 0.009µL of the positive hybridisation control. This was subsequently added to the SNPlex hybridisation plate. The PCR products were bound to the hybridisation plate by transferring 1.5µL from each well of the diluted PCR product to the Hybridisation Plate and incubating at room temperature on a rotary shaker for 1 hour. In this reaction, the biotinylated amplicons from the PCR products are bound to the streptavidin coat on the hybridisation plate. The plate was centrifugated at 1000RPM for 1min, the supernatant was removed, and the plate was washed three times with 100µL of diluted SNPlex Wash buffer. 50µL of 0.1N sodium hydroxide (NaOH) was added to each sample in the hybridisation plate, the plate was covered and incubated for 30 mins on a rotary shaker at room temperature.

The hybridisation plate was removed from the rotary shaker, spun at 1000RMP for 1 min, and the supernatant was removed. The hybridisation plate was washed five times with diluted 100µL SNPlex hybridisation wash buffer. An incubation oven was equilibrated to 37$^{o}$C. A hybridisation master-mix was prepared with 0.05µL of SNPlex ZipChute mix, 11.25µL of SNPlex Denaturant and 13.7µL of SNPlex ZipChute dilution buffer for each sample. 25µL of hybridisation master-mix was added to each sample in the hybridisation plate. The plate was covered and incubated at 37$^{o}$C on a rotary shaker for 1 hour.

A sample loading mix was prepared, which contained, for each sample, 0.59µL of SNPlex size standard and 16.91µL of SNPlex sample loading reagent. The hybridisation plate was briefly spun and the supernatant was removed. The plate was washed four times with 100µL of diluted hybridisation buffer. The plate was spun upside down at 100RPM for 1 min on a stack of paper towels. 17.5µL of SNPlex sample loading mix was added to each well of the hybridisation plate. The plate was covered and incubated at 37$^{o}$C for 30 mins. 7.5µL of the products in the hybridisation plate was transferred to a new 384-well optical reaction plate. The plate was read on the Applied Biosystem 3730xl DNA Analyzer. The results were analysed with GeneMapper 4.0 software. The manufacturer default settings were used to analyse the data.

### 2.10.4: OpenArray genotyping

The TaqMan® OpenArray™ genotyping system is another mid-range genotyping platform from Applied Biosystems. The 32-plex format was used for these

experiments, with which up to 96 samples can be genotyped on a single OpenArray Genotyping plate. However, 16-, 64-, 128-, 192- and 256-plex formats are also available for 144, 48, 24, 16 and 12 samples, respectively. Each OpenArray genotyping plate consists of 48 subarrays and each subarray is comprised of 64 through-holes. The hydrophobic and hydrophilic coatings ensure that the through-holes can retain 33nL reaction volume. The OpenArray genotyping plates were manufactured with the 32 TaqMan SNP assays, with an assay in each through-hole.

2ul of TaqMan® OpenArray™ Master-Mix was mixed with each well of a 384-well PCR sample plate (1ul of 50ng/μL genomic DNA or 2ul of 50ng/μL of WGA-DNA). The master-mix-DNA solutions were transferred from the sample plate to the OpenArray Genotyping plates. To do this, the sample plate was divided into eight different sections, consisting of 12-well by 4-well areas as shown in Figure 2.2. Plate guides were used to ensure that the appropriate sections of the sample plate were transferred to the OpenArray plate. The plate guide was placed over the sample plate. A tip block was placed over the appropriate section of the plate guide. OpenArray Loader tips were placed into each whole of the tip block. The tip block, with the loader tips inserted, was slid up and down approximately 50 times until the tips were filled to 1mm above the bottom edge of the tip block.

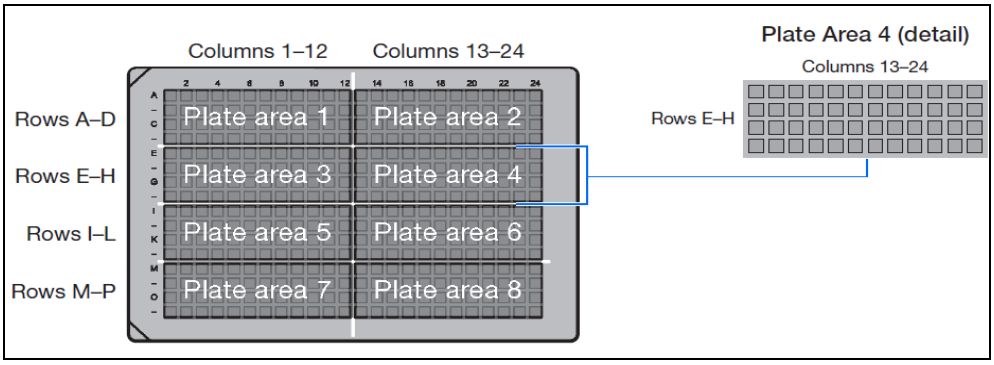



Figure 2.2: Sections of the OpenArray sample plate

Each OpenArray sample plate section contains 48 wells.

The samples in the loader tips were transferred to the OpenArray plate with the auto loader. The loaded OpenArray plate was placed into an OpenArray genotyping case, which was filled with immersion fluid. The genotyping case was sealed with glue on the OpenArray case sealing station. The genotyping case was cleaned and placed in the Bio-Rad thermal cycler for the following conditions:

| **Step** | **Temperature and Time** |
|---|---|
| RAMP | 0.8°C/ second to 95.5°C |
| Hold | 91.0°C for 10:00 mins |
| ANNEALING & EXTENDING: 50 CYCLES | 0.5°C/s RAMP to 51.0°C |
| | Hold 51.0°C for 23s |
| | 0.8°C/s RAMP to 53.5°C |
| | Hold 53.5°C for 30s |
| | 0.8°C/s RAMP to 54.5°C |
| | Hold 54.5°C for 13s |
| | 0.8°C/s to 97.0°C |
| | Hold 97.0°C to 22s |
| | 0.8°C/s RAMP to 92.0°C |
| | Hold 92.0°C for 7s |
| Hold | 20°C for 5 mins |
| Hold | 4°C for Forever |

s - seconds, mins - minutes

The genotyping case was “imaged” with the OpenArray™ NT Imager software. The genotypes were manually called after training by Applied Biosystems OpenArray technical expert.

## 2.11: Sequencing

DNA sequencing of 95 MALOVA cases amplified with REPLI-g was performed with the help of Mark Cox. The samples were sequenced for two regions of *BRCA1*, 316705 (exon 11), and 316700 (exon 13), in order to further investigate discordances between the genomic control DNA and corresponding REPLI-g DNA.

10ng of the amplified DNA was sequenced using the BRCA1 (v1) Variant SeqR kit (Applied Biosystems). 95-well plates were used. For each sample, a reaction mix containing: 5µL of 2x AmpliTaq Gold® Master mix, 1.6 µL of 50% UltraPure™ Glycerol, 1.0µL (0.6µM/µL) of Forward VariantSEQr RSA primer, 1.0µL (0.6 µM/µL) of Reverse VariantSEQr™ RSA primer, and 1µL of distilled water, was prepared. 10µl of the reaction mix was added to each sample. The plates were covered, vortexed, pulse centrifuged and thermocycled on AB9700 cyclers. The PCR cycling conditions were: heat activation at 96ºC for 5 min, followed by 40 cycles of 94ºC for 30 sec, 60ºC for 45 sec, and 72ºC for 45 sec; final extension of 72ºC for 10 min, and cooled to 4ºC. The PCR reactions products were cleaned by adding 2µL of ExoSAP-IT® (USB Corporation), and incubating at 37ºC for 30 min followed by heat inactivation at 80ºC for 15 min.

The forward Sequencing Master Mix contains the M13 Forward primer and the reverse contains the M13 Reverse primer. The forward and reverse sequencing reaction mix was then prepared by mixing 4µL of BigDye® Terminator Mix v1.1 with 1µL of 3.2pmol/µL M13 forward or reverse primer, 3µL of deionised water. 8µL of the sequencing mix was added to 2µL of the PCR product. The sequencing

mixture was thermocycled at 96 ºC for 1 min, 25 cycles of 96 ºC for 10 sec, 50 ºC for 5 sec, and 60 ºC for 4 min, then cooled to 4 ºC. The sequencing reaction was cleaned-up by mixing 2.5µL of 125mM EDTA and 30µL of 100% ethanol with the sequencing products, and incubating at room temperature for 15 min. The DNA plate was subsequently centrifuged at 2500xg for 30 min at 4$^{o}$C. The supernatant was removed, 30 µL of 70% ethanol was added to each sample, and the plate was centrifuged upside-down at 1650xg for 15 min at 4$^{o}$C. The supernatant was removed, the plate was left to air dry, and the pellet was re-suspended in 10µL of Hi-Dye Formamide. The sequences were analysed (after performing electrophoresis on the 3730xl DNA Analyzer with POP-7™) with SeqScape® v2.5 software (Applied Biosystems).

## 2.12: Genotyping quality control

All genotyping for the association studies was conducted in 384-well plate format. Each plate contained at least one non-template negative test control (NTC) and twelve duplicate samples, which accounted for 3% of the total proportion of samples. Studies were excluded from analysis if the concordance between the duplicate samples was less than 98%. Genotyping of sample plates were either repeated, where possible, or excluded from analysis if the NTCs failed. For the oncogene and the BCAC, mismatch repair, cell cycle control, DNA repair pathways of the AML method, studies with call rates less than 90% were excluded from analysis. For the MMCT-18 study, and OCAC genotyping SNPs, the OCAC genotyping quality control guidelines were used. These guidelines encompassed the above, with the addition that 384-well DNA sample plates with call rates less than 90% were to be excluded from analysis.

## 2.13: Statistical methods

All statistical analysis was performed on the STATA version 8.2 statistical package (College Station, TX, www.stata.com).

### 2.13.1: Genetic Susceptibility

Due to the lack of apparent drive for evolution in humans, it is expected that populations within their respective ethnic groups are in Hardy-Weinberg equilibrium (HWE). Therefore, deviation from HWE was assessed in controls for each study population for each assay using standard $\chi^2$ test. The $\chi^2$ test measures the extent to which observed values differ from the expected proportion of genotypes (Norman and Streiner, 2008). For each population set, $\chi^2$ (1 degree of freedom) was used to assess deviations from genotype frequencies of the control subjects from those expected under Hardy-Weinberg equilibrium (HWE). For the polymorphisms analysed from the oncogene and the BCAC, mismatch repair, cell cycle control, DNA repair pathways, sample sets which were significantly out of HWE ($P<0.05$), had genotype clustering was evaluated and sample sets with good quality clustering were included in the analysis. For the genotyping data from the MMCT-18 candidate genes, the sample sets were excluded from analysis if they deviated from HWE at $P<10^{-4}$.

Logistic regression is a statistical model, which can be used for predicting the probability of the occurrence of an event, taking into account risk factors, for example, the chance of a person having a heart attack is dependent on their age, sex and body mass index. In genetic association studies, logistic regression is used to

produce a model to predict the probability that an individual will be affected by ovarian cancer, given their genotype. This is done by using the genotyping results to determine the frequency of each genotype in the cases and the controls and comparing the frequencies to ascertain whether there are significant differences between cases and controls for each genotype.

Associations between invasive epithelial ovarian cancer and each SNP were assessed using two tests; the one-degree of freedom Cochran–Armitage trend test and the general two-degrees of freedom $\chi^2$ test (heterogeneity test). The $\chi^2$ test for trend was stratified by study to account for any differences within the sample sets. Unconditional logistic regression was used to assess the relationship between each tSNP and risk of ovarian cancer for each population set, then the pooled samples (stratified by study) with the primary test of association being a test for trend (P-trend). The dependent variable (outcome/risk of ovarian cancer), can be estimated with the logistic regression model, which is generally written as:

$$\ln(\text{ODDS}) = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + .... + \beta_K\chi_K$$

which can be re-written, this study, as:

$$\text{Ln(risk of being affected)} = \beta_{\text{int}\,ercept} + \beta_{genotype} + \beta_{sample\_set}$$

A "Do-file", which contains a file with a list of commands for STATA to run when requested, was used to analysing groups of SNPs. However, the same results could be obtained by using the following STATA commands:

- xi:logistic status i.set i.SNP
- est sto A
- xi:logistic status i.set if(SNP!=.)
- lrtest A

The above commands are comprised of unconditional logistic regression with terms for disease status (whether case or control) and sample set with and without a term for genotype-study interaction. The likelihood ratio test (lrtest) assesses the fit of the model with genotypes nested within a model without a term for genotypes. The P-value <0.05 suggests that the null hypothesis that there is no association between the genotypes and risk of disease should be rejected.

Homogeneity between the samples sets was also assessed to ensure there were no statistically significant differences in the distribution of genotypes within the different population set. Homogeneity between studies was tested with likelihood ratio tests to compare the logistic regression models with and without a genotype-stratus interaction term. Statistically significant (P<0.05) heterogeneity between sample sets was usually caused by a study with different minor allele frequency for a particular polymorphism. Should this arise, the study would be excluded and the test for homogeneity repeated. If there was still statistically significant heterogeneity between studies, the genotyping data for the polymorphism would be excluded.

Trend tests are used for categorical data analysis. In genetic association studies, affected status is categorical (an individual is a case or a control); the genotypes are also categorical and ordered – for example, if the common allele at a SNP site is denoted as “A” and the rare allele is “a”, an individual is either homozygous for the common allele (AA), heterozygous (Aa) or homozygous for the rare allele (aa). The odds ratio (OR) is the ratio of the odds of an event (developing ovarian cancer) occurring in one group (homozygotes of the common allele [*y*]) compared with to

the odds of the same event occurring in another group (heterozygous [*z*]). The OR can be calculated with the formula

$$OR = \frac{y/(1-y)}{z/(1-z)}$$

Which can be simplified to:

$$OR = \frac{y(1-z)}{z(1-y)}$$

(Elston *et al.* 2002).

The ovarian cancer risks associated with being a heterozygote or a rare homozygote were estimated as OR with associated 95% confidence intervals by unconditional logistic regression with the common homozygote as the baseline comparator.

### 2.13.2:Haplotype definition and analysis

The confidence interval option (Gabriel *et al.* 2002) of the Haploview programme was used, with some minor adjustments to include adjacent SNPs, to define the haplotype blocks of the candidate genes. However, the cumulative frequency of the common haplotypes was maintain at >90%. Only tSNPs successfully genotyped were used to define the haplotype blocks, using the HapMap reference genotyping data. Quality control checks were performed on the haplotype frequencies per study, based on the genotype data.

The TagSNPs programme (Stram *et al*. 2003) was used to model multi-marker haplotypes from aggressive SNP tagging and also haplotypes of each gene when haplotype analysis was performed. TagSNPs implements an expectation-substitution

approach to account for the uncertainty caused by the unphased genotype data (Stram *et al*. 2003). The programme outputs estimates of the haplotype dosages for each haplotype for each individual, and logistic regression is performed on the estimates.

The genotyping data for nominally significant tSNPs were modelled with the log-additive, co-dominant, dominant and recessive genetic models and compared with likelihood ratio tests to ascertain the genetic model of best fit. This analysis was only conducted with associations between a SNP and ovarian cancer overall (not with histological subtypes).

The aetiology of ovarian cancer is very heterogeneous, and it has been demonstrated that mutation in particular genes are predominantly found in specific histological subtypes of the disease (Christie and Oehler 2006). Due to this heterogeneity, analysis was also restricted to the major histological subtypes of ovarian cancer (serous, endometrioid, mucinous and clear cell). Univariate unconditional logistic regression was used to test for associations of the germline polymorphisms and haplotypes of the candidate genes. The global effects of haplotypes of each gene or haplotype block were assessed with logistic regression and likelihood ratio tests. Models with and without the multiplicative effects of the haplotypes (minus the most common haplotype) were evaluated.

### 2.13.3: Admixture maximum likelihood test

The admixture maximum likelihood (AML) test is a method which was created for assessing the overall evidence for an excess of statistically significant associations

between the genetic polymorphisms and risk of diseases (such as breast or ovarian cancer) in case-control studies (Tyrer *et al*. 2006). The AML test was used to this effect with the genotyping data from candidate genes from DNA repair, mismatch repair, cell cycle control, oncogenes associated with ovarian cancer, differentially expressed genes from *in vitro* functional experiments, and candidate SNPs from the OCAC.

The AML method concurrently estimates the proportion of underlying false hypotheses, as well as testing the global null hypothesis of no association between the polymorphisms and risk of disease. The method does this by formulating the alternative hypothesis based on the probability ($\alpha$) that a given SNP is associated with disease and the estimated effect size of the polymorphism. The calculated $\chi^2$ statistic of a SNP associated with disease is distributed, asymptotically, as a non-central $\chi^2$ distribution with the usual degrees of freedom and a non-centrality parameter, $\eta$.

The non-centrality parameter, which is closely related to the contribution of the SNP to the genetic variance of the trait, is a measure of the size of effect of the polymorphism. The AML method assumes that the non-centrality parameter for all variants is the same, in order to make the model more parsimonious. Thus, the non-centrality parameter will be estimated. However, this is required because power will increase if the non-centrality parameter is the same for associated SNPs, as fewer parameters need to be optimised.

If η is assumed to be the same for each associated SNP, then α and η can be estimated by maximum likelihood, and a test of the null hypothesis, of no association, can then be obtained as a likelihood ratio test. In the instances where some variants were correlated, as many of the polymorphisms are in these studies, the same procedure can generate pseudo-maximum likelihood estimates, as if the germline variants are not correlated. The statistical significance of the AML test can then be determined by simulation testing. One thousand permutations were used to ascertain the significance of the AML test based on the ovarian cancer genotyping data.

The genomic control method for adjusting for cryptic population stratification was used on all polymorphisms analysed. Population stratification refers to the differences with in populations which may lead to false positive associations between genes or polymorphisms and disease risk. The genomic control approach, which is described in Devlin *et al*. (2001), estimates and takes into consideration the "over dispersion" of statistics used to evaluate association when there is population stratification. The genomic control approach involves estimating and taking into account the degree of over-dispersion caused by population stratification, by analysing polymorphisms, including some associated with disease risk, throughout the genome (Devlin *et al*. 2001).

Genotyping data from breast cancer case-control samples from the genome-wide association study (Easton *et al.* 2007; Hunter *et al.* 2007) were used to estimate the degree of over-dispersion of statistics, also known as inflation test statistic (Pharoah *et al*. 2007). The genotyping data consisted of 280 randomly selected, unlinked

polymorphisms from 4,037 breast cases and 4,012 controls. A more conservative inflation statistic (10%), than the one estimated from the genomic controls for the breast cancer study was used to adjust the P-trend for cryptic population stratification.

### 2.13.4: Survival analysis

There was a variable time between diagnosis and patient recruitment, therefore subjects were only considered to be at risk from the date of recruitment (blood draw). This provides an unbiased estimate of the relative hazard, provided that the proportional hazard assumption is not violated. The assumptions of proportional hazard are that the hazard ratios are constant over time and the hazard ratios are proportional within the different genotype groups across time. The survival period was defined as starting at date of blood draw.

All-cause mortality was the only end-point collected; censoring was at the date a participant was last known to be alive or at 10 years after diagnosis if the participant was still alive. Log-log survival curves were used to check that the assumptions of proportional hazards were met. The primary tests were likelihood ratio test for trend (1 degree of freedom), based on the number of rare alleles carried.

The Cox regression for survival analysis, stratified by study, was used to estimate the hazard ratio (HR) per rare allele carried. The Cox regression for survival analysis (also known as proportional-hazards regression) models the effect of variables (genotypes), over the time an event (death) takes to occur, or within a specified time

period. The model produces estimates of hazard ratios for the explanatory variables (genotypes).

The STATA commands for performing Cox regression survival analyses are as follows:

- stset yearoutcome, failure(outcome) enter(yearenter) exit(time 10)
- xi:stcox SNP i.study

the first command informs STATA of the names of the appropriate variable names which are essential for performing survival analysis (for example information regarding whether an event (death/"failure") has occurred can be found from the "outcome" variable; yearenter=([date of blood draw] – [date of diagnosis])/365.25; yearoutcome = ([date last seen or date of death] – [date of diagnosis])/365.25). Survival over 10 years was investigated. The second command runs the Cox regression test on the selected SNP, stratified by study.

The hazard ratios of all variables were adjusted for prognostic factors; age at diagnosis, tumour stage, tumour grade and histological subtype, where survival modelling showed that the prognostic factor significantly affected chances of survival. The inclusion of these prognostic factors in the survival models is known as multivariate survival analysis. These factors (age at diagnosis, tumour stage, tumour grade and histological subtype) are known to affect patient survival. The adjustments were made in order to observe whether the association remained after adjustments for known prognostic factors.

Univariate Cox regression analysis, stratified by study, was also used when the genotyping data was restricted to the four major histological subtypes of epithelial ovarian cancer (serous, endometrioid, mucinous and clear cell) to determine if the tSNPs were associated with survival in individuals with the particular histology. Clinical factors such as age, tumour stage and grade are known to affect survival. Therefore, they must be adjusted for in order to ascertain whether statistically significant associations were attributable to the SNP, and not to the clinical factors. The effects of clinical factors are discussed in more detail in Chapter 3. The effects of the clinical factors were tested with:

- xi: stcox i.agegroup i.grade i.stage i.set

the dummy variables generated by the command (for statistically associated factors) could be saved by renaming the variable name. For example:

- rename _Istage_2 stage2

Multivariate survival analysis was performed with terms for the statistically associated clinical factors. For example:

- xi:stcox SNP stage2 i.study

As with the susceptibility analysis, the TagSNPs programme was used to estimate haplotype dosages of each individual, for the survival analysis. The haplotype dose was based on the maximum likelihood of haplotypes of the candidates. Cox regression analysis, stratified by study, was also used to assess the effect of each haplotype dose on survival. In the STATA command, the name of the SNP was replaced with the name of the haplotype to assess its affect.

### 2.13.5: Kaplan-Meier survival estimates

Kaplan-Meier survival estimates are used to plot survival curves in order to illustrate survival over a period of time of the different groups (of genotypes, or clinical factors) being analysed and compared. The graphs can be generated in STATA with the command:

- xi: sts graph, by(var)

(var refers to variable, such as genotype, age at diagnosis, tumour histology, grade or stage).

The Kaplan-Meier survival estimator of the survival function at time *t* (or the probability of surviving up to time *t*) is calculated by the following formula:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

$\hat{S}(t)$ is the survival function at time *t*

$n_i$ corresponds to the number "at risk" just prior to time $t_i$,

and $d_i$, the number of deaths at time $t_i$ (Hosmer *et al*. 2008).

# Chapter 3: Results - The effects of common SNPs and haplotypes variants of oncogenes and functional candidate genes on the risk of ovarian cancer

## 3.1: Introduction

Hypothesis:

Common germline genetic variants in candidate genes connected with ovarian cancer development can influence the risk of epithelial ovarian cancer.

Aims:

(1) To determine if there is an effect of common variants and haplotypes of candidate oncogenes on the risk of invasive epithelial ovarian cancer.

(2) To determine if there is an effect of common variants and haplotypes of functional candidate genes (associated with neoplastic suppression of ovarian cancer cell lines) on predisposition to ovarian cancer.

(3) To use the admixture maximum likelihood test to assess if a significant number of associations have been found from ovarian cancer association studies.

Objectives:

(1) To use two-tailed unconditional logistic regression analysis to evaluate associations between common variants and haplotypes in 2 sets of candidate genes (oncogenes: *BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA*; and functional candidates:

*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) and risk of invasive epithelial ovarian cancer.

(2) To investigate associations between the variant and haplotypes of these genes on the 4 major ovarian cancer histological subtypes: serous, endometrioid, mucinous and clear cell.

(3) To use the admixture maximum likelihood experiment-wise test for association to evaluate the overall evidence of association between 340 common variants (in 74 genes and 10 regions without known genes or open reading frames) and risk of ovarian cancer.

## 3.2: Investigation of the effect of candidate oncogenes on risk of ovarian cancer

Oncogenes, such as *MYC*, *KRAS*, *BRAF* and *ERBB2* have been shown to be mutated or amplified in ovarian tumours. However, it is not known whether germline variants of the normal copies of these genes may predict a woman's risk of ovarian cancer. The following describes the selection of candidate oncogenes and the results of the analyses of the common polymorphisms and haplotypes of the genes on ovarian cancer risk.

### 3.2.1: Candidate oncogene and tSNP selection

The aim of the candidate oncogene selection was to find genes, with implications in ovarian cancer development, which would fit into a single iPLEX SNP multiplex genotyping run, and the minimum numbers of tSNPs remaining for genotyping by the TaqMan platform. The iPLEX genotyping platform can genotype up to 27 variants in a single reaction. Furthermore, the cost of genotyping with the TaqMan

platform increased with rising numbers of tSNPs which required genotyping with the platform, therefore, the number of SNPs which could not be genotyped on iPLEX had to be kept to a minimum. It became apparent that although *BCL2* and *KIT* were good candidate oncogenes for the study, too many tSNPs, 86 and 30, respectively, would need to be genotyped in order to genotype enough SNPs to gain genotyping data for the whole genes. Therefore, they were excluded from the selection. There was only 1 tSNP for *MUC1* (and 2 tSNPs for *MYC*), which suggested that the genes were either insufficiently tagged or there was not enough genotyping data available from HapMap for efficient tagging. Therefore, these genes were also excluded. Table 3.1 shows the number of tSNPs for all the candidate genes initially selected for tagging.

Table 3.1: Number of tagging SNPs of candidate oncogenes

| Gene | Total SNPs | No. criteria SNPs | No. tSNPs | |
|---|---|---|---|---|
| *AKT2* | 33 | 17 | 4 | |
| *BCL2* | 374 | 170 | 86 | Excluded |
| *BRAF* | 158 | 75 | 9 | |
| *CTNNB1* | 88 | 22 | 11 | |
| *ERBB2* | 16 | 6 | 3 | |
| *KIT* | 147 | 71 | 30 | Excluded |
| *KRAS* | 59 | 46 | 11 | |
| *MDM2* | 50 | 10 | 5 | |
| *MUC1* | 3 | 1 | 1 | Excluded |
| *MUC2* | 17 | 14 | 10 | |
| *MYC* | 15 | 2 | 2 | Excluded |
| *NMI** | 45 | 25 | 6 | |
| *PIK3CA* | 53 | 36 | 11 | |

Criteria SNPs – Minor allele frequency ≥0.05; HWE > 0.01; * NMI was erroneously selected from HapMap due to its interaction with *MYC* – its data is presented under *CMYC*, an alias of *MYC*. The mistake was not realised until the samples had been genotyped with iPLEX.

The National Center for Biotechnology Information (NCBI) Single Nucleotide Polymorphism website www.ncbi.nlm.nih.gov/projects/SNP/, also known as dbSNP, was used to find the Fasta sequence of each tSNP of *AKT2*, *MUC2*, *BRAF*, *KRAS*,

*NMI*, *PIK3CA*, *MDM2*, *ERBB2* and *CTNNB1*. The iPLEX Assay Design software was used to design panels based on four different combinations of candidate oncogenes (see Table 3.2).

Table 3.2: SNP panels from iPLEX assay design software

| Gene | No. tSNPs in panel | No. tSNPs not in panel | Total |
|---|---|---|---|
| **Panel 1** | | | |
| *ERBB2* | 2 | 1 | 3 |
| *KRAS* | 7 | 4 | 11 |
| *NMI* | 6 | - | 6 |
| *PIK3CA* | 11 | | 11 |
| **Panel 2** | | | |
| *BRAF* | 8 | 2 | 10 |
| *ERBB2* | 3 | - | 3 |
| *KRAS* | 7 | 4 | 11 |
| *NMI* | 7 | 2 | 9 |
| **Panel 3** | | | |
| *BRAF* | 6 | 3 | 9 |
| *ERBB2* | 3 | 0 | 3 |
| *KRAS* | 8 | 3 | 11 |
| *NM* | 5 | 1 | 6 |
| *PIK3CA* | 5 | 6 | 11 |
| **Panel 4** | | | |
| *ERBB2* | 2 | 1 | 3 |
| *MDM2* | 3 | 2 | 5 |
| *MUC2* | 7 | 3 | 10 |
| *KRAS* | 5 | 6 | 11 |
| *BRAF* | 5 | 5 | 10 |

Panel – refers to iPLEX panel.

The best iPLEX assay design pool, Panel 3, contained tSNPs of *BRAF*, *NMI*, *ERBB2*, *KRAS* and *PIK3CA*, which were also considered to be good candidates due to their known or predicted involvement in ovarian cancer development. Therefore, Panel 3 was selected for genotyping. The iPLEX assay pool comprised of 27 tSNPs. The best panel (with the most important genes and minimum number of tSNPs which would have to genotyped with TaqMan) was chosen.

### 3.2.2: Oncogenes - Samples and methods

The stage 1 samples consisted of GEOCS, MALOVA, SEARCH, UKOPS and USC (A) series of population-based studies. Stage 2 samples, comprising of the DOVE, HOPE and USC (B) studies, were used for the validation of significant findings from stage 1. The numbers of samples for each study are listed in Table 3.3.

Table 3.3: Samples used in oncogene study

| Study | Controls | Total cases | Histology | | | |
|---|---|---|---|---|---|---|
| | | | Serous | Endometrioid | Mucinous | Clear cell |
| GEOCS | 429 | 327 | 166 | 47 | 29 | 23 |
| MALOVA | 1221 | 446 | 275 | 56 | 43 | 33 |
| SEARCH | 855 | 730 | 254 | 130 | 94 | 62 |
| UKOPS | 271 | 116 | 65 | 20 | 10 | 12 |
| USC (A) | 224 | 197 | 115 | 22 | 16 | 8 |
| **Total stage 1** | **3000** | **1816** | **875** | **275** | **192** | **138** |
| DOVE | 716 | 584 | 303 | 86 | 18 | 30 |
| HOPE | 636 | 276 | 157 | 39 | 13 | 21 |
| USC (B) | 360 | 237 | 161 | 33 | 19 | 14 |
| **Total stage 2** | **1792** | **1097** | **621** | **158** | **50** | **65** |
| **Total stages 1 & 2** | **4713** | **2913** | **1496** | **433** | **242** | **203** |

All study individuals included in the analysis were non-Hispanic Whites.

In total, 40 tSNPs were selected to tag the common germline variants of the candidate oncogenes (*BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA*). A combination of iPLEX and TaqMan were used to genotype the first stage samples, and only Taqman was used to genotype second stage samples. Whole genome amplified samples were also genotyped on iPLEX – the results for these and problems with the iPLEX platform are discussed in Chapter 5.

### 3.2.3: Quality control

There were 12 duplicate samples per 384-well plate. Studies with less than 98% concordance between the total number of duplicate samples were excluded from analysis. Sample sets which had a call rate <90% for a SNP were also excluded from analysis. There were no studies which were out of Hardy-Weinberg equilibrium.

Of the 27 tSNPs genotyped on iPLEX, 2 (rs10842513 and rs6944385) failed quality control due to failed assays (failed PCR) and one was monomorphic (rs11047917). The GEOCS study performed extremely poorly on iPLEX, with only 5 tSNPs from the panel achieving call rates greater than 90%. The reason for the poor call rates may have been due to degraded non-amplified DNA, which had been plated and dried a year prior to the experimental work. These samples were used because the iPLEX manufacturers, Sequenom, advised the use of non-amplified DNA for the study, and these were the only remaining non-amplified DNA for GEOCS. Consequently, the iPLEX genotyping results from the GEOCS sample set were excluded from analysis.

TaqMan assays of all polymorphisms not genotyped on iPLEX were manufactured, where possible and GEOCS, UKOPS and USC (A) were genotyped with the TaqMan platform. Of the 40 tSNPs required to be genotyped in order to get full gene coverage, 34 were successfully genotyped with iPLEX and TaqMan. The six remaining tagging variants had either failed assay design, manufacture or probe testing and could not be efficiently genotyped by any other polymorphism.

### 3.2.4: Associations between candidate genes and ovarian cancer risk

Two-tailed unconditional logistic regression was used to determine the effect of common variants, and haplotypes of the oncogenes on risk of ovarian cancer. A SNP or haplotype is said to be associated with ovarian cancer susceptibility when there is a significant difference in the frequency of genotypes or haplotypes between cases and controls. Two-tailed unconditional logistic regression was used because no assumptions were made about the effect of the tSNP or haplotype prior to analysis – i.e. no assumptions were made about whether a SNP/haplotype would increase or decrease predisposition to ovarian cancer.

Odds ratios are used as a measure of the effect of the variant or haplotype on the risk of disease. An odds ratio <1 corresponds to a reduced risk of disease, and odds ratios >1 - increased risk. The genetics and histological pathology of ovarian cancer suggest different aetiologies for the histological subtypes of the disease. In order to establish the effects of the candidate oncogenes on the risk of serous, endometrioid, mucinous and clear cell subtypes, the logistic regression analysis was restricted to these particular subtypes. The results from these tests are below. The results reported below are of the stage 1, unless otherwise stated. All reported P-values are 2-tailed. The numbers of samples for each genotype group are shown in Appendices II-A and II-B for the oncogene and MMCT-18 candidate genes, respectively, and the genotype and haplotype-specific risks for all oncogenes are shown in Appendices III-A to III-J.

Initially, the associations between the risk of ovarian cancer and the common variants of candidate genes were assessed with all ovarian cancer cases pooled. The statistically significant association from this analysis was validated with addition samples.

There was no evidence of association between risk of ovarian cancer and common variants of *BRAF, ERBB2, KRAS* and *PIK3CA* when all cases were combined (Table 3.4). The rare allele of a tSNP in *NMI* (rs11683487) showed evidence of association with reduced risk of ovarian cancer (heterozygous odds ratio [OR] with 95% confidence intervals [CI] 0.80 [95% CI 0.69-0.93] homozygous OR 0.87 [0.71-1.02], P = 0.038). The HetOR is the odds ratio for individuals heterozygous for the variant, and HomOR is the odds ratio for individuals homozygous for the rare allele. Both HetOR and HomOR are compared with the homozygotes of the common allele. The association is a result of the rare allele of *NMI* rs11683487 being more frequent in the controls (46.3%) compared to cases (43.9%). There were 1,464 cases and 2,564 controls successfully genotyped for this SNP.

The genetic model of best fit for the *NMI* rs11683487 tSNP was a dominant model (rare allele carriers vs common allele homozygotes) because the odds ratios were similar in the heterozygotes and rare homozygotes. The similarity of the odds ratios of the heterozygotes and rare homozygotes suggested that this was a result of either a single or 2 copies of the rare allele. When the heterozygotes and rare homozygotes were grouped for analysis, the odds ratio for the dominant model was OR=0.81 (0.71- 0.94), P=0.004 for all cases analysed.

Table 3.4: Genotype-specific risks of pooled stage 1 oncogene data

| Gene | SNP name | Cases | Controls | MAF | HetOR§ | HomOR§ | *P*-Trend |
|---|---|---|---|---|---|---|---|
| *BRAF* | rs10487888 | 1680 | 2694 | 0.47 | 1.09 (0.93 - 1.28) | 1.02 (0.86 - 1.22) | 0.9 |
| | rs1733832 | 1159 | 2043 | 0.06 | 1.08 (0.86 - 1.36) | 3.39 (0.96 - 11.89) | 0.2 |
| | rs1267622 | 1751 | 2880 | 0.24 | 0.99 (0.87 - 1.12) | 0.97 (0.75 - 1.26) | 0.79 |
| | rs13241719 | 1602 | 2488 | 0.31 | 0.98 (0.85 - 1.12) | 0.86 (0.69 - 1.08) | 0.27 |
| | rs17695623 | 1744 | 2901 | 0.07 | 0.97 (0.81 - 1.16) | 1.14 (0.52 - 2.46) | 0.86 |
| | rs17161747 | 1771 | 2909 | 0.5 | 1.13 (0.93 - 1.38) | 1.29 (0.57 - 2.93) | 0.18 |
| | rs17623382 | 1764 | 2900 | 0.12 | 1.01 (0.87 - 1.17) | 1.01 (0.61 - 1.66) | 0.9 |
| | rs6944385 | 1758 | 2893 | 0.14 | 1.14 (0.99 - 1.32) | 0.99 (0.66 - 1.50) | 0.14 |
| | rs1267622, rs6944385; AA | 1786 | 2948 | 0.76 | 1.02 (0.79 - 1.33) | 1.04 (0.80 - 1.34) | 0.77 |
| *ERBB2* | rs2952155 | 1667 | 2678 | 0.24 | 1.01 (0.89 - 1.15) | 1.11 (0.84 - 1.47) | 0.57 |
| | rs2952156 | 1766 | 2912 | 0.29 | 0.97 (0.86 - 1.10) | 1.15 (0.89 - 1.49) | 0.74 |
| | rs1801200 | 1766 | 2916 | 0.22 | 1.04 (0.92 - 1.19) | 1.01 (0.77 - 1.31) | 0.64 |
| *KRAS* | rs12305513 | 1788 | 2934 | 0.1 | 0.87 (0.74 - 1.03) | 0.71 (0.38 - 1.31) | 0.053 |
| | rs12822857 | 1751 | 2901 | 0.47 | 1.01 (0.88 - 1.17) | 0.94 (0.80 - 1.12) | 0.53 |
| | rs10842508 | 1776 | 2935 | 0.25 | 0.97 (0.86 - 1.10) | 0.95 (0.73 - 1.22) | 0.57 |
| | rs12579073 | 1765 | 2900 | 0.48 | 0.97 (0.84 - 1.12) | 0.92 (0.78 - 1.09) | 0.36 |
| | rs10842513 | 1770 | 2878 | 0.09 | 1.03 (0.87 - 1.21) | 0.93 (0.50 - 1.74) | 0.86 |
| | rs4623993 | 1748 | 2892 | 0.16 | 0.96 (0.83 - 1.10) | 1.13 (0.77 - 1.67) | 0.85 |
| | rs6487464 | 1763 | 2895 | 0.38 | 1.04 (0.91 - 1.18) | 0.99 (0.82 - 1.19) | 0.94 |
| | rs10842514 | 1757 | 2886 | 0.44 | 0.98 (0.86 - 1.13) | 1.08 (0.91 - 1.29) | 0.42 |
| | rs11047917 | 1476 | 2456 | 0.06 | 0.92 (0.75 - 1.14) | 1.62 (0.57 - 4.57) | 0.71 |
| | rs4623993, rs12579073; TC | 1717 | 2818 | 0.1 | 0.96 (0.80 - 1.15) | 0.94 (0.56 - 1.57) | 0.63 |
| | rs12822857, rs10842508; AC | 1730 | 2857 | 0.23 | 0.99 (0.87 - 1.13) | 1.04 (0.80 - 1.36) | 0.93 |
| | rs12822857, rs10842514; GT | 1715 | 2806 | 0.4 | 1.04 (0.91 - 1.20) | 1.12 (0.94 - 1.34) | 0.23 |
| | rs12822857, rs12579073, rs6487464; GAC | 1689 | 2746 | 0.39 | 1.04 (0.89 - 1.21) | 1.06 (0.88 - 1.29) | 0.51 |
| *NMI* | rs394884 | 1708 | 2852 | 0.15 | 1.01 (0.88 - 1.17) | 1.40 (0.84 - 2.32) | 0.47 |
| | rs11551174 | 1159 | 2040 | 0.06 | 0.96 (0.76 - 1.23) | 1.23 (0.45 - 3.38) | 0.92 |
| | rs289831 | 1665 | 2718 | 0.13 | 1.05 (0.91 - 1.22) | 1.08 (0.61 - 1.89) | 0.48 |
| | rs3771886 | 1764 | 2927 | 0.41 | 1.03 (0.90 - 1.18) | **1.19 (1.00 - 1.42)** | 0.075 |
| | **rs11683487** | 1464 | 2564 | 0.46 | **0.80 (0.69 - 0.93)** | 0.87 (0.71 - 1.02) | **0.038** |
| | rs2113509 | 1776 | 2944 | 0.13 | 1.05 (0.91 - 1.21) | 1.16 (0.68 - 1.97) | 0.42 |
| *PIK3CA* | rs2865084 | 1164 | 2046 | 0.06 | 1.14 (0.89 - 1.45) | **0.43 (0.37 - 0.50)** | 0.29 |
| | rs7621329 | 1749 | 2818 | 0.16 | 0.99 (0.86 - 1.13) | 1.23 (0.86 - 1.77) | 0.64 |
| | rs1517586 | 1739 | 2908 | 0.1 | 0.98 (0.83 - 1.15) | 0.77 (0.42 - 1.40) | 0.54 |
| | rs2699905 | 1741 | 2855 | 0.27 | 1.01 (0.88 - 1.15) | 0.89 (0.71 - 1.11) | 0.49 |
| | rs7641889 | 1779 | 2939 | 0.07 | 0.89 (0.74 - 1.07) | 1.28 (0.58 - 2.84) | 0.38 |
| | rs7651265 | 1794 | 2883 | 0.1 | 0.89 (0.76 - 1.04) | 1.58 (0.89 - 2.80) | 0.54 |
| | rs7640662 | 1765 | 2916 | 0.15 | 1.02 (0.89 - 1.17) | 0.85 (0.57 - 1.27) | 0.86 |
| | rs2677760 | 1762 | 2925 | 0.49 | 1.01 (0.87 - 1.16) | 1.04 (0.88 - 1.23) | 0.67 |

MAF – minor allele frequency; Het: heterozygous; Hom – homozygous; OR= odds ratio (with associated 95% confidence intervals in parentheses); § compared with common homozygous; Emboldened tSNP, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Validation of association with *NMI* rs11683487 association

The association found between the rare allele of *NMI* rs11683487 and a decrease in the risk of invasive epithelial ovarian cancer was further investigated by performing a second stage of genotyping. Three additional populations from the USA (USC B); DOVE and HOPE), which comprised of an extra 1,097 cases and 1,712 controls were used in this second stage. There was no evidence of association between the rare allele of *NMI* rs11683487 and risk of ovarian cancer with the stage 2 samples alone (dominant model: OR= 0.01 [0.85-1.20]; $P_{dominant}$=0.92). When the data from both stages was subsequently combined and analysed with unconditional logistic regression the association with rs11683487 was weaker, but still statistically significant (OR= 0.89 [0.80– 0.99]; $P_{dominant}$ =0.0317; Table 3.5).

Table 3.5: The effect of *NMI* rs11683487 on the risk of ovarian cancer in Stages 1 & 2 cases

| Stage | Controls | Cases | Dominant model | |
|---|---|---|---|---|
| | | | **OR[§] (95% CI)** | **P-value** |
| 1 | 2564 | 1464 | **0.81 (0.71-0.94)** | **0.004** |
| 2 | 1712 | 1097 | 0.01 (0.85-1.2) | 0.92 |
| 1 & 2 | 4276 | 2561 | **0.87 (0.8-0.99)** | **0.0317** |

OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

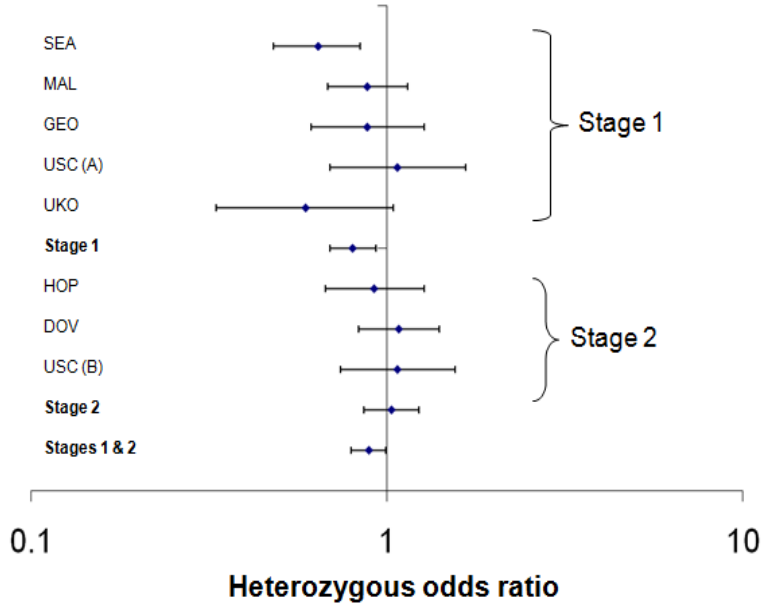
The forest plot in Figure 3.3 shows the effect size and the corresponding confidence intervals of the variant per study and stage(s). Forest plots are plotted on the natural logarithmic scale with the odds ratio represented by the diamond, the corresponding 95% confidence intervals shown as the horizontal lines, and no effect on risk symbolised by a vertical line. Odds ratios left of the vertical line (<1) imply a decrease in odds ratio, therefore protective effect of the rare variant, and those on the

right (>1) – increased risk of disease. Statistically significant associations have confidence intervals not crossing the vertical line.

The *NMI* SNP rs11683487 tags eight other SNPs with $r^2 > 0.8$ (rs3854012, rs3771882, rs4665150, rs1048135, rs11730, rs13004590, rs12987765 and rs17798290). rs11683487 tags rs1048135 with an $r^2=1$ (perfectly correlated). rs1048135 is a non-synonymous coding SNP and the rare (G) allele codes for a leucine instead of serine. The programme PMut (Ferrer-Costa *et al*. 2005), predicted that the rare allele (coding for leucine) had a pathological significance score of 3/10 and was classed as 'damaging' using the SIFT programme (Ng and Henikoff 2001; Ng and Henikoff 2002) . The bioinformatics tool, PupaSNP (http://pupasuite.bioinfo.cipf.es/) (Conde *et al.* 2006; Reumers *et al.* 2007) also suggested that this allele may disrupt the binding of exonic splicing enhancers. In addition, PupaSNP indicated that rs11683487 and rs11730 may have transcription and translation regulatory functions, and that rs11730 may affect exon splicing.

Combinations of SNPs, in a haplotype, have also been suggested ble to affect an individual's disease risk. A haplotype is a combination of alleles at multiple loci on the same chromosome, which are transmitted as a single unit (haplotype block). A haplotype block comprises a region of a chromosome which is unlikely to undergo recombination. Figure 3.2 shows the two haplotype blocks of *KRAS*, which was the only oncogene with more than 1 haplotype block. The variants shown in the figure were those genotyped in this study.

**A**




**B**

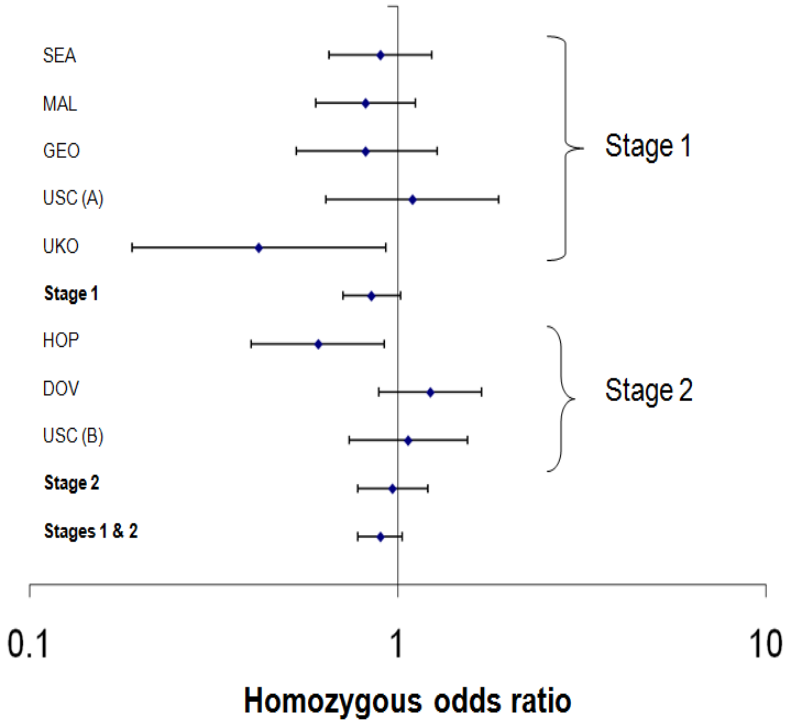



Figure 3.1: Forest plots of tSNP rs11683487 in the *NMI* gene in ovarian cancer case-control populations

In the haplotype analysis, the frequencies of the haplotypes in cases and healthy controls are compared to ascertain whether there is statistically significant difference. A common haplotype occurs in a population with a frequency of ≥5%. In haplotype analysis, these common haplotypes were analysed as an individual entity, and the rare haplotypes were grouped. The overall effect of the gene on disease risk was also assessed in the global analysis.

Haplotype block 1 of *KRAS* comprised of 3 tagging variants, spanning 6 kilobases (kb), and haplotype block 2 consisted of the 6 tSNPs over 20kb. It is important to note that the variants shown in the haplotypes are tagging SNPs, and therefore encompass 50 common SNPs captured with $r^2$=0.8 in the *KRAS* gene, (see Figure 3.2).

There were 4 common haplotypes of *KRAS* block 1, and 11 of haplotype block 2 (Figure 3.3). Within each representation of haplotype, the common allele is denoted by "0", and the rare by "1". The most frequent haplotype of *KRAS* block 1, h100, includes the rare allele of rs12305513, and the common alleles of rs12822857 and rs10842508, respectively. The "h" in front of the alleles represents "haplotype", and is used to distinguish between haplotypes and other strings of "0" and "1" which may occur in a document.

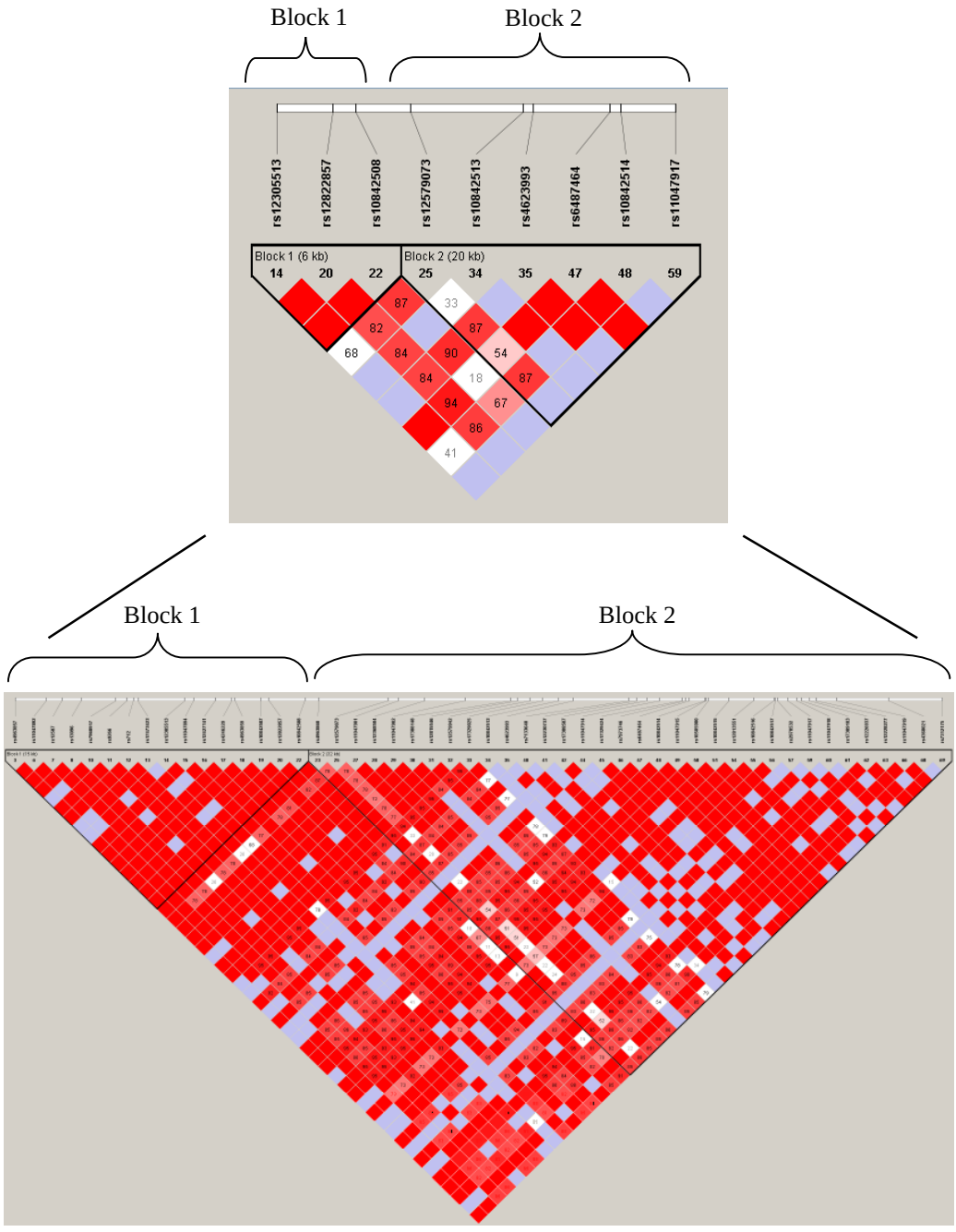



Figure 3.2: Haplotype blocks of *KRAS* SNPs genotyped

(a) Haplotypes blocks based on genotyped *KRAS* tSNPs; (b) the total number of SNPs in the haplotypes (captured by the tSNPs); colour scheme: standard (D'/LOD) – white (D'<1, LOD<2), shades of pink/red (D'<1, LOD≥2), blue (D'=1, LOD<2) and bright red (D'=1, LOD≥2), numbers shown in sqares (LD values) are based on D'.

**Block 1**

| rs12305513 | rs12822857 | rs10842508 | Freq (%) |
|---|---|---|---|
| 1 | 0 | 0 | 52.1 |
| 0 | 0 | 0 | 22.8 |
| 0 | 0 | 1 | 15.1 |
| 1 | 0 | 1 | 9.5 |

**Block 2**

| rs12579073 | rs10842513 | rs4623993 | rs6487464 | rs10842514 | rs11047917 | Freq (%) |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 0 | 30.6 |
| 1 | 0 | 0 | 0 | 1 | 0 | 12.9 |
| 1 | 0 | 0 | 1 | 0 | 0 | 11.9 |
| 1 | 0 | 1 | 1 | 0 | 0 | 10.6 |
| 0 | 1 | 0 | 0 | 0 | 0 | 5.5 |
| 0 | 0 | 0 | 1 | 0 | 0 | 5.4 |
| 1 | 0 | 0 | 0 | 0 | 0 | 5.4 |
| 0 | 0 | 1 | 1 | 0 | 0 | 4.7 |
| 1 | 0 | 0 | 1 | 0 | 1 | 4.3 |
| 1 | 1 | 0 | 0 | 0 | 0 | 3.2 |
| 0 | 0 | 0 | 0 | 0 | 0 | 3.1 |

Figure 3.3: Common haplotypes of *KRAS*

"0" – common allele of SNP; "1" – rare allele.

Associations between haplotypes of the oncogenes and risk of ovarian cancer were also evaluated. There was no evidence of association between risk of ovarian cancer and haplotypes of *KRAS* and *PIK3CA* (Table 3.6).

Statistically significant associations were found between common haplotypes of *BRAF*, *ERBB2* and *NMI* and risk of ovarian cancer at the 5% significance level. The h00001 haplotype of *NMI* correlated with a decreased risk of ovarian cancer (OR=0.91 [0.84-0.99], P=0.0276). Another haplotype of *NMI*, h00010, was also marginally associated with the risk of ovarian cancer (P=0.043; Table 3.6).

Table 3.6: Haplotype analysis results for *BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA*

| Gene | Haplotype | Freq (%) in controls | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|
| ***BRAF*** | h10000000 | 21.5 | 1.07 (0.97-1.19) | 0.182 | **0.005** |
| | h10010000 | 19.4 | 0.88 (0.76-1.01) | 0.07 | |
| | h00000000 | 17.3 | 0.90 (0.78-1.04) | 0.15 | |
| | h10010010 | 11.8 | 0.96 (0.82-1.12) | 0.57 | |
| | **h00100000** | 10.3 | **0.81 (0.68-0.95)** | **0.012** | |
| | h00101001 | 6.8 | 0.94 (0.78-1.13) | 0.49 | |
| | h01100001 | 6.1 | 1.15 (0.95-1.39) | 0.14 | |
| | h00000100 | 5.2 | 1.08 (0.88-1.31) | 0.48 | |
| ***ERBB2*** | h000 | 53.6 | 0.95 (0.87-1.04) | 0.284 | **0.034** |
| | **h110** | 16.3 | **1.19 (1.03-1.37)** | **0.016** | |
| | **h001** | 16 | **1.17 (1.02-1.34)** | **0.022** | |
| | h010 | 6.6 | 0.99 (0.82-1.20) | 0.9 | |
| | h111 | 6.5 | 0.84 (0.68-1.05) | 0.12 | |
| ***KRAS* haplotype block 1** | h100 | 52.1 | 1.02 (0.94-1.11) | 0.66 | 0.16 |
| | h000 | 22.8 | 1.00 (0.9 – 1.11) | 0.99 | |
| | h001 | 15.1 | 1.03 (0.91 – 1.16) | 0.67 | |
| | h101 | 9.5 | 0.89 (0.77 – 1.04) | 0.15 | |
| *KRAS* haplotype block 2 | h000010 | 30.6 | 1.04 (0.95-1.15) | 0.389 | 0.56 |
| | h100010 | 12.9 | 1.00 (0.85-1.17) | 0.98 | |
| | h100100 | 11.9 | 1.03 (0.88-1.19) | 0.75 | |
| | h101100 | 10.6 | 0.98 (0.84-1.14) | 0.77 | |
| | h010000 | 5.5 | 1.01 (0.82-1.25) | 0.91 | |
| | h000100 | 5.4 | 0.92 (0.73-1.18) | 0.52 | |
| | h100000 | 5.4 | 0.81 (0.65-1.01) | 0.06 | |
| | h001100 | 4.7 | 0.96 (0.74-1.24) | 0.74 | |
| | h100101 | 4.3 | 0.82 (0.64-1.05) | 0.11 | |
| | h110000 | 3.2 | 0.94 (0.70-1.27) | 0.69 | |
| | h000000 | 3.1 | 0.89 (0.64-1.23) | 0.48 | |
| ***NMI*** | **h00001** | 45.9 | **0.91 (0.84-0.99)** | **0.027** | **0.26** |
| | **h00010** | 33.7 | **1.11 (1.003-1.22)** | **0.043** | |
| | h10100 | 11.8 | 1.09 (0.95-1.25) | 0.22 | |
| | h01010 | 5.7 | 1.05 (0.84-1.3) | 0.67 | |
| *PIK3CA* | h00000001 | 48.2 | 1.02 (0.93-1.11) | 0.713 | 0.69 |
| | h00010010 | 14.8 | 1.01 (0.89-1.14) | 0.91 | |
| | h00000000 | 10.2 | 0.94 (0.81-1.09) | 0.39 | |
| | h00110000 | 9.7 | 0.98 (0.84-1.14) | 0.79 | |
| | h01001100 | 6.6 | 0.94 (0.79-1.12) | 0.5 | |
| | h01000100 | 4 | 1.02 (0.82-1.28) | 0.84 | |
| | h11000000 | 3.9 | 1.20 (0.97-1.48) | 0.102 | |

OR - odds ratio, CI - confidence interval; SNP order in haplotypes is 5' to 3' of the genes – *BRAF*: rs10487888, rs1733832, rs1267622, rs13241719, rs17695623, rs17161747, rs17623382, rs6944385; *ERBB2*: rs2952155, rs2952156, rs1801200; *KRAS* (block 1): rs12305513, rs12822857, rs10842508; *KRAS* (block 2): rs12579073, rs10842513, rs4623993, rs6487464, rs10842514, rs11047917; *NMI*: rs394884, rs11551174, rs289831, rs3771886, rs11683487; *PIK3CA*: rs2865084, rs7621329, rs1517586, rs2699905, rs7641889, rs7651265, rs7640662, rs2677760.

Although no associations were found between the tSNPs of *BRAF* and *ERBB2*, correlations were found with the combinations of tSNPs from these oncogenes and risk of ovarian cancer. There was an association between the h00100000 haplotype of *BRAF* and a reduced risk of all ovarian cancer cases (OR=0.81 [0.68- 0.95], P=0.012). Statistically significant *ERBB2* haplotype-effects were observed with the risk of ovarian cancer. Both associated haplotypes, h001 and h110, correlated with an increase in ovarian cancer risk when all cases were analysed (OR=1.17[1.02-1.34], P=0.022; and (OR=1.19 [1.03-1.37], P=0.016), respectively. The fact that both h110 and h001 were associated with an increased risk of ovarian cancer could not be explained because there was no common allele which could explain the associations found with h110 and h001; the haplotypes contained opposite alleles at all loci. HapMap genotype data was used to investigation whether these haplotypes shared a common untagged variant, however, this was without success. Globally, haplotypes of *BRAF* and *ERBB2* were associated with ovarian cancer susceptibility (P=0.005; and P=0.034, respectively).

Ovarian cancer is known to be heterogeneous, and the aetiologies of the four major histological subtypes are believed to be through different pathways. It is, therefore, of considerable interest to evaluate the effect of the oncogene variants and haplotypes on the risk of the serous, mucinous, endometrioid and clear cell histological subtypes of ovarian cancer. It is important to treat the results with caution because the numbers of the individual histological subtypes are far less than when all cases are grouped; therefore the power to detect true positive associations is reduced. The power to detect an association with the recessive, dominant and co-dominant genetic models with 3,000 controls, 875 serous ovarian cancer cases for a

SNP with MAF=0.46 and an effect size of 1.2 at the 5% significance level are 40%, 55% and 92%, respectively. Serous is the most common histological subtype, therefore the power to detect associations for the remaining common subtypes are further reduced. The results will be described on a gene-by-gene basis.

The association between the rare allele of *NMI* rs11683487 and reduced risk of ovarian cancer remained when the analysis was restricted to the serous and mucinous subtypes from the stage 1 data; (HetOR=0.81 [0.67-0.98], HomOR=0.80 [0.63-1.01], P=0.0377; HetOR=0.67 [0.47-0.96], HomOR=0.62 [0.39-0.99], P=0.0269) respectively. However, the associations were not as significant as for all cases - Table 3.7. As with all cases, the associations were not independently validated in with the stage 2 data, however the association with the mucinous subtype remained when the results from stages 1 and 2 were pooled.

The associations observed between the h00001 and h00010 haplotypes of *NMI* and risk of ovarian cancer also remained when the analysis was restricted to the serous subtype (see Table 3.8). The effect size was these associations were marginally stronger; however the significance was reduced, suggesting these are likely to be false positives. None the less, all of the associations found with the *NMI* haplotypes were supported by the single tSNP findings with *NMI* rs11683487, which was in the last position of the haplotype.

Table 3.7: The effect of *NMI* rs11683487 on the risk of ovarian cancer in Stages 1 & 2 cases

| Stage | Controls | Cases | Histology | Dominant model | |
|---|---|---|---|---|---|
| | | | | OR[§] (95% CI) | P-value |
| **Stage 1** | 2564 | 1464 | **All** | **0.81 (0.71-0.94)** | **0.004** |
| | | 713 | **Serous** | **0.8 (0.67-0.95)** | **0.0112** |
| | | 154 | **Mucinous** | **0.81 (0.68-0.96)** | **0.0163** |
| Stage 2 | 1712 | 1097 | All | 0.01 (0.85-1.2) | 0.92 |
| | | 711 | Serous | 1.02 (0.84-1.25) | 0.8319 |
| | | 50 | Mucinous | 1.38 (0.73-2.62) | 0.314 |
| **Stages 1 & 2** | 4276 | 2561 | **All** | **0.87 (0.8-0.99)** | **0.0317** |
| | | 1424 | Serous | 0.89 (0.78-1.02) | 0.0824 |
| | | 204 | **Mucinous** | **0.84 (0.71-0.99)** | **0.0419** |

MAF – minor allele frequency; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.8: Haplotype analysis results for *NMI* (P<0.05)

| Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|
| h00001 | 45.9 | **All** | **0.91 (0.84-0.99)** | **0.0276** | 0.26 |
| | | **Serous** | **0.89 (0.8-1)** | **0.048** | |
| | | Endometrioid | 1.05 (0.87-1.25) | 0.631 | |
| | | Mucinous | 0.84 (0.68-1.05) | 0.128 | |
| | | Clear cell | 0.96 (0.74-1.23) | 0.729 | |
| h00010 | 33.7 | **All** | **1.11 (1-1.22)** | **0.043** | |
| | | **Serous** | **1.13 (1.01-1.28)** | **0.041** | |
| | | Endometrioid | 0.9 (0.74-1.1) | 0.305 | |
| | | Mucinous | 1.11 (0.88-1.4) | 0.361 | |
| | | Clear cell | 0.97 (0.74-1.28) | 0.832 | |

OR - odds ratio, CI - confidence interval; in the haplotypes, †: '0'= common allele and '1'= rare allele; SNP order in haplotypes is 5' to 3' of the gene : rs394884, rs11551174, rs289831, rs3771886, rs11683487.

## Associations between *KRAS* and ovarian cancer susceptibility

When the effects of *KRAS* common variants and haplotypes on ovarian cancer risk were assessed, there was evidence suggesting that 3 variants of *KRAS* were associated with risk of the mucinous histological subtype of invasive epithelial

ovarian cancer. The frequency of the rare allele of rs10842514 was significantly greater in cases of mucinous subtype compared to healthy controls. The minor allele frequency (MAF) of the rare allele, "T", of the variant in the controls was 0.44, however, 0.54 the cases. The rare allele of rs10842514 was associated with a 2.02-fold increase in the risk of the mucinous subtype (heterozygous odds ratio [OR] with 95% confidence intervals [CI] =1.13 [95% CI 0.78-1.64], homozygous OR=2.02 [1.35-3.01], P-trend = $6x10^{-4}$), see Table 3.9. This tSNP is in intron 2 of *KRAS*, and to date, it does not tag any other SNP within or flanking *KRAS*. The association was not found with other subtypes of ovarian cancer or all subtypes combined, therefore if the association is true positive, it is unique to the mucinous subtype.

An association was also found between the mucinous histological subtype and two other polymorphisms of *KRAS*. Those homozygous for the rare allele of rs12822857 had a reduced risk of mucinous ovarian cancer. The heterozygous genotype of rs6487464 was also associated with a 0.61-fold reduction in the risk of the subtype in individuals heterozygous for the tSNP; see Table 3.9. However, it is still unknown how mutations or variation of *KRAS* may result in the mucinous subtype. The results for all the tSNPs genotyped for *KRAS* can be found in Appendix III-E.

The associations between these variants and risk of mucinous subtype of ovarian cancer are of particular interest because mutations in *KRAS* have previously been reported to be associated with this subtype. Gemignani *et al* found that of 22 mucinous ovarian carcinomas, 50% had mutations in the *KRAS* oncogene, compared with 4 (5%) of 82 non-mucinous tumours (Gemignani *et al*, 2003).

Table 3.9: Genotype-specific risks of *KRAS* tSNPs (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR (95% CI)* | HomOR (95% CI)* | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| *KRAS* | rs12822857 | 0.47 | 2901 | 1751 | All | 1.01 (0.88-1.17) | 0.94 (0.80-1.12) | 0.5281 |
| | | | | 835 | Serous | 1.08 (0.90-1.30) | 0.97 (0.78-1.20) | 0.8167 |
| | | | | 268 | Endometrioid | 0.99 (0.73-1.33) | 1.06 (0.75-1.50) | 0.7605 |
| | | | | 187 | **Mucinous** | 0.74 (0.53-1.04) | **0.63 (0.41-0.96)** | **0.0232** |
| | | | | 132 | Clear cell | 1.14 (0.75-1.74) | 1.04 (0.63-1.72) | 0.8398 |
| | rs6487464 | 0.38 | 2895 | 1763 | All | 1.04 (0.91-1.18) | 0.99 (0.82-1.19) | 0.9408 |
| | | | | 836 | Serous | 1.09 (0.92-1.29) | 0.98 (0.77-1.24) | 0.8783 |
| | | | | 269 | Endometrioid | 1.13 (0.86-1.5) | 1.15 (0.79-1.67 | 0.3878 |
| | | | | 192 | **Mucinous** | **0.61 (0.44-0.85)** | 0.76 (0.50-1.18) | **0.0379** |
| | | | | 136 | Clear cell | 1.02 (0.70-1.48) | 0.95 (0.56-1.61) | 0.8918 |
| | rs10842514 | 0.44 | 2886 | 1757 | All | 0.98 (0.86-1.13) | 1.08 (0.91-1.29) | 0.4153 |
| | | | | 835 | Serous | 1.04 (0.87-1.24) | 1.09 (0.88-1.36) | 0.4379 |
| | | | | 269 | Endometrioid | 0.85 (0.64-1.13) | 0.97 (0.68-1.38) | 0.7294 |
| | | | | 188 | **Mucinous** | 1.13 (0.78-1.64) | **2.02 (1.35-3.01)** | **0.0006** |
| | | | | 134 | Clear cell | 0.95 (0.64-1.40) | 0.79 (0.47-1.33) | 0.4026 |

MAF: minor allele frequency; Cons: controls; Het: heterozygous; Hom: homozygous; OR: odds ratio; CI: confidence interval, * compared with common homozygote. Enboldened OR and P-trend values are statistically significant.

Table 3.10: Haplotype analysis results for *KRAS* (P<0.05)

| Gene/ block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *KRAS* haplotype block 2 | h000010 | 30.6 | All | 1.04 (0.95-1.15) | 0.389 |
| | | | Serous | 1.07 (0.94-1.21) | 0.306 |
| | | | Endometrioid | 0.99 (0.81-1.21) | 0.916 |
| | | | **Mucinous** | **1.3 (1.03-1.64)** | **0.025** |
| | | | Clear cell | 0.79 (0.59-1.06) | 0.121 |
| | h000000 | 3.1 | All | 0.89 (0.64-1.23) | 0.48 |
| | | | Serous | 1.08 (0.75-1.58) | 0.67 |
| | | | **Endometrioid** | **0.36 (0.14-0.92)** | **0.033** |
| | | | **Mucinous** | **0.3 (0.09-0.99)** | **0.049** |
| | | | Clear cell | 0.61 (0.21-1.78) | 0.365 |

OR - odds ratio, CI - confidence interval, In the haplotypes, †: '0'= common allele and '1'= rare allele; SNP order in haplotypes is 5' to 3' of the genes –*KRAS* (block 2): rs12579073, rs10842513, rs4623993, rs6487464, rs10842514, rs11047917

When the frequencies of the haplotypes of *KRAS* haplotype block 1 and 2 were compared between cases and controls, statistically significant differences were found between 2 haplotypes of *KRAS* block 2 and the risk of the mucinous subtype of ovarian cancer. The h000010 haplotype of *KRAS* block 2 was associated with a 1.3-fold increase in the risk of the mucinous subtype (OR=1.3 (1.03-1.64), P=0.025. Conversely, the h000000 haplotype of the same block was associated with a decreased risk of the subtype, OR=0.3 (0.09-0.99), P=0.049. The associations between h000010 and h000000 and risk of mucinous ovarian cancer are supported by the presence of the common allele of rs6487464 in the fourth position of the haplotype, and the rare allele of rs10842514 in the fifth position. This is because the rare allele of rs6487464 was associated with reduced risk, and the rare allele of rs10842514 with increased risk. However, this haplotype is not the only one with this combination of alleles of rs6487464 and rs10842514. Assuming that the association found between the rare allele of rs10842514 and increased risk of the mucinous subtype is a true positive, it is feasible that the associations between the

h000000 and h000010 haplotypes of *KRAS* block 2 were due, primarily to the alleles of rs10842514.

Analysis of the effect of haplotype on disease risk can sometimes elucidate associations which could not be identified through individual SNP analysis. It is believed that the combinations of alleles, rather than individual variant may affect disease risk. Such an association was found between the h000000 haplotype of *KRAS* and reduced risk of endometrioid ovarian cancer. This association was also found with the mucinous histological subtype. The association between h000000 and risk of the endometrioid subtype was marginally more significant than the mucinous subtype (OR=0.36 [0.14-0.92], P=0.033), with similar odds ratio (Table 3.10 – page 145). See Appendix III-F for the results for all the common and combined rare haplotypes of *KRAS*.

**Associations between *BRAF* and ovarian cancer susceptibility**

There was evidence suggesting that three variants of *BRAF* also influenced the risk of mucinous histological subtype of invasive epithelial ovarian cancer. The rare alleles of two of the variants, rs1267622 and rs17695623, were associated with a decreased risk of the subtype (see Table 3.11). The rare allele of *BRAF* rs10487888, conversely, was associated with an increased risk of the subtype; HetOR=1.32 (0.86-2.03), HomOR=1.61 (1.03-2.53), P=0.0357 (Table 3.11). The associations with these variant were also unique to the subtype. The results for all the *BRAF* common variants can be found in Appendix III-A.

There were 8 common (haplotype frequency >5% in control samples) *BRAF* haplotypes (see Appendix III-A). Analysis showed that a haplotype with a 6.8% frequency was associated with a reduced risk of mucinous ovarian cancer (OR=0.54 [0.31-0.93], P=0.027). The *BRAF* tSNPs, rs10487888, rs1267622 and rs17695623 are in the first, third and eighth positions, respectively, of the haplotype. rs10487888 tags rs1267622 with $r^2$=0.318, and rs17695623 with $r^2$=0.097; and rs1267622 and rs17695623 tag each other with $r^2$=0.306.

The haplotype showing significant association with mucinous ovarian cancer contained the common allele of rs10487888, and the rare alleles of rs1267622 and rs17695623, both of which were associated with decreased risk (Table 3.11), which may explain the association. The combination of the alleles supported a reduction in risk of the mucinous subtype. However, this combination of alleles of the above variants also occurred in another haplotype of *BRAF* (h01100001), which had an odds ratio of less than 1, but was not significantly associated with risk of mucinous ovarian cancer (OR=0.93 [0.58-1.48], P=0.76); Table 3.12.

There was an association between the h00100000 haplotype of *BRAF* and a reduced risk of all ovarian cancer cases (OR=0.81 [0.68- 0.95], P=0.012). The association with this haplotype, which occurred at a frequency of 10.3% in controls, remained when analysis was restricted to serous only cases – OR=0.8 (0.66-0.98), P=0.028 - Table 3.12. Although there were no other statistically significant associations of this haplotype and the other histological subtypes of ovarian cancer, the odds ratios for endometrioid, mucinous and clear cell histological subtypes were similar to that of serous (Table 3.12).

Table 3.11: Genotype-specific risks of common *BRAF* tSNPs (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR (95% CI) | HomOR (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| *BRAF* | rs10487888 | 0.47 | 2694 | 1680 | All | 1.09 (0.93-1.28) | 1.02 (0.86-1.22) | 0.902 |
| | | | | 804 | Serous | 1.21 (0.99-1.49) | 1.09 (0.87-1.37) | 0.5747 |
| | | | | 251 | Endometriod | 0.84 (0.61-1.17) | 0.88 (0.61-1.25) | 0.5007 |
| | | | | 180 | **Mucinous** | 1.32 (0.86-2.03) | **1.61 (1.03-2.53)** | **0.0357** |
| | | | | 125 | Clear cell | 1.04 (0.66-1.64) | 0.88 (0.52-1.48) | 0.5947 |
| | rs1267622 | 0.24 | 2880 | 1751 | All | 0.99 (0.87-1.12) | 0.97 (0.75-1.26) | 0.7894 |
| | | | | 831 | Serous | 1.04 (0.88-1.22) | 0.90 (0.64-1.26) | 0.9055 |
| | | | | 268 | Endometriod | 0.76 (0.58-1.01) | 1.06 (0.63-1.77) | 0.2725 |
| | | | | 187 | **Mucinous** | **0.67 (0.48-0.94)** | 0.71 (0.35-1.43) | **0.0278** |
| | | | | 135 | Clear cell | 1.24 (0.87-1.78) | 1.12 (0.53-2.37) | 0.3392 |
| | rs17695623 | 0.07 | 2901 | 1744 | All | 0.97 (0.81-1.16) | 1.14 (0.52-2.46) | 0.8642 |
| | | | | 829 | Serous | 1.04 (0.83-1.31) | 1.19 (0.45-3.10) | 0.6437 |
| | | | | 264 | Endometriod | 0.99 (0.68-1.45) | 0.63 (0.08-4.85) | 0.8421 |
| | | | | 186 | **Mucinous** | **0.47 (0.26-0.86)** | 0.79 (0.10-6.08) | **0.0191** |
| | | | | 135 | Clear cell | 1.25 (0.78-2.03) | 1.37 (0.18-10.56) | 0.3393 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.12: Haplotype analysis results for *BRAF* (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | OR± (95% CI)† | P-value |
|---|---|---|---|---|---|
| *BRAF* | h10010000 | 19.4 | All | 0.88 (0.76-1.01) | 0.07 |
| | | | Serous | 0.99 (0.86-1.15) | 0.921 |
| | | | Endometrioid | 1.04 (0.82-1.31) | 0.769 |
| | | | Mucinous | 0.92 (0.69-1.22) | 0.547 |
| | | | **Clear cell** | **0.67 (0.46-0.97)** | **0.033** |
| | h00100000 | 10.3 | **All** | **0.81 (0.68-0.95)** | **0.012** |
| | | | **Serous** | **0.8 (0.66-0.98)** | **0.028** |
| | | | Endometrioid | 0.88 (0.64-1.21) | 0.439 |
| | | | Mucinous | 0.87 (0.59-1.27) | 0.463 |
| | | | Clear cell | 0.86 (0.55-1.34) | 0.508 |
| | h00101001 | 6.8 | All | 0.94 (0.78-1.13) | 0.49 |
| | | | Serous | 1.09 (0.88-1.35) | 0.444 |
| | | | Endometrioid | 0.99 (0.69-1.42) | 0.947 |
| | | | **Mucinous** | **0.54 (0.31-0.93)** | **0.027** |
| | | | Clear cell | 1.36 (0.87-2.11) | 0.175 |

OR - odds ratio; CI - confidence interval; in the haplotypes, '0'= common allele and '1'= rare allele; SNP order in haplotypes is 5' to 3' of the genes – *BRAF*: rs10487888, rs1733832, rs1267622, rs13241719, rs17695623, rs17161747, rs17623382, rs6944385.

Another haplotype of *BRAF*, h10010000, was associated with a reduced risk of the clear cell subtype – OR=0.67 (0.46-0.97), P=0.033. The combined rare haplotypes of *BRAF* were also associated with a decrease in the risk of serous ovarian cancer (P=0.038; see Table 3.12), however the frequency of the combined rare haplotype was less than 1. Therefore only a small number of samples with the rare haplotypes would have had a reduced risk of the subtype, if the association is a true positive.

## Associations between *ERBB2* and ovarian cancer susceptibility

Of the 3 common tagging variants of ERRB2 genotyped, there was evidence of association between rs1801200 and risk of the endometrioid subtype of ovarian cancer. The rare allele of *ERBB2* rs1801200 was associated with an increased risk of endometrioid disease (HetOR=1.16 [0.88-1.52], HomOR=1.71 [1.05-2.76],

P=0.0389; Table 3.13). This variant was not statistically associated with all cases or the other subtypes of ovarian cancer. The heterozygous risk for clear cell cases was OR=1.51, and the confidence intervals did not cross 1, (1.05-2.17), however this correlation was not statistically significant, although the P-value was close to significance at the 5% level (P=0.0564). The results of the *ERBB2* variants are in Appendix III-C.

The *ERBB2* rs1801200 polymorphism is a non-synonymous coding SNP. The "A" allele, which is the major form of the SNP, encodes isoleucine, while the "G" allele encodes valine. The polymorphism is conserved in mice and the sequence is also predicted to be an exonic splicing enhancer.

Statistically significant *ERBB2* haplotype-effects were observed with the risk of ovarian cancer – see Appendix III-D for the results of all the *ERBB2* haplotypes. h001 and h110, which were associated with the risk of all subtypes combined were also clear cell and mucinous subtypes, respectively, see Table 3.14.

The *ERBB2* variant which was associated with ovarian cancer risk, rs1801200, was in the last position of the haplotypes. The fact that both h110 and h001 were associated with an increase in risk of ovarian cancer could not be explained by this variant because it would suggest that the 2 different alleles both result in increased risk of disease. The other loci of the haplotype also contained opposite alleles, therefore, there was no common allele which could explain the associations found with h110 and h001.

Table 3.13: Genotype-specific risks of common *ERBB2* tSNPs (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-Trend |
|---|---|---|---|---|---|---|---|---|
| *ERBB2* | rs1801200 | 0.22 | 2916 | 1766 | All | 1.04 (0.92-1.19) | 1.01 (0.77-1.31) | 0.6401 |
| | | | | 847 | Serous | 1.04 (0.88-1.22) | 0.86 (0.60-1.23) | 0.8257 |
| | | | | 263 | **Endometriod** | 1.16 (0.88-1.52) | **1.71 (1.05-2.76)** | **0.0389** |
| | | | | 188 | Mucinous | 0.79 (0.57-1.11) | 0.82 (0.41-1.66) | 0.2007 |
| | | | | 134 | Clear cell | **1.51 (1.05-2.17)** | 1.30 (0.61-2.76) | 0.0564 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.14: Haplotype analysis results for *ERBB2* (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | OR± (95% CI)† | P-value* |
|---|---|---|---|---|---|
| *ERBB2* | h110 | 16.3 | **All** | **1.19 (1.03-1.37)** | **0.016** |
| | | | Serous | 1 (0.85-1.19) | 0.964 |
| | | | Endometrioid | 1 (0.75-1.34) | 0.982 |
| | | | **Mucinous** | **1.39 (1.02-1.9)** | **0.036** |
| | | | Clear cell | 0.94 (0.63-1.4) | 0.752 |
| | h001 | 16 | **All** | **1.17 (1.02-1.34)** | **0.022** |
| | | | Serous | 1.08 (0.92-1.27) | 0.329 |
| | | | Endometrioid | 1.25 (0.97-1.62) | 0.079 |
| | | | Mucinous | 0.88 (0.63-1.23) | 0.466 |
| | | | **Clear cell** | **1.6 (1.15-2.21)** | **0.005** |

OR - odds ratio, CI -confidence interval; in the haplotypes: '0'= common allele and '1'= rare allele; SNP order in haplotypes: rs2952155, rs2952156, rs1801200.

Both the size of the effect and the significance between the h001 haplotype of *ERBB2* and increased risk of ovarian cancer became stronger when analysis was restricted to the clear cell subtype only (OR=1.6 [1.15-2.21], P=0.005). The mucinous subtype was associated with the other significant haplotype, h110, OR=1.39 (1.02-1.9), P=0.036 (see Table 3.14).

**Associations between *PIK3CA* and ovarian cancer susceptibility**

There was no evidence of association between tSNPs *PIK3CA* and susceptibility to invasive epithelial ovarian cancer when all the cases of stage 1 samples were analysed (see Appendix III-I). However, a statistically significant association was found between the rare allele of rs2865084 and risk of the endometrioid subtype when analysis was restricted to this subtype, – HetOR=1.6 (1.03-2.5), HomOR=0.3 (0.22-0.42), P=0.0344; Table 3.15. Interestingly, for all but the clear cell subtype, there was a suggestion that the heterozygotes for the tSNP had an increased risk of disease, however all rare homozygotes had a reduced risk of the disease (the confidence interval did not cross 1) – Appendix III-I. The odds ratios for all cases and the all individual subtypes did not cross 1, which suggest an association with rare homozygosity of the variant despite P>0.05 for all but the endometrioid samples – see Table 3.15. *PIK3CA* rs2865084 tags 4 other SNPs. All of these SNPs are intronic, however they are all conserved. rs2865084 is in a transcription factor binding site, upstream of the gene. Analysis of the *PIK3CA* haplotypes showed that the h11000000 haplotype was associated with a marginally increased risk of the endometrioid subtype (P<0.05, Table 3.16). However, corrections for multiple testing would attenuate the association. The odds ratios for all the *PIK3CA* haplotypes are shown in Appendix III-J.

Table 3.15: Genotype-specific risks of common *PIK3CA* tSNPs (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-Trend |
|---|---|---|---|---|---|---|---|---|
| *PIK3CA* | rs2865084 | 0.06 | 2046 | 1164 | All | 1.14 (0.89-1.45) | **0.43 (0.37-0.50)** | 0.294 |
| | | | | 525 | Serous | 1.13 (0.83-1.55) | **0.77 (0.63-0.93)** | 0.428 |
| | | | | 183 | **Endometrioid** | **1.60 (1.03-2.50)** | **0.30 (0.22-0.42)** | **0.034** |
| | | | | 135 | Mucinous | 1.32 (0.77-2.25) | **0.32 (0.22-0.46)** | 0.309 |
| | | | | 95 | Clear cell | 0.51 (0.21-1.28) | **0.37 (0.24-0.57)** | 0.147 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.16: Haplotype analysis results for *PIK3CA* (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *PIK3CA* | h11000000 | 3.9 | All | 1.20 (0.97-1.48) | 0.102 | 0.69 |
| | | | Serous | 1.29 (0.99-1.67) | 0.055 | |
| | | | **Endometrioid** | **1.49 (1-2.22)** | **0.049** | |
| | | | Mucinous | 1.07 (0.63-1.82) | 0.795 | |
| | | | Clear cell | 0.91 (0.47-1.79) | 0.793 | |

OR - odds ratio; CI - confidence interval, Freq=frequency in controls; In the haplotypes, '0'= common allele and '1'= rare allele; SNP order in haplotypes is 5' to 3' of *PIK3CA*: rs2865084, rs7621329, rs1517586, rs2699905, rs7641889, rs7651265, rs7640662, rs2677760.

## 3.3: The Effect of tagging SNPs and haplotypes of functional candidate genes on risk of ovarian cancer

Deletions on chromosomes 4, 5, 6, 13, 14, 15 and 18 have frequently been observed in primary ovarian cancer cell lines with the metaphase comparative genomic hybridisation procedure. There is also strong evidence from *in vitro* assays suggesting that the incorporation of a normal chromosome 18 into 2 ovarian cancer cancer cell lines can lead to the suppression of the neoplastic phenotype (Dafou *et al.* 2009). The micro-cell mediated chromosome transfer of a normal chromosome 18 (MMCT-18) has also been demonstrated to result in the suppression of the tumourigenic phenotype in prostate and pancreatic cancer cell lines (Padalecki *et al.* 2001; Lefter *et al.* 2002; Gagnon *et al.* 2006). These observations suggest that chromosome 18 harbours tumour suppressor genes which may contribute to these suppressions (Lefter *et al.* 2004; Dafou *et al.* 2009).

Although there has been some degree of success in identifying germline polymorphisms associated with ovarian cancer predisposition with the candidate gene approach, almost all associations have been borderline significant with small effect sizes. It is possible that the limited success may be because the candidate gene selection process is usually based on a predicted function or mutations found in tumours rather than on "functional" evidence. The microcell-mediated chromosome transfer (MMCT) technique was used in the hope of identifying putative functionally relevant genes involved in ovarian cancer aetiology.

The MMCT method was used to transfer normal human chromosome 18 (MMCT-18) into ovarian cancer cell lines. *In vitro* and *in vivo* assays were used to ascertain

the effect of the transferred chromosome on the neoplastic phenotype of the resultant recipient/donor hybrids. Differences in gene expression were assessed between the parental cancer cell lines, and the neoplastically suppressed hybrid cells, in order to establish if the incorporation of chromosome 18 resulted in a phenotypic change which could be correlated to biological and/or molecular function.

### 3.3.1: Gene and tSNP selection of functional candidate genes

The global differential gene expression results from the micro-cell mediated chromosome 18 transfer (MMCT-18) were obtained for each cell line containing columns for: gene ID, probe ID, gene name, cytoband, fold change (hybrid/parental), $\log_2$ of the fold change and P-value. Although the gene expression results were not validated with quantitative (Q)-PCR, the experiments were conducted in triplicate and the results corrected for multiple testing. The gene lists contained data on 32,878 probes for 32,000 genes; there was more than 1 probe for some genes due to different transcripts for the genes being available. Genes from all chromosomes were analysed because genes have dynamic interactions with others throughout the whole genome. It is feasible that although a gene on chromosome 18 may not directly affect tumourigenicity, the gene may interact with or regulate other genes which may have an effect on the suppression of tumour growth.

For the selection of candidate genes, the gene list from the parental and hybrid cell lines of TOV112D was combined with the gene list from TOV21G expression data, matching gene and probe IDs. A simple colour scheme was created for fold change for easy visualisation and data sorting. Green corresponded to down-regulation of gene expression in hybrids in comparison with their respective parental cell line, and

red for up-regulation of gene expression of hybrids. Probes with missing data for both cell lines were excluded. Genes with inconsistent fold change directions within and between cell lines were also excluded. For probes where there were missing data for 1 of the parental/hybrid cell lines, the data from the other parental/hybrid trio was used. Gene lists were created for genes which were:

- up-regulated in the hybrids of both cell lines (62 genes);
- down-regulated in the hybrids of both cell lines (993 genes);
- up-regulated in TOV21G hybrids (264 genes);
- down-regulated in TOV21G hybrids (1089 genes);
- up-regulated in TOV112D hybrids (72 genes);
- down-regulated in TOV112D hybrids (312 genes);
- up-regulated in TOV21G hybrids chromosome 18 only (31 genes);
- down-regulated in TOV21G hybrids chromosome 18 only (44 genes);
- up-regulated in TOV112D hybrids chromosome 18 only (24 genes);
- down-regulated in TOV112D hybrids chromosome 18 only (53 genes);
- up-regulated in TOV112D hybrids chromosome 18 break-point region (5 genes);
- down-regulated in TOV112D hybrids chromosome 18 break-point region (8 genes).

Figure 3.4 shows a flow chart of the candidate gene selection process.

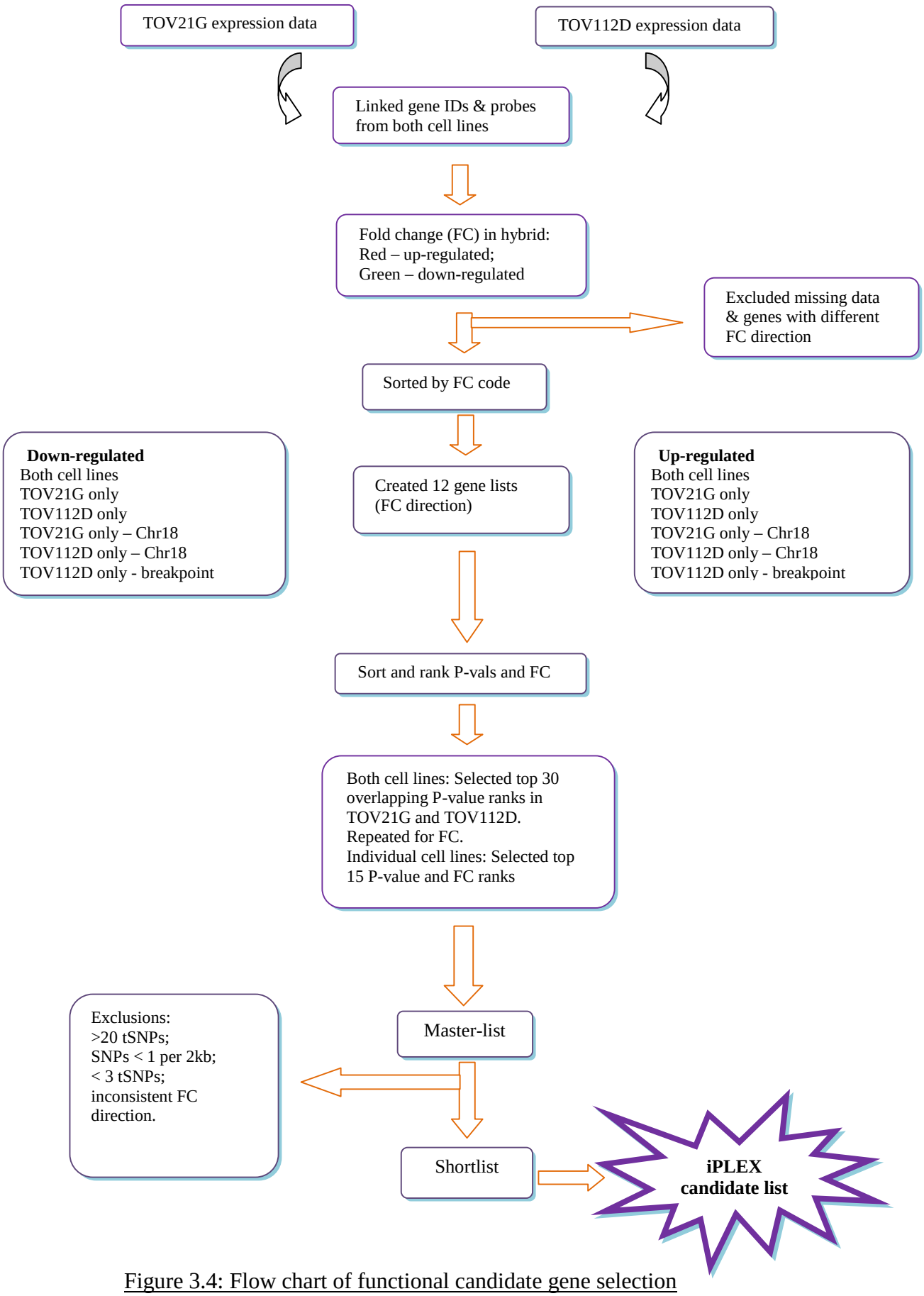



Figure 3.4: Flow chart of functional candidate gene selection

The genes from these lists were ranked by p-values and differential gene expression fold change. From the gene lists with concordant fold change in both cell-lines, the top 30 ranking P-values and fold changes from both cell lines were selected for the master list. From the other gene lists, the top 15 ranking genes according to P-value and fold change were selected for the master list. There were 192 genes in the master list (Appendix I), of which there were some genes that were duplicated from the different lists. The gene size and functions for the genes in the master list were obtained from Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/entrez) and Genecards (http://genecards.org/). The genes were subsequently tagged as described in Chapter 2.

The master list was narrowed down by excluding genes with less than one common variant (MAF $\geq$ 0.05) per 2kb of gene; less than 3 tagging SNPs (tSNPs); greater than 20 tSNPs. Genes were excluded from selection if the function of the gene was unknown. Genes with inconsistent fold change data (i.e. up-regulated in one hybrid, but down-regulated in the other hybrid of the same cell-line) were also discarded. This resulted in the exclusion of 107 genes/probes from the master list. Many genes from those excluded would have been very interesting to study (such as *APOBEC3C*, *TUBA1*, *HIF1A*, *ANXA13*, *DIO2*, *C20orf100*, *LIX1*, *C18orf34*, *GAMP*, *TPM3*, *DNAJB1*, *HEY1*, *SERINC2*, *RARB*, *RCDHB2*, *EPS15L2*, *SLC38A6* and *MAPT*) – see Appendix I for the functions of these genes.

Genes were then selected based on their function and role in ovarian and other malignancies, using literature searches and Oncomine (http://www.oncomine.org/). The hypothesis was that common variants and haplotypes of differentially expressed

candidate genes selected based on *in vitro* and *in vivo* functional evidence may affect a woman's risk of ovarian cancer.

At least 1 candidate gene was selected from each gene list. The genes selected from each list were: one from "down-regulated in TOV112D hybrids" (*RGC32* [FC=3, $P=1.79x10^{-6}$]); two from "up-regulated in TOV112D hybrids" (*FILIP1L* [FC=4.9, $P=3.44x10^{-5}$], *AXIN2* [FC=5.2, P=0.0027]); one from "down-regulated in the TOV21G hybrids" (*SQSTM1* [FC=109.9, $P=1.29x10^{-5}$]); two from "up-regulated in TOV21G hybrids" (*CASP5* [FC=6.9, $P=1.7x10^{-3}$], *STAG3* [FC=8.6, $P=6.4x10^{-5}$); two from "down-regulated in both cell lines" (*RUVBL1* [TOV21G: FC=46.7, $P=2.22x10^{-6}$; TOV112D: FC=2, $P=9.49x10^{-6}$], *SFRS9* [TOV21G: FC=76.9, P=0.0028; TOV112D: FC=4.3, P=0.0019]); three from "up-regulated in both cell-lines" (*AIFM2* [TOV21G: FC=3.4, P=0.0014; TOV112D: FC=1.8, P=0.003], *AKTIP* [TOV21G: FC=5.7, P=0.006; TOV112D: FC=3, P=0.0067], *EIF4B* [TOV21G: FC=8.8, P=0.09; TOV112D: FC=4, P=0.0014]); and one from "chromosome 18, up-regulated in TOV112D" (*RBBP8* [FC=2.9, P=0.023]).

iPLEX Gold Assay Design software was used to create multiplex panels with the tSNPs from these genes. The candidate genes selected for genotyping were based on the multiplex levels within two panels, how many SNPs from each gene were included in the panel; the rankings of the gene within the lists; and the function of the gene. The candidate genes selected from the MMCT gene expression results for iPLEX Gold genotyping were *AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *EIF4B*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1*, *SFRS9*, *SQSTM1* and STAG3 (see Table 3.17).

Table 3.17: Candidate "functional" genes from MMCT-18 study

| Regulation in hybrids | Gene | Cytoband | Function | Size (bp) | No. SNPs | tSNPs |
|---|---|---|---|---|---|---|
| Down in TOV112D hybrids | *RGC32* | 13q14.11 | Cell cycle progression regulation. Induced by p53 in response to DNA damage. | 13,323 | 17 | 8 |
| Up in TOV112D hybrids | *FILIP1L* | 3q12.1 | Down regulated in ovarian cancer. | 281,369 | 135 | 8 |
| | *AXIN2* | 17q23-q24 | Inhibitor of β-catenin in the Wnt signalling pathway. In region of LOH in breast & other cancers. | 33,084 | 14 | 12 |
| Up in TOV21G hybrids | *CASP5* | 11q22.2-q22.3 | Regulation of apoptosis. | 14,729 | 17 | 9 |
| | *STAG3* | 7q22.1 | Component of cohesin complex. Chromosome segregation. | 43,764 | 28 | 3 |
| Down in TOV21G hybrids | *SQSTM1* | 5q35 | May be involved in cell differentiation, apoptosis, immune response and regulation of K(+) channels | 17,181 | 15 | 10 |
| Down in TOV112D & TOV21G hybrids | *RUVBL1* | 3q21 | Interacts with MYC. Forms a complex which may be involved in cell growth. | 42,857 | 29 | 7 |
| | *SFRS9* | 12q24.31 | Plays a role in constitutive splicing and can modulate the selection of alternative splice sites. | 8,087 | 5 | 3 |
| Up in TOV112D & TOV21G hybrids | *AIFM2* | 10q22.1 | TP53-induced apoptosis. Overexpression has been shown to induce apoptosis. | 34,711 | 17 | 13 |
| | *AKTIP* | 16q12.2 | Apoptosis. Protein interacts directly with PKB/Akt and modulates PKB activity by enhancing the phosphorylation of PKB's regulatory sites. | 11,978 | 7 | 4 |
| | *EIF4B* | 12q13.13 | Translation initiation factor. | 35,770 | 40 | 8 |
| Chr18 up in TOV112D hybrids | *RBBP8* | 18q11.2 | RB1 binding protein. Believed to modulate the functions of BRCA1 in transcriptional regulation, DNA repair, and/or cell cycle checkpoint control. Has been proposed that this gene may be a tumour suppressor acting in the same pathway as BRCA1. | 93,155 | 39 | 4 |

SNP selection criteria: minor allele frequency ≥ 0.05, Hardy-Weinberg equilibrium p-value ≥ 0.01 using HapMap Data Release 22.

The sequences of the tagging SNP of the differentially expressed candidate genes were formatted for the iPLEX Gold Assay design software, in differing gene combinations to assess the most efficient multiplexes for panels. The assay panel design chosen contained the most interesting candidate genes in terms of known function (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*), would result in 10 tSNPs in total being genotyped by TaqMan if all the polymorphisms were successfully genotyped on the iPLEX Gold system. The functions of the candidate genes that were chosen for the association study are summarised in Table 3.17. The panel also contained assays for *BRCA1* (rs799917) and *BRCA2* (rs144848) SNPs, which had been genotyped and sequenced. These variants were used for quality control purposes. The panel selected comprised of a 27-plex and 33-plex.

### 3.3.2: MMCT-18 samples and methods

Due to the poor performance of the GEOCS samples on the iPLEX platform in the oncogene study, it was excluded from the stage 1 set of samples. More SEARCH and UKOPS samples were available for stage 1 than for the oncogene study. Additional SEARCH cases and UKOPS cases and controls were also available for stage 2 validation. These extra samples came from ongoing participant recruitment and preparation of samples. The sample sets used in this "functional" study are shown in Table 3.18. The histological subtype data for BAVARIA and the new SEARCH and UKOPS cases are currently not available.

Three different populations were used in stage 1 of this study. These populations consisted of: (1) The Danish MALOVA study; (2) The UK SEARCH study; (3) the

UKOPS study from the UK. In total, there were 1,799 invasive epithelial ovarian cancer cases and 3,045 unaffected controls in this series. Stage 2 samples were used to validate findings from stage 1. See Table 3.18 for the numbers of cases and controls genotyped from each population set.

Table 3.18: Ovarian cancer case-control populations included in functional study

| Study | Controls | Total cases | Serous | Endometrioid | Mucinous | Clear cell |
|---|---|---|---|---|---|---|
| MALOVA | 1221 | 446 | 275 | 56 | 43 | 33 |
| SEARCH | 1229 | 847 | 328 | 138 | 104 | 83 |
| UKOPS | 595 | 506 | 246 | 84 | 48 | 49 |
| **Total stage 1** | **3045** | **1799** | **849** | **278** | **195** | **165** |
| AUS | 1122 | 768 | 464 | 105 | 27 | 50 |
| BAV | 234 | 228 | Unknown | Unknown | Unknown | Unknown |
| DOVE | 716 | 584 | 303 | 86 | 18 | 30 |
| GEOCS | 429 | 327 | 166 | 47 | 29 | 23 |
| GER | 416 | 218 | 103 | 21 | 21 | 6 |
| HAWAII | 158 | 70 | 36 | 11 | 2 | 6 |
| HOPE | 603 | 280 | 159 | 40 | 13 | 21 |
| JAC | 593 | 603 | 300 | 62 | 49 | 12 |
| NCOCS | 747 | 622 | 242 | 50 | 33 | 22 |
| USC | 546 | 391 | 161 | 33 | 19 | 14 |
| UKO-P2 | 467 | 553 | Unknown | Unknown | Unknown | Unknown |
| **Total stage 2** | **6031** | **4590** | **1934** | **455** | **211** | **184** |
| **Total stages 1 & 2** | **9076** | **6389** | **2783** | **733** | **406** | **349** |

The stage 2 samples comprised of case-control studies from Australia (AUS), Germany (BAVARIA; and GER), Poland (JAC), the United Kingdom (UKO (B) - consisting of 368 SEARCH cases; and 185 UKOPS cases and 467 controls), and the United States of America (DOVE; GEOCS; HAWAII; HOPE; NCOCS and USC).

A total of 68 tSNP were identified from the nine candidate genes. Of these, five tSNPs (*AIFM2* rs2271695; *AXIN2* rs2240308 and rs4128941; *RGC32* rs3783197;

and *RUVBL1* rs13091198) failed assay design, manufacture, probe testing or QC. Therefore, the stage 1 samples were successfully genotyped with 63 tSNPs from *AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3* with a combination of iPLEX Gold and Taqman SNP genotyping platforms. Overall 95% of the candidate genes were covered by the tSNPs which were successfully genotyped (288 variants covered by the genotyped tSNPs out of a total 303 variants).

### 3.3.3: Ovarian cancer risks associated with common genetic variation in functional candidate genes

The genotype distributions for the tagging SNPs of the MMCT-18 candidate genes are tabulated in Appendix II-B. The results of the logistic regression are shown in Appendix IV-A to IV-R.

When the effects of the common variants of the MMCT-18 candidate genes on predisposition of ovarian cancer were assessed, there was no evidence of association with *AIFM2*, *AKTIP*, *FILIP1L*, *RBBP8*, *RGC32* and *STAG3*. A tSNP from *AXIN2*, *CASP5* and *RUVBL1* were associated with risk of ovarian cancer. The rare allele of *AXIN2* rs11079571 was associated with an increase in ovarian cancer risk (HetOR=1.23 [1-1.51], HomOR=1.73 [0.99-3.01], P=0.0383). The rare allele of *CASP5* rs518604 was associated with an increase in ovarian cancer risk; (HetOR= 1.39 [1.06-1.81], HomOR=1.44 [1.05-1.97], P=0.0124), Table 3.19. This association remained when analysis was restricted to serous only cases.

Table 3.19: Genotype-specific risks of MMCT-18 candidate genes

| Gene | tSNP | MAF | Controls | All cases | Serous cases | All subtypes | | | Serous subtype | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | HetOR | HomOR | P-trend | HetOR | HomOR | P-trend |
| *AIFM2* | rs2394655 | 0.04 | 2924 | 1751 | 827 | 1.01 (0.80-1.27) | 1.37 (0.42-4.42) | 0.7773 | 1.12 (0.84-1.49) | 1.03 (0.21-5.04) | 0.4297 |
| | rs7908957 | 0.13 | 2873 | 1719 | 817 | 0.92 (0.79-1.07) | 1.13 (0.73-1.75) | 0.5342 | 0.91 (0.75-1.11) | 0.91 (0.50-1.67) | 0.3812 |
| | rs1053495 | 0.08 | 2704 | 1697 | 790 | 0.97 (0.81-1.16) | 0.72 (0.35-1.51) | 0.457 | 0.96 (0.75-1.21) | 0.88 (0.35-2.18) | 0.6293 |
| | rs2894111 | 0.28 | 2861 | 1770 | 835 | 0.99 (0.87-1.13) | 0.93 (0.74-1.16) | 0.5545 | 1.02 (0.87-1.20) | 0.86 (0.64-1.17) | 0.5896 |
| | rs2394656 | 0.19 | 1703 | 913 | 506 | 0.96 (0.80-1.15) | 0.85 (0.55-1.34) | 0.4114 | 0.93 (0.75-1.17) | 0.85 (0.50-1.47) | 0.3965 |
| | rs6480440 | 0.23 | 1140 | 422 | 556 | 1.01 (0.79-1.29) | 0.88 (0.55-1.42) | 0.7992 | 1.07 (0.80-1.43) | 0.75 (0.40-1.39) | 0.8126 |
| | rs2280201 | 0.12 | 1783 | 1313 | 261 | 1.05 (0.88-1.25) | 0.96 (0.53-1.71) | 0.59 | 1.11 (0.88-1.40) | 0.97 (0.44-2.12) | 0.5089 |
| | rs10999147 | 0.08 | 2395 | 1285 | 600 | **1.25 (1.03-1.51)** | 0.48 (0.16-1.47) | 0.2055 | 1.13 (0.88-1.46) | 0.79 (0.23-2.75) | 0.6396 |
| | rs3750772 | 0.05 | 2944 | 1743 | 831 | 1.02 (0.83-1.24) | 2.28 (0.86-6.04) | 0.54 | 1.12 (0.88-1.44) | **3.43 (1.23-9.58)** | 0.1043 |
| | rs4295944 | 0.41 | 1784 | 1335 | 567 | 0.98 (0.83-1.15) | 1.09 (0.88-1.34) | 0.4913 | 0.95 (0.77-1.18) | 1.11 (0.84-1.46) | 0.6042 |
| | rs2394644 | 0.12 | 1685 | 1324 | 561 | 1.05 (0.88-1.26) | 1.03 (0.60-1.79) | 0.5245 | 1.02 (0.80-1.29) | 1.11 (0.55-2.25) | 0.7103 |
| | rs10999152 | 0.17 | 2610 | 1618 | 760 | 1.14 (0.99-1.32) | 0.96 (0.68-1.36) | 0.2066 | 1.17 (0.98-1.40) | 1.01 (0.65-1.57) | 0.2171 |
| *AKTIP* | rs9931702 | 0.44 | 1722 | 917 | 506 | 0.96 (0.80-1.16) | 1.01 (0.80-1.27) | 0.9734 | 1.10 (0.87-1.38) | 1.12 (0.85-1.50) | 0.4039 |
| | rs17801966 | 0.14 | 1469 | 828 | 450 | 1.06 (0.87-1.31) | 0.74 (0.38-1.46) | 0.9282 | 1.14 (0.89-1.46) | 0.90 (0.40-2.00) | 0.4718 |
| | rs7189819 | 0.32 | 2923 | 1745 | 825 | 0.93 (0.82-1.06) | 0.92 (0.75-1.14) | 0.2796 | 0.99 (0.84-1.17) | 1.02 (0.78-1.33) | 0.867 |
| | rs3743772 | 0.06 | 1093 | 413 | 256 | 0.90 (0.62-1.32) | 2.02 (0.32-12.59) | 0.6778 | 1.12 (0.73-1.72) | 1.72 (0.17-17.14) | 0.6424 |
| ***AXIN2*** | rs11868547 | 0.49 | 1717 | 919 | 509 | 0.93 (0.76-1.13) | 1.02 (0.81-1.29) | 0.8178 | 0.92 (0.72-1.17) | 0.99 (0.74-1.31) | 0.9949 |
| | rs7591 | 0.37 | 2881 | 1779 | 838 | 1.09 (0.96-1.24) | 1.05 (0.87-1.27) | 0.4234 | **1.19 (1.01-1.41)** | 1.11 (0.87-1.41) | 0.1463 |
| | rs4074947 | 0.2 | 2898 | 1775 | 840 | **1.16 (1.02-1.32)** | 0.92 (0.67-1.24) | 0.2189 | **1.21 (1.03-1.43)** | 0.76 (0.50-1.16) | 0.323 |
| | rs7210356 | 0.11 | 2974 | 1777 | 838 | 1.01 (0.87-1.17) | 0.97 (0.56-1.69) | 0.8864 | 1.09 (0.90-1.32) | 1.08 (0.54-2.14) | 0.3518 |
| | rs11655966 | 0.26 | 1779 | 1301 | 552 | 0.92 (0.79-1.07) | 1.07 (0.80-1.44) | 0.9064 | 0.99 (0.81-1.22) | 1.10 (0.74-1.63) | 0.6218 |
| | rs4541111 | 0.49 | 1770 | 1297 | 554 | 1.09 (0.91-1.30) | 0.97 (0.78-1.19) | 0.806 | 1.10 (0.87-1.40) | 1.03 (0.78-1.36) | 0.755 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | rs4791171 | 0.28 | 2109 | 1185 | 539 | 1.08 (0.92-1.25) | 1.19 (0.92-1.55) | 0.1238 | 1.05 (0.86-1.29) | 1.23 (0.88-1.73) | 0.2499 |
| | **rs11079571** | 0.15 | 1206 | 839 | 326 | **1.23 (1.00-1.51)** | 1.73 (0.99-3.01) | **0.0383** | 1.22 (0.92-1.63) | 1.74 (0.84-3.63) | 0.1127 |
| | rs3923087 | 0.21 | 2910 | 1780 | 843 | 1.05 (0.93-1.20) | 1.26 (0.95-1.68) | 0.1545 | 1.05 (0.89-1.24) | 1.10 (0.75-1.60) | 0.4525 |
| | rs3923086 | 0.41 | 2935 | 1753 | 833 | 0.91 (0.80-1.04) | 1.07 (0.90-1.27) | 0.813 | **0.84 (0.71-1.00)** | 1.02 (0.81-1.27) | 0.6828 |
| ***CASP5*** | **rs518604** | 0.44 | 1195 | 438 | 270 | **1.39 (1.06-1.81)** | **1.44 (1.05-1.97)** | **0.0124** | 1.36 (0.98-1.88) | 1.45 (0.99-2.11) | **0.0313** |
| | rs523104 | 0.47 | 1199 | 824 | 320 | 1.03 (0.83-1.29) | 1.07 (0.82-1.39) | 0.7689 | 0.86 (0.65-1.16) | 0.80 (0.55-1.15) | 0.1294 |
| | rs3181328 | 0.09 | 1206 | 829 | 319 | 0.95 (0.73-1.23) | 1.16 (0.50-2.72) | 0.7779 | 0.93 (0.65-1.33) | 0.68 (0.18-2.64) | 0.7345 |
| | rs17446518 | 0.12 | 1177 | 803 | 311 | 0.88 (0.69-1.12) | 1.28 (0.54-3.02) | 0.5052 | 0.87 (0.63-1.22) | 1.01 (0.28-3.69) | 0.5292 |
| | rs9651713 | 0.11 | 2898 | 1730 | 819 | 0.99 (0.85-1.15) | 1.22 (0.70-2.13) | 0.836 | 0.96 (0.78-1.17) | 1.55 (0.81-2.96) | 0.8167 |
| | rs3181175 | 0.18 | 2379 | 1282 | 597 | 0.96 (0.83-1.12) | 1.23 (0.83-1.82) | 0.9331 | 0.93 (0.76-1.13) | 1.34 (0.82-2.19) | 0.9482 |
| | rs3181174 | 0.08 | 2962 | 1780 | 840 | 0.96 (0.80-1.14) | 1.21 (0.55-2.65) | 0.7967 | 1.09 (0.88-1.36) | 1.12 (0.41-3.09) | 0.4061 |
| | rs2282657 | 0.35 | 1478 | 852 | 462 | 1.08 (0.90-1.30) | 0.89 (0.67-1.20) | 0.7645 | 1.16 (0.92-1.45) | 1.11 (0.78-1.56) | 0.3615 |
| | rs507879 | 0.54 | 2839 | 1768 | 835 | 1.02 (0.87-1.20) | 1.00 (0.84-1.19) | 0.9144 | 1.09 (0.89-1.34) | 1.00 (0.79-1.26) | 0.8497 |
| *FILIP1L* | rs796977 | 0.33 | 1166 | 437 | 269 | 1.07 (0.84-1.35) | 1.33 (0.94-1.89) | 0.1458 | 0.98 (0.74-1.31) | 1.25 (0.83-1.90) | 0.4593 |
| | rs793477 | 0.13 | 2646 | 1653 | 771 | 1.02 (0.88-1.18) | 0.88 (0.53-1.43) | 0.9373 | 1.11 (0.92-1.35) | 1.08 (0.59-1.99) | 0.3224 |
| | rs793446 | 0.4 | 2947 | 1773 | 838 | 1.05 (0.92-1.20) | 1.11 (0.93-1.32) | 0.3207 | 0.95 (0.80-1.13) | 1.07 (0.86-1.34) | 0.842 |
| | rs3921767 | 0.07 | 2859 | 1773 | 840 | 1.00 (0.84-1.20) | 0.69 (0.29-1.61) | 0.6908 | 0.90 (0.71-1.15) | 0.19 (0.03-1.43) | 0.1194 |
| | rs17338680 | 0.11 | 2989 | 1786 | 574 | 1.06 (0.91-1.24) | 0.77 (0.46-1.28) | 0.985 | 0.88 (0.69-1.14) | 0.98 (0.44-2.20) | 0.3051 |
| | rs9864437 | 0.22 | 2972 | 1786 | 843 | 0.98 (0.86-1.12) | 1.18 (0.91-1.53) | 0.6077 | 0.96 (0.81-1.14) | **1.47 (1.08-2.01)** | 0.2249 |
| | rs6788750 | 0.41 | 2532 | 1414 | 710 | 0.98 (0.84-1.13) | 0.96 (0.79-1.17) | 0.7028 | 1.07 (0.88-1.29) | 1.12 (0.88-1.44) | 0.3295 |
| | rs12494994 | 0.17 | 2347 | 1273 | 594 | 1.14 (0.98-1.33) | 0.88 (0.59-1.33) | 0.433 | 0.95 (0.78-1.17) | 0.52 (0.27-0.99) | 0.097 |
| ***RBBP8*** | rs7239066 | 0.11 | 2366 | 1272 | 594 | 0.80 (0.67-0.96) | 1.05 (0.57-1.95) | 0.0645 | 0.99 (0.79-1.23) | 1.18 (0.55-2.53) | 0.7647 |
| | rs11082221 | 0.04 | 2937 | 1748 | 826 | 1.12 (0.89-1.41) | 1.13 (0.31-4.07) | 0.2974 | 1.17 (0.88-1.57) | 0.55 (0.07-4.59) | 0.3355 |
| | **rs4474794** | 0.37 | 2895 | 1764 | 829 | 0.94 (0.82-1.07) | 0.88 (0.72-1.06) | 0.2066 | **0.83 (0.70-0.98)** | 0.80 (0.63-1.03) | **0.0323** |
| | rs9304261 | 0.24 | 888 | 346 | 215 | 1.07 (0.82-1.40) | 0.67 (0.38-1.17) | 0.5163 | 1.08 (0.79-1.49) | 0.37 (0.16-0.88) | 0.2176 |
| *RGC32* | rs10467472 | 0.13 | 2887 | 1769 | 839 | 0.99 (0.86-1.15) | 1.05 (0.68-1.64) | 0.9822 | 0.91 (0.75-1.10) | 0.74 (0.39-1.41) | 0.2126 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | rs3783194 | 0.11 | 2723 | 1690 | 788 | 1.02 (0.88-1.19) | 1.03 (0.57-1.89) | 0.8873 | 0.89 (0.72-1.09) | 0.70 (0.29-1.69) | 0.1363 |
| | rs11618371 | 0.1 | 2959 | 1771 | 835 | 1.01 (0.87-1.18) | 1.31 (0.74-2.32) | 0.5158 | 1.03 (0.85-1.26) | 1.33 (0.66-2.70) | 0.4921 |
| | rs9532824 | 0.08 | 2892 | 1782 | 841 | 0.96 (0.81-1.15) | 0.47 (0.19-1.16) | 0.3412 | 0.94 (0.74-1.18) | 0.46 (0.14-1.54) | 0.3245 |
| | rs995845 | 0.26 | 2365 | 1274 | 595 | 1.01 (0.87-1.16) | 1.09 (0.82-1.45) | 0.8121 | 0.89 (0.74-1.09) | 1.06 (0.73-1.52) | 0.4695 |
| | rs9594551 | 0.15 | 2863 | 1766 | 833 | 1.01 (0.88-1.16) | 1.10 (0.74-1.65) | 0.611 | 1.09 (0.91-1.29) | 0.97 (0.56-1.65) | 0.4095 |
| | rs975590 | 0.24 | 2940 | 1749 | 828 | 1.02 (0.90-1.16) | 0.89 (0.68-1.17) | 0.8348 | 1.03 (0.87-1.21) | 0.84 (0.59-1.21) | 0.7862 |
| ***RUVBL1*** | rs9860614 | 0.11 | 2966 | 1777 | 839 | 1.05 (0.90-1.22) | 1.44 (0.89-2.31) | 0.2094 | 1.16 (0.96-1.40) | 1.03 (0.52-2.02) | 0.1425 |
| | **rs13063604** | 0.23 | 1724 | 1266 | 537 | 1.14 (0.97-1.34) | **1.39 (1.02-1.89)** | **0.0192** | **1.42 (1.15-1.74)** | **1.63 (1.10-2.42)** | **0.0002** |
| | rs3732402 | 0.38 | 2382 | 1280 | 596 | **1.20 (1.03-1.40)** | 1.06 (0.85-1.31) | 0.207 | **1.35 (1.10-1.64)** | 1.14 (0.86-1.52) | 0.0677 |
| | **rs7650365** | 0.49 | 2672 | 1645 | 769 | 1.08 (0.93-1.26) | 0.86 (0.72-1.03) | 0.1081 | 0.97 (0.80-1.17) | **0.74 (0.58-0.93)** | **0.009** |
| | rs4857836 | 0.27 | 2993 | 1787 | 845 | 1.05 (0.93-1.19) | 0.97 (0.76-1.23) | 0.8219 | 1.11 (0.95-1.31) | 1.04 (0.77-1.41) | 0.3458 |
| | rs9821568 | 0.15 | 2911 | 1733 | 820 | 0.97 (0.84-1.11) | 0.93 (0.63-1.37) | 0.5613 | 0.99 (0.82-1.18) | 0.57 (0.31-1.06) | 0.2966 |
| ***STAG3*** | rs11762932 | 0.22 | 2965 | 1787 | 846 | 1.03 (0.90-1.17) | 1.06 (0.80-1.40) | 0.6327 | 1.03 (0.88-1.22) | 1.06 (0.74-1.52) | 0.6639 |
| | rs2246713 | 0.49 | 1765 | 1295 | 549 | 0.97 (0.81-1.15) | 0.96 (0.78-1.19) | 0.6593 | 0.95 (0.75-1.20) | 0.87 (0.66-1.16) | 0.2436 |
| | **rs1637001** | 0.28 | 2967 | 1784 | 843 | **0.86 (0.76-0.98)** | 0.92 (0.73-1.16) | 0.0692 | **0.84 (0.71-0.99)** | 0.77 (0.56-1.05) | **0.0177** |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

The confidence intervals of both the heterozygous and homozygous odds ratios of *CASP5* rs518604 crossed 1, however the lower confidence intervals were close to 1 (HetOR=1.36 [0.98-1.88], HomOR=1.45 [0.99-2.11], P=0.0313; Table 3.19). The *AXIN2* rs11079571 and *CASP5* rs518604 were only successfully genotyped for 839 cases and 1,206 controls, and 438 cases and 1,195 controls, respectively with the iPLEX Gold platform. TaqMan assays of the variants failed to produce callable genotype clusters, therefore the associations could not be further investigated with the remaining stage 1 and stage 2 samples. Neither *AXIN2* nor *CASP5* variants were tagged by any other SNP with $r^2 \geq 0.8$, therefore it was not possible to manufacture Taqman assays of SNPs in LD. Due to the relatively small numbers of samples successfully genotyped for these variants and the borderline significance of the P-values, there is a possibility that the associations are chance findings.

The rare allele of *RUVBL1* rs13063604 was also associated with an increased risk of ovarian cancer; HetOR=1.14 (0.97-1.34), and the HomOR=1.39 (1.02-1.89), P=0.0192. This association became stronger when the analysis was restricted to the serous histological subtype (P=0.0002). There was also evidence suggesting that the rare allele of *RUVBL1*, rs7650365, was associated with a decreased risk of the serous subtype (Table 3.20).

Table 3.20: Genotype-specific risks of variants of *RUVBL1* (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR (95% CI)* | HomOR (95% CI)* | P-trend |
|---|---|---|---|---|---|---|---|---|
| *RUVBL1* | rs13063604 | 0.25 | 1724 | 1266 | **All** | 1.14 (0.97-1.34) | **1.39 (1.02-1.89)** | **0.0192** |
| | | | | 537 | **Serous** | **1.42 (1.15-1.74)** | **1.63 (1.10-2.42)** | **0.0002** |
| | | | | 207 | Endometrioid | 0.74 (0.53-1.04) | 1.29 (0.73-2.31) | 0.3904 |
| | | | | 143 | Mucinous | 0.95 (0.64-1.41) | 1.39 (0.69-2.83) | 0.5473 |
| | | | | 124 | Clear cell | 1.22 (0.83-1.80) | 1.07 (0.48-2.40) | 0.4113 |
| | rs7650365 | 0.46 | 2672 | 1645 | All | 1.08 (0.93-1.26) | 0.86 (0.72-1.03) | 0.1081 |
| | | | | 769 | **Serous** | 0.97 (0.80-1.17) | **0.74 (0.58-0.93)** | **0.009** |
| | | | | 256 | Endometrioid | 1.12 (0.83-1.53) | 0.74 (0.50-1.09) | 0.1777 |
| | | | | 175 | Mucinous | 1.04 (0.70-1.53) | 1.25 (0.81-1.93) | 0.371 |
| | | | | 155 | Clear cell | 1.08 (0.73-1.60) | 0.93 (0.58-1.48) | 0.7821 |
| | rs9821568 | 0.15 | 2911 | 1733 | All | 0.97 (0.84-1.11) | 0.93 (0.63-1.37) | 0.5613 |
| | | | | 820 | Serous | 0.99 (0.82-1.18) | 0.57 (0.31-1.06) | 0.2966 |
| | | | | 269 | **Endometrioid** | 1.14 (0.86-1.52) | **2.09 (1.16-3.78)** | **0.0286** |
| | | | | 186 | Mucinous | 0.86 (0.60-1.23) | 0.62 (0.19-2.00) | 0.1981 |
| | | | | 161 | Clear cell | 0.68 (0.45-1.02) | 0.85 (0.30-2.38) | 0.0967 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

**Validation of *RUVBL1* results**

The two SNPs with the strongest associations with ovarian cancer risk (*RUVBL1* rs13063604 and rs7650365) were genotyped in additional stage 2 samples, which included a total of 2,636 cases and 6,164 controls. Stage 2 comprised of samples from AUS, BAVARIA, DOVE, GEOCS, GER, HAW, HOPE, JAC, NCOCS, UKO-P2 (UKOPS cases and controls, and SEARCH cases) and USC. The associations with risk of the serous subtype were not validated in the stage 2 samples alone (P>0.05; refer to Table 3.21). The association between increased risk of serous ovarian cancer and the rare allele of rs13063604 HetOR=1.13 (1.00-1.27), HomOR= 1.22 (0.9-1.56), P=0.0191 remained statistically significant after combining stages 1 and 2 genotyping data. The rs13063604 variant tags nine other SNPs. Two of these, rs1057220 and rs1057156, are located in the 3' untranslated region (3'UTR), and they are predicted to be exonic splicing enhancers.

The association between the rare allele of rs7650365 and reduced risk of the serous subtype was no longer statistically significant after stages 1 and 2 were combined (HetOR=0.94 [0.86-1.04), HomOR=0.92 [0.82-1.03), P=0.142); Table 3.21.

The rare allele of *RBBP8* rs4474794 was found to be associated with a decrease in risk of serous histological subtype when logistic regression analysis was restricted to the subtype, HetOR=0.83 (0.70-0.98), HomOR=0.80 (0.63-1.03), P=0.0323; Appendix IV-L. The rs4474794 tSNP tags 17 other variants with $r^2 \geq 0.8$. Five of these SNPs are conserved in mice, however, all of the SNPs are intronic, and there are no known functions which could explain the association.

Table 3.21: Genotype-specific risks of *RUVBL1* rs13063604 and rs7650365 (by genotyping stage)

| tSNP | **Study** | **Controls** | **Cases** | **Histology** | **HetOR* (95% CI)** | **HomOR* (95% CI)** | **P-trend** |
|---|---|---|---|---|---|---|---|
| **rs13063604** | Stage 1 | 1724 | 1266 | All | 1.14 (0.97-1.34) | **1.39 (1.02-1.89)** | **0.019** |
| | Stage 2 | 2639 | 1915 | | 1.04 (0.92-1.18) | 1.09 (0.85-1.41) | 0.402 |
| | Stage 1 & 2 | 4363 | 3181 | | 1.08 (0.98-1.19) | 1.19 (0.98-1.45) | **0.033** |
| | Stage 1 | 1724 | 537 | Serous | **1.42 (1.15-1.74)** | **1.63 (1.10-2.42)** | **0.0002** |
| | Stage 2 | 2639 | 1218 | | 1.00 (0.86-1.16) | 1.05 (0.77-1.42) | 0.83 |
| | Stage 1 & 2 | 4363 | 1755 | | **1.13 (1.00-1.27)** | **1.22 (0.96-1.56)** | **0.019** |
| **rs7650365** | Stage 1 | 2672 | 1645 | All | 1.08 (0.93-1.26) | 0.86 (0.72-1.03) | 0.11 |
| | Stage 2 | 5885 | 4437 | | 0.97 (0.88-1.06) | 0.97 (0.87-1.09) | 0.624 |
| | Stage 1 & 2 | 8778 | 6129 | | 1.02 (0.94-1.1) | 0.96 (0.87-1.05) | 0.404 |
| | Stage 1 | 2672 | 769 | Serous | 0.97 (0.80-1.17) | **0.74 (0.58-0.93)** | **0.009** |
| | Stage 2 | 5885 | 2534 | | 0.94 (0.84-1.05) | 0.99 (0.87-1.13) | 0.858 |
| | Stage 1 & 2 | 8778 | 3303 | | 0.94 (0.86-1.04) | 0.92 (0.82-1.03) | 0.142 |

Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; * compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

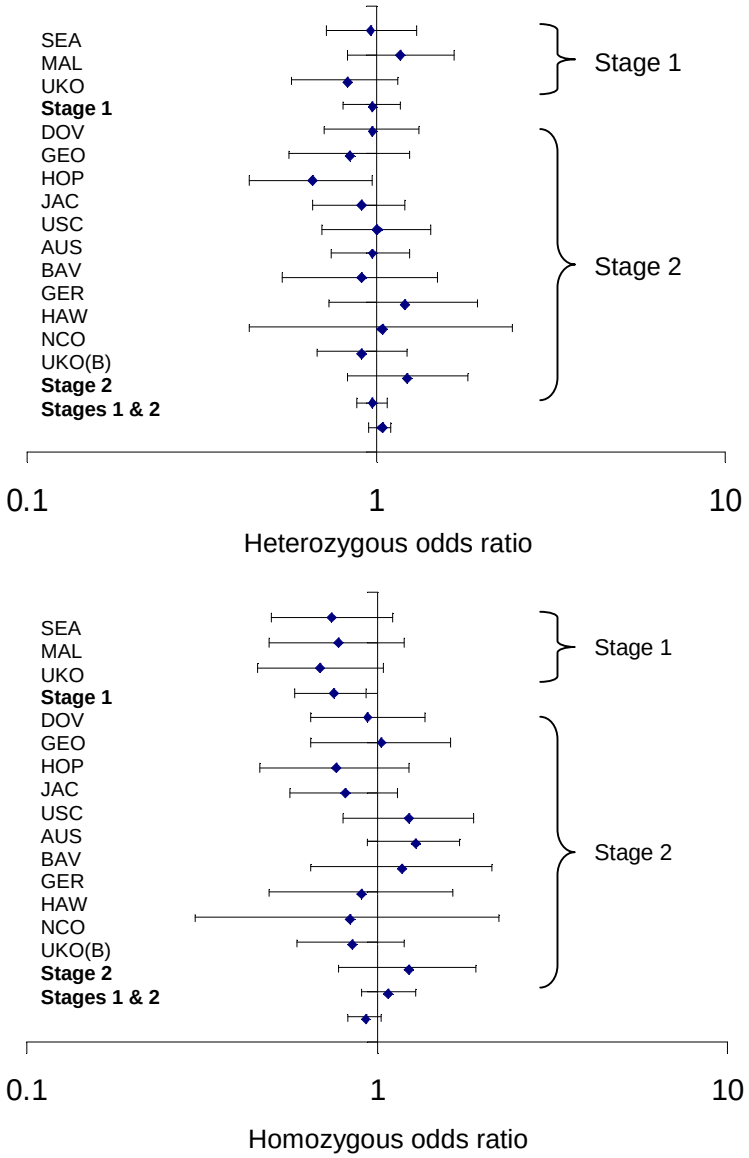



Figure 3.5: Forest plots of *RUVBL1* rs7650365 (serous subtype)

The genotype- and haplotype-specific results of *STAG3* are shown in Appendix IV-Q and V-R. The rare allele of rs1637001 was associated with a reduced risk of the serous histological subtype (HetOR=0.84 (0.71-0.99), HomOR=0.77 (0.56-1.05), P=0.0177). The heterozygous odd ratio for rs1637001 for all histological subtypes correlated to a decrease in the risk of ovarian cancer, HetOR=0.86 (0.76-0.98),

HomOR=0.92 (0.73-1.16), P=0.0692. However, the correlation was not statistically significant.

Associations were also found with the haplotypes of *AXIN2*, *CASP5* and *RUVBL1* and ovarian cancer susceptibility – see Table 3.22. The h1111 haplotype of *AXIN2* block 2 was associated with an increased risk of ovarian cancer - OR=1.21 (1.03-1.42), P=0.023. This association was also found when analysis was restricted to the serous subtype (OR=1.19 [1.01-1.39], P=0.037) – Appendix IV-F. The rs11079571 variant, which was associated with disease risk, was in the second position of the haplotypes of this block (*AXIN2* block 2). The association between the rare allele of rs11079571 and increased risk of ovarian cancer was consistent with the correlation of the h1111 haplotype of *AXIN2* block 2 and increased risk of ovarian cancer. No other common haplotypes of *AXIN2* block 2 contained the rare allele of rs11079571 (see Appendix IV-F).

The h000 haplotype of *CASP5* block 1, which contained rs518604 in the first position, was associated with a reduced risk of ovarian cancer, OR=0.72 (0.56-0.94), P=0.015. Conversely, the h100 haplotype of the same block was associated with an increased risk of ovarian cancer of all subtypes, OR=1.13 (1.03-1.24), P=0.012; Table 3.22. The associations were consistent with the effects of the rs518604 alleles. When the global effects of *CASP5* haplotype block 1 on ovarian cancer susceptibility were investigated, a highly significant association was found, $P=8.43x10^{-6}$. Although no individual haplotype of *RUVBL1* was found to be associated with the risk of ovarian cancer, *RUVBL1* haplotypes were globally associated with ovarian cancer predisposition, P=0.0016.

Table 3.22: MMCT-18 susceptibility - haplotype results (all subtypes)

| Gene/haplotype block | Haplotype | Freq (%) | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|
| *AIFM2* block 1 | h0000000 | 70 | 1.04 (0.9-1.2) | 0.60 | 0.93 |
| | h0001011 | 7 | 1.22 (0.94-1.59) | 0.13 | |
| | h0001100 | 4 | 0.8 (0.54-1.17) | 0.24 | |
| | h1111110 | 4 | 1 (0.71-1.42) | 1.00 | |
| *AIFM2* block 2 | h00000 | 36 | 0.92 (0.83-1.02) | 0.12 | 0.80 |
| | h00001 | 4 | 1.06 (0.83-1.35) | 0.67 | |
| | h00011 | 7 | 1.12 (0.92-1.36) | 0.25 | |
| | h00100 | 39 | 1.01 (0.91-1.11) | 0.86 | |
| *AKTIP* | h0000 | 55 | 1.01 (0.88-1.16) | 0.90 | 0.37 |
| | h1010 | 30 | 1 (0.86-1.16) | 1.00 | |
| | h1100 | 8 | 0.97 (0.75-1.25) | 0.80 | |
| | h1101 | 6 | 1 (0.75-1.33) | 1.00 | |
| *AXIN2* block 1 | h000001 | 15 | 0.9 (0.77-1.04) | 0.15 | 0.19 |
| | h010011 | 6 | 0.9 (0.72-1.12) | 0.36 | |
| | h010111 | 11 | 1.05 (0.89-1.23) | 0.58 | |
| | h011001 | 10 | 1.15 (0.98-1.36) | 0.08 | |
| | h011010 | 4 | 0.96 (0.73-1.26) | 0.75 | |
| | h011011 | 6 | 1.08 (0.87-1.33) | 0.49 | |
| | h100000 | 45 | 1.03 (0.93-1.13) | 0.60 | |
| *AXIN2* block 2 | h0000 | 53 | 0.98 (0.87-1.1) | 0.75 | 0.085 |
| | h0001 | 15 | 0.93 (0.79-1.11) | 0.43 | |
| | h1001 | 6 | 0.98 (0.77-1.26) | 0.87 | |
| | h1011 | 7 | 1.06 (0.86-1.3) | 0.60 | |
| | **h1111** | 13 | **1.21 (1.03-1.42)** | **0.023** | |
| ***CASP5* block 1** | h001 | 10 | 0.9 (0.77-1.06) | 0.22 | **$8.4 \times 10^{-6}$** |
| | h010 | 44 | 0.99 (0.9-1.09) | 0.79 | |
| | **h100** | 43 | **1.13 (1.03-1.24)** | **0.012** | |
| *CASP5* block 2 | h000000 | 10 | 0.87 (0.71-1.07) | 0.18 | 0.25 |
| | h000001 | 49 | 1.02 (0.9-1.14) | 0.77 | |
| | h000010 | 13 | 1.1 (0.93-1.3) | 0.29 | |
| | h001110 | 7 | 0.97 (0.77-1.22) | 0.79 | |
| | h011010 | 10 | 1.15 (0.95-1.39) | 0.16 | |
| | h100001 | 5 | 1.01 (0.82-1.26) | 0.90 | |
| *FILIP1L* block 1 | h00000 | 46 | 0.95 (0.84-1.08) | 0.45 | 0.59 |
| | h00110 | 7 | 0.94 (0.74-1.2) | 0.62 | |
| | h01000 | 12 | 0.97 (0.79-1.19) | 0.78 | |
| | h10100 | 23 | 1.08 (0.93-1.24) | 0.31 | |
| | h10101 | 11 | 1.11 (0.91-1.35) | 0.29 | |
| *FILIP1L* block 2 | h000 | 19 | 0.94 (0.83-1.08) | 0.38 | 0.76 |
| | h001 | 17 | 1.04 (0.92-1.17) | 0.56 | |
| | h010 | 41 | 0.98 (0.88-1.08) | 0.64 | |

| Gene/haplotype block | Haplotype | Freq (%) | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|
| | h100 | 23 | 1.04 (0.93-1.16) | 0.53 | |
| *RBBP8* | h0000 | 62 | 1.12 (0.98-1.27) | 0.09 | 0.64 |
| | h0010 | 02 | 0.88 (0.71-1.1) | 0.27 | |
| | h0011 | 23 | 0.92 (0.8-1.07) | 0.27 | |
| | h1010 | 7 | 0.83 (0.64-1.07) | 0.15 | |
| *RGC32* | h0000000 | 42 | 0.98 (0.89-1.08) | 0.67 | 0.63 |
| | h0000011 | 5 | 0.93 (0.74-1.17) | 0.54 | |
| | h0000100 | 10 | 1.04 (0.89-1.21) | 0.61 | |
| | h0001001 | 7 | 0.92 (0.75-1.12) | 0.41 | |
| | h0010011 | 8 | 1.1 (0.93-1.31) | 0.28 | |
| | h0100100 | 11 | 1.01 (0.87-1.18) | 0.85 | |
| | h1000000 | 8 | 1 (0.84-1.2) | 0.96 | |
| ***RUVBL1*** | h000000 | 13 | 0.91 (0.77-1.07) | 0.25 | **0.0016** |
| | h000100 | 48 | 0.96 (0.86-1.07) | 0.49 | |
| | h001011 | 15 | 0.98 (0.84-1.15) | 0.83 | |
| | h011010 | 12 | 1.15 (0.97-1.36) | 0.11 | |
| | h111000 | 10 | 1.17 (0.99-1.4) | 0.07 | |
| *STAG3* | h000 | 51 | 1.06 (0.96-1.16) | 0.26 | 0.098 |
| | h011 | 27 | 0.94 (0.85-1.05) | 0.29 | |
| | h110 | 21 | 1.03 (0.92-1.16) | 0.63 | |

Freq – frequency; OR – odds ratio; CI – confidence interval; SNP order in haplotypes is 5' to 3' of the genes – *AIFM2* (block 1): rs2394655, rs7908957, rs1053495, rs2894111, rs2394656, rs6480440, rs2280201. *AIFM2* (block 2): rs10999147, rs3750772, rs4295944, rs2394644, rs10999152. *ATKIP*: rs9931702, rs17801966, rs7189819, rs3743772. *AXIN2* (block 1): rs11868547, rs7591, rs4074947, rs7210356, rs11655966, rs4541111. *AXIN2* (block 2): rs4791171, rs11079571, rs3923087, rs3923086. *CASP5* (block 1): rs518604, rs523104, rs3181328. *CASP5* (block 2): rs17446518, rs9651713, rs3181175, rs3181174, rs2282657, rs507879. *FILIP1L* (block 1): rs796977, rs793477, rs793446, rs3921767, rs17338680. *FILIP1L* (block 2): rs9864437, rs6788750, rs12494994. *RBBP8*: rs7239066, rs11082221, rs4474794, rs9304261. *RGC32*: rs10467472, rs3783194, rs11618371, rs9532824, rs995845, rs9594551, rs975590. *RUVBL1*: rs9860614, rs13063604, rs3732402, rs7650365, rs4857836, rs9821568. *STAG3*: rs11762932, rs2246713, rs1637001.

## Associations between the MMCT-18 candidates and risk of the major histological subtypes of ovarian cancer

As shown with the candidate oncogenes, statistically significant associations may be found between candidate genes and the histological subtypes of ovarian cancer, which may not be detected with the analysis of all samples. However, the results should be treated with caution as the numbers of samples are further reduced. The

following describes statistically significant associations with tables of the results of all subtypes combined and individually.

## Associations between *AIFM2* and ovarian cancer susceptibility

A haplotype of *AIFM2* block 2 was associated with the risk of mucinous ovarian cancer (Table 3.23). The h00100 of *AIFM2* block 2 was associated with an increased risk of the subtype, OR=1.26 (1.02-1.55), P=0.034. See Appendix IV-B for the logistic regression results for all the common *AIFM2* haplotypes.

Table 3.23: Haplotype-specific results of *AIFM2* (P<0.05)

| Haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *AIFM2* haplotype block 2 | h00100 | 39 | All | 1.01 (0.91-1.11) | 0.856 | 0.7949 |
| | | | Serous | 1.02 (0.91-1.14) | 0.702 | |
| | | | Endometrioid | 0.96 (0.8-1.15) | 0.637 | |
| | | | **Mucinous** | **1.26 (1.02-1.55)** | **0.034** | |
| | | | Clear cell | 0.91 (0.72-1.15) | 0.437 | |

SNP order in haplotypes is 5' to 3' of the gene – haplotype block 2: rs10999147, rs3750772, rs4295944, rs2394644, rs10999152.

## Association between *RGC32* and ovarian cancer susceptibility

The only association between risk of ovarian cancer and *RGC32* was with a common variant with a minor allele frequency of 0.11. The rare allele of rs3783194 was associated with a 1.5-fold increase in the risk of the clear cell histological subtype – HetOR=1.5 (1.04-2.17), HomOR=1.99 (0.59-6.7), P=0.0206; see Table 3.24. Currently, it is not known if this variant tags any other SNPs in the gene or the regulatory regions up- or downstream of the gene. See Appendix IV-M for the results for the other common tSNPs of this gene.

There was also evidence of an association between a haplotype of *RGC32*, h0100100, and increased risk of clear cell ovarian cancer (OR=1.53 [1.11-2.11], P=0.01), see Table 3.25. This haplotype comprised of the rare allele of *RGC32* rs3783194, which was associated with an increase risk of clear cell ovarian cancer in the second position, thus lending support to the association with the h0100100 haplotype. No other common haplotype contained the rare allele of rs378319.

**Associations between *RBBP8* and ovarian cancer susceptibility**

In addition to the association between the *RBBP8* rs4474794 variant and a decrease in risk of serous histological subtype, the h0000 haplotype of the gene was associated with a marginal increase in the risk of the subtype – OR=1.13 (1.01-1.27), P=0.032. These haplotype-specific associations were concordant with the tSNP findings. The results for all tSNPs and haplotypes of *RBBP8* are shown in Appendices IV-K and IV-L, respectively.

**Associations between *AXIN2* and ovarian cancer susceptibility**

As well as the association between h1111, of *AXIN2* haplotype block 2 and increased risk of ovarian cancer in general – OR=1.21 (1.03-1.42), P=0.023. This association was also found when analysis was restricted to the serous subtype (OR=1.19 [1.01-1.39], P=0.037) – see Table 3.26. The rs11079571 variant, which was associated with disease risk, was in the second position of the haplotypes of this block (*AXIN2* block 2). The association between the rare allele of rs11079571 and increased risk of ovarian cancer was consistent with the correlation of the h1111 haplotype of *AXIN2* block 2 and increased risk of ovarian cancer. The rare allele of rs11079571 was not present in any other common haplotype (Appendix IV-F).

Table 3.24: Genotype-specific risks of *RGC32* (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR (95% CI)* | HomOR (95% CI)* | P-trend |
|---|---|---|---|---|---|---|---|---|
| *RGC32* | rs3783194 | 0.11 | 2723 | 1690 | All | 1.02 (0.88-1.19) | 1.03 (0.57-1.89) | 0.8873 |
| | | | | 788 | Serous | 0.89 (0.72-1.09) | 0.70 (0.29-1.69) | 0.1363 |
| | | | | 264 | Endometrioid | 1.13 (0.83-1.54) | 1.50 (0.51-4.36) | 0.4304 |
| | | | | 184 | Mucinous | 1.09 (0.75-1.59) | 1.08 (0.25-4.66) | 0.7964 |
| | | | | 155 | **Clear cell** | **1.50 (1.04-2.17)** | 1.99 (0.59-6.70) | **0.0206** |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.25: Haplotype-specific results of *RGC32* (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *RGC32* | h0100100 | 10.8 | All | 1.01 (0.87-1.18) | 0.851 | 0.6294 |
| | | | Serous | 0.89 (0.74-1.07) | 0.212 | |
| | | | Endometrioid | 1.02 (0.76-1.36) | 0.89 | |
| | | | Mucinous | 1.01 (0.72-1.41) | 0.962 | |
| | | | **Clear cell** | **1.53 (1.11-2.11)** | **0.01** | |

SNP order in haplotypes is 5' to 3' of the gene: rs10467472, rs3783194, rs11618371, rs9532824, rs995845, rs9594551, rs975590.

Statistically significant associations were also found with 2 haplotypes of *AXIN2* block 1, which had opposing effects on the risk of the serous histological subtype. The h000001 haplotype was associated with a reduced risk of the disease, OR=0.81 (0.68-0.97), P=0.018, (see Table 3.26). However, the h011001 haplotype of *AXIN2* block 1 was associated with an increased risk of the serous subtype (OR=1.21 [1.01-1.45], P=0.041). The results of the remaining haplotypes are shown in Appendix IV-F.

**Associations between *FILIP1L* and ovarian cancer susceptibility**

Although there was no evidence suggesting that a common variant of *FILIP1L* was associated with overall risk of ovarian cancer, statistically significant associations were found when the analysis was restricted to the endometrioid and mucinous histological subtypes. These associations were found with 3 variants of the gene, rs793446, rs17338680 and rs12494994, of which the rare alleles of all the variants were associated with increased risk of the endometrioid histological subtype (see Table 3.27 on page 181). The risks associated with carrying at least 1 of the rare alleles of rs793446, rs17338680 or rs12494994 ranged from 1.36 (for the rs793446 tSNP) to 1.71 (for rs17338680) for the endometrioid histological subtype (Table 3.27). The rs12494994 variant had the strongest association; HetOR=1.48 (1.08-2.04), HomOR=2.16 (1.13-4.12), P=0.0024. The heterozygous genotype of rs12494994 was also correlated with a 1.57-fold increase in the risk of the mucinous histological subtype, however, the connection was not statistically significant (P=0.2574). See Appendix IV-I for the genotype-specific results for all *FILIP1L* variants.

Table 3.26: Haplotype-specific risks of common *AXIN2* (P<0.05)

| Gene/haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *AXIN2* haplotype block 1 | h000001 | 14.6 | All | 0.9 (0.77-1.04) | 0.148 | 0.185 |
| | | | **Serous** | **0.81 (0.68-0.97)** | **0.018** | |
| | | | Endometrioid | 0.95 (0.73-1.24) | 0.7 | |
| | | | Mucinous | 1.08 (0.8-1.46) | 0.632 | |
| | | | Clear cell | 0.97 (0.7-1.36) | 0.868 | |
| | h011001 | 10.4 | All | 1.15 (0.98-1.36) | 0.082 | |
| | | | **Serous** | **1.21 (1.01-1.45)** | **0.041** | |
| | | | Endometrioid | 1.16 (0.87-1.54) | 0.312 | |
| | | | Mucinous | 1.22 (0.87-1.7) | 0.246 | |
| | | | Clear cell | 0.86 (0.57-1.29) | 0.458 | |
| *AXIN2* haplotype block | h1111 | 12.8 | **All** | **1.21 (1.03-1.42)** | **0.023** | 0.0847 |
| | | | **Serous** | **1.19 (1.01-1.39)** | **0.037** | |
| | | | Endometrioid | 1.08 (0.83-1.4) | 0.572 | |
| | | | Mucinous | 1.13 (0.83-1.53) | 0.434 | |
| | | | Clear cell | 0.96 (0.68-1.36) | 0.838 | |

SNP order in haplotypes is 5' to 3' of the genes –*AXIN2* haplotype block 1: rs11868547, rs7591, rs4074947, rs7210356, rs11655966, rs4541111. *AXIN2* haplotype block 2: rs4791171, rs11079571, rs3923087, rs3923086.

The correlation between rs793446 and rs17338680 (which are both in intron 4 of the gene) is $r^2$=0.173; rs793446 and rs12494994 is $r^2$=0.407; and rs17338680 and rs12494994 (intron 1) is $r^2$=0.404. The rs793446 variant tags 28 other SNPs with r2□0.8; rs17338680 tags 6 other SNPs and rs12494994 tags 11 other variants with r2□0.8. All of these SNPs are in the introns of the gene, and approximately 70% of them are conserved in mice.

Associations were also found between risk of endometrioid ovarian cancer and haplotypes of *FILIP1L*, see Table 3.28 for the significant associations, and Appendix IV-J for all results. *FILIP1L* comprises 2 haplotype blocks. The h10101 haplotype of haplotype block 1 had the strongest association with disease risk, with a 1.56-fold increase in odds; OR=1.56 (1.22-2.01), P=5.01x$10^{-4}$. The haplotype, which had a frequency of 10.7%, contained the rare alleles of rs793446 and rs17338680 in the third and last positions of the haplotype, respectively. The rare alleles of the variants were associated with an increased risk of endometrioid ovarian cancer, thus, the haplotype result was supported by the individual common tSNP findings.

There was also evidence suggesting a statistically significant association between a variant of *FILIP1L* haplotype block 2 and risk of the endometrioid subtype. The h001 haplotype was associated with an increased risk of the subtype; OR=1.37 (1.1-1.69), P=0.004 (see Table 3.28). This association was also supported by the individual SNP results - the rare allele of rs12494994, which was correlated to increased risk of the endometrioid histological subtype which was in the last position of the haplotype.

Table 3.27: Genotype-specific risks of common *FILIP1L* variants (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| *FILIP1L* | rs793446 | 0.41 | 2947 | 1773 | All | 1.05 (0.92-1.20) | 1.11 (0.93-1.32) | 0.3207 |
| | | | | 838 | Serous | 0.95 (0.80-1.13) | 1.07 (0.86-1.34) | 0.842 |
| | | | | 274 | **Endometrioid** | **1.36 (1.02-1.81)** | **1.52 (1.05-2.20)** | **0.0262** |
| | | | | 194 | Mucinous | 1.14 (0.82-1.58) | 0.93 (0.58-1.48) | 0.8885 |
| | | | | 164 | Clear cell | 1.05 (0.74-1.50) | 1.11 (0.69-1.77) | 0.6725 |
| | rs17338680 | 0.11 | 2989 | 1786 | All | 1.06 (0.91-1.24) | 0.77 (0.46-1.28) | 0.985 |
| | | | | 574 | Serous | 0.88 (0.69-1.14) | 0.98 (0.44-2.20) | 0.3051 |
| | | | | 221 | **Endometrioid** | **1.71 (1.24-2.36)** | 0.79 (0.19-3.39) | **0.0073** |
| | | | | 196 | Mucinous | 1.23 (0.86-1.76) | 0.35 (0.05-2.58) | 0.7109 |
| | | | | 133 | Clear cell | 1.04 (0.67-1.63) | 1.12 (0.26-4.81) | 0.8406 |
| | rs12494994 | 0.18 | 2347 | 1273 | All | 1.14 (0.98-1.33) | 0.88 (0.59-1.33) | 0.433 |
| | | | | 594 | Serous | 0.95 (0.78-1.17) | 0.52 (0.27-0.99) | 0.097 |
| | | | | 193 | **Endometrioid** | **1.48 (1.08-2.04)** | **2.16 (1.13-4.12)** | **0.0024** |
| | | | | 145 | Mucinous | **1.57 (1.10-2.25)** | 0.25 (0.03-1.83) | 0.2574 |
| | | | | 113 | Clear cell | 1.20 (0.79-1.82) | 1.69 (0.71-4.03) | 0.1986 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.28: Haplotype-specific risks of *FILIP1L* (P<0.05)

| Gene/hap block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *FILIP1L* haplotype block 1 | h10101 | 10.7 | All | 1.11 (0.91-1.35) | 0.29 | 0.5938 |
| | | | Serous | 0.94 (0.78-1.13) | 0.514 | |
| | | | **Endometrioid** | **1.56 (1.22-2.01)** | **$5.01x10^{-4}$** | |
| | | | Mucinous | 1.09 (0.78-1.51) | 0.617 | |
| | | | Clear cell | 1.25 (0.89-1.75) | 0.204 | |
| *FILIP1L* haplotype block 2 | h001 | 17.1 | All | 1.04 (0.92-1.17) | 0.562 | 0.7565 |
| | | | Serous | 0.88 (0.76-1.02) | 0.092 | |
| | | | **Endometrioid** | **1.37 (1.1-1.69)** | **0.004** | |
| | | | Mucinous | 1.03 (0.79-1.36) | 0.815 | |
| | | | Clear cell | 1.17 (0.88-1.56) | 0.269 | |

SNP order in haplotypes is 5' to 3' of the gene –*FILIP1L* (block 1): rs796977, rs793477, rs793446, rs3921767, rs17338680. *FILIP1L* (block 2): rs9864437, rs6788750, rs12494994.

Table 3.29: Haplotype-specific risks of *STAG3* (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *STAG3* | h000 | 50.7 | All | 1.06 (0.96-1.16) | 0.257 | 0.0979 |
| | | | **Serous** | **1.12 (1.01-1.25)** | **0.039** | |
| | | | Endometrioid | 1.06 (0.89-1.27) | 0.523 | |
| | | | Mucinous | 0.97 (0.79-1.2) | 0.805 | |
| | | | Clear cell | 1 (0.8-1.25) | 0.996 | |
| | h011 | 26.8 | All | 0.94 (0.85-1.05) | 0.29 | |
| | | | **Serous** | **0.88 (0.78-1)** | **0.046** | |
| | | | Endometrioid | 0.89 (0.73-1.09) | 0.251 | |
| | | | Mucinous | 1.22 (0.97-1.53) | 0.084 | |
| | | | Clear cell | 0.89 (0.69-1.16) | 0.401 | |

Freq – frequency; OR – odds ratio; CI – confidence interval; SNP order in haplotypes is 5' to 3' of the gene: rs11762932, rs2246713, rs1637001.

The *STAG3* rs1637001 variant was in the last position of a haplotype block comprising the 3 tSNPs genotyped. Associations were found between the serous histological subtype and two haplotypes of *STAG3*, Table 3.29. The h000 haplotype was associated with a marginal, 1.12-fold, increase in the risk of serous ovarian cancer (OR=1.12 [1.01-1.25], P=0.039). Conversely, the h011 haplotype was associated with a reduced risk of the serous subtype (0.88 [0.78-1], P=0.046). This

association was also supported by the association between the rare allele of rs1637001 tSNP (in the last position of the haplotype) and the decrease in risk of serous ovarian cancer. The rs1637001 polymorphism tags 17 other variants with $r^2 \geq 0.8$. Three of these SNPs, rs1623264, rs1727134 and rs1727128, are located in transcription factor binding sites. Several of the other polymorphisms may be involved in splicing.

**Associations between *AKTIP* and ovarian cancer susceptibility**

There was evidence of an association between the *AKTIP* gene and ovarian cancer susceptibility. The rare allele of rs718919 was associated with risk of the mucinous and clear cell histological subtypes. The associated risks were: HetOR=0.87 (0.64-1.19), HomOR=0.42 (0.21-0.84), P=0.0247 for the mucinous subtype; and HetOR=0.62 (0.44-0.87), HomOR=0.78 (0.45-1.35), P=0.0412 for the clear cell (Table 3.30). See Appendix IV-C for the genotype-specific susceptibility results of all the common variants of *AKTIP*. A haplotype of *AKTIP*, h1010, was also associated with a 0.73- and 0.77-fold decrease in the risk of both the mucinous and the clear cell subtypes, respectively (see Table 3.31). rs718919, the rare allele of which was associated with a reduced risk of mucinous and clear cell subtypes, was in the third position of the haplotype. These associations were concordant with each other. An additional association was found with the h1101 haplotype of *AKTIP* and the serous subtype. The h1101 haplotype was associated with increased risk of the serous subtype (OR=1.29 [1.01-1.66], P=0.044). Although the correlations were not statistically significant, with the exception of the mucinous subtype, the odds ratios for all cases of ovarian cancer, the endometrioid and clear cell subtypes were greater than 1 (see Table 3.31).

Table 3.30: Genotype-specific risks of an *AKTIP* tSNP (P<0.05)

| Gene | tSNP | MAF | Controls | Cases | Histology | HetOR (95% CI)* | HomOR (95% CI)* | P-trend |
|---|---|---|---|---|---|---|---|---|
| *AKTIP* | rs7189819 | 0.3 | 2923 | 1745 | All | 0.93 (0.82-1.06) | 0.92 (0.75-1.14) | 0.2796 |
| | | | | 825 | Serous | 0.99 (0.84-1.17) | 1.02 (0.78-1.33) | 0.867 |
| | | | | 271 | Endometrioid | 0.95 (0.73-1.24) | 1.06 (0.69-1.61) | 0.9177 |
| | | | | 186 | **Mucinous** | 0.87 (0.64-1.19) | **0.42 (0.21-0.84)** | **0.0247** |
| | | | | 163 | **Clear cell** | **0.62 (0.44-0.87)** | 0.78 (0.45-1.35) | **0.0412** |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

Table 3.31: Haplotype-specific risks for *AKTIP* (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value | Global P-value |
|---|---|---|---|---|---|---|
| *AKTIP* | h1010 | 30.4 | All | 1 (0.86-1.16) | 0.996 | 0.3703 |
| | | | Serous | 0.99 (0.88-1.11) | 0.858 | |
| | | | Endometrioid | 0.97 (0.8-1.18) | 0.766 | |
| | | | **Mucinous** | **0.73 (0.58-0.94)** | **0.013** | |
| | | | **Clear cell** | **0.77 (0.59-1)** | **0.047** | |
| | h1101 | 5.7 | All | 1 (0.75-1.33) | 0.999 | |
| | | | **Serous** | **1.29 (1.01-1.66)** | **0.044** | |
| | | | Endometrioid | 1.02 (0.65-1.59) | 0.927 | |
| | | | Mucinous | 0.94 (0.54-1.65) | 0.835 | |
| | | | Clear cell | 1.37 (0.82-2.3) | 0.227 | |

SNP order in haplotypes is 5' to 3' of the genes –*ATKIP*: rs9931702, rs17801966, rs7189819, rs3743772.

## 3.4: Admixture Maximum Likelihood test results

A large number of statistical tests are involved in the analysis of genetic association studies, however multiple testing corrections such as the Bonferoni correction are too stringent and do not take into account the correlation between SNPs. It has been suggested that the adjustment for "experiment-wise" type I error is more appropriate method for testing the global null hypothesis of no association within an experiment. These methods evaluate whether a greater than expected proportion of statistically significant associations are detect within an experiment. The admixture maximum likelihood (AML) test is a reportedly robust method for testing the global null hypothesis. The AML test simultaneously estimates the proportion of associated SNPs and their effect size. The AML test was used to evaluate the SNP genotyping data from 12 previous ovarian cancer case-control association studies for global evidence of associations between 340 SNPs from 84 genes and 10 chromosomal regions and the risk of the disease. The test was used to establish whether there was a statistically significant difference in the proportion of associations found from genetic susceptibility association studies of ovarian cancer and that which would have been found by chance.

### 3.4.1: Samples and methods

Genotyping data of 340 SNPs from three population-based case-control study series were analysed with the admixture maximum likelihood test. The studies comprised of up to 1,491 invasive epithelial ovarian cancer cases and 3,145 healthy controls from the GEOCS, MALOVA and SEARCH sample sets. The vast majority (>250) of the SNPs were tagging SNPs identified from 84 candidate genes from pathways, such as the cell cycle control, mismatch repair, DNA repair, oncogene and

differentially expressed genes with described functions from functional studies (microcell-mediated chromosome 18 transfer [MMCT-18) group), which have been implicated with ovarian cancer development.

Candidate SNPs from 10 different regions on chromosomes 2, 3, 5, 8, 11, 12 and 17 were also analysed. These variants had originally been selected for validation from the Ovarian Cancer Association Consortium (OCAC), Breast Cancer Association Consortium (BCAC) or the breast cancer genome-wide association study due to associations with breast or ovarian cancer. Associations between individual SNPs and ovarian cancer risk with unconditional logistic regression, and the heterogeneity and trend test. The admixture maximum likelihood test was used on groups of SNPs to determine whether the proportion of associations found were greater than that which would be expected.

All variants analysed were allocated into a group based on known or putative function, or the research consortia from which the candidate SNP had come. There were a total of 7 groups – BCAC (16 SNPs), cell cycle control (101 tSNPs), DNA repair (28 SNPs), mismatch repair (43 tSNPs), MMCT (consisting of differentially expressed genes from functional tumour suppression experiments- 63 tSNPs), OCAC (55 SNPs) and ovarian cancer oncogenes (34 tSNPs). The genotype distribution for all SNPs analysed are shown in Appendix II-B.

### 3.4.2: Logistic regression results (unadjusted)

When the trend model was used to test for association, 22 (6.5%) of the 340 SNPs were significantly associated with ovarian cancer risk at the 5% level, and 5 SNPs

(1.5%) were significant at the 1% level. Of the 5 most significant SNPs with the trend model, two variants were from the BCAC group (rs2107425 on chromosome 11p15.5 and rs3817198 of *LSP1*), another two were from the cell cycle group (*CDKN1B* rs2066827 and CDK6 rs8) one SNP was from the OCAC group (*ESR1* rs9322336).

Adjustments for population stratification by genomic control were made in order to ensure that the associations found were due to the variants analysed, rather than underlying structure of the population. Logistic regression analyses were also stratified by sample sets to account for population stratification. Following adjustments of genomic controls for population stratification, 18 (5.3%) of the 340 variants were now statistically significant at the 5% level. This was a reduction of 4 SNPs compared to the unadjusted findings. The same number of SNPs (5 [1.5%]) were significant at the 1% level, after adjustments for population stratification (Table 3.32).

After analysis with the heterogeneity test, 17 (5%) of the SNPs were significant at the 5% level, 6 SNPs (1.8%) were significant at the 1% level, and one at the 0.001% significance level. After adjusting for population stratification, 15 of the 17 SNPs significantly associated with ovarian cancer risk at the 5% level, remained significant. One of the 6 SNPs, significant at the 1% level with the heterogeneity test, was no longer significant, thus 5 (1.5%) SNPs remained significant at the 1% level. The only SNP, which reached 0.001% level of significance, remained at the same level after adjustments for population stratification.

Both the heterogeneity and trend tests detected associations between the same 9 SNPs and risk of epithelial ovarian cancer. Although there was a slight attenuation in the P-values after adjustments for population stratification, the significant associations remained. The unadjusted and adjusted trend test results for each SNP are shown in Appendix VI. The results of the trend test are illustrated in Figure 3.6 as a quantile-quantile (Q-Q) plot. Q-Q plots are probability plots used for comparing 2 probability distributions. In order for the probability distributions to be compared, the quantiles of the distributions are plotted against each other.

The Q-Q plot in Figure 3.6 shows the ordered observed trend test statistics plotted against the expected trend $\chi^2$ results given the rank. The line of equivalence is the straight line through the plot. This line is used as a reference for no difference between the observed and expected $\chi^2$ values, given the rank. Deviation from the line of equivalence suggests differences between the observed and expected $\chi^2$ values. In Figure 3.6, the plots of both the unadjusted and adjusted trend test results suggested that a greater proportion of associations were found than expected. In the Q-Q plot shown in Figure 3.6, the plot followed the line of equivalence for the first 240 SNPs, and then started to deviate. This indicates that a modest number of SNPs were associated with ovarian cancer risk.

Table 3.32: AML - SNPs with significant associations (trend test for association)

| Group | Gene location | SNP | MAF | Controls | Cases | HetOR‡ (95% CI) | HomOR‡ (95% CI) | Unadjust *P*-het* | Adjust. P-het§ | Unadjusted *P*-trend | Adjusted *P*-trend§ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BCAC | 11p15.5 | rs2107425 | 0.32 | 1460 | 2463 | **0.71 (0.62-0.82)** | 0.88 (0.70-1.10) | **$1.28x10^{-5}$** | **$2.17x10^{-5}$** | **0.0012** | **0.0019** |
| OCAC | *ESR1* | rs9322336 | 0.23 | 1453 | 2464 | **0.81 (0.70-0.93)** | 0.73 (0.52-1.02) | **0.005** | **0.006** | **0.0013** | **0.0021** |
| BCAC | *LSP1* | rs3817198 | 0.3 | 1457 | 2435 | **1.16 (1.01-1.34)** | **1.40 (1.11-1.75)** | **0.006** | **0.009** | **0.0016** | **0.0026** |
| Cell cycle | *CDKN1B* | rs2066827 | 0.26 | 1481 | 2484 | 0.88 (0.77-1.01) | **0.68 (0.51-0.90)** | **0.011** | **0.019** | **0.0035** | **0.0053** |
| Cell cycle | *CDK6* | rs8 | 0.21 | 1473 | 2481 | **1.17 (1.02-1.35)** | **1.44 (1.04-1.99)** | **0.015** | **0.013** | **0.0039** | **0.0059** |
| Mismatch | *PMS2* | rs7797466 | 0.18 | 1305 | 1968 | **1.18 (1.01-1.38)** | 1.38 (0.96-2.00) | **0.039** | **0.044** | **0.0108** | **0.0142** |
| Cell cycle | *CCND1* | rs603965 | 0.44 | 1476 | 2464 | 1.06 (0.91-1.23) | **1.28 (1.06-1.55)** | **0.027** | **0.032** | **0.013** | **0.0178** |
| MMCT-18 | *RUVBL1* | rs13063604 | 0.22 | 564 | 785 | 1.23 (0.98-1.56) | **1.54 (1.00-2.39)** | 0.0556 | 0.058 | **0.016** | **0.0181** |
| OCAC | *PGR* | rs1042838 | 0.14 | 1424 | 2408 | **1.25 (1.07-1.46)** | 1.09 (0.73-1.64) | **0.019** | **0.023** | **0.0161** | **0.0215** |
| Cell cycle | *CCND1* | rs7178 | 0.07 | 1480 | 2491 | **1.24 (1.04-1.49)** | 1.24 (0.50-3.04) | 0.063 | 0.072 | **0.021** | **0.0278** |
| OCAC | *IL18* | rs1834481 | 0.25 | 1449 | 2435 | 0.89 (0.77-1.02) | 0.77 (0.59-1.01) | 0.074 | 0.083 | **0.0227** | **0.0295** |
| Cell cycle | *CCND1* | rs602652 | 0.46 | 1468 | 2493 | 1.13 (0.97-1.32) | **1.24 (1.03-1.49)** | 0.074 | 0.084 | **0.0235** | **0.0307** |
| OCAC | IGF2 | rs4320932 | 0.2 | 1473 | 2402 | **0.84 (0.73-0.97)** | 0.86 (0.60-1.22) | 0.0529 | 0.061 | **0.0243** | **0.0314** |
| MMCT-18 | *CASP5* | rs518604 | 0.44 | 1041 | 2029 | 1.11 (0.93-1.33) | **1.27 (1.02-1.58)** | 0.0987 | 0.072 | **0.032** | **0.0387** |
| Cell cycle | *CCND1* | rs3212879 | 0.49 | 1472 | 2491 | **0.85 (0.73-0.99)** | **0.82 (0.68-0.99)** | 0.063 | 0.108 | **0.0321** | **0.0409** |
| DNA | *XRCC2* | rs3218536 | 0.08 | 1337 | 1787 | 0.88 (0.72-1.08) | **0.23 (0.07-0.79)** | **0.014** | **0.017** | **0.0364** | **0.0439** |
| Cell cycle | *CCND1* | rs3212891 | 0.46 | 1475 | 2476 | **0.86 (0.74-1.00)** | **0.83 (0.69-1.00)** | 0.082 | 0.092 | **0.0376** | **0.0472** |
| Mismatch | *PMS1* | rs256563 | 0.12 | 1456 | 2446 | 2.50 (0.99-6.33) | 2.15 (0.84-5.48) | **0.0435** | 0.134 | **0.04** | 0.05 |
| BCAC | 8q24.21 | rs10808556 | 0.4 | 1462 | 2453 | 1.15 (0.99-1.33) | 1.20 (0.99-1.46) | 0.1071 | 0.119 | **0.0446** | 0.0552 |
| Cell cycle | *CDKN2A* | rs3731257 | 0.26 | 1480 | 2476 | 0.89 (0.78-1.03) | 0.80 (0.60-1.07) | 0.1345 | 0.148 | **0.0451** | 0.056 |
| Cell cycle | *CCNE1* | rs3218036 | 0.31 | 1476 | 2481 | 1.07 (0.93-1.23) | **1.27 (1.01-1.59)** | 0.1126 | 0.125 | **0.0458** | 0.0567 |
| OCAC | IGF2 | rs1003483 | 0.49 | 1459 | 2407 | **1.20 (1.02-1.40)** | **1.20 (1.00-1.44)** | 0.0611 | 0.07 | **0.0473** | 0.0581 |

‡ compared with common homozygous; HetOR – heterozygous odds ratio, HomOR – homozygous odds ratio; CI – confidence interval; *P-heterogeneity; § Adjusted for population stratification.

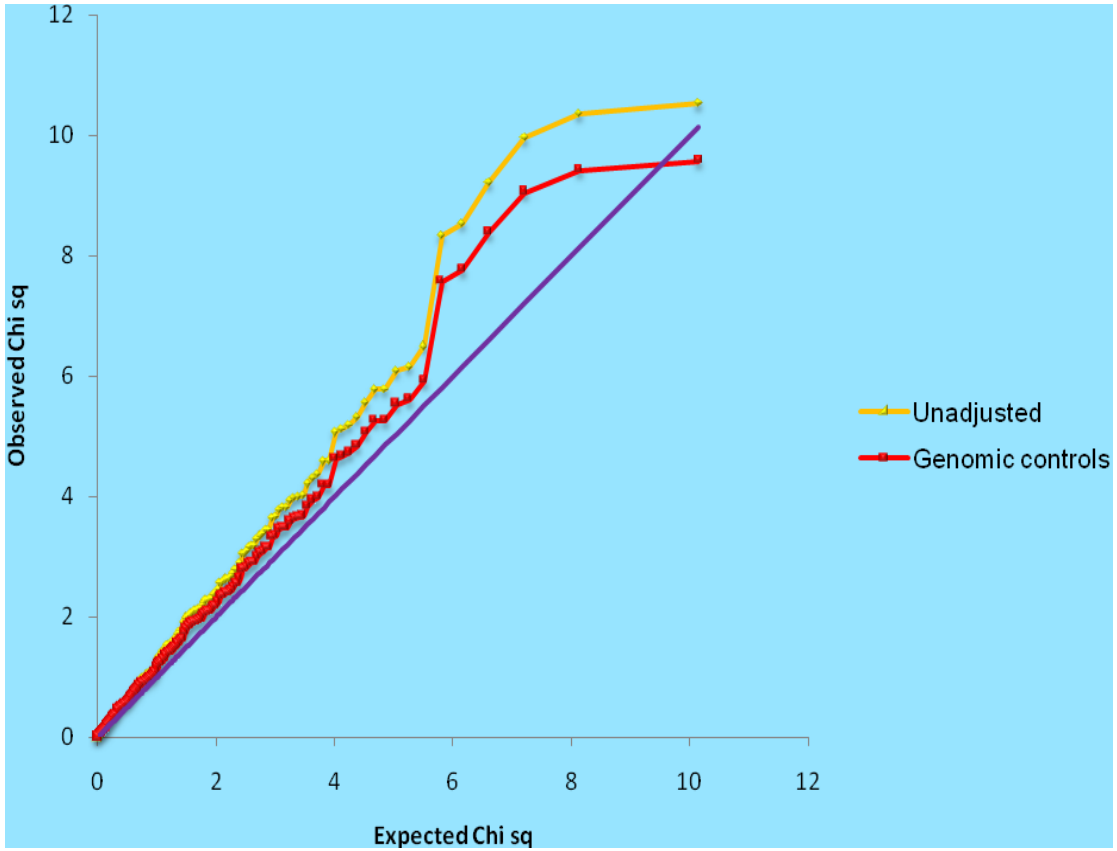



Figure 3.6: Quantile-quantile plot of the univariate trend test results

### 3.4.3: AML results (adjusted for population stratification)

The genomic control method for adjusting for cryptic population stratification was used on the variants analysed. Genotyping data from breast cancer case-control samples from the genome-wide association study (Easton *et al.* 2007; Hunter *et al.* 2007) were used to estimate the degree of over-dispersion of statistics, also known as inflation test statistic (Pharoah *et al*. 2007). The results from the breast cancer were used to estimate the level of stratification within Caucasian populations. A more conservative inflation statistic of 10%, was used to adjust the P-trend for cryptic population stratification.

Eight of the 22 SNPs that were statistically significant at the 5% level with the trend test belonged to the mitotic cell cycle control pathway group (Table 3.32). This group consisted of 101 SNPs from 15 genes, which have been demonstrated to be involved in the regulation of progression through the cell cycle. The most significant tSNP in the cell cycle group was rs2066827 in the cyclin-dependent kinase inhibitor 1B (*CDKN1B*) gene. The rare allele of this SNP was associated with a decrease in the risk of ovarian cancer (HetOR=0.88 [0.77-1.01], HomOR=0.68 (0.51-0.9), adjusted (for population stratification) P-het=0.019, adjusted P-trend=0.0059. The rs2066827 variant (the fourth most significant SNP with the trend model) is a missense SNP located in exon 1 of *CDKN1B*. The common allele encodes a valine amino acid, and the rare allele, which has a frequency of 26%, glycine.

Five variants from the 55 SNPs in the OCAC group were significant at the 5% level after adjustments for population stratification. The most significant SNP from the

OCAC group, rs9322336, was from the oestrogen receptor gene (*ESR1*). This variant was the second most significant of the SNPs analysed with the trend test, and the third most significant with the heterogeneity test. The rare allele of this variant was also associated with a reduced risk of ovarian cancer – HetOR= 0.81 (0.70-0.93), HomOR=0.73 (0.52-1.02), adjusted P-trend=0.0021.

Three of the most significant SNPs were from the BCAC group, which comprised the 16 variants which were identified from genome wide association studies to be strongly associated with breast cancer risk. One of these variants, rs2107425, was associated with a decrease in risk of ovarian cancer (hetOR=0.71 [0.62-0.82], HomOR=0.88 [0.70-1.10], adjusted P-trend=0.0019). rs2107425 is located on chromosome 11p15.5 in a region with no known genes or open reading frame. This variant had the strongest association with ovarian cancer risk, with both the trend and heterogeneity models – and the p-value for the heterogeneity test reached a level of significance deemed to provide definitive evidence of association ($P<1x10^{-4}$) in case-control association studies, however, not enough for genome-wide significance ($P<1x10^{-7}$).

Two variants from the functional candidate genes (of 63 SNPs) and the DNA mismatch repair pathway (of 43) groups were also statistically significant. The most significant SNP from the functional candidate group was the intronic rs13063604 in the *RUVBL1* gene on chromosome 3. *RUVBL1* rs13063604 was associated in an increased risk of ovarian cancer with the trend model (HetOR=1.23 (0.98-1.56), HomOR=1.54 (1.00-2.39), adjusted P-het=0.058, P-trend=0.0181. Incidentally an association was also found with this SNP with 1,755 serous histological subtype

cases and 4,363 controls from 7 different population based case-control series including SEARCH (DOVE, GEOCS, HOPE, JAC, UKOPS and USC). A single variant from *XRCC2* gene of the DNA double strand break repair pathway group (28 SNPs), rs3218536, was also among the significant associations with the trend test, with a correlation with a reduction in ovarian cancer risk (HetOR=0.88 [0.72-1.08], HomOR=0.23 [0.07-0.79], adjusted P-het=0.017, adjusted P-trend=0.0439.

Of the 15 associations identified with the heterogeneity test at the 5% significance level after adjustments for population stratification, 2 variants were from the BCAC group, 7 were from the cell cycle control pathway, one from the DNA repair pathway, two from the mismatch repair pathway, two from the MMCT-18 functional group, two from the OCAC and one from the oncogene pathway (of 34 tSNPs). Seven of the associations found with the heterogeneity test were not identified with the trend test.

The AML method was used to test for association of the SNPs according to functional group, biological pathway or genotyping group. There was evidence suggesting that the breast cancer associated group of SNPs, identified by genome wide association studies, was significantly associated with ovarian cancer risk (*P*-trend = 0.0028; adjusted *P*-trend = 0.0059). The statistically significant findings suggest that there were a greater number of variants observed to be associated with ovarian cancer risk, than that would have been expected by chance.

Table 3.33: AML experiment-wise test results for genotyping groups

| Pathway / Group | Genes/ regions‡ | No. SNPs | LR *P*-trend of most significant SNP* | AML *P*-het* | AML *P*-trend* | Reference with original single SNP analysis using logistic regression |
|---|---|---|---|---|---|---|
| BCAC† | 5 (5§) | 16 | 0.0012 | 0.0003 | 0.0028 | (Song *et al*. 2009a) |
| OCAC† | 36 (6§) | 55 | 0.0014 | 0.863 | 0.806 | (Palmieri *et al.* 2008; Pearce *et al.* 2008; Ramus *et al.* 2008b) |
| MMCT-18 | 9 | 63 | 0.016 | 0.609 | 0.468 | (Notaridou *et al.* 2010) |
| Cell cycle control | 15 | 101 | 0.0035 | 0.274 | 0.225 | (Dicioccio *et al.* 2004; Song *et al.* 2006b; Gayther *et al.* 2007) |
| Mismatch repair | 7 | 43 | 0.0106 | 0.706 | 0.702 | (Song *et al.* 2006a) |
| DNA repair | 7 | 28 | 0.0374 | 0.366 | 0.444 | (Auranen *et al.* 2005; Song *et al.* 2007) |
| Ovarian Cancer Oncogenes | 5 | 34 | 0.0671 | 0.524 | 0.528 | (Quaye *et al.* 2009) |
| **Total** | **84 (10)** | **340** | | **0.051** | **0.068** | |

*Based on GEOCS, MALOVA and SEARCH genotypes; ‡ SNPs in regions with no known genes or open reading frames are in parenthesis;† candidate genes identified from the Breast Cancer Association Consortium (BCAC) and Ovarian Cancer Association Consortium (OCAC); § different SNPs from 8q24.21 were genotyped in both BCAC and OCAC sets; LR – logistic regression; AML – admixture maximum likelihood; het – heterogeneity.

There was no evidence that there were a significant proportion of variants from the remaining groups (cell cycle control, DNA repair, mismatch repair, MMCT-18, OCAC and ovarian cancer oncogenes) associated with ovarian cancer risk than that which would have been expected by chance. When the genotyping data from all groups were combined and analysed, the AML experiment-wise test for association was not significant for either the heterogeneity test ($P$=0.051) or the trend test ($P$=0.068). This suggests that there is a trend towards a proportion of the SNPs evaluated being associated with disease, however this is not statistically significant, and the effect sizes were too small to detect for individual SNPs. Table 3.33 shows the results of the AML experiment-wise tests summarised for the complete set of

SNPs categorised according to functional group, biological pathway or genotyping group.

## 3.5: Summary

The effects of 34 tSNPs of *BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA* on susceptibility of ovarian cancer were evaluated with 1,816 invasive epithelial ovarian cancer cases and 3,000 unaffected controls. There was evidence of association between risk of ovarian cancer and all the candidate genes. Three tSNPs of both *BRAF* and *KRAS* were associated with the risk of the mucinous histological subtype. The associations between the mucinous subtype and *BRAF* and *KRAS* also extended to haplotypes of these oncogenes. These findings are of particular interest because *KRAS* and, to a lesser extent, *BRAF* mutations are predominantly found in mucinous ovarian tumours, and these mutations are early events in the development of the disease. Moremover, a haplotype of *BRAF*, h00100000, was associated with a decrease in the risk all subtypes of ovarian cancer. This association remained when the analysis was restricted to the serous subtype.

Common polymorphic variants of *ERBB2* (non-synonymous coding SNP, rs1801200) and *PIK3CA* (rs2865084) were marginally associated with risk of the endometrioid subtype. There was also evidence suggesting that two haplotypes of *ERBB2*, h110 and h001, which had opposite alleles at every position, were associated with increased risk of ovarian cancer (all subtypes). These associations may be caused by an unknown polymorphism which tags both haplotypes.

Furthermore, the haplotypes of *BRAF* and *ERBB2* were globally associated with ovarian cancer susceptibility, (P=0.005 and P=0.034, respectively).

A statistically significant association was found between the rs11683487 variant of *NMI* and ovarian cancer. This SNP was associated with serous and endometrioid subtypes when the analysis was restricted to the histological subtypes. The finding was not replicated with additional 1,097 cases and 1,712 controls in stage 2. However, when the genotyping data from both stages of the study were combined, the association with all histological subtypes, and the mucinous subtype remained. Two haplotypes of *NMI* were also associated with all subtypes combined. There results were also found when analysis was restricted to the serous subtype.

The effects of 63 tSNPs and haplotypes of candidate genes (from differentially expressed genes with described function from *in vitro* neoplastic suppression studies) on the risk of ovarian cancer were analysed with 1,799 ovarian cancer cases and 3,045 controls. There was evidence of association between ovarian cancer susceptibility and all of the differentially expressed genes (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*). A common variant of *AXIN2* (11079571), *CASP5* (rs518604) and *RUVBL1* (rs13063604) were associated with the risk of ovarian cancer when it is considered as a single disease. The TaqMan probes for the *CASP5* and *AXIN2* SNPs failed probe testing, therefore they could not be validated with additional samples.

The association between the rare allele of *CASP5* rs518604 and increased risk of ovarian cancer remained when analysis was restricted to the serous subtype. The

*CASP5* block 1 haplotypes, h100 and h000, were also associated with the risk of ovarian cancer (P=0.012 and P=0.015, respectively). h100 and h010 of *CASP5* haplotype block 1 were also associated with the risk of the serous subtype. The haplotypes of *CASP5* block 1 were globally, strongly, associated with the risk of ovarian cancer ($P=8.43x10^{-6}$).

*RUVBL1* rs13063604 was not independently validated with the stage 2 samples (4,590 cases and 6,031 controls) alone, however, the association remained statistically significant when the genotyping data from stages 1 and 2 were combined (P=0.033). rs13063604 and another tSNP of *RUVBL1*, rs7650365, were associated with risk of the serous subtype with stage 1 samples, P=0.002 and P=0.009, respectively. Neither of these associations were independently validated with stage 2 samples and only the association between the rare allele of rs13063604 and increased risk of the serous subtype remained when the data from the 2 genotyping stages were combined (HetOR=1.13 [1-1.27], HomOR=1.22 [0.96-1.56], P=0.019). Two haplotypes of *RUVBL1* were also associated with the risk of the serous subtype, and globally, the haplotypes of *RUVBL1* were associated with ovarian cancer susceptibility (P=0.0016).

Associations were also found between the risk of the serous histological subtype of ovarian cancer and common a variant of *RBBP8* and *STAG3*, and haplotypes of *AKTIP*, *AXIN2*, as well as *RBBP8* and *STAG3*. Interestingly, associations were found between three tSNPs and two haplotypes of *FILIP1L* and the risk of endometrioid ovarian cancer. The tSNPs and the SNPs they tag were intronic, and some were conserved in mice. A common tagging variant of *AKTIP* was associated

with risk of mucinous and clear cell histological subtypes. There was also evidence of association between a haplotype of *AKTIP* and risk of serous ovarian cancer. Another haplotype of *AKTIP* was associated with reduced susceptibility to mucinous and clear cell disease.

Although many of the associations appear to be of great interest, it is important to take into consideration that the results are based on relatively small samples, particulary when the analyses were restricted to the histological subtypes. Many statistical tests were performed in the analyses, however, there was no correction for multiple testing, which may render many, if not, all associations statistically significant. Nonetheless, the tagging approach of genetic associations attempts to identify markers, rather than the causal genetic locus.

The admixture test (AML) was used to establish whether there was a statistically significant difference in the proportion of associations found from genetic susceptibility association studies of ovarian cancer and that which would have been found by chance. A modest number of SNPs were associated with predisposition of ovarian cancer. When the AML method was used to evaluate SNPs which were grouped according to their proposed function, biological pathway or validation study, only the BCAC group was statistically significant for an excess of positive associations. The SNPs within this group were those which were highly associated with risk of breast cancer. Three (19%) out of 16 tSNPs in the BCAC group were significantly associated with risk of ovarian cancer. The experiment-wise test of the 340 SNPs analysed was not significant (P=0.068).

# Chapter 4: Results - Common germline variants in candidate ovarian cancer genes and survival of patients with invasive epithelial ovarian cancer

## 4.1: Introduction

Hypothesis:

Common germline genetic variants in candidate genes associated with ovarian cancer development can influence the clinical outcome (survival) of patients diagnosed with invasive epithelial ovarian cancer.

Aims:

(1) To evaluate the effect of tSNPs and haplotypes from candidate oncogenes on all-cause mortality of ovarian cancer patients.

(2) To investigate the effect of tSNPs and haplotypes in a series of "functional" candidates identified from *in vitro* studies on all-cause survival of ovarian cancer patients.

Objectives

(1) To assess the effects of common germline genetic variants and haplotypes of candidate oncogenes and functional genes on clinical outcome of ovarian cancer patients using univariate Cox regression survival analysis.

(2) To evaluate the effects of the common germline variants and haplotypes in candidate genes on overall survival after restricting the analysis to the major histological subtypes of ovarian cancer (serous, endometrioid, mucinous and clear cell).

(3) To examine the effects of other prognostic factors such as age at diagnosis, tumour histological subtype, grade and stage on clinical outcome.

(4) To investigate the effects of the common genetic variants and haplotypes of candidate genes, after adjustments for confounding prognostic factors, on clinical outcome (using multivariate Cox regression survival analysis).

The effects of overall survival for tSNPs and haplotypes of candidate oncogenes and a series of functional candidate genes identified from in vitro modelling studies in patients with ovarian cancer over a 10 year period were investigated. The oncogenes (*BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA*) and functional candidates (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) were selected because of their putative role in ovarian cancer development.

Cox regression survival analysis was used to establish the effects of the genetic variants and haplotypes on all-cause mortality in ovarian cancer patients. Cox regression survival analysis was also used to evaluate the effects of clinical, prognostic factors on patient survival in order to make appropriate adjustments for these potentially confounding factors. Clinical factors which were found to be significantly associated with all cause mortality were adjusted for all common variants and haplotypes, in order to determine if true associations with the genetic

factors existed. Survival analysis was performed on a total of 2,021 invasive epithelial ovarian cancer cases.

## 4.2: Survival analyses of variants and haplotypes of candidate oncogenes

Thirty-four tSNPs identified in the candidate oncogenes (*BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA*) were genotyped in a total of 1,572 invasive epithelial ovarian cancer cases from 4 population-based series: GEOCS (327 cases), MALOVA (445 cases), SEARCH (708 cases) and UKOPS (92 cases). Together, these cohorts included a total of 662 deaths in 6,467 person-years at risk. The time at risk was calculated by the summation of the time (years) from entry into the study until an individual died or was censored from the study. The effects of the tSNPs on all-cause mortality were investigated using Cox regression survival analysis. All reported values are based on likelihood ratio test for trend (1 degree of freedom).

### 4.2.1:Univariate survival analysis results of *BRAF*

There was evidence of a statistically significant association between a common genetic variant of *BRAF*, rs6944385, and all-cause mortality of ovarian cancer patients, with the univariate survival model. The univariate survival model contained terms for the common variant, stratified by population set because there were significant differences in the survival of patients in the different data sets. The rare allele of rs6944385 was associated with poor survival (per-rare allele hazard ratio (HR) =1.19 (95% confidence interval 1.03-1.38, P=0.021). The rare allele of the tSNP resulted in a 1.19-fold increase in mortality, compared with the common

allele. This suggested that the addition of a rare allele increases the hazard ratio by 1.19. Thus for the additive model, heterozygotes have an increased hazard of 1.19, and rare homozygotes have a 2.38-fold increase in hazard, compared with common homozygotes. The hazard ratio measures the effect of the explanatory factor (allele) on the risk (hazard) of death.

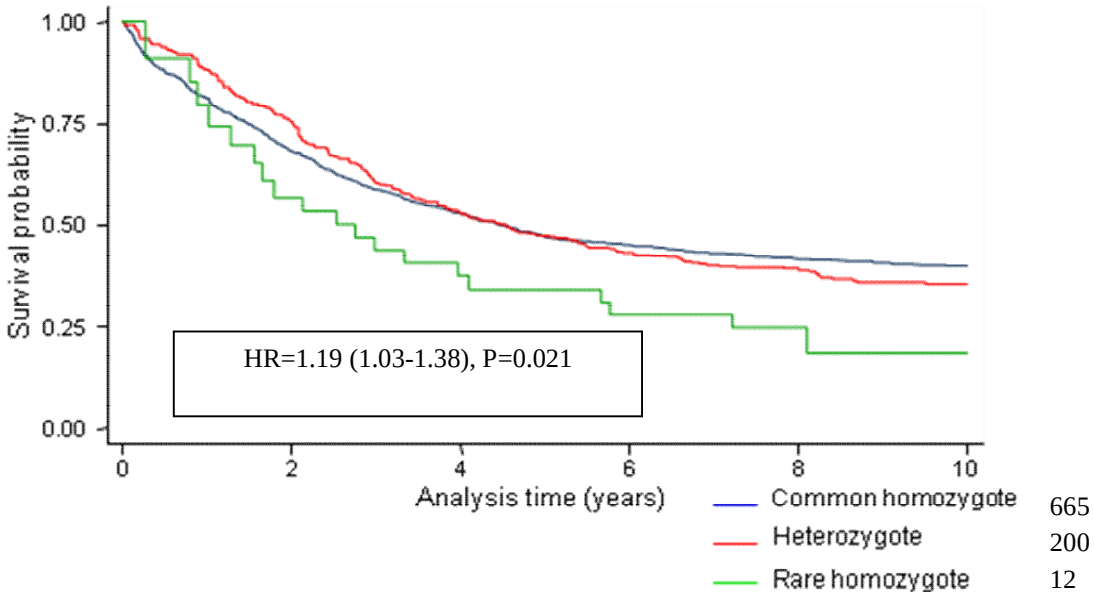



Figure 4.1: Kaplan-Meier survival estimates of *BRAF* rs6944385 (all cases)

Numbers following the keys are individuals still at risk after 10 years.

Figure 4.1 shows the plot of the Kaplan-Meier survival estimates of the different genotypes of *BRAF* rs6944385 over a 10-year period. Kaplan-Meier survival estimates were used to illustrate the survival function of ovarian cancer patients grouped according to their genotype for a particular tSNP. The survivor curves are step functions that decrease (step-down) at the time points when patients die (Everitt and Palmer 2005). The figure clearly shows a worse survival associated with rare

homozygotes of rs6944385 compared with common homozygotes and heterozygotes. Half of the patients homozygous for the rare allele of rs6944385 survived for 2.5 years after diagnosis, which was approximately 1.5 years less than the survival of the common homozygotes and heterozygotes.

The association observed with the *BRAF* rs6944385 variant and survival of all histological subtypes combined was more significant, with increased hazard when analysis was restricted to the clear cell histological subtype. The rare variant of rs6944385 was associated with a 2.2-fold increase in mortality compared with the common allele, HR=2.22 (1.18-4.17), P=0.014, see Table 4.1.

Table 4.1: Univariate Cox regression results of *BRAF* rs6944385, by histology

| Gene | tSNP | MAF | Histology | Cases | Univariate | |
|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value |
| *BRAF* | rs6944385 | 0.14 | **All** | 1758 | **1.19 (1.03-1.38)** | **0.021** |
| | | | Serous | 840 | 0.97 (0.79-1.2) | 0.804 |
| | | | Endometrioid | 268 | 1.31 (0.84-2.07) | 0.235 |
| | | | Mucinous | 187 | 0.83 (0.4-1.73) | 0.614 |
| | | | **Clear cell** | 124 | **2.22 (1.18-4.17)** | **0.014** |

HR - Hazard ratio; CI - confidence interval; MAF- minor allele frequency; Emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

Although associations may not be found with individual tSNPs, different combinations of SNPs forming haplotypes may affect survival from ovarian cancer. When the effects of the *BRAF* haplotypes on survival from ovarian cancer were evaluated, none of the common haplotypes of *BRAF* were statistically associated with survival from ovarian cancer. However, the 95% confidence intervals of all histological subtypes of the h01100001 haplotype did not cross 1, HR=1.21 (1-1.46),

P=0.055. This suggests that the haplotype may be marginally correlated with increased mortality of ovarian cancer.

### 4.2.2: Univariate survival analysis results of *KRAS*

When Cox regression survival analysis was used to assess the effect of *KRAS* variants on the survival from ovarian cancer, a statistically significant association was found between a common tSNP, rs10842513, and all-cause mortality of serous histological subtype cases. The rare allele of rs10842513 was associated with poor survival (HR=1.38 (1.09-1.75), P=0.008), see Table 4.2. Although, this variant was not statistically associated with increased mortality of patients with clear cell ovarian cancer, the 95% confidence interval did not cross 1, which suggests a marginal correlation (Table 4.2).

Table 4.2: Univariate Cox regression results of common tSNPs of *KRAS* (P<0.05)

| Gene | tSNP | MAF | Histology | Cases | Univariate | |
|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value |
| *KRAS* | rs10842513 | 0.09 | All | 1770 | 1.18 (0.98-1.42) | 0.08 |
| | | | **Serous** | 846 | **1.38 (1.09-1.75)** | **0.008** |
| | | | Endometrioid | 271 | 1.19 (0.67-2.1) | 0.552 |
| | | | Mucinous | 187 | 0.7 (0.29-1.69) | 0.432 |
| | | | Clear cell | 132 | **2.02 (1-4.1)** | 0.052 |
| | rs4623993 | 0.16 | All | 1748 | 0.93 (0.8-1.09) | 0.378 |
| | | | Serous | 834 | 0.89 (0.73-1.08) | 0.236 |
| | | | Endometrioid | 242 | 0.83 (0.51-1.36) | 0.463 |
| | | | **Mucinous** | 187 | **1.79 (1.02-3.15)** | **0.044** |
| | | | Clear cell | 136 | 0.93 (0.46-1.89) | 0.835 |

HR - Hazard ratio; CI - confidence interval; MAF- minor allele frequency; Emboldened histologies are statistically associated with survival; emboldened HR are statistically significant or the CI does not cross 1.

Another variant of *KRAS*, rs4623993, was associated with survival of mucinous cases. The rare allele of *KRAS* rs4623993 was associated with poor survival of

individuals with mucinous ovarian cancer, HR=1.79 (1.02-3.15), P=0.044 (see Table 4.2).

The effect of both rs10842513 and rs4623993 on the mucinous subtype, although not statistically significant for the previous tSNP, was the opposite of the other histological subtypes (see Table 4.2). This difference in the ratios was also found with another tSNP of *KRAS*, rs4623993, which was significantly associated with the increased risk of the mucinous subtype. These marked differences between mucinous disease and the other subtypes may be a result of putative involvement of the *KRAS* gene in the development of mucinous ovarian cancer.

Table 4.3: Univariate Cox regression results of *KRAS* haplotypes (P<0.05)

| Gene | Haplotype† | Freq (%) | Histology | Univariate | |
|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value |
| *KRAS* haplotype block 2 | h010000 | 5.9 | All | 1.27 (0.99-1.62) | 0.056 |
| | | | **Serous** | **1.69 (1.21-2.36)** | **0.002** |
| | | | Endometrioid | 1.21 (0.59-2.48) | 0.599 |
| | | | Mucinous | 0.66 (0.17-2.55) | 0.55 |
| | | | Clear cell | 2.81 (0.95-8.33) | 0.062 |
| | h001100 | 3.7 | All | 1.02 (0.79-1.31) | 0.902 |
| | | | Serous | 0.87 (0.62-1.21) | 0.411 |
| | | | Endometrioid | 0.99 (0.44-2.18) | 0.971 |
| | | | **Mucinous** | **3.24 (1.55-6.74)** | **0.002** |
| | | | Clear cell | 2.42 (0.6-9.66) | 0.212 |
| | h000000 | 2.6 | All | 1.26 (0.87-1.82) | 0.219 |
| | | | Serous | 0.96 (0.61-1.53) | 0.872 |
| | | | Endometrioid | 2.47 (0.84-7.23) | 0.099 |
| | | | **Mucinous** | **6.59 (1.37-31.62)** | **0.018** |
| | | | Clear cell | 1.53 (0.32-7.36) | 0.593 |

†: '0'= common allele and '1'= rare allele; HR - Hazard ratio; CI - confidence interval; Emboldened HR are statistically significant or the CI does not cross 1; Emboldened haplotypes are statistically significant; SNP order in haplotypes (5' to 3' of the genes) –*KRAS* - block 2: rs12579073, rs10842513, rs4623993, rs6487464, rs10842514, rs11047917.

Statistically significant associations were found between 3 haplotypes of *KRAS* block 2 and survival of patients with serous and mucinous histological subtypes of ovarian cancers. The h010000 haplotype, which has a frequency of 5.9%, was associated with poor survival of serous cases (HR=1.69 [1.21-2.36], P=0.002), see Table 4.3. This association was supported by the tSNP results – the rare allele of *KRAS* rs10842513, which was associated with poor survival of patients with serous ovarian cancer, was in the second position of the *KRAS* haplotype block 2.

The other 2 haplotypes of *KRAS* block 2, h001100 and h000000, were also associated with poor survival, however, of the mucinous histological subtype in these instances, (see Table 4.3). These associations were concordant with the univariate analysis results of *KRAS* rs4623993 (third position of the *KRAS* haplotype block 2), which was associated with survival from mucinous disease.

### 4.2.3:Univariate survival analysis results of *PIK3CA*

An association was found between a tSNP and haplotype of *PIK3CA* when the effects of the gene on survival were evaluated. The rare allele of *PIK3CA* rs7651265 was associated with poor survival of clear cell ovarian cancer cases – HR=2.25 (1.06-4.79), P=0.035, see Table 4.4. The h11000000 haplotype of *PIK3CA* was also associated with poor survival from the endometrioid subtype, HR=2.19 (1.1-4.37), P=0.026 (Table 4.5).

Table 4.4: Univariate Cox regression results of *PIK3CA* rs7651265 (by histology)

| Gene | tSNP | MAF | Histology | Cases | Univariate | |
|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value |
| *PIK3CA* | rs7651265 | 0.1 | All | 1794 | 1.07 (0.9-1.26) | 0.449 |
| | | | Serous | 828 | 1.05 (0.85-1.29) | 0.678 |
| | | | Endometrioid | 267 | 0.97 (0.56-1.67) | 0.913 |
| | | | Mucinous | 189 | 1.66 (0.79-3.46) | 0.179 |
| | | | **Clear cell** | 135 | **2.25 (1.06-4.79)** | **0.035** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; Emboldened histologies are statistically associated with survival; emboldened HR are statistically significant.

Table 4.5: Univariate Cox regression results of a *PIK3CA* haplotype (by histology)

| Gene | Haplotype† | Freq (%) | Histology | Univariate | |
|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value |
| *PIK3CA* | h11000000 | 4.9 | All | 1.1 (0.86-1.42) | 0.444 |
| | | | Serous | 1.06 (0.78-1.45) | 0.701 |
| | | | **Endometrioid** | **2.19 (1.1-4.37)** | **0.026** |
| | | | Mucinous | 0.76 (0.24-2.46) | 0.651 |
| | | | Clear cell | 0.38 (0.04-3.48) | 0.394 |

†: '0'= common allele and '1'= rare allele; HR - Hazard ratio; CI - confidence interval; Emboldened HR are statistically significant or the CI does not cross 1; Emboldened histologies are statistically significant; SNP order in haplotypes is 5' to 3' of *PIK3CA*: rs2865084, rs7621329, rs1517586, rs2699905, rs7641889, rs7651265, rs7640662, rs2677760.

### 4.2.3: The influence of clinical prognostic factors on survival

Clinical factors such as age at diagnosis, tumour histological subtype, grade and stage are known to influence survival from ovarian cancer. Therefore, these factors are used clinically to predict a patient's chances of survival. It is possible that these prognostic factors confound the results from the univariate analyses, either by masking statistically significant associations, or creating false positive associations.

Cox regression survival analysis was used to ascertain the effects of the prognostic factors on the samples within the dataset. As expected, statistically significant associations were found between survival from ovarian cancer and all the prognostic factors (age at diagnosis, tumour histological subtype, grade and stage). Cox regression survival modelling showed that survival from ovarian cancer decreased with increasing age; those between aged between 50 and 59 years had a 1.67 fold increase in all-cause mortality (HR=1.67 (1.01-2.77), P=0.047 compared with those aged less than 40 years. Individuals in the greater than 60 years age group had the worse survival, compared with the under 40 year olds (see Table 4.6).
Figure 4.2 shows the Kaplan-Meier survival curves for the prognostic factors. Individuals with mucinous, endometrioid or clear cell histological subtypes of invasive epithelial ovarian cancer survived for longer than those with the serous subtype (see Table 4.6).

Tumour grades 2 (moderately differentiated tumour) and 3 (poorly differentiated, more malignant tumour) were also significantly associated with poor survival (HR=1.47 (1.11-1.96), P=0.008; HR=1.6 (1.21-2.11), P=0.001, respectively). Advanced stage tumours, which comprised of tumours that have spread to lymph nodes or metastasised to distant locations, had the strongest effect on survival from ovarian cancer (HR=4.08 (3.15-5.29), P=$1.57 \times 10^{-26}$) when compared with localised, early stage tumours. The Cox regression survival analysis results for the clinical factors are summarised in Table 4.6.

The samples analysed included both incident cases (patients recruited into their respective studies before diagnosis of the ovarian cancer) and prevalent cases

(sufferers recruited after diagnosis of disease). There was potential survival bias between incident and prevalent cases because prevalent cases are likely to have received treatment before recruitment into the studies, and individuals with poor chances of survival would have died before recruitment. It was thus expected that incident cases would have poor survival compared with prevalent cases. Although the Kaplan-Meier curves showed that incident cases had a slightly higher mortality rate compared with prevalent cases, the difference in mortality was not statistically significant (HR=1.04 [0.66-1.63], P=0.871).

Table 4.6: Results of univariate Cox regression survival analysis of clinical prognostic factors (oncogene dataset)

| **Prognostic factor** | **No. cases** | **HR (95% CI)** | **P-value** | **5-year survival rate** | **10-year survival rate** |
|---|---|---|---|---|---|
| **Histological subtype** | | | | | |
| Serous | 735 (47%) | 1 | | 40% | 30% |
| **Endometrioid** | 249 (16%) | **0.4 (0.3-0.52)** | **$2.56x10^{-11}$** | 70% | 65% |
| **Mucinous** | 170 (11%) | **0.4 (0.29-0.56)** | **$9.29x10^{-8}$** | 60% | 63% |
| **Clear cell** | 126 (6%) | **0.4 (0.28-0.59)** | **$1.75x10^{-6}$** | 65% | 60% |
| **Age at diagnosis (years)** | | | | | |
| < 40 | 100 (6%) | 1 | | 75% | 63% |
| 40-49 | 306 (19%) | 1.49 (0.88-2.52) | 0.137 | 60% | 48% |
| **50-59** | 586 (37%) | **1.67 (1.01-2.77)** | **0.047** | 50% | 43% |
| **≥ 60** | 580 (37%) | **2.26 (1.37-3.73)** | **0.002** | 38% | 30% |
| **Tumour grade*** | | | | | |
| 1 | 260 (17%) | 1 | | 68% | 60% |
| **2** | 398 (25%) | **1.47 (1.11-1.96)** | **0.008** | 52% | 40% |
| **3** | 540 (34%) | **1.6 (1.21-2.11)** | **0.001** | 38% | 30% |
| **Tumour stage** | | | | | |
| Localised | 531 (34%) | 1 | | 80% | 68% |
| **Advanced**[§] | 736 (47%) | **4.08 (3.15-5.29)** | **$1.57x10^{-26}$** | 28% | 18% |

N=1,572; HR – hazard ratio; CI – confidence interval; * Tumour grades (1= well differentiated – low grade; 2= moderately differentiated; 3= poorly differentiated (high grade). § spread to regional lymph nodes or distant metastases; emboldened prognostic factors are significantly associated with survival from ovarian cancer.

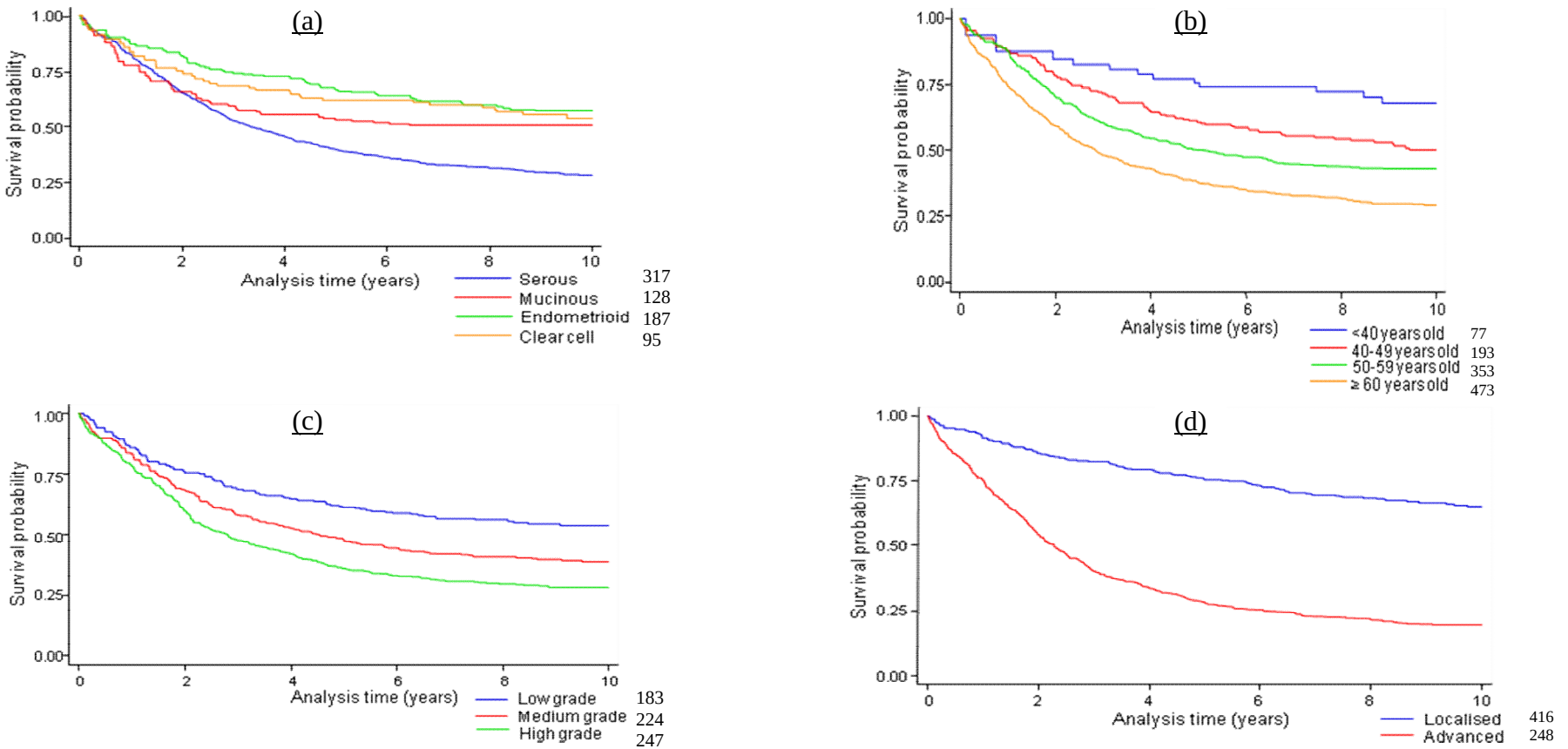



Figure 4.2: Kaplan-Meier survival curves

by (a) histological subtype; (b) age-group at diagnosis; (c) tumour grade; (d) tumour stage; numbers following the keys are individuals still at risk after 10 years.

### 4.2.4: Multivariate survival analysis results of oncogene variants

The results of all variants were adjusted for the prognostic factors which were significantly associated with survival from ovarian cancer (age at diagnosis ≥ 50 years; mucinous, endometrioid and clear cell histological subtypes; tumour grades 2 and 3; and advanced stage disease). The statistical modelling of the survival data, with adjustments for confounding (clinical prognostic) factors is known as multivariate analysis. The results of the univariate and multivariate survival analysis for the common tagging polymorphisms and haplotypes of the candidate oncogenes are tabulated in Appendices VII-A to VII-J.

There was no evidence of association between the common genetic variants or haplotype of *ERBB2* or *NMI*. The univariate and multivariate Cox regression analysis results for *ERBB2* and *NMI* are shown in Appendices VII-C to VII-D, and VII-G to VII-H, respectively. The associations found in the univariate survival analysis of the *PIK3CA* variants or haplotypes were no longer statistically significant after adjustments for prognostic factors. The survival results of the tSNPs and haplotypes can be found in Appendix VII-I and VII-J, respectively.

### 4.2.5: Multivariate survival analysis results of *BRAF* oncogene

When multivariate Cox regression survival analysis was used to assess the effects of the common variants of candidate oncogenes on survival from epithelial ovarian cancer, the association between the rare allele of *BRAF* rs6944385 and all-cause survival of all subtypes combined became stronger – adjusted (for prognostic factors) per-rare allele HR=1.25 (1.05-1.5), P=0.013. However, the association of

the variant with the clear cell histological subtype was no longer significant (see Table 4.7). The univariate and multivariate Cox regression survival results for all the common variants and haplotypes of the *BRAF* oncogene are shown in Appendix VII-A and VII-B, respectively.

Additional associations, which were not identified with the univariate analyses, were found between all-cause survival and variants of *BRAF*. The rare allele of *BRAF* rs1267622 was associated with poor survival of ovarian cancer patients (adjusted HR=1.19 (1.03-1.38), P=0.02). The rare allele of *BRAF* rs13241719 and the AA haplotype of rs1267622:rs6944385 were associated with better survival of all histological subtypes combined in the multivariate analyses (adjusted HR=0.79 (0.67-0.93), P=0.004; and adjusted HR=0.84 (0.72-0.97), P=0.018, respectively). The rs13241719 variant was also associated with the serous histological subtype when the analysis was restricted to the individual subtypes (see Table 4.7).

The *BRAF* variants rs1267622, rs13241719 and rs6944385 are correlated. The $r^2$ between *BRAF* rs1267622 and rs13241719 is 0.116; rs1267622 and rs6944385 - $r^2$=0.339; and rs13241719 and rs6944385 - $r^2$=0.039. A likelihood ratio test was performed with and without terms for the three *BRAF* variants, adjusted for the prognostic factors. This test was used to evaluate whether a model with all three variants was statistically significant, compared with a model without the variants.

Table 4.7: Univariate and multivariate Cox regression results of *BRAF* tSNPs, by histology

| Gene | tSNP | MAF | Histology | Cases | Univariate* | | Multivariate*§ | | Diff HR (%)‡ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *BRAF* | rs1267622 | 0.23 | **All** | 1751 | 1.12 (0.99-1.27) | 0.077 | **1.19 (1.03-1.38)** | **0.02** | 6 |
| | | | Serous | 831 | 1.03 (0.87-1.22) | 0.727 | 1.2 (1-1.4) | 0.134 | **17** |
| | | | Endometrioid | 268 | 1.07 (0.72-1.59) | 0.733 | 1.1 (0.71-1.71) | 0.655 | 3 |
| | | | Mucinous | 187 | 1.08 (0.64-1.82) | 0.764 | 0.89 (0.54-1.49) | 0.663 | **18** |
| | | | Clear cell | 123 | 1.27 (0.7-2.3) | 0.429 | 1 (0.51-1.98) | 0.997 | **21** |
| | rs13241719 | 0.31 | **All** | 1602 | 0.97 (0.85-1.1) | 0.606 | **0.79 (0.67-0.93)** | **0.004** | **19** |
| | | | **Serous** | 733 | 0.94 (0.79-1.12) | 0.507 | **0.8 (0.6-0.9)** | **0.006** | **15** |
| | | | Endometrioid | 246 | 0.77 (0.49-1.21) | 0.258 | 0.79 (0.48-1.29) | 0.339 | 3 |
| | | | Mucinous | 176 | 0.99 (0.58-1.69) | 0.98 | 1.05 (0.63-1.74) | 0.852 | 6 |
| | | | Clear cell | 135 | 1.12 (0.59-2.15) | 0.723 | 1.4 (0.64-3.06) | 0.404 | **25** |
| | rs6944385 | 0.14 | **All** | 1758 | **1.19 (1.03-1.38)** | **0.021** | **1.25 (1.05-1.5)** | **0.013** | 5 |
| | | | Serous | 840 | 0.97 (0.79-1.2) | 0.804 | 1.1 (0.9-1.3) | 0.516 | **13** |
| | | | Endometrioid | 268 | 1.31 (0.84-2.07) | 0.235 | 1.43 (0.87-2.35) | 0.156 | 9 |
| | | | Mucinous | 187 | 0.83 (0.4-1.73) | 0.614 | 0.76 (0.36-1.62) | 0.477 | 8 |
| | | | Clear cell | 124 | **2.22 (1.18-4.17)** | **0.014** | 1.93 (0.95-3.92) | 0.07 | **13** |
| | rs1267622, rs6944385; AA | 76† | **All** | 1786 | 0.89 (0.79-1.01) | 0.076 | **0.84 (0.72-0.97)** | **0.018** | 6 |
| | | | Serous | 724 | 0.97 (0.82-1.14) | 0.708 | **0.9 (0.7-1)** | 0.115 | 7 |
| | | | Endometrioid | 246 | 0.9 (0.61-1.34) | 0.611 | 0.87 (0.57-1.34) | 0.532 | 3 |
| | | | Mucinous | 169 | 0.94 (0.56-1.58) | 0.82 | 1.18 (0.7-1.98) | 0.528 | **26** |
| | | | Clear cell | 126 | 0.79 (0.44-1.43) | 0.434 | 1 (0.51-1.98) | 0.999 | **27** |

* stratified by study; HR - Hazard ratio; CI - confidence interval; MAF- minor allele frequency; † Haplotype frequency; § adjusted for prognostic factors (histology [where appropriate], age, stage and grade);‡: difference in HR after multivariate analysis – values ≥ 10: prognostic factors were confounding. Emboldened tSNP names are statistically associated with survival after adjustments; emboldened HR are statistically significant or the CI does not cross 1.

Table 4.8: Univariate and multivariate Cox regression results of *BRAF* haplotypes (P<0.05)

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *BRAF* | h10010000 | 18.8 | **All** | 0.95 (0.81-1.1) | 0.493 | **0.8 (0.66-0.95)** | **0.014** | **16** |
| | | | Serous | 0.95 (0.79-1.16) | 0.633 | 1.1 (0.9-1.4) | 0.415 | **16** |
| | | | Endometrioid | 0.88 (0.54-1.42) | 0.591 | 0.88 (0.53-1.48) | 0.629 | 0 |
| | | | Mucinous | 0.63 (0.31-1.28) | 0.204 | 0.62 (0.32-1.23) | 0.173 | 2 |
| | | | Clear cell | 1.38 (0.61-3.08) | 0.438 | 2.38 (0.92-6.15) | 0.074 | **72** |
| | h10010010 | 12.2 | All | 1.1 (0.94-1.3) | 0.238 | 1 (0.83-1.21) | 0.96 | 9 |
| | | | **Serous** | 1.12 (0.91-1.37) | 0.296 | **0.8 (0.6-1)** | **0.037** | **29** |
| | | | Endometrioid | 0.74 (0.36-1.5) | 0.401 | 0.74 (0.36-1.54) | 0.425 | 0 |
| | | | Mucinous | 1.5 (0.81-2.8) | 0.199 | 1.91 (0.96-3.78) | 0.065 | **27** |
| | | | Clear cell | 0.79 (0.35-1.75) | 0.557 | 0.62 (0.23-1.7) | 0.351 | **22** |
| | h01100001 | 7.1 | **All** | **1.21 (1-1.46)** | 0.055 | **1.43 (1.14-1.8)** | **0.002** | **18** |
| | | | Serous | 1.1 (0.85-1.42) | 0.483 | 0.9 (0.7-1.3) | 0.636 | **18** |
| | | | **Endometrioid** | 1.3 (0.71-2.4) | 0.393 | **2.04 (1.05-3.99)** | **0.036** | **57** |
| | | | Mucinous | 0.81 (0.33-1.99) | 0.652 | 0.9 (0.38-2.1) | 0.804 | **11** |
| | | | Clear cell | 1.86 (0.84-4.13) | 0.127 | 1.92 (0.74-4.96) | 0.179 | 3 |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; HR: Hazard ratio; CI: confidence interval; Emboldened HR are statistically significant or the CI does not cross 1; Emboldened haplotypes are statistically significant after adjustments for clinical factors; SNP order in haplotypes is 5' to 3' of the genes – *BRAF*: rs10487888, rs1733832, rs1267622, rs13241719, rs17695623, rs17161747, rs17623382, rs6944385;

The likelihood ratio test including all three tSNPs, adjusted for prognostic factors, was statistically significant (P=0.0147), however, when forward stepwise regression procedure was used, only rs13241719 was retained in the final model (P=0.009). The forward stepwise regression procedure involved the modelling of the variants (rs1267622, rs13241719 and rs6944385), one-by-one, and retaining the tSNP which was statistically significant (rs13241719 in this instance). The prognostic factors were included in the model because the association between rs1267622 and rs13241719 were found with the multivariate survival analysis.

*BRAF* rs1267622 tags rs4726020 with $r^2$ =1. According to Pupasuite, both SNPs are intronic and the dbSNP database showed rs1267622 is in intron 3 of the oncogene, and rs4726020 is in intron 1. rs13241719 (intron 2) is not known to tag any other SNP within *BRAF*. rs6944385 (intron 1) tags rs9648716 (intron 1) with $r^2$=1. None of these SNPs are predicted to have functions that could explain their association with survival from ovarian cancer.

The correlation between the h01100001 haplotype of *BRAF* and all-cause mortality of the combined subtypes of ovarian cancer cases became statistically significant after adjustments for the prognostic factors, adjusted HR=1.43 (1.14-1.8), P=0.002, (see Table 4.8). This haplotype was also associated with poor survival, when analysis was restricted to the endometrioid histological subtype. The haplotype was associated with a 2.04-fold increase in hazard of the endometrioid subtype – Table 4.8. These results were confounded by the clinical prognostic factors, by at least 18%.

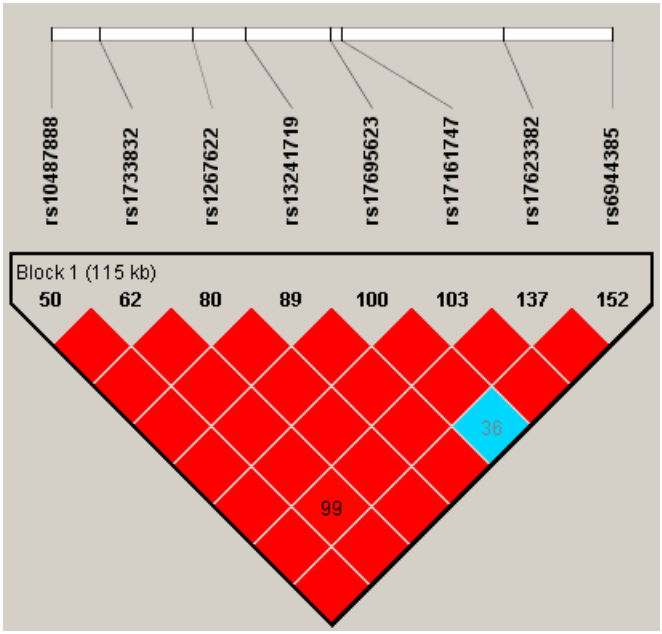



Figure 4.3: tSNPs in *BRAF* haplotype block

Colour scheme: standard (D'/LOD) – white (D'<1, LOD<2), shades of pink/red (D'<1, LOD≥2), blue (D'=1, LOD<2) and bright red (D'=1, LOD≥2), numbers shown in squares (LD values) are based on D'.

Two other, previously undetected, associations were found between haplotypes of *BRAF* and survival from ovarian cancer after adjustments for prognostic factors. The h10010000 haplotype was associated with improved survival of all histological subtypes (adjusted HR=0.8 [0.66-0.95], P=0.014; Table 4.8). There was also evidence suggesting that this haplotype was associated with improved survival of serous cases, when the Cox regression survival analysis was restricted to the histological subtype – Table 4.8. This association was supported by the multivariate result of *BRAF* rs13241719, which was in the fourth position of the haplotype (Figure 4.3).

### 4.2.6: Multivariate survival analysis results of *KRAS* oncogene

An association was found between the rs10842513 variant of *KRAS* and poor survival of ovarian cancer, after adjustments for the clinical prognostic factors (Table 4.9). A statistically significant association was also found with this variant and the serous histological subtype in the univariate analysis (see Table 4.9). Although the association with the serous subtype was no longer statistically significant, the 95% confidence interval did not cross 1.

The difference between the hazard ratios of the univariate and multivariate analyses of the serous subtype for the rs10842513 variant were not significant (<10%), therefore the hazard ratio was not confounded by the prognostic factors. The rs10842513 SNP, which is located in intron 2 of *KRAS*, is not known to tag another SNP within the oncogene. Although the SNP is currently not predicted to have a "function", it is conserved in mice.

The h001100 haplotype of *KRAS* block 2 remained significantly associated with reduced survival of sufferers with the mucinous subtype (adjusted HR=2.74 [1.27-5.9], P=0.01). The h100010 haplotype of block 2 was also associated with poor survival of mucinous disease – see Table 4.10. This association was found after adjustments for the prognostic factors had been made.

Table 4.9: Univariate and multivariate survival results of *KRAS* rs10842513

| Gene | tSNP | MAF | Histology | Cases | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *KRAS* | rs10842513 | 0.09 | **All** | 1770 | 1.18 (0.98-1.42) | 0.08 | **1.25 (1.01-1.55)** | **0.039** | 6 |
| | | | Serous | 846 | **1.38 (1.09-1.75)** | **0.008** | **1.3 (1-1.6)** | 0.091 | 6 |
| | | | Endometrioid | 271 | 1.19 (0.67-2.1) | 0.552 | 1.47 (0.79-2.74) | 0.227 | **24** |
| | | | Mucinous | 187 | 0.7 (0.29-1.69) | 0.432 | 0.74 (0.29-1.87) | 0.521 | 6 |
| | | | Clear cell | 132 | **2.02 (1-4.1)** | 0.052 | 1.71 (0.81-3.58) | 0.156 | **15** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; § adjusted for prognostic factors (histology, age, stage and grade, where appropriate); emboldened HR are statistically significant or the CI does not cross 1.

Table 4.10: Univariate and multivariate survival results of *KRAS* haplotype block 2 (P<0.05)

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *KRAS* haplotype block 2 | h100010 | 10.7 | All | 0.96 (0.83-1.12) | 0.625 | 0.95 (0.81-1.12) | 0.556 | 1 |
| | | | Serous | 0.9 (0.74-1.08) | 0.255 | 0.9 (0.6-1.3) | 0.479 | 0 |
| | | | Endometrioid | 0.86 (0.51-1.45) | 0.571 | 0.82 (0.47-1.43) | 0.481 | 5 |
| | | | **Mucinous** | 1.32 (0.79-2.22) | 0.288 | **1.79 (1.03-3.13)** | **0.04** | **36** |
| | | | Clear cell | 0.95 (0.47-1.91) | 0.881 | 0.87 (0.39-1.93) | 0.723 | 8 |
| | h001100 | 3.7 | All | 1.02 (0.79-1.31) | 0.902 | 1.05 (0.79-1.41) | 0.722 | 3 |
| | | | Serous | 0.87 (0.62-1.21) | 0.411 | 1.1 (0.8-1.5) | 0.631 | **26** |
| | | | Endometrioid | 0.99 (0.44-2.18) | 0.971 | 1.21 (0.52-2.82) | 0.652 | **22** |
| | | | **Mucinous** | **3.24 (1.55-6.74)** | **0.002** | **2.74 (1.27-5.9)** | **0.01** | **15** |
| | | | Clear cell | 2.42 (0.6-9.66) | 0.212 | 3.42 (0.65-18) | 0.146 | **41** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; HR: Hazard ratio; CI: confidence interval; Emboldened HR are statistically significant or the CI does not cross 1; SNP order in haplotypes is 5' to 3' of the genes –*KRAS* haplotype block 2: rs12579073, rs10842513, rs4623993, rs6487464, rs10842514, rs11047917.

## 4.3: Survival analyses of variants and haplotypes of functional candidates

Sixty-three tSNPs from nine differentially expressed genes (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) selected from the microcell-mediated transfer of chromosome 18 (MMCT-18) into two ovarian cancer cell-lines were genotyped with the MALOVA (446 cases), SEARCH (847 cases) and UKOPS (401 cases) population-based sample sets. There were 617 deaths in 5,885 person-years at risk. Cox regression survival analysis was used to ascertain the effect of the tSNPs and haplotypes of the functional candidates on the survival of ovarian cancer patients.

### 4.3.1: Association between clinical prognostic factors and survival for "functional" candidate genes

The results from the survival analysis of tSNPs and haplotypes of candidate oncogenes demonstrated that the adjustment for clinical prognostic factors was critical in determining associations which were not confounded by prognostic factors. Therefore, Cox regression survival analysis was used to assess the effects of the prognostic factors (age at diagnosis, tumour histological subtype, grade and stage) on survival from ovarian cancer of individuals from the MALOVA, SEARCH and UKOPS population-based studies used in this analysis. The results of the effect of the prognostic factors on survival from ovarian cancer are summarised in Table 4.11.

Table 4.11: Cox regression survival analysis results of clinical prognostic factors (MMCT-18 dataset)

| Prognostic factor | No. cases (%) | HR (95% CI) | P-value |
|---|---|---|---|
| **Histological subtype** | | | |
| Serous | 796 (47%) | Reference | |
| Mucinous | 185 (11%) | 1.15 (0.78-1.69) | 0.49 |
| Endometrioid | 262 (16%) | 0.79 (0.57-1.09) | 0.148 |
| Clear cell | 153 (9%) | 0.75 (0.44-1.26) | 0.278 |
| **Age at diagnosis (years)** | | | |
| < 40 | 72 (4%) | Reference | |
| 40-49 | 270 (16%) | 1.41 (0.66-3) | 0.37 |
| 50-59 | 636 (38%) | 1.79 (0.87-3.68) | 0.115 |
| **≥ 60** | 716 (42%) | **2.44 (1.19-4.97)** | **0.014** |
| **Tumour grade*** | | | |
| 1 | 250 (15%) | Reference | |
| **2** | 400 (24%) | **1.36 (1.01-1.82)** | **0.041** |
| **3** | 518 (31%) | **1.38 (1.04-1.85)** | **0.028** |
| **Tumour stage§** | | | |
| Localised | 492 (29%) | Reference | |
| **Advanced§** | 744 (44%) | **3.99 (3.01-5.02)** | **$4.04 \times 10^{-22}$** |

N=1,694; * 1= well differentiated – low grade, 2= moderately differentiated - medium grade, 3= poorly differentiated (high grade); § spread to regional lymph nodes or distant metastases; emboldened prognostic factors are significantly associated with survival from ovarian cancer.

Contrary to the findings with the samples analysed in the oncogene study, there were no statistically significant associations between survival from ovarian cancer and tumour histological subtypes (P>0.05). This may have been as a result of the absence of the GEOCS and additional samples in the SEARCH and UKOPS sample populations. Individuals in the > 60 years old age group had a significantly increased mortality rate compared with those in the <40 years age group (HR=2.44 [1.19-4.97], P=0.014). There were statistically significant differences in the mortality of individuals with low grade tumours and those with intermediate and high grade tumours (HR=1.36 [1.01-1.82], P=0.041; HR=1.38 [1.04-1.85], P=0.028, respectively). In concordance with the samples in the oncogene study, advanced stage disease had the biggest effect on survival from ovarian cancer, with an

approximately 4-fold increase in mortality compared with localised, early stage disease (HR=3.99 (3.01-5.02), P=$4.04 \times 10^{-22}$).

### 4.3.2: Effect of "functional" candidate ovarian cancer genes on survival of ovarian cancer patients

Univariate and multivariate Cox regression survival analysis results for all common variants and haplotypes of this series of functional candidate genes (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) are tabulated in Appendix VIII-A to VIII-R.

There was no evidence of association between survival from ovarian cancer and the common tSNPs or haplotypes of *AKTIP*, *AXIN2* or *STAG3*. However, associations were found between survival from clear cell, and endometrioid ovarian cancers and the combined rare haplotypes of *AKTIP* and *STAG3*, respectively (Appendix VIII-D, and VIII-R, respectively).

#### Multivariate survival analysis results of *AIFM2*

After adjustments for the prognostic factors, two variants of *AIFM2* were significantly associated with survival from histological subtypes of invasive epithelial ovarian cancer. The rare allele of *AIFM2* rs2394655 was associated with increased mortality of patients with the mucinous subtype (adjusted per-rare allele HR=3.05 [1.03-8.98], P=0.043). This association was also found with the univariate survival analysis (see Table 4.12). The rare allele of *AIFM2* rs2280201 was associated with poor survival of the endometrioid subtype, adjusted HR=2.03 (1.13-3.65), P=0.018.

Associations were also found between haplotypes of *AIFM2* and survival of ovarian cancer cases. The associations were found with both haplotype block of *AIFM2*. The h0001011 haplotype of *AIFM2* block 1 was associated with increased mortality of endometrioid patients after adjustments for prognostic factors (adjusted HR=2.76 [1.36-5.59], P=0.005). The variant of *AIFM2*, rs2280201, was in the last position of haplotype block 1. The rare allele of this variant was associated with poor survival of endometrioid cases, and thus supports the findings of the haplotype analysis. The h01011 haplotype of *AIFM2* block 2 was also associated with poor survival of endometrioid cases, adjusted HR=5.31 (2.04-13.8), P=0.001, see Table 4.13.

There was evidence suggesting that the h1111110 haplotype of *AIFM2* block 1 was associated with increased mortality of mucinous patients (adjusted HR=3.02 (1.02-8.91), P=0.045; Table 4.13). This association was supported by the single variant results. The rare allele of rs2394655, which was associated with poor survival, was the first position of the haplotype. The rs2394655 variant is not known to tag another SNP. The variant, which is conserved in mice, is located in the 3' untranslated region of the gene, and it is predicted to be an exonic splicing enhancer.

Furthermore, the h00001 haplotype of *AIFM2* block 2 was associated with poor survival of those with clear cell disease, adjusted HR=2.29 (1.23-4.28), P=0.009; see Table 4.13.

Table 4.12: Univariate and multivariate survival results of *AIFM2* tSNPs (P<0.05)

| Gene | tSNP | MAF | Histology | Cases | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *AIFM2* | rs2394655 | 0.04 | All | 1751 | 1.04 (0.78-1.39) | 0.788 | 1 (0.72-1.4) | 0.986 | 4 |
| | | | Serous | 827 | 0.97 (0.68-1.39) | 0.878 | 0.91 (0.61-1.36) | 0.652 | 6 |
| | | | Endometrioid | 269 | 0.22 (0.03-1.56) | 0.129 | 0.32 (0.04-2.35) | 0.262 | **45** |
| | | | **Mucinous** | 189 | **4.88 (1.96-12.15)** | **0.001** | **3.05 (1.03-8.98)** | **0.043** | **38** |
| | | | Clear cell | 150 | 1.08 (0.55-2.11) | 0.824 | 1.27 (0.6-2.73) | 0.532 | **18** |
| | rs2280201 | 0.12 | All | 1313 | 0.93 (0.77-1.11) | 0.392 | 0.95 (0.77-1.17) | 0.617 | 2 |
| | | | Serous | 556 | 0.81 (0.63-1.03) | 0.08 | 0.87 (0.66-1.13) | 0.296 | 7 |
| | | | **Endometrioid** | 216 | 1.44 (0.84-2.45) | 0.182 | **2.03 (1.13-3.65)** | **0.018** | **41** |
| | | | Mucinous | 146 | 0.93 (0.49-1.78) | 0.833 | 2.02 (0.96-4.24) | 0.065 | **117** |
| | | | Clear cell | 150 | 1.01 (0.71-1.45) | 0.94 | 0.86 (0.56-1.33) | 0.496 | **15** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; § adjusted for prognostic factors (histology, age, stage and grade, where appropriate); Emboldened histological subtypes - variants are statistically associated with survival after adjustments; emboldened HR are statistically significant or the CI does not cross 1.

Table 4.13: Effects of *AIFM2* haplotypes on survival from ovarian cancer (P<0.05)

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *AIFM2* haplotype block 1 | h0001011 | 7 | All | 0.90 (0.71-1.13) | 0.348 | 1 (0.77-1.3) | 0.98 | **11** |
| | | | Serous | 0.80 (0.60-1.08) | 0.149 | 0.86 (0.62-1.19) | 0.352 | 7 |
| | | | **Endometrioid** | 1.51 (0.81-2.8) | 0.191 | **2.76 (1.36-5.59)** | **0.005** | **83** |
| | | | Mucinous | 0.88 (0.37-2.12) | 0.776 | 1.98 (0.69-5.7) | 0.204 | **125** |
| | | | Clear cell | 0.87 (0.52-1.47) | 0.602 | 0.86 (0.48-1.56) | 0.626 | 1 |
| | h1111110 | 4 | All | 1.02 (0.76-1.38) | 0.879 | 0.99 (0.7-1.41) | 0.957 | 3 |
| | | | Serous | 0.91 (0.62-1.32) | 0.612 | 0.82 (0.53-1.27) | 0.379 | 10 |
| | | | Endometrioid | 0.34 (0.07-1.71) | 0.191 | 0.5 (0.1-2.58) | 0.408 | **47** |
| | | | **Mucinous** | **4.87 (1.95-12.17)** | **0.001** | **3.02 (1.02-8.91)** | **0.045** | **38** |
| | | | Clear cell | 1.15 (0.58-2.27) | 0.69 | 1.57 (0.71-3.48) | 0.27 | **37** |
| *AIFM2* haplotype block 2 | h00001 | 4 | All | 1.07 (0.79-1.45) | 0.651 | 1.21 (0.86-1.71) | 0.279 | **13** |
| | | | Serous | 1.18 (0.79-1.75) | 0.414 | 1.37 (0.89-2.11) | 0.155 | **16** |
| | | | Endometrioid | 0.27 (0.05-1.34) | 0.11 | 0.33 (0.07-1.51) | 0.153 | **22** |
| | | | Mucinous | 0.47 (0.11-1.97) | 0.304 | 1.31 (0.29-5.91) | 0.725 | **179** |
| | | | **Clear cell** | **1.77 (1-3.12)** | 0.05 | **2.29 (1.23-4.28)** | **0.009** | **29** |
| | h01011 | 2 | All | 0.93 (0.64-1.35) | 0.702 | 1.16 (0.77-1.74) | 0.476 | **25** |
| | | | Serous | 0.73 (0.45-1.17) | 0.188 | 1.04 (0.63-1.73) | 0.879 | **42** |
| | | | **Endometrioid** | **2.74 (1.07-7.04)** | **0.036** | **5.31 (2.04-13.8)** | **0.001** | **94** |
| | | | Mucinous | - | - | - | - | - |
| | | | Clear cell | 1 (0.43-2.36) | 0.992 | 0.62 (0.22-1.73) | 0.36 | **38** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', haplotype block 1: rs2394655, rs7908957, rs1053495, rs2894111, rs2394656, rs6480440, rs2280201; haplotype block 2: rs10999147, rs3750772, rs4295944, rs2394644, rs10999152.

**Multivariate survival analysis results of *CASP5***

An association between the rare allele of *CASP5* rs2282657 and reduced mortality of clear cell patients became stronger after adjustments in the multivariate analysis, adjusted HR=0.68 (0.48-0.96), P=0.029, see Table 4.14. The *CASP5* rs2282657 variant, an intronic SNP, tags 2 other intronic SNPs, all of which are conserved in mice.

The h000011 haplotype of *CASP5* block 2 remained associated with reduced mortality of clear cell patients after adjustments for prognostic factors, adjusted HR=0.57 (0.34-0.97), P=0.037; see Table 4.15. This association is concordant with the presence of the rare allele of *CASP5* rs2282657, in the fifth position of the haplotype block, which was associated with improved survival of clear cell cases.

The combined rare haplotypes of *CASP5* block 1 were also associated with poor survival of all histological subtypes ($P=8.85x10^{-5}$), and the serous and clear cell subtypes when the multivariate analysis was restricted to the subtypes (see Appendix VIII-H). Despite the strength of the association with all histological subtypes, the combined rare haplotypes have a frequency of 4%, and it is not possible to definitively ascertain the haplotype responsible for the association. However, if the causative haplotype was found, only a very small number of cases are likely to carry the haplotype.

Table 4.14: Effect of *CASP5* rs2282657 on survival from ovarian cancer

| Gene | tSNP | MAF | Histology | Cases | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *CASP5* | rs2282657 | 0.35 | All | 852 | 0.94 (0.83-1.06) | 0.329 | 0.94 (0.81-1.09) | 0.442 | 0 |
| | | | Serous | 462 | 1.10 (0.95-1.28) | 0.216 | 1.12 (0.93-1.34) | 0.247 | 2 |
| | | | Endometrioid | 128 | 0.82 (0.55-1.22) | 0.327 | 0.78 (0.5-1.2) | 0.254 | 5 |
| | | | Mucinous | 80 | 0.75 (0.47-1.19) | 0.224 | 0.92 (0.57-1.48) | 0.735 | **23** |
| | | | **Clear cell** | 73 | **0.76 (0.57-1)** | **0.049** | **0.68 (0.48-0.96)** | **0.029** | **11** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

Table 4.15: Univaraite and multivariate survival results of *CASP5* haplotype (P<0.05)

| Haplotype block | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *CASP5* (haplotype block 2) | h000011 | 13 | All | 1.02 (0.87-1.20) | 0.821 | 1.01 (0.83-1.23) | 0.916 | 1 |
| | | | Serous | 1.11 (0.91-1.35) | 0.298 | 1.08 (0.86-1.36) | 0.493 | 3 |
| | | | Endometrioid | 1.05 (0.63-1.77) | 0.847 | 1.17 (0.66-2.07) | 0.583 | **11** |
| | | | Mucinous | 1.69 (0.92-3.1) | 0.089 | 1.64 (0.85-3.15) | 0.139 | 3 |
| | | | **Clear cell** | **0.62 (0.4-0.97)** | **0.034** | **0.57 (0.34-0.97)** | **0.037** | 8 |

Freq – frequency; HR – hazard ratio; CI – confidence interval; †: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *CASP5* haplotype block 2: rs17446518, rs9651713, rs3181175, rs3181174, rs2282657, rs507879.

**Multivariate survival analysis results of *RGC32***

There was also evidence of an association between poor survival from the serous histological subtype and the rare allele of *RGC32* rs3783194 (adjusted HR=1.44 (1.12-1.86), P=0.005); Table 4.16. This SNP is located in intron 2 of the gene, and to date, it is not known if it tags another variant. The rare allele of another genetic variant of *RGC32*, rs995845, was also associated with poor survival of endometrioid patients after using the multivariate Cox regression survival analysis (adjusted HR=1.8 (1.03-3.14), P=0.039); see Appendix VIII-M.

Despite the associations found between the common genetic variants of *RGC32* and survival from ovarian cancer, no statistically significant associations were found between common haplotypes of the gene, and survival from the disease. The results of the haplotype-specific effects are given in Appendix VIII-N.

**Multivariate survival analysis results of *FILIP1L***

When multivariate Cox regression survival analysis was used to determine the effects of common tSNPs from *FILIP1L* on survival of ovarian cancer patients, statistically significant associations were found with two variants. The rare allele of *FILIP1L* rs3921767 was associated with poor survival of ovarian cancer patients regardless of the histology of the tumour, adjusted (for prognostic factors) per-rare allele HR=1.39 (1.07-1.81), P=0.014 - Table 4.17.

Table 4.16: Univariate and multivariate survival results of *RGC32* tSNPs (P<0.05)

| Gene | tSNP | MAF | Histology | Cases | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RGC32* | rs3783194 | 0.11 | All | 1690 | 0.95 (0.79-1.14) | 0.567 | 1.11 (0.9-1.36) | 0.342 | **17** |
| | | | **Serous** | 788 | 1.11 (0.88-1.41) | 0.359 | **1.44 (1.12-1.86)** | **0.005** | **30** |
| | | | Endometrioid | 264 | 1.09 (0.65-1.84) | 0.742 | 1.12 (0.61-2.05) | 0.713 | 3 |
| | | | Mucinous | 184 | 0.63 (0.29-1.35) | 0.232 | 0.64 (0.24-1.73) | 0.38 | 2 |
| | | | Clear cell | 155 | 0.81 (0.55-1.19) | 0.282 | 0.76 (0.5-1.16) | 0.202 | 6 |
| | rs995845 | 0.2 | All | 1274 | 0.96 (0.85-1.08) | 0.488 | 1.13 (0.93-1.38) | 0.218 | **18** |
| | | | Serous | 595 | 1.03 (0.88-1.21) | 0.682 | 1.25 (0.97-1.61) | 0.082 | **21** |
| | | | **Endometrioid** | 193 | 1.55 (0.93-2.6) | 0.093 | **1.8 (1.03-3.14)** | **0.039** | **16** |
| | | | Mucinous | 146 | 1.09 (0.58-2.03) | 0.797 | 0.75 (0.36-1.54) | 0.43 | **31** |
| | | | Clear cell | 112 | 0.94 (0.66-1.34) | 0.716 | 1.04 (0.69-1.56) | 0.851 | **11** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; § adjusted for prognostic factors (histology, age, stage and grade, where appropriate); Emboldened histological subtypes - variants are statistically associated with survival after adjustments; emboldened HR are statistically significant or the CI does not cross 1.

Conversely, the rare allele of *FILIP1L* rs9864437 was associated with better survival of mucinous cases alone (adjusted HR=0.46 (0.23-0.91), P=0.027), see Table 4.17. An additional association was found between the rare allele of another tSNP of *FILIP1L*, rs793446, and reduced mortality of mucinous cases, adjusted HR=0.57 (0.33-0.99), P=0.046. See Appendix VIII-I and VIII-J for the univariate and multivariate Cox regression analysis results for all common variants and haplotypes, respectively, of *FILIP1L*.

Analysis of the effects of the haplotypes of *FILIP1L* on survival from ovarian cancer also showed statistically significant associations. A total of 4 haplotypes of *FILIP1L* were associated with survival from ovarian cancer, 2 from haplotype block 1, and the other 2 from block 2. The h00110 haplotype of *FILIP1L* block 1 was associated with reduced mortality, of all histological subtypes, after adjustments for prognostic factors, adjusted HR=1.36 (1.04-1.77), P=0.024 (Table 4.18). The association with h00110 haplotype of *FILIP1L* block 1 was supported by the effect of the rare allele of the rs3921767 variant, which was in the fourth position of the haplotype.

The remaining 3 *FILIP1L* haplotypes were all associated with survival of patients with the mucinous subtype. One of these haplotypes was from block 1 of the gene, and the other 2 were from haplotype block 2. The h10100 haplotype of *FILIP1L* block 1, which had a frequency of 21%, was associated with reduced mortality (adjusted HR=0.44 (0.21-0.9), P=0.024) – see Table 4.18 for the Cox regression survival analysis results for the combined and individual subtypes for this haplotype. This association was also in concordance with the single variant results. The rare

allele of rs793446, which was correlated with reduced mortality, was in the third position of *FILIP1L* haplotype block 1.

As shown in Table 4.18, the 2 haplotypes of *FILIP1L* block 2, h000 and h100, had opposing effects on survival from the mucinous subtype. The h000 haplotype was associated with poor survival of mucinous cases (adjusted HR=1.96 [1.15-3.33], P=0.013). Conversely, h100 of the same haplotype block was associated with reduced mortality (adjusted HR=0.46 [0.23-0.91], P=0.026). These associations were also supported by the single variant results (*FILIP1L* rs9864437 was in the first position of the block 2 haplotypes).

**Multivariate survival analysis results of *RBBP8***

There was evidence suggesting that the rare alleles of two tSNPs of *RBBP8*, rs4474794 and rs9304261, were associated with better survival of ovarian cancer (adjusted HR=0.86 (0.74-0.99), P=0.034; adjusted HR=0.83 (0.7-0.99), P=0.038), respectively - Table 4.19. The Kaplan-Meier survival curves for these two variants are shown in Figure 4.4.

The association between the *RBBP8* tSNPs and survival from ovarian cancer was also identified in the univariate analysis. *RBBP8* rs4474794 and rs9304261 are correlated with $r^2$=0.56. *RBBP8* rs4474794 was retained in the final model after the forward stepwise regression (P=0.035). An interaction between rs4474794 and rs9304261 was statistically significant (adjusted HR=0.95 (0.9-0.99), P=0.036), and the likelihood ratio test of this interaction was also significant (P=0.031).

Table 4.17: Univariate and multivariate survival results of *FILIP1L* tSNPs (P<0.05)

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *FILIP1L* | rs793446 | 0.41 | 1773 | All | 0.96 (0.85-1.07) | 0.461 | 1.04 (0.9-1.19) | 0.628 | 8 |
| | | | 838 | Serous | 1.02 (0.88-1.18) | 0.816 | 1.08 (0.91-1.28) | 0.399 | 6 |
| | | | 274 | Endometrioid | 1.08 (0.74-1.59) | 0.69 | 1.15 (0.77-1.71) | 0.501 | 6 |
| | | | 194 | **Mucinous** | 0.65 (0.41-1.03) | 0.065 | **0.57 (0.33-0.99)** | **0.046** | **12** |
| | | | 164 | Clear cell | 0.97 (0.76-1.25) | 0.83 | 0.99 (0.74-1.32) | 0.932 | 2 |
| | rs3921767 | 0.07 | 1773 | **All** | 1.03 (0.83-1.29) | 0.786 | **1.39 (1.07-1.81)** | **0.014** | **35** |
| | | | 840 | Serous | 0.98 (0.71-1.34) | 0.895 | 1.28 (0.89-1.84) | 0.186 | **31** |
| | | | 276 | Endometrioid | 0.99 (0.5-1.93) | 0.967 | 1.23 (0.59-2.57) | 0.576 | **24** |
| | | | 191 | Mucinous | 1.03 (0.45-2.34) | 0.949 | 1.09 (0.44-2.73) | 0.849 | 6 |
| | | | 166 | Clear cell | 1.29 (0.88-1.89) | 0.196 | 1.59 (0.99-2.58) | 0.057 | **23** |
| | rs9864437 | 0.22 | 1786 | All | 0.96 (0.84-1.09) | 0.515 | 0.93 (0.8-1.09) | 0.366 | 3 |
| | | | 843 | Serous | 1.03 (0.88-1.20) | 0.708 | 1 (0.84-1.2) | 0.964 | 3 |
| | | | 278 | Endometrioid | 1.07 (0.69-1.68) | 0.752 | 0.97 (0.6-1.57) | 0.892 | 9 |
| | | | 195 | **Mucinous** | **0.45 (0.25-0.82)** | **0.009** | **0.46 (0.23-0.91)** | **0.027** | 2 |
| | | | 165 | Clear cell | 0.89 (0.66-1.2) | 0.455 | 0.87 (0.63-1.2) | 0.396 | 2 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; § adjusted for prognostic factors (histology, age, stage and grade, where appropriate); Emboldened histological subtypes - variants are statistically associated with survival after adjustments; emboldened HR are statistically significant or the CI does not cross 1.

Table 4.18: Univariate and multivariate survival results of *FILIP1L* haplotype block 2 (P<0.05)

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *FILIP1L* (haplotype block 1) | h10100 | 21 | All | 0.96 (0.84-1.09) | 0.52 | 0.94 (0.8-1.1) | 0.424 | 2 |
| | | | Serous | 1.03 (0.89-1.21) | 0.667 | 0.94 (0.67-1.31) | 0.697 | 9 |
| | | | Endometrioid | 1.12 (0.71-1.78) | 0.618 | 1 (0.61-1.63) | 0.998 | **11** |
| | | | **Mucinous** | **0.42 (0.22-0.78)** | **0.006** | **0.44 (0.21-0.9)** | **0.024** | 5 |
| | | | Clear cell | 0.88 (0.65-1.19) | 0.419 | 0.88 (0.64-1.22) | 0.447 | 0 |
| | h00110 | 7 | **All** | 1.02 (0.81-1.27) | 0.871 | **1.36 (1.04-1.77)** | **0.024** | **33** |
| | | | Serous | 0.95 (0.69-1.30) | 0.742 | 1.22 (0.85-1.76) | 0.283 | **28** |
| | | | Endometrioid | 1 (0.5-2) | 0.996 | 1.22 (0.58-2.55) | 0.604 | **22** |
| | | | Mucinous | 1.01 (0.44-2.31) | 0.977 | 1.08 (0.43-2.69) | 0.877 | 7 |
| | | | Clear cell | 1.29 (0.88-1.89) | 0.197 | 1.61 (0.99-2.6) | 0.053 | **25** |
| *FILIP1L* (haplotype block 2) | h100 | 22 | All | 0.96 (0.84-1.09) | 0.506 | 0.94 (0.8-1.09) | 0.42 | 2 |
| | | | Serous | 1.03 (0.89-1.21) | 0.668 | 0.92 (0.77-1.1) | 0.371 | **11** |
| | | | Endometrioid | 1.08 (0.69-1.69) | 0.727 | 0.97 (0.6-1.57) | 0.902 | **10** |
| | | | Mucinous | **0.45 (0.25-0.82)** | **0.009** | **0.46 (0.23-0.91)** | **0.026** | 2 |
| | | | Clear cell | 0.88 (0.65-1.18) | 0.385 | 0.87 (0.63-1.21) | 0.407 | 1 |
| | h000 | 19 | All | 1.10 (0.94-1.28) | 0.223 | 1.11 (0.92-1.33) | 0.292 | 1 |
| | | | Serous | 1.07 (0.87-1.31) | 0.506 | 0.64 (0.26-1.55) | 0.321 | **40** |
| | | | Endometrioid | 0.87 (0.53-1.44) | 0.594 | 0.98 (0.59-1.64) | 0.946 | **13** |
| | | | **Mucinous** | 1.72 (1.09-2.72) | 0.019 | **1.96 (1.15-3.33)** | **0.013** | **14** |
| | | | Clear cell | 1.13 (0.82-1.58) | 0.456 | 1.09 (0.74-1.6) | 0.664 | 4 |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *FILIP1L* -block 1: rs796977, rs793477, rs793446, rs3921767, rs17338680. *FILIP1L* - block 2: rs9864437, rs6788750, rs12494994.

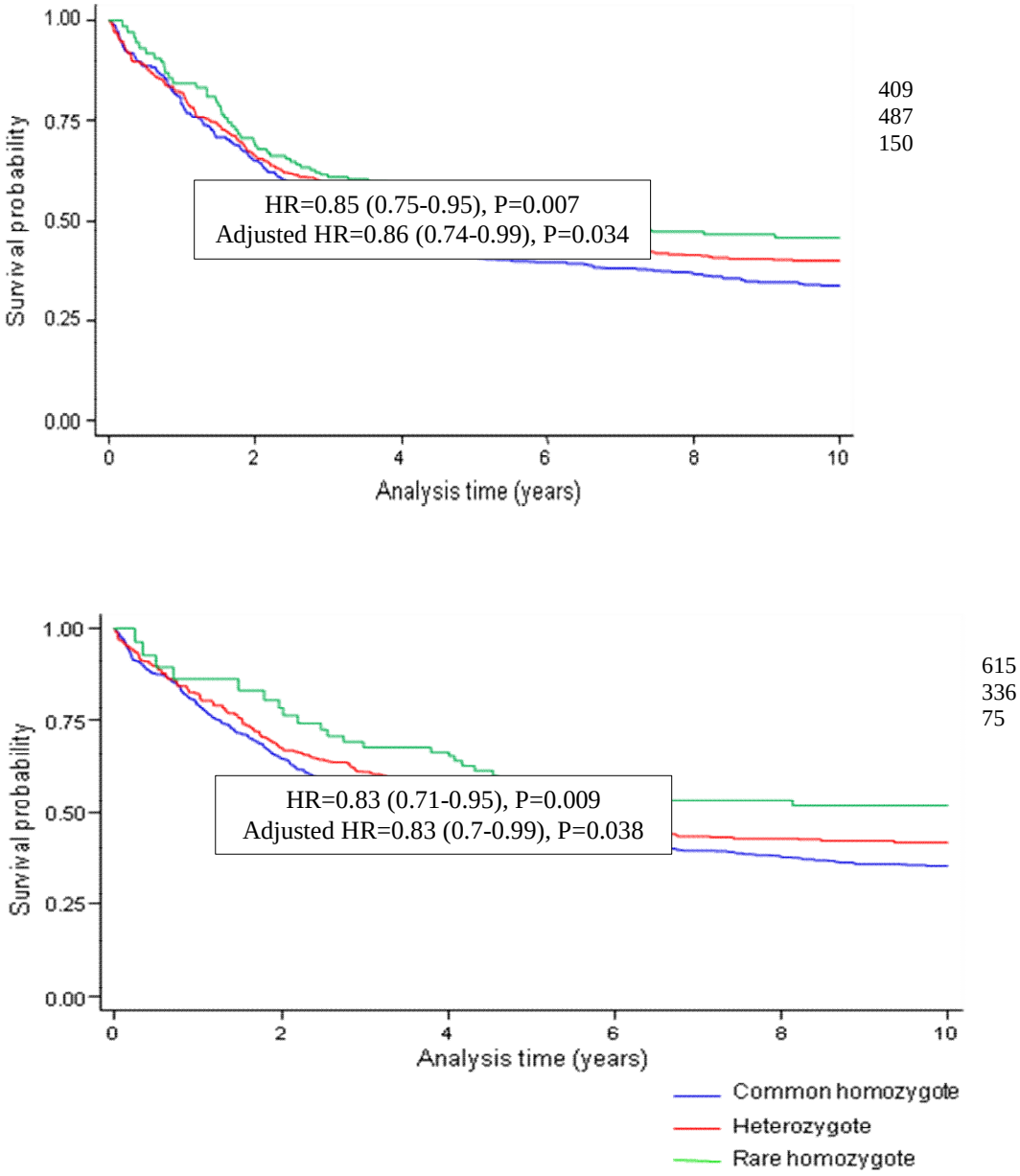



Figure 4.4: Kaplan-Meier survival estimates of *RBBP8* (a) rs4474794; (b) rs9304261 (all subtypes combined)

Numbers following the keys are individuals still at risk after 10 years.

Table 4.19 Univariate and multivariate survival results of *RBBP8* tSNPs (P<0.05)

| Gene | tSNP | MAF | Histology | Cases | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RBBP8* | rs4474794 | 0.36 | **All** | 1764 | **0.85 (0.75-0.95)** | **0.007** | **0.86 (0.74-0.99)** | **0.034** | 1 |
| | | | Serous | 829 | 0.88 (0.75-1.02) | 0.098 | 0.85 (0.71-1.01) | 0.065 | 3 |
| | | | Endometrioid | 271 | 0.8 (0.53-1.19) | 0.265 | 0.86 (0.56-1.31) | 0.479 | 7 |
| | | | Mucinous | 193 | 0.67 (0.42-1.05) | 0.079 | 0.83 (0.51-1.36) | 0.465 | **24** |
| | | | Clear cell | 165 | 0.91 (0.71-1.18) | 0.484 | 0.98 (0.73-1.32) | 0.899 | 8 |
| | rs9304261 | 0.22 | **All** | 346 | **0.83 (0.71-0.95)** | **0.009** | **0.83 (0.7-0.99)** | **0.038** | 0 |
| | | | Serous | 215 | 0.87 (0.72-1.05) | 0.143 | 0.82 (0.66-1.02) | 0.073 | 6 |
| | | | Endometrioid | 44 | 0.87 (0.56-1.35) | 0.536 | 0.99 (0.62-1.6) | 0.982 | **14** |
| | | | Mucinous | 33 | 0.61 (0.35-1.05) | 0.074 | 0.81 (0.44-1.49) | 0.497 | **33** |
| | | | Clear cell | 21 | 0.84 (0.62-1.13) | 0.242 | 0.86 (0.61-1.22) | 0.401 | 2 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; § adjusted for prognostic factors (histology, age, stage and grade, where appropriate); Emboldened histological subtypes - variants are statistically associated with survival after adjustments; emboldened HR are statistically significant or the CI does not cross 1.

Table 4.20: Univariate and multivariate survival results of *RBBP8* haplotypes (P<0.05)

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RBBP8* | h0000 | 62 | **All** | **1.16 (1.03-1.31)** | **0.015** | **1.17 (1.01-1.34)** | **0.032** | 1 |
| | | | Serous | 1.12 (0.97-1.30) | 0.134 | 0.02 (0-27735) | 0.589 | **98** |
| | | | Endometrioid | 1.26 (0.85-1.86) | 0.257 | 1.15 (0.76-1.75) | 0.511 | 9 |
| | | | Mucinous | 1.57 (1-2.47) | 0.05 | 1.22 (0.75-2) | 0.422 | **22** |
| | | | Clear cell | 1.06 (0.83-1.35) | 0.659 | 1 (0.75-1.32) | 0.98 | 6 |
| | h0011 | 23 | **All** | **0.81 (0.71-0.94)** | **0.005** | **0.82 (0.69-0.98)** | **0.029** | 1 |
| | | | Serous | 0.84 (0.70-1.02) | 0.079 | 0.99 (0.53-1.85) | 0.976 | **18** |
| | | | Endometrioid | 0.83 (0.53-1.3) | 0.422 | 0.94 (0.58-1.52) | 0.789 | **13** |
| | | | Mucinous | 0.64 (0.37-1.08) | 0.096 | 0.86 (0.48-1.53) | 0.614 | **34** |
| | | | Clear cell | 0.85 (0.64-1.13) | 0.261 | 0.87 (0.63-1.22) | 0.429 | 2 |
| | h0010 | 3 | All | **0.60 (0.39-0.93)** | **0.022** | 0.75 (0.45-1.25) | 0.275 | **25** |
| | | | **Serous** | 0.75 (0.41-1.36) | 0.347 | **1.2 (1.01-1.42)** | **0.041** | **60** |
| | | | Endometrioid | 0.35 (0.08-1.56) | 0.169 | 0.54 (0.12-2.46) | 0.429 | **54** |
| | | | Mucinous | 0.56 (0.13-2.31) | 0.419 | 1.03 (0.15-7.23) | 0.976 | **84** |
| | | | Clear cell | 0.7 (0.31-1.58) | 0.388 | 0.78 (0.28-2.18) | 0.637 | **11** |

†: '0'= common allele and '1'= rare allele; Freq – frequency; HR – hazard ratio; CI – confidence interval; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *RBBP8*: rs7239066, rs11082221, rs4474794, rs9304261.

As shown in Table 4.19, the univariate Cox regression survival analysis also detected the associations between survival and the tSNPs of *RBBP8* (rs4474794 and rs9304261). As well as being associated with better survival from ovarian cancer overall, the rare allele of rs4474794 was also significantly associated with a decreased risk of serous ovarian cancer. The rs4474794 variant and the 17 SNPs it tags were intronic, with no predicted functions.

Two haplotypes of *RBBP8*, h0000 and h0011, had opposing effects on survival of the patients. The h0000 haplotype was associated with poor survival (adjusted HR=1.17 (1.01-1.34), P=0.032). However, the h0011 haplotype of *RBBP8* was associated with improved survival from the disease, adjusted HR=0.82 (0.69-0.98), P=0.029. These associations were also identified with the univariate analyses, and were supported by the tSNP results (see Table 4.20). The variants of *RBBP8*, rs4474794 and rs9304261, which were shown to affect survival from ovarian cancer, were in the third and fourth positions, respectively, of the haplotypes.

The association between the h0000 haplotype of *RBBP8* and poor survival of all ovarian cancer patients was also found when multivariate analysis was restricted to serous only samples (adjusted for prognostic factors HR=1.2 (1.01-1.42), P=0.041). Although, the association between reduced mortality and the h0011 haplotype of *RBBP8* was attenuated after restriction to the serous subtype, the confidence interval did not cross 1 (adjusted HR=0.81 (0.65-1), P=0.054); Table 4.20.

Table 4.21: Cox regresssion results of *RUVBL1* rs4857836

| Gene | tSNP | MAF | Histology | Cases | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RUVBL1* | rs4857836 | 0.2 | **All** | 1787 | 0.98 (0.86-1.12) | 0.758 | **0.81 (0.67-0.98)** | **0.03** | **17** |
| | | | **Serous** | 845 | 1.01 (0.86-1.19) | 0.879 | **0.75 (0.59-0.95)** | **0.018** | **26** |
| | | | Endometrioid | 278 | 1.12 (0.67-1.88) | 0.66 | 0.92 (0.52-1.61) | 0.762 | **18** |
| | | | Mucinous | 195 | 0.8 (0.42-1.55) | 0.513 | 0.9 (0.42-1.95) | 0.797 | **13** |
| | | | Clear cell | 165 | 0.97 (0.68-1.38) | 0.863 | 0.84 (0.55-1.28) | 0.406 | **13** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

However, rs9304261, which is downstream of the gene, is currently known to tag 14 SNPs with $r^2$>0.8, half of these are conserved in mice. One of these SNPs, rs930910, is upstream of the gene in a transcription factor binding site. Another of the tagged SNPs, rs1902921, has a predicted triplex-forming sequence.

### 4.3.3:Multivariate survival analysis results of *RUVBL1*

When multivariate Cox regression survival analysis was used to determine the effects of common tSNPs and haplotypes from the *RUVBL1* gene, statistically significant associations were found between survival and a tSNP and 2 haplotypes and survival from ovarian cancer. The rare allele of *RUVBL1* rs4857836 was associated with reduced mortality of all cases, adjusted HR=0.81 (0.67-0.98), P=0.03. The size of the effect of the variant and the significance increased when the analysis was restricted to the serous only histological cases (see Table 4.21). The rs4857836 variant was not significantly associated with survival from the other major histological subtypes of ovarian cancer when the analysis was restricted (to the subtypes). However, similarly to the serous and all subtypes, the hazard ratios for the individual subtypes (endometrioid, mucinous and clear cell) were less than 1 (see Table 4.21).

## 4.4: Summary

Cox regression survival analysis of a model with terms for all of the prognostic factors showed that although all of the factors significantly affected survival from ovarian cancer, there were of varying effect sizes. Advanced tumour stage had the strongest effect on survival, with a four times increased hazard ratio compared with localised early stage disease.

Multivariate survival analysis with terms for prognostic factors suggested that the results of univariate analyses (with genotypes as the only explanatory variable) may have been masked by the unaccounted clinical factors. When multivariate Cox regression survival analysis was used to evaluate the affect of tSNPs and haplotypes of candidate oncogenes on the survival of 1,572 invasive epithelial ovarian cancer cases, associations were found between *BRAF* and *KRAS* and clinical outcome of invasive epithelial ovarian cancer patients. Three common tagging variants of *BRAF* (rs1267622, rs13241719 and rs6944385), and the AA haplotype of rs1267622 and rs6944385 influenced survival of all cases. When the analysis was restricted to the histological subtypes, an association was found between the survival of serous cases and the rs13241719 polymorphism. Associations were also found between haplotypes of *BRAF* (h10010000 and h01100001) and survival of all cases. The h01100001 haplotype was also associated with poor survival of endometrioid patients. Furthermore, an additional haplotype of *BRAF*, h10010010, was associated with improved survival of serous cases. The rare allele of *KRAS* rs10842513 and two haplotypes of the oncogene were associated with poor survival of all ovarian cancer cases.

Multivariate Cox regression survival analysis was also used to evaluate the affects of candidate genes selected from *in vitro* tumour suppression studies on the clinical outcome of ~1,700 ovarian cancer patients. Associations were identified between a polymorphism (rs2280201) and 2 haplotypes (h0001011 of haplotype block 1 and h01011 of haplotype block 2) of *AIFM2* and survival of endometrioid cases, P=0.018, P=0.005 and P=0.001, respectively. This gene was also associated with the

survival of those with mucinous and clear cell ovarian cancer. Moreover, the rare allele of a variant, and haplotypes of *CASP5* were associated with increased survival of patients of clear cell ovarian cancer.

*FILIP1L*, *RUVBL1* and *RBBP8* influenced the clinical of individuals with epithelial ovarian cancer. A common variant and haplotype of *FILIP1L* were associated with the survival of all cases of ovarian cancer combined. Two tSNPs and haplotypes of *FILIP1L* were associated with the survival of mucinous as well. The rare allele of *RUVBL1* rs4857836 was associated with longer survival of all hstological subtypes of the disease. This association became stronger when the analysis was restricted to the serous subtype. The rare alleles of *RBBP8* rs4474794 and rs9304261 were marginally associated with improved survival of all subtypes (P=0.034 and P=0.038, respectively). Two haplotypes of *RBBP8* were also associated with the clinical outcome of all cases. Furthermore, another hapltoyep of *RBBP8* was associated with poor survival of patients with the serous subtype.

# Chapter 5: Results
# Evaluating whole genome amplification methods and SNP multiplex genotyping platforms

## 5.1: Introduction

Aims:

(1) To evaluate the ease of use and quality of whole genome amplification methods.

(2) To evaluate the performance of non-amplified and whole amplified DNA on multiplex SNP genotyping platforms.

Objectives:

(1) To assess the call rates and concordance of non-amplified, genomic DNA and whole genome amplified DNA.

(2) To evaluate the performance of multiplex SNP genotyping platforms with genomic and whole genome amplified DNA.

(3) To examine the SNP pass rates, call rates and concordance between genomic and amplified DNA.

## 5.2: Whole genome amplification

Whole genome amplification methods are used to increase the amount of DNA available for research, and some types of diagnoses (ie preimplantation diagnostics). A whole genome amplification method which produces good quality amplified DNA, that generates good call rates and consistent, accurate genotypes could be used instead of the limited, non-amplified genomic DNA. In order to evaluate the

magnitude of DNA amplification, quality of genotype calls, and concordance between genotypes of the whole genome amplified samples and non-amplified genomic samples, 95 samples were amplified with GenomePlex™, GenomiPhi, PEP and RepliG™. PEP is a random 15-mer method of PCR-based DNA amplification. This method and GenomePlex uses the Taq polymerase enzyme for extension. GenomiPhi uses the Phi29 (Φ29) DNA polymerase, and REPLI-g uses a modified Φ29 enzyme. GenomePlex is a PCR-based method, which generates a library from the template DNA and subsequently amplifies the library. PEP amplification has been used to amplify GEOCS and SEARCH study samples for several years and all Taqman assays with samples from those two studies were performed with PEP-amplified samples. Although 100ng of starting DNA was used for GenomePlex, GenomiPhi and RepliG amplification, 20ng, the recommended amount, was used for PEP. The focus of the research was primarily on the GenomePlex, GenomiPhi and RepliGTM methods.

### 5.2.1: Comparison of the ease of use of whole genome amplification methods

In terms of ease of the amplification procedure, GenomiPhi was the simplest. The protocol involved adding the amplification mix and enzyme to the DNA and two incubation steps. The PEP protocol was also very easy to perform, with only a PCR reaction following the addition of the reaction mixture to the template DNA. The GenomePlex method was the most time consuming of the four amplification methods during the preparation steps (Table 5.1). There were three separate stages involving addition of reagents followed by incubation steps. This contrasted with only 1 stage each for GenomiPhi, PEP and REPLI-g. PEP was the only protocol without a separate denaturing DNA step. Overall, GenomePlex required the least

amount of time to perform the whole procedure. There was an insufficient amount of the Library preparation enzyme in the GenomePlex amplification kit, therefore five samples could not be successfully amplified with the method. In total 90 samples were amplified with the GenomePlex method. It is possible that some of the enzyme may have evaporated during the pipetting into the sample mixture, because a master-mix was not made with the enzyme and only 1µl was to be aliquotted into the samples.

Table 5.1: Comparison of whole genome amplification methods

| | **GenomePlex** | **GenomiPhi** | **PEP** | **REPLI-g** |
|---|---|---|---|---|
| Pattern | Thermal | Isothermal | Thermal | Isothermal |
| Ease of performance | Least easy | Easiest | Moderate | Moderate |
| Time required | 6 hours | 17 hours | 13 hours | 17 hours |
| Template DNA concentration used | 100ng | 100ng | 100ng | 100ng |
| Amplified DNA yield | 9µg | 15µg | 1.6µg | 189µg |
| Fold increase | 90 | 150 | 80 | 1894 |

The REPLI-g-amplified samples were the most difficult to pipette after amplification, before the DNA of each sample was quantified. This was because some samples were very viscous. As a consequence, a 1 in 10 dilution was needed prior to DNA quantification.

### 5.2.2: Quantities of whole genome amplified products

PICO-green was used to evaluate the amount of DNA produced by each whole genome amplification method. As shown in Table 5.2, the REPLI-g method produced the greatest increase in amplified DNA, with an average yield of approximately 190µg. This approximated to 1890-fold increase overall, in the

quantity of DNA with this method. However, there was also a very large range in the quantities of DNA produced, from 53µg to 579µg total yield with a standard deviation of 110. The standard deviations for the total yield for the other methods were less than 4. The fold increases in DNA quantities after GenomePlex and PEP amplification were relatively similar (90 and 80, respectively). The average total yield produced for GenomiPhi-amplified samples was 15µg, which was a 150-fold increase in the amount of DNA.

Table 5.2: Average fold increase in DNA quantities after WGA

| | **GenomePlex** | **GenomiPhi** | **PEP** | **REPLI-g** |
|---|---|---|---|---|
| Template DNA concentration used | 100ng | 100ng | 20ng | 100ng |
| Amplified DNA yield | 9µg | 15µg | 1.6µg | 189µg |
| Fold increase | 90 | 150 | 80 | 1894 |
| Standard deviation | 2.2 | 3.5 | 1.9 | 109.7 |

The yields of GenomiPhi and REPLI-g exceeded that which was expected for the method. This may have been due to the increased amount of input template DNA than suggested by the protocol (10ng). 100ng of template DNA was used in order to reduce the likelihood of allele dropout. Approximately double the expected amount of amplified DNA was generated by the GenomiPhi method, and more than 4 times by REPLI-g.

## 5.3: Comparison of SNP multiplex genotyping platforms

There were advantages and disadvantages for the procedures of all of the SNP multiplex genotyping platforms tested (iPLEX, OpenArray and SNPlex). This makes it difficult to say which platform was easiest to use. One of the biggest

advantages that OpenArray had over SNPlex and iPLEX was that any combination of SNPs could be included in a panel. The panels of SNPs for each iPLEX and SNPlex reaction needed to the designed to ensure reactions could not occur between the reaction products and the allele/SNP masses, so that they could be differentiated from each other. The ease of use of the SNP multiplexing platforms are summarised in Table 5.3.

### 5.3.1:OpenArray

Overall, the OpenArray platform was the most straight-forward to use after the transfer of the samples on to the TaqMan OpenArray plates. However, the transfer of the samples onto the plate was labour intensive and could not be automated with robotics. This process, thus had an increased chance of operator error. Before the transfer of the samples, each 384-well sample plate was divided into eight sections, as illustrated in Figure 2.2, and only one section, comprised of 48 samples, could be transferred at any one time. Two different plate guides, (one for plate areas 1, 3, 6 and 8, and the other for plate areas 2, 4, 5 and 7) were needed for transferring the samples from the sample plate on to the TaqMan OpenArray plate. Furthermore, only 96 samples could fit onto an OpenArray plate, therefore, five plates were required for genotyping the GenomePlex, GenomiPhi, PEP and REPLI-g products, and the corresponding genomic DNA.

There is no easy way of tracking the sample plates or sections after the samples have been transferred onto the OpenArray plates – the manual suggested writing the barcodes on the plates before transferring the samples. This created another area where operator error could occur. The serial number of the OpenArray plate also

had to be manually entered into the software before performing the imaging run. Again, there was potential for operator error when entering the OpenArray plate serial number.

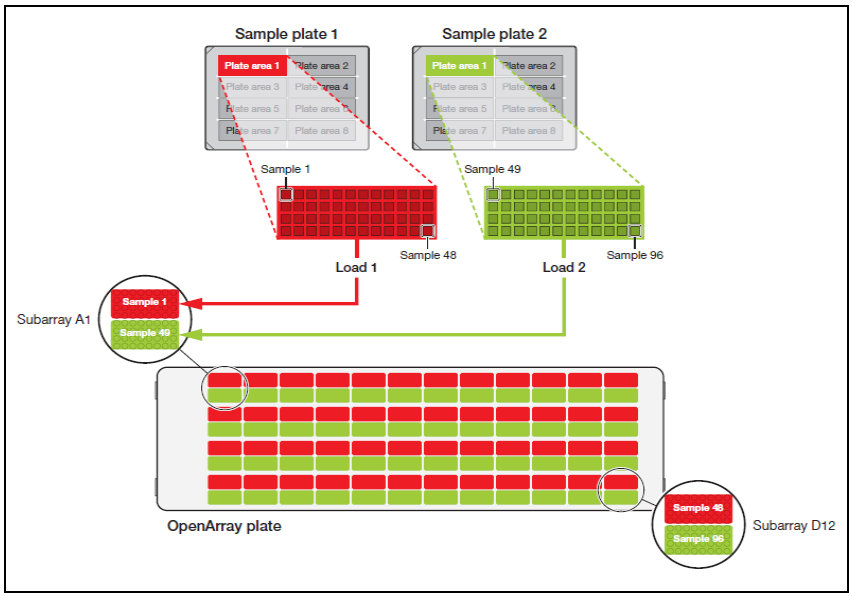



Figure 5.1: The transfer of sections of sample plates to an OpenArray plate

Each OpenArray sample plate section contains 48 wells, and each OpenArray plate can hold 96 samples when the 32-plex option is used. Normally the 4 sections of a sample plate are transferred to an OpenArray plate.

The OpenArray SNP Genotyping analysis software was reasonably easy to use. Sample well positions were included in the sample information files which were needed for the plate "set-up" files prior to imaging. However, the OpenArray SNP Genotyping Analysis software neither displayed, nor outputted the well positions of samples, which made it difficult to ascertain areas of the sample plates which failed genotyping.

Table 5.3: Ease of use of SNP multiplex genotyping platforms

| | **iPLEX** | **OpenArray** | **SNPlex** |
|---|---|---|---|
| Company | Sequenom | Applied Biosystems | Applied Biosystems |
| Multiplex level tested | 27 | 32 | 48 |
| Panel design | SNPs must work in panel | Any SNP combination | SNPs must work in panel |
| DNA required (ng) | 10 | 125 | 50-100 |
| DNA per SNP (ng) | 0.34 | 3.91 | 1.04-2.08 |
| Total no. samples/plate | 384 | 96 | 384 |
| Type of procedure | Extensive post-PCR processing | Extensive sample preparation | Extensive post-PCR processing |
| SNP/allele detection | MALDI-TOF | Fluorescence | Mass and fluorescence |
| Experiment time | 2 days | 2 days | 3 days |
| Ease of procedure | Easiest | Least easy | Moderately easy |
| Ease of analysis | Easiest | Moderately easy | Least easy |
| SNP pass rate (gDNA)* | 24 (100%)[§] | 32 (100%) | 29 (60%) |

* Overall rate for non-amplified genomic DNA only. [§] Three SNPs with insufficient extend primers failed.

Applied Biosystems recommended manually calling the genotypes on OpenArray. The advantage of using the OpenArray genotyping software was that it allowed the importing and analysis of multiple OpenArray plates in the same "genotyping project". This compensated for having to manually call the genotypes, and also allowed the analysis of multiple plates, provided there was minimal plate-to-plate variation in genotyping quality. The OCAC quality control guidelines required that at least 95 ovarian cancer samples were placed on each 384-well plate in order to eliminate any possible case/control genotyping bias. However, the sample information file, and the fact that only 96 samples could be genotyped on a 32-plex format, which would make it impossible to conform to the guidelines for genotyping studies. Adjustments to incorporate the OCAC guidelines would involve re-organisation of all sample plates layouts and the corresponding templates used for quality control purposes.

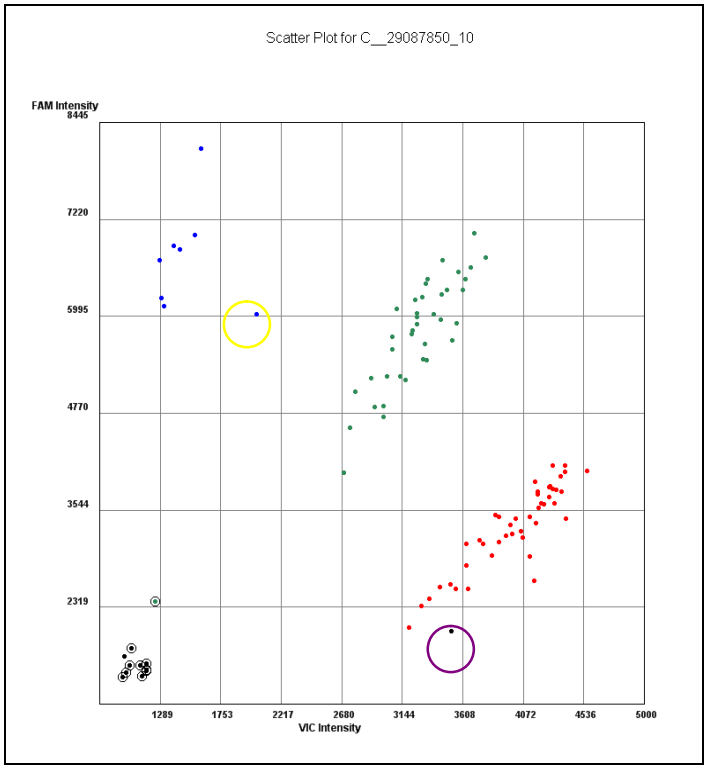



Figure 5.2: OpenArray cluster (auto-call)

Dots: Blue – homozygotes for FAM allele; green – heterozygotes; red – homozygotes for VIC allele; black – undetermined genotype; purple circle – sample could not be given a user call without another VIC homozygote becoming "uncalled", however a user call (to undetermined) could be made for the genotype circled in yellow.

Despite the manual calling, there was a considerable amount of automation in the calling when "cluster centre" and exclusion bars were used. There were instances when some samples could not be called with the cluster centre and exclusion bars, without a logical reason (as demonstrated in Figure 5.2). The "Draw" function of the OpenArray genotyping analysis software was supposedly available for modifying the genotype cluster shapes. However, the genotype calling from the Draw function could not be exported, which rendered the function useless.

### 5.3.2: SNPlex

The SNPlex genotyping platform was a fairly standard procedure. Although the procedure was carried out manually, the vast majority of the steps could be

automated with liquid handling robotics. However, there was a risk of contamination, and/or operator error with the extensive post-PCR process.

Manual calls could not be made per se, with the GeneMapper software. The clustering parameters could be modified in order to make the callings more, or less stringent. The quality of the experiments could be evaluated by checking the allelic ladders, and the software had an internal quality control that failed assays with call rates less than 80%, therefore it was not possible to find the exact call rates, or genotypes of poor performing assays.

### 5.3.2:iPLEX

The iPLEX genotyping platform was the preferred SNP multiplex genotyping platform. The platform performed the least level of multiplexing level compared with SNPlex and OpenArray, however all steps of the protocol were highly automatable, and relatively simple to perform. One of the few problems with the iPLEX procedure was the use of a dimple plate for transferring Clean Resin into the diluted extend products. There was a chance that less than 6mg of Clean resin was dispensed into each well of the dimple plate, as the method used is not particularly accurate. There was also a risk that the sample and dimple plates did not completely align when the resin was transferred to the sample plate, so small amounts of resin may not have entered the wells of the sample plate.

The iPLEX genotype analysis software, TyperAnalyzer, was very user-friendly and relatively easy to use. It was possible to find an approximate call rate for each SNP prior to data output, and colour coding of the “traffic light” plot indicates the overall

performance of each sample. Although auto-calling function of the software was fairly accurate, user (manual) calls could also be made. The software gave a "Status" for each genotype, either "conservative", "moderate", "aggressive", "low probability", "bad spectrum", "user" or "No-alleles". User-calls could be made for clear and distinct genotype clusters, like that shown in Figure 5.3.

Perfect clustering Possible user calls

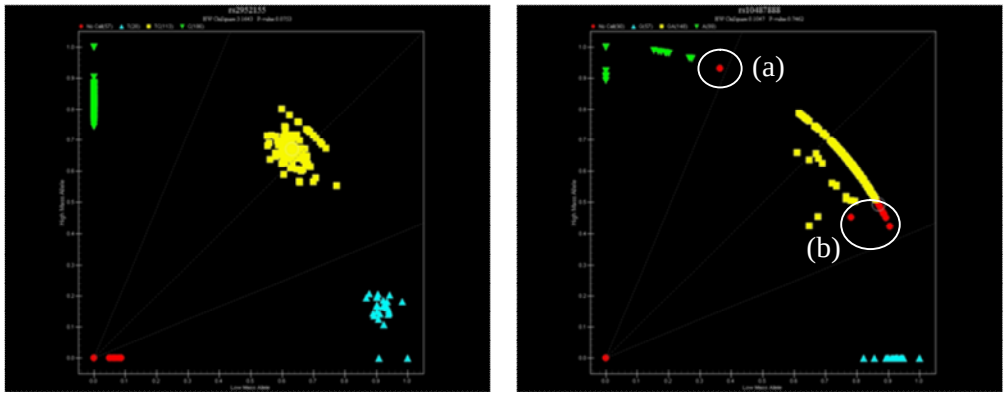



Figure 5.3: Examples of iPLEX clustering

Green triangle – homozygotes of high mass allele; yellow squares – heterozygote; blue triangle – homozygote of low mass allele; red dots – uncalled genotypes; (a) auto-call did not call the sample because the spectra was noisy, however, it was clear that the genotype was homozygous for the high mass allele; (b) samples were not auto-called because the allele peak heights were not equal, but the peaks were high enough to call them heterozygous.

The spectrum of each reaction could be viewed to determine the reason behind the auto-caller not calling some samples, and whether the genotype could be manually called. Examples of spectra of genotypes giving conservative, aggressive and user calls statuses are shown in Figure 5.4. Genotypes with conservative or moderate call status were considered accurate, and those with bad spectra or no-alleles were considered as failed assays. However the software automatically assigned genotypes for those conservatively, moderately or aggressively called.

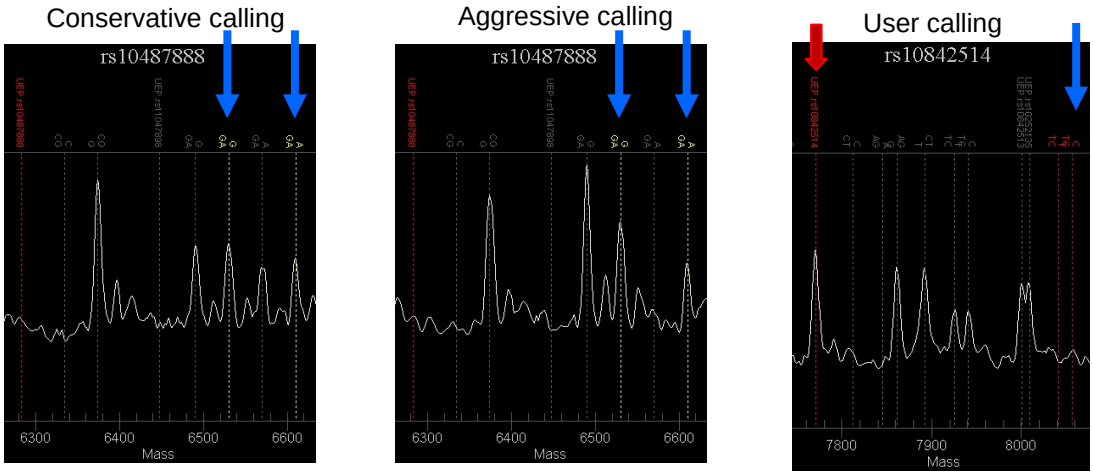



Figure 5.4: Examples of iPLEX genotype call statuses

Blue arrow – allele peaks; red arrow – unextended primer peak (UEP); the allele peak of the user calling cluster was not auto-call because it was low and the UEP was very high – suggesting a high proportion of the extend primers had not been incorporated in the reaction.

iPLEX was the only multiplexing platform on which some non-template negative test controls (NTCs) failed. Failed NTCs were found in four assays with the genomic DNA and GenomePlex, PEP and REPLI-g. Most of the failed NTCs were with rs11551174, which had a high call rate, and all the NTCs were homozygous for the G allele. The failed NTCs may have been caused by dimer pairs forming in the iPLEX reaction, which were the same mass as the GG genotype of rs11551174. Failed NTCs appears to be an artefact of the platform, and are not necessarily an indication of DNA contamination, due to concordant duplicates. However failed NTCs are a still a major concern.

## 5.4: Concordance of WGA products with genomic DNA on Taqman

The WGA-amplified DNA and their corresponding genomic DNA were genotyped with four TaqMan assays in order to compare the clustering, call rates and

concordance. The average call rate for the genomic DNA was 97%. Samples amplified with PEP had the highest call rates, averaging 100% for the 4 TaqMan SNPs genotyped. GenomePlex also had excellent call rates, which averaged 98%. The call rates of GenomiPhi were not as good as GenomePlex, but they were all greater than 90% (averaging 93%). Although REPLI-g amplification produced the greatest yield of amplified DNA, it had the poorest call rate on the Taqman genotyping platform.

The highest call rate for the REPLI-g-amplified DNA genotyped on TaqMan was 87%, and the average call rate for the assays combined was 82%, which is below the level accepted by the Ovarian Cancer Association Consortium (OCAC). Refer to Appendix IX-A for the individual call rates and concordance of the WGA samples compared with non-amplified genomic DNA genotyped on the TaqMan platform. REPLI-g amplified DNA also produced the worst clusters, as shown in Figure 5.5, which explains the poor call rates.

No discordances were found between the genotypes of genomic DNA and GenomePlex, GenomiPhi and PEP-amplified DNA on TaqMan. It should be taken into consideration that only a small number of assays were tested. However, the SNPs genotyped on TaqMan had high minor allele frequencies (MAFs). A total of four discordances between the genotype of genomic DNA samples and their corresponding REPLI-g-amplified DNA were found on TaqMan (for rs602652, rs3217869 and rs10487888). However, the concordance levels were >98% for two of the polymorphisms, therefore the genotyping results for these (rs602652 and rs3217869) were acceptable under quality control (QC). The genomic DNA was

homozygous for the rare allele of rs602652; however, the corresponding REPLI-g amplified DNA was homozygous for the common allele of the SNP. This discordance, is known as a "miscall" and cannot be explained by unequal amplification of the alleles.

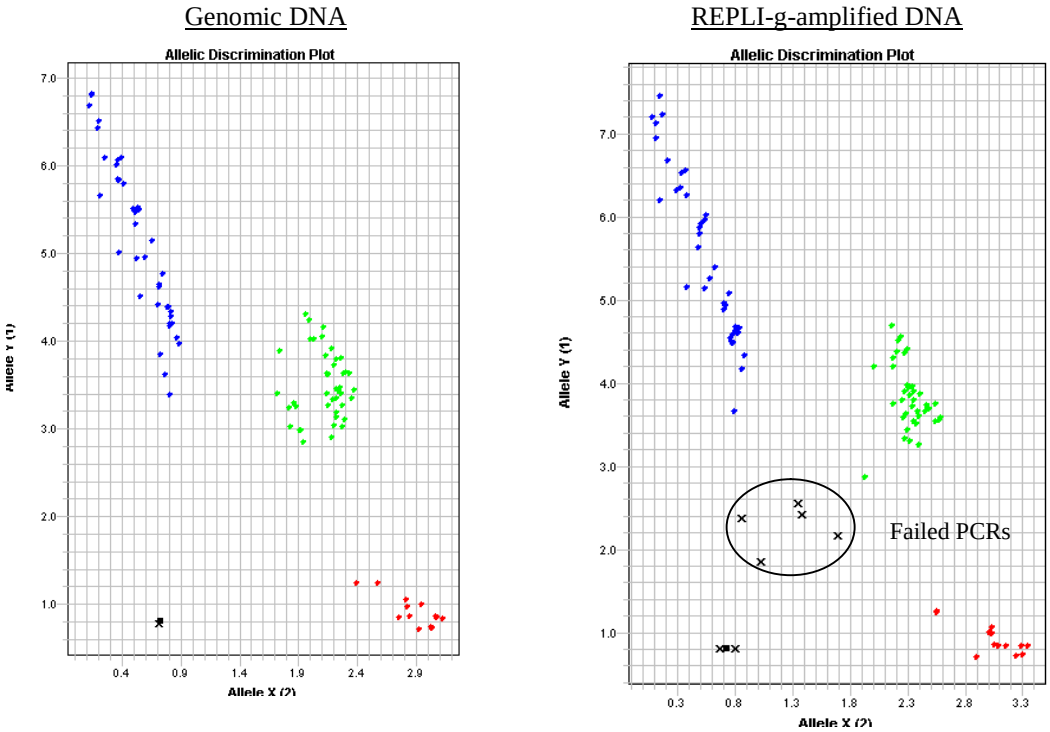



Figure 5.5: Clustering of genomic and corresponding REPLI-g-amplified samples with rs602652

Clusters: blue – common homozygous; green – heterozygous; red – rare homozygous; black x – failed reactions; black square – non-template control.

Miscall discordances consist of the genotype of the genomic DNA being homozygous for an allele of a SNP, and the corresponding WGA DNA being homozygous for the other allele, or heterozygous for the same SNP as demonstrated in Figure 5.6. The other type of discordance is known as "allele-drop", which involves the loss of an allele in the WGA DNA (when the genomic DNA is

heterozygous), suggesting unequal amplification of an allele at a heterozygous locus (see Figure 5.6). This type of discordance was found with rs10487888.

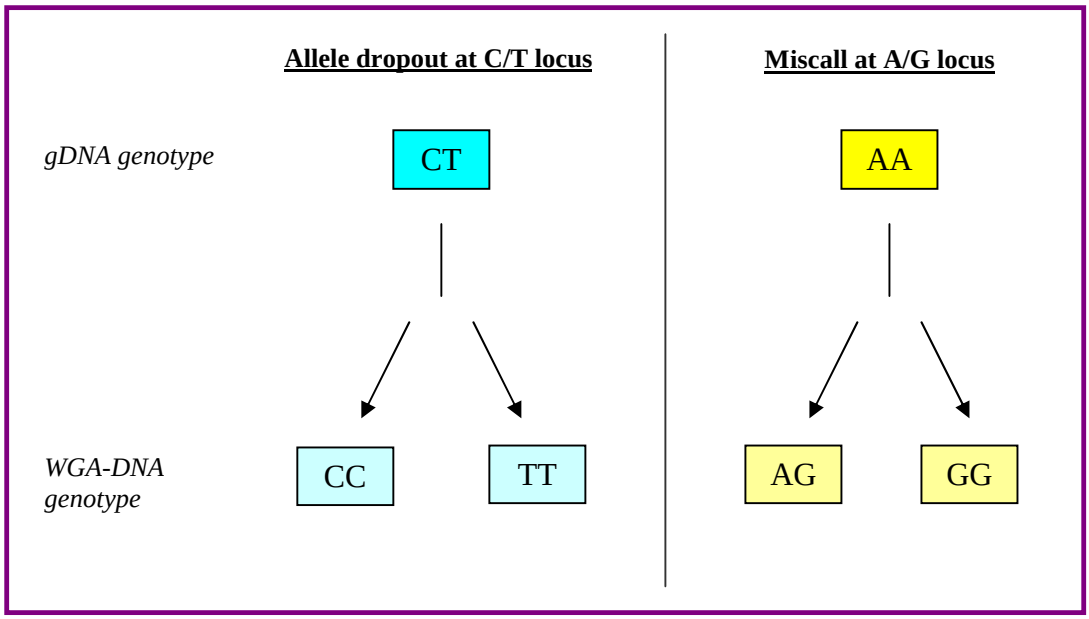



Figure 5.6: Types of discordances

The clustering of REPLI-g amplified DNA suggested that there was unequal amplification of the alleles. Ninety-five other samples were amplified with REPLI-g, and sequenced in order to investigate unequal amplification of the alleles further. *BRCA1* sequencing data was already available for the matching genomic DNA of these samples. Therefore, the sequences of the genomic and REPLI-g amplified DNAs for *BRCA1* regions 316705 in exon 11, and 316700 in exon 13 of the Variant SeqR kit, which contained SNPs rs16941 and rs1060915, respectively.

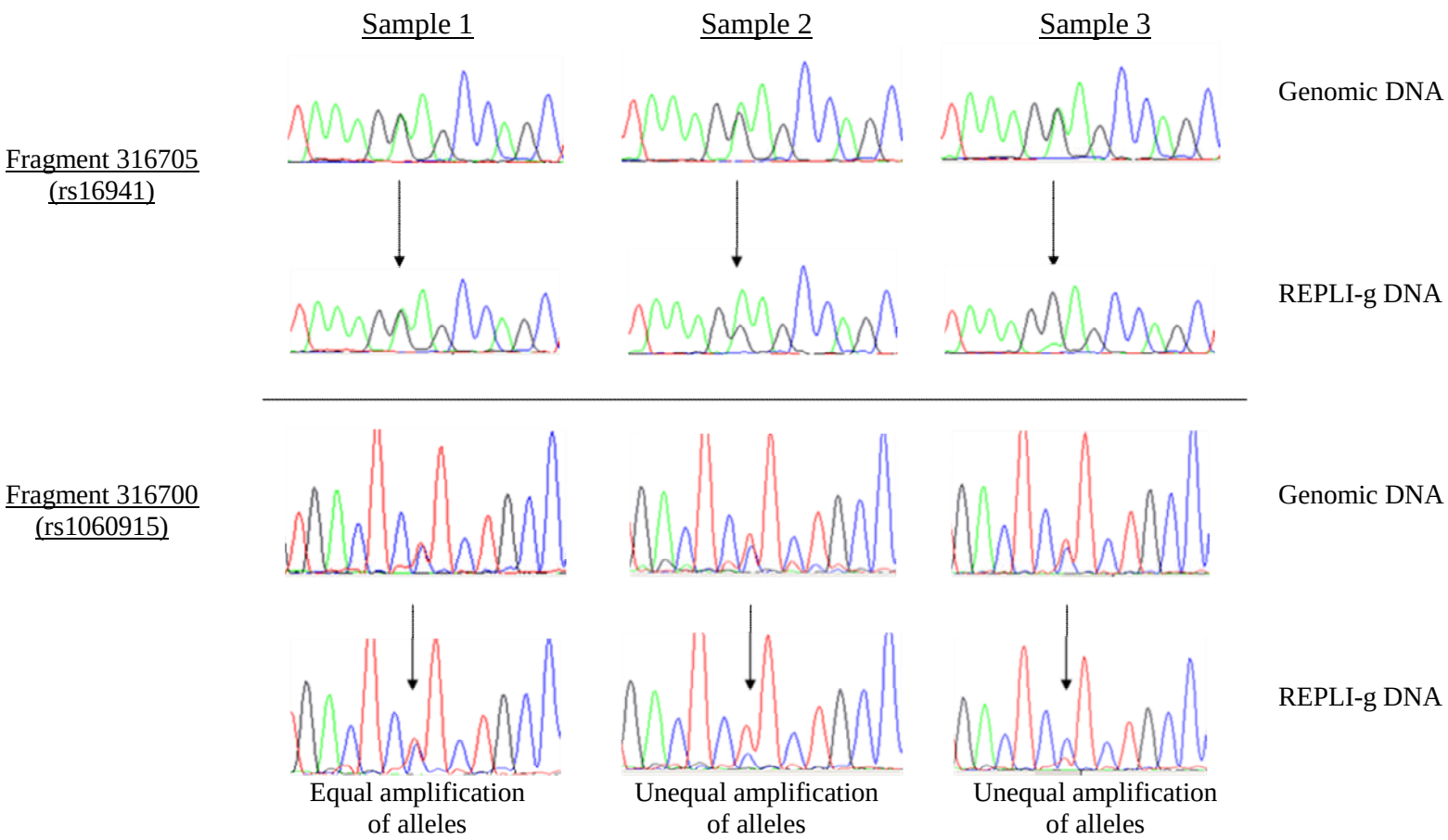



Figure 5.7: Unequal amplification of REPLI-g amplified DNA

Figure 5.7 shows electropherograms of gDNA and corresponding sequences for REPLI-g amplified DNA with equal and unequal amplification of alleles. The green, blue, black and red peaks of the electropherograms correspond to bases A, C, G and T, respectively. Overlapping peaks are heterozygous genotypes. The SNPs are in LD ($r^2$=1) and have a minor allele frequency of 0.35. There was 97% concordance between the genotypes of genomic and REPLI-g DNA for the SNPs. There were 44 heterozygous individuals for these SNPs and 3 of them (6.8%) showed unequal amplification of the linked alleles of the polymorphisms.

## 5.5: The performance of genomic and amplified DNA on SNP multiplex genotyping platforms

The next aim of the study was to investigate the performance of the whole genome amplified products on SNP multiplex genotyping platforms. In order to execute this aim, the WGA DNA and their matching genomic DNA were genotyped on iPLEX, SNPlex and OpenArray. The resultant call rates and concordances were evaluated. Assays which failed were excluded from all, but the SNP pass rate calculation.

### 5.5.1:Call rates

The 95 non-amplified genomic DNA samples and their corresponding GenomePlex, GenomiPhi, PEP and REPLI-g-amplified DNA were genotyped on the iPLEX genotyping platform with a 27-plex panel. The performances of the genomic and WGA DNA on the platform were assessed by calculating the SNP pass rate (the number of SNPs from the panel which produced callable genotype clusters), and the per SNP assay call rates (the number of samples successfully assigned a genotype).

From the 27-plex panel, three assays (rs6944385, rs10842514 and rs10252135) with insufficient amounts of extend primers, consistently failed for all the WGA products and the genomic DNA. The auto-calls from the iPLEX TyperAnalyzer software were evaluated. The call rates per assay/SNP are shown in Appendix IX-B. Refer to Table 5.4 for the summary of the call rates from iPLEX genotyping.

Overall, the GenomePlex, GenomiPhi and RepliG genotypes had call rates >95% for more assays (23 tSNPs each) than the genomic DNA (17 tSNPs). PEP-amplified DNA had the poorest call rates with only 10 assays with call rates greater than 90% and an average of call rate of 89%, see Table 5.4. The average call rates for the genomic and GenomePlex, GenomiPhi and RepliG-amplified samples were >95%.

Table 5.4: iPLEX call rates by DNA amplification method

| Call rate (%) | Genomic | GenomePlex[§] | GenomiPhi | PEP | REPLI-g |
|---|---|---|---|---|---|
| <80 | 1 | 0 | 0 | 9 | 0 |
| 80-89 | 1 | 1 | 1 | 5 | 1 |
| 90-94 | 5 | 0 | 0 | 2 | 0 |
| ≥95 | 17 | 23 | 23 | 8 | 23 |
| Mean | 97% | 99% | 99% | 81%* | 99% |

The 3 assays which failed for all amplification methods and gDNA were not included in the analysis;. § n=90, there were 95 samples for all other methods.

## 5.5.2:Concordance rates from iPLEX genotyping

The fidelity of the amplification methods were assessed with the iPLEX platform by comparing the genotypes of genomic DNA with the matching WGA DNA. GenomiPhi results had the lowest level of discordance. There were only 2 discordances in the 1,872 genotypes called in both the GenomiPhi-amplified DNA and their corresponding genomic DNA. This averaged to a concordance rate of

99.9%. The greatest level of discordance was found between PEP-amplified DNA and their corresponding genomic DNA (1.62%).

Table 5.5: Overall discordance per WGA method on iPLEX platform

| | **GenomePlex** | **GenomiPhi** | **PEP** | **REPLI-g** |
|---|---|---|---|---|
| Total genotypes | 1872 | 1998 | 1423 | 1972 |
| Total discordant | 5 | 2 | 23 | 20 |
| No. discord SNPs | 4 | 2 | 7 | 13 |
| Discordance (%) | 0.3 | 0.1 | 1.6 | 1 |

As shown in Table 5.5, when the overall discordance per WGA method is considered, none of the methods resulted in discordance greater than or equal to 2%. Therefore all the WGA methods would have passed the OCAC quality control criteria for the concordance between duplicates.

When the discordances per assay were considered, only the genotypes of GenomiPhi DNA were acceptable in terms of discordance rates (<2%) for the 24 SNPs. The discordance rates of the SNPs genotyped on iPLEX are shown in Appendix IX-C. Of the 19 SNPs successfully genotyped with PEP-amplified DNA, 4 assays were discordant for > 2% of the samples. The highest discordance rate was 19% for rs3771886 for the PEP DNA, however the call rate was very poor (32%). The second highest level of discordance (13%) for the PEP samples was for the for the rs10487888 SNP, which had a call rate of 85% (see Appendix IX-C). Discordances were not restricted to SNPs with poor call rates, refer to Appendices IX-B and IX-C. The call rates of the genomic and PEP DNAs for rs1801200 was 98%, however, 2% of the genotypes were discordant. Furthermore, 8 of the 13 SNPs with discordances were REPLI-g DNA with call rates >95%. The greatest number of SNPs with

discordances was found with the REPLI-g-amplified DNA (13 out of 22 SNPs [~60%]). Of these, 7 SNPs were discordant for ≥2% of genotypes (see Appendix IX-C). Moreover, only one assay (rs17623382) had a call rate <90% (REPLI-g vs genomic DNA). Discordances were found between genomic DNA GenomePlex for four of the SNPs. However, only one (rs11551174) was discordant for >2% of the genotypes (see Appendix IX-C). Of the 4 assays genotyped with GenomePlex-amplified DNA, there was one, with call rate >90%, with >2% discordance. GenomiPhi-amplified DNA were the only one which did not have discordances >2% for any of the assays; see Appendix IX-C.

All of the discordances found with GenomePlex DNA were allele dropouts except one of the rs11551174 discordances, where the genomic DNA was homozygous for the "G" allele and the amplified sample was heterozygous for the SNP. The PEP DNA failed for this assay. The other WGA DNA genotypes (GenomiPhi and REPLI-g) were also heterozygous for the SNP, which suggests that the genomic DNA was incorrect. The genomic DNA was discordant for all WGA DNA for another SNP, rs17623382, which also indicates that the genomic DNA was incorrect. When the genotypes of these two SNPs for the particular sample are excluded, there are no longer discordances with GenomiPhi DNA on the iPLEX platform. The vast majority of the discordances with PEP and REPLI-g DNA were allele dropouts, suggesting unequal amplification of the alleles (refer to Table5.6).

Table 5.6: Types of discordances found with iPLEX (by WGA method)

| SNP | GenomePlex | | GenomiPhi | | PEP | | REPLI-g | |
|---|---|---|---|---|---|---|---|---|
| | **Dropout** | **Miscall** | **Dropout** | **Miscall** | **Dropout** | **Miscall** | **Dropout** | **Miscall** |
| rs10487888 | 0 | 0 | 0 | 0 | 9 | 2 | 1 | 1 |
| rs11047917 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 |
| rs11551174 | 1 | 1* | 0 | 1* | 0 | 0 | 1 | 1* |
| rs12305513 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| rs12822857 | 0 | 0 | 0 | 0 | Failed | | 1 | 0 |
| rs17623382 | 1* | 0 | 1* | 0 | Failed | | 1* | 1 |
| rs1801200 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 0 |
| rs2161841 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| rs2699905 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| rs2952155 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| rs3771882 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| rs3771886 | 0 | 0 | 0 | 0 | 6 | 0 | 1 | 0 |
| rs3854012 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| rs4623993 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **Total** | **4 (3)** | **1 (0)** | **1 (0)** | **1 (0)** | **19** | **4** | **14 (13)** | **4 (3)** |

Dropout – Genotype of genomic DNA is heterozygous. Miscall – genotype of genomic DNA is homozygote for an allele. * Genomic DNA appeared to be incorrect; Total discordance in parenthesis are the totals when the incorrect genomic results are excluded.

### 5.5.3:Assays with discordances in more than one amplification method

At least one DNA sample amplified with GenomePlex and REPLI-g was also discordant for the assays with discordances in the GenomiPhi-amplified DNA. There was also at least one discordant REPLI-g-amplified DNA sample with each assay that had a discordant genotype for the GenomePlex-amplified DNA.

The discordances found with GenomePLEX-amplified samples on the iPLEX genotyping platform occurred in different samples – each sample was only discordant for 1 SNP. rs11551174 was the only SNP with discordance in more than 1 sample. No sample was discordant in more than 1 WGA method. However, three and four samples amplified with PEP and REPLI-g, respectively, were discordant for 2 or more assays. This suggests that there may have been either a problem with the amplification of these samples with the WGA method, or with the genotyping of the individual samples.

### 5.5.4:The performance of gDNA and WGA-DNA on SNPlex

The automatically assigned genotypes of the genomic and WGA-DNA from the SNPlex platform were analysed with the GeneMapper software. No assays were successfully genotyped with the PEP-amplified DNA on the platform. However, it is worth noting that the GeneMapper software automatically fails samples which fail for >80 of the SNPs, and also if the per assay call rate is below 80%. Therefore, it was not possible to find the exact call rates of the assays which failed the internal quality control criteria. The REPLI-g DNA had the highest SNP pass rate (73%) and aside from PEP DNA samples, GenomiPhi had the lowest SNP pass rate; see Table 5.7. Twenty-nine SNPs were successfully genotyped with genomic DNA, and there

was an average genotype call rate of 97% for these SNPs. In comparison, the GenomiPhi DNA, which performed badly on SNPlex, had the lowest average call rate of 85%. The SNP pass and call rates are summarised in Table 5.7; the individual SNP call rates for the genomic and WGA DNA from the SNPlex genotyping are tabulated in Appendix IX-D.

All but 1 of the SNPs which passed the GeneMapper internal QC had SNP genotype call rates >95% for the genomic DNA. GenomePlex was the only amplification method with SNPs with call rates >95 on SNPlex. However, this comprised only 2 out of the 13 SNPs which passed genotyping on SNPlex (see Table 5.7).

Table 5.7: SNPlex assay pass rates (by WGA method)

| | **Genomic** | **Genomeplex** | **GenomiPhi** | **PEP** | **REPLI-g** |
|---|---|---|---|---|---|
| **SNP pass rates** | | | | | |
| DNA conc (ng/ìL) | 50 | 100 | 100 | 100 | 100 |
| No. SNPs passed | 29 | 13 | 9 | 0 | 35 |
| % SNPs passed | 60.4 | 27.1 | 18.8 | 0.0 | 72.9 |
| No. of calls | 2647 | 1147 | 722 | 0 | 3010 |
| **Call rates*** | | | | | |
| <80% | 19 | 39 | 35 | 48 | 13 |
| 80-89% | 0 | 0 | 9 | 0 | 3 |
| 90-94% | 1 | 11 | 0 | 0 | 32 |
| >95% | 28 | 2 | 0 | 0 | 0 |
| **Mean** | **97%** | **94%** | **85%** | **Fail** | **91%** |

gDNA – genomic DNA; conc – concentration; the call rates for all assays genotyped with PEP amplified DNA were less than 80%; * based on SNPs with >80% call rate (pass).

### 5.5.5:Concordance between gDNA and WGA-DNA on SNPlex

When the concordance between genomic DNA and the WGA-DNA genotyped on SNPlex was evaluated, there was evidence of discordances between the genomic DNA and the corresponding amplified DNA for all the WGA methods. Overall,

DNA amplified with GenomePlex appeared to have the greatest proportion of discordant genotypes, when the autocall genotypes from the SNPlex platform were analysed. There were 217 discordance samples from those with genotypes for both genomic DNA and the amplified DNA out of 699 genotypes. This accounted for 31% of the genotypes. This contrasted sharply with the discordances found with the GenomiPhi and REPLI-g-amplified DNA 0.7% and 3.6%, respectively.

Upon inspection of the genotype clusters of the assays for the genomic DNA and WGA-DNA, it became apparent that some assays which were deemed as good quality, were indeed failed assays. rs751340, rs2286216, rs927221 and rs1713423 failed to produce callable clusters for GenomePlex-amplified DNA. The clusters of rs751340, rs2286216 and rs1713423 for all the samples are shown in Figure 5.8. However, the GeneMapper software assigned genotypes to the samples, resulting in approximately a third of GenomePlex-amplified DNA samples (61%) being discordant for these variants alone (see Appendix IX-E). Figure 5.8 shows some of these clusters with poor quality and miscalled clustering.

When the failed assays were excluded, the overall concordance rates were 98%, 99% and 96% for GenomePlex, GenomiPhi and REPLI-g, respectively. See Appendix IX-E for the discordance rates for each SNP. Of the assays which passed for both gDNA and WGA-DNA, >98% concordance was observed for 2 assays out of 8 genotyped in both genomic and GenomePlex, 8 of 9 assays with GenomiPhi, and 3 of the 25 assays with REPLI-g-amplified DNA (see Appendix IX-E).

All five discordances of the GenomiPhi amplified DNA were allele dropouts (heterozygous for genomic DNA, homozygous for the WGA-DNA). Of the 7 discordances found between the GenomePlex-amplified DNA and genomic DNA, only 2 were allele dropouts (see Appendix IX-E), and the 4 of the remaining discordances were miscalls of the same SNP (rs1419755). Refer to Appendix IX-F for the types of discordances for the successfully genotyped SNPs. Only 25 (33%) of the REPLI-g discordances were allele dropouts (see Appendix IX-F), the remaining discordances were miscalls. However, some of these may be attributable to inadequate genotype calling of the GeneMapper software.

A sample amplified with GenomePlex was discordant for 2 assays on the SNPlex platform. Although, there were 76 discordances with REPLI-g DNA on SNPlex, these occurred with 10 samples. The discordances did not correlate with amplification yield. Of the 10 REPLI-g-amplified samples with discordances, only 2 samples were discordance for a single assay. The other 8 REPLI-g-amplified samples were discordant for at least 3 assays each. Thus, these 8 samples were discordant in 74 instances. There were 4 instances when the same discordances were found with GenomiPhi and REPLI-g DNA and on each occasion, the discordance was an allele dropout. There is a possibility that the same alleles were preferentially amplified with both GenomiPhi and REPLI-g methods at the 4 loci, however, due to the poor quality genotypes from the SNPlex platform, it is more likely that either the genomic or WGA genotypes were incorrect.

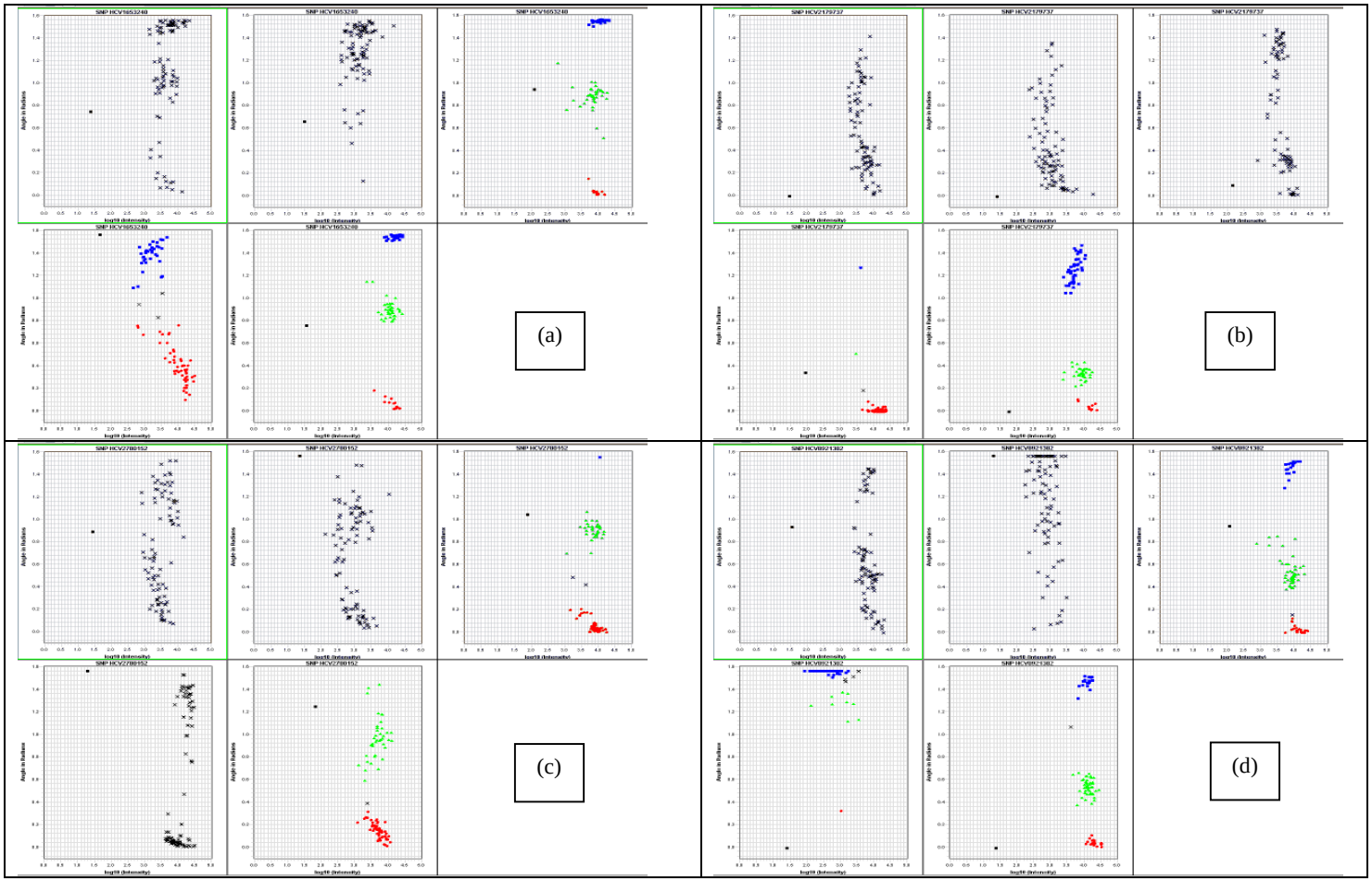



Figure 5.8: Discrepant auto-calling of SNPlex platform

(a-d) Clusters top left to right: GenomiPhi, PEP, and REPLI-g; bottom L to R: GenomePlex and genomic DNA.
(a) rs751340
(b) rs2286216
(c) rs1861606
(d) rs1713423
Clusters: blue – homozygous for allele 1; green – heterozygous; red – homozygous for allele 2; black square – non-template test control; black "x" – failed genotyping.

### 5.5.6:The performance of gDNA and WGA-DNA on OpenArray

The samples amplified with GenomePlex, GenomiPhi, PEP and REPLI-g, and their corresponding genomic DNA were genotyped with a 32-assay panel on the Open array multiplex genotyping platform. The samples were initially prepared manually. The genotyping was performed with Applied Biosystems' staff. The technical representative for the platform recommended manually calling the genotypes on the TaqMan OpenArray™ SNP Genotyping Analysis Software.

As on SNPlex, the PEP-amplified DNA failed to produce callable clusters on the OpenArray platform. Refer to Appendix IX-G for the call rates for each assay for the genomic, and GenomePlex, GenomiPhi and REPLI-g. There were callable clusters for all the assays with the genomic DNA, however, 10, 4 and 3 SNPs failed for GenomePlex, GenomiPhi and REPLI-g amplified DNA, respectively. As shown in Table 5.8, all but one of the SNPs had call rates >90% for the genomic DNA, with an average call rate of 97%. However, none of the SNPs genotyped with amplified DNA resulted in average call rates >90% (see Table 5.8).

Table 5.8: OpenArray call rates*

| Call rate (%) | Genomic | GenomePlex | GenomiPhi | REPLI-g |
|---|---|---|---|---|
| <80 | 0 | 10 | 5 | 10 |
| 80-89 | 1 | 14 | 13 | 22 |
| 90-94 | 2 | 7 | 14 | 0 |
| ≥ 95 | 29 | 1 | 0 | 0 |
| **Average** | **97%** | **88%** | **89%** | **82%** |

N=93 *based on manually called SNPs)

When the overall concordance rates between genomic DNA, and their corresponding whole genome amplified DNA were assessed, again, samples amplified with REPLI-

g had the lowest concordance (92.7%). The overall concordance for GenomePlex and GenomiPhi was 97.1% for 1,693 and 2,270 genotypes, respectively. Therefore, none of the overall concordance rates for the amplification methods met the OCAC criteria of ≥98%; (see Appendix IX-H for the discordance rate for each assay).

Only 5 SNPs out of the 21 successfully genotyped OpenArray SNPs had concordance rates >98 for the GeomePlex amplified DNA. Again, DNA amplified with GenomiPhi were concordance for >98% of genotypes, for the highest proportion of SNPs (12/29 variants). As with iPLEX and SNPlex, REPLI-g amplified DNA resulted in the highest proportion of SNPs with discordance rates >2%. Of the 29 polymorphisms successfully genotyped with REPLI-g DNA, only 2 assays had acceptable concordance rates (see Appendix IX-H).

As observed with iPLEX and SNPlex, there were both allele dropouts and miscall discordances between the genomic DNA and the matching amplified DNA on the OpenArray platform; refer to Appendix IX-I for the number of each type of discordance per SNP. However, for all of the WGA methods, the majority of discordances were miscalls, which suggests there was a serious problem with using DNA amplified with PEP, GenomePlex, GenomiPhi or REPLI-g WGA on this platform. The total numbers of allele dropouts and miscalls for the WGA products which were successfully genotyped on OpenArray are shown below:

| GenomePlex | | GenomiPhi | | REPLI-g | |
|---|---|---|---|---|---|
| Dropout | Miscall | Dropout | Miscall | Dropout | Miscall |
| 12 | 38 | 15 | 51 | 72 | 85 |

There were several occasions (18) when the genomic DNA was homozygous for an allele, and the corresponding genotypes of the WGA DNA samples heterozygous for

all WGA methods. This occurred with 15 different assays, and 9 different samples, and suggests that some, if not all, of the genomic genotypes may be incorrect.

### 5.5.7: Reproducibility of the OpenArray genotyping data

Twelve duplicates of the WGA DNA and their corresponding genomic DNA were genotyped on OpenArray and compared with each other in order to ascertain reproducibility of genotyping data from the OpenArray platform. No discordances were found between the GenomePlex-amplified duplicates; but this method had the lowest SNP pass rate (59%). The concordance rates between the duplicates (by WGA method) are tabulated in Appendix IX-J. When all genotypes were considered, acceptable levels of discordance (<2%) were found with the genomic, GenomePlex and GenomiPhi DNA; see Table 5.9.

Table 5.9: Reproducibility of genotypes from OpenArray platform

| | **Genomic** | **GenomePlex** | **GenomiPhi** | **REPLI-g** |
|---|---|---|---|---|
| No. SNPs | 29 | 19 | 27 | 26 |
| No. genotypes | 316 | 206 | 299 | 253 |
| Average call rate | 90.8% | 90.4% | 95.8% | 81.1% |
| Discordance | 5 (1.6%) | 0 (0%) | 2 (0.7%) | **20 (7.9%)** |

The discordances in the duplicated genomic DNA were with 5 different SNPs. Due to the small numbers of duplicates assessed, each discordant pair will reduce the concordance rate by at least 8%. Again, REPLI-g had the greatest amount of discordances with 20 discordances out of 253 genotypes (8%, see Table 5.9). These REPLI-g discordances occurred in 14 SNPs, therefore the duplicates were concordant for only 12 SNPs.

### 5.5.8:Comparison of rs10487888 genotypes from iPLEX and OpenArray

One of the assays genotyped on OpenArray, rs10487888, was also genotyped with iPLEX Gold with genomic DNA. Comparisons of the genotyping results of the gDNA and WGA samples were made.

There was 100% concordance between genomic DNA genotyped on iPLEX Gold and GenomePlex and GenomiPhi DNA genotyped on OpenArray. However, 2 discordances were found between the genomic DNA which was genotyped on both platforms and 4 with REPLI-g (Table 5.10). All but 1 of the discordances found were allele dropouts in the OpenArray genotypes.

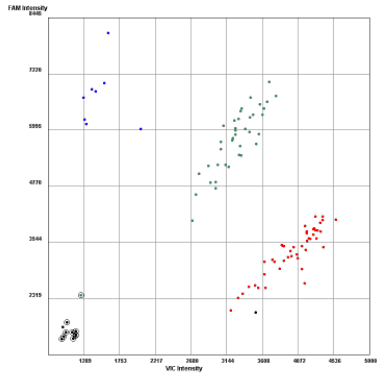
Table 5.10: Comparison of iPLEX Gold with OpenArray genotypes for rs10487888

| | **Genomic** | **GenomePlex** | **GenomiPhi** | **REPLI-g** |
|---|---|---|---|---|
| No. samples called | 89 | 83 | 90 | 89 |
| Call rate | 95.7% | 90.2% | 97.8% | 96.7% |
| No. discordant | 2 | 0 | 0 | 4 |
| Concordance | 97.8% | 100 | 100 | 95.5% |

Applied Biosystems recommended manually calling the genotypes, however this is time consuming, particularly as the number of SNPs and samples increases. Although auto-calling is able to correctly assign genotypes, based on clustering in some instances, there were also gross miscallings in others (see Figure 5.9).

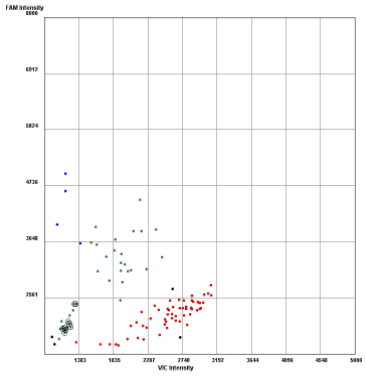
Of the auto-called genomic DNA assays, 10 (31%) of the SNPs with callable clusters were incorrectly called. Two assays with poor clustering were auto-called when they should not have been. The systematic miscalling also occurred with WGA samples.

Correct auto-calling

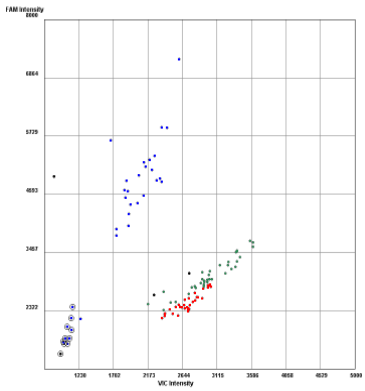


Assay 8

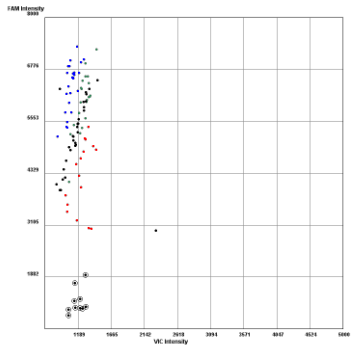
Incorrect auto-calling (below)



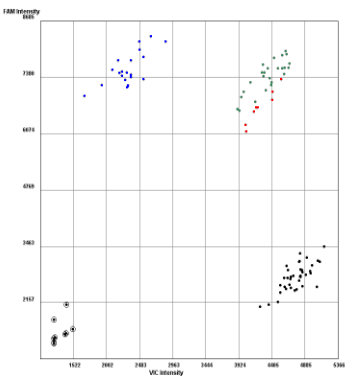
rs28665122



rs12722489



rs2066845



rs1050152

Figure 5.9: Examples of auto-calling with OpenArray Genotype Analysis software

Dots: Red – homozygous for VIC allele; green – heterozygous; blue – homozygous for FAM allele; black – uncalled; black dot with circle – non-template test control.

A meeting was conducted with Applied Biosytems staff to discuss the results from the OpenArray platform. During the meeting it became apparent that, as with the test runs on SNPlex, the OpenArray platform was still in its developmental stages. Although the company claims to have genotyped a great number of SNPs (>100) on the platform, only 20 samples were used in each experiment. The use of such small numbers of samples makes it difficult to gauge the quality of clusters, and also to assess reproducibility. Applied Biosystems suggested genotyping some of the SNPs with the same genomic samples on the TaqMan platform.

During the meeting it also became clear that the technical staff did not have faith in the SNPlex platform. The poor performance of the samples on the SNPlex and OpenArray platforms with SNPs which had been validated and optimised by Applied Biosystems, and the lack of explanation for the inadequate auto-calling, as well as low genotyping pass rates, makes OpenArray (and SNPlex) unsuitable for high-throughput genotyping with the research samples used.

## 5.6: Direct comparison of the multiplexing methods

As shown in Table 5.11, iPLEX genotypes produced the best average call rates and the fewest discordances of the SNP multiplex genotyping platforms. Aside from the genotypes from iPLEX, the REPLI-g amplified DNA always produced the most discordances, possibly due to over-amplification of the DNA, see Table 5.12. The OpenArray genotyping system was the worst performing platform in terms of genotyping call rate and discordances. On OpenArray, only GenomePlex DNA had an average call rate >90. None of the WGA methods had average concordances >98%. Moreover, most of the miscall discordances were identified on OpenArray,

and SNPlex platforms (see Table 5.12). When the criteria of 90% call rate and <2% discordance are taken into consideration, only 1 SNP, each for the GenomePlex-amplified DNA would fulfil both call rate and discordance on the SNPlex and OpenArray platforms, despite the relatively high numbers of SNPs genotyped (see Table 5.13). However, the criteria were fulfilled for 20 and 21 SNPs for GenomePlex and GenomiPhi-amplified DNA, respectively.

Table 5.11: Average call rate and discordances for each method

| | gDNA (n=95) | GenomePlex (n=90) | | GenomiPhi (n=95) | | PEP (n=95) | | REPLI-g (n=95) | |
|---|---|---|---|---|---|---|---|---|---|
| | Call rate | Call rate | Dis | Call rate | Dis | Call rate | Dis | Call rate | Dis |
| Taqman (n=5) | 97 | 97 | 0 | 94 | 0 | 97 | 0 | 82 | 1 |
| iPLEX (n=22) | 97 | 99 | 0 | 99 | 0 | **81** | **2** | 99 | 1 |
| SNPlex (n=29) | 96 | 94 | **2** | **84** | 1 | **Fail** | **Fail** | 90 | **4** |
| Open Array (n=32) | 95 | 91 | **3** | **87** | **3** | **Fail** | **Fail** | **80** | **7** |

Dis – discordance; call rates and discordances (%); Bold pass call ≤90% and discordance ≥2%.

Table 5.12: Types of discordances identified with each WGA method and platform

| | | GenomePlex (n=90) | | GenomiPhi (n=95) | | PEP (n=95) | | REPLI-g (n=95) | |
|---|---|---|---|---|---|---|---|---|---|
| | No. SNPs passed | Drop out | Miss-call | Drop out | Miss-call | Drop out | Miss-call | Drop out | Miss-call |
| TaqMan | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 |
| iPLEX | 24 | 4 | 1 | 1 | 1 | 19 | 4 | 14 | 4 |
| SNPlex | 29 | 2 | 5 | 5 | 0 | Fail | Fail | 25 | 51 |
| Open Array | 32 | 12 | 38 | 15 | 51 | Fail | Fail | 72 | 85 |

Number of assays passed for genomic DNA samples

Table 5.13: Number of SNPs with call rates ≥90% and discordances <2% for each method

| | gDNA (n=95)* | GenomePlex (n=90) | GenomiPhi (n=95) | PEP (n=95) | REPLI-g (n=95) |
|---|---|---|---|---|---|
| Taqman (n=5) | 5 | 5 | 5 | 5 | 0 |
| iplex (n=22) | 21 | 20 | 21 | 9 | 14 |
| SNPlex (n=29) | 29 | 1 | 0 | Fail | 0 |
| Open Array (n=32) | 22 | 1 | 0 | Fail | 0 |

* call rates only

## 5.7: Genotyping on iPLEX gold system

The iPLEX Gold platform, which is an upgraded version of iPLEX, is highly automated and accurately advertises high through-put genotyping. The only major difference between iPLEX and iPLEX Gold is the number of SNPs which can be genotyped in a single reaction, up to 29 and up to 40, respectively. This upgrade was achieved by increasing the mass range in which alleles/SNPs can be detected by the mass spectrometer. The work flow is the same as that of iPLEX - straightforward and simple to follow.

The automated genotype calling of iPLEX and iPLEX Gold was correct for the majority of genotypes, however an error was found with SNP rs3783197 (see Figure 5.10). The TyperAnalyzer software had obviously called the genotypes based on the peaks of the mass spectra, however, the plot clearly shows that there is an increase in the high mass molecules, resulting in a shift of the blue cluster (common homozygous). A similar shift was observed for rs6788750, however the shift was not as pronounced as it was for rs3783197.

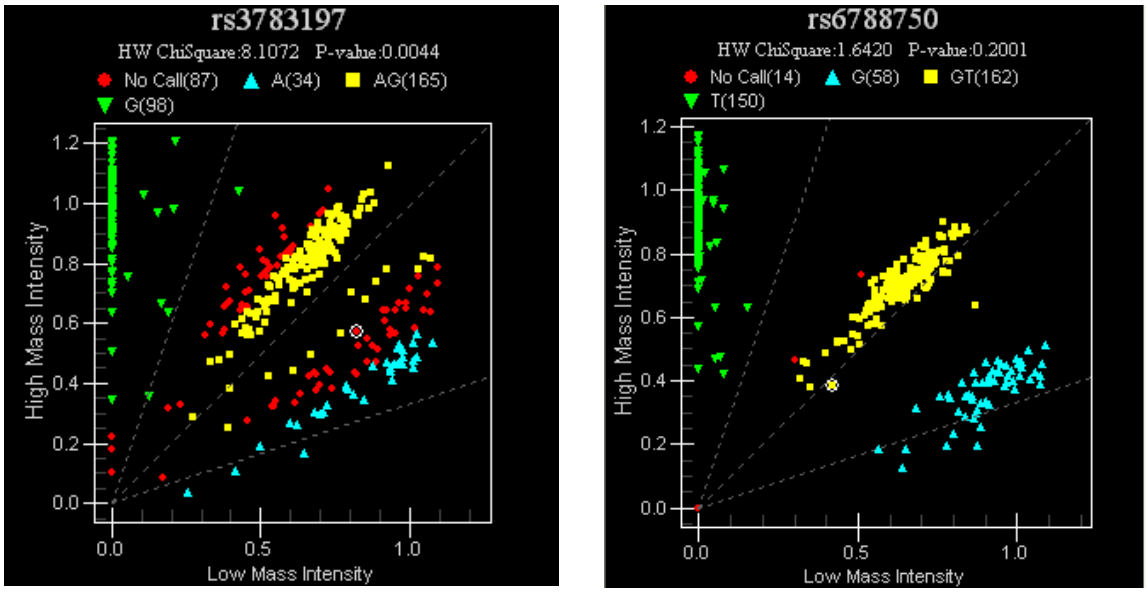



Figure 5.10: Shift in clusters on iPLEX clusters

The concordance between genomic and WGA amplified DNA on the iPLEX platform, was overall, acceptable for all WGA methods. Comparison of genotypes of the 95 samples from iPLEX and TaqMan were also concordant for rs17623382 and rs3771886. Acceptable concordance and call rates were found for MALOVA and SEARCH DNA with the iPLEX method with the variants from the candidate oncogenes.

However, there were instances where the NTCs failed QC. These genotypes may have been caused by primer dimers in the reactions, or as a result of poor desalting during the cleaning step. This is of great concern because it suggests that a genotype could be attained for a low concentration DNA sample, which would, otherwise, have failed. There was no correlation between failed NTCs and the assigned genotypes. The failed NTCs occurred with some, not all SNPs, which suggests the locus of the SNP may also affect whether an NTC fails.

There were also instances on both iPLEX and iPLEX Gold, when common SNPs, with minor allele frequencies >0.05, appeared to be monomorphic. On iPLEX Gold, the rs2271695, which has a MAF=5.8% in Caucasian population, only had one cluster. The rs17695623 polymorphism, MAF=0.125, also appeared to be monomorphic, according to the platform. A TaqMan assay of this SNP produced 3 clusters. A possible explanation for this may be that there was insufficient differences between the masses of the extend primers for the alleles of the polymorphism. Poor de-salting may also explain the lack of mass separation.

Despite the common SNPs, which are classed as monomorphic on the iPLEX platform, and the failed NTCs, acceptable call rates and concordances were found with the vast majority of SNPs. The quality of the genotyping with iPLEX Gold was not as good as iPLEX. The intensity of the allele peaks were not as high as those from the iPLEX reactions, and there were worrying levels of discordances and poor call rates were found with tSNPs from the MMCT-18 candidate genes with iPLEX Gold. However, these may have been due to a sub-optimal run.

In spite of acceptable call rates, discordance rates greater than 2% were observed between duplicate genomic DNA samples on the iPLEX Gold platform for two tSNPs, rs523104 and rs7650365. TaqMan assays of these SNPs were manufactured, and used to genotype MALOVA, SERACH and UKOPS population series. When the genotypes from iPLEX Gold and TaqMan were compared, discordance rates of 6.1% (241/3,924 calls) for rs523104, and 5.8% (256/4,417 calls) for rs7650365 were found. The genotypes from TaqMan were reproducible, however, do to high costs, the iPLEX experiments could not be repeated. It is likely that the high proportion of

"aggressive calling" caused by poor cluster quality may have been responsible for some of the discordances identified with the iPLEX Gold genotypes for the two SNPs.

Genotyping data for some of the SNPs genotyped on iPLEX Gold was also available from the Illumina Infinium 610K array. The Infinium 610K array platform was used to genotype the SNPs analysed for the ovarian cancer genome-wide association studies (Song *et al.* 2009a) Comparisons were made between genotyping from iPLEX Gold, TaqMan and Infinium SNPs for the MMCT-18 candidate SNPs, where possible.

Genotype data was available for rs523104 and rs7650365 was available from the Infinium platform, as well iPLEX Gold and TaqMan. The genotypes were compared. Of the samples genotyped on all three platforms, the discordance rates are summarised below:

| **Comparison** | **Discordances** | |
|---|---|---|
| | **rs523104** | **rs7650365** |
| Infinium vs iPLEX Gold | 58/1092 (5.3%) | 88/1231 (7.1%) |
| Infinium vs TaqMan | 24/1275 (1.9%) | 15/1262 (1.9%) |
| iPLEX Gold vs TaqMan | 64/1063 (6%) | 75/1169 (6%) |

There were discordances between the genotypes from iPLEX, Infinium and TaqMan. However, the lowest discordance rates were between genotypes from Infinium and TaqMan for both rs523104 and rs7650365. Despite these discordances between Infinium and TaqMan, the rates were just under 2% for both SNPs. Although the call rate for UKOPS samples with rs7650365 on iPLEX Gold was 94%, and there were no discordances between the 47 successfully genotyped pairs of duplicate genomic DNA, discordance rates of >2% were found when the genotypes of

Infinium and iPLEX Gold, and Infinium and Taqman were compared, see Table 5.15.

The distributions of the genotypes are shown in Table 5.14 and Table 5.15, for rs523104 and rs7650365, respectively. The discordances of rs523104 are predominantly in UKOPS samples, and rs7650365 in SEARCH samples. Interestingly, 22 (55%) of the discordances between genotypes from iPLEX Gold and Infinium were common homozygotes according to the Infinium genotypes of SEARCH samples, and heterozygous by iPLEX Gold. A similar distribution was found between the rs523104 genotypes by iPLEX Gold and TaqMan, with 59% of the discordances being heterozygous for the later platform and common homozygous for the Taqman platform; Table 5.14.

Disproportional distributions of the discordances were also found between genotypes of rs7650365 from different genotyping platforms. 97% of the discordances consisted of rare homozygotes from iPLEX and heterozygotes from the Infinium or Taqman platforms.

Additional discordances were found when data from iPLEX Gold and Infinium, for other variants were compared. The discordances ranged from 0 to 2.4%, when SEARCH and UKOPS genotypes were combined, see Table 5.16. Appendix IX-K shows the discordance rates for SEARCH and UKOPS, separately for the SNPs listed in Table 5.16. It can be seen in Appendix IX-K that most of the discordances were with UKOPS DNA samples. Furthermore, discordances (3% of 1334 genotypes) were also found between TaqMan and Infinium genotypes for rs2894111.

Table 5.14: Distribution of discordances of rs523104 (Infinium vs iPLEX Gold vs Taqman)

| Call | | SEARCH | UKOPS |
|---|---|---|---|
| **Infinium** | **iPLEX** | **No. Discord (%)*** | **No. Discord (%)*** |
| **GG** | **CG** | **22 (55%)** | 11 (27.5%) |
| GG | CC | 1 (2.5%) | 2 (5%) |
| CG | GG | 1 (2.5%) | 8 (20%) |
| **CG** | **CC** | **15 (37.5%)** | 10 (25%) |
| CC | GG | 0 (0%) | 2 (5%) |
| CC | CG | 1 (2.5%) | 7 (17.5%) |
| **Total** | | **40/638 (6.3%)** | **40/454 (8.8%)** |
| | | | |
| **Infinium** | **TaqMan** | | |
| GG | CG | 1 (16.7%) | 4 (22.2%) |
| GG | CC | 0 (0%) | 3 (16.7%) |
| CG | GG | 1 (16.7%) | 3 (16.7%) |
| **CG** | **CC** | **3 (50%)** | 4 (22.2%) |
| CC | GG | 0 (0%) | 3 (16.7%) |
| CC | CG | 1 (16.7%) | 1 (5.6%) |
| **Total** | | **6/813 (0.74%)** | **18/462 (3.9%)** |
| | | | |
| **iPLEX** | **TaqMan** | | |
| GG | CG | 1 (2.9%) | 5 (16.7%) |
| GG | CC | 0 (0%) | 1 (3.3%) |
| **CG** | **GG** | **20 (58.8%)** | **7 (23.3%)** |
| **CG** | **CC** | 0 (0%) | **9 (30%)** |
| CC | GG | 1 (2.9%) | 1 (3.3%) |
| **CC** | **CG** | **12 (35.3%)** | **7 (23.3%)** |
| **Total** | | **34/620 (5.5%)** | **30/443 (9.3%)** |

Discord: * % of discordances; iPLEX- iPLEX Gold

Table 5.15: Distribution of discordances of rs7650365 (Infinium vs iPLEX Gold vs Taqman)

| Call | | SEARCH | UKOPS |
|---|---|---|---|
| **Infinium** | **iPLEX** | **No. Discord (%)*** | **No. Discord (%)*** |
| AA | AG | 2 (2.8%) | 2 (11.8%) |
| AA | GG | 0 (0%) | 3 (17.7%) |
| **AG** | **AA** | 0 (0%) | **6 (35.3%)** |
| **AG** | **GG** | **69 (97.2%)** | 3 (22.2%) |
| GG | AA | 0 (0%) | 2 (17.7%) |
| GG | AG | 0 (0%) | 1 (5.9%) |
| **Total** | | **71/765 (9.23%)** | **17/466 (3.6%)** |
| | | | |
| **Infinium** | **TaqMan** | | |
| AA | AG | 1 (33.3%) | 1 (8.3%) |
| AA | GG | 0 (0%) | 3 (25%) |
| AG | AA | 1 (33.3%) | 4 (33.3%) |
| AG | GG | 1 (33.3%) | 1 (8.3%) |
| GG | AA | 0 (0%) | 2 (16.7%) |
| GG | AG | 0 (0%) | 1 (8.3%) |
| **Total** | | **3/810 (0.37%)** | **12/452 (2.7%)** |
| | | | |
| **iPLEX** | **TaqMan** | | |
| AA | AG | 0 (0%) | 2 (33.3%) |
| AA | GG | 0 (0%) | 0 (0%) |
| AG | AA | 2 (2.9%) | 2 (33.3%) |
| AG | GG | 0 (0%) | 0 (0%) |
| GG | AA | 0 (0%) | 0 (0%) |
| **GG** | **AG** | **67 (97.1%)** | 2 (33.3%) |
| **Total** | | **69/742 (9.3%)** | **6/427 (1.4%)** |

Discord: *% of discordances; iPLEX- iPLEX Gold

The quality of the iPLEX Gold genotyping was investigated further by comparing the genomic DNA genotypes of *BRCA1* rs799917 from sequencing of 467 samples and the corresponding iPLEX Gold results. Of the 442 samples successfully genotyped with sequencing and iPLEX Gold, there were 3 discordances (0.68%). The fact that some NTCs were automatically assigned genotypes with high confidence on the iPLEX and iPLEX gold platforms was of particular concern because samples which may have otherwise failed genotyping may appear to have

been successfully genotyped. The randomness of the genotypes of the failed NTCs also suggests that these artefacts of the platform may be locus or mass dependent.

Table 5.16: Discordances between iPLEX Gold and Infinium

| SNP | Total no. of genotypes | No. of discordances | Discordance % |
|---|---|---|---|
| rs6788750 | 962 | 19 | 2.0 |
| rs7650365 | 1263 | 16 | 1.3 |
| rs2280201 | 1292 | 7 | 0.5 |
| rs2394644 | 1302 | 8 | 0.6 |
| rs3181175 | 834 | 4 | 0.5 |
| rs3783194 | 1321 | 10 | 0.8 |
| rs3923086 | 1291 | 23 | 1.8 |
| rs793477 | 1285 | 21 | 1.6 |
| rs12494994 | 829 | 1 | 0.1 |
| rs9860614 | 1321 | 15 | 1.1 |
| rs10999147 | 1322 | 6 | 0.5 |
| rs3181328 | 1284 | 8 | 0.6 |
| rs2282657 | 1321 | 15 | 1.1 |
| rs7189819 | 1286 | 31 | 2.4 |
| rs4541111 | 1274 | 20 | 1.6 |
| rs4791171 | 828 | 1 | 0.1 |

A possible explanation of the discordances with the iPLEX Gold genotyping platform could be due to inadequate desalting of the extend products (single base extension products). No two alleles for an iPLEX assay are within 15Daltons (Da) of each other. The inadequate desalting of samples can cause the formation of sodium and potassium adducts, which are 22Da and 38Da, respectively. The presence of adducts in the nano-dispensed single extension products may make accurate heterozygote allele discrimination, particularly for A/C (24Da) and C/G (40Da) SNPs, difficult. Adducts usually have smaller peak areas than allele peaks. However, it is likely that some of the peaks of the adducts were assigned genotypes because of low peaks of the sample genotypes. Evaluation of the SEARCH

genotyping plates showed that desalting of the single base extensions were suboptimal.

The comparison of TaqMan vs iPLEX Gold vs Infinium genotyping platforms showed that there may be discordances in genotypes from different genotyping platforms of the same samples and assays. These discordances were sometimes observed even when the genotyping data passed quality control criteria. This worrying issue should be taken into consideration when different genotyping platforms are used for a study. The OCAC minimise the problem by ensuring that the same CEPH samples are genotyped for the same SNP(s) at all the genotyping centres and the genotype results are compared with each other and the results from HapMap.

## 5.8: Summary

Ninety-five DNA samples were amplified with GenomePlex, GenomiPhi, PEP and REPLI-g. The resulting amplified products were genotyped on TaqMan, and SNP multiplex genotyping platforms (iPLEX, OpenArray and SNPlex). The DNA amplification procedures were simple for all of the methods. Of the WGA methods, REPLI-g produced the greatest amount of amplified DNA. The iPLEX was the most automatable and straight-forward SNP multiplex genotyping platform, compared with OpenArray and SNPlex. The iPLEX platform also produced the most concordant results when comparisons were made between the genotypes of the amplified DNA and their corresponding genomic DNA.

Although there was ~1,800 fold increase of the amount of DNA with the REPLI-g method, the products also resulted in the highest rates of discordances on iPLEX, SNPlex and OpenArray multiplex SNP genotyping platforms, as well as on TaqMan. It is possible that unequal and over- amplification of the DNA was responsible for some, if not all of these discordances. Despite excellent genotype call rates and concordance on TaqMan, PEP-amplified DNA performed badly on all of the multiplex genotyping platforms tested. GenomePlex and GenomiPhi –amplified DNA gave the best results, in terms of SNP pass rates, genotype call rates and concordance with non-amplified DNA. Disproportionate amounts of miscall discordances were found with OpenArray and SNPlex platforms. These discordances cannot be explained by unequal amplification of the DNA alone, and they suggest problems with the multiplexing platforms. GenomePlex and GenomiPhi genotyped on iPLEX was the best combination of WGA method and SNP multiplex platform identified with the evaluations.

# Chapter 6: Discussion and conclusions

Although ovarian cancer is relatively rare, globally, it is the seventh most common cause of cancer death amongst women, with ~125,000 deaths a year worldwide. This is because the disease is usually diagnosed in the advanced stages, when the chances of survival are drastically reduced. Despite the poor survival rate, the aetiology of ovarian cancer is still poorly understood and the known genetic causes are responsible for approximately 10% of all cases. Several moderate predisposition genes, and multiple low risk (low penetrance) genes may account for some of the remaining cases which are not explained by the known susceptibility genes. The work presented in this thesis aimed to investigate this hypothesis.

This thesis reports the results from investigations of candidate genes which may affect ovarian cancer susceptibility with a case-control association study design; and the influence of these candidate genes on survival of ovarian cancer patients with multivariate Cox regression survival analysis. The performance of possible solutions to the issues of limited amounts of DNA from study participants and the increasing numbers of SNPs genotyped for association studies, respectively, were also investigated.

## The effect of oncogene common germline SNP variants and haplotypes on ovarian cancer susceptibility

The activation of proto-oncogenes through somatic mutation is a common feature of cancer. These mutations may result in un-regulated proliferation of cells, leading to

a neoplastic phenotype (Rhim 1988; Croce 2008). The activation of different oncogenes have been demonstrated to cause the development of different types of cancer. For example, mutations in the *MYC* gene are linked to the development of chondrosarcoma and osteosarcoma (Castresana *et al.* 1992; Ladanyi *et al.* 1993).

Oncogenes, such as *BRAF*, *ERBB2*, *KRAS* and *PIK3CA*, among others are crucial in the development of malignancies, and have been shown to be mutated in ovarian cancer (Shayesteh *et al.* 1999; Hellstrom *et al.* 2001; Gemignani *et al.* 2003; Sieben *et al.* 2004).

The alteration of the oncogenes in ovarian cancer leads to the question of whether germline common polymorphisms or combinations of alleles (haplotypes) of these oncogenes may predispose some women to developing ovarian cancer and/or affect a sufferer's chances of survival. In order to attempt to answer this question, thirty-four tagging SNPs (tSNPs) from *BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA* were genotyped in five different population-based case-control series. Logistic regression was used to evaluate associations between the tSNPs and haplotypes of the candidate oncogenes and the risk of developing ovarian cancer, or survival from the disease.

When all cases were combined, and ovarian cancer was treated as a single disease, there was evidence of association between a tSNP of *NMI*, rs11683487, and risk of ovarian cancer. The rare allele of the polymorphism, which has a minor allele frequency (MAF) of 0.46, was associated with a reduced risk of the disease. Even though the association between *NMI* rs11683487 and risk of ovarian cancer was not validated with an additional 1,097 cases and 1,712 unaffected controls, the

association was marginally significant when the data from stages 1 and 2 of genotyping were combined to increase the statistical power of the study ($P_{dominant}$=0.0419). The association remained when analysis was restricted to the serous and mucinous histological subtypes of ovarian cancer. The rs11683487 variant tags 8 other SNPs with $r^2 \geq 0.8$. One of these, rs1048135, is a non-synonymous coding SNP. The rare, G, allele of rs1048135 codes for a leucine instead of serine. The PMut (http://mmb2.pcb.ub.es:8080/PMut/ ) (Ferrer-Costa *et al.* 2005) and SIFT (http://sift.jcvi.org/ ) (Ng and Henikoff 2001) programmes classified the rare variant of this SNP as having "pathological significance", and "damaging", respectively. Another bioinformatics programme, PupaSuite (http://pupasuite.bioinfo.cipf.es/) (Conde *et al.* 2006; Reumers *et al.* 2008), also suggested that the rare allele of rs1048135 may disrupt the binding of exonic splicing enhancers. This disruption may result in alternative splicing of the gene. PupaSuite also predicted that the rs11683487 variant and another of the tagged polymorphisms, rs11730, may influence the regulation of transcription and translation, and that rs11730 may affect exon splicing.

The *NMI* gene, which was erroneously chosen as a candidate due to its interaction with *MYC*, may be important in ovarian cancer development. *NMI* interacts with the *NMYC*, *MYC*, *MAX* and *FOS* oncogenes and has been shown to form a complex with *MYC* and *BRCA1* (Li, H. *et al.* 2002). This complex has been demonstrated to inhibit the MYC-induced human telomerase reverse transcriptase gene (hTERT) promoter activity in breast cancer (Li, H. *et al.* 2002). The formation of this NMI-MYC-BRCA1 complex suggests that *NMI* may be involved in breast and ovarian cancers.

Ovarian cancer is a very heterogeneous disease and it has been suggested that the different histological subtypes are in fact different diseases of the ovary and should be treated as such. There is evidence in support of this suggestion: certain genetic alterations are found predominantly in particular histological subtypes of ovarian cancer (Christie and Oehler 2006). Of note is the high proportion of *KRAS* mutations in mucinous ovarian tumours (Cuatrecasas *et al.* 1997; Gemignani *et al.* 2003). There is also a correlation between germline mutations of *BRCA1* and *BRCA2* with the susceptibility of the serous histology of ovarian cancer (Lakhani *et al*. 2004).

The numbers of ovarian cancer cases of the 4 major histological subtypes, particularly that of endometrioid, mucinous and clear cell subtypes were small and there was insufficient statistical power to accurately detect associations. However, due to the compelling evidence for the heterogeneity of ovarian cancer, and the associations of some genes with specific histological subtypes, it was worth evaluating the effect of the candidate oncogenes on the risk of the 4 major histological subtypes of ovarian cancer. The identification of potential associations could be validated by consortia, such as the Ovarian Cancer Association Consortium (OCAC).

Although there was no evidence of association between the polymorphisms of *BRAF* and predisposition to ovarian cancer when all cases were considered, statistically significant associations were detected when the analysis was restricted to the mucinous histological subtype. Three tSNPs of *BRAF* (rs10487888, rs1267622 and

rs1769623) were associated with susceptibility of the mucinous subtype (P-trend < 0.05). These tSNPs were intronic polymorphisms, which tagged other intronic variants. To date, these associated SNPs and the SNPs they tag are not known (or predicted) to be functional. When PupaSuite PupaSNP (http://pupasuite.bioinfo.cipf.es/) was used to evaluate the predicted functions of these SNPs, the sequences of a *BRAF* variant, rs9640168, which correlated with rs10487888 ($r^2$=0.934), suggested that it was located in putative triplex forming sequences. Triple helices are long sequences containing only purines or only pyrimidines in a given strand that have the potential to form additional hydrogen bonds with functional groups of the major groove of a DNA double helix, resulting in a triple helical structure. Triple helices may cause replication blocking, subsequently leading to DNA recombination and mutation (Guntaka *et al.* 2003; Patel *et al.* 2004). Many of the other polymorphisms are conserved in mice.

There was evidence of an association between a haplotype, h10010000, of *BRAF* and reduced risk of clear cell ovarian cancer. Furthermore *BRAF* haplotypes were globally associated with predisposition to ovarian cancer (P=0.005).

Associations were also identified between *KRAS* and predisposition to the mucinous histological subtype of ovarian cancer. rs6487464 and rs10842514 of *KRAS* were associated with susceptibility of the mucinous subtype (P-trend < 0.05). Although the associations were found with a limited number of samples, they are of particular interest because *KRAS* mutations are found in 34% of mucinous ovarian adenomas and carcinomas (COSMIC, http://www.sanger.ac.uk/genetics/CGP/cosmic/), and the mutations have been shown to be early events in ovarian cancer development.

Haplotypes (h000010 and h00000) *KRAS* block 2 were associated with susceptibility of mucinous ovarian cancer. These haplotypes comprised of the SNPs which were associated with risk.

BRAF and KRAS are components of the mitogen-activated protein kinase (MAPK) pathway. The MAPK pathway transmits signals for processes such as cell proliferation and cell survival from the cytoplasm to the nucleus (Hingorani *et al*. 2003). The pathway is activated by growth stimulating signals (Pohl *et al*. 2005), and mutations in *BRAF* or *KRAS* lead to the continuous activation of the MAPK. The activation of the MAPK pathway activates downstream cellular targets, including both cellular and nuclear proteins (Pohl *et al*. 2005). It has been demonstrated that the inhibition of the mitogen/extracellular signal-reguated kinase (MEK) pathway, a downstream effector of the MAPK pathway in cell lines with *BRAF* or *KRAS* mutations, results in the suppression of cell growth and promotion of apoptosis (Hingorani *et al*. 2003). It is conceivable then, that a functional germline variant in either of these genes could influence a multitude of downstream targets that may affect the biological and clinical characteristics of ovarian cancers.

The rare allele of a variant of *ERBB2*, rs1801200, with MAF=0.22, was associated with increased risk of endometrioid ovarian cancer. Another group has published an association between the rare allele of rs1801200 and an increased risk of endometrioid ovarian cancer (Pinto *et al.* 2005). This independent validation of the result found with this work is of great importance and it would be of interest to establish if the amalgamation of the data would increase the statistical significance of the association. This polymorphism, also known as I655V, is a non-synonymous

coding, and the rare allele, "G", codes for valine, instead of isoleucine. The amino acid change was predicted to be tolerated with a score of 0.75 by SIFT (http://sift.jcvi.org/www/SIFT_seq_submit2.html). rs1801200 is also conserved in mice and is predicted to enhance exonic splicing. The coding variant is located in the region of the gene which is involved in the dimerisation and activation of the ERBB2 receptor. Fleishman's group demonstrated that the rare allele of the variant destabilises the active dimmer formation (Fleishman *et al.* 2002). The rs1801200 polymorphism has also been associated with the risk of breast cancer (Montgomery *et al.* 2005; Rutter *et al.* 2003). However other studies have not been able to replicate this result (Benusiglio *et al.* 2005).

Two haplotypes of *ERBB2*, h110 and h001, were associated with increased risk of ovarian cancer. The fact that the haplotypes contain the opposite allele at each SNP loci is surprising. There was nothing from HapMap genotyping data to suggest that these putative susceptibility haplotypes shared an untagged common variant, however, it is feasible that an unknown or rare polymorphism tags both haplotypes. There was evidence suggesting that the global test of haplotype effect was also significant for the *ERBB2* gene ($P\text{-}_{global}=0.034$).

*ERBB2* is involved in cell proliferation and cell differentiation (Wu, Y. *et al.* 2004), and is over-expressed in approximately 40% of ovarian tumours (Hogdall *et al.* 2003). This suggests that the over-expression of *ERBB2* leads to tumour growth advantage (Hellstrom *et al.* 2001). Although the association identified with the risk of ovarian cancer and this gene was marginal, and corrections for multiple testing would render the association insignificant, it is likely that the amplification of the

gene is more important in predicting disease risk, rather than a SNP, which would not indicate the amplification of the gene. Although no statistically significant associations were found between *ERBB2* and survival from ovarian cancer with this project, over-expression of the oncogene has been correlated with poor survival of MALOVA samples in another study (Hogdall *et al*. 2003). An association has also been previously published between the rare allele of the polymorphism which was associated with increased risk of endometrioid ovarian cancer, rs1801200, and poor survival of cases (Pinto *et al*. 2005). However, this association was not replicated with this study.

A borderline association was also found between the rare allele of *PIK3CA* rs2865084 and decreased risk of endometrioid ovarian cancer. This SNP is upstream of the gene, and is predicted to generate a new transcription factor binding site. A new transcription binding site has the potential of affecting the transcription of the gene.

**The effect of "functional" common germline SNP variants and haplotypes on ovarian cancer susceptibility**

Although there has been some success in the identification of ovarian cancer predisposition variants through the candidate pathway/candidate gene approach, it is clear that other ways of identifying genes which affect ovarian cancer susceptibility and survival are needed. To this effect, a functional approach (micro-cell mediated chromosome transfer) was used to identify differentially expressed genes, which may be involved in the development of ovarian neoplasm, from an *in vitro* model of ovarian cancer suppression. Nine candidate genes with described functions were

selected from differential expression data of the parental ovarian cancer cell lines and their corresponding hybrid clones (which had incorporated normal human chromosome 18). The associations between 63 tSNPs from nine candidate differentially expressed "functional" genes (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) and susceptibility of invasive epithelial ovarian cancer were evaluated.

An association was found between the rare allele of a variant of *CASP5*, rs518604, and predisposition to ovarian cancer. This association became stronger when the analysis was restricted to the serous histological subtype. However, the results could not be validated with additional samples because TaqMan assays for this SNP and those it tags, could not be successfully manufactured. Associations were also found between the risk of ovarian cancer and haplotypes of *CASP5*. The associations with the *CASP5* haplotypes were supported by the individual SNP finding. The global test for association of haplotype effect on ovarian cancer predisposition was highly significant for *CASP5* ($8.43 \times 10^{-6}$). This was the only association which reached the level of significance suggested for case-control genetic association studies ($P<10^{-5}$) (Thomas *et al*. 2005). The involvement of CASP5 in apoptosis and inflammation makes it a plausible ovarian cancer susceptibility gene, despite the small numbers of samples analysed.

The rare allele of *RUVBL1* rs13063604 tSNP correlated with an increased risk of all histological subtypes of ovarian cancer. This association became stronger when analysis was restricted to the serous subtype (P=0.002). The rare allele of another *RUVBL1* SNP, rs7650365, had the opposite effect on serous ovarian cancer

(P=0.009). Up to additional 2,636 cases and 6,164 controls were genotyped with assays of rs13063604 and rs7650365, for a second genotyping stage of the study, to ascertain if the significant finding was reproducible. The associations were not validated with the stage 2 samples alone, however when the data from the two stages were combined, the association between rs13063604 and ovarian cancer remained. This polymorphism tags 9 other variants with $r^2 \geq 0.8$. Two of them, rs1057220 and rs1057156, which are tagged with $r^2=1$, are located in the 3' untranslated region of the gene and they are predicted to be exonic splicing enhancers. Globally, *RUVBL1* haplotypes also had a significant effect on the risk of ovarian cancer (P=0.0016).

*RUVBL1* (also known as *pontin*), was down-regulated in the reverted hybrids of both cancer cell lines, and thus up-regulated in the parental neoplastic ovarian cancer cell lines. Although, to date, there are few publications reporting the over-expression of *RUVBL1* in malignancies, data from the Oncomine database (http://www.oncomine.org) shows over-expression of the gene in breast, colon, bladder, liver and other malignancies (Dehan *et al.* 2007; Lauscher *et al.* 2007; Rousseau *et al.* 2007; Huber *et al.* 2008; Haurie *et al.* 2009). If the associations from this research are true, then they could be attributed to its interactions with other genes which are implicated in cancer. *RUVBL1* interacts with the transcriptional activation domain of MYC and also β-catenin (Bauer *et al.* 1998; Wood *et al.* 2000). *RUVBL1* is necessary for Tip60 activity, which is involved in DNA damage repair (Jha *et al*. 2008). Venteicher and colleagues demonstrated that *RUVBL1* interacts with the human telomerase reverse transcriptase (hTERT) and is involved in the assembly and function of the telomerase complex; and the abrogation of *RUVBL1* expression has been shown to induce premature senescence (Venteicher *et al*. 2008).

Furthermore, *RUVBL1* is a component of the NuA4 histone acetyltransferase complex, which may activate the transcription of cell growth, DNA repair and apoptosis genes (Jha *et al*. 2008).

The rare alleles of three *FILIP1L* polymorphisms, rs793446, rs17338680 and rs12494994, were associated with increased risk of the endometrioid histological subtype. However, no functions were identified for these SNPs or those they tag, other than the conservation of many of the SNP sequences in mice.

*FILIPL1*, which is located on chromosome 3q12.1, was up-regulated in the hybrids of TOV112D, and thus down-regulated in the malignant parental cell line. The *FILIP1L* gene has also been reported to be down-regulated in ovarian cancer by another group (Mok *et al*. 1994), which support the results from the gene expression data of the MMCT-18 TOV112D hybrid clones. The *FILIP1L* gene is also conserved in yeast, and although little is known about the gene, it has been shown to be implicated in a variety of cellular functions (Hwang and Murray 1997). It has been shown that although *FILIP1L* was present in normal ovarian surface epithelial cells, it was predominantly absent from the cell lines tested (Mok *et al*. 1994). The gene is part of a subunit, Doc1p/Apc10, which is involved in substrate recognition by anaphase-promoting co-activator complexes (Passmore *et al*. 2003). Tandle's group showed that *FILIP1L* may be involved in mediating some of the effects of the pro-inflammatory cytokine endothelial monocyte activating polypeptide-II (Tandle *et al*. 2005). There is also evidence suggesting that the FILIP1L protein is down-regulated in human prostate cancer cell lines and it may be involved in the regulation of senescence (Schwarze *et al*. 2002).

In 2009, the results from the ovarian cancer genome wide association study (GWAS) were published by the Ovarian Cancer Association Consortium. Virtually all of the UKOPS and SEARCH cases used in these projects were also analysed in the GWAS, along the cases from 5 other UK-based ovarian cancer studies, totalling 1,890 cases and 2,353 controls. 14 tSNPs genotyped in the oncogene study were also analysed in the GWAS, however, none of them were associated with ovarian cancer risk at the genome-wide significance level ($P \leq 5x10^{-8}$).

Despite the lack of association of the candidate oncogenes at the genome-wide significance level, some similarities were identified between the results of the GWAS and candidate gene studies. Of these, *NMI* rs11683487, which was nominally associated with ovarian cancer risk, had relatively similar odds ratio and P-values in the 2 approaches (candidate gene result: HetOR = 0.80 (0.69-0.93) HomOR= 0.87 (0.71-1.02), P= 0.0379; GWAS per rare allele: OR=0.86, P=0.027). However, this is indicative of no association for this variant when both experiments are taken into consideration.

From the MMCT-18 study, no similarities were identified between the results of the nominally significant associations from this project and those from the first stage GWAS data when all histological subtypes were analysed. However, when the analysis was restricted to the serous histological subtype, similarities were found with *CASP5* rs518604 (candidate gene result: HetOR= 1.36 (0.98-1.88), HomOR= 1.45 (0.99-2.11), P=0.0313; GWAS per rare allele: OR=1.12, P=0.047) and rs523104 (candidate gene result: HetOR= 0.86 (0.65-1.16), HomOR= 0.80 (0.55-

1.15), P= 0.1294; GWAS per rare allele OR=1.12, P=0.044), and *FILIP1L* rs12494994 (candidate gene result: HetOR= 0.95 (0.78-1.17, HomOR = 0.52 (0.27-0.99), P=0.097; GWAS per rare allele OR=0.84, P=0.02). *CASP5* rs518604 is currently being investigated further with additional samples. Although *FILIP1L* rs12494994 was not associated with the serous ovarian cancer susceptibility in the candidate gene approach, the homozygous odds ratio did not cross 1. Incidentally, the rare allele of *FILIP1L* rs12494994 was associated with an increased risk of the endometrioid histological subtype (P=0.0024) with the candidate gene data.

The ovarian cancer GWAS was a 3-staged experimental design which further investigated the most statistically significant associations from the previous stage, using samples from the UK, Australia, USA, Denmark, Poland, Germany and Canada. It is likely that the discrepancies in the associations identified with the candidate gene approach and GWAS could be due to the samples used; SEARCH and UKOPS are the only studies common to both approaches. Although none of the SNPs analysed in this project reached genome-wide significance, only a limited number of the SNPs evaluated with the candidate gene approach were also analysed with the GWAS. However, the results from both the candidate gene and GWAS suggest that there is strong no evidence for association between the tSNPs analysed in this project and the risk of ovarian cancer. The 12 strongest associations from the GWAS, with $P<10^{-8}$, were all located on chromosome 9p22.2 (Song *et al*. 2009). Eight of these SNPs were in *BCN2*, a DNA-binding zinc finger protein, and the remaining SNPs were within 45kb upstream of the gene. *BCN2* would not have been an obvious candidate for an ovarian cancer association study because there is very little evidence to suggest its involvement in the development of the disease.

However, this finding strongly suggests that the germline common variants of genes which have been demonstrated to be important in the initiation and development of cancer may not be useful in predicting an individual's disease risk.

rs3814113, the strongest associated SNP from the ovarian cancer GWAS, was 44kb from *BCN2*, the nearest gene, with per rare allele OR of 0.79 (0.75-0.84), $P=2.47x10^{-17}$. The identification of strongly associated SNPs in so-called gene-deserts have also been observed in the GWAS for other diseases, which suggests genes already implicated in the development of disease may not necessarily be the best predictors of disease risk. However, it is possible that rare SNPs and/or copy number variants may also be associated with disease risk. Ovarian cancer is a complex heterogeneous disease which arises through various genetic and genetic factors. Therefore interactions between genes and the environment should, ultimately, be considered when evaluating a woman's risk of the disease.

**Global evaluation of associations with admixture maximum likelihood method**

Numerous statistically significant associations have been identified between candidate genes and ovarian cancer risk and survival over the past decade; however, it is not known if an excess of significant associations were identified compared to the proportion which would be expected by chance (Goodman *et al.* 2001; Song *et al.* 2006; Gayther *et al.* 2007; Harley *et al.* 2008; Kelemen *et al.* 2008; Ramus *et al.* 2008). The admixture maximum likelihood (AML) approach was used to evaluate the overall evidence of excess of positive associations with SNP genotyping data for 340 SNPs in 94 candidate genes or chromosomal regions. The polymorphisms were

genotyped in approximately 1,500 cases and 3,100 unaffected controls from three population-based ovarian cancer case-control series.

Twenty-two out of the 340 SNPs analysed with the AML test were significantly associated (P-trend<0.05) with ovarian cancer risk. This number was reduced to 18 polymorphisms when the results were adjusted for population stratification. The 3 most significant SNPs had adjusted (for population stratification) P-values <0.003. These SNPs were rs2107425 (adjusted P=0.0019) from chromosome 11p15.5, rs9322336 (adjusted P=0.0021) from the oestrogen receptor (*ESR1*) gene and rs3817198 (adjusted P=0.0026) from the lymphocyte-specific protein 1 (*LSP1*) gene.

The most strongly associated SNP was from the BCAC group of SNPs which is located on chromosome 11p15.5, in a region which does not contain genes or open reading frames. The 11p15.5 region, despite the lack of genes, loss of heterogeneity of this locus has been observed, and it is said to contain tumour suppressive properties malignancies and is associates with breast, lung, bladder and stomach cancers (Viel *et al.* 1992; Gudmundsson *et al.* 1995; Shaw and Knowles 1995; Baffa *et al.* 1996; Bepler *et al.* 1998; Karnik *et al.* 1998; Xu *et al.* 2001). rs2107425 was chosen as a candidate SNP due to its strong association with breast cancer susceptibility. This association was identified from the breast cancer genome-wide association study (Easton *et al*. 2007). This SNP tags another, rs2251375, with $r^2=1$, which is located in a region that is conserved in mice. *ESR1* rs9322336, which is in the second intron of the gene, is not known to tag any other common SNP. The *ESR1* gene encodes a ligand activated transcription factor, which is able to bind hormones and DNA. The gene is associated with ovarian and breast cancers (Imura

*et al.* 2006; Dunning *et al.* 2009). The third most significant SNP from the analysis, rs3817198 of the *LSP1* gene, was also identified from the breast cancer association genome-wide association study and chosen because of its strong association with the risk of breast cancer (Easton *et al.* 2007).

It is likely that the associations would no longer be statistically significant after adjustments for multiple corrections. It may be more appropriate to assess the experiment-wise significance of either subsets of polymorphisms investigated, or the totality of all SNPs analysed. The AML test was used to this effect on subsets of SNPs based on their function, or their genotyping group.

There was no evidence for an overall association between common genetic variation in the 94 candidate genes or regions and risk of ovarian cancer, when the genotyping data was analysed with AML, P-trend=0.068. There was evidence of a statistically significant association between tSNPs identified from the breast cancer genome-wide association study and evidence of an excess of positive associations over the proportion expected by chance from this group of 16 common variants (P=0.0028).

Although the AML did not identify an excess of statistically significant associations, the associations found should not be disregarded. The associations are very modest, and should be treated with caution, however, there is evidence that the SNPs (or the chromosomal regions in which they are located) are associated with the risk of other cancers.

**Effects of germline variants and haplotypes of candidate oncogenes on survival of ovarian cancer cases**

Comparisons between the univariate and multivariate results of all the tSNPs and haplotypes demonstrated the importance of performing the multivariate analysis with adjustments for prognostic factors such as age at diagnosis, tumour histology, grade and stage for all genetic variants being investigated. Prognostic factors that had a statistically significant affect on survival can mask the influences of the SNPs, which can result in types I and II statistical errors. This was demonstrated with the tSNP of *RUVBL1*, rs4857836. The rare allele of the common variant was not significantly associated with survival with the univariate Cox regression analysis (per-rare allele hazard ratio [HR]=0.98 [0.86-1.12], P=0.758), however there was evidence of association between the rare allele of rs4857836 and better survival after adjustments for the prognostic factors (adjusted HR=0.8 [0.67-0.98], P=0.003). Conversely, the opposite was also observed, where the results of the univariate analysis was significant (*KRAS* haplotype block 2 h010000, HR=1.69 (1.21-2.36), P=0.002); but the association was no longer significant after adjustments for prognostic factors (adjusted HR=0.9 (0.6-1.13), P=0.523). There were 17% and 47% differences between the hazard ratios of the univariate and multivariate analysis results for rs4857836 and haplotype h010000, respectively, which is substantial.

These findings suggest, although associations may be identified with the univatiate Cox regression analysis, correction for prognostic factors is required for all the variants analysed in order to ascertain more accurate associations between the genetic variants and survival from the disease. Numerous publications, including work from this thesis, have reported survival analysis results in ovarian and other

malignancies with adjustments for prognostic factors of only the variants found to be associated with survival from the univariate analysis (Mann *et al.* 2008; Quaye *et al.* 2008; Koessler *et al.* 2009; Quaye *et al.* 2009; Udler *et al.* 2009). However, it is likely that some associations were overlooked by not conducting multivariate Cox regression survival analysis of all the polymorphisms analysed.

The multivariate survival analysis of the oncogene variants showed that associations between *BRAF* polymorphisms and ovarian cancer were not restricted to predisposition to ovarian cancer. Correlations were also found between haplotypes and tSNPs of *BRAF* and all-cause survival of patients with ovarian cancer. Contrary to the susceptibility results, the associations were observed when all ovarian cancer cases were combined. Although the rare allele of *BRAF* rs1267622, was associated with a reduced risk of mucinous subtype, it was also associated with poor survival of all ovarian cancer cases combined. These associations suggest that although the common variant may influence the risk of only the mucinous subtype, the tSNP may also be useful for predicting the survival of ovarian cancer patients, regardless of the ovarian tumour histological subtype. There was also evidence of an association between *BRAF* h10010000 and better survival of all cases combined (P=0.014). This haplotype was also associated with a reduction in the rest risk of clear cell ovarian cancer.

Although no statistically significant associations were found between *ERBB2* and survival from ovarian cancer with this project, over-expression of the oncogene has been correlated with poor survival of MALOVA samples in another study (Hogdall *et al*. 2003). An association has also been previously published between the rare

allele of the polymorphism which was associated with increased risk of endometrioid ovarian cancer, rs1801200, and poor survival of cases (Pinto *et al*. 2005). However, this association was not replicated in this study. Subtle differences in the genetics or lifestyle of the Portuguese population used in the Pinto study and the research presented in this thesis may explain the lack of validation of Pinto's group's study. Alternatively, the result from the Pinto paper may have been a false positive; there were a total of 129 ovarian cancer patients included in the Pinto study in comparison with 1,766 cases genotyped for rs1801200 in this study.

**Effects of germline variants and haplotypes of "functional" candidates on survival of ovarian cancer cases**

When the effects of the MMCT-18 common genetic variants on survival of ovarian cancer patients were evaluated, associations were identified with various candidate genes which correlated with suppression of the tumourigenic phenotype. Of note, the rare alleles of two tSNPs of *RBBP8*, rs4474794 and rs9304261, were associated with better survival of all ovarian cancer patients after adjustments for prognostic factors. Incidentally, the rare allele of rs4474794 was also associated with reduced risk of serous ovarian cancer. Haplotypes of *RBBP8* were also associated with the risk of ovarian cancer.

The *RBBP8* protein, which is also known as CtBP-interacting protein (CtIP), has been shown to interact with the retinoblastoma protein and the BRCA protein C-terminal region domains of the *BRCA1* gene and a variety of other proteins which are involved in the regulation of the cell cycle and transcription (Fusco *et al.* 1998; Meloni *et al.* 1999; Li, S. *et al.* 2000; Zheng *et al.* 2000). Chen *et al* have suggested

that a complex containing *RBBP8*, *BRCA1* and *MRN* is cell-cycle dependent and is involved in the activation of homologous recombination double strand DNA repair in the S and $G_2$ phases of the cell cycle (Sartori *et al.* 2007; Chen *et al.* 2008). There is also evidence suggesting that the *RBBP8* protein is resistant to DNA double strand break-inducing agents (Sartori *et al*. 2007). Furthermore, the expression of *RBBP8* has been shown to be elevated in the majority of oestrogen receptor alpha (ER) positive breast cancer cell lines. However, this gene was down-regulated in the TOV112D cell line. Nonetheless, the over-expression of the gene is associated with patient response to neoadjuvant endocrine therapy (Wu, M. *et al.* 2007). Moreover, it was been demonstrated that tamoxifen resistance in breast cancer cells is conferred through the silencing of the *RBBP8* gene (Wu, M. *et al.* 2007).

rs793446 and two other polymorphisms (rs3921767 and rs9864437) of *FILIP1L* were also associated with ovarian cancer survival. However, the rs793446 and rs9864437 were associated with survival from mucinous ovarian cancer. There was no striking functional evidence to explain the observed associations between ovarian cancer susceptibility and survival. However the fact that many of the sequences of the tagged SNPs were conserved in mice suggests that the region containing the gene may be functionally important. Furthermore, it has been demonstrated that when the FILIP1L protein is over-expressed in endothelial cells, there is an increase in apoptosis and inhibition of cell proliferation and migration (Au *et al.* 2002; Passmore *et al.* 2003).

Another polymorphism of *RUVBL1*, rs4857836, which tags 2 other SNPs (rs4857837 and rs7641133), was associated with survival from ovarian cancer;

adjusted (for prognostic factors) HR=0.81 (0.67-0.98), P=0.03. This association became stronger when the analysis was restricted to the serous subtype, adjusted HR= 0.75 (0.59-0.95), P=0.018.

There was a 5-fold and 7-fold increase in the expression of *FILIP1L*, and *RBBP8*, respectively, in TOV112D$^{+18}$ hybrid cell lines, compared to parental cancer cell lines. This suggested that the genes behave like tumour suppressor genes. There was an average 25-fold decrease in the expression of *RUVBL1* in both TOV21G$^{+18}$ and TOV112D$^{+18}$ hybrid cell lines compared with parental cancer cell lines, suggesting the gene behaves like an oncogene. Although the gene expression changes for *FILIP1L*, *RBBP8* and *RUVBL1* are likely to be attributed, either directly or indirectly, to the transfer of chromosome 18 into the hybrid cell lines, *RBBP8* was the only one located on the transferred chromosome. Fluorescence *in situ* hybridization analysis showed that *RBBP8* was on the 10Mb fragment of human chromosome 18 that was transferred into the TOV112D cells, rather than the whole chromosome. The gene expression data has shown that the *RBBP8* gene had the greatest fold increase in expression compared to the other genes in the transferred region (Quaye *et al.* 2009).

The rare allele of another variant of *CASP5*, rs2282657, was associated with better survival of clear cell ovarian cancer patients. The sequence of the SNP indicates that it is located within a splice site of the gene. *CASP5* was up-regulated in the hybrid clones showing suppressed (reverted) neoplastic phenotype, therefore, it was down-regulated in the parental ovarian cancer cell line (TOV21G). CASP5 is involved in apoptosis and inflammation signaling (Eckhart *et al.* 2006), the protein is a

component of the NALP1 inflammasome, and it is involved in the maturation and secretion of interleukin-1β following simulation by lipopolysaccharide, when it is part of this complex (Martinon *et al*. 2002).The expression of CASP5 is also regulated by interferon-γ (Lin *et al.* 2000). Mutations in this gene have been observed in leukaemia, lymphoma and colon cancer (Takeuchi *et al*. 2003). *CASP5* is also of interest because it forms a complex with *MYC* and *MAX* oncogenes and has been demonstrated to cleave MAX, which is important for cell growth, differentiation and apoptosis (Krippner-Heidenreich *et al*. 2001). *CASP5* also appears to be a target gene in the microsatellite mutator pathway for cancer (Offman *et al.* 2005).

**Whole genome DNA amplification and multiplex SNP genotyping platforms**

The final topic of this thesis was whole genome amplification and SNP multiplex genotyping platforms. The concentration of DNA available for research from each patient is limited. The relative rarity and devastating effects of ovarian cancer makes it difficult to recruit ovarian cancer patients for studies. Whole genome amplification (WGA) is a method through which DNA concentrations can be increased. The use of WGA amplified DNA has the potential of increasing the numbers of samples available for research because more DNA from cumulatively increasing numbers of study participants would substantially increase the statistical power of genetic association studies.

The number of common genetic polymorphisms within the human genome and the numbers of these SNPs from candidate genes, which need to be evaluated, make the single SNP reactions labourious and unappealing. This is especially so because of

the availability of SNP multiplex genotyping platforms. Four WGA methods, GenomePlex, GenomiPhi, PEP and REPLI-g were used to amplify 95 samples; and the performance of the WGA products and their performance was evaluated with three SNP multiplex genotyping platforms, iPLEX, SNPlex and OpenArray.

REPLI-g generated the greatest amount of amplified material and PEP, the lowest fold increase in amplified DNA. PEP-amplified DNA performed poorly on all of the SNP multiplex genotyping platforms evaluated. REPLI-g DNA were the only products with average call rates <90% on TaqMan genotyping platform. The call rates for the genomic DNA were generally higher than those of the amplified DNA.

Of the SNP multiplex genotyping platforms tested, iPLEX generated the best quality genotyping results. However, there are issues with failed non-template negative test controls, which seems to be an artefact of the platform. This could result in a sample with low concentration apparently yielding an incorrect genotype. A small proportion of common SNPs genotyped on iPLEX appear to be monomorphic despite MAF>0.05. The reason behind this is still unclear. It is possible that contaminating salt adducts prevent the discrimination of genotypes of the SNP.

Discordances were found between genomic DNA and matching amplified products from all of the WGA methods. However, fewer discordances were identified with the GenomePlex and GenomiPhi amplified DNA. GenomePlex and GenomiPhi appeared to provide the best balance between quantity of amplified DNA and performance on SNP multiplexing platforms.

Other studies have reported similar findings of lower call rates of amplified DNA and discordances between WGA DNA and genomic DNA (Tranah *et al.* 2003; Bergen *et al.* 2005; Berthier-Schaad *et al.* 2007; Moore *et al.* 2007; Cunningham *et al.* 2008; Xing *et al.* 2008).

Although discordances between genotypes of the amplified DNA and non-amplified genomic DNA were observed with every WGA method on all of the SNP multiplex genotyping platforms, the vast majority were found with REPLI-g-amplified DNA. When PEP-amplified DNA is not considered, genotypes of the WGA DNA from the iPLEX platform were the most concordant with those of the corresponding non-amplified genomic DNA. None of the amplification products had an average discordance rates ≥2% on the iPLEX platform, however the average discordance rates ranged from 1% (for GenomiPhi-DNA on SNPlex) to 7% (REPLI-g on OpenArray) for the other 2 multiplex platforms investigated. These results, like many reported in the literature, are misleading because they are not indicative of the discordances for each polymorphism genotyped. Cunningham's group reported a >99% average concordance rate between WGA and their corresponding genomic DNA when they were genotyped on the Illumina GoldenGate BeadArray platform, however, only 1 (0.9%) of the 116 pairs of WGA and genomic DNA was concordant for all the 1,536 SNPs successfully genotyped (Cunningham *et al*. 2008).

Although the majority of discordances, both allele dropout and miscall were found between REPLI-g amplified DNA and genomic DNA, other researchers have not found this. Xing *et al*. report "excellent" (overall 98.7%) concordance between genomic and matching REPLI-g DNA based on genotyping data from the

Affymetrix 250K array platform, however, the genotypes are of 4 individuals, which is too small a number to accurately ascertain the performance of WGA DNA (Xing *et al*. 2008). Another study reported 100% concordance between genomic DNA and the corresponding DNA amplified with DOP-PCR, ligation-mediated PCR and a strand displacement amplification method, for the 10 SNPs genotyped, however, again, only 4 samples were analysed (Lee *et al*. 2008).

Talseth-Palmer *et al*. have demonstrated gains and losses of GenomePlex and GenomiPhi amplified DNA with array comparative genomic hybridisation, compared with genomic DNA. They also report that the discordances appear to be random and are not reproducible (Talseth-Palmer *et al*. 2008). The results from the OpenArray platform, which was the only genotyping platform where the reproducibility of the genotyping was assessed, also suggested that there is a lack reproducibility of the genotyping results in both genomic and amplified DNA on the platform.

**Limitations**

The studies within this project had differing statistical powers to detect associations. Where all ovarian cancer cases were analysed, there was 97% power at the 5% significance level to detect a co-dominant allele with a minor allele frequency of 0.3 that confers an odds ratio of 1.2, and 96% power to detect a dominant allele with a minor allele frequency of 0.1 that confers an odds ratio of 1.3. However, statistical power depends on the sample size, the minor allele frequency, the risks conferred, and the genetic model. Therefore, the statistical power to detect associations when

analysis was restricted to the histological subtypes of ovarian cancer was greatly reduced.

It is possible that some associations were missed because some tSNPs from the candidate genes could not genotyped such as *BRAF* (1 tSNP not genotyped), *KRAS* (2 tSNPs), *PIK3CA* (3 tSNPs), *AIFM2* (1tSNP), *AXIN2* (2 tSNPs), *RGC32* (1 tSNP) and *RUVBL1* (1 tSNP).

The findings reported should be treated with caution because they could be chance findings. The results have not been adjusted for multiple testing, which may diminish the vast majority of the associations found. Unfortunately, associations between germline genetic variants and other clinical features of disease, such as disease recurrence, and response and resistance of chemotherapy could not be assessed in this project. This is because the data for the collections are epidemiological, rather than clinical. Therefore there is no access to the clinical information, other than the ones mentioned.

Genomic controls from a breast cancer study were used to estimate the inflation of the test statistic used to adjust for cryptic population stratification. It is possible that the stratification observed in the breast cancer study was not a true reflection of that from the samples analysed with the admixture maximum likelihood test. Although, a very conservative inflation test of 10% was used to adjust the results, it is nonetheless, possible that the value is an over-estimation, or under-estimation for the ovarian cancer studies. The test statistics from the ovarian cancer GWAS showed that there were marginal increases in the estimated inflation factor ($\lambda_{1000}$=1.026) with

the stage 1 samples which were exclusively Caucasian Britons and $\lambda_{1000}$=1.005 with the European, Australian and North American non-Hispanic Caucasians (Song *et al.* 2009b). Therefore the 1.1 inflation factor used to adjust for cryptic population stratifications in this thesis was likely to be a gross over-estimation.

It is also believed that the use of prevalent samples in the survival analysis may be a weakness of the study. However, Cox regression survival analysis of the follow-up data showed that, although the prevalent samples appeared to have better survival than incident cases, the difference was not statistically significant (P>0.05). Left truncation of the data was also used in the analysis, which controlled for any bias of the hazard ratio estimates which may have arisen. Although, as a result of smaller events (deaths) occurring with prevalent samples, the exclusion of prevalent samples would reduce the overall sample size and number of events. Therefore the inclusion of prevalent samples increases the statistical power to detect associations. The inclusion of prevalent samples may also be considered as a form of adjustment, as they may generate more conservative associations. Azzato *et al*. (2009) have also reported that they did not find significant bias in the hazard ratios of incident and prevalent cases of breast cancer when survival analysis, with left truncation, was conducted on data on clinical stage, histopathological grade and oestrogen receptor.

The analysis of data based on all-cause mortality, rather than mortality from ovarian cancer is another limitation of this study. It is likely that some of the affect participants die or will die from causes other than ovarian cancer. However, this issue it not of great importance as the vast majority of the cases, will sadly die from

ovarian cancer. The small number of patients who die from other causes should be too small to greatly affect the results.

The numbers of histological subtypes of epithelial ovarian cancer, which include serous, endometrioid, mucinous and clear cell, are limited, as was the power to detect association with a reasonable degree of confidence. However, the results from the susceptibility and survival analyses re-affirms the heterogeneity of ovarian cancer aetiology. The results were in concordance with mutational analyses, which have found that mutations of some genes are restricted to particular histological subtypes of ovarian cancer. This is particularly true when the susceptibility and survival results for candidate genes such as *BRAF*, *KRAS* and *FILIP1L* are considered. These results need to be validated with additional samples.

The results from these studies provide proof of principle for the theory that SNPs may influence predisposition and the survival outcome of ovarian cancer. Ultimately highly significant SNPs with strong effects may be used clinically to predict a woman's risk of ovarian cancer, or survival from the disease. However, the vast majority of associations identified by these studies have been limited to marginal significance, which are considerably less than the P-value suggested for candidate gene association studies ($P=10^{-4}$).

The additional limitations of this study and others of ovarian cancer susceptibility are the fact that the sizes of the effects have generally been <2, which suggest that the findings from these studies are unlikely to be translated and implemented to the clinical setting. None-the-less, the genes which have been selected for the

association studies have been plausible candidates and the results have enriched ovarian cancer genetic research.

Moreover, although SNP genotyping data from the HapMap Project is invaluable to SNP and haplotype association studies, the research is ongoing and is not immune from error. HapMap data from Release 20 for the BRAF oncogene suggested that one of the tSNPs genotyped, rs1267622, tagged rs7384384 with r2=1. However, recent data releases from HapMap no longer include rs7384384, and dbSNP gives the error message:

> "This snp_id was merged into rs4726020
>
> refSNP cluster id(rs): *rs7384384* is an invalid snp_id value.
>
> Note that rs# is not contiguous due to user withdraws and merging of clusters."

This message suggests that both rs numbers were assigned to the same SNP. The policy in these instances is to keep the lower rs number and "retire" the higher number. There have also been a few occasions when HapMap has announced errors with its data. For example:

> "2008-02-21: Incorrect position for merged SNPs in rel #23
>
> The position of ~24,500 SNPs was inadvertently entered incorrectly in HapMap release #23 bulk files (genotypes and frequencies). A complete list of affected SNPs can be found here. Errors are being corrected and new genotypes and frequency files will be made available shortly under HapMap release #23a."

and

> "2007-12-12: **Assembly errors in rel#22 phased files**
>
> Files with errors have been removed from public view and will be replaced with correct files. Nonetheless, the files continue to be under scrutiny. An official announcement will be made when these files are officially approved for general use."

These announcements highlight the importance of regularly visiting the HapMap website for such updates, and also conducting validation studies in order to confirm the associations identified.

A new release of HapMap SNP genotying data, Release 21, became available after the completion of the genotyping of the samples with the oncogene probes. The oncogenes selected for the study were re-tagged with the more up-to-date data to establish the efficiency of the tagging based on the tSNPs successfully genotyped. As shown in Appendix Y, although there was data available for more SNPs ("All SNPs" in the appendix table) for all of the genes, with the exception of *ERBB2* (which had a tagging efficiency reduced, from 100% to 80% with the new data) the tagging efficiency of the tSNP from the other genes remained unchanged. Moreover, the overall tagging efficiency of all the genes combined stayed at 94%.

## Future work

The advent of genome-wide association studies has led to the usefulness of the candidate gene approach of genetic association studies being questioned. This is likely to be due to the limited success of the approach. These questions are of particular importance since some of the most significant associations found from the breast and ovarian cancer genome-wide association studies were located at regions

without genes or open reading frames. To date, the only possible explanation for the associations may be that of long-range regulation. Fine-mapping is a technique in which the SNPs tagged by the significant tSNP and other neighbouring SNPs are genotyped in order to elucidate the individual SNP responsible for the association. It is expected that the P-value and effect (odds ratio) would be greatest at the "causal" polymorphism. Similar results can also be obtained through sequencing the region(s) significantly associated with ovarian cancer susceptibility or survival.

It is feasible that the fine-mapping technique could be used to find "causal" SNPs from the positive associations found from these and other studies. However, all the results need to be corrected for multiple testing and validated before such a step is taken. The validation of some of the findings could increase the power to detect associations. It would be of great interest to run the AML method on all genotyping data, restricted by histological subtype, to evaluate the overall evidence of positive associations over the proportion expected by chance.

There were insufficient numbers of samples for investigating gene-gene or gene-environment interactions. However, it would be of great interest to conduct these analyses as the SNPs are unlikely to cause ovarian cancer without interacting with other factors. Therefore these tests should be considered in the future. Stronger, more significant associations may be found if the analyses of these interactions were performed.

## Conclusions

In conclusion, 34 tSNPs of four oncogenes (*BRAF*, *ERBB2*, *KRAS* and *PIK3CA*) and a putative oncogene (*NMI*), and 63 tSNPS from 9 differentially expressed genes (*AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*) from *in vitro* neoplastic suppression experiments were genotyped in invasive ovarian cancer case-control series. Associations were identified between polymorphisms and haplotypes of *NMI*, *CASP5*, and *RUVBL1* and disease risk when ovarian cancer is considered as a single disease. Additional associations were found with many of the other genes when analysis was restricted to the histological subtypes of ovarian cancer. Of note, associations were found between mucinous ovarian cancer susceptibility and survival and haplotypes and variants of *BRAF* and *KRAS*; and risk of endometrioid disease and variants of *FILIP1L*.

Associations were also identified between *RBBP8*, *RUVBL1* and *FILIP1L* and clinical outcome of ovarian cancer patients. Additional associations were found when the survival analyses were restricted to the major histological subtypes of ovarian cancer. Although the results should be treated with caution, they should be further investigated. The identification of strongly associated polymorphisms candidate genes could used for targeted screening of individuals at high risk of ovarian cancer, the prediction of response to therapy or prognosis, and/or more effective treatment.

The genotyping of GenomePlex and GenomiPhi amplified DNA on the iPLEX system was the best combination of WGA method and SNP multiplex genotyping

platform. However, these results should be confirmed with replication of the investigations.

# References

Abubaker, J., Z. Jehan, P. Bavi, M. Sultana, S. Al-Harbi, M. Ibrahim, A. Al-Nuaim, M. Ahmed, T. Amin, M. Al-Fehaily, O. Al-Sanea, F. Al-Dayel, S. Uddin and K. S. Al-Kuraya (2007). "Clinicopathological analysis of papillary thyroid cancer with PIK3CA alterations in a Middle Eastern population." *J Clin Endocrinol Metab*.

Agorastos, T., S. Masouridou, A. F. Lambropoulos, S. Chrisafi, D. Miliaras, K. Pantazis, T. C. Constantinides, A. Kotsis and I. Bontis (2004). "P53 codon 72 polymorphism and correlation with ovarian and endometrial cancer in Greek women." *Eur J Cancer Prev* **13**(4): 277-280.

Aktas, D., I. Guney, M. Alikasifoglu, K. Yuce, E. Tuncbilek and A. Ayhan (2002). "CYP1A1 gene polymorphism and risk of epithelial ovarian neoplasm." *Gynecol Oncol* **86**(2): 124-128.

Au, S. W., X. Leng, J. W. Harper and D. Barford (2002). "Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the Doc1/Apc10 subunit." *J Mol Biol* **316**(4): 955-968.

Auranen, A., A. B. Spurdle, X. Chen, J. Lipscombe, D. M. Purdie, J. L. Hopper, A. Green, C. S. Healey, K. Redman, A. M. Dunning, P. D. Pharoah, D. F. Easton, B. A. Ponder, G. Chenevix-Trench and K. L. Novik (2003). "BRCA2 Arg372Hispolymorphism and epithelial ovarian cancer risk." *Int J Cancer* **103**(3): 427-430.

Baffa, R., M. Negrini, B. Mandes, M. Rugge, G. N. Ranzani, S. Hirohashi and C. M. Croce (1996). "Loss of heterozygosity for chromosome 11 in adenocarcinoma of the stomach." *Cancer Res* **56**(2): 268-272.

Bauer, A., O. Huber and R. Kemler (1998). "Pontin52, an interaction partner of beta-catenin, binds to the TATA box binding protein." *Proc Natl Acad Sci U S A* **95**(25): 14787-14792.

Benusiglio, P. R., F. Lesueur, C. Luccarini, D. M. Conroy, M. Shah, D. F. Easton, N. E. Day, A. M. Dunning, P. D. Pharoah and B. A. Ponder (2005). "Common ERBB2 polymorphisms and risk of breast cancer in a white British population: a case-control study." *Breast Cancer Res* **7**(2): R204-209.

Bepler, G., K. M. Fong, B. E. Johnson, K. C. O'Briant, L. A. Daly, P. V. Zimmerman, M. A. Garcia-Blanco and B. Peterson (1998). "Association of chromosome 11 locus D11S12 with histology, stage, and metastases in lung cancer." *Cancer Detect Prev* **22**(1): 14-19.

Berchuck, A., J. M. Schildkraut, R. M. Wenham, B. Calingaert, S. Ali, A. Henriott, S. Halabi, G. C. Rodriguez, D. Gertig, D. M. Purdie, L. Kelemen, A. B. Spurdle, J. Marks and G. Chenevix-Trench (2004). "Progesterone receptor promoter +331A polymorphism is associated with a reduced risk of

endometrioid and clear cell ovarian cancers." *Cancer Epidemiol Biomarkers Prev* **13**(12): 2141-2147.

Bergen, A. W., Y. Qi, K. A. Haque, R. A. Welch and S. J. Chanock (2005). "Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance." *BMC Biotechnol* **5**: 24.

Berthier-Schaad, Y., W. H. Kao, J. Coresh, L. Zhang, R. G. Ingersoll, R. Stephens and M. W. Smith (2007). "Reliability of high-throughput genotyping of whole genome amplified DNA in SNP genotyping studies." *Electrophoresis* **28**(16): 2812-2817.

Boon, T. (1993). "Teaching the immune system to fight cancer." *Sci Am* **268**(3): 82-89.

Cantley, L. C. (2002). "The phosphoinositide 3-kinase pathway." *Science* **296**(5573): 1655-1657.

Castresana, J. S., C. Barrios, L. Gomez and A. Kreicbergs (1992). "Amplification of the c-myc proto-oncogene in human chondrosarcoma." *Diagn Mol Pathol* **1**(4): 235-238.

Chen, L., C. J. Nievera, A. Y. Lee and X. Wu (2008). "Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair." *J Biol Chem* **283**(12): 7713-7720.

Christie, M. and M. K. Oehler (2006). "Molecular pathology of epithelial ovarian cancer." *J Br Menopause Soc* **12**(2): 57-63.

Conde, L., J. M. Vaquerizas, H. Dopazo, L. Arbiza, J. Reumers, F. Rousseau, J. Schymkowitz and J. Dopazo (2006). "PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes." *Nucleic Acids Res* **34**(Web Server issue): W621-625.

Cramer, D. W., R. F. Liberman, L. Titus-Ernstoff, W. R. Welch, E. R. Greenberg, J. A. Baron and B. L. Harlow (1999). "Genital talc exposure and risk of ovarian cancer." *Int J Cancer* **81**(3): 351-356.

Croce, C. M. (2008). "Oncogenes and cancer." *N Engl J Med* **358**(5): 502-511.

Cuatrecasas, M., A. Villanueva, X. Matias-Guiu and J. Prat (1997). "K-ras mutations in mucinous ovarian tumors: a clinicopathologic and molecular study of 95 cases." *Cancer* **79**(8): 1581-1586.

Cuatrecasas, M., N. Erill, E. Musulen, I. Costa, X. Matias-Guiu and J. Prat (1998). "K-ras mutations in nonmucinous ovarian epithelial tumors: a molecular analysis and clinicopathologic study of 144 patients." *Cancer* **82**(6): 1088-1095.

Cunningham, J. M., T. A. Sellers, J. M. Schildkraut, Z. S. Fredericksen, R. A. Vierkant, L. E. Kelemen, M. Gadre, C. M. Phelan, Y. Huang, J. G. Meyer, V. S. Pankratz and E. L. Goode (2008). "Performance of amplified DNA in an Illumina GoldenGate BeadArray assay." *Cancer Epidemiol Biomarkers Prev* **17**(7): 1781-1789.

Cupples, L. A., H. T. Arruda, E. J. Benjamin, R. B. D'Agostino, Sr., S. Demissie, A. L. DeStefano, J. Dupuis, K. M. Falls, C. S. Fox, D. J. Gottlieb, D. R. Govindaraju, C. Y. Guo, N. L. Heard-Costa, S. J. Hwang, S. Kathiresan, D. P. Kiel, J. M. Laramie, M. G. Larson, D. Levy, C. Y. Liu, K. L. Lunetta, M. D. Mailman, A. K. Manning, J. B. Meigs, J. M. Murabito, C. Newton-Cheh, G. T. O'Connor, C. J. O'Donnell, M. Pandey, S. Seshadri, R. S. Vasan, Z. Y. Wang, J. B. Wilk, P. A. Wolf, Q. Yang and L. D. Atwood (2007). "The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports." *BMC Med Genet* **8 Suppl 1**: S1.

De Roock, W., H. Piessevaux, J. De Schutter, M. Janssens, G. De Hertogh, N. Personeni, B. Biesmans, J. L. Van Laethem, M. Peeters, Y. Humblet, E. Van Cutsem and S. Tejpar (2007). "KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab." *Ann Oncol*.

Dehan, E., A. Ben-Dor, W. Liao, D. Lipson, H. Frimer, S. Rienstein, D. Simansky, M. Krupsky, P. Yaron, E. Friedman, G. Rechavi, M. Perlman, A. Aviram-Goldring, S. Izraeli, M. Bittner, Z. Yakhini and N. Kaminski (2007). "Chromosomal aberrations and gene expression profiles in non-small cell lung cancer." *Lung Cancer* **56**(2): 175-184.

Devlin, B., K. Roeder and S. A. Bacanu (2001). "Unbiased methods for population-based association studies." *Genet Epidemiol* **21**(4): 273-284.

Dunning, A. M., C. S. Healey, C. Baynes, A. T. Maia, S. Scollen, A. Vega, R. Rodriguez, N. L. Barbosa-Morais, B. A. Ponder, Y. L. Low, S. Bingham, C. A. Haiman, L. Le Marchand, A. Broeks, M. K. Schmidt, J. Hopper, M. Southey, M. W. Beckmann, P. A. Fasching, J. Peto, N. Johnson, S. E. Bojesen, B. Nordestgaard, R. L. Milne, J. Benitez, U. Hamann, Y. Ko, R. K. Schmutzler, B. Burwinkel, P. Schurmann, T. Dork, T. Heikkinen, H. Nevanlinna, A. Lindblom, S. Margolin, A. Mannermaa, V. M. Kosma, X. Chen, A. Spurdle, J. Change-Claude, D. Flesch-Janys, F. J. Couch, J. E. Olson, G. Severi, L. Baglietto, A. L. Borresen-Dale, V. Kristensen, D. J. Hunter, S. E. Hankinson, P. Devilee, M. Vreeswijk, J. Lissowska, L. Brinton, J. Liu, P. Hall, D. Kang, K. Y. Yoo, C. Y. Shen, J. C. Yu, H. Anton-Culver, A. Ziogoas, A. Sigurdson, J. Struewing, D. F. Easton, M. Garcia-Closas, M. K. Humphreys, J. Morrison, P. D. Pharoah, K. A. Pooley and G. Chenevix-Trench (2009). "Association of ESR1 gene tagging SNPs with breast cancer risk." *Hum Mol Genet* **18**(6): 1131-1139.

Easton, D. F., K. A. Pooley, A. M. Dunning, P. D. Pharoah, D. Thompson, D. G. Ballinger, J. P. Struewing, J. Morrison, H. Field, R. Luben, N. Wareham, S. Ahmed, C. S. Healey, R. Bowman, K. B. Meyer, C. A. Haiman, L. K. Kolonel, B. E. Henderson, L. Le Marchand, P. Brennan, S. Sangrajrang, V. Gaborieau, F. Odefrey, C. Y. Shen, P. E. Wu, H. C. Wang, D. Eccles, D. G. Evans, J. Peto, O. Fletcher, N. Johnson, S. Seal, M. R. Stratton, N. Rahman, G. Chenevix-Trench, S. E. Bojesen, B. G. Nordestgaard, C. K. Axelsson, M. Garcia-Closas, L. Brinton, S. Chanock, J. Lissowska, B. Peplonska, H. Nevanlinna, R. Fagerholm, H. Eerola, D. Kang, K. Y. Yoo, D. Y. Noh, S. H.

Ahn, D. J. Hunter, S. E. Hankinson, D. G. Cox, P. Hall, S. Wedren, J. Liu, Y. L. Low, N. Bogdanova, P. Schurmann, T. Dork, R. A. Tollenaar, C. E. Jacobi, P. Devilee, J. G. Klijn, A. J. Sigurdson, M. M. Doody, B. H. Alexander, J. Zhang, A. Cox, I. W. Brock, G. MacPherson, M. W. Reed, F. J. Couch, E. L. Goode, J. E. Olson, H. Meijers-Heijboer, A. van den Ouweland, A. Uitterlinden, F. Rivadeneira, R. L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, J. L. Hopper, M. McCredie, M. Southey, G. G. Giles, C. Schroen, C. Justenhoven, H. Brauch, U. Hamann, Y. D. Ko, A. B. Spurdle, J. Beesley, X. Chen, A. Mannermaa, V. M. Kosma, V. Kataja, J. Hartikainen, N. E. Day, D. R. Cox and B. A. Ponder (2007). "Genome-wide association study identifies novel breast cancer susceptibility loci." *Nature* **447**(7148): 1087-1093.

Eckhart, L., C. Kittel, S. Gawlas, F. Gruber, M. Mildner, B. Jilma and E. Tschachler (2006). "Identification of a novel exon encoding the amino-terminus of the predominant caspase-5 variants." *Biochem Biophys Res Commun* **348**(2): 682-688.

Everitt, B. and C. R. Palmer *The encyclopaedic companion to medical statistics*, London : Hodder Arnold ; 2005.

Fathalla, M. F. (1971). "Incessant ovulation--a factor in ovarian neoplasia?" *Lancet* **2**(7716): 163.

Ferrer-Costa, C., J. L. Gelpi, L. Zamakola, I. Parraga, X. de la Cruz and M. Orozco (2005). "PMUT: a web-based tool for the annotation of pathological mutations on proteins." *Bioinformatics* **21**(14): 3176-3178.

Fleishman, S. J., J. Schlessinger and N. Ben-Tal (2002). "A putative molecular-activation switch in the transmembrane domain of erbB2." *Proc Natl Acad Sci U S A* **99**(25): 15937-15940.

Folsom, A. R., J. P. Anderson and J. A. Ross (2004). "Estrogen replacement therapy and ovarian cancer." *Epidemiology* **15**(1): 100-104.

Forbes, S., J. Clements, E. Dawson, S. Bamford, T. Webb, A. Dogan, A. Flanagan, J. Teague, R. Wooster, P. A. Futreal and M. R. Stratton (2006). "Cosmic 2005." *Br J Cancer* **94**(2): 318-322.

French, J. E., G. D. Lacks, C. Trempus, J. K. Dunnick, J. Foley, J. Mahler, R. R. Tice and R. W. Tennant (2001). "Loss of heterozygosity frequency at the Trp53 locus in p53-deficient (+/-) mouse tumors is carcinogen-and tissue-dependent." *Carcinogenesis* **22**(1): 99-106.

Fritz, P., C. M. Cabrera, J. Dippon, A. Gerteis, W. Simon, W. E. Aulitzky and H. van der Kuip (2005). "c-erbB2 and topoisomerase IIalpha protein expression independently predict poor survival in primary human breast cancer: a retrospective study." *Breast Cancer Res* **7**(3): R374-384.

Fruman, D. A., R. E. Meyers and L. C. Cantley (1998). "Phosphoinositide kinases." *Annu Rev Biochem* **67**: 481-507.

Fusco, C., A. Reymond and A. S. Zervos (1998). "Molecular cloning and characterization of a novel retinoblastoma-binding protein." *Genomics* **51**(3): 351-358.

Gayther, S. A., H. Song, S. J. Ramus, S. K. Kjaer, A. S. Whittemore, L. Quaye, J. Tyrer, D. Shadforth, E. Hogdall, C. Hogdall, J. Blaeker, R. DiCioccio, V. McGuire, P. M. Webb, J. Beesley, A. C. Green, D. C. Whiteman, M. T. Goodman, G. Lurie, M. E. Carney, F. Modugno, R. B. Ness, R. P. Edwards, K. B. Moysich, E. L. Goode, F. J. Couch, J. M. Cunningham, T. A. Sellers, A. H. Wu, M. C. Pike, E. S. Iversen, J. R. Marks, M. Garcia-Closas, L. Brinton, J. Lissowska, B. Peplonska, D. F. Easton, I. Jacobs, B. A. Ponder, J. Schildkraut, C. L. Pearce, G. Chenevix-Trench, A. Berchuck and P. D. Pharoah (2007). "Tagging single nucleotide polymorphisms in cell cycle control genes and susceptibility to invasive epithelial ovarian cancer." *Cancer Res* **67**(7): 3027-3035.

Gayther, S. A., P. Russell, P. Harrington, A. C. Antoniou, D. F. Easton and B. A. Ponder (1999). "The contribution of germline BRCA1 and BRCA2 mutations to familial ovarian cancer: no evidence for other ovarian cancer-susceptibility genes." *Am J Hum Genet* **65**(4): 1021-1029.

Gemignani, M. L., A. C. Schlaerth, F. Bogomolniy, R. R. Barakat, O. Lin, R. Soslow, E. Venkatraman and J. Boyd (2003). "Role of KRAS and BRAF gene mutations in mucinous ovarian carcinoma." *Gynecol Oncol* **90**(2): 378-381.

Gertig, D. M., D. J. Hunter, D. W. Cramer, G. A. Colditz, F. E. Speizer, W. C. Willett and S. E. Hankinson (2000). "Prospective study of talc use and ovarian cancer." *J Natl Cancer Inst* **92**(3): 249-252.

Goodman, M. T., K. McDuffie, L. N. Kolonel, K. Terada, T. A. Donlon, L. R. Wilkens, C. Guo and L. Le Marchand (2001). "Case-control study of ovarian cancer and polymorphisms in genes involved in catecholestrogen formation and metabolism." *Cancer Epidemiol Biomarkers Prev* **10**(3): 209-216.

Gross, A. J. and P. H. Berg (1995). "A meta-analytical approach examining the potential relationship between talc exposure and ovarian cancer." *J Expo Anal Environ Epidemiol* **5**(2): 181-195.

Gudmundsson, J., R. B. Barkardottir, G. Eiriksdottir, T. Baldursson, A. Arason, V. Egilsson and S. Ingvarsson (1995). "Loss of heterozygosity at chromosome 11 in breast cancer: association of prognostic factors with genetic alterations." *Br J Cancer* **72**(3): 696-701.

Guntaka, R. V., B. R. Varma and K. T. Weber (2003). "Triplex-forming oligonucleotides as modulators of gene expression." *Int J Biochem Cell Biol* **35**(1): 22-31.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." *Cell* **100**(1): 57-70.

Harley, I., B. Rosen, H. A. Risch, K. Siminovitch, M. E. Beiner, J. McLaughlin, P. Sun and S. A. Narod (2008). "Ovarian cancer risk is associated with a

common variant in the promoter sequence of the mismatch repair gene MLH1." *Gynecol Oncol* **109**(3): 384-387.

Haurie, V., L. Menard, A. Nicou, C. Touriol, P. Metzler, J. Fernandez, D. Taras, P. Lestienne, C. Balabaud, P. Bioulac-Sage, H. Prats, J. Zucman-Rossi and J. Rosenbaum (2009). "Adenosine triphosphatase pontin is overexpressed in hepatocellular carcinoma and coregulated with reptin through a new posttranslational mechanism." *Hepatology* **50**(6): 1871-1883.

Hellstrom, I., G. Goodman, J. Pullman, Y. Yang and K. E. Hellstrom (2001). "Overexpression of HER-2 in ovarian carcinomas." *Cancer Res* **61**(6): 2420-2423.

Hingorani, S. R., M. A. Jacobetz, G. P. Robertson, M. Herlyn and D. A. Tuveson (2003). "Suppression of BRAF(V599E) in human melanoma abrogates transformation." *Cancer Res* **63**(17): 5198-5202.

Ho, E. S., C. R. Lai, Y. T. Hsieh, J. T. Chen, A. J. Lin, M. H. Hung and F. S. Liu (2001). "p53 mutation is infrequent in clear cell carcinoma of the ovary." *Gynecol Oncol* **80**(2): 189-193.

Hogdall, E. V., L. Christensen, S. K. Kjaer, J. Blaakaer, J. E. Bock, E. Glud, B. Norgaard-Pedersen and C. K. Hogdall (2003). "Distribution of HER-2 overexpression in ovarian carcinoma tissue and its prognostic value in patients with ovarian carcinoma: from the Danish MALOVA Ovarian Cancer Study." *Cancer* **98**(1): 66-73.

Hogdall, E. V., C. K. Hogdall, J. Blaakaer, L. Christensen, J. E. Bock, J. Vuust, E. Glud and S. K. Kjaer (2003). "K-ras alterations in Danish ovarian tumour patients. From the Danish "Malova" Ovarian Cancer study." *Gynecol Oncol* **89**(1): 31-36.

Huber, O., L. Menard, V. Haurie, A. Nicou, D. Taras and J. Rosenbaum (2008). "Pontin and reptin, two related ATPases with multiple roles in cancer." *Cancer Res* **68**(17): 6873-6876.

Hulla, J. E., J. E. French and J. K. Dunnick (2001). "Chromosome 11 allelotypes reflect a mechanism of chemical carcinogenesis in heterozygous p53-deficient mice." *Carcinogenesis* **22**(1): 89-98.

Huncharek, M., J. F. Geschwind and B. Kupelnick (2003). "Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies." *Anticancer Res* **23**(2C): 1955-1960.

Hwang, L. H. and A. W. Murray (1997). "A novel yeast screen for mitotic arrest mutants identifies DOC1, a new gene involved in cyclin proteolysis." *Mol Biol Cell* **8**(10): 1877-1887.

Imura, M., S. Yamashita, L. Y. Cai, J. Furuta, M. Wakabayashi, T. Yasugi and T. Ushijima (2006). "Methylation and expression analysis of 15 genes and three normally-methylated genes in 13 Ovarian cancer cell lines." *Cancer Lett* **241**(2): 213-220.

Jasmine, F., H. Ahsan, I. L. Andrulis, E. M. John, J. Chang-Claude and M. G. Kibriya (2008). "Whole-genome amplification enables accurate genotyping for microarray-based high-density single nucleotide polymorphism array." *Cancer Epidemiol Biomarkers Prev* **17**(12): 3499-3508.

Jha, S., E. Shibata and A. Dutta (2008). "Human Rvb1/Tip49 is required for the histone acetyltransferase activity of Tip60/NuA4 and for the downregulation of phosphorylation on H2AX after DNA damage." *Mol Cell Biol* **28**(8): 2690-2700.

Kaklamani, V. G., N. Hou, Y. Bian, J. Reich, K. Offit, L. S. Michel, W. S. Rubinstein, A. Rademaker and B. Pasche (2003). "TGFBR1*6A and cancer risk: a meta-analysis of seven case-control studies." *J Clin Oncol* **21**(17): 3236-3243.

Karnik, P., M. Paris, B. R. Williams, G. Casey, J. Crowe and P. Chen (1998). "Two distinct tumor suppressor loci within chromosome 11p15 implicated in breast cancer progression and metastasis." *Hum Mol Genet* **7**(5): 895-903.

Kato, S., S. Iida, T. Higuchi, T. Ishikawa, Y. Takagi, M. Yasuno, M. Enomoto, H. Uetake and K. Sugihara (2007). "PIK3CA mutation is predictive of poor survival in patients with colorectal cancer." *Int J Cancer* **121**(8): 1771-1778.

Kelemen, L. E., T. A. Sellers, J. M. Schildkraut, J. M. Cunningham, R. A. Vierkant, V. S. Pankratz, Z. S. Fredericksen, M. K. Gadre, D. N. Rider, M. Liebow and E. L. Goode (2008). "Genetic variation in the one-carbon transfer pathway and ovarian cancer risk." *Cancer Res* **68**(7): 2498-2506.

Kim, Y. T., T. Y. Kim, D. S. Lee, S. J. Park, J. Y. Park, S. J. Seo, H. S. Choi, H. J. Kang, S. Hahn, C. H. Kang, S. W. Sung and J. H. Kim (2008). "Molecular changes of epidermal growth factor receptor (EGFR) and KRAS and their impact on the clinical outcomes in surgically resected adenocarcinoma of the lung." *Lung Cancer* **59**(1): 111-118.

Koessler, T., E. M. Azzato, B. Perkins, R. J. Macinnis, D. Greenberg, D. F. Easton and P. D. Pharoah (2009). "Common germline variation in mismatch repair genes and survival after a diagnosis of colorectal cancer." *Int J Cancer* **124**(8): 1887-1891.

Krippner-Heidenreich, A., R. V. Talanian, R. Sekul, R. Kraft, H. Thole, H. Ottleben and B. Luscher (2001). "Targeting of the transcription factor Max during apoptosis: phosphorylation-regulated cleavage by caspase-5 at an unusual glutamic acid residue in position P1." *Biochem J* **358**(Pt 3): 705-715.

Lacey, J. V., Jr., P. J. Mink, J. H. Lubin, M. E. Sherman, R. Troisi, P. Hartge, A. Schatzkin and C. Schairer (2002). "Menopausal hormone replacement therapy and risk of ovarian cancer." *Jama* **288**(3): 334-341.

Ladanyi, M., C. K. Park, R. Lewis, S. C. Jhanwar, J. H. Healey and A. G. Huvos (1993). "Sporadic amplification of the MYC gene in human osteosarcomas." *Diagn Mol Pathol* **2**(3): 163-167.

Lakhani, S. R., S. Manek, F. Penault-Llorca, A. Flanagan, L. Arnout, S. Merrett, L. McGuffog, D. Steele, P. Devilee, J. G. Klijn, H. Meijers-Heijboer, P. Radice, S. Pilotti, H. Nevanlinna, R. Butzow, H. Sobol, J. Jacquemier, D. S. Lyonet, S. L. Neuhausen, B. Weber, T. Wagner, R. Winqvist, Y. J. Bignon, F. Monti, F. Schmitt, G. Lenoir, S. Seitz, U. Hamman, P. Pharoah, G. Lane, B. Ponder, D. T. Bishop and D. F. Easton (2004). "Pathology of ovarian cancers in BRCA1 and BRCA2 carriers." *Clin Cancer Res* **10**(7): 2473-2481.

Laplace-Marieze, V., N. Presneau, V. Sylvain, F. Kwiatkowski, A. Lortholary, A. Hardouin and Y. J. Bignon (1999). "Systematic sequencing of the BRCA-1 coding region for germ-line mutation detection in 70 French high-risk families." *Int J Oncol* **14**(5): 971-977.

Lauscher, J. C., C. Loddenkemper, L. Kosel, J. Grone, H. J. Buhr and O. Huber (2007). "Increased pontin expression in human colorectal cancer tissue." *Hum Pathol* **38**(7): 978-985.

Lee, Y. S., C. N. Tsai, C. L. Tsai, S. D. Chang, D. W. Hsueh, C. T. Liu, C. C. Ma, S. H. Lin, T. H. Wang and H. S. Wang (2008). "Comparison of whole genome amplification methods for further quantitative analysis with microarray-based comparative genomic hybridization." *Taiwan J Obstet Gynecol* **47**(1): 32-41.

Li, H., T. H. Lee and H. Avraham (2002). "A novel tricomplex of BRCA1, Nmi, and c-Myc inhibits c-Myc-induced human telomerase reverse transcriptase gene (hTERT) promoter activity in breast cancer." *J Biol Chem* **277**(23): 20965-20973.

Li, S., N. S. Ting, L. Zheng, P. L. Chen, Y. Ziv, Y. Shiloh, E. Y. Lee and W. H. Lee (2000). "Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response." *Nature* **406**(6792): 210-215.

Lichtenstein, P., N. V. Holm, P. K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe and K. Hemminki (2000). "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland." *N Engl J Med* **343**(2): 78-85.

Lievre, A., J. B. Bachet, D. Le Corre, V. Boige, B. Landi, J. F. Emile, J. F. Cote, G. Tomasic, C. Penna, M. Ducreux, P. Rougier, F. Penault-Llorca and P. Laurent-Puig (2006). "KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer." *Cancer Res* **66**(8): 3992-3995.

Lin, X. Y., M. S. Choi and A. G. Porter (2000). "Expression analysis of the human caspase-1 subfamily reveals specific regulation of the CASP5 gene by lipopolysaccharide and interferon-gamma." *J Biol Chem* **275**(51): 39920-39926.

Mann, A., E. Hogdall, S. J. Ramus, R. A. DiCioccio, C. Hogdall, L. Quaye, V. McGuire, A. S. Whittemore, M. Shah, D. Greenberg, D. F. Easton, B. A. Ponder, S. K. Kjaer, S. A. Gayther, D. J. Thompson, P. D. Pharoah and H. Song (2008). "Mismatch repair gene polymorphisms and survival in invasive ovarian cancer patients." *Eur J Cancer* **44**(15): 2259-2265.

Marsh, S. (2005). "Thymidylate synthase pharmacogenetics." *Invest New Drugs* **23**(6): 533-537.

Martinon, F., K. Burns and J. Tschopp (2002). "The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta." *Mol Cell* **10**(2): 417-426.

Meloni, A. R., E. J. Smith and J. R. Nevins (1999). "A mechanism for Rb/p130-mediated transcription repression involving recruitment of the CtBP corepressor." *Proc Natl Acad Sci U S A* **96**(17): 9574-9579.

Miki, Y., J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L. M. Bennett, W. Ding and et al. (1994). "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1." *Science* **266**(5182): 66-71.

Mills, P. K., D. G. Riordan, R. D. Cress and H. A. Young (2004). "Perineal talc exposure and epithelial ovarian cancer risk in the Central Valley of California." *Int J Cancer* **112**(3): 458-464.

Mok, S. C., K. K. Wong, R. K. Chan, C. C. Lau, S. W. Tsao, R. C. Knapp and R. S. Berkowitz (1994). "Molecular cloning of differentially expressed genes in human epithelial ovarian cancer." *Gynecol Oncol* **52**(2): 247-252.

Montgomery, K. G., D. M. Gertig, S. W. Baxter, R. L. Milne, G. S. Dite, M. R. McCredie, G. G. Giles, M. C. Southey, J. L. Hopper and I. G. Campbell (2003). "The HER2 I655V polymorphism and risk of breast cancer in women < age 40 years." *Cancer Epidemiol Biomarkers Prev* **12**(10): 1109-1111.

Moore, L. E., A. W. Bergen, K. A. Haque, Y. Qi, P. Castle, S. J. Chanock, K. Egan, P. Newcomb, L. Titus-Ernstoff, J. Alguacil, N. Rothman and M. Garcia-Closas (2007). "Whole genome amplification of buccal cytobrush DNA collected for molecular epidemiology studies." *Biomarkers* **12**(3): 303-312.

Muller, C. I., C. W. Miller, W. K. Hofmann, M. E. Gross, C. S. Walsh, N. Kawamata, Q. T. Luong and H. P. Koeffler (2007). "Rare mutations of the PIK3CA gene in malignancies of the hematopoietic system as well as endometrium, ovary, prostate and osteosarcomas, and discovery of a PIK3CA pseudogene." *Leuk Res* **31**(1): 27-32.

Ng, P. C. and S. Henikoff (2001). "Predicting deleterious amino acid substitutions." *Genome Res* **11**(5): 863-874.

Offman, J., K. Gascoigne, F. Bristow, P. Macpherson, M. Bignami, I. Casorelli, G. Leone, L. Pagano, S. Sica, O. Halil, D. Cummins, N. R. Banner and P. Karran (2005). "Repeated sequences in CASPASE-5 and FANCD2 but not NF1 are targets for mutation in microsatellite-unstable acute leukemia/myelodysplastic syndrome." *Mol Cancer Res* **3**(5): 251-260.

Pan, X., A. E. Urban, D. Palejev, V. Schulz, F. Grubert, Y. Hu, M. Snyder and S. M. Weissman (2008). "A procedure for highly specific, sensitive, and unbiased whole-genome amplification." *Proc Natl Acad Sci U S A* **105**(40): 15499-15504.

Passmore, L. A., E. A. McCormack, S. W. Au, A. Paul, K. R. Willison, J. W. Harper and D. Barford (2003). "Doc1 mediates the activity of the anaphase-promoting complex by contributing to substrate recognition." *EMBO J* **22**(4): 786-796.

Patel, H. P., L. Lu, R. T. Blaszak and J. J. Bissler (2004). "PKD1 intron 21: triplex DNA formation and effect on replication." *Nucleic Acids Res* **32**(4): 1460-1468.

Pharoah, P. D., D. F. Easton, D. L. Stockton, S. Gayther and B. A. Ponder (1999). "Survival in familial, BRCA1-associated, and BRCA2-associated epithelial ovarian cancer. United Kingdom Coordinating Committee for Cancer Research (UKCCCR) Familial Ovarian Cancer Study Group." *Cancer Res* **59**(4): 868-871.

Pinard, R., A. de Winter, G. J. Sarkis, M. B. Gerstein, K. R. Tartaro, R. N. Plant, M. Egholm, J. M. Rothberg and J. H. Leamon (2006). "Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing." *BMC Genomics* **7**: 216.

Pinto, D., D. Pereira, C. Portela, J. L. da Silva, C. Lopes and R. Medeiros (2005). "The influence of HER2 genotypes as molecular markers in ovarian cancer outcome." *Biochem Biophys Res Commun* **335**(4): 1173-1178.

Pohl, G., C. L. Ho, R. J. Kurman, R. Bristow, T. L. Wang and M. Shih Ie (2005). "Inactivation of the mitogen-activated protein kinase pathway as a potential target-based therapy in ovarian serous tumors with KRAS or BRAF mutations." *Cancer Res* **65**(5): 1994-2000.

Quaye, L., D. Dafou, S. J. Ramus, H. Song, A. Gentry-Maharaj, M. Notaridou, E. Hogdall, S. K. Kjaer, L. Christensen, C. Hogdall, D. F. Easton, I. Jacobs, U. Menon, P. D. Pharoah and S. A. Gayther (2009). "Functional complementation studies identify candidate genes and common genetic variants associated with ovarian cancer survival." *Hum Mol Genet* **18**(10): 1869-1878.

Quaye, L., S. A. Gayther, S. J. Ramus, R. A. Di Cioccio, V. McGuire, E. Hogdall, C. Hogdall, J. Blaakr, D. F. Easton, B. A. Ponder, I. Jacobs, S. K. Kjaer, A. S. Whittemore, C. L. Pearce, P. D. Pharoah and H. Song (2008). "The effects of common genetic variants in oncogenes on ovarian cancer survival." *Clin Cancer Res* **14**(18): 5833-5839.

Ramus, S. J., A. Fishman, P. D. Pharoah, S. Yarkoni, M. Altaras and B. A. Ponder (2001). "Ovarian cancer survival in Ashkenazi Jewish patients with BRCA1 and BRCA2 mutations." *Eur J Surg Oncol* **27**(3): 278-281.

Ramus, S. J., R. A. Vierkant, S. E. Johnatty, M. C. Pike, D. J. Van Den Berg, A. H. Wu, C. L. Pearce, U. Menon, A. Gentry-Maharaj, S. A. Gayther, R. A. Dicioccio, V. McGuire, A. S. Whittemore, H. Song, D. F. Easton, P. D. Pharoah, M. Garcia-Closas, S. Chanock, J. Lissowska, L. Brinton, K. L. Terry, D. W. Cramer, S. S. Tworoger, S. E. Hankinson, A. Berchuck, P. G. Moorman, J. M. Schildkraut, J. M. Cunningham, M. Liebow, S. K. Kjaer, E. Hogdall, C. Hogdall, J. Blaakaer, R. B. Ness, K. B. Moysich, R. P. Edwards,

M. E. Carney, G. Lurie, M. T. Goodman, S. Wang-Gohrke, S. Kropp, J. Chang-Claude, P. M. Webb, X. Chen, J. Beesley, G. Chenevix-Trench and E. L. Goode (2008). "Consortium analysis of 7 candidate SNPs for ovarian cancer." *Int J Cancer* **123**(2): 380-388.

Reumers, J., L. Conde, I. Medina, S. Maurer-Stroh, J. Van Durme, J. Dopazo, F. Rousseau and J. Schymkowitz (2008). "Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNPeffect and PupaSuite databases." *Nucleic Acids Res* **36**(Database issue): D825-829.

Rhim, J. S. (1988). "Viruses, oncogenes, and cancer." *Cancer Detect Prev* **11**(3-6): 139-149.

Rodriguez, C., A. V. Patel, E. E. Calle, E. J. Jacob and M. J. Thun (2001). "Estrogen replacement therapy and ovarian cancer mortality in a large prospective study of US women." *Jama* **285**(11): 1460-1465.

Rousseau, B., L. Menard, V. Haurie, D. Taras, J. F. Blanc, F. Moreau-Gaudry, P. Metzler, M. Hugues, S. Boyault, S. Lemiere, X. Canron, P. Costet, M. Cole, C. Balabaud, P. Bioulac-Sage, J. Zucman-Rossi and J. Rosenbaum (2007). "Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma." *Hepatology* **46**(4): 1108-1118.

Roymans, D. and H. Slegers (2001). "Phosphatidylinositol 3-kinases in tumor progression." *Eur J Biochem* **268**(3): 487-498.

Rubin, S. C., I. Benjamin, K. Behbakht, H. Takahashi, M. A. Morgan, V. A. LiVolsi, A. Berchuck, M. G. Muto, J. E. Garber, B. L. Weber, H. T. Lynch and J. Boyd (1996). "Clinical and pathological features of ovarian cancer in women with germ-line mutations of BRCA1." *N Engl J Med* **335**(19): 1413-1416.

Rutter, J. L., N. Chatterjee, S. Wacholder and J. Struewing (2003). "The HER2 I655V polymorphism and breast cancer risk in Ashkenazim." *Epidemiology* **14**(6): 694-700.

Samowitz, W. S., C. Sweeney, J. Herrick, H. Albertsen, T. R. Levin, M. A. Murtaugh, R. K. Wolff and M. L. Slattery (2005). "Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers." *Cancer Res* **65**(14): 6063-6069.

Samuels, Y. and K. Ericson (2006). "Oncogenic PI3K and its role in cancer." *Curr Opin Oncol* **18**(1): 77-82.

Sankaranarayanan, R. and J. Ferlay (2006). "Worldwide burden of gynaecological cancer: the size of the problem." *Best Pract Res Clin Obstet Gynaecol* **20**(2): 207-225.

Sartori, A. A., C. Lukas, J. Coates, M. Mistrik, S. Fu, J. Bartek, R. Baer, J. Lukas and S. P. Jackson (2007). "Human CtIP promotes DNA end resection." *Nature* **450**(7169): 509-514.

Schwarze, S. R., S. E. DePrimo, L. M. Grabert, V. X. Fu, J. D. Brooks and D. F. Jarrard (2002). "Novel pathways associated with bypassing cellular

senescence in human prostate epithelial cells." *J Biol Chem* **277**(17): 14877-14883.

Shaw, M. E. and M. A. Knowles (1995). "Deletion mapping of chromosome 11 in carcinoma of the bladder." *Genes Chromosomes Cancer* **13**(1): 1-8.

Shayesteh, L., Y. Lu, W. L. Kuo, R. Baldocchi, T. Godfrey, C. Collins, D. Pinkel, B. Powell, G. B. Mills and J. W. Gray (1999). "PIK3CA is implicated as an oncogene in ovarian cancer." *Nat Genet* **21**(1): 99-102.

Sieben, N. L., P. Macropoulos, G. M. Roemen, S. M. Kolkman-Uljee, G. Jan Fleuren, R. Houmadi, T. Diss, B. Warren, M. Al Adnani, A. P. De Goeij, T. Krausz and A. M. Flanagan (2004). "In ovarian neoplasms, BRAF, but not KRAS, mutations are restricted to low-grade serous tumours." *J Pathol* **202**(3): 336-340.

Simes, R. J. (1986). "An improved Bonferroni procedure for multiple tests of significance." *Biometrika* **73**: 751–754.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. J. Hudson, A. Montpetit, A. V. Pshezhetsky, M. Prentki, B. I. Posner, D. J. Balding, D. Meyre, C. Polychronakos and P. Froguel (2007). "A genome-wide association study identifies novel risk loci for type 2 diabetes." *Nature* **445**(7130): 881-885.

Song, H., S. J. Ramus, S. Kruger Kjaer, R. A. Dicioccio, G. Chenevix-Trench, C. L. Pearce, E. Hogdall, A. S. Whittemore, V. McGuire, C. Hogdall, J. Blaakaer, A. H. Wu, D. J. Van Den Berg, D. O. Stram, U. Menon, A. Gentry-Maharaj, I. J. Jacobs, P. M. Webb, J. Beesley, X. Chen, M. A. Rossing, J. A. Doherty, J. Chang-Claude, S. Wang-Gohrke, M. T. Goodman, G. Lurie, P. J. Thompson, M. E. Carney, R. B. Ness, K. Moysich, E. L. Goode, R. A. Vierkant, J. M. Cunningham, S. Anderson, J. M. Schildkraut, A. Berchuck, E. S. Iversen, P. G. Moorman, M. Garcia-Closas, S. Chanock, J. Lissowska, L. Brinton, H. Anton-Culver, A. Ziogas, W. R. Brewster, B. A. Ponder, D. F. Easton, S. A. Gayther and P. D. Pharoah (2009a). "Association between invasive ovarian cancer susceptibility and 11 best candidate SNPs from breast cancer genome-wide association study." *Hum Mol Genet*.

Song, H., S. J. Ramus, D. Shadforth, L. Quaye, S. K. Kjaer, R. A. Dicioccio, A. M. Dunning, E. Hogdall, C. Hogdall, A. S. Whittemore, V. McGuire, F. Lesueur, D. F. Easton, I. J. Jacobs, B. A. Ponder, S. A. Gayther and P. D. Pharoah (2006). "Common variants in RB1 gene and risk of invasive ovarian cancer." *Cancer Res* **66**(20): 10220-10226.

Song, H., S. J. Ramus, J. Tyrer, K. L. Bolton, A. Gentry-Maharaj, E. Wozniak, H. Anton-Culver, J. Chang-Claude, D. W. Cramer, R. DiCioccio, T. Dork, E. L. Goode, M. T. Goodman, J. M. Schildkraut, T. Sellers, L. Baglietto, M. W. Beckmann, J. Beesley, J. Blaakaer, M. E. Carney, S. Chanock, Z. Chen, J. M. Cunningham, E. Dicks, J. A. Doherty, M. Durst, A. B. Ekici, D. Fenstermacher, B. L. Fridley, G. Giles, M. E. Gore, I. De Vivo, P. Hillemanns, C. Hogdall, E. Hogdall, E. S. Iversen, I. J. Jacobs, A. Jakubowska, D. Li, J. Lissowska, J. Lubinski, G. Lurie, V. McGuire, J.

McLaughlin, K. Medrek, P. G. Moorman, K. Moysich, S. Narod, C. Phelan, C. Pye, H. Risch, I. B. Runnebaum, G. Severi, M. Southey, D. O. Stram, F. C. Thiel, K. L. Terry, Y. Y. Tsai, S. S. Tworoger, D. J. Van Den Berg, R. A. Vierkant, S. Wang-Gohrke, P. M. Webb, L. R. Wilkens, A. H. Wu, H. Yang, W. Brewster, A. Ziogas, R. Houlston, I. Tomlinson, A. S. Whittemore, M. A. Rossing, B. A. Ponder, C. L. Pearce, R. B. Ness, U. Menon, S. K. Kjaer, J. Gronwald, M. Garcia-Closas, P. A. Fasching, D. F. Easton, G. Chenevix-Trench, A. Berchuck, P. D. Pharoah and S. A. Gayther (2009b). "A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2." *Nat Genet* **41**(9): 996-1000.

Sorensen, K. M., C. Jespersgaard, J. Vuust, D. Hougaard, B. Norgaard-Pedersen and P. S. Andersen (2007). "Whole genome amplification on DNA from filter paper blood spot samples: an evaluation of selected systems." *Genet Test* **11**(1): 65-71.

Stratton, J. F., P. Pharoah, S. K. Smith, D. Easton and B. A. Ponder (1998). "A systematic review and meta-analysis of family history and risk of ovarian cancer." *Br J Obstet Gynaecol* **105**(5): 493-499.

Takeuchi, S., N. Takeuchi, A. C. Fermin, H. Taguchi and H. P. Koeffler (2003). "Frameshift mutations in caspase-5 and other target genes in leukemia and lymphoma cell lines having microsatellite instability." *Leuk Res* **27**(4): 359-361.

Talseth-Palmer, B. A., N. A. Bowden, A. Hill, C. Meldrum and R. J. Scott (2008). "Whole genome amplification and its impact on CGH array profiles." *BMC Res Notes* **1**: 56.

Tan, D. S., C. Rothermundt, K. Thomas, E. Bancroft, R. Eeles, S. Shanley, A. Ardern-Jones, A. Norman, S. B. Kaye and M. E. Gore (2008). ""BRCAness" Syndrome in Ovarian Cancer: A Case-Control Study Describing the Clinical Features and Outcome of Patients With Epithelial Ovarian Cancer Associated With BRCA1 and BRCA2 Mutations." *J Clin Oncol*.

Tandle, A. T., C. Mazzanti, H. R. Alexander, D. D. Roberts and S. K. Libutti (2005). "Endothelial monocyte activating polypeptide-II induced gene expression changes in endothelial cells." *Cytokine* **30**(6): 347-358.

Thomas, D. C., R. W. Haile and D. Duggan (2005). "Recent developments in genomewide association scans: a workshop summary and review." *Am J Hum Genet* **77**(3): 337-345.

Thomas, G., K. B. Jacobs, M. Yeager, P. Kraft, S. Wacholder, N. Orr, K. Yu, N. Chatterjee, R. Welch, A. Hutchinson, A. Crenshaw, G. Cancel-Tassin, B. J. Staats, Z. Wang, J. Gonzalez-Bosquet, J. Fang, X. Deng, S. I. Berndt, E. E. Calle, H. S. Feigelson, M. J. Thun, C. Rodriguez, D. Albanes, J. Virtamo, S. Weinstein, F. R. Schumacher, E. Giovannucci, W. C. Willett, O. Cussenot, A. Valeri, G. L. Andriole, E. D. Crawford, M. Tucker, D. S. Gerhard, J. F. Fraumeni, Jr., R. Hoover, R. B. Hayes, D. J. Hunter and S. J. Chanock

(2008). "Multiple loci identified in a genome-wide association study of prostate cancer." *Nat Genet* **40**(3): 310-315.

Tranah, G. J., P. J. Lescault, D. J. Hunter and I. De Vivo (2003). "Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies." *Biotechnol Lett* **25**(13): 1031-1036.

Udler, M. S., E. M. Azzato, C. S. Healey, S. Ahmed, K. A. Pooley, D. Greenberg, M. Shah, A. E. Teschendorff, C. Caldas, A. M. Dunning, E. A. Ostrander, N. E. Caporaso, D. Easton and P. D. Pharoah (2009). "Common germline polymorphisms in COMT, CYP19A1, ESR1, PGR, SULT1E1 and STS and survival after a diagnosis of breast cancer." *Int J Cancer* **125**(11): 2687-2696.

Venteicher, A. S., Z. Meng, P. J. Mason, T. D. Veenstra and S. E. Artandi (2008). "Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly." *Cell* **132**(6): 945-957.

Viel, A., F. Giannini, L. Tumiotto, F. Sopracordevole, M. C. Visentin and M. Boiocchi (1992). "Chromosomal localisation of two putative 11p oncosuppressor genes involved in human ovarian tumours." *Br J Cancer* **66**(6): 1030-1036.

Villafranca, E., Y. Okruzhnov, M. A. Dominguez, J. Garcia-Foncillas, I. Azinovic, E. Martinez, J. J. Illarramendi, F. Arias, R. Martinez Monge, E. Salgado, S. Angeletti and A. Brugarolas (2001). "Polymorphisms of the repeated sequences in the enhancer region of the thymidylate synthase gene promoter may predict downstaging after preoperative chemoradiation in rectal cancer." *J Clin Oncol* **19**(6): 1779-1786.

Volinia, S., I. Hiles, E. Ormondroyd, D. Nizetic, R. Antonacci, M. Rocchi and M. D. Waterfield (1994). "Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene." *Genomics* **24**(3): 472-477.

Weiss, N. S., T. Homonchuk and J. L. Young, Jr. (1977). "Incidence of the histologic types of ovarian cancer: the U.S. Third National Cancer Survey, 1969-1971." *Gynecol Oncol* **5**(2): 161-167.

Wenham, R. M., B. Calingaert, S. Ali, K. McClean, R. Whitaker, R. Bentley, J. M. Lancaster, J. Schildkraut, J. Marks and A. Berchuck (2003). "Matrix metalloproteinase-1 gene promoter polymorphism and risk of ovarian cancer." *J Soc Gynecol Investig* **10**(6): 381-387.

Whittemore, A. S., R. R. Balise, P. D. Pharoah, R. A. Dicioccio, I. Oakley-Girvan, S. J. Ramus, M. Daly, M. B. Usinowicz, K. Garlinghouse-Jones, B. A. Ponder, S. Buys, R. Senie, I. Andrulis, E. John, J. L. Hopper and M. S. Piver (2004). "Oral contraceptive use and ovarian cancer risk among carriers of BRCA1 or BRCA2 mutations." *Br J Cancer* **91**(11): 1911-1915.

Woenckhaus, J., K. Steger, K. Sturm, K. Munstedt, F. E. Franke and I. Fenic (2007). "Prognostic value of PIK3CA and phosphorylated AKT expression in ovarian cancer." *Virchows Arch* **450**(4): 387-395.

Wood, M. A., S. B. McMahon and M. D. Cole (2000). "An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc." *Mol Cell* **5**(2): 321-330.

Wu, M., D. R. Soler, M. C. Abba, M. I. Nunez, R. Baer, C. Hatzis, A. Llombart-Cussac, A. Llombart-Bosch and C. M. Aldaz (2007). "CtIP silencing as a novel mechanism of tamoxifen resistance in breast cancer." *Mol Cancer Res* **5**(12): 1285-1295.

Wu, Y., R. A. Soslow, D. S. Marshall, M. Leitao and B. Chen (2004). "Her-2/neu expression and amplification in early stage ovarian surface epithelial neoplasms." *Gynecol Oncol* **95**(3): 570-575.

Xing, J., W. S. Watkins, Y. Zhang, D. J. Witherspoon and L. B. Jorde (2008). "High fidelity of whole-genome amplified DNA on high-density single nucleotide polymorphism arrays." *Genomics* **92**(6): 452-456.

Xu, X. L., L. C. Wu, F. Du, A. Davis, M. Peyton, Y. Tomizawa, A. Maitra, G. Tomlinson, A. F. Gazdar, B. E. Weissman, A. M. Bowcock, R. Baer and J. D. Minna (2001). "Inactivation of human SRBC, located within the 11p15.5-p15.4 tumor suppressor region, in breast and lung cancers." *Cancer Res* **61**(21): 7943-7949.

Yeager, M., N. Orr, R. B. Hayes, K. B. Jacobs, P. Kraft, S. Wacholder, M. J. Minichiello, P. Fearnhead, K. Yu, N. Chatterjee, Z. Wang, R. Welch, B. J. Staats, E. E. Calle, H. S. Feigelson, M. J. Thun, C. Rodriguez, D. Albanes, J. Virtamo, S. Weinstein, F. R. Schumacher, E. Giovannucci, W. C. Willett, G. Cancel-Tassin, O. Cussenot, A. Valeri, G. L. Andriole, E. P. Gelmann, M. Tucker, D. S. Gerhard, J. F. Fraumeni, Jr., R. Hoover, D. J. Hunter, S. J. Chanock and G. Thomas (2007). "Genome-wide association study of prostate cancer identifies a second risk locus at 8q24." *Nat Genet* **39**(5): 645-649.

Zanke, B. W., C. M. Greenwood, J. Rangrej, R. Kustra, A. Tenesa, S. M. Farrington, J. Prendergast, S. Olschwang, T. Chiang, E. Crowdy, V. Ferretti, P. Laflamme, S. Sundararajan, S. Roumy, J. F. Olivier, F. Robidoux, R. Sladek, A. Montpetit, P. Campbell, S. Bezieau, A. M. O'Shea, G. Zogopoulos, M. Cotterchio, P. Newcomb, J. McLaughlin, B. Younghusband, R. Green, J. Green, M. E. Porteous, H. Campbell, H. Blanche, M. Sahbatou, E. Tubacher, C. Bonaiti-Pellie, B. Buecher, E. Riboli, S. Kury, S. J. Chanock, J. Potter, G. Thomas, S. Gallinger, T. J. Hudson and M. G. Dunlop (2007). "Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24." *Nat Genet* **39**(8): 989-994.

Zheng, L., S. Li, T. G. Boyer and W. H. Lee (2000). "Lessons learned from BRCA1 and BRCA2." *Oncogene* **19**(53): 6159-6175.

# Appendices

## Appendix I: MMCT-18 master-list

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrids TOV112D | *C20orf100* | 20q13.12 | | 1 | | 4 | 153,449 | 167 | 60 | TOX high mobility group box family member 2. Granulosa cell HMG box protein. Putative transcriptional activator involved in the hypothalamo-pituitary-gonadal system. |
| Down in hybrids TOV112D | *FAM19A5* | 22q13.32 | | 2 | | 8 | 269,793 | 216 | 83 | Family with sequence similarity 19 (chemokine (C-C motif)-like), member A5. Unknown function. |
| Down in hybrids TOV112D | *PDGFRL* | 8p22-p21.3 | | 3 | | 91 | 65,917 | 124 | 56 | Platelet-derived growth factor receptor-like. Mutations in gene, or deletion of a chromosomal segment containing this gene, are associated with sporadic hepatocellular carcinomas, colorectal cancers, and non-small cell lung cancers. May function as tumour suppressor. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrids TOV112D | *DIO2* | 14q24.2-q24.3 | | 4 | | 1 | 14,656 | 13 | 5 | Deiodinase, iodothyronine, type II. |
| Down in hybrids TOV112D | *RGC32* | 13q14.11 | | 5 | | 80 | 13,323 | 17 | 8 | Believed to regulate cell cycle progression. Induced by p53 in response to DNA damage, or by sublytic levels of complement system proteins that result in activation of the cell cycle. |
| Down in hybrids TOV112D | *APCDD1* | 18p11.22 | | 6 | | 14 | 34,154 | 42 | 13 | Adenomatosis polyposis coli down-regulated 1. May play a role in colorectal tumorigenesis. May be a developmental target gene of the Wnt/β-catenin pathway |
| Down in hybrids TOV112D | *SLITRK6* | 13q31.1 | | 7 | | 55 | 6,561 | 5 | 4 | SLIT and NTRK-like family, member 6. SLITRKs are expressed predominantly in neural tissues and have neurite-modulating activity. |
| Down in hybrids TOV112D | *TCBA1* | 6q21 | | 8 | | 19 | 1,021,734 | 1093 | 279 | Na+/K+ transporting ATPase interacting 2. T-cell lymphoma breakpoint associated target 1. |
| Down in hybrids TOV112D | *CXXC4* | 4q22-q24 | | 9 | | 11 | 26,485 | 5 | 3 | CXXC finger 4. May be in the Wnt receptor signalling pathway. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrids TOV112D | *CLEC11A* | 19q13.3 | | 10 | | 13 | 2,376 | None | | Unknown function. |
| Down in hybrids TOV112D | *MME* | 3q25.1-q25.2 | | 15 | | 9 | 104,033 | 83 | 35 | Membrane metallo-endopeptidase. Gene encodes a common acute lymphocytic leukemia antigen that is an important cell surface marker in the diagnosis of human acute lymphocytic leukemia (ALL). This protein is present on leukemic cells of pre-B phenotype, which represent 85% of cases of ALL. Also found on variety of normal tissues. |
| Down in hybrids TOV112D | *EN1* | 2q13-q21 | | 17 | | 2 | 5,993 | 1 | 1 | Engrailed homeobox 1. Homeobox-containing genes are believed to be involved in controlling development. The human engrailed homologs 1 and 2 encode homeodomain-containing proteins and have been implicated in the control of pattern formation during development of the central nervous system. |
| Down in hybrids TOV112D | *C21orf94* | 21q21.3 | | 19 | | 7 | 9,572 | 14 | 9 | Uncharacterized protein |
| Down in hybrids TOV112D | *ARMCX2* | Xq21.33-q22.2 | | 22 | | 5 | 4,609 | 2 | 2 | Armadillo repeat containing, X-linked 2. Arm protein lost in epithelial cancers. Gene encodes a |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | member of the ALEX family of proteins and may play a role in tumour suppression. The encoded protein contains a potential N-terminal transmembrane domain and a single Armadillo (arm) repeat. Other proteins containing the arm repeat are involved in development, maintenance of tissue integrity, and tumorigenesis. |
| Down in hybrids TOV112D | *CRABP1* | 15q24 | | 28 | | 3 | 7,878 | 2 | 2 | Cellular retinoic acid binding protein 1. Cellular retinoic acid-binding protein is assumed to play an important role in retinoic acid-mediated differentiation and proliferation processes. |
| Down in hybrids TOV112D | *MAB21L1* | 13q13 | | 38 | | 6 | 2,511 | 1 | 1 | Mab-21-like 1 (C. elegans). This gene is similar to the MAB-21 cell fate-determining gene found in C. elegans. May be involved in eye and cerebellum development, and it has been proposed that expansion of a trinucleotide repeat region in the 5' UTR may play a role in a variety of psychiatric disorders. |
| Down in hybrids TOV112D | *CSN3* | 4q21.1 | | 46 | | 10 | 8,838 | 20 | 7 | Casein kappa. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrids TOV112D | *ANKFN1* | 17q23.2 | | 1 | | 7 | 329,171 | 261 | 51 | Ankyrin-repeat and fibronectin type III domain containing 1. Unknown function. |
| Up in hybrids TOV112D | *CDH12* | 5p14-p13 | | 2 | | 13 | 1,102,756 | 792 | 133 | Cadherin 12. Gene encodes an integral membrane protein that mediates calcium-dependent cell-cell adhesion. |
| Up in hybrids TOV112D | *CXCL14* | 5q31 | | 3 | | 10 | 8,594 | 11 | 6 | Chemokine (C-X-C motif) ligand 14. |
| Up in hybrids TOV112D | *LIX1* | 5q15 | | 4 | | 2 | 51,002 | 36 | 7 | Lix1 homolog (mouse). Limb expression 1. Unknown function |
| Up in hybrids TOV112D | *C18orf34* | 18q12.1 | | 5 | | 5 | | | | Hypothetical protein |
| Up in hybrids TOV112D | *PRAC* | 17q21 | | 6 | | 6 | 801 | 1 | 1 | This gene is reported to be specifically expressed in prostate, rectum and distal colon. Sequence analysis suggests that it may play a regulatory role in the nucleus. |
| Up in hybrids TOV112D | *FILIP1L (DOC1)* | 3q12.1 | | 7 | | 9 | 281,369 | 135 | 8 | Filament A interacting protein 1-like. GPBP-interacting protein 90; down-regulated in ovarian cancer 1. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrids TOV112D | *GNAT3* | 7q21.11 | | 8 | | 18 | 53,255 | Not in hapmap | | Guanine nucleotide binding protein, alpha transducing 3. |
| Up in hybrids TOV112D | *OLFM3* | 1p22 | | 9 | | 16 | 194,456 | 250 | 69 | Olfactomedin 3. Expressed in brain and retina; may be a candidate gene for disorders involving the anterior segment of the eye and the retina. |
| Up in hybrids TOV112D | *TAIP-2* | 2q24.3 | | 10 | | 29 | 211,055 | 86 | 21 | Family with sequence similarity 130, member A2. TGF beta induced apotosis protein. |
| Up in hybrids TOV112D | *NELL2* | 12q13.11-q13.12 | | 13 | | 4 | 368,073 | 319 | 49 | NEL-like 2 (chicken). Neural epidermal growth factor-like 2. Gene encodes a cytoplasmic protein that contains epidermal growth factor (EGF) -like repeats. The encoded heterotrimeric protein may be involved in cell growth regulation and differentiation. |
| Up in hybrids TOV112D | *AXIN2* | 17q23-q24 | | 19 | | 8 | 33,084 | 14 | 12 | Axin 2 (conductin, axil). Inhibitor β-catenin in the Wnt signalling pathway. In region of frequent loss of heterozygosity in breast cancer, neuroblastoma, and other tumors. Mutations in this gene have been associated with colorectal cancer with defective mismatch repair. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrids TOV112D | *PSMAL* | 11q14.3 | | 22 | | 1 | 66,834 | 17 | 4 | Growth-inhibiting protein 26. |
| Up in hybrids TOV112D | *STMN2* | 8q21.13 | | 24 | | 3 | 54,996 | 86 | 24 | Stathmin-like 2. May play a role in neuronal differentiation, and in modulating membrane interaction with the cytoskeleton during neurite outgrowth |
| Up in hybrid TOV21G | *C1orf116* | 1q32.1 | 1 | | 28 | | 14,226 | 5 | 5 | Specifically androgen-regulated protein. Unknown function |
| Up in hybrid TOV21G | *C11orf17* | 11p15.3 | 2 | | 120 | | 8,925 | 10 | 4 | Breast cancer associated gene 3; koyt binding protein 1; koyt binding protein 2; koyt binding protein 3; protein kinase A-interacting protein 1. |
| Up in hybrid TOV21G | *STAG3* | 7q22.1 | 3 | | 31 | | 43,764 | 28 | 3 | Stromal antigen 3. Encoded protein is a component of the cohesion complex during chromosome segregation. |
| Up in hybrid TOV21G | *RLN1* | 9p24.1 | 4 | | 50 | | 4,904 | 8 | 3 | Relaxin 1. Unknown function. |
| Up in hybrid TOV21G | *VTCN1* | 1p13.1 | 5 | | 35 | | 67,347 | 54 | 22 | V-set domain containing T cell activation inhibitor 1. Expressed on the surface of antigen-presenting cells and interact with ligands on T lymphocytes. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrid TOV21G | *CASP5* | 11q22.2-q22.3 | 6 | | 40 | | 14,729 | 17 | 9 | Caspase 5. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis. Overexpression of the active form of this enzyme induces apoptosis in fibroblasts. Max, a central component of the Myc/Max/Mad transcription regulation network important for cell growth, differentiation, and apoptosis, is cleaved by this protein. Target gene in the microsatellite mutator pathway for cancer. |
| Up in hybrid TOV21G | *STK17A* | 7p12-p14 | 7 | | 103 | | 42,996 | 34 | 11 | Gene encodes an autophosphorylated nuclear protein which acts as a positive regulator of apoptosis. |
| Up in hybrid TOV21G | *CXorf6* | Xq28 | 8 | | 86 | | 68,729 | 87 | 35 | Hypothetical protein. Putative DNA binding protein, expressed in skeletal muscle, brain, heart. May be involved in gonadal function. |
| Up in hybrid TOV21G | *F2* | 11p11-q12 | 9 | | 12 | | 20,300 | 13 | 4 | Coagulation factor II (thrombin). Involved in first step of the coagulation cascade which ultimately results in the stemming of blood loss. Also plays a role in maintaining vascular integrity during development and postnatal life. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrid TOV21G | *C10orf33* | 10q24.2 | 10 | | 61 | | 31,619 | 88 | 9 | Oxidoreductase activity. |
| Up in hybrid TOV21G | *ANXA13* | 8q24.13 | 16 | | 5 | | 56,613 | 98 | 40 | Annexin A13. May play a role in the regulation of cellular growth and in signal transduction pathways. |
| Up in hybrid TOV21G | *FLJ20701* | 2q36.3 | 25 | | 2 | | 247,291 | 339 | 85 | Phosphotyrosine interaction domain containing 1. Increases proliferation of preadipocytes. |
| Up in hybrid TOV21G | *SLC17A2* | 6p21.3 | 28 | | 8 | | 17,857 | 24 | 7 | Solute carrier family 17 (sodium phosphate), member 2. May be involved in actively transporting phosphate into cells. |
| Up in hybrid TOV21G | *MPL* | 1p34 | 31 | | 9 | | 16,660 | 3 | 1 | Myeloproliferative leukemia virus oncogene. Encodes a transmembrane domain. Important in megakaryocyte and platelet formation. |
| Up in hybrid TOV21G | *WDR78* | 1p31.2 | 51 | | 7 | | 112,002 | 77 | 23 | Unknown function |
| Up in hybrid TOV21G | *PRDM14* | 8p21-p12 | 65 | | 6 | | 19,542 | 16 | 14 | PRDM14 mRNA is overexpressed in about 2/3 of breast cancers; moreover, immunohistochemical analysis showed that expression of PRDM14 protein is also up-regulated. Regulation of transcription. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrid TOV21G | *PNOC* | 8p21 | 70 | | 10 | | 26,219 | 49 | 12 | Prepronociceptin. Protein in part of the neuropeptide signalling pathway. May be involved in neuronal differentiation and development. |
| Up in hybrid TOV21G | *LOC400942* | 2p25.1 | 84 | | 3 | | | Not in Hapmap | | Hypothetical protein. Unknown function. |
| Up in hybrid TOV21G | *C14orf110* | 14q32.33 | 114 | | 4 | | 4,163 | Not in Hapmap | | Hypothetical protein. Unknown function. |
| Up in hybrid TOV21G | *LOC283677* | 15q24.1 | 120 | | 1 | | 116,832 | 32 | 9 | Hypothetical protein. Unknown function. |
| Down in hybrid TOV21G | *HIF1A* | 14q21-q24 | 1 | | 12 | | 52,737 | 16 | 6 | Hypoxia-inducible factor 1. Transcription factor. |
| Down in hybrid TOV21G | *IPO7* | 11p15.4 | 2 | | 330 | | 60,871 | 20 | 6 | Importin 7. RAN binding protein 7. The importin-α/β complex and the GTPase Ran mediate nuclear import of proteins. |
| Down in hybrid TOV21G | *KIAA0895* | 7p14.1 | 3 | | 48 | | 65,863 | | | Hypothetical protein. Unknown function. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrid TOV21G | *HPS6* | 10q24.32 | 4 | | 183 | | 2,648 | 2 | 2 | Hermansky-Pudlak syndrome 6. This intronless gene encodes a protein that may play a role in organelle biogenesis associated with melanosomes, platelet dense granules, and lysosomes. |
| Down in hybrid TOV21G | *FBXW5* | 9q34.3 | 5 | | 28 | | 4,286 | 1 | 1 | F-box and WD repeat domain containing 5. The F-box proteins constitute one of the four subunits of ubiquitin protein ligase complex, which function in phosphorylation-dependent ubiquitination. |
| Down in hybrid TOV21G | *APOBEC3C* | 22q13.1-q13.2 | 6 | | 3 | | 18,110 | 6 | 3 | Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C. May be RNA editing enzymes and have roles in growth or cell cycle control. |
| Down in hybrid TOV21G | *BCAR1* | 16q22-q23 | 7 | | 64 | | 22,575 | 17 | 7 | Breast cancer anti-estrogen resistance 1. Docking protein which plays a central coordinating role for signalling related to cell adhesion. Implicated in induction of cell migration. Overexpression confers antiestrogen resistance on breast cancer cells. |
| Down in hybrid TOV21G | *CTPS2* | Xp22 | 8 | | 182 | | 124,937 | 31 | 10 | Cytidine 5'-triphosphate synthetase 2. Cancer cells that exhibit increased cell proliferation also exhibit an increased activity of this encoded |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | protein. |
| Down in hybrid TOV21G | *LOC130951* | 2p13.1 | 9 | | 286 | | | | | Hypothetical protein. Unknown function |
| Down in hybrid TOV21G | *NMT1* | 17q21.31 | 10 | | 34 | | 47,704 | 41 | 10 | N-myristoyltransferase 1. Adds a myristoyl group to the N-terminal glycine residue of certain cellular and viral proteins |
| Down in hybrid TOV21G | *DDIT4* | 10pter-q26.12 | 13 | | 10 | | 2,120 | 1 | 1 | DNA-damage-inducible transcript 4. |
| Down in hybrid TOV21G | *SQSTM1* | 5q35 | 44 | | 9 | | 17,181 | 15 | 10 | Sequestosome 1. Paget disease of bone 3. May be involved in cell differentiation, apoptosis, immune response and regulation of K(+) channels. |
| Down in hybrid TOV21G | *PDZK1IP1* | 1p33 | 77 | | 6 | | 7,455 | 1 | 1 | PDZK1 interacting protein 1. Epithelial protein up-regulated in carcinoma. |
| Down in hybrid TOV21G | *TUBA1* | 2q35 | 106 | | 1 | | 4,205 | 4 | 1 | Encodes an α-tubulin, a major component of microtubules. Microtubules of the eukaryotic cytoskeleton perform essential and diverse functions. Highly conserved. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrid TOV21G | *IL6* | 7p21 | 151 | | 8 | | 6,113 | 10 | 4 | Interleukin 6. Involved in the regulation of immune response. |
| Down in hybrid TOV21G | *NNMT* | 11q23.1 | 236 | | 7 | | 54,685 | 13 | 7 | Nicotinamide N-methyltransferase. Protein is involved in the metabolism of drugs and xenobiotic compounds by the liver. |
| Down in hybrid TOV21G | *CXCL1* | 4q21 | 261 | | 4 | | 14,276 | 1 | 1 | Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha). Oncogene involved in regulation of cell trafficking of leukocytes. Also play fundamental roles in the development, homeostasis, and function of the immune system, and have effects on cells of the central nervous system as well as on endothelial cells involved in angiogenesis or angiostasis. |
| Down in hybrid TOV21G | *PPP1CA* | 11q13 | 332 | | 5 | | 3,750 | 1 | 1 | Protein phosphatase 1, catalytic subunit, alpha isoform. Encoded protein is one of the three catalytic subunits of protein phosphatase 1 (PP1). PP1 is a serine/threonine specific protein phosphatase known to be involved in the regulation of a variety of cellular processes, such as cell division, glycogen metabolism, muscle contractility, protein |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | synthesis, and HIV-1 viral transcription. |
| Down in hybrid TOV21G | *SPP1* | 4q21-q25 | 456 | | 2 | | 7,766 | 10 | 4 | Secreted phosphoprotein 1. May be involved in cell-matrix interaction. Sequence suggests that the protein acts as a cytokine involved in enhancing production of interferon-γ and interleukin-12 and reducing production of interleukin-10 and is essential in the pathway that leads to type I immunity. |
| Up in hybrids both lines | *RARB* | 3p24 | 11 | 22 | 25 | 2 | 423,531 | 245 | 102 | Retinoic acid receptor-β. This receptor controls cell function by directly regulating gene expression. |
| Up in hybrids both lines | *PRO1843* | 12q13.13 | 42 | 9 | 9 | 3 | 35,770 | 40 | 8 | Eukaryotic translation initiation factor 4B. Required for the binding of mRNA to ribosomes. |
| Up in hybrids both lines | *HEY1* | 8q21 | 15 | 16 | 16 | 4 | 3,760 | 2 | 2 | Hairy/enhancer-of-split related with YRPW motif 1. Transcriptional repressors. Implicated in Notch signaling pathway; and nervous system development. |
| Up in hybrids both lines | *AKTIP (FTS)* | 16q12.2 | 18 | 14 | 15 | 6 | 11,978 | 7 | 4 | AKT interacting protein. Fused toes homolog. Regulates apoptosis. This protein interacts directly with serine/threonine kinase protein kinase B (PKB)/Akt and modulates |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | PKB activity by enhancing the phosphorylation of PKB's regulatory sites. |
| Up in hybrids both lines | *IGHM* | 14q32.33 | 43 | 46 | 28 | 7 | 4,285 | Not in hapmap | | Immunoglobulin heavy constant mu. Protein has transmembrane receptor activity, which is implicated in activation of MAPK activity. |
| Up in hybrids both lines | *NMB* | 15q22-qter | 44 | 24 | 11 | 8 | 3,442 | 7 | 4 | neuromedin B. Stimulates smooth muscle contraction in a manner similar to that of bombesin. hormone activity. signal transduction. neuropeptide signaling pathway. cell-cell signalling. |
| Up in hybrids both lines | *MRPL46* | 15q24-q25 | 12 | 23 | 44 | 9 | 7,910 | 1 | 1 | Encoded a subunit of mammalian mitochondrial ribosomal proteins, which help in protein synthesis within the mitochondrion. |
| Up in hybrids both lines | *SERINC2* | 1p35.1 | 19 | 27 | 22 | 10 | 25,113 | 8 | 3 | Serine incorporator 2. Positive regulation of transferase activity. |
| Up in hybrids both lines | *PCDHB2* | 5q31 | 2 | 29 | 19 | 12 | 2,745 | 1 | 1 | Protocadherin β 2. Specific functions are unknown but they most likely play a critical role in the establishment and function of specific cell-cell neural connections. |
| Up in hybrids both lines | *KRT8L2* | 3q25.33 | 39 | 15 | 4 | 13 | 1,688 | Not on hapmap | | keratin 8 pseudogene 12. Unknown function. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Up in hybrids both lines | *SP110* | 2q37.1 | 8 | 20 | 33 | 14 | 51,036 | 87 | 29 | SP110 nuclear body protein. The protein can function as an activator of gene transcription and may serve as a nuclear hormone receptor coactivator. May also be involved in ribosome biogenesis and induction of myeloid cell differentiation. |
| Up in hybrids both lines | *EPS15L2* | 7p12.3 | 16 | 25 | 18 | 15 | 2,046 | Not on hapmap | | Epidermal growth factor receptor pathway substrate 15-like 2. Pseudo gene. Unknown function. |
| Up in hybrids both lines | *EPHX2* | 8p21-p12 | 7 | 35 | 24 | 16 | 53,860 | 50 | 11 | Epoxide hydrolase 2, cytoplasmic. Plays role in xenobiotic metabolism by degrading potentially toxic epoxides. |
| Up in hybrids both lines | *LOC92689* | 4p14 | 40 | 21 | 29 | 18 | 77,923 | 100 | 24 | Family with sequence similarity 114, member A1. Nervous system over-expressed protein. May play a role in neuronal cell development (By similarity). Hypothetical protein. |
| Up in hybrids both lines | *SRGAP1* | 12q14.2 | 25 | 58 | 14 | 19 | 299,033 | 170 | 60 | SLIT-ROBO Rho GTPase activating protein 1. Together with CDC42 seems to be involved in the pathway mediating the repulsive signalling of Robo and Slit proteins in neuronal migration. |
| Up in hybrids both lines | *FLJ10826* | 16q12.2 | 28 | 1 | 57 | 20 | 27,588 | | | 2-oxoglutarate and iron-dependent oxygenase domain containing 1. Hypothetical protein. Unknown |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | function. |
| Up in hybrids both lines | *SLC38A6* | 14q23.1 | 1 | 18 | 20 | 21 | 102,550 | 22 | 5 | Solute carrier family 38, member 6. |
| Up in hybrids both lines | *AIFM2 (AMID)* | 10q22.1 | 10 | 11 | 27 | 23 | 34,711 | 17 | 13 | Apoptosis-inducing factor, mitochondrion-associated, 2. Induction of apoptosis. Induced by tumour suppressor protein p53 in colon cancer cells. Down-regulated in a wide range of human tumours. |
| Up in hybrids both lines | *PLXNA1* | 3q21.3 | 38 | 51 | 2 | 24 | 48,730 | 16 | 9 | Plexin A1. Plays a role in axon guidance, invasive growth and cell migration. |
| Up in hybrids both lines | *LHX6* | 9q33.2 | 35 | 52 | 3 | 26 | 26,221 | 21 | 12 | LIM homeobox 6. The encoded protein may function as a transcriptional regulator and may be involved in the control of differentiation and development of neural and lymphoid cells. |
| Up in hybrids both lines | *IL4* | 5q31.1 | 26 | 47 | 8 | 28 | 8,995 | 15 | 3 | Interleukin 4. Regulation of B cell proliferation. |
| Up in hybrids both lines | *MAPT* | 17q21.1 | 3 | 13 | 17 | 30 | 133,923 | 188 | 27 | Microtubule-associated protein tau. Promotes microtubule assembly and stability, and might be involved in the establishment and maintenance |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | of neuronal polarity. |
| Up in hybrids both lines | *C20orf6* | 20p12.1 | 30 | 12 | 39 | 34 | 70,563 | 70 | 14 | ESF1, nucleolar pre-rRNA processing protein, homolog (S. cerevisiae). May constitute a novel regulatory system for basal transcription. Negatively regulates ABT1 (By similarity). Transcription. regulation of transcription, DNA-dependent. |
| Up in hybrids both lines | *FLJ22662* | 12p13.1 | 6 | 2 | 30 | 35 | 64,196 | 41 | 13 | Hypothetical protein. Unknown function. |
| Up in hybrids both lines | *TMEM45A* | 3q12.2 | 27 | 19 | 10 | 51 | 84,822 | 77 | 14 | Transmembrane protein 45A. Unknown function |
| Down in hybrids both cell lines | *TPM3* | 1q21.2 | 19 | 87 | 7 | 9 | 35,776 | 11 | 6 | Tropomyosin 3. Gene encodes a member of the tropomyosin family of actin-binding proteins involved in the contractile system of striated and smooth muscles and the cytoskeleton of non-muscle cells. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrids both cell lines | *SFRS9* | 12q24.31 | 91 | 68 | 8 | 11 | 8,087 | 5 | 3 | Splicing factor, arginine/serine-rich 9. Plays a role in constitutive splicing and can modulate the selection of alternative splice sites. |
| Down in hybrids both cell lines | *RUVBL1* | 3q21 | 8 | 7 | 10 | 67 | 42,857 | 29 | 7 | Interacts with MYC. Forms a complex which may be required for the activation of transcriptional programs associated with oncogene and proto-oncogene mediated growth induction, tumor suppressor mediated growth arrest and replicative senescence, apoptosis, and DNA repair. |
| Down in hybrids both cell lines | *GAMT* | 19p13.3 | 15 | 124 | 14 | 5 | 4,464 | 0 | 0 | Guanidinoacetate N-methyltransferase. Converts guanidoacetate to creatine. Important in creatine biosynthetic process. |
| Down in hybrids both cell lines | *DNAJB1* | 19p13.2 | 7 | 71 | 18 | 29 | 3,619 | 6 | 4 | DnaJ (Hsp40) homolog, subfamily B, member 1. Interacts with HSP70 and can stimulate its ATPase activity. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Down in hybrids both cell lines | *SIP1* | 14q13 | 26 | 9 | 26 | 75 | 22,689 | 10 | 4 | Survival of motor neuron protein interacting protein 1. The SMN complex plays an essential role in spliceosomal snRNP assembly in the cytoplasm and is required for pre-mRNA splicing in the nucleus. |
| Down in hybrids both cell lines | *DKFZP686A10121* | 7q21.13 | 20 | 26 | 59 | 79 | 44,329 | 70 | 10 | GTP-binding protein 10 (putative). Hypothetical protein. Unknown function. |
| Down in hybrids both cell lines | *BTG3* | 21q21.1-q21.2 | 21 | 23 | 104 | 64 | 19,294 | 5 | 3 | BTG family, member 3. Regulation of progression through mitotic cell cycle. Putatively involved in neurogenesis in the central nervous system. |
| Down in hybrids both cell lines | *CTSL* | 9q21-q22 | 16 | 18 | 105 | 66 | 5,874 | 3 | 2 | Cathepsin L1. Encoded protein plays a major role in intracellular protein catabolism. |
| CHR18 TOV21G down in hybrids | *ZNF532* | 18q21.32 | 1 | | 21 | | 123,648 | 73 | 28 | Zinc finger protein 532. Nucleic acid binding activity. Transcription activity. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 TOV21G down in hybrids | *ACAA2* | 18q21.1 | 2 | | 3 | | 30,376 | 26 | 8 | Acetyl-Coenzyme A acyltransferase 2. Catalyses the last step of the mitochondrial fatty acid beta-oxidation spiral. |
| CHR18 TOV21G down in hybrids | *RALBP1* | 18p11.3 | 3 | | 6 | | 63,106 | 47 | 15 | RalA binding protein 1. Can catalyse transport of glutathione conjugates and xenobiotics, and may contribute to the multidrug resistance phenomenon. Serves as a scaffold protein that brings together proteins forming an endocytotic complex during interphase and also with CDC2 to switch off endocytosis. |
| CHR18 TOV21G down in hybrids | *RAB31* | 18p11.3 | 4 | | 15 | | 154,284 | 174 | 85 | RAB31, member RAS oncogene family. Predominantly expressed in melanocytes. Signal transduction. Small GTPase mediated signal transduction. |
| CHR18 TOV21G down in hybrids | *C18orf55* | 18q22.3 | 5 | | 8 | | 10,445 | 14 | 3 | TIM21-like protein, mitochondrial precursor. May participate in the translocation of transit peptide-containing proteins across the mitochondrial inner membrane (By similarity). Hypothetical protein. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 TOV21G down in hybrids | *KNTC2* | 18p11.32 | 6 | | 9 | | 45,055 | 63 | 11 | NDC80 homolog, kinetochore complex component (S. cerevisiae). Mitotic sister chromatid segregation. Spindle organization and biogenesis. Vesicle-mediated transport. Phosphoinositide-mediated signalling. |
| CHR18 TOV21G down in hybrids | *CCDC5* | 18q21.1 | 7 | | 27 | | 24,002 | 12 | 6 | Coiled-coil domain containing 5 (spindle associated). Regulator of spindle function and integrity during the metaphase-anaphase transition. |
| CHR18 TOV21G down in hybrids | *LAMA1* | 18p11.31 | 8 | | 10 | | 175,929 | 246 | 104 | Laminin, α 1. Mediates the attachment, migration and organization of cells into tissues during embryonic development. |
| CHR18 TOV21G down in hybrids | *ATP5A1* | 18q12-q21 | 9 | | 1 | | 20,090 | 15 | 4 | ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle. This gene encodes a subunit of mitochondrial ATP synthase. Involved in ATP biosynthetic process. |
| CHR18 TOV21G down in hybrids | *DYM* | 18q12-q21.1 | 10 | | 7 | | 416,908 | 237 | 47 | Dymeclin. Protein is involved in normal skeletal development and brain function. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 TOV21G down in hybrids | *ENOSF1* | 18p11.32 | 11 | | 25 | | 38,729 | 68 | 16 | Enolase superfamily member 1. catalytic activity. transferase activity. |
| CHR18 TOV21G down in hybrids | *PQLC1* | 18q23 | 12 | | 22 | | 49,131 | 22 | 9 | PQ loop repeat containing 1. Unknown function. |
| CHR18 TOV21G down in hybrids | *LPIN2* | 18p11.31 | 13 | | 14 | | 94,954 | 81 | 21 | Lipin 2. Unknown function. |
| CHR18 TOV21G down in hybrids | *SMAD4* | 18q21.1 | 14 | | 32 | | 54,803 | 20 | 3 | SMAD family member 4. Common mediator of signal transduction by TGF-beta (transforming growth factor) superfamily; May act as a tumor suppressor. Negative regulation of cell proliferation. |
| CHR18 TOV21G down in hybrids | *DSG2* | 18q12.1 | 15 | | 12 | | 48,686 | 51 | 9 | Desmoglein 2. Component of intercellular desmosome junctions. Involved in the interaction of plaque proteins and intermediate filaments mediating cell-cell adhesion. |
| CHR18 TOV21G down in hybrids | *MRLC2* | 18p11.31 | 16 | | 2 | | 16,160 | 12 | 3 | Myosin regulatory light chain MRLC2. Plays an important role in regulation of both smooth muscle and non-muscle cell contractile activity. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 TOV21G down in hybrids | *RPL17* | 18q21 | 19 | | 13 | | 4,053 | 5 | 4 | Ribosomal protein L17. This gene encodes a ribosomal protein that is a component of the 60S subunit. Amino acid translation. |
| CHR18 TOV21G down in hybrids | *NDUFV2* | 18p11.31-p11.2 | 21 | | 4 | | 31,632 | 22 | 7 | NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa. NADH dehydrogenase (ubiquinone) activity. |
| CHR18 TOV21G down in hybrids | *C18orf10* | 18q12.2 | 26 | | 11 | | 33,125 | 28 | 8 | Hypothetical protein. Unknown function |
| CHR18 TOV21G down in hybrids | *RPL17* | 18q21 | 27 | | 5 | | | | | Same as above |
| CHR18 TOV21G up in hybrids | *SDCCAG33* | 18q22.3 | 1 | | 19 | | 79,171 | 79 | 30 | Tee-shirt zinc finger homeobox 1. May be involved in transcriptional regulation of developmental processes. |
| CHR18 TOV21G up in hybrids | *C18orf22* | 18q23 | 2 | | 17 | | 12,035 | 12 | 5 | rRNA processing. Hypothetical protein. Unknown function. |
| CHR18 TOV21G up in hybrids | *TCF4* | 18q21.1 | 3 | | 1 | | 408,217 | 183 | 62 | Transcription factor-4 (immunoglobulin transcription factor-2). Transcription factor that |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | binds to the immunoglobulin enchancer Mu-E5/KE5-motif. |
| CHR18 TOV21G up in hybrids | *KIAA1632* | 18q12.3 | 4 | | 6 | | 119,732 | 87 | 14 | Hypothetical protein. Unknown function |
| CHR18 TOV21G up in hybrids | *KIAA1012* | 18q12.1 | 5 | | 21 | | 113,167 | 66 | 11 | May play a role in vesicular transport from endoplasmic reticulum to Golgi. |
| CHR18 TOV21G up in hybrids | *MYO5B* | 18q21 | 6 | | 7 | | 372,284 | 519 | 108 | Motor activity. Actin, calmodulin and nucleotide binding protein. |
| CHR18 TOV21G up in hybrids | *EMILIN2* | 18p11.3 | 7 | | 5 | | 67,063 | 57 | 39 | Elastin microfibril interfacer 2. May be responsible for anchoring smooth muscle cells to elastic fibres, formation of elastic fibres and vessel assembly regulation. |
| CHR18 TOV21G up in hybrids | *RALBP1* | 18p11.3 | 8 | | 9 | | | | | Same as above |
| CHR18 TOV21G up in hybrids | *CYB5* | 18q23 | 9 | | 24 | | 38,695 | 52 | 9 | Cytochrome b5 type A (microsomal). A membrane bound hemoprotein which functions as an electron carrier |
| CHR18 TOV21G up in hybrids | *RALBP1* | 18p11.3 | 10 | | 22 | | | | | Same as above |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 TOV21G up in hybrids | *ZNF396* | 18q12 | 11 | | 16 | | 10,641 | 13 | 4 | Zinc finger protein 396. Isoforms 1 and 2 act as DNA-dependent transcriptional repressors. |
| CHR18 TOV21G up in hybrids | *C18orf12* | 18q21.1 | 12 | | 10 | | Unknown | Not on hapmap | | Hypothetical protein. Unknown function. |
| CHR18 TOV21G up in hybrids | *BCL2* | 18q21.33\|18q21.3 | 13 | | 12 | | 196,783 | 189 | 66 | B-cell CLL/lymphoma 2. Regulation of progression through cell cycle and apoptosis. |
| CHR18 TOV21G up in hybrids | *ONECUT2* | 18q21.1-q21.2 | 14 | | 3 | | 55,613 | 67 | 13 | One cut homeobox 2. Transcriptional activator of target genes, which include genes involved in melanocyte and hepatocyte differentiation. |
| CHR18 TOV21G up in hybrids | *EPB41L3* | 18p11.32 | 15 | | 2 | | 238,253 | 115 | 32 | Erythrocyte membrane protein band 4.1-like 3. Differentially expressed in adenocarcinoma of the lung. Critical growth regulator in the pathogenesis of meningiomas. |
| CHR18 TOV21G up in hybrids | *NAPG* | 18p11.22 | 16 | | 11 | | 24,342 | 26 | 3 | N-ethylmaleimide-sensitive factor attachment protein γ. Required for vesicular transport between the endoplasmic reticulum and the Golgi apparatus. |
| CHR18 TOV21G up in hybrids | *TWSG1* | 18p11.3 | 18 | | 15 | | 67,568 | 46 | 12 | Twisted gastrulation homolog 1. May be involved in dorsoventral axis formation. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 TOV21G up in hybrids | *LIPG* | 18q21.1 | 19 | | 4 | | 30,852 | 17 | 8 | Lipase, endothelial. The protein encoded by this gene has substantial phospholipase activity and may be involved in lipoprotein metabolism and vascular biology. |
| CHR18 TOV21G up in hybrids | *SLC14A2* | 18q12.1-q21.1 | 21 | | 8 | | 68,307 | 88 | 29 | Solute carrier family 14 (urea transporter), member 2. Mediates urea transport in kidney. |
| CHR18 TOV21G up in hybrids | *FBXO15* | 18q22.3 | 22 | | 13 | | 74,345 | 81 | 9 | F-box protein 15. Involved in the ubiquitin cycle. |
| CHR18 TOV21G up in hybrids | *FAM59A* | 18q12.1 | 23 | | 14 | | 202,985 | 165 | 34 | Family with sequence similarity 59, member A. Unknown function. |
| CHR18 Up in hybrids TOV112D | *C18orf34* | 18q12.1 | | 1 | | 2 | 503,097 | 334 | 34 | Hypothetical protein. DNA unwinding during replication (by similarity. DNA topological change |
| CHR18 Up in hybrids TOV112D | *SLC39A6* | 18q12.2 | | 2 | | 7 | 20,007 | 29 | 7 | Solute carrier family 39 (zinc transporter), member 6. Zinc ion transporter. |
| CHR18 Up in hybrids TOV112D | *NOL4* | 18q12 | | 3 | | 5 | 372,458 | 210 | 66 | Nucleolar protein 4. Transition metal ion binding activity. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 Up in hybrids TOV112D | *KNTC2* | 18p11.32 | | 4 | | 4 | | | | Same as above |
| CHR18 Up in hybrids TOV112D | *IMPACT* | 18q11.2-q12.1 | | 5 | | 18 | 26,811 | 29 | 8 | Impact homolog (mouse). Involved in ubiquitin cycle and protein modification process. |
| CHR18 Up in hybrids TOV112D | *CCDC5* | 18q21.1 | | 6 | | 16 | | | | Same as above |
| CHR18 Up in hybrids TOV112D | *C18orf51* | 18q22.3 | | 7 | | 1 | 22,217 | 35 | 14 | Hypothetical protein. Unknown function |
| CHR18 Up in hybrids TOV112D | *LOC390773\|RPL17* | 18q21 | | 8 | | 22 | | | | Same as above |
| CHR18 Up in hybrids TOV112D | *KIAA0863* | 18q23 | | 9 | | 17 | 31,318 | 34 | 6 | ADNP homeobox 2. May be involved in transcriptional regulation. |
| CHR18 Up in hybrids TOV112D | *NDUFV2* | 18p11.31-p11.2 | | 10 | | 8 | | | | Same as above |
| CHR18 Up in hybrids TOV112D | *PTPN2* | 18p11.3-p11.2 | | 11 | | 14 | 98,855 | 48 | 13 | Protein tyrosine phosphatase, non-receptor type 2. The protein encoded is member of the protein |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | tyrosine phosphatase (PTP) family. PTPs are known to regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. |
| CHR18 Up in hybrids TOV112D | *RBBP8* | 18q11.2 | | 12 | | 3 | 93,155 | 39 | 4 | Retinoblastoma binding protein 8. The protein a ubiquitously expressed nuclear protein. Found among several proteins that bind directly to retinoblastoma protein, which regulates cell proliferation. This protein complexes with transcriptional co-repressor CTBP. It is also associated with BRCA1 and is thought to modulate the functions of BRCA1 in transcriptional regulation, DNA repair, and/or cell cycle checkpoint control. It is suggested that this gene may itself be a tumour suppressor acting in the same pathway as BRCA1. |
| CHR18 Up in hybrids TOV112D | *RAB31* | 18p11.3 | | 13 | | 10 | | | | Same as above |
| CHR18 Up in hybrids TOV112D | *RPL17* | 18q21 | | 14 | | 6 | | | | Same as above |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 Up in hybrids TOV112D | *C18orf49* | 18q21.33 | | 15 | | 13 | unknown | Not in hapmap | | Hypothetical protein. Unknown function. |
| CHR18 Up in hybrids TOV112D | *MIB1* | 18q11.2 | | 17 | | 15 | 129,369 | 31 | 10 | Mindbomb homolog 1. Regulates the Delta-mediated Notch signaling by ubiquitinating the intracellular domain of Delta, leading to endocytosis of Delta receptors. |
| CHR18 Up in hybrids TOV112D | *MRLC2* | 18p11.31 | | 20 | | 9 | | | | Same as above |
| CHR18 Up in hybrids TOV112D | *OSBPL1A* | 18q11.1 | | 21 | | 12 | 235,782 | 158 | 69 | Oxysterol binding protein-like 1A. This gene encodes a member of the oxysterol-binding protein (OSBP) family, a group of intracellular lipid receptors. |
| CHR18 Up in hybrids TOV112D | *DYM* | 18q12-q21.1 | | 23 | | 11 | | | | Same as above |
| CHR18 Down in hybrids TOV112D | *CDH2* | 18q11.2 | | 1 | | 1 | 226,257 | 303 | 59 | Cadherin 2, type 1, N-cadherin (neuronal). Encoded protein is a calcium dependent cell-cell adhesion glycoprotein. The protein functions during gastrulation and is required for establishment of left-right asymmetry. May be involved in neuronal recognition mechanism. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 Down in hybrids TOV112D | *COLEC12* | 18pter-p11.3 | | 5 | | 2 | 181,334 | 288 | 127 | Protein is a scavenger receptor, a cell surface glycoprotein that can bind to carbohydrate antigens on microorganisms facilitating their recognition and removal. May also participate in removing oxidatively damaged or apoptotic cells. phosphate transport. |
| CHR18 Down in hybrids TOV112D | *APCDD1* | 18p11.22 | | 3 | | 3 | 34,070 | 55 | 16 | Adenomatosis polyposis coli down-regulated 1. May play a role in colorectal tumorigenesis. May be a developmental target gene of the Wnt/β-catenin pathway |
| CHR18 Down in hybrids TOV112D | *FAM38B* | 18p11.22 | | 4 | | 4 | 26,965 | 39 | 13 | Family with sequence similarity 38, member B. Unknown function |
| CHR18 Down in hybrids TOV112D | *EPB41L3* | 18p11.32 | | 11 | | 5 | | | | Same as above |
| CHR18 Down in hybrids TOV112D | *LOC284214* | 18p11.31 | | 19 | | 6 | unknown | Not in hapmap | | Hypothetical protein. Unknown function |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 Down in hybrids TOV112D | *FHOD3* | 18q12 | | 27 | | 7 | 482,317 | 527 | 141 | Formin homology 2 domain containing 3. Involved in actin cytoskeleton organization and biogenesis. |
| CHR18 Down in hybrids TOV112D | *KATNAL2* | 18q21.1 | | 14 | | 8 | 100,877 | 51 | 14 | Katanin p60 subunit A-like 2. ATP binding activity. |
| CHR18 Down in hybrids TOV112D | *EPB41L3* | 18p11.32 | | 20 | | 9 | | | | Same as above |
| CHR18 Down in hybrids TOV112D | *ZNF532* | 18q21.32 | | 40 | | 10 | | | | Same as above |
| CHR18 Down in hybrids TOV112D | *PMAIP1* | 18q21.32 | | 16 | | 11 | 4,301 | 3 | 3 | Phorbol-12-myristate-13-acetate-induced protein 1. Adult T cell leukemia-derived PMA-responsive. |
| CHR18 Down in hybrids TOV112D | *P15RS* | 18q12.2 | | 24 | | 12 | 77,729 | 41 | 10 | Cyclin-dependent kinase 2B-inhibitor-related protein. May act as a negative regulator of cyclin D1 (CCND1) and cyclin E (CCNE1) in the cell cycle. Up-regulated in cells overexpressing CDKN2B. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 Down in hybrids TOV112D | *SMCHD1* | 18p11.32 | | 15 | | 13 | 148,281 | Not on Hapmap | | Structural maintenance of chromosomes flexible hinge domain containing 1. |
| CHR18 Down in hybrids TOV112D | *C18orf17* | 18q11.2 | | 33 | | 14 | 140,255 | 38 | 20 | Hypothetical protein. Unknown function |
| CHR18 Down in hybrids TOV112D | *NEDD4L* | 18q21 | | 2 | | 15 | 353,592 | 446 | 116 | Neural precursor cell expressed, developmentally down-regulated 4-like. Unknown function |
| CHR18 Down in hybrids TOV112D | *DSC3* | 18q12.1 | | 13 | | 19 | 51,659 | 84 | 27 | Desmocollin 3. The protein encoded by this gene is a calcium-dependent glycoprotein. Found primarily in epithelial cells, they are required for cell adhesion and desmosome formation. |
| CHR18 Down in hybrids TOV112D | *PHLPP* | 18q21.33 | | 7 | | 23 | 264,933 | 146 | 46 | PH domain and leucine rich repeat protein phosphatise. Protein regulates the balance between cell survival and apoptosis. May act as a negative regulator of K-Ras signalling in the membrane rafts. |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| CHR18 Down in hybrids TOV112D | *TGIF* | 18p11.3 | | 8 | | 24 | 46,338 | 37 | 19 | TGFB-induced factor homeobox 1. The protein is an active transcriptional co-repressor of SMAD2 and may participate in the transmission of nuclear signals during development and in the adult. Negative regulation of transcription from RNA polymerase II promoter. |
| CHR18 Down in hybrids TOV112D | *YES1* | 18p11.31-p11.21 | | 10 | | 27 | 90,740 | 52 | 17 | V-yes-1 Yamaguchi sarcoma viral oncogene homolog 1. This gene is the cellular homolog of the Yamaguchi sarcoma virus oncogene. The encoded protein has tyrosine kinase activity and belongs to the src family of proteins. |
| CHR18 Down in hybrids TOV112D | *YES1* | 18p11.31-p11.21 | | 6 | | 33 | | | | Same as above |
| CHR18 Down in hybrids TOV112D | *LOC441805* | 18p11.32 | | 9 | | 40 | | | | Gene record has been discontinued. |
| CHR18 Down in hybrids TOV112D | *PTPN2* | 18p11.3-p11.2 | | 12 | | 51 | | | | Same as above |

| List | Gene | Cytoband | 21G pval hybrid rank | 112D pval hybrid rank | 21G FC* hybrid rank | 112D FC* hybrid rank | Gene Size (bp) | No. Criteria SNPs§ | No. tSNPs | Function |
|---|---|---|---|---|---|---|---|---|---|---|
| Breakpoint region TOV112D | *OSBPL1A* | 18q11.1 | | 21 | | 12 | | | | Same as above |
| Breakpoint region TOV112D | *RBBP8* | 18q11.2 | | 12 | | 3 | 93,154 | 40 | 4 | Same as above |
| Breakpoint region TOV112D | *SNRPD1* | 18q11.2 | | 16 | | 20 | 18,108 | 5 | 1 | Small nuclear ribonucleoprotein D1 polypeptide 16kDa. Sm-D autoantigen. Protein may act as a charged protein scaffold to promote SNRNP assembly or strengthen SNRNP-SNRNP interactions with RNA. |
| Breakpoint region TOV112D | *MIB1* | 18q11.2 | | 17 | | 15 | 129,369 | 31 | 10 | Same as above. |
| Breakpoint region TOV112D | *ZNF521* | 18q11.2 | | 19 | | 19 | 290,226 | 220 | 65 | Zinc finger protein 521. Transcription factor involved in regulation of transcription. |
| Breakpoint region TOV112D | *IMPACT* | 18q11.2-q12.1 | | 5 | | 18 | 26,810 | 29 | 8 | Highly conserved. Unknown function |

* FC - expresion fold change between hybrids and parental; § Criteria SNPs - minor allele frequency ≥ 0.05; Hardy-Weinberg ≥ 0.01
Ranks of differential gene expression of hybrid clones over parental expression: 21G – TOV21G, 112D – TOV112D, pval – P-value,

## Appendix II-A: Genotype distributions of tagging SNPs in *BRAF*, *ERBB2*, *KRAS*, *NMI* and *PIK3CA* (by study)

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| *BRAF* | rs10487888 | GEOCS | 88 | 157 | 72 | **317** | 98 | 205 | 114 | **417** |
| | | MALOVA | 92 | 219 | 132 | **443** | 274 | 582 | 352 | **1208** |
| | | SEARCH | 124 | 372 | 221 | **717** | 173 | 419 | 262 | **854** |
| | | USC | 57 | 86 | 50 | **193** | 67 | 90 | 58 | **215** |
| *BRAF* | rs1733832 | MALOVA | 377 | 58 | 3 | **438** | 1047 | 150 | 3 | **1200** |
| | | SEARCH | 629 | 78 | 4 | **711** | 754 | 88 | 1 | **843** |
| *BRAF* | rs1267622 | GEOCS | 174 | 111 | 26 | **311** | 250 | 141 | 21 | **412** |
| | | MALOVA | 268 | 151 | 17 | **436** | 681 | 427 | 75 | **1183** |
| | | SEARCH | 416 | 258 | 34 | **708** | 497 | 312 | 43 | **852** |
| | | UKOPS | 54 | 35 | 8 | **97** | 129 | 82 | 13 | **224** |
| | | USC | 106 | 64 | 18 | **188** | 117 | 69 | 23 | **209** |
| *BRAF* | rs13241719 | GEOCS | 166 | 124 | 26 | **316** | 202 | 178 | 39 | **419** |
| | | MALOVA | 126 | 120 | 28 | **274** | 370 | 322 | 73 | **765** |
| | | SEARCH | 316 | 333 | 63 | **712** | 374 | 370 | 108 | **852** |
| | | UKOPS | 54 | 31 | 15 | **100** | 112 | 101 | 25 | **238** |
| | | USC | 106 | 68 | 16 | **190** | 118 | 80 | 16 | **214** |
| *BRAF* | rs17695623 | GEOCS | 282 | 30 | 3 | **315** | 364 | 42 | 2 | **408** |
| | | MALOVA | 367 | 53 | 1 | **421** | 1024 | 165 | 5 | **1194** |
| | | SEARCH | 598 | 99 | 3 | **700** | 713 | 105 | 7 | **825** |
| | | UKOPS | 86 | 17 | 0 | **103** | 216 | 36 | 0 | **252** |
| | | USC | 176 | 15 | 4 | **195** | 191 | 30 | 1 | **222** |
| *BRAF* | rs17161747 | GEOCS | 282 | 34 | 4 | **320** | 373 | 41 | 5 | **419** |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | MALOVA | 371 | 64 | 4 | **439** | 1065 | 132 | 0 | **1197** |
| | | SEARCH | 647 | 66 | 0 | **713** | 771 | 76 | 4 | **851** |
| | | UKOPS | 92 | 5 | 1 | **98** | 215 | 18 | 2 | **235** |
| | | USC | 173 | 15 | 2 | **190** | 187 | 18 | 2 | **207** |
| | | GEOCS | 239 | 71 | 12 | **322** | 326 | 84 | 8 | **418** |
| | | MALOVA | 322 | 111 | 7 | **440** | 884 | 282 | 23 | **1189** |
| *BRAF* | rs10281173 | SEARCH | 526 | 175 | 10 | **711** | 651 | 185 | 17 | **853** |
| | | UKOPS | 65 | 26 | 3 | **94** | 169 | 50 | 6 | **225** |
| | | USC | 140 | 41 | 7 | **188** | 155 | 57 | 8 | **220** |
| | | GEOCS | 255 | 62 | 3 | **320** | 318 | 98 | 5 | **421** |
| | | MALOVA | 331 | 89 | 12 | **432** | 917 | 234 | 17 | **1168** |
| *BRAF* | rs17623382 | SEARCH | 542 | 163 | 7 | **712** | 646 | 188 | 16 | **850** |
| | | UKOPS | 78 | 17 | 1 | **96** | 189 | 46 | 5 | **240** |
| | | USC | 149 | 41 | 3 | **193** | 179 | 40 | 2 | **221** |
| | | GEOCS | 237 | 74 | 12 | **323** | 325 | 86 | 8 | **419** |
| | | MALOVA | 318 | 112 | 6 | **436** | 888 | 284 | 24 | **1196** |
| *BRAF* | rs6944385 | SEARCH | 524 | 177 | 10 | **711** | 646 | 185 | 18 | **849** |
| | | UKOPS | 61 | 28 | 3 | **92** | 161 | 49 | 5 | **215** |
| | | USC | 138 | 42 | 6 | **186** | 150 | 55 | 9 | **214** |
| | | GEOCS | 182 | 112 | 14 | **308** | 239 | 136 | 29 | **404** |
| | | MALOVA | 263 | 159 | 21 | **443** | 700 | 458 | 49 | **1207** |
| *ERBB2* | rs2952155 | SEARCH | 391 | 272 | 50 | **713** | 485 | 322 | 44 | **851** |
| | | USC | 105 | 78 | 10 | **193** | 117 | 88 | 11 | **216** |
| *ERBB2* | rs2952156 | GEOCS | 152 | 162 | 0 | **314** | 194 | 219 | 0 | **413** |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | MALOVA | 216 | 186 | 41 | **443** | 581 | 536 | 90 | **1207** |
| | | SEARCH | 323 | 316 | 75 | **714** | 405 | 365 | 83 | **853** |
| | | UKOPS | 48 | 49 | 0 | **97** | 94 | 129 | 1 | **224** |
| | | USC | 83 | 104 | 0 | **187** | 92 | 124 | 0 | **216** |
| | | GEOCS | 190 | 97 | 16 | **303** | 245 | 120 | 19 | **384** |
| | | MALOVA | 255 | 162 | 26 | **443** | 695 | 427 | 84 | **1206** |
| *ERBB2* | rs1801200 | SEARCH | 408 | 259 | 43 | **710** | 507 | 297 | 40 | **844** |
| | | UKOPS | 71 | 31 | 3 | **105** | 159 | 85 | 19 | **263** |
| | | USC | 123 | 61 | 10 | **194** | 146 | 64 | 9 | **219** |
| | | GEOCS | 267 | 49 | 7 | **323** | 340 | 78 | 3 | **421** |
| | | MALOVA | 363 | 78 | 2 | **443** | 990 | 201 | 7 | **1198** |
| *KRAS* | rs12305513 | SEARCH | 599 | 112 | 5 | **716** | 682 | 157 | 14 | **853** |
| | | UKOPS | 89 | 13 | 0 | **102** | 205 | 45 | 2 | **252** |
| | | USC | 158 | 32 | 2 | **192** | 166 | 40 | 5 | **211** |
| | | GEOCS | 97 | 147 | 75 | **319** | 127 | 202 | 89 | **418** |
| | | MALOVA | 109 | 218 | 110 | **437** | 298 | 599 | 288 | **1185** |
| *KRAS* | rs12822857 | SEARCH | 211 | 347 | 150 | **708** | 250 | 395 | 198 | **843** |
| | | UKOPS | 32 | 56 | 16 | **104** | 69 | 122 | 56 | **247** |
| | | USC | 57 | 75 | 40 | **172** | 65 | 86 | 57 | **208** |
| *KRAS* | rs10842508 | GEOCS | 199 | 102 | 21 | **322** | 239 | 157 | 25 | **421** |
| | | MALOVA | 247 | 164 | 27 | **438** | 670 | 460 | 65 | **1195** |
| | | SEARCH | 403 | 269 | 43 | **715** | 491 | 304 | 58 | **853** |
| | | UKOPS | 65 | 33 | 6 | **104** | 146 | 88 | 15 | **249** |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | USC | 107 | 68 | 11 | **186** | 121 | 78 | 18 | **217** |
| *KRAS* | rs12579073 | GEOCS | 91 | 150 | 80 | **321** | 113 | 213 | 94 | **420** |
| | | MALOVA | 117 | 216 | 104 | **437** | 305 | 586 | 287 | **1178** |
| | | SEARCH | 203 | 343 | 157 | **703** | 237 | 401 | 206 | **844** |
| | | UKOPS | 28 | 53 | 19 | **100** | 62 | 126 | 49 | **237** |
| | | USC | 58 | 91 | 44 | **193** | 62 | 100 | 59 | **221** |
| *KRAS* | rs10842513 | GEOCS | 274 | 46 | 4 | **324** | 346 | 74 | 3 | **423** |
| | | MALOVA | 362 | 69 | 2 | **433** | 937 | 200 | 15 | **1152** |
| | | SEARCH | 591 | 115 | 5 | **711** | 720 | 121 | 5 | **846** |
| | | UKOPS | 78 | 22 | 0 | **100** | 197 | 41 | 4 | **242** |
| | | USC | 155 | 31 | 5 | **191** | 185 | 29 | 2 | **216** |
| *KRAS* | rs4623993 | GEOCS | 238 | 68 | 9 | **315** | 289 | 118 | 6 | **413** |
| | | MALOVA | 312 | 115 | 11 | **438** | 841 | 325 | 27 | **1193** |
| | | SEARCH | 496 | 186 | 19 | **701** | 615 | 203 | 19 | **837** |
| | | UKOPS | 63 | 26 | 2 | **91** | 161 | 59 | 8 | **228** |
| | | USC | 147 | 41 | 5 | **193** | 153 | 60 | 7 | **220** |
| *KRAS* | rs6487464 | GEOCS | 130 | 141 | 52 | **323** | 171 | 185 | 66 | **422** |
| | | MALOVA | 150 | 212 | 73 | **435** | 409 | 563 | 201 | **1173** |
| | | SEARCH | 280 | 324 | 107 | **711** | 343 | 383 | 127 | **853** |
| | | UKOPS | 34 | 46 | 9 | **89** | 94 | 108 | 30 | **232** |
| | | USC | 79 | 87 | 28 | **194** | 85 | 95 | 35 | **215** |
| *KRAS* | rs10842514 | GEOCS | 97 | 157 | 67 | **321** | 134 | 195 | 88 | **417** |
| | | MALOVA | 144 | 201 | 87 | **432** | 386 | 577 | 208 | **1171** |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | SEARCH | 213 | 333 | 171 | **717** | 253 | 421 | 178 | **852** |
| | | UKOPS | 29 | 48 | 19 | **96** | 77 | 110 | 48 | **235** |
| | | USC | 60 | 84 | 37 | **181** | 67 | 97 | 46 | **210** |
| | | GEOCS | 271 | 41 | 2 | **314** | 365 | 41 | 1 | **407** |
| *KRAS* | rs11047917 | MALOVA | 386 | 54 | 3 | **443** | 1052 | 147 | 5 | **1204** |
| | | SEARCH | 645 | 62 | 2 | **709** | 741 | 102 | 2 | **845** |
| | | GEOCS | 216 | 85 | 3 | **304** | 281 | 110 | 8 | **399** |
| | | MALOVA | 321 | 99 | 5 | **425** | 909 | 245 | 11 | **1165** |
| *NMI* | rs394884 | SEARCH | 507 | 170 | 8 | **685** | 620 | 198 | 8 | **826** |
| | | UKOPS | 69 | 28 | 3 | **100** | 171 | 80 | 3 | **254** |
| | | USC | 132 | 41 | 10 | **183** | 141 | 64 | 4 | **209** |
| *NMI* | rs11551174 | MALOVA | 394 | 41 | 1 | **436** | 1048 | 141 | 5 | **1194** |
| | | SEARCH | 629 | 78 | 6 | **713** | 762 | 80 | 4 | **846** |
| | | GEOCS | 231 | 83 | 3 | **317** | 300 | 107 | 10 | **417** |
| | | MALOVA | 328 | 85 | 6 | **419** | 955 | 222 | 15 | **1192** |
| *NMI* | rs289831 | SEARCH | 545 | 158 | 9 | **712** | 668 | 176 | 7 | **851** |
| | | UKOPS | 69 | 22 | 0 | **91** | 106 | 45 | 0 | **151** |
| | | USC | 84 | 29 | 2 | **115** | 78 | 29 | 1 | **108** |
| | | GEOCS | 98 | 166 | 56 | **320** | 164 | 187 | 69 | **420** |
| | | MALOVA | 146 | 199 | 92 | **437** | 384 | 584 | 229 | **1197** |
| *NMI* | rs3771886 | SEARCH | 250 | 327 | 129 | **706** | 294 | 420 | 134 | **848** |
| | | UKOPS | 29 | 51 | 21 | **101** | 99 | 120 | 30 | **249** |
| | | USC | 66 | 84 | 39 | **189** | 70 | 100 | 43 | **213** |
| *NMI* | rs11683487 | GEOCS | 91 | 123 | 56 | **270** | 111 | 170 | 83 | **364** |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | MALOVA | 140 | 192 | 83 | **415** | 356 | 549 | 257 | **1162** |
| | | SEARCH | 162 | 204 | 124 | **490** | 164 | 321 | 138 | **623** |
| | | UKOPS | 32 | 46 | 12 | **90** | 45 | 113 | 40 | **198** |
| | | USC | 66 | 87 | 41 | **194** | 78 | 96 | 44 | **218** |
| | | GEOCS | 237 | 82 | 2 | **321** | 300 | 106 | 13 | **419** |
| | | MALOVA | 340 | 91 | 7 | **438** | 957 | 231 | 14 | **1202** |
| *NMI* | rs2113509 | SEARCH | 548 | 156 | 9 | **713** | 666 | 174 | 6 | **846** |
| | | UKOPS | 75 | 28 | 1 | **104** | 189 | 68 | 1 | **258** |
| | | USC | 141 | 44 | 4 | **189** | 162 | 56 | 1 | **219** |
| *PIK3CA* | rs2865084 | MALOVA | 393 | 48 | 0 | **441** | 1083 | 115 | 0 | **1198** |
| | | SEARCH | 630 | 79 | 0 | **709** | 758 | 83 | 0 | **841** |
| | | GEOCS | 211 | 103 | 10 | **324** | 293 | 116 | 10 | **419** |
| | | MALOVA | 311 | 114 | 15 | **440** | 814 | 318 | 30 | **1162** |
| *PIK3CA* | rs7621329 | SEARCH | 487 | 195 | 19 | **701** | 583 | 237 | 23 | **843** |
| | | UKOPS | 65 | 14 | 5 | **84** | 130 | 46 | 4 | **180** |
| | | USC | 129 | 55 | 6 | **190** | 147 | 62 | 6 | **215** |
| | | GEOCS | 243 | 64 | 3 | **310** | 337 | 61 | 7 | **405** |
| | | MALOVA | 351 | 71 | 4 | **426** | 956 | 215 | 13 | **1184** |
| *PIK3CA* | rs1517586 | SEARCH | 580 | 109 | 7 | **696** | 697 | 141 | 7 | **845** |
| | | UKOPS | 85 | 14 | 2 | **101** | 209 | 39 | 6 | **254** |
| | | USC | 165 | 30 | 0 | **195** | 176 | 41 | 3 | **220** |
| *PIK3CA* | rs2699905 | GEOCS | 157 | 113 | 25 | **295** | 231 | 124 | 37 | **392** |
| | | MALOVA | 264 | 144 | 32 | **440** | 658 | 440 | 103 | **1201** |
| | | SEARCH | 392 | 248 | 70 | **710** | 487 | 288 | 72 | **847** |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | UKOPS | 50 | 34 | 8 | **92** | 117 | 53 | 30 | **200** |
| | | USC | 126 | 57 | 10 | **193** | 118 | 75 | 23 | **216** |
| | | GEOCS | 277 | 46 | 0 | **323** | 360 | 56 | 3 | **419** |
| | | MALOVA | 393 | 43 | 3 | **439** | 1044 | 144 | 8 | **1196** |
| *PIK3CA* | rs7641889 | SEARCH | 627 | 82 | 6 | **715** | 731 | 115 | 2 | **848** |
| | | UKOPS | 87 | 12 | 1 | **100** | 231 | 27 | 1 | **259** |
| | | USC | 168 | 22 | 1 | **191** | 192 | 24 | 1 | **217** |
| | | GEOCS | 257 | 52 | 3 | **312** | 314 | 82 | 5 | **401** |
| | | MALOVA | 348 | 80 | 6 | **434** | 932 | 235 | 15 | **1182** |
| *PIK3CA* | rs7651265 | SEARCH | 568 | 128 | 12 | **708** | 657 | 186 | 5 | **848** |
| | | UKOPS | 79 | 14 | 2 | **95** | 205 | 34 | 0 | **239** |
| | | USC | 145 | 43 | 1 | **189** | 179 | 33 | 1 | **213** |
| | | GEOCS | 223 | 98 | 1 | **322** | 312 | 102 | 9 | **423** |
| | | MALOVA | 335 | 94 | 11 | **440** | 857 | 320 | 26 | **1203** |
| *PIK3CA* | rs7640662 | SEARCH | 493 | 200 | 19 | **712** | 619 | 202 | 25 | **846** |
| | | UKOPS | 65 | 20 | 2 | **87** | 164 | 53 | 10 | **227** |
| | | USC | 149 | 39 | 5 | **193** | 155 | 58 | 4 | **217** |
| | | GEOCS | 91 | 165 | 59 | **315** | 100 | 203 | 104 | **407** |
| | | MALOVA | 98 | 218 | 115 | **431** | 315 | 605 | 268 | **1188** |
| *PIK3CA* | rs2677760 | SEARCH | 209 | 322 | 179 | **710** | 230 | 416 | 201 | **847** |
| | | UKOPS | 23 | 56 | 21 | **100** | 67 | 122 | 72 | **261** |
| | | USC | 45 | 104 | 46 | **195** | 62 | 116 | 44 | **222** |

AA – common homozygous; Aa – heterozygous; aa – rare homozygous;

## Appendix II-B: Genotype distributions of tSNPs in *AIFM2*, *AKTIP*, *AXIN2*, *CASP5*, *FILIP1L*, *RBBP8*, *RGC32*, *RUVBL1* and *STAG3*

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| *AIFM2* | rs2394655 | MALOVA | 406 | 34 | 1 | 441 | 1053 | 91 | 5 | 1149 |
| | | SEARCH | 768 | 59 | 1 | 828 | 1123 | 85 | 0 | 1208 |
| | | UKOPS | 443 | 36 | 3 | 482 | 526 | 39 | 2 | 567 |
| *AIFM2* | rs7908957 | MALOVA | 352 | 78 | 8 | 438 | 862 | 246 | 19 | 1127 |
| | | SEARCH | 612 | 175 | 20 | 807 | 883 | 275 | 21 | 1179 |
| | | UKOPS | 363 | 103 | 8 | 474 | 443 | 113 | 11 | 567 |
| *AIFM2* | rs1053495 | MALOVA | 308 | 42 | 1 | 351 | 767 | 124 | 8 | 899 |
| | | SEARCH | 727 | 109 | 7 | 843 | 1046 | 163 | 11 | 1220 |
| | | UKOPS | 430 | 70 | 3 | 503 | 504 | 77 | 4 | 585 |
| *AIFM2* | rs2894111 | MALOVA | 234 | 179 | 30 | 443 | 589 | 405 | 91 | 1085 |
| | | SEARCH | 437 | 333 | 70 | 840 | 600 | 505 | 108 | 1213 |
| | | UKOPS | 253 | 194 | 40 | 487 | 297 | 223 | 43 | 563 |
| *AIFM2* | rs2394656 | MALOVA | 303 | 125 | 12 | 440 | 746 | 351 | 37 | 1134 |
| | | UKOPS | 310 | 143 | 20 | 473 | 376 | 167 | 26 | 569 |
| *AIFM2* | rs6480440 | MALOVA | 259 | 138 | 25 | 422 | 697 | 368 | 75 | 1140 |
| *AIFM2* | rs2280201 | MALOVA | 342 | 94 | 6 | 442 | 928 | 212 | 22 | 1162 |
| | | SEARCH | 631 | 187 | 13 | 831 | 925 | 260 | 23 | 1208 |
| | | UKOPS | 372 | 101 | 9 | 482 | 451 | 118 | 6 | 575 |
| *AIFM2* | rs10999147 | MALOVA | 373 | 65 | 2 | 440 | 1017 | 152 | 4 | 1173 |
| | | SEARCH | 695 | 148 | 2 | 845 | 1032 | 178 | 12 | 1222 |
| | | UKOPS | 418 | 77 | 5 | 500 | 488 | 78 | 9 | 575 |
| *AIFM2* | rs3750772 | MALOVA | 383 | 57 | 3 | 443 | 1043 | 129 | 5 | 1177 |
| | | SEARCH | 740 | 72 | 3 | 815 | 1059 | 128 | 3 | 1190 |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | UKOPS | 435 | 47 | 3 | 485 | 529 | 48 | 0 | 577 |
| *AIFM2* | rs4295944 | MALOVA | 160 | 198 | 81 | 439 | 423 | 533 | 217 | 1173 |
| | | SEARCH | 284 | 400 | 158 | 842 | 404 | 602 | 204 | 1210 |
| | | UKOPS | 164 | 238 | 91 | 493 | 201 | 269 | 104 | 574 |
| *AIFM2* | rs2394644 | SEARCH | 635 | 183 | 14 | 832 | 863 | 231 | 22 | 1116 |
| | | UKOPS | 373 | 108 | 11 | 492 | 438 | 122 | 9 | 569 |
| *AIFM2* | rs10999152 | MALOVA | 221 | 105 | 14 | 340 | 602 | 241 | 32 | 875 |
| | | SEARCH | 535 | 239 | 23 | 797 | 806 | 322 | 41 | 1169 |
| | | UKOPS | 329 | 135 | 17 | 481 | 401 | 143 | 22 | 566 |
| *AKTIP* | rs9931702 | MALOVA | 145 | 211 | 85 | 441 | 362 | 553 | 235 | 1150 |
| | | UKOPS | 149 | 224 | 103 | 476 | 183 | 278 | 111 | 572 |
| *AKTIP* | rs17801966 | MALOVA | 265 | 78 | 5 | 348 | 670 | 214 | 14 | 898 |
| | | UKOPS | 344 | 128 | 8 | 480 | 426 | 130 | 15 | 571 |
| *AKTIP* | rs7189819 | MALOVA | 212 | 182 | 45 | 439 | 557 | 471 | 124 | 1152 |
| | | SEARCH | 398 | 354 | 73 | 825 | 538 | 536 | 126 | 1200 |
| | | UKOPS | 236 | 190 | 55 | 481 | 275 | 242 | 54 | 571 |
| *AKTIP* | rs3743772 | MALOVA | 371 | 40 | 2 | 413 | 971 | 119 | 3 | 1093 |
| *AXIN2* | rs11868547 | MALOVA | 114 | 204 | 118 | 436 | 268 | 578 | 298 | 1144 |
| | | SEARCH | 185 | 314 | 141 | 640 | 293 | 456 | 238 | 987 |
| | | UKOPS | 127 | 249 | 107 | 483 | 160 | 297 | 116 | 573 |
| *AXIN2* | rs7591 | MALOVA | 175 | 196 | 70 | 441 | 492 | 524 | 157 | 1173 |
| | | SEARCH | 308 | 412 | 117 | 837 | 446 | 512 | 168 | 1126 |
| | | UKOPS | 194 | 239 | 68 | 501 | 219 | 277 | 86 | 582 |
| *AXIN2* | rs4074947 | MALOVA | 274 | 143 | 22 | 439 | 721 | 345 | 43 | 1109 |
| | | SEARCH | 508 | 303 | 27 | 838 | 775 | 383 | 55 | 1213 |
| | | UKOPS | 297 | 181 | 20 | 498 | 356 | 188 | 32 | 576 |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| *AXIN2* | rs7210356 | MALOVA | 348 | 90 | 2 | 440 | 934 | 225 | 18 | 1177 |
| | | SEARCH | 658 | 173 | 12 | 843 | 972 | 235 | 13 | 1220 |
| | | UKOPS | 395 | 92 | 7 | 494 | 447 | 125 | 5 | 577 |
| *AXIN2* | rs11655966 | SEARCH | 448 | 312 | 66 | 826 | 648 | 475 | 83 | 1206 |
| | | UKOPS | 272 | 173 | 30 | 475 | 313 | 228 | 32 | 573 |
| *AXIN2* | rs4541111 | SEARCH | 203 | 430 | 191 | 824 | 332 | 574 | 296 | 1202 |
| | | UKOPS | 132 | 234 | 107 | 473 | 141 | 289 | 138 | 568 |
| *AXIN2* | rs4791171 | MALOVA | 173 | 136 | 36 | 345 | 475 | 352 | 69 | 896 |
| | | SEARCH | 412 | 351 | 77 | 840 | 619 | 491 | 103 | 1213 |
| *AXIN2* | rs11079571 | SEARCH | 573 | 238 | 28 | 839 | 869 | 310 | 27 | 1206 |
| *AXIN2* | rs3923087 | MALOVA | 285 | 138 | 17 | 440 | 757 | 332 | 40 | 1129 |
| | | SEARCH | 498 | 289 | 49 | 836 | 741 | 417 | 50 | 1208 |
| | | UKOPS | 295 | 180 | 29 | 504 | 338 | 203 | 32 | 573 |
| *AXIN2* | rs3923086 | MALOVA | 171 | 208 | 64 | 443 | 422 | 570 | 165 | 1157 |
| | | SEARCH | 274 | 392 | 160 | 826 | 409 | 585 | 211 | 1205 |
| | | UKOPS | 166 | 216 | 102 | 484 | 175 | 296 | 102 | 573 |
| *CASP5* | rs518604 | MALOVA | 105 | 225 | 108 | 438 | 369 | 568 | 258 | 1195 |
| | | SEARCH | 262 | 411 | 158 | 831 | 375 | 628 | 201 | 1204 |
| *CASP5* | rs523104 | SEARCH | 223 | 424 | 177 | 824 | 330 | 617 | 252 | 1199 |
| *CASP5* | rs3181328 | MALOVA | 365 | 75 | 2 | 442 | 933 | 193 | 22 | 1148 |
| | | SEARCH | 702 | 116 | 11 | 829 | 1013 | 179 | 14 | 1206 |
| | | UKOPS | 398 | 78 | 1 | 477 | 491 | 71 | 9 | 571 |
| *CASP5* | rs17446518 | SEARCH | 647 | 146 | 10 | 803 | 930 | 235 | 12 | 1177 |
| | | UKOPS | 360 | 106 | 22 | 488 | 451 | 97 | 26 | 574 |
| *CASP5* | rs9651713 | MALOVA | 363 | 73 | 4 | 440 | 936 | 197 | 9 | 1142 |
| | | SEARCH | 634 | 173 | 11 | 818 | 943 | 235 | 11 | 1189 |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | UKOPS | 367 | 97 | 8 | 472 | 427 | 130 | 10 | 567 |
| *CASP5* | rs3181175 | MALOVA | 298 | 128 | 13 | 439 | 774 | 359 | 31 | 1164 |
| | | SEARCH | 554 | 257 | 32 | 843 | 800 | 379 | 36 | 1215 |
| *CASP5* | rs3181174 | MALOVA | 370 | 68 | 3 | 441 | 986 | 179 | 9 | 1174 |
| | | SEARCH | 729 | 103 | 4 | 836 | 1036 | 165 | 6 | 1207 |
| | | UKOPS | 436 | 63 | 4 | 503 | 508 | 72 | 1 | 581 |
| *CASP5* | rs2282657 | MALOVA | 131 | 183 | 37 | 351 | 363 | 416 | 119 | 898 |
| | | SEARCH | 363 | 366 | 114 | 843 | 510 | 586 | 123 | 1219 |
| | | UKOPS | 214 | 234 | 52 | 500 | 241 | 274 | 65 | 580 |
| *CASP5* | rs507879 | MALOVA | 90 | 236 | 115 | 441 | 235 | 570 | 324 | 1129 |
| | | SEARCH | 167 | 407 | 251 | 825 | 239 | 609 | 346 | 1194 |
| | | UKOPS | 104 | 251 | 147 | 502 | 110 | 249 | 157 | 516 |
| *FILIP1L* | rs796977 | MALOVA | 188 | 188 | 61 | 437 | 541 | 488 | 137 | 1166 |
| | | SEARCH | 622 | 190 | 13 | 825 | 912 | 261 | 26 | 1199 |
| | | UKOPS | 362 | 108 | 8 | 478 | 436 | 123 | 9 | 568 |
| *FILIP1L* | rs793446 | MALOVA | 145 | 205 | 88 | 438 | 407 | 536 | 216 | 1159 |
| | | SEARCH | 292 | 415 | 135 | 842 | 438 | 582 | 196 | 1216 |
| | | UKOPS | 180 | 233 | 80 | 493 | 213 | 274 | 85 | 572 |
| *FILIP1L* | rs3921767 | MALOVA | 379 | 59 | 1 | 439 | 958 | 151 | 6 | 1115 |
| | | SEARCH | 718 | 114 | 4 | 836 | 1062 | 148 | 9 | 1219 |
| | | UKOPS | 444 | 51 | 3 | 498 | 453 | 69 | 3 | 525 |
| *FILIP1L* | rs17338680 | MALOVA | 349 | 81 | 9 | 439 | 954 | 200 | 30 | 1184 |
| | | SEARCH | 677 | 162 | 5 | 844 | 979 | 225 | 17 | 1221 |
| | | UKOPS | 390 | 105 | 8 | 503 | 460 | 116 | 8 | 584 |
| *FILIP1L* | rs9864437 | MALOVA | 265 | 135 | 39 | 439 | 685 | 415 | 66 | 1166 |
| | | SEARCH | 507 | 300 | 38 | 845 | 739 | 418 | 64 | 1221 |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | UKOPS | 312 | 161 | 29 | 502 | 372 | 184 | 29 | 585 |
| *FILIP1L* | rs6788750 | MALOVA | 161 | 212 | 67 | 440 | 418 | 534 | 192 | 1144 |
| | | SEARCH | 165 | 241 | 88 | 494 | 270 | 402 | 146 | 818 |
| | | UKOPS | 161 | 239 | 80 | 480 | 186 | 291 | 93 | 570 |
| *FILIP1L* | rs12494994 | MALOVA | 294 | 121 | 19 | 434 | 781 | 314 | 40 | 1135 |
| | | SEARCH | 562 | 260 | 17 | 839 | 845 | 324 | 43 | 1212 |
| *RBBP8* | rs7239066 | MALOVA | 352 | 83 | 8 | 443 | 889 | 255 | 15 | 1159 |
| | | SEARCH | 674 | 146 | 9 | 829 | 942 | 251 | 14 | 1207 |
| | | UKOPS | 365 | 111 | 5 | 481 | 469 | 99 | 6 | 574 |
| *RBBP8* | rs11082221 | MALOVA | 403 | 37 | 2 | 442 | 1058 | 96 | 3 | 1157 |
| | | SEARCH | 777 | 50 | 1 | 828 | 1134 | 72 | 2 | 1208 |
| | | UKOPS | 432 | 45 | 1 | 478 | 534 | 37 | 1 | 572 |
| *RBBP8* | rs4474794 | MALOVA | 187 | 197 | 51 | 435 | 444 | 507 | 155 | 1106 |
| | | SEARCH | 342 | 378 | 115 | 835 | 479 | 571 | 165 | 1215 |
| | | UKOPS | 201 | 229 | 64 | 494 | 230 | 261 | 83 | 574 |
| *RBBP8* | rs9304261 | MALOVA | 207 | 122 | 17 | 346 | 531 | 291 | 66 | 888 |
| *RGC32* | rs10467472 | MALOVA | 344 | 90 | 9 | 443 | 853 | 228 | 18 | 1099 |
| | | SEARCH | 628 | 193 | 16 | 837 | 914 | 282 | 23 | 1219 |
| | | UKOPS | 367 | 112 | 10 | 489 | 426 | 131 | 12 | 569 |
| *RGC32* | rs3783194 | MALOVA | 272 | 71 | 6 | 349 | 728 | 184 | 8 | 920 |
| | | SEARCH | 662 | 171 | 11 | 844 | 951 | 254 | 13 | 1218 |
| | | UKOPS | 394 | 102 | 1 | 497 | 465 | 112 | 8 | 585 |
| *RGC32* | rs11618371 | MALOVA | 349 | 83 | 8 | 440 | 956 | 201 | 11 | 1168 |
| | | SEARCH | 664 | 168 | 9 | 841 | 971 | 233 | 11 | 1215 |
| | | UKOPS | 403 | 82 | 5 | 490 | 459 | 110 | 7 | 576 |
| *RGC32* | rs9532824 | MALOVA | 383 | 59 | 2 | 444 | 929 | 161 | 11 | 1101 |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| | | SEARCH | 720 | 119 | 2 | 841 | 1037 | 178 | 8 | 1223 |
| | | UKOPS | 437 | 58 | 2 | 497 | 506 | 58 | 4 | 568 |
| *RGC32* | rs995845 | MALOVA | 230 | 171 | 35 | 436 | 636 | 451 | 76 | 1163 |
| | | SEARCH | 468 | 317 | 53 | 838 | 657 | 467 | 78 | 1202 |
| *RGC32* | rs9594551 | MALOVA | 306 | 120 | 14 | 440 | 771 | 291 | 22 | 1084 |
| | | SEARCH | 630 | 194 | 14 | 838 | 912 | 286 | 21 | 1219 |
| | | UKOPS | 339 | 135 | 14 | 488 | 396 | 145 | 19 | 560 |
| *RGC32* | rs975590 | MALOVA | 252 | 169 | 19 | 440 | 658 | 449 | 62 | 1169 |
| | | SEARCH | 474 | 304 | 46 | 824 | 700 | 424 | 72 | 1196 |
| | | UKOPS | 287 | 171 | 27 | 485 | 343 | 196 | 36 | 575 |
| *RUVBL1* | rs9860614 | MALOVA | 341 | 86 | 7 | 434 | 895 | 253 | 22 | 1170 |
| | | SEARCH | 649 | 175 | 19 | 843 | 973 | 233 | 13 | 1219 |
| | | UKOPS | 386 | 107 | 7 | 500 | 457 | 114 | 6 | 577 |
| *RUVBL1* | rs13063604 | SEARCH | 450 | 283 | 55 | 788 | 709 | 383 | 63 | 1155 |
| | | UKOPS | 266 | 176 | 36 | 478 | 333 | 202 | 34 | 569 |
| *RUVBL1* | rs3732402 | MALOVA | 163 | 218 | 57 | 438 | 450 | 544 | 171 | 1165 |
| | | SEARCH | 290 | 425 | 127 | 842 | 478 | 559 | 180 | 1217 |
| *RUVBL1* | rs7650365 | MALOVA | 88 | 185 | 76 | 349 | 242 | 423 | 233 | 898 |
| | | SEARCH | 206 | 445 | 168 | 819 | 313 | 594 | 297 | 1204 |
| | | UKOPS | 141 | 222 | 114 | 477 | 151 | 270 | 149 | 570 |
| *RUVBL1* | rs4857836 | MALOVA | 237 | 180 | 25 | 442 | 658 | 462 | 69 | 1189 |
| | | SEARCH | 443 | 341 | 60 | 844 | 648 | 482 | 90 | 1220 |
| | | UKOPS | 264 | 195 | 42 | 501 | 306 | 226 | 52 | 584 |
| *RUVBL1* | rs9821568 | MALOVA | 321 | 114 | 5 | 440 | 855 | 271 | 20 | 1146 |
| | | SEARCH | 604 | 190 | 22 | 816 | 866 | 301 | 30 | 1197 |
| | | UKOPS | 344 | 117 | 16 | 477 | 397 | 152 | 19 | 568 |

| Gene | tSNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| *STAG3* | rs11762932 | MALOVA | 266 | 146 | 28 | 440 | 705 | 396 | 64 | 1165 |
| | | SEARCH | 518 | 287 | 39 | 844 | 753 | 412 | 55 | 1220 |
| | | UKOPS | 304 | 179 | 20 | 503 | 362 | 192 | 26 | 580 |
| *STAG3* | rs2246713 | SEARCH | 218 | 415 | 189 | 822 | 290 | 621 | 286 | 1197 |
| | | UKOPS | 123 | 248 | 102 | 473 | 162 | 285 | 121 | 568 |
| *STAG3* | rs1637001 | MALOVA | 263 | 148 | 29 | 440 | 608 | 472 | 87 | 1167 |
| | | SEARCH | 451 | 324 | 68 | 843 | 612 | 505 | 99 | 1216 |
| | | UKOPS | 269 | 193 | 39 | 501 | 321 | 222 | 41 | 584 |

AA – common homozygous; Aa – heterozygous; aa – rare homozygous;

## Appendix III-A: Genotype specific ratios of *BRAF* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR[§] (95% CI) | HomOR[§] (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 1680 | All | 1.09 (0.93-1.28) | 1.02 (0.86-1.22) | 0.902 |
| | | | | 804 | Serous | 1.21 (0.99-1.49) | 1.09 (0.87-1.37) | 0.5747 |
| ***BRAF*** | **rs10487888** | 0.47 | 2694 | 251 | Endometrioid | 0.84 (0.61-1.17) | 0.88 (0.61-1.25) | 0.5007 |
| | | | | 180 | **Mucinous** | 1.32 (0.86-2.03) | **1.61 (1.03-2.53)** | **0.0357** |
| | | | | 125 | Clear cell | 1.04 (0.66-1.64) | 0.88 (0.52-1.48) | 0.5947 |
| | | | | 1159 | All | 1.08 (0.86-1.36) | 3.39 (0.96-11.89) | 0.1985 |
| | | | | 525 | Serous | 1.15 (0.86-1.53) | 2.99 (0.66-13.46) | 0.1847 |
| *BRAF* | rs1733832 | 0.06 | 2043 | 182 | Endometrioid | 1.25 (0.79-1.98) | 3.18 (0.33-30.71) | 0.2301 |
| | | | | 135 | Mucinous | 0.58 (0.29-1.16) | **15.23 (3.16-73.29)** | 0.9086 |
| | | | | 95 | Clear cell | 1.18 (0.63-2.20) | 5.89 (0.62-55.68) | 0.335 |
| | | | | 1751 | All | 0.99 (0.87-1.12) | 0.97 (0.75-1.26) | 0.7894 |
| | | | | 831 | Serous | 1.04 (0.88-1.22) | 0.90 (0.64-1.26) | 0.9055 |
| ***BRAF*** | **rs1267622** | 0.24 | 2880 | 268 | Endometrioid | 0.76 (0.58-1.01) | 1.06 (0.63-1.77) | 0.2725 |
| | | | | 187 | **Mucinous** | **0.67 (0.48-0.94)** | 0.71 (0.35-1.43) | **0.0278** |
| | | | | 135 | Clear cell | 1.24 (0.87-1.78) | 1.12 (0.53-2.37) | 0.3392 |
| | | | | 1602 | All | 0.98 (0.85-1.12) | 0.86 (0.69-1.08) | 0.2715 |
| | | | | 733 | Serous | 1.01 (0.85-1.20) | 0.91 (0.68-1.23) | 0.7063 |
| *BRAF* | rs13241719 | 0.31 | 2488 | 246 | Endometrioid | 0.93 (0.70-1.23) | 0.88 (0.56-1.39) | 0.5238 |
| | | | | 176 | Mucinous | 1.14 (0.82-1.58) | 0.98 (0.57-1.67) | 0.7484 |
| | | | | 123 | Clear cell | 0.79 (0.54-1.16) | 0.57 (0.28-1.16) | 0.0734 |
| ***BRAF*** | **rs17695623** | 0.07 | 2901 | 1744 | All | 0.97 (0.81-1.16) | 1.14 (0.52-2.46) | 0.8642 |
| | | | | 829 | Serous | 1.04 (0.83-1.31) | 1.19 (0.45-3.10) | 0.6437 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 264 | Endometrioid | 0.99 (0.68-1.45) | 0.63 (0.08-4.85) | 0.8421 |
| | | | | 186 | **Mucinous** | **0.47 (0.26-0.86)** | 0.79 (0.10-6.08) | **0.0191** |
| | | | | 135 | Clear cell | 1.25 (0.78-2.03) | 1.37 (0.18-10.56) | 0.3393 |
| | | | | 1771 | All | 1.13 (0.93-1.38) | 1.29 (0.57-2.93) | 0.1802 |
| | | | | 847 | Serous | 1.04 (0.80-1.34) | 1.43 (0.53-3.82) | 0.5816 |
| *BRAF* | rs17161747 | 0.5 | 2909 | 272 | Endometrioid | 1.03 (0.67-1.58) | 2.85 (0.91-8.91) | 0.3223 |
| | | | | 191 | Mucinous | 1.14 (0.70-1.84) | 0.96 (0.12-7.47) | 0.6521 |
| | | | | 132 | Clear cell | 1.42 (0.84-2.41) | 1.55 (0.20-12.12) | 0.1738 |
| | | | | 1764 | All | 1.01 (0.87-1.17) | 1.01 (0.61-1.66) | 0.9035 |
| | | | | 841 | Serous | 0.97 (0.80-1.18) | 0.98 (0.51-1.86) | 0.7935 |
| *BRAF* | rs17623382 | 0.12 | 2900 | 270 | Endometrioid | 0.87 (0.64-1.20) | 0.22 (0.03-1.62) | 0.1391 |
| | | | | 186 | Mucinous | 1.08 (0.76-1.55) | 1.74 (0.67-4.51) | 0.3566 |
| | | | | 134 | Clear cell | 0.90 (0.58-1.40) | 1.86 (0.65-5.30) | 0.8512 |
| | | | | 1758 | All | 1.14 (0.99-1.32) | 0.99 (0.66-1.50) | 0.1407 |
| | | | | 840 | Serous | **1.25 (1.05-1.50)** | 0.78 (0.44-1.38) | 0.114 |
| *BRAF* | rs6944385 | 0.14 | 2893 | 268 | Endometrioid | 0.98 (0.72-1.33) | 1.18 (0.53-2.62) | 0.9193 |
| | | | | 187 | Mucinous | **0.63 (0.42-0.95)** | 0.88 (0.31-2.47) | 0.057 |
| | | | | 133 | Clear cell | 1.41 (0.96-2.09) | 1.54 (0.54-4.34) | 0.0691 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix III-B: Haplotype-specific risks of *BRAF* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *BRAF* | h10000000 | 21.5 | All | 1.07 (0.97-1.19) | 0.182 |
| | | | Serous | 1.07 (0.93-1.24) | 0.325 |
| | | | Endometrioid | 1 (0.8-1.27) | 0.966 |
| | | | Mucinous | 1.28 (0.99-1.66) | 0.059 |
| | | | Clear cell | 1.24 (0.91-1.69) | 0.174 |
| ***BRAF*** | **h10010000** | 19.4 | All | 0.88 (0.76-1.01) | 0.07 |
| | | | Serous | 0.99 (0.86-1.15) | 0.921 |
| | | | Endometrioid | 1.04 (0.82-1.31) | 0.769 |
| | | | Mucinous | 0.92 (0.69-1.22) | 0.547 |
| | | | **Clear cell** | **0.67 (0.46-0.97)** | **0.033** |
| *BRAF* | h00000000 | 17.3 | All | 0.90 (0.78-1.04) | 0.15 |
| | | | Serous | 0.96 (0.83-1.12) | 0.605 |
| | | | Endometrioid | 1.06 (0.84-1.35) | 0.612 |
| | | | Mucinous | 0.98 (0.73-1.31) | 0.881 |
| | | | Clear cell | 0.83 (0.57-1.19) | 0.303 |
| *BRAF* | h10010010 | 11.8 | All | 0.96 (0.82-1.12) | 0.57 |
| | | | Serous | 1.01 (0.85-1.2) | 0.931 |
| | | | Endometrioid | 0.87 (0.65-1.17) | 0.36 |
| | | | Mucinous | 1.19 (0.87-1.62) | 0.282 |
| | | | Clear cell | 1.03 (0.7-1.51) | 0.895 |
| ***BRAF*** | **h00100000** | 10.3 | **All** | **0.81 (0.68-0.95)** | **0.012** |
| | | | **Serous** | **0.8 (0.66-0.98)** | **0.028** |
| | | | Endometrioid | 0.88 (0.64-1.21) | 0.439 |
| | | | Mucinous | 0.87 (0.59-1.27) | 0.463 |
| | | | Clear cell | 0.86 (0.55-1.34) | 0.508 |
| ***BRAF*** | **h00101001** | 6.8 | All | 0.94 (0.78-1.13) | 0.49 |
| | | | Serous | 1.09 (0.88-1.35) | 0.444 |
| | | | Endometrioid | 0.99 (0.69-1.42) | 0.947 |
| | | | **Mucinous** | **0.54 (0.31-0.93)** | **0.027** |
| | | | Clear cell | 1.36 (0.87-2.11) | 0.175 |
| *BRAF* | h01100001 | 6.1 | All | 1.15 (0.95-1.39) | 0.14 |
| | | | Serous | 1.2 (0.97-1.5) | 0.095 |
| | | | Endometrioid | 1.17 (0.81-1.69) | 0.398 |
| | | | Mucinous | 0.93 (0.58-1.48) | 0.76 |
| | | | Clear cell | 1.32 (0.81-2.16) | 0.262 |
| *BRAF* | h00000100 | 5.2 | All | 1.08 (0.88-1.31) | 0.48 |
| | | | Serous | 1.11 (0.88-1.4) | 0.382 |
| | | | Endometrioid | 1.17 (0.81-1.69) | 0.405 |
| | | | Mucinous | 1.11 (0.7-1.75) | 0.661 |
| | | | Clear cell | 1.36 (0.84-2.22) | 0.213 |
| ***BRAF*** | **Rare** | 0.6 | **All** | **0.6 (0.4-0.91)** | **0.007** |
| | | | **Serous** | **0.52 (0.28-0.97)** | **0.038** |
| | | | Endometrioid | 0.52 (0.17-1.61) | 0.258 |
| | | | Mucinous | 0.87 (0.38-1.99) | 0.748 |
| | | | Clear cell | 0.73 (0.2-2.62) | 0.626 |

OR - odds ratio, CI - confidence interval; SNP order in haplotype (5' to 3') – rs10487888, rs1733832, rs1267622, rs13241719, rs17695623, rs17161747, rs17623382, rs6944385.

## Appendix III-C: Genotype specific risks of *ERBB2* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| *ERBB2* | rs2952155 | 0.24 | 2678 | 1667 | All | 1.01 (0.89-1.15) | 1.11 (0.84-1.47) | 0.5745 |
| | | | | 795 | Serous | 0.93 (0.79-1.10) | 0.97 (0.67-1.40) | 0.4832 |
| | | | | 250 | Endometrioid | 1.06 (0.81-1.40) | 1.26 (0.73-2.20) | 0.4297 |
| | | | | 177 | Mucinous | 1.32 (0.96-1.81) | 1.11 (0.54-2.27) | 0.175 |
| | | | | 126 | Clear cell | 0.95 (0.65-1.39) | 1.20 (0.56-2.55) | 0.9093 |
| *ERBB2* | rs2952156 | 0.29 | 2912 | 1766 | All | 0.97 (0.86-1.10) | 1.15 (0.89-1.49) | 0.7416 |
| | | | | 840 | Serous | 0.92 (0.79-1.08) | 1.24 (0.90-1.72) | 0.8924 |
| | | | | 269 | Endometrioid | 0.95 (0.73-1.23) | 1.05 (0.61-1.78) | 0.8719 |
| | | | | 186 | Mucinous | 1.25 (0.92-1.71) | 0.81 (0.39-1.65) | 0.5535 |
| | | | | 135 | Clear cell | 0.89 (0.63-1.28) | 0.89 (0.41-1.90) | 0.5537 |
| ***ERBB2*** | **rs1801200** | 0.22 | 2916 | 1766 | All | 1.04 (0.92-1.19) | 1.01 (0.77-1.31) | 0.6401 |
| | | | | 847 | Serous | 1.04 (0.88-1.22) | 0.86 (0.60-1.23) | 0.8257 |
| | | | | 263 | **Endometrioid** | 1.16 (0.88-1.52) | **1.71 (1.05-2.76)** | **0.0389** |
| | | | | 188 | Mucinous | 0.79 (0.57-1.11) | 0.82 (0.41-1.66) | 0.2007 |
| | | | | 134 | Clear cell | **1.51 (1.05-2.17)** | 1.30 (0.61-2.76) | 0.0564 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix III-D: Haplotype-specific risks of *ERBB2* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *ERBB2* | h000 | 53.6 | All | 0.95 (0.87-1.04) | 0.284 |
| | | | Serous | 0.99 (0.88-1.11) | 0.808 |
| | | | Endometrioid | 0.85 (0.7-1.04) | 0.108 |
| | | | Mucinous | 1.03 (0.82-1.3) | 0.792 |
| | | | Clear cell | 0.87 (0.67-1.15) | 0.331 |
| ***ERBB2*** | **h110** | 16.3 | **All** | **1.19 (1.03-1.37)** | **0.016** |
| | | | Serous | 1 (0.85-1.19) | 0.964 |
| | | | Endometrioid | 1 (0.75-1.34) | 0.982 |
| | | | **Mucinous** | **1.39 (1.02-1.9)** | **0.036** |
| | | | Clear cell | 0.94 (0.63-1.4) | 0.752 |
| ***ERBB2*** | **h001** | 16 | **All** | **1.17 (1.02-1.34)** | **0.022** |
| | | | Serous | 1.08 (0.92-1.27) | 0.329 |
| | | | Endometrioid | 1.25 (0.97-1.62) | 0.079 |
| | | | Mucinous | 0.88 (0.63-1.23) | 0.466 |
| | | | **Clear cell** | **1.6 (1.15-2.21)** | **0.005** |
| *ERBB2* | h010 | 6.6 | All | 0.99 (0.82-1.20) | 0.9 |
| | | | Serous | 1.1 (0.87-1.38) | 0.435 |
| | | | Endometrioid | 0.86 (0.57-1.3) | 0.474 |
| | | | Mucinous | 0.63 (0.37-1.1) | 0.102 |
| | | | Clear cell | 0.65 (0.34-1.24) | 0.193 |
| *ERBB2* | h111 | 6.5 | All | 0.84 (0.68-1.05) | 0.12 |
| | | | Serous | 0.85 (0.64-1.14) | 0.275 |
| | | | Endometrioid | 1.36 (0.89-2.06) | 0.151 |
| | | | Mucinous | 0.85 (0.48-1.48) | 0.563 |
| | | | Clear cell | 0.85 (0.44-1.64) | 0.626 |
| ***ERBB2*** | **Rare** | 5.6 | All | 0.7 (0.43-1.15) | 0.157 |
| | | | **Serous** | **0.46 (0.23-0.95)** | **0.035** |
| | | | Endometrioid | 1.14 (0.48-2.74) | 0.761 |
| | | | Mucinous | 0.61 (0.14-2.63) | 0.507 |
| | | | Clear cell | 0.93 (0.22-3.92) | 0.923 |

OR - odds ratio, CI - confidence interval; SNP order in haplotype (5' to 3') - rs2952155, rs2952156, rs1801200.

## Appendix III-E: Genotype specific ratios of *KRAS* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR[§] (95% CI) | HomOR[§] (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| *KRAS* | rs12305513 | 0.1 | 2934 | 1788 | All | 0.87 (0.74-1.03) | 0.71 (0.38-1.31) | 0.0526 |
| | | | | 852 | Serous | 0.86 (0.70-1.06) | 0.89 (0.42-1.89) | 0.1677 |
| | | | | 272 | Endometrioid | 0.75 (0.53-1.07) | 0.83 (0.25-2.76) | 0.1342 |
| | | | | 189 | Mucinous | 0.89 (0.6-1.33) | 1.22 (0.36-4.06) | 0.7436 |
| | | | | 136 | Clear cell | 1.06 (0.69-1.65) | 0.61 (0.08-4.56) | 0.9825 |
| ***KRAS*** | **rs12822857** | 0.47 | 2901 | 1751 | All | 1.01 (0.88-1.17) | 0.94 (0.80-1.12) | 0.5281 |
| | | | | 835 | Serous | 1.08 (0.90-1.30) | 0.97 (0.78-1.20) | 0.8167 |
| | | | | 268 | Endometrioid | 0.99 (0.73-1.33) | 1.06 (0.75-1.50) | 0.7605 |
| | | | | 187 | **Mucinous** | 0.74 (0.53-1.04) | **0.63 (0.41-0.96)** | **0.0232** |
| | | | | 132 | Clear cell | 1.14 (0.75-1.74) | 1.04 (0.63-1.72) | 0.8398 |
| *KRAS* | rs10842508 | 0.25 | 2935 | 1776 | All | 0.97 (0.86-1.10) | 0.95 (0.73-1.22) | 0.5789 |
| | | | | 841 | Serous | 0.95 (0.81-1.12) | 1.03 (0.74-1.41) | 0.7882 |
| | | | | 273 | Endometrioid | 1.05 (0.80-1.37) | 1.26 (0.77-2.05) | 0.4115 |
| | | | | 190 | Mucinous | 0.93 (0.68-1.27) | 0.46 (0.20-1.06) | 0.1345 |
| | | | | 136 | Clear cell | 0.93 (0.64-1.34) | 1.12 (0.57-2.21) | 0.9696 |
| *KRAS* | rs12579073 | 0.48 | 2900 | 1765 | All | 0.97 (0.84-1.12) | 0.92 (0.78-1.09) | 0.3591 |
| | | | | 836 | Serous | 1.01 (0.84-1.22) | 0.88 (0.70-1.10) | 0.2746 |
| | | | | 269 | Endometrioid | 0.95 (0.70-1.28) | 1.02 (0.72-1.44) | 0.9418 |
| | | | | 190 | Mucinous | 0.72 (0.51-1.01) | 0.74 (0.49-1.10) | 0.1074 |
| | | | | 135 | Clear cell | 1.29 (0.84-2.01) | 1.32 (0.80-2.18) | 0.2772 |
| *KRAS* | rs10842513 | 0.09 | 2878 | 1770 | All | 1.03 (0.87-1.21) | 0.93 (0.50-1.74) | 0.8581 |
| | | | | 846 | Serous | 0.91 (0.74-1.13) | 0.91 (0.41-2.01) | 0.4011 |
| | | | | 271 | Endometrioid | 1.28 (0.93-1.77) | 0.86 (0.20-3.67) | 0.2085 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 187 | Mucinous | 1.04 (0.69-1.57) | 0.62 (0.08-4.66) | 0.9898 |
| | | | | 137 | Clear cell | 1.26 (0.81-1.95) | 1.77 (0.41-7.57) | 0.2217 |
| | | | | 1748 | All | 0.96 (0.83-1.10) | 1.13 (0.77-1.67) | 0.845 |
| | | | | 834 | Serous | 0.93 (0.77-1.11) | 1.06 (0.64-1.76) | 0.5753 |
| *KRAS* | rs4623993 | 0.16 | 2892 | 268 | Endometrioid | 1.17 (0.88-1.53) | 1.21 (0.54-2.69) | 0.2623 |
| | | | | 187 | Mucinous | 0.73 (0.50-1.06) | 1.07 (0.42-2.72) | 0.2037 |
| | | | | 132 | Clear cell | 0.80 (0.52-1.23) | **2.55 (1.18-5.50)** | 0.5715 |
| | | | | 1763 | All | 1.04 (0.91-1.18) | 0.99 (0.82-1.19) | 0.9408 |
| | | | | 836 | Serous | 1.09 (0.92-1.29) | 0.98 (0.77-1.24) | 0.8783 |
| ***KRAS*** | **rs6487464** | 0.38 | 2895 | 269 | Endometrioid | 1.13 (0.86-1.5) | 1.15 (0.79-1.67 | 0.3878 |
| | | | | 192 | **Mucinous** | **0.61 (0.44-0.85)** | 0.76 (0.50-1.18) | **0.0379** |
| | | | | 136 | Clear cell | 1.02 (0.70-1.48) | 0.95 (0.56-1.61) | 0.8918 |
| | | | | 1757 | All | 0.98 (0.86-1.13) | 1.08 (0.91-1.29) | 0.4153 |
| | | | | 835 | Serous | 1.04 (0.87-1.24) | 1.09 (0.88-1.36) | 0.4379 |
| ***KRAS*** | **rs10842514** | 0.44 | 2886 | 269 | Endometrioid | 0.85 (0.64-1.13) | 0.97 (0.68-1.38) | 0.7294 |
| | | | | 188 | **Mucinous** | 1.13 (0.78-1.64) | **2.02 (1.35-3.01)** | **0.0006** |
| | | | | 134 | Clear cell | 0.95 (0.64-1.40) | 0.79 (0.47-1.33) | 0.4026 |
| | | | | 1476 | All | 0.92 (0.75-1.14) | 1.62 (0.57-4.57) | 0.7116 |
| | | | | 685 | Serous | 0.99 (0.76-1.29) | 1.91 (0.57-6.40) | 0.7757 |
| *KRAS* | rs11047917 | 0.06 | 2456 | 231 | Endometrioid | 0.68 (0.42-1.11) | 1.4 (0.17-11.64) | 0.171 |
| | | | | 163 | Mucinous | 0.76 (0.44-1.32) | 1.97 (0.24-16.40) | 0.4739 |
| | | | | 118 | Clear cell | 1.20 (0.69-2.06) | 2.87 (0.35-23.69) | 0.354 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix III-F: Haplotype-specific risks of *KRAS* on ovarian cancer susceptibility

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *KRAS* haplotype block 1 | h100 | 52.1 | All | 1.02 (0.94-1.11) | 0.66 |
| | | | Serous | 1.01 (0.9-1.12) | 0.894 |
| | | | Endometrioid | 0.92 (0.77-1.1) | 0.366 |
| | | | Mucinous | 1.23 (0.99-1.52) | 0.061 |
| | | | Clear cell | 0.97 (0.76-1.24) | 0.81 |
| *KRAS* haplotype block 1 | h000 | 22.8 | All | 1.00 (0.9 – 1.11) | 0.99 |
| | | | Serous | 1.03 (0.9-1.18) | 0.643 |
| | | | Endometrioid | 1.04 (0.84-1.29) | 0.736 |
| | | | Mucinous | 0.92 (0.71-1.2) | 0.53 |
| | | | Clear cell | 1.13 (0.84-1.52) | 0.406 |
| *KRAS* haplotype block 1 | h001 | 15.1 | All | 1.03 (0.91 – 1.16) | 0.67 |
| | | | Serous | 1.04 (0.89-1.22) | 0.591 |
| | | | Endometrioid | 1.25 (0.98-1.59) | 0.068 |
| | | | Mucinous | 0.79 (0.57-1.09) | 0.152 |
| | | | Clear cell | 0.98 (0.69-1.4) | 0.917 |
| *KRAS* haplotype block 1 | h101 | 9.5 | All | 0.89 (0.77 – 1.04) | 0.15 |
| | | | Serous | 0.88 (0.72-1.07) | 0.195 |
| | | | Endometrioid | 0.85 (0.62-1.18) | 0.329 |
| | | | Mucinous | 0.9 (0.62-1.31) | 0.584 |
| | | | Clear cell | 0.9 (0.58-1.4) | 0.652 |
| ***KRAS* haplotype block 1** | **Rare** | 0.1 | **All** | 0.43 (0.17-1.06) | **0.0465** |
| | | | Serous | 0.54 (0.19-1.55) | 0.253 |
| | | | Endometrioid | 0.53 (0.08-3.74) | 0.527 |
| | | | Mucinous | 0.7 (0.09-5.2) | 0.724 |
| | | | Clear cell | 0.02 (0-19174) | 0.587 |
| ***KRAS* haplotype block 2** | **h000010** | 30.6 | All | 1.04 (0.95-1.15) | 0.389 |
| | | | Serous | 1.07 (0.94-1.21) | 0.306 |
| | | | Endometrioid | 0.99 (0.81-1.21) | 0.916 |
| | | | **Mucinous** | **1.3 (1.03-1.64)** | **0.025** |
| | | | Clear cell | 0.79 (0.59-1.06) | 0.121 |
| *KRAS* haplotype block 2 | h100010 | 12.9 | All | 1.00 (0.85-1.17) | 0.98 |
| | | | Serous | 0.97 (0.82-1.16) | 0.772 |
| | | | Endometrioid | 1 (0.72-1.37) | 0.979 |
| | | | Mucinous | 1.37 (0.99-1.89) | 0.058 |
| | | | Clear cell | 1.07 (0.71-1.62) | 0.74 |
| *KRAS* haplotype block 2 | h100100 | 11.9 | All | 1.03 (0.88-1.19) | 0.75 |
| | | | Serous | 1.08 (0.9-1.3) | 0.381 |
| | | | Endometrioid | 1.16 (0.88-1.53) | 0.301 |
| | | | Mucinous | 0.78 (0.54-1.13) | 0.195 |
| | | | Clear cell | 0.87 (0.57-1.33) | 0.521 |
| *KRAS* haplotype block 2 | h101100 | 10.6 | All | 0.98 (0.84-1.14) | 0.77 |
| | | | Serous | 0.96 (0.8-1.16) | 0.703 |
| | | | Endometrioid | 1.15 (0.86-1.53) | 0.337 |

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| | | | Mucinous | 0.75 (0.51-1.1) | 0.143 |
| | | | Clear cell | 1.21 (0.82-1.79) | 0.326 |
| *KRAS* haplotype block 2 | h010000 | 5.5 | All | 1.01 (0.82-1.25) | 0.91 |
| | | | Serous | 0.92 (0.7-1.2) | 0.526 |
| | | | Endometrioid | 1.11 (0.74-1.66) | 0.629 |
| | | | Mucinous | 0.78 (0.45-1.35) | 0.375 |
| | | | Clear cell | 1.36 (0.8-2.29) | 0.254 |
| *KRAS* haplotype block 2 | h000100 | 5.4 | All | 0.92 (0.73-1.18) | 0.52 |
| | | | Serous | 0.91 (0.68-1.2) | 0.487 |
| | | | Endometrioid | 1.09 (0.66-1.82) | 0.731 |
| | | | Mucinous | 1.09 (0.61-1.95) | 0.775 |
| | | | Clear cell | 1.13 (0.59-2.16) | 0.708 |
| *KRAS* haplotype block 2 | h100000 | 5.4 | All | 0.81 (0.65-1.01) | 0.06 |
| | | | Serous | 0.8 (0.6-1.07) | 0.136 |
| | | | Endometrioid | 0.68 (0.42-1.11) | 0.123 |
| | | | Mucinous | 0.61 (0.33-1.12) | 0.11 |
| | | | Clear cell | 1.07 (0.6-1.92) | 0.817 |
| *KRAS* haplotype block 2 | h001100 | 4.7 | All | 0.96 (0.74-1.24) | 0.74 |
| | | | Serous | 0.95 (0.7-1.29) | 0.759 |
| | | | Endometrioid | 1 (0.57-1.78) | 0.988 |
| | | | Mucinous | 1.15 (0.62-2.12) | 0.664 |
| | | | Clear cell | 0.66 (0.26-1.65) | 0.369 |
| *KRAS* haplotype block 2 | h100101 | 4.3 | All | 0.82 (0.64-1.05) | 0.11 |
| | | | Serous | 0.92 (0.66-1.27) | 0.616 |
| | | | Endometrioid | 0.82 (0.48-1.38) | 0.45 |
| | | | Mucinous | 0.94 (0.53-1.69) | 0.845 |
| | | | Clear cell | 1.33 (0.71-2.46) | 0.371 |
| *KRAS* haplotype block 2 | h110000 | 3.2 | All | 0.94 (0.70-1.27) | 0.69 |
| | | | Serous | 0.78 (0.53-1.16) | 0.224 |
| | | | Endometrioid | 1.16 (0.63-2.16) | 0.633 |
| | | | Mucinous | 1.12 (0.54-2.35) | 0.761 |
| | | | Clear cell | 1.68 (0.84-3.36) | 0.139 |
| *KRAS* haplotype block 2 | h000000 | 3.1 | All | 0.89 (0.64-1.23) | 0.48 |
| | | | Serous | 1.08 (0.75-1.58) | 0.67 |
| | | | **Endometrioid** | **0.36 (0.14-0.92)** | **0.033** |
| | | | **Mucinous** | **0.3 (0.09-0.99)** | **0.049** |
| | | | Clear cell | 0.61 (0.21-1.78) | 0.365 |
| ***KRAS* haplotype block 2** | **Rare** | 1.7 | All | 1.27 (0.93-1.73) | 0.131 |
| | | | **Serous** | **1.55 (1.09-2.2)** | **0.014** |
| | | | Endometrioid | 0.93 (0.42-2.07) | 0.861 |
| | | | Mucinous | 0.83 (0.31-2.23) | 0.712 |
| | | | Clear cell | 1.14 (0.44-2.96) | 0.793 |

OR - odds ratio, CI - confidence interval; SNP order in haplotype (5' to 3') haplotype block 1 - rs12305513, rs12822857, rs10842508; haplotype block 2 - rs12579073, rs10842513, rs4623993, rs6487464, rs10842514, rs11047917.

## Appendix III-G: Genotype specific risks of *NMI* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR[§] (95% CI) | HomOR[§] (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 1708 | All | 1.01 (0.88-1.17) | 1.40 (0.84-2.32) | 0.474 |
| | | | | 809 | Serous | 0.98 (0.81-1.18) | 1.56 (0.85-2.87) | 0.6587 |
| *NMI* | rs394884 | 0.15 | 2852 | 260 | Endometrioid | 1.01 (0.75-1.36) | 1.97 (0.81-4.82) | 0.4804 |
| | | | | 184 | Mucinous | 0.98 (0.69-1.39) | 0.45 (0.06-3.31) | 0.6601 |
| | | | | 129 | Clear cell | 1.26 (0.85-1.86) | 0.68 (0.09-5.03) | 0.3847 |
| | | | | 1159 | All | 0.96 (0.76-1.23) | 1.23 (0.45-3.38) | 0.9163 |
| | | | | 524 | Serous | 0.90 (0.65-1.24) | 1.23 (0.33-4.58) | 0.6308 |
| *NMI* | rs11551174 | 0.06 | 2040 | 185 | Endometrioid | 1.25 (0.78-1.99) | 1.18 (0.14-9.63) | 0.3592 |
| | | | | 133 | Mucinous | 1.05 (0.59-1.87) | 1.61 (0.20-13.17) | 0.7416 |
| | | | | 95 | Clear cell | 0.69 (0.31-1.51) | 2.41 (0.30-19.70) | 0.608 |
| | | | | 1665 | All | 1.05 (0.91-1.22) | 1.08 (0.61-1.89) | 0.4843 |
| | | | | 792 | Serous | 1.04 (0.86-1.27) | 1.47 (0.78-2.79) | 0.3594 |
| *NMI* | rs289831 | 0.13 | 2718 | 258 | Endometrioid | 1.13 (0.84-1.53) | 1.07 (0.32-3.58) | 0.4472 |
| | | | | 176 | Mucinous | 0.89 (0.61-1.31) | 0.48 (0.06-3.53) | 0.4077 |
| | | | | 132 | Clear cell | 1.23 (0.82-1.84) | 0.71 (0.09-5.26) | 0.448 |
| | | | | 1764 | All | 1.03 (0.90-1.18) | **1.19 (1.00-1.42)** | 0.0753 |
| | | | | 843 | Serous | 1.10 (0.92-1.31) | **1.25 (1.00-1.56)** | 0.0546 |
| *NMI* | rs3771886 | 0.41 | 2927 | 266 | Endometrioid | 0.85 (0.65-1.12) | 0.85 (0.58-1.24) | 0.2876 |
| | | | | 191 | Mucinous | 1.02 (0.73-1.43) | **1.51 (1.01-2.27)** | 0.0747 |
| | | | | 132 | Clear cell | 0.92 (0.62-1.35) | 0.90 (0.53-1.53) | 0.6514 |
| ***NMI*** | **rs11683487** | 0.46 | 2564 | 1464 | **All** | **0.80 (0.69-0.93)** | 0.87 (0.71-1.02) | **0.0379** |
| | | | | 713 | **Serous** | **0.81 (0.67-0.98)** | 0.80 (0.63-1.01) | **0.0377** |
| | | | | 227 | Endometrioid | 0.82 (0.59-1.13) | 1.09 (0.75-1.57) | 0.7567 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 154 | **Mucinous** | **0.67 (0.47-0.96)** | **0.62 (0.39-0.99)** | **0.0269** |
| | | | | 107 | Clear cell | 0.97 (0.62-1.52) | 0.90 (0.52-1.57) | 0.7186 |
| | | | | 1776 | All | 1.05 (0.91-1.21) | 1.16 (0.68-1.97) | 0.4222 |
| | | | | 843 | Serous | 1.03 (0.86-1.25) | 1.37 (0.73-2.57) | 0.457 |
| *NMI* | rs2113509 | 0.13 | 2944 | 272 | Endometrioid | 1.21 (0.91-1.62) | 1.02 (0.31-3.38) | 0.2417 |
| | | | | 190 | Mucinous | 0.90 (0.62-1.31) | 1.38 (0.42-4.61) | 0.8066 |
| | | | | 138 | Clear cell | 1.21 (0.81-1.79) | 0.65 (0.09-4.82) | 0.5121 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix III-H: Haplotype-specific risks of *NMI* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| | | | **All** | **0.91 (0.84-0.99)** | **0.027** |
| | | | **Serous** | **0.89 (0.8-1)** | **0.048** |
| ***NMI*** | **h00001** | 45.9 | Endometrioid | 1.05 (0.87-1.25) | 0.631 |
| | | | Mucinous | 0.84 (0.68-1.05) | 0.128 |
| | | | Clear cell | 0.96 (0.74-1.23) | 0.729 |
| | | | All | **1.11 (1.003-1.22)** | **0.043** |
| | | | Serous | **1.13 (1.01-1.28)** | **0.041** |
| ***NMI*** | **h00010** | 33.7 | Endometrioid | 0.9 (0.74-1.1) | 0.305 |
| | | | Mucinous | 1.11 (0.88-1.4) | 0.361 |
| | | | Clear cell | 0.97 (0.74-1.28) | 0.832 |
| | | | All | 1.09 (0.95-1.25) | 0.22 |
| | | | Serous | 1.04 (0.87-1.23) | 0.686 |
| *NMI* | h10100 | 11.8 | Endometrioid | 1.14 (0.87-1.5) | 0.346 |
| | | | Mucinous | 1 (0.71-1.4) | 0.979 |
| | | | Clear cell | 1.06 (0.72-1.57) | 0.774 |
| | | | All | 1.05 (0.84-1.3) | 0.67 |
| | | | Serous | 0.98 (0.73-1.32) | 0.915 |
| *NMI* | h01010 | 5.7 | Endometrioid | 1.01 (0.63-1.63) | 0.951 |
| | | | Mucinous | 1.23 (0.75-2.03) | 0.418 |
| | | | Clear cell | 0.85 (0.43-1.69) | 0.643 |
| | | | All | 1.11 (0.87-1.43) | 0.399 |
| | | | Serous | 0.94 (0.66-1.33) | 0.713 |
| *NMI* | Rare | 1.9 | Endometrioid | 0.85 (0.47-1.56) | 0.607 |
| | | | Mucinous | 1.4 (0.78-2.51) | 0.257 |
| | | | Clear cell | 1.7 (0.9-3.19) | 0.101 |

OR - odds ratio, CI - confidence interval; SNP order in haplotype (5' to 3') - rs394884, rs11551174, rs289831, rs3771886, rs11683487.

## Appendix III-I: Genotype specific risks of *PIK3CA* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR[§] (95% CI) | HomOR[§] (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 1164 | All | 1.14 (0.89-1.45) | **0.43 (0.37-0.50)** | 0.294 |
| | | | | 525 | Serous | 1.13 (0.83-1.55) | **0.77 (0.63-0.93)** | 0.4279 |
| ***PIK3CA*** | **rs2865084** | 0.06 | 2046 | 183 | **Endometrioid** | **1.60 (1.03-2.50)** | **0.30 (0.22-0.42)** | **0.0344** |
| | | | | 135 | Mucinous | 1.32 (0.77-2.25) | **0.32 (0.22-0.46)** | 0.3093 |
| | | | | 95 | Clear cell | 0.51 (0.21-1.28) | 0.37 (0.24-0.57) | 0.1465 |
| | | | | 1749 | All | 0.99 (0.86-1.13) | 1.23 (0.86-1.77) | 0.6387 |
| | | | | 834 | Serous | 1.04 (0.88-1.24) | **1.59 (1.04-2.43)** | 0.1222 |
| *PIK3CA* | rs7621329 | 0.16 | 2818 | 268 | Endometrioid | 0.95 (0.71-1.27) | 1.44 (0.73-2.87) | 0.7763 |
| | | | | 186 | Mucinous | 0.90 (0.64-1.27) | 0.39 (0.09-1.60) | 0.2353 |
| | | | | 135 | Clear cell | 1.05 (0.71-1.54) | 0.57 (0.14-2.37) | 0.826 |
| | | | | 1739 | All | 0.98 (0.83-1.15) | 0.77 (0.42-1.40) | 0.5448 |
| | | | | 827 | Serous | 0.98 (0.79-1.20) | 0.78 (0.36-1.70) | 0.6412 |
| *PIK3CA* | rs1517586 | 0.1 | 2908 | 267 | Endometrioid | 0.90 (0.64-1.28) | 0.62 (0.15-2.61) | 0.4212 |
| | | | | 183 | Mucinous | 1.16 (0.79-1.71) | 0.49 (0.07-3.64) | 0.7202 |
| | | | | 134 | Clear cell | 0.82 (0.50-1.35) | 1.22 (0.29-5.16) | 0.5886 |
| | | | | 1741 | All | 1.01 (0.88-1.15) | 0.89 (0.71-1.11) | 0.4877 |
| | | | | 825 | Serous | 1.00 (0.84-1.18) | 0.84 (0.63-1.13) | 0.4006 |
| *PIK3CA* | rs2699905 | 0.27 | 2855 | 266 | Endometrioid | 0.97 (0.73-1.28) | 1.08 (0.70-1.67) | 0.8835 |
| | | | | 184 | Mucinous | 1.09 (0.80-1.50) | 0.72 (0.39-1.33) | 0.65 |
| | | | | 135 | Clear cell | 1.02 (0.70-1.48) | 1.04 (0.57-1.90) | 0.893 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR§ (95% CI) | HomOR§ (95% CI) | *P*-trend |
|---|---|---|---|---|---|---|---|---|
| *PIK3CA* | rs7641889 | 0.07 | 2939 | 1779 | All | 0.89 (0.74-1.07) | 1.28 (0.58-2.84) | 0.377 |
| | | | | 845 | Serous | 0.90 (0.71-1.14) | 2.22 (0.96-5.14) | 0.989 |
| | | | | 273 | Endometrioid | 0.75 (0.50-1.14) | 0.78 (0.10-6.05) | 0.1854 |
| | | | | 192 | Mucinous | 0.88 (0.56-1.40) | 1.14 (0.15-8.78) | 0.655 |
| | | | | 136 | Clear cell | 1.04 (0.63-1.74) | 1.68 (0.22-12.93) | 0.741 |
| *PIK3CA* | rs7651265 | 0.1 | 2883 | 1794 | All | 0.89 (0.76-1.04) | 1.58 (0.89-2.80) | 0.5447 |
| | | | | 828 | Serous | 0.89 (0.73-1.09) | **2.08 (1.09-3.98)** | 0.9988 |
| | | | | 267 | Endometrioid | 0.90 (0.65-1.24) | 1.80 (0.61-5.31) | 0.8684 |
| | | | | 189 | Mucinous | 0.89 (0.61-1.30) | 0.63 (0.08-4.69) | 0.4749 |
| | | | | 136 | Clear cell | 0.98 (0.64-1.51) | 0.90 (0.12-6.78) | 0.9072 |
| *PIK3CA* | rs7640662 | 0.15 | 2916 | 1765 | All | 1.02 (0.89-1.17) | 0.85 (0.57-1.27) | 0.8572 |
| | | | | 842 | Serous | 0.99 (0.83-1.18) | 0.58 (0.31-1.07) | 0.3077 |
| | | | | 268 | Endometrioid | 1.15 (0.87-1.54) | 1.46 (0.74-2.91) | 0.1748 |
| | | | | 188 | Mucinous | 1.02 (0.73-1.44) | 0.61 (0.19-1.96) | 0.7189 |
| | | | | 135 | Clear cell | 1.04 (0.69-1.55) | 1.41 (0.55-3.58) | 0.5889 |
| *PIK3CA* | rs2677760 | 0.49 | 2925 | 1762 | All | 1.01 (0.87-1.16) | 1.04 (0.88-1.23) | 0.663 |
| | | | | 836 | Serous | 0.92 (0.76-1.10) | 0.96 (0.77-1.19) | 0.6718 |
| | | | | 268 | Endometrioid | 1.08 (0.80-1.47) | 1.02 (0.71-1.46) | 0.9136 |
| | | | | 189 | Mucinous | 1.29 (0.89-1.89) | 1.37 (0.89-2.11) | 0.1488 |
| | | | | 134 | Clear cell | 1.03 (0.68-1.56) | 0.90 (0.55-1.49) | 0.7044 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix III-J: Haplotype-specific risks of *PIK3CA* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *PIK3CA* | h00000001 | 48.2 | All | 1.02 (0.93-1.11) | 0.713 |
| | | | Serous | 0.99 (0.89-1.11) | 0.866 |
| | | | Endometrioid | 1.01 (0.84-1.21) | 0.914 |
| | | | Mucinous | 1.17 (0.95-1.45) | 0.146 |
| | | | Clear cell | 0.98 (0.76-1.26) | 0.871 |
| *PIK3CA* | h00010010 | 14.8 | All | 1.01 (0.89-1.14) | 0.91 |
| | | | Serous | 0.95 (0.81-1.11) | 0.543 |
| | | | Endometrioid | 1.13 (0.89-1.45) | 0.31 |
| | | | Mucinous | 0.93 (0.69-1.26) | 0.649 |
| | | | Clear cell | 1.08 (0.77-1.51) | 0.67 |
| *PIK3CA* | h00000000 | 10.2 | All | 0.94 (0.81-1.09) | 0.39 |
| | | | Serous | 1 (0.83-1.2) | 0.996 |
| | | | Endometrioid | 0.82 (0.59-1.13) | 0.223 |
| | | | Mucinous | 0.85 (0.59-1.25) | 0.414 |
| | | | Clear cell | 1.1 (0.73-1.65) | 0.657 |
| *PIK3CA* | h00110000 | 9.7 | All | 0.98 (0.84-1.14) | 0.79 |
| | | | Serous | 0.97 (0.81-1.17) | 0.782 |
| | | | Endometrioid | 0.86 (0.63-1.19) | 0.374 |
| | | | Mucinous | 1.08 (0.76-1.53) | 0.664 |
| | | | Clear cell | 0.89 (0.58-1.38) | 0.612 |
| *PIK3CA* | h01001100 | 6.6 | All | 0.94 (0.79-1.12) | 0.5 |
| | | | Serous | 0.98 (0.79-1.23) | 0.881 |
| | | | Endometrioid | 0.84 (0.57-1.24) | 0.372 |
| | | | Mucinous | 0.81 (0.51-1.29) | 0.369 |
| | | | Clear cell | 1.02 (0.62-1.68) | 0.932 |
| *PIK3CA* | h01000100 | 4 | All | 1.02 (0.82-1.28) | 0.84 |
| | | | Serous | 1.06 (0.8-1.4) | 0.698 |
| | | | Endometrioid | 0.98 (0.61-1.58) | 0.932 |
| | | | Mucinous | 0.82 (0.45-1.5) | 0.522 |
| | | | Clear cell | 1.17 (0.63-2.16) | 0.625 |
| ***PIK3CA*** | **h11000000** | 3.9 | All | 1.20 (0.97-1.48) | 0.102 |
| | | | Serous | 1.29 (0.99-1.67) | 0.055 |
| | | | **Endometrioid** | **1.49 (1-2.22)** | **0.049** |
| | | | Mucinous | 1.07 (0.63-1.82) | 0.795 |
| | | | Clear cell | 0.91 (0.47-1.79) | 0.793 |
| *PIK3CA* | Rare | 1.6 | All | 0.94 (0.73-1.21) | 0.633 |
| | | | Serous | 0.96 (0.69-1.33) | 0.795 |
| | | | Endometrioid | 1.07 (0.68-1.68) | 0.786 |
| | | | Mucinous | 0.66 (0.27-1.58) | 0.345 |
| | | | Clear cell | 0.71 (0.27-1.9) | 0.499 |

OR - odds ratio, CI - confidence interval; SNP order in haplotype (5' to 3') – rs2865084, rs7621329, rs1517586, rs2699905, rs7641889, rs7651265, rs7640662, rs2677760.

## Appendix IV-A: Genotype specific ratios of *AIFM2* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| *AIFM2* | rs2394655 | 0.04 | 2924 | 1751 | All | 1.01 (0.80-1.27) | 1.37 (0.42-4.42) | 0.7773 |
| | | | | 827 | Serous | 1.12 (0.84-1.49) | 1.03 (0.21-5.04) | 0.4297 |
| | | | | 269 | Endometrioid | 0.65 (0.36-1.16) | 2.21 (0.26-18.68) | 0.2592 |
| | | | | 189 | Mucinous | 1.11 (0.64-1.94) | 4.09 (0.48-35.13) | 0.4945 |
| | | | | 164 | Clear cell | 1.53 (0.90-2.58) | 3.71 (0.44-31.65) | 0.064 |
| *AIFM2* | rs7908957 | 0.13 | 2873 | 1719 | All | 0.92 (0.79-1.07) | 1.13 (0.73-1.75) | 0.5342 |
| | | | | 817 | Serous | 0.91 (0.75-1.11) | 0.91 (0.50-1.67) | 0.3812 |
| | | | | 264 | Endometrioid | **0.71 (0.50-1.00)** | **2.06 (1.01-4.18)** | 0.5702 |
| | | | | 184 | Mucinous | 1.03 (0.71-1.48) | 1.84 (0.75-4.49) | 0.4123 |
| | | | | 159 | Clear cell | 1.21 (0.84-1.76) | 1.05 (0.32-3.44) | 0.3951 |
| *AIFM2* | rs1053495 | 0.07 | 2704 | 1697 | All | 0.97 (0.81-1.16) | 0.72 (0.35-1.51) | 0.457 |
| | | | | 790 | Serous | 0.96 (0.75-1.21) | 0.88 (0.35-2.18) | 0.6293 |
| | | | | 267 | Endometrioid | **0.64 (0.41-0.98)** | 0.84 (0.19-3.65) | 0.0622 |
| | | | | 185 | Mucinous | 1.11 (0.71-1.71) | 0.52 (0.06-4.13) | 0.9899 |
| | | | | 156 | Clear cell | 1.38 (0.89-2.12) | 1.51 (0.35-6.57) | 0.1241 |
| *AIFM2* | rs2894111 | 0.28 | 2861 | 1770 | All | 0.99 (0.87-1.13) | 0.93 (0.74-1.16) | 0.5545 |
| | | | | 835 | Serous | 1.02 (0.87-1.20) | 0.86 (0.64-1.17) | 0.5896 |
| | | | | 276 | Endometrioid | 0.87 (0.67-1.14) | 1.26 (0.82-1.94) | 0.8942 |
| | | | | 192 | Mucinous | 1.04 (0.76-1.42) | 1.03 (0.59-1.80) | 0.8834 |
| | | | | 164 | Clear cell | 1.13 (0.81-1.57) | 1.27 (0.73-2.22) | 0.3776 |
| *AIFM2* | rs2394656 | 0.19 | 1703 | 913 | All | 0.96 (0.80-1.15) | 0.85 (0.55-1.34) | 0.4114 |
| | | | | 506 | Serous | 0.93 (0.75-1.17) | 0.85 (0.50-1.47) | 0.3965 |
| | | | | 136 | Endometrioid | 0.91 (0.61-1.35) | 0.87 (0.33-2.26) | 0.5861 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 85 | Mucinous | 1.11 (0.69-1.77) | 0.58 (0.14-2.44) | 0.9529 |
| | | | | 78 | Clear cell | 1.13 (0.69-1.86) | 0.99 (0.30-3.31) | 0.7857 |
| | | | | 422 | All | 1.01 (0.79-1.29) | 0.88 (0.55-1.42) | 0.7992 |
| | | | | 261 | Serous | 1.07 (0.80-1.43) | 0.75 (0.40-1.39) | 0.8126 |
| *AIFM2* | rs6480440 | 0.24 | 1140 | 56 | Endometrioid | 0.66 (0.35-1.27) | 1.50 (0.61-3.69) | 0.9252 |
| | | | | 42 | Mucinous | 1.20 (0.62-2.33) | 1.56 (0.52-4.64) | 0.3371 |
| | | | | 28 | Clear cell | 1.14 (0.52-2.53) | 0.55 (0.07-4.17) | 0.8263 |
| | | | | 1313 | All | 1.05 (0.88-1.25) | 0.96 (0.53-1.71) | 0.59 |
| | | | | 556 | Serous | 1.11 (0.88-1.40) | 0.97 (0.44-2.12) | 0.5089 |
| *AIFM2* | rs2280201 | 0.12 | 1783 | 216 | Endometrioid | 1.03 (0.73-1.47) | 1.72 (0.67-4.42) | 0.3743 |
| | | | | 195 | Mucinous | 1.03 (0.67-1.60) | 1.01 (0.24-4.32) | 0.9688 |
| | | | | 164 | Clear cell | 1.39 (0.97-1.99) | 0.40 (0.05-2.93) | 0.2643 |
| | | | | 1285 | All | 1.25 (1.03-1.51) | 0.48 (0.16-1.47) | 0.2055 |
| | | | | 600 | Serous | 1.13 (0.88-1.46) | 0.79 (0.23-2.75) | 0.6396 |
| *AIFM2* | rs10999147 | 0.09 | 2395 | 277 | Endometrioid | 1.07 (0.75-1.52) | 0.37 (0.05-2.74) | 0.8708 |
| | | | | 194 | Mucinous | 0.87 (0.56-1.36) | 1.17 (0.27-5.02) | 0.606 |
| | | | | 165 | Clear cell | 1.43 (0.95-2.15) | 0.64 (0.09-4.81) | 0.2092 |
| | | | | 1743 | All | 1.02 (0.83-1.24) | 2.28 (0.86-6.04) | 0.54 |
| | | | | 831 | Serous | 1.12 (0.88-1.44) | **3.43 (1.23-9.58)** | 0.1043 |
| *AIFM2* | rs3750772 | 0.06 | 2944 | 266 | Endometrioid | 1.14 (0.76-1.71) | 2.63 (0.32-21.75) | 0.4623 |
| | | | | 186 | Mucinous | 0.81 (0.47-1.41) | 3.02 (0.36-25.58) | 0.5111 |
| | | | | 163 | Clear cell | 1.07 (0.63-1.80) | 2.96 (0.36-24.53) | 0.6195 |
| *AIFM2* | rs4295944 | 0.42 | 1784 | 1335 | All | 0.98 (0.83-1.15) | 1.09 (0.88-1.34) | 0.4913 |
| | | | | 567 | Serous | 0.95 (0.77-1.18) | 1.11 (0.84-1.46) | 0.6042 |
| | | | | 220 | Endometrioid | 0.87 (0.64-1.20) | 0.88 (0.57-1.35) | 0.3373 |
| | | | | 149 | Mucinous | 1.18 (0.78-1.79) | 1.50 (0.90-2.50) | 0.1196 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 133 | Clear cell | 0.81 (0.55-1.20) | 0.86 (0.51-1.44) | 0.413 |
| *AIFM2* | rs2394644 | 0.13 | 1685 | 1324 | All | 1.05 (0.88-1.26) | 1.03 (0.60-1.79) | 0.5245 |
| | | | | 561 | Serous | 1.02 (0.80-1.29) | 1.11 (0.55-2.25) | 0.7103 |
| | | | | 218 | Endometrioid | 1.13 (0.80-1.59) | 1.64 (0.67-4.03) | 0.1308 |
| | | | | 149 | Mucinous | 0.64 (0.39-1.06) | 0.42 (0.06-3.15) | 0.0727 |
| | | | | 133 | Clear cell | 1.45 (0.97-2.16) | 1.44 (0.43-4.83) | 0.073 |
| *AIFM2* | rs10999152 | 0.18 | 2610 | 1618 | All | 1.14 (0.99-1.32) | 0.96 (0.68-1.36) | 0.2066 |
| | | | | 760 | Serous | 1.17 (0.98-1.40) | 1.01 (0.65-1.57) | 0.2171 |
| | | | | 251 | Endometrioid | 1.18 (0.88-1.58) | 1.51 (0.82-2.78) | 0.0767 |
| | | | | 170 | Mucinous | 0.83 (0.57-1.20) | 0.50 (0.15-1.61) | 0.1659 |
| | | | | 153 | Clear cell | **1.46 (1.03-2.08)** | 1.22 (0.52-2.87) | 0.0666 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-B: Haplotype-specific risks of *AIFM2* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *AIFM2* (haplotype block 1) | h0000000 | 69.8 | All | 1.04 (0.9-1.2) | 0.602 |
| | | | Serous | 1.05 (0.93-1.18) | 0.465 |
| | | | Endometrioid | 0.99 (0.82-1.2) | 0.941 |
| | | | Mucinous | 1 (0.8-1.24) | 0.966 |
| | | | Clear cell | 0.95 (0.75-1.21) | 0.672 |
| *AIFM2* (haplotype block 1) | h0001011 | 6.5 | All | 1.22 (0.94-1.59) | 0.13 |
| | | | Serous | 1.11 (0.9-1.37) | 0.348 |
| | | | Endometrioid | 1.15 (0.82-1.61) | 0.418 |
| | | | Mucinous | 1.25 (0.85-1.84) | 0.253 |
| | | | Clear cell | 1.44 (0.97-2.14) | 0.073 |
| *AIFM2* (haplotype block 1) | h1111110 | 4.2 | All | 1 (0.71-1.42) | 0.997 |
| | | | Serous | 1.11 (0.84-1.47) | 0.458 |
| | | | Endometrioid | 0.74 (0.43-1.27) | 0.272 |
| | | | Mucinous | 1.11 (0.64-1.9) | 0.714 |
| | | | Clear cell | 1.5 (0.89-2.51) | 0.127 |
| *AIFM2* (haplotype block 1) | h0001100 | 3.5 | All | 0.8 (0.54-1.17) | 0.241 |
| | | | Serous | 0.91 (0.67-1.24) | 0.537 |
| | | | Endometrioid | 1.16 (0.74-1.83) | 0.511 |
| | | | Mucinous | 0.97 (0.55-1.72) | 0.919 |
| | | | Clear cell | 0.88 (0.46-1.66) | 0.689 |
| *AIFM2* (haplotype block 1) | h0001010 | 2.8 | All | 0.98 (0.66-1.46) | 0.929 |
| | | | Serous | 1.02 (0.74-1.42) | 0.885 |
| | | | Endometrioid | 1.09 (0.65-1.8) | 0.751 |
| | | | Mucinous | 0.76 (0.38-1.52) | 0.43 |
| | | | Clear cell | 0.76 (0.36-1.59) | 0.463 |
| *AIFM2* (haplotype block 1) | h0101111 | 2.8 | All | 1.06 (0.7-1.6) | 0.801 |
| | | | Serous | 1 (0.72-1.38) | 0.976 |
| | | | Endometrioid | 1.12 (0.69-1.83) | 0.646 |
| | | | Mucinous | 1.22 (0.7-2.13) | 0.481 |
| | | | Clear cell | 0.8 (0.39-1.65) | 0.544 |
| *AIFM2* (haplotype block 1) | h0111110 | 2.8 | All | 0.79 (0.5-1.24) | 0.309 |
| | | | Serous | 0.8 (0.57-1.11) | 0.184 |
| | | | Endometrioid | 0.69 (0.38-1.24) | 0.209 |
| | | | Mucinous | 1.21 (0.71-2.09) | 0.481 |
| | | | Clear cell | 1.35 (0.78-2.35) | 0.289 |
| *AIFM2* (haplotype block 1) | Rare | 1 | All | 0.98 (0.62-1.55) | 0.926 |
| | | | Serous | 0.84 (0.67-1.05) | 0.128 |
| | | | Endometrioid | 1 (0.71-1.4) | 0.993 |
| | | | Mucinous | 0.68 (0.42-1.08) | 0.101 |
| | | | Clear cell | 0.65 (0.39-1.09) | 0.104 |
| ***AIFM2* (haplotype block 2)** | **h00100** | 39 | All | 1.01 (0.91-1.11) | 0.856 |
| | | | Serous | 1.02 (0.91-1.14) | 0.702 |
| | | | Endometrioid | 0.96 (0.8-1.15) | 0.637 |
| | | | **Mucinous** | **1.26 (1.02-1.55)** | **0.034** |
| | | | Clear cell | 0.91 (0.72-1.15) | 0.437 |
| *AIFM2* (haplotype block 2) | h00000 | 36.3 | All | 0.92 (0.83-1.02) | 0.119 |
| | | | Serous | 0.92 (0.82-1.04) | 0.197 |
| | | | Endometrioid | 1.01 (0.84-1.22) | 0.907 |
| | | | Mucinous | 0.94 (0.75-1.18) | 0.6 |
| | | | Clear cell | 0.99 (0.78-1.26) | 0.939 |

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *AIFM2* (haplotype block 2) | h00011 | 6.7 | All | 1.12 (0.92-1.36) | 0.253 |
| | | | Serous | 1.11 (0.89-1.39) | 0.359 |
| | | | Endometrioid | 1.18 (0.83-1.68) | 0.346 |
| | | | Mucinous | 0.77 (0.47-1.26) | 0.294 |
| | | | Clear cell | 1.3 (0.84-2) | 0.24 |
| *AIFM2* (haplotype block 2) | h00001 | 4.1 | All | 1.06 (0.83-1.35) | 0.665 |
| | | | Serous | 0.97 (0.73-1.3) | 0.849 |
| | | | Endometrioid | 1.05 (0.67-1.67) | 0.823 |
| | | | Mucinous | 1.12 (0.66-1.91) | 0.675 |
| | | | Clear cell | 1.13 (0.63-2.02) | 0.686 |
| *AIFM2* (haplotype block 2) | h10000 | 3.2 | All | 1.15 (0.84-1.58) | 0.387 |
| | | | Serous | 1.18 (0.83-1.67) | 0.352 |
| | | | Endometrioid | 0.73 (0.37-1.42) | 0.351 |
| | | | Mucinous | 1.11 (0.56-2.17) | 0.719 |
| | | | Clear cell | 1.14 (0.55-2.35) | 0.719 |
| *AIFM2* (haplotype block 2) | h10011 | 2.3 | All | 0.98 (0.68-1.4) | 0.895 |
| | | | Serous | 1.01 (0.65-1.56) | 0.97 |
| | | | Endometrioid | 0.64 (0.29-1.44) | 0.286 |
| | | | Mucinous | 0.41 (0.13-1.26) | 0.119 |
| | | | Clear cell | 1.1 (0.47-2.54) | 0.833 |
| *AIFM2* (haplotype block 2) | h01011 | 2.2 | All | 1.1 (0.8-1.53) | 0.551 |
| | | | Serous | 1.33 (0.93-1.9) | 0.112 |
| | | | Endometrioid | 1.53 (0.88-2.65) | 0.129 |
| | | | Mucinous | 0.66 (0.27-1.62) | 0.362 |
| | | | Clear cell | 0.81 (0.32-2.03) | 0.65 |
| *AIFM2* (haplotype block 2) | h10100 | 2.2 | All | 1.33 (0.92-1.93) | 0.126 |
| | | | Serous | 1.06 (0.68-1.65) | 0.792 |
| | | | Endometrioid | 1.43 (0.76-2.68) | 0.263 |
| | | | Mucinous | 1.24 (0.57-2.67) | 0.591 |
| | | | Clear cell | 0.83 (0.31-2.22) | 0.709 |
| *AIFM2* (haplotype block 2) | Rare | 0.3 | All | 0.61 (0.19-2) | 0.418 |
| | | | Serous | 1.05 (0.79-1.4) | 0.732 |
| | | | Endometrioid | 1.04 (0.65-1.65) | 0.88 |
| | | | Mucinous | 0.67 (0.35-1.31) | 0.246 |
| | | | Clear cell | 1.43 (0.84-2.42) | 0.183 |

SNP order in haplotypes is 5' to 3' of *AIFM2* (block 1): rs2394655, rs7908957, rs1053495, rs2894111, rs2394656, rs6480440, rs2280201. *AIFM2* (block 2): rs10999147, rs3750772, rs4295944, rs2394644, rs10999152.

## Appendix IV-C: Genotype specific ratios of *AKTIP* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| *AKTIP* | **rs9931702** | 0.44 | 1722 | 917 | All | 0.96 (0.80-1.16) | 1.01 (0.80-1.27) | 0.9734 |
| | | | | 506 | Serous | 1.10 (0.87-1.38) | 1.12 (0.85-1.50) | 0.4039 |
| | | | | 137 | Endometrioid | 0.89 (0.59-1.34) | 1.06 (0.65-1.73) | 0.8928 |
| | | | | 86 | Mucinous | 0.83 (0.52-1.34) | 0.58 (0.29-1.14) | 0.1045 |
| | | | | 79 | Clear cell | **0.52 (0.31-0.88)** | 0.72 (0.39-1.34) | 0.1424 |
| *AKTIP* | rs17801966 | 0.15 | 1469 | 828 | All | 1.06 (0.87-1.31) | 0.74 (0.38-1.46) | 0.9282 |
| | | | | 450 | Serous | 1.14 (0.89-1.46) | 0.90 (0.40-2.00) | 0.4718 |
| | | | | 125 | Endometrioid | 0.81 (0.50-1.31) | 0.96 (0.28-3.35) | 0.3441 |
| | | | | 75 | Mucinous | 1.05 (0.61-1.83) | 0.58 (0.08-4.37) | 0.8683 |
| | | | | 78 | Clear cell | 1.13 (0.66-1.91) | 0.58 (0.08-4.40) | 0.9369 |
| ***AKTIP*** | **rs7189819** | 0.3 | 2923 | 1745 | All | 0.93 (0.82-1.06) | 0.92 (0.75-1.14) | 0.2796 |
| | | | | 825 | Serous | 0.99 (0.84-1.17) | 1.02 (0.78-1.33) | 0.867 |
| | | | | 271 | Endometrioid | 0.95 (0.73-1.24) | 1.06 (0.69-1.61) | 0.9177 |
| | | | | 186 | **Mucinous** | 0.87 (0.64-1.19) | **0.42 (0.21-0.84)** | **0.0247** |
| | | | | 163 | **Clear cell** | **0.62 (0.44-0.87)** | 0.78 (0.45-1.35) | **0.0412** |
| *AKTIP* | rs3743772 | 0.07 | 1093 | 413 | All | 0.90 (0.62-1.32) | 2.02 (0.32-12.59) | 0.6778 |
| | | | | 256 | Serous | 1.12 (0.73-1.72) | 1.72 (0.17-17.14) | 0.6424 |
| | | | | 54 | Endometrioid | 0.67 (0.24-1.90) | 11.42 (1.04-125.41) | 0.9435 |
| | | | | 43 | Mucinous | 0.41 (0.10-1.71) | 8.79 (0.86-89.58) | 0.6759 |
| | | | | 26 | Clear cell | 0.34 (0.05-2.55) | **22.33 (1.82-273.37)** | 0.9875 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-D: Haplotype-specific risks of *AKTIP* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *AKTIP* | h0000 | 54.7 | All | 1.01 (0.88-1.16) | 0.9 |
| | | | Serous | 0.92 (0.83-1.03) | 0.153 |
| | | | Endometrioid | 1.04 (0.87-1.25) | 0.634 |
| | | | Mucinous | 1.2 (0.97-1.49) | 0.086 |
| | | | Clear cell | 1.13 (0.9-1.41) | 0.299 |
| ***AKTIP*** | **h1010** | 30.4 | All | 1 (0.86-1.16) | 0.996 |
| | | | Serous | 0.99 (0.88-1.11) | 0.858 |
| | | | Endometrioid | 0.97 (0.8-1.18) | 0.766 |
| | | | **Mucinous** | **0.73 (0.58-0.94)** | **0.013** |
| | | | **Clear cell** | **0.77 (0.59-1)** | **0.047** |
| *AKTIP* | h1100 | 7.9 | All | 0.97 (0.75-1.25) | 0.804 |
| | | | Serous | 0.93 (0.76-1.15) | 0.524 |
| | | | Endometrioid | 0.77 (0.54-1.11) | 0.158 |
| | | | Mucinous | 0.86 (0.57-1.29) | 0.46 |
| | | | Clear cell | 0.97 (0.63-1.49) | 0.891 |
| ***AKTIP*** | **h1101** | 5.7 | All | 1 (0.75-1.33) | 0.999 |
| | | | **Serous** | **1.29 (1.01-1.66)** | **0.044** |
| | | | Endometrioid | 1.02 (0.65-1.59) | 0.927 |
| | | | Mucinous | 0.94 (0.54-1.65) | 0.835 |
| | | | Clear cell | 1.37 (0.82-2.3) | 0.227 |
| ***AKTIP*** | **Rare** | 0.02 | All | 73.28 (0.84-6372.38) | 0.059 |
| | | | **Serous** | **1.8 (1.3-2.49)** | **$3.71x10^{-4}$** |
| | | | Endometrioid | 1.53 (0.93-2.52) | 0.093 |
| | | | **Mucinous** | **2.27 (1.39-3.71)** | **0.001** |
| | | | Clear cell | 1.52 (0.81-2.86) | 0.19 |

SNP order in haplotypes is 5' to 3' of *ATKIP*: rs9931702, rs17801966, rs7189819, rs3743772.

## Appendix IV-E: Genotype specific ratios of *AXIN2* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. | No. | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| *AXIN2* | rs11868547 | 0.48 | 1717 | 919 | All | 0.93 (0.76-1.13) | 1.02 (0.81-1.29) | 0.8178 |
| | | | | 509 | Serous | 0.92 (0.72-1.17) | 0.99 (0.74-1.31) | 0.9949 |
| | | | | 136 | Endometrioid | 0.87 (0.56-1.35) | 1.21 (0.75-1.97) | 0.4357 |
| | | | | 85 | Mucinous | 0.91 (0.54-1.55) | 0.95 (0.51-1.77) | 0.89 |
| | | | | 81 | Clear cell | 0.81 (0.47-1.38) | 1.06 (0.57-1.96) | 0.9215 |
| *AXIN2* | rs7591 | 0.38 | 2881 | 1779 | All | 1.09 (0.96-1.24) | 1.05 (0.87-1.27) | 0.4234 |
| | | | | 838 | Serous | **1.19 (1.01-1.41)** | 1.11 (0.87-1.41) | 0.1463 |
| | | | | 277 | Endometrioid | 0.80 (0.61-1.06) | 1.04 (0.72-1.50) | 0.6889 |
| | | | | 195 | Mucinous | 1.29 (0.93-1.77) | 0.80 (0.48-1.32) | 0.9404 |
| | | | | 165 | Clear cell | 0.84 (0.59-1.18) | 1.05 (0.66-1.66) | 0.8548 |
| *AXIN2* | rs4074947 | 0.22 | 2898 | 1775 | All | 1.16 (1.02-1.32) | 0.92 (0.67-1.24) | 0.2189 |
| | | | | 840 | Serous | **1.21 (1.03-1.43)** | 0.76 (0.50-1.16) | 0.323 |
| | | | | 276 | Endometrioid | 1.10 (0.84-1.44) | 1.20 (0.68-2.11) | 0.4121 |
| | | | | 192 | Mucinous | 1.03 (0.75-1.43) | 0.96 (0.45-2.01) | 0.948 |
| | | | | 163 | Clear cell | 1.02 (0.72-1.44) | 1.20 (0.59-2.44) | 0.6962 |
| *AXIN2* | rs7210356 | 0.11 | 2974 | 1777 | All | 1.01 (0.87-1.17) | 0.97 (0.56-1.69) | 0.8864 |
| | | | | 838 | Serous | 1.09 (0.90-1.32) | 1.08 (0.54-2.14) | 0.3518 |
| | | | | 277 | Endometrioid | 0.88 (0.64-1.22) | 1.09 (0.33-3.61) | 0.5776 |
| | | | | 193 | Mucinous | 1.05 (0.73-1.52) | 1.19 (0.34-4.20) | 0.5978 |
| | | | | 165 | Clear cell | 0.71 (0.46-1.10) | 1.54 (0.46-5.13) | 0.3368 |
| *AXIN2* | rs11655966 | 0.27 | 1779 | 1301 | All | 0.92 (0.79-1.07) | 1.07 (0.80-1.44) | 0.9064 |
| | | | | 552 | Serous | 0.99 (0.81-1.22) | 1.10 (0.74-1.63) | 0.6218 |
| | | | | 213 | Endometrioid | 0.84 (0.61-1.15) | 1.35 (0.79-2.28) | 0.865 |
| | | | | 144 | Mucinous | 0.95 (0.66-1.39) | 0.40 (0.15-1.07) | 0.4605 |
| | | | | 130 | Clear cell | 0.78 (0.53-1.15) | 1.30 (0.68-2.47) | 0.8989 |
| *AXIN2* | rs4541111 | 0.48 | 1770 | 1297 | All | 1.09 (0.91-1.30) | 0.97 (0.78-1.19) | 0.806 |
| | | | | 554 | Serous | 1.10 (0.87-1.40) | 1.03 (0.78-1.36) | 0.755 |

| Gene | tSNP | MAF | No. | No. | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 214 | Endometrioid | 0.93 (0.65-1.31) | 1.04 (0.70-1.55) | 0.894 |
| | | | | 141 | Mucinous | 1.32 (0.85-2.04) | 0.64 (0.36-1.15) | 0.3682 |
| | | | | 128 | Clear cell | 0.92 (0.60-1.39) | 0.77 (0.46-1.28) | 0.3251 |
| | | | | 1185 | All | 1.08 (0.92-1.25) | 1.19 (0.92-1.55) | 0.1238 |
| | | | | 539 | Serous | 1.05 (0.86-1.29) | 1.23 (0.88-1.73) | 0.2499 |
| *AXIN2* | rs4791171 | 0.3 | 2109 | 180 | Endometrioid | 1.04 (0.75-1.45) | 1.30 (0.76-2.20) | 0.3845 |
| | | | | 133 | Mucinous | 1.37 (0.94-2.00) | 1.07 (0.54-2.11) | 0.2561 |
| | | | | 111 | Clear cell | 0.88 (0.58-1.32) | 1.01 (0.50-2.01) | 0.7639 |
| | | | | 839 | **All** | **1.23 (1.00-1.51)** | 1.73 (0.99-3.01) | **0.0383** |
| | | | | 326 | Serous | 1.22 (0.92-1.63) | 1.74 (0.84-3.63) | 0.1127 |
| ***AXIN2*** | **rs11079571** | 0.17 | 1206 | 137 | Endometrioid | 1.05 (0.69-1.59) | 2.32 (0.92-5.87) | 0.3712 |
| | | | | 104 | Mucinous | 1.03 (0.63-1.69) | 1.77 (0.54-5.80) | 0.5048 |
| | | | | 83 | Clear cell | 1.54 (0.95-2.50) | 2.00 (0.58-6.89) | 0.0625 |
| | | | | 1780 | All | 1.05 (0.93-1.20) | 1.26 (0.95-1.68) | 0.1545 |
| | | | | 843 | Serous | 1.05 (0.89-1.24) | 1.10 (0.75-1.60) | 0.4525 |
| *AXIN2* | rs3923087 | 0.22 | 2910 | 275 | Endometrioid | 0.96 (0.73-1.26) | **1.68 (1.02-2.79)** | 0.3043 |
| | | | | 193 | Mucinous | 1.03 (0.75-1.42) | 1.17 (0.59-2.32) | 0.657 |
| | | | | 164 | Clear cell | 0.84 (0.59-1.20) | 1.60 (0.85-3.00) | 0.8136 |
| | | | | 1753 | All | 0.91 (0.80-1.04) | 1.07 (0.90-1.27) | 0.813 |
| | | | | 833 | Serous | **0.84 (0.71-1.00)** | 1.02 (0.81-1.27) | 0.6828 |
| *AXIN2* | rs3923086 | 0.42 | 2935 | 267 | Endometrioid | 0.89 (0.67-1.18) | 1.02 (0.71-1.47) | 0.9432 |
| | | | | 188 | Mucinous | 1.18 (0.83-1.66) | 1.22 (0.78-1.90) | 0.2639 |
| | | | | 163 | Clear cell | 0.81 (0.57-1.16) | 1.00 (0.64-1.57) | 0.793 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-F: Haplotype-specific risks of *AXIN2* on ovarian cancer susceptibility

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *AXIN2* (haplotype block 1) | h100000 | 44.9 | All | 1.03 (0.93-1.13) | 0.599 |
| | | | Serous | 1.01 (0.9-1.12) | 0.907 |
| | | | Endometrioid | 1.09 (0.92-1.3) | 0.316 |
| | | | Mucinous | 0.95 (0.77-1.17) | 0.639 |
| | | | Clear cell | 1.06 (0.85-1.32) | 0.6 |
| ***AXIN2* (haplotype block 1)** | **h000001** | 14.6 | All | 0.9 (0.77-1.04) | 0.148 |
| | | | **Serous** | **0.81 (0.68-0.97)** | **0.018** |
| | | | Endometrioid | 0.95 (0.73-1.24) | 0.7 |
| | | | Mucinous | 1.08 (0.8-1.46) | 0.632 |
| | | | Clear cell | 0.97 (0.7-1.36) | 0.868 |
| *AXIN2* (haplotype block 1) | h010111 | 10.7 | All | 1.05 (0.89-1.23) | 0.58 |
| | | | Serous | 1.07 (0.9-1.28) | 0.421 |
| | | | Endometrioid | 0.89 (0.65-1.2) | 0.431 |
| | | | Mucinous | 1.13 (0.81-1.57) | 0.471 |
| | | | Clear cell | 0.85 (0.57-1.26) | 0.414 |
| ***AXIN2* (haplotype block 1)** | **h011001** | 10.4 | All | 1.15 (0.98-1.36) | 0.082 |
| | | | **Serous** | **1.21 (1.01-1.45)** | **0.041** |
| | | | Endometrioid | 1.16 (0.87-1.54) | 0.312 |
| | | | Mucinous | 1.22 (0.87-1.7) | 0.246 |
| | | | Clear cell | 0.86 (0.57-1.29) | 0.458 |
| *AXIN2* (haplotype block 1) | h011011 | 6 | All | 1.08 (0.87-1.33) | 0.49 |
| | | | Serous | 1.03 (0.81-1.31) | 0.803 |
| | | | Endometrioid | 1.37 (0.97-1.92) | 0.075 |
| | | | Mucinous | 0.66 (0.39-1.14) | 0.137 |
| | | | Clear cell | 1.2 (0.76-1.89) | 0.433 |
| *AXIN2* (haplotype block 1) | h010011 | 5.5 | All | 0.9 (0.72-1.12) | 0.355 |
| | | | Serous | 1.08 (0.85-1.37) | 0.547 |
| | | | Endometrioid | 0.86 (0.57-1.31) | 0.495 |
| | | | Mucinous | 0.86 (0.53-1.41) | 0.553 |
| | | | Clear cell | 1.01 (0.61-1.66) | 0.984 |
| *AXIN2* (haplotype block 1) | h011010 | 3.5 | All | 0.96 (0.73-1.26) | 0.751 |
| | | | Serous | 0.96 (0.71-1.3) | 0.797 |
| | | | Endometrioid | 0.7 (0.4-1.23) | 0.216 |
| | | | Mucinous | 1.03 (0.58-1.82) | 0.916 |
| | | | Clear cell | 1.38 (0.8-2.4) | 0.251 |
| ***AXIN2* (haplotype block 1)** | **Rare** | 0.8 | All | 0.83 (0.49-1.41) | 0.488 |
| | | | Serous | 0.86 (0.67-1.11) | 0.249 |
| | | | **Endometrioid** | **0.58 (0.35-0.96)** | **0.033** |
| | | | Mucinous | 0.97 (0.61-1.53) | 0.887 |

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| | | | Clear cell | 0.85 (0.5-1.44) | 0.546 |
| *AXIN2* (haplotype block 2) | h0000 | 53.4 | All | 0.98 (0.87-1.1) | 0.75 |
| | | | Serous | 1.03 (0.92-1.15) | 0.637 |
| | | | Endometrioid | 1.02 (0.85-1.22) | 0.833 |
| | | | Mucinous | 0.93 (0.75-1.15) | 0.48 |
| | | | Clear cell | 1.04 (0.82-1.3) | 0.767 |
| *AXIN2* (haplotype block 2) | h0001 | 14.5 | All | 0.93 (0.79-1.11) | 0.431 |
| | | | Serous | 0.89 (0.75-1.05) | 0.167 |
| | | | Endometrioid | 0.85 (0.64-1.12) | 0.24 |
| | | | Mucinous | 1.23 (0.92-1.65) | 0.157 |
| | | | Clear cell | 1.06 (0.76-1.47) | 0.744 |
| ***AXIN2* (haplotype block 2)** | **h1111** | 12.8 | **All** | **1.21 (1.03-1.42)** | **0.023** |
| | | | **Serous** | **1.19 (1.01-1.39)** | **0.037** |
| | | | Endometrioid | 1.08 (0.83-1.4) | 0.572 |
| | | | Mucinous | 1.13 (0.83-1.53) | 0.434 |
| | | | Clear cell | 0.96 (0.68-1.36) | 0.838 |
| *AXIN2* (haplotype block 2) | h1011 | 7 | All | 1.06 (0.86-1.3) | 0.596 |
| | | | Serous | 0.95 (0.76-1.19) | 0.656 |
| | | | Endometrioid | 1.28 (0.93-1.77) | 0.134 |
| | | | Mucinous | 0.83 (0.53-1.31) | 0.43 |
| | | | Clear cell | 1.2 (0.79-1.83) | 0.395 |
| *AXIN2* (haplotype block 2) | h1001 | 5.8 | All | 0.98 (0.77-1.26) | 0.873 |
| | | | Serous | 0.92 (0.73-1.17) | 0.508 |
| | | | Endometrioid | 0.95 (0.64-1.4) | 0.78 |
| | | | Mucinous | 0.98 (0.63-1.54) | 0.943 |
| | | | Clear cell | 0.74 (0.43-1.27) | 0.27 |
| *AXIN2* (haplotype block 2) | h1000 | 2.7 | All | 0.84 (0.59-1.21) | 0.355 |
| | | | Serous | 0.91 (0.61-1.35) | 0.638 |
| | | | Endometrioid | 0.7 (0.34-1.44) | 0.327 |
| | | | Mucinous | 0.78 (0.35-1.75) | 0.549 |
| | | | Clear cell | 0.6 (0.22-1.62) | 0.314 |
| *AXIN2* (haplotype block 2) | Rare | 0.1 | All | 1.67 (0.59-4.69) | 0.332 |
| | | | Serous | 0.91 (0.69-1.21) | 0.511 |
| | | | Endometrioid | 0.92 (0.58-1.45) | 0.708 |
| | | | Mucinous | 0.88 (0.51-1.53) | 0.649 |
| | | | Clear cell | 1 (0.57-1.74) | 0.992 |

SNP order in haplotypes is 5' to 3' of *AXIN2* (block 1): rs11868547, rs7591, rs4074947, rs7210356, rs11655966, rs4541111. *AXIN2* (block 2): rs4791171, rs11079571, rs3923087, rs3923086.

## Appendix IV-G: Genotype specific ratios of *CASP5* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 438 | **All** | **1.39 (1.06-1.81)** | **1.44 (1.05-1.97)** | **0.0124** |
| | | | | 270 | **Serous** | 1.36 (0.98-1.88) | 1.45 (0.99-2.11) | **0.0313** |
| ***CASP5*** | **rs518604** | 0.46 | 1195 | 56 | Endometrioid | 1.12 (0.58-2.15) | 1.40 (0.67-2.92) | 0.3451 |
| | | | | 42 | Mucinous | 1.45 (0.65-3.22) | 2.00 (0.84-4.77) | 0.0958 |
| | | | | 32 | Clear cell | 1.63 (0.71-3.73) | 0.72 (0.21-2.42) | 0.8022 |
| | | | | 824 | All | 1.03 (0.83-1.29) | 1.07 (0.82-1.39) | 0.7689 |
| | | | | 320 | Serous | 0.86 (0.65-1.16) | 0.80 (0.55-1.15) | 0.1294 |
| *CASP5* | rs523104 | 0.46 | 1199 | 131 | Endometrioid | 1.24 (0.78-1.97) | 1.45 (0.84-2.48) | 0.0826 |
| | | | | 102 | Mucinous | 1.00 (0.60-1.66) | 0.87 (0.46-1.66) | 0.8518 |
| | | | | 81 | Clear cell | 1.01 (0.60-1.71) | 0.81 (0.40-1.60) | 0.5623 |
| | | | | 829 | All | 0.95 (0.73-1.23) | 1.16 (0.50-2.72) | 0.7779 |
| | | | | 319 | Serous | 0.93 (0.65-1.33) | 0.68 (0.18-2.64) | 0.7345 |
| *CASP5* | rs3181328 | 0.09 | 1206 | 273 | Endometrioid | 0.85 (0.59-1.24) | 0.54 (0.13-2.25) | 0.2291 |
| | | | | 102 | Mucinous | 0.97 (0.52-1.80) | **3.90 (1.06-14.34)** | 0.5649 |
| | | | | 83 | Clear cell | 1.10 (0.60-2.00) | 1.12 (0.14-8.75) | 0.6481 |
| | | | | 803 | All | 0.88 (0.69-1.12) | 1.28 (0.54-3.02) | 0.5052 |
| | | | | 311 | Serous | 0.87 (0.63-1.22) | 1.01 (0.28-3.69) | 0.5292 |
| *CASP5* | rs17446518 | 0.11 | 1177 | 130 | Endometrioid | 0.81 (0.50-1.33) | 0.79 (0.10-6.24) | 0.3787 |
| | | | | 97 | Mucinous | 1.15 (0.66-1.99) | **4.12 (1.10-15.36)** | 0.2146 |
| | | | | 81 | Clear cell | 0.70 (0.37-1.32) | 1.20 (0.15-9.47) | 0.3467 |
| *CASP5* | rs9651713 | 0.11 | 2898 | 1730 | All | 0.99 (0.85-1.15) | 1.22 (0.70-2.13) | 0.836 |
| | | | | 819 | Serous | 0.96 (0.78-1.17) | 1.55 (0.81-2.96) | 0.8167 |
| | | | | 269 | Endometrioid | 1.16 (0.86-1.58) | 1.80 (0.68-4.77) | 0.2196 |
| | | | | 183 | Mucinous | 0.86 (0.57-1.28) | 1.10 (0.26-4.69) | 0.4883 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 162 | Clear cell | 1.09 (0.74-1.61) | 1.13 (0.27-4.83) | 0.6561 |
| *CASP5* | rs3181175 | 0.19 | 2379 | 1282 | All | 0.96 (0.83-1.12) | 1.23 (0.83-1.82) | 0.9331 |
| | | | | 597 | Serous | 0.93 (0.76-1.13) | 1.34 (0.82-2.19) | 0.9482 |
| | | | | 194 | Endometrioid | 1.01 (0.73-1.39) | 1.66 (0.80-3.44) | 0.4992 |
| | | | | 147 | Mucinous | 1.05 (0.73-1.51) | 0.72 (0.22-2.37) | 0.8575 |
| | | | | 115 | Clear cell | 0.98 (0.65-1.48) | 1.19 (0.42-3.37) | 0.9266 |
| *CASP5* | rs3181174 | 0.07 | 2962 | 1780 | All | 0.96 (0.80-1.14) | 1.21 (0.55-2.65) | 0.7967 |
| | | | | 840 | Serous | 1.09 (0.88-1.36) | 1.12 (0.41-3.09) | 0.4061 |
| | | | | 278 | Endometrioid | 0.82 (0.56-1.21) | 0.83 (0.11-6.43) | 0.3188 |
| | | | | 195 | Mucinous | 0.87 (0.56-1.37) | 1.11 (0.14-8.64) | 0.6099 |
| | | | | 165 | Clear cell | 0.64 (0.37-1.10) | 1.20 (0.16-9.26) | 0.151 |
| *CASP5* | rs2282657 | 0.35 | 1478 | 852 | All | 1.08 (0.90-1.30) | 0.89 (0.67-1.20) | 0.7645 |
| | | | | 462 | Serous | 1.16 (0.92-1.45) | 1.11 (0.78-1.56) | 0.3615 |
| | | | | 128 | Endometrioid | 1.12 (0.76-1.66) | 0.87 (0.45-1.69) | 0.7802 |
| | | | | 80 | Mucinous | 1.02 (0.64-1.64) | 0.63 (0.26-1.52) | 0.4369 |
| | | | | 73 | Clear cell | 1.06 (0.65-1.73) | 0.35 (0.11-1.18) | 0.1926 |
| *CASP5* | rs507879 | 0.46 | 2839 | 1768 | All | 1.02 (0.87-1.20) | 1.00 (0.84-1.19) | 0.9144 |
| | | | | 835 | Serous | 1.09 (0.89-1.34) | 1.00 (0.79-1.26) | 0.8497 |
| | | | | 276 | Endometrioid | 0.99 (0.72-1.36) | 0.83 (0.58-1.20) | 0.2713 |
| | | | | 194 | Mucinous | 0.84 (0.57-1.23) | 1.01 (0.67-1.53) | 0.7951 |
| | | | | 164 | Clear cell | 1.05 (0.70-1.60) | 0.97 (0.61-1.54) | 0.8762 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-H: Haplotype-specific risks of CASP5 on ovarian cancer susceptibility

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| ***CASP5*** **(haplotype block 1)** | **h010** | 43.8 | All | 0.99 (0.9-1.09) | 0.79 |
| | | | **Serous** | **0.89 (0.79-0.99)** | **0.033** |
| | | | Endometrioid | 1.07 (0.89-1.28) | 0.475 |
| | | | Mucinous | 0.82 (0.66-1.01) | 0.068 |
| | | | Clear cell | 1.02 (0.81-1.29) | 0.836 |
| ***CASP5*** **(haplotype block 1)** | **h100** | 43.3 | **All** | **1.13 (1.03-1.24)** | **0.012** |
| | | | **Serous** | **1.2 (1.08-1.35)** | **0.001** |
| | | | Endometrioid | 1.1 (0.92-1.32) | 0.289 |
| | | | Mucinous | 1.18 (0.95-1.46) | 0.126 |
| | | | Clear cell | 1 (0.79-1.26) | 0.983 |
| *CASP5* (haplotype block 1) | h001 | 9.8 | All | 0.9 (0.77-1.06) | 0.215 |
| | | | Serous | 0.92 (0.76-1.12) | 0.421 |
| | | | Endometrioid | 0.74 (0.53-1.06) | 0.098 |
| | | | Mucinous | 1.13 (0.8-1.6) | 0.476 |
| | | | Clear cell | 0.94 (0.63-1.41) | 0.764 |
| ***CASP5*** **(haplotype block 1)** | **h000** | 2.3 | **All** | **0.72 (0.56-0.94)** | **0.015** |
| | | | Serous | 0.91 (0.59-1.4) | 0.653 |
| | | | Endometrioid | 0.77 (0.37-1.59) | 0.477 |
| | | | Mucinous | 0.84 (0.36-1.99) | 0.697 |
| | | | Clear cell | 1.32 (0.65-2.71) | 0.443 |
| *CASP5* (haplotype block 1) | h110 | 0.8 | All | 0.74 (0.54-1.02) | 0.069 |
| | | | Serous | 0.65 (0.41-1.02) | 0.061 |
| | | | Endometrioid | 0.37 (0.14-1.01) | 0.052 |
| | | | Mucinous | 1.16 (0.6-2.24) | 0.653 |
| | | | Clear cell | 0.46 (0.14-1.48) | 0.193 |
| *CASP5* (haplotype block 2) | h000001 | 48.7 | All | 1.02 (0.9-1.14) | 0.768 |
| | | | Serous | 1.01 (0.9-1.13) | 0.904 |
| | | | Endometrioid | 0.96 (0.8-1.15) | 0.628 |
| | | | Mucinous | 1.01 (0.81-1.25) | 0.952 |
| | | | Clear cell | 1.03 (0.82-1.3) | 0.791 |
| *CASP5* (haplotype block 2) | h000010 | 12.8 | All | 1.1 (0.93-1.3) | 0.285 |
| | | | Serous | 1.07 (0.91-1.26) | 0.43 |
| | | | Endometrioid | 1.28 (0.99-1.64) | 0.059 |
| | | | Mucinous | 1.05 (0.77-1.44) | 0.752 |
| | | | Clear cell | 1.17 (0.84-1.63) | 0.348 |
| *CASP5* (haplotype block 2) | h000000 | 10.1 | All | 0.87 (0.71-1.07) | 0.179 |
| | | | Serous | 0.9 (0.74-1.1) | 0.319 |
| | | | Endometrioid | 1.01 (0.75-1.38) | 0.931 |
| | | | Mucinous | 0.98 (0.68-1.42) | 0.931 |

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| | | | Clear cell | 0.98 (0.66-1.46) | 0.923 |
| *CASP5* (haplotype block 2) | h011010 | 9.6 | All | 1.15 (0.95-1.39) | 0.162 |
| | | | Serous | 1.02 (0.85-1.23) | 0.833 |
| | | | Endometrioid | 1.28 (0.97-1.69) | 0.076 |
| | | | Mucinous | 0.9 (0.63-1.31) | 0.593 |
| | | | Clear cell | 1.17 (0.82-1.68) | 0.393 |
| *CASP5* (haplotype block 2) | h001110 | 6.5 | All | 0.97 (0.77-1.22) | 0.787 |
| | | | Serous | 1.1 (0.88-1.37) | 0.396 |
| | | | Endometrioid | 0.82 (0.56-1.22) | 0.335 |
| | | | Mucinous | 0.79 (0.49-1.27) | 0.331 |
| | | | Clear cell | 0.76 (0.45-1.28) | 0.305 |
| *CASP5* (haplotype block 2) | h100001 | 5.2 | All | 1.01 (0.82-1.26) | 0.902 |
| | | | Serous | 0.97 (0.75-1.26) | 0.821 |
| | | | Endometrioid | 0.77 (0.48-1.22) | 0.27 |
| | | | Mucinous | 1.12 (0.69-1.8) | 0.651 |
| | | | Clear cell | 0.76 (0.42-1.39) | 0.38 |
| *CASP5* (haplotype block 2) | h100010 | 2.6 | All | 1.07 (0.79-1.45) | 0.663 |
| | | | Serous | 0.88 (0.57-1.34) | 0.548 |
| | | | Endometrioid | 0.79 (0.39-1.61) | 0.517 |
| | | | Mucinous | 1.39 (0.7-2.76) | 0.343 |
| | | | Clear cell | 0.72 (0.28-1.82) | 0.485 |
| ***CASP5* (haplotype block 2)** | **Rare** | 0.9 | **All** | **0.28 (0.12-0.69)** | **0.005** |
| | | | Serous | 0.84 (0.6-1.17) | 0.293 |
| | | | **Endometrioid** | **0.45 (0.22-0.89)** | **0.022** |
| | | | Mucinous | 1.1 (0.62-1.95) | 0.742 |
| | | | Clear cell | 0.75 (0.36-1.55) | 0.433 |

SNP order in haplotypes is 5' to 3' of *CASP5* (block 1): rs518604, rs523104, rs3181328. *CASP5* (block 2): rs17446518, rs9651713, rs3181175, rs3181174, rs2282657, rs507879.

## Appendix IV-I: Genotype specific ratios of *FILIP1L* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 437 | All | 1.07 (0.84-1.35) | 1.33 (0.94-1.89) | 0.1458 |
| | | | | 269 | Serous | 0.98 (0.74-1.31) | 1.25 (0.83-1.90) | 0.4593 |
| *FILIP1L* | rs796977 | 0.33 | 1166 | 56 | Endometrioid | 1.67 (0.93-3.02) | 1.51 (0.62-3.68) | 0.1545 |
| | | | | 43 | Mucinous | 1.20 (0.62-2.34) | 1.77 (0.72-4.38) | 0.2813 |
| | | | | 31 | Clear cell | 0.93 (0.42-2.03) | 1.38 (0.49-3.91) | 0.6504 |
| | | | | 1653 | All | 1.02 (0.88-1.18) | 0.88 (0.53-1.43) | 0.9373 |
| | | | | 771 | Serous | 1.11 (0.92-1.35) | 1.08 (0.59-1.99) | 0.3224 |
| *FILIP1L* | rs793477 | 0.13 | 2646 | 257 | Endometrioid | 0.92 (0.67-1.27) | 0.60 (0.18-1.97) | 0.4293 |
| | | | | 175 | Mucinous | 1.16 (0.80-1.67) | 0.66 (0.16-2.76) | 0.8392 |
| | | | | 155 | Clear cell | 1.04 (0.70-1.54) | 1.79 (0.69-4.61) | 0.4497 |
| | | | | 1773 | All | 1.05 (0.92-1.20) | 1.11 (0.93-1.32) | 0.3207 |
| | | | | 838 | Serous | 0.95 (0.80-1.13) | 1.07 (0.86-1.34) | 0.842 |
| ***FILIP1L*** | **rs793446** | 0.41 | 2947 | 274 | **Endometrioid** | **1.36 (1.02-1.81)** | **1.52 (1.05-2.20)** | **0.0262** |
| | | | | 194 | Mucinous | 1.14 (0.82-1.58) | 0.93 (0.58-1.48) | 0.8885 |
| | | | | 164 | Clear cell | 1.05 (0.74-1.50) | 1.11 (0.69-1.77) | 0.6725 |
| | | | | 1773 | All | 1.00 (0.84-1.20) | 0.69 (0.29-1.61) | 0.6908 |
| | | | | 840 | Serous | 0.90 (0.71-1.15) | 0.19 (0.03-1.43) | 0.1194 |
| *FILIP1L* | rs3921767 | 0.07 | 2859 | 276 | Endometrioid | 1.09 (0.75-1.57) | 1.16 (0.26-5.15) | 0.6936 |
| | | | | 191 | Mucinous | 1.12 (0.72-1.73) | 0.90 (0.12-6.92) | 0.7966 |
| | | | | 166 | Clear cell | 1.00 (0.62-1.61) | 1.94 (0.44-8.55) | 0.7117 |
| ***FILIP1L*** | **rs17338680** | 0.11 | 2989 | 1786 | All | 1.06 (0.91-1.24) | 0.77 (0.46-1.28) | 0.985 |
| | | | | 574 | Serous | 0.88 (0.69-1.14) | 0.98 (0.44-2.20) | 0.3051 |
| | | | | 221 | **Endometrioid** | **1.71 (1.24-2.36)** | 0.79 (0.19-3.39) | **0.0073** |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 196 | Mucinous | 1.23 (0.86-1.76) | 0.35 (0.05-2.58) | 0.7109 |
| | | | | 133 | Clear cell | 1.04 (0.67-1.63) | 1.12 (0.26-4.81) | 0.8406 |
| | | | | 1786 | All | 0.98 (0.86-1.12) | 1.18 (0.91-1.53) | 0.6077 |
| | | | | 843 | Serous | 0.96 (0.81-1.14) | **1.47 (1.08-2.01)** | 0.2249 |
| *FILIP1L* | rs9864437 | 0.22 | 2972 | 278 | Endometrioid | 1.05 (0.80-1.37) | 0.86 (0.47-1.59) | 0.8102 |
| | | | | 195 | Mucinous | 0.99 (0.72-1.36) | 0.91 (0.45-1.85) | 0.782 |
| | | | | 165 | Clear cell | 0.81 (0.57-1.15) | 0.98 (0.49-1.98) | 0.3807 |
| | | | | 1414 | All | 0.98 (0.84-1.13) | 0.96 (0.79-1.17) | 0.7028 |
| | | | | 710 | Serous | 1.07 (0.88-1.29) | 1.12 (0.88-1.44) | 0.3295 |
| *FILIP1L* | rs6788750 | 0.41 | 2532 | 226 | Endometrioid | 0.84 (0.62-1.14) | 0.72 (0.47-1.11) | 0.1277 |
| | | | | 140 | Mucinous | 1.01 (0.69-1.48) | 0.86 (0.50-1.46) | 0.6811 |
| | | | | 131 | Clear cell | 0.91 (0.61-1.35) | 1.15 (0.70-1.89) | 0.7396 |
| | | | | 1273 | All | 1.14 (0.98-1.33) | 0.88 (0.59-1.33) | 0.433 |
| | | | | 594 | Serous | 0.95 (0.78-1.17) | **0.52 (0.27-0.99)** | 0.097 |
| ***FILIP1L*** | **rs12494994** | 0.18 | 2347 | 193 | **Endometrioid** | **1.48 (1.08-2.04)** | **2.16 (1.13-4.12)** | **0.0024** |
| | | | | 145 | Mucinous | **1.57 (1.10-2.25)** | 0.25 (0.03-1.83) | 0.2574 |
| | | | | 113 | Clear cell | 1.20 (0.79-1.82) | 1.69 (0.71-4.03) | 0.1986 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-J: Haplotype-specific risks of *FILIP1L* on ovarian cancer susceptibility

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *FILIP1L* (haplotype block 1) | h00000 | 46.1 | All | 0.95 (0.84-1.08) | 0.445 |
| | | | Serous | 0.97 (0.87-1.08) | 0.572 |
| | | | Endometrioid | 0.86 (0.72-1.03) | 0.092 |
| | | | Mucinous | 0.98 (0.79-1.2) | 0.823 |
| | | | Clear cell | 0.9 (0.71-1.13) | 0.351 |
| *FILIP1L* (haplotype block 1) | h10100 | 22.5 | All | 1.08 (0.93-1.24) | 0.311 |
| | | | Serous | 1.12 (0.98-1.27) | 0.087 |
| | | | Endometrioid | 1 (0.81-1.24) | 0.972 |
| | | | Mucinous | 1 (0.78-1.29) | 1 |
| | | | Clear cell | 0.88 (0.67-1.17) | 0.39 |
| *FILIP1L* (haplotype block 1) | h01000 | 11.5 | All | 0.97 (0.79-1.19) | 0.782 |
| | | | Serous | 1.08 (0.91-1.27) | 0.383 |
| | | | Endometrioid | 0.93 (0.71-1.23) | 0.62 |
| | | | Mucinous | 1.1 (0.81-1.49) | 0.527 |
| | | | Clear cell | 1.17 (0.85-1.6) | 0.347 |
| ***FILIP1L* (haplotype block 1)** | **h10101** | 10.7 | All | 1.11 (0.91-1.35) | 0.29 |
| | | | Serous | 0.94 (0.78-1.13) | 0.514 |
| | | | **Endometrioid** | **1.56 (1.22-2.01)** | **$5.01 \times 10^{-4}$** |
| | | | Mucinous | 1.09 (0.78-1.51) | 0.617 |
| | | | Clear cell | 1.25 (0.89-1.75) | 0.204 |
| *FILIP1L* (haplotype block 1) | h00110 | 7.4 | All | 0.94 (0.74-1.2) | 0.617 |
| | | | Serous | 0.83 (0.66-1.04) | 0.111 |
| | | | Endometrioid | 1.05 (0.75-1.47) | 0.791 |
| | | | Mucinous | 1.05 (0.7-1.56) | 0.819 |
| | | | Clear cell | 1.06 (0.69-1.63) | 0.779 |
| *FILIP1L* (haplotype block 1) | Rare | 0.4 | All | 0.34 (0.11-1.03) | 0.056 |
| | | | Serous | 0.81 (0.53-1.25) | 0.345 |
| | | | Endometrioid | 0.61 (0.27-1.36) | 0.226 |
| | | | Mucinous | 0.15 (0.02-1.01) | 0.051 |
| | | | Clear cell | 1.11 (0.53-2.33) | 0.784 |
| *FILIP1L* (haplotype block 2) | h010 | 41 | All | 0.98 (0.88-1.08) | 0.644 |
| | | | Serous | 1.08 (0.96-1.2) | 0.197 |
| | | | Endometrioid | 0.83 (0.69-1) | 0.053 |
| | | | Mucinous | 0.96 (0.77-1.18) | 0.672 |
| | | | Clear cell | 1.03 (0.82-1.29) | 0.823 |
| *FILIP1L* (haplotype block 2) | h100 | 22.6 | All | 1.04 (0.93-1.16) | 0.528 |
| | | | Serous | 1.09 (0.96-1.24) | 0.167 |
| | | | Endometrioid | 0.98 (0.79-1.22) | 0.877 |
| | | | Mucinous | 0.99 (0.77-1.27) | 0.952 |

| Gene/ haplotype block | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| | | | Clear cell | 0.9 (0.69-1.19) | 0.473 |
| *FILIP1L* (haplotype block 2) | h000 | 18.9 | All | 0.94 (0.83-1.08) | 0.383 |
| | | | Serous | 0.9 (0.78-1.04) | 0.16 |
| | | | Endometrioid | 0.98 (0.78-1.23) | 0.873 |
| | | | Mucinous | 1.07 (0.82-1.38) | 0.615 |
| | | | Clear cell | 0.93 (0.7-1.25) | 0.649 |
| ***FILIP1L* (haplotype block 2)** | **h001** | 17.1 | All | 1.04 (0.92-1.17) | 0.562 |
| | | | Serous | 0.88 (0.76-1.02) | 0.092 |
| | | | **Endometrioid** | **1.37 (1.1-1.69)** | **0.004** |
| | | | Mucinous | 1.03 (0.79-1.36) | 0.815 |
| | | | Clear cell | 1.17 (0.88-1.56) | 0.269 |

SNP order in haplotypes is 5' to 3' of *FILIP1L* (block 1): rs796977, rs793477, rs793446, rs3921767, rs17338680. *FILIP1L* (block 2): rs9864437, rs6788750, rs12494994.

## Appendix IV-K: Genotype specific ratios of *RBBP8* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 1272 | All | 0.80 (0.67-0.96) | 1.05 (0.57-1.95) | 0.0645 |
| | | | | 594 | Serous | 0.99 (0.79-1.23) | 1.18 (0.55-2.53) | 0.7647 |
| *RBBP8* | rs7239066 | 0.11 | 2366 | 188 | Endometrioid | 0.70 (0.47-1.06) | 0.74 (0.17-3.18) | 0.1083 |
| | | | | 144 | Mucinous | 0.60 (0.37-0.97) | 0.88 (0.20-3.92) | 0.1098 |
| | | | | 115 | Clear cell | 0.64 (0.38-1.09) | 1.31 (0.30-5.64) | 0.2167 |
| | | | | 1748 | All | 1.12 (0.89-1.41) | 1.13 (0.31-4.07) | 0.2974 |
| | | | | 826 | Serous | 1.17 (0.88-1.57) | 0.55 (0.07-4.59) | 0.3355 |
| *RBBP8* | rs11082221 | 0.04 | 2937 | 271 | Endometrioid | 1.27 (0.80-2.02) | 1.56 (0.18-13.31) | 0.2858 |
| | | | | 188 | Mucinous | 1.38 (0.81-2.36) | 1.92 (0.22-16.89) | 0.1282 |
| | | | | 165 | Clear cell | 1.10 (0.60-2.02) | 3.56 (0.40-31.29) | 0.4888 |
| | | | | 1764 | All | 0.94 (0.82-1.07) | 0.88 (0.72-1.06) | 0.2066 |
| | | | | 829 | **Serous** | **0.83 (0.70-0.98)** | 0.80 (0.63-1.03) | **0.0323** |
| ***RBBP8*** | **rs4474794** | 0.36 | 2895 | 271 | Endometrioid | 1.16 (0.88-1.52) | 1.03 (0.69-1.53) | 0.5615 |
| | | | | 193 | Mucinous | 1.11 (0.81-1.53) | 0.97 (0.60-1.57) | 0.8372 |
| | | | | 165 | Clear cell | 1.03 (0.73-1.46) | 1.27 (0.80-2.01) | 0.3447 |
| | | | | 346 | All | 1.07 (0.82-1.40) | 0.67 (0.38-1.17) | 0.5163 |
| | | | | 215 | Serous | 1.08 (0.79-1.49) | **0.37 (0.16-0.88)** | 0.2176 |
| ***RBBP8*** | **rs9304261** | 0.22 | 888 | 44 | Endometrioid | 1.34 (0.71-2.52) | 0.67 (0.15-2.91) | 0.8077 |
| | | | | 33 | Mucinous | 1.26 (0.59-2.69) | 1.90 (0.62-5.83) | 0.2486 |
| | | | | 21 | Clear cell | 0.58 (0.19-1.83) | **3.22 (1.09-9.46)** | 0.1763 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-L: Haplotype-specific risks of *RBBP8* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| ***RBBP8*** | **h0000** | 62.3 | All | 1.12 (0.98-1.27) | 0.092 |
| | | | **Serous** | **1.13 (1.01-1.27)** | **0.032** |
| | | | Endometrioid | 0.99 (0.83-1.19) | 0.932 |
| | | | Mucinous | 0.98 (0.8-1.21) | 0.865 |
| | | | Clear cell | 0.91 (0.72-1.13) | 0.389 |
| *RBBP8* | h0011 | 23 | All | 0.92 (0.8-1.07) | 0.273 |
| | | | **Serous** | **0.88 (0.77-1)** | 0.051 |
| | | | Endometrioid | 0.99 (0.81-1.22) | 0.958 |
| | | | Mucinous | 1.05 (0.83-1.34) | 0.667 |
| | | | Clear cell | 1.11 (0.86-1.43) | 0.415 |
| *RBBP8* | h1010 | 7.2 | All | 0.83 (0.64-1.07) | 0.147 |
| | | | Serous | 0.93 (0.75-1.16) | 0.542 |
| | | | Endometrioid | 1.02 (0.72-1.45) | 0.891 |
| | | | Mucinous | 0.76 (0.48-1.2) | 0.235 |
| | | | Clear cell | 1.07 (0.7-1.65) | 0.742 |
| *RBBP8* | h1110 | 4.4 | All | 1.16 (0.85-1.58) | 0.338 |
| | | | Serous | 1.2 (0.9-1.59) | 0.211 |
| | | | Endometrioid | 1.11 (0.69-1.78) | 0.66 |
| | | | Mucinous | 1.32 (0.79-2.22) | 0.284 |
| | | | Clear cell | 1.1 (0.61-2.01) | 0.745 |
| *RBBP8* | h0010 | 2 | All | 0.88 (0.71-1.1) | 0.266 |
| | | | Serous | 0.77 (0.53-1.11) | 0.156 |
| | | | Endometrioid | 1.18 (0.73-1.91) | 0.502 |
| | | | Mucinous | 1.15 (0.65-2.03) | 0.624 |
| | | | Clear cell | 1.05 (0.56-2) | 0.873 |
| *RBBP8* | Rare | 0.1 | All | 0.74 (0.25-2.15) | 0.579 |
| | | | Serous | 0.85 (0.52-1.39) | 0.517 |
| | | | Endometrioid | 0.52 (0.19-1.38) | 0.189 |
| | | | Mucinous | 0.74 (0.28-1.92) | 0.532 |
| | | | Clear cell | 0.64 (0.21-1.99) | 0.441 |

SNP order in haplotypes is 5' to 3' of *RBBP8*: rs7239066, rs11082221, rs4474794, rs9304261.

## Appendix IV-M: Genotype specific ratios of *RGC32* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| *RGC32* | rs10467472 | 0.13 | 2887 | 1769 | All | 0.99 (0.86-1.15) | 1.05 (0.68-1.64) | 0.9822 |
| | | | | 839 | Serous | 0.91 (0.75-1.10) | 0.74 (0.39-1.41) | 0.2126 |
| | | | | 273 | Endometrioid | 1.04 (0.77-1.40) | 0.62 (0.19-2.02) | 0.8091 |
| | | | | 191 | Mucinous | 0.88 (0.61-1.27) | 1.69 (0.70-4.06) | 0.9293 |
| | | | | 164 | Clear cell | 0.98 (0.67-1.44) | 1.97 (0.82-4.71) | 0.4683 |
| ***RGC32*** | **rs3783194** | 0.11 | 2723 | 1690 | All | 1.02 (0.88-1.19) | 1.03 (0.57-1.89) | 0.8873 |
| | | | | 788 | Serous | 0.89 (0.72-1.09) | 0.70 (0.29-1.69) | 0.1363 |
| | | | | 264 | Endometrioid | 1.13 (0.83-1.54) | 1.50 (0.51-4.36) | 0.4304 |
| | | | | 184 | Mucinous | 1.09 (0.75-1.59) | 1.08 (0.25-4.66) | 0.7964 |
| | | | | 155 | **Clear cell** | **1.50 (1.04-2.17)** | 1.99 (0.59-6.70) | **0.0206** |
| *RGC32* | rs11618371 | 0.11 | 2959 | 1771 | All | 1.01 (0.87-1.18) | 1.31 (0.74-2.32) | 0.5158 |
| | | | | 835 | Serous | 1.03 (0.85-1.26) | 1.33 (0.66-2.70) | 0.4921 |
| | | | | 275 | Endometrioid | 0.95 (0.69-1.32) | 1.48 (0.51-4.31) | 0.9104 |
| | | | | 193 | Mucinous | 1.03 (0.71-1.51) | 0.55 (0.07-4.09) | 0.871 |
| | | | | 164 | Clear cell | 1.12 (0.76-1.66) | 1.31 (0.31-5.62) | 0.4759 |
| *RGC32* | rs9532824 | 0.07 | 2892 | 1782 | All | 0.96 (0.81-1.15) | 0.47 (0.19-1.16) | 0.3412 |
| | | | | 841 | Serous | 0.94 (0.74-1.18) | 0.46 (0.14-1.54) | 0.3245 |
| | | | | 276 | Endometrioid | 1.22 (0.86-1.72) | 0.54 (0.07-4.07) | 0.4241 |
| | | | | 196 | Mucinous | 1.23 (0.82-1.83) | 0.93 (0.12-7.06) | 0.288 |
| | | | | 164 | Clear cell | 0.94 (0.58-1.50) | 0.82 (0.11-6.13) | 0.7749 |
| *RGC32* | rs995845 | 0.2 | 2365 | 1274 | All | 1.01 (0.87-1.16) | 1.09 (0.82-1.45) | 0.8121 |
| | | | | 595 | Serous | 0.89 (0.74-1.09) | 1.06 (0.73-1.52) | 0.4695 |
| | | | | 193 | Endometrioid | 0.99 (0.73-1.36) | 1.04 (0.56-1.90) | 0.9679 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 146 | Mucinous | 0.98 (0.69-1.40) | 0.93 (0.45-1.92) | 0.8142 |
| | | | | 112 | Clear cell | 1.41 (0.95-2.10) | 1.45 (0.70-3.03) | 0.1034 |
| | | | | 1766 | All | 1.01 (0.88-1.16) | 1.10 (0.74-1.65) | 0.611 |
| | | | | 833 | Serous | 1.09 (0.91-1.29) | 0.97 (0.56-1.65) | 0.4095 |
| *RGC32* | rs9594551 | 0.15 | 2863 | 275 | Endometrioid | 1.02 (0.77-1.36) | 0.78 (0.31-1.99) | 0.9957 |
| | | | | 193 | Mucinous | 0.95 (0.67-1.34) | 0.46 (0.11-1.90) | 0.5421 |
| | | | | 163 | Clear cell | 1.03 (0.72-1.49) | 0.82 (0.25-2.66) | 0.9809 |
| | | | | 1749 | All | 1.02 (0.90-1.16) | 0.89 (0.68-1.17) | 0.8348 |
| | | | | 828 | Serous | 1.03 (0.87-1.21) | 0.84 (0.59-1.21) | 0.7862 |
| *RGC32* | rs975590 | 0.23 | 2940 | 268 | Endometrioid | 1.08 (0.83-1.41) | 1.03 (0.60-1.78) | 0.5433 |
| | | | | 188 | Mucinous | 1.11 (0.81-1.51) | 0.65 (0.30-1.44) | 0.9566 |
| | | | | 161 | Clear cell | 1.08 (0.78-1.51) | 0.41 (0.15-1.12) | 0.465 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-N: Haplotype-specific risks of *RGC32* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| *RGC32* | h0000000 | 41.5 | All | 0.98 (0.89-1.08) | 0.674 |
| | | | Serous | 1.11 (0.99-1.25) | 0.063 |
| | | | Endometrioid | 0.93 (0.78-1.13) | 0.478 |
| | | | Mucinous | 1.07 (0.86-1.33) | 0.549 |
| | | | Clear cell | 0.89 (0.7-1.12) | 0.32 |
| ***RGC32*** | **h0100100** | 10.8 | All | 1.01 (0.87-1.18) | 0.851 |
| | | | Serous | 0.89 (0.74-1.07) | 0.212 |
| | | | Endometrioid | 1.02 (0.76-1.36) | 0.89 |
| | | | Mucinous | 1.01 (0.72-1.41) | 0.962 |
| | | | **Clear cell** | **1.53 (1.11-2.11)** | **0.01** |
| *RGC32* | h0000100 | 10.4 | All | 1.04 (0.89-1.21) | 0.608 |
| | | | Serous | 1.09 (0.91-1.3) | 0.352 |
| | | | Endometrioid | 0.97 (0.72-1.31) | 0.826 |
| | | | Mucinous | 0.8 (0.55-1.17) | 0.25 |
| | | | Clear cell | 0.78 (0.52-1.18) | 0.239 |
| *RGC32* | h1000000 | 8.1 | All | 1 (0.84-1.2) | 0.957 |
| | | | Serous | 0.92 (0.74-1.15) | 0.466 |
| | | | Endometrioid | 1.12 (0.8-1.56) | 0.511 |
| | | | Mucinous | 1.11 (0.75-1.64) | 0.603 |
| | | | Clear cell | 1.11 (0.73-1.69) | 0.635 |
| *RGC32* | h0010011 | 7.8 | All | 1.1 (0.93-1.31) | 0.278 |
| | | | Serous | 1.15 (0.95-1.4) | 0.146 |
| | | | Endometrioid | 0.98 (0.7-1.36) | 0.892 |
| | | | Mucinous | 0.87 (0.58-1.32) | 0.521 |
| | | | Clear cell | 1.15 (0.78-1.71) | 0.478 |
| *RGC32* | h0001001 | 6.7 | All | 0.92 (0.75-1.12) | 0.405 |
| | | | Serous | 0.92 (0.73-1.16) | 0.501 |
| | | | Endometrioid | 1.15 (0.81-1.63) | 0.433 |
| | | | Mucinous | 1.16 (0.77-1.75) | 0.484 |
| | | | Clear cell | 0.94 (0.58-1.53) | 0.816 |
| *RGC32* | h0000011 | 4.9 | All | 0.93 (0.74-1.17) | 0.54 |
| | | | Serous | 0.97 (0.75-1.26) | 0.829 |
| | | | Endometrioid | 0.96 (0.63-1.47) | 0.849 |
| | | | Mucinous | 0.66 (0.37-1.19) | 0.169 |
| | | | Clear cell | 0.85 (0.49-1.51) | 0.587 |
| *RGC32* | h1000100 | 4.1 | All | 0.99 (0.79-1.26) | 0.958 |
| | | | Serous | 0.9 (0.67-1.22) | 0.502 |
| | | | Endometrioid | 0.83 (0.51-1.36) | 0.458 |
| | | | Mucinous | 1.03 (0.61-1.75) | 0.905 |

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| | | | Clear cell | 1.24 (0.73-2.1) | 0.429 |
| ***RGC32*** | **Rare** | 0.8 | All | 1.15 (0.66-2.02) | 0.621 |
| | | | **Serous** | **0.61 (0.42-0.89)** | **0.009** |
| | | | Endometrioid | 0.8 (0.46-1.39) | 0.428 |
| | | | Mucinous | 0.7 (0.34-1.41) | 0.317 |
| | | | Clear cell | 0.74 (0.35-1.55) | 0.422 |

SNP order in haplotypes is 5' to 3' of *RGC32*: rs10467472, rs3783194, rs11618371, rs9532824, rs995845, rs9594551, rs975590.

## Appendix IV-O: Genotype specific ratios of *RUVBL1* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 1777 | All | 1.05 (0.90-1.22) | 1.44 (0.89-2.31) | 0.2094 |
| | | | | 839 | Serous | 1.16 (0.96-1.40) | 1.03 (0.52-2.02) | 0.1425 |
| *RUVBL1* | rs9860614 | 0.12 | 2966 | 276 | Endometrioid | 0.90 (0.65-1.24) | 1.76 (0.72-4.27) | 0.9619 |
| | | | | 193 | Mucinous | 0.85 (0.58-1.25) | 1.73 (0.59-5.11) | 0.8531 |
| | | | | 162 | Clear cell | 0.85 (0.56-1.29) | 1.97 (0.69-5.67) | 0.983 |
| | | | | 1266 | **All** | 1.14 (0.97-1.34) | **1.39 (1.02-1.89)** | **0.0192** |
| | | | | 537 | **Serous** | **1.42 (1.15-1.74)** | **1.63 (1.10-2.42)** | **0.0002** |
| ***RUVBL1*** | **rs13063604** | 0.25 | 1724 | 207 | Endometrioid | 0.74 (0.53-1.04) | 1.29 (0.73-2.31) | 0.3904 |
| | | | | 143 | Mucinous | 0.95 (0.64-1.41) | 1.39 (0.69-2.83) | 0.5473 |
| | | | | 124 | Clear cell | 1.22 (0.83-1.80) | 1.07 (0.48-2.40) | 0.4113 |
| | | | | 1280 | All | 1.20 (1.03-1.40) | 1.06 (0.85-1.31) | 0.207 |
| | | | | 596 | Serous | **1.35 (1.10-1.64)** | 1.14 (0.86-1.52) | 0.0677 |
| *RUVBL1* | rs3732402 | 0.4 | 2382 | 194 | Endometrioid | 1.25 (0.90-1.74) | 1.24 (0.79-1.94) | 0.236 |
| | | | | 147 | Mucinous | 0.91 (0.64-1.31) | 0.70 (0.40-1.24) | 0.1882 |
| | | | | 114 | Clear cell | 1.00 (0.67-1.49) | 0.61 (0.31-1.20) | 0.2799 |
| | | | | 1645 | All | 1.08 (0.93-1.26) | 0.86 (0.72-1.03) | 0.1081 |
| | | | | 769 | **Serous** | 0.97 (0.80-1.17) | **0.74 (0.58-0.93)** | **0.009** |
| ***RUVBL1*** | **rs7650365** | 0.46 | 2672 | 256 | Endometrioid | 1.12 (0.83-1.53) | 0.74 (0.50-1.09) | 0.1777 |
| | | | | 175 | Mucinous | 1.04 (0.70-1.53) | 1.25 (0.81-1.93) | 0.371 |
| | | | | 155 | Clear cell | 1.08 (0.73-1.60) | 0.93 (0.58-1.48) | 0.7821 |
| *RUVBL1* | rs4857836 | 0.2 | 2993 | 1787 | All | 1.05 (0.93-1.19) | 0.97 (0.76-1.23) | 0.8219 |
| | | | | 845 | Serous | 1.11 (0.95-1.31) | 1.04 (0.77-1.41) | 0.3458 |
| | | | | 278 | Endometrioid | 1.05 (0.80-1.37) | 1.38 (0.89-2.14) | 0.2742 |

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| | | | | 195 | Mucinous | 0.85 (0.62-1.16) | 0.76 (0.40-1.44) | 0.1432 |
| | | | | 165 | Clear cell | 0.89 (0.63-1.24) | 0.91 (0.49-1.69) | 0.5199 |
| | | | | 1733 | All | 0.97 (0.84-1.11) | 0.93 (0.63-1.37) | 0.5613 |
| | | | | 820 | Serous | 0.99 (0.82-1.18) | 0.57 (0.31-1.06) | 0.2966 |
| ***RUVBL1*** | **rs9821568** | 0.15 | 2911 | 269 | **Endometrioid** | 1.14 (0.86-1.52) | **2.09 (1.16-3.78)** | **0.0286** |
| | | | | 186 | Mucinous | 0.86 (0.60-1.23) | 0.62 (0.19-2.00) | 0.1981 |
| | | | | 161 | Clear cell | 0.68 (0.45-1.02) | 0.85 (0.30-2.38) | 0.0967 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-P: Haplotype-specific risks of *RUVBL1* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| ***RUVBL1*** | **h000100** | 48 | All | 0.96 (0.86-1.07) | 0.491 |
| | | | **Serous** | **0.76 (0.62-0.93)** | **0.0315** |
| | | | Endometrioid | 0.75 (0.54-1.05) | 0.137 |
| | | | Mucinous | 1.22 (0.83-1.79) | 0.3395 |
| | | | Clear cell | 0.9 (0.6-1.38) | 0.624 |
| ***RUVBL1*** | **h001011** | 14.5 | All | 0.98 (0.84-1.15) | 0.833 |
| | | | Serous | 0.86 (0.64-1.17) | 0.429 |
| | | | **Endometrioid** | **1.61 (1.03-2.52)** | **0.0465** |
| | | | Mucinous | 0.59 (0.32-1.11) | 0.1085 |
| | | | Clear cell | 0.605 (0.32-1.17) | 0.1645 |
| *RUVBL1* | h000000 | 13.3 | All | 0.91 (0.77-1.07) | 0.247 |
| | | | Serous | 1.22 (0.9-1.67) | 0.3585 |
| | | | Endometrioid | 1.43 (0.89-2.32) | 0.1805 |
| | | | Mucinous | 1.23 (0.7-2.18) | 0.485 |
| | | | Clear cell | 1.51 (0.85-2.75) | 0.165 |
| ***RUVBL1*** | **h011010** | 11.7 | All | 1.15 (0.97-1.36) | 0.114 |
| | | | **Serous** | **1.44 (1.04-1.98)** | **0.027** |
| | | | Endometrioid | 0.79 (0.45-1.39) | 0.401 |
| | | | Mucinous | 0.91 (0.48-1.75) | 0.751 |
| | | | Clear cell | 1.32 (0.7-2.51) | 0.395 |
| *RUVBL1* | h111000 | 9.8 | All | 1.17 (0.99-1.4) | 0.071 |
| | | | Serous | 1.4 (0.99-2) | 0.1915 |
| | | | Endometrioid | 0.95 (0.55-1.67) | 0.8565 |
| | | | Mucinous | 0.96 (0.49-1.9) | 0.7185 |
| | | | Clear cell | 1.02 (0.51-2.1) | 0.74 |
| *RUVBL1* | Rare | 0.8 | All | 0.99 (0.54-1.8) | 0.963 |
| | | | Serous | 0.98 (0.69-1.38) | 0.892 |
| | | | Endometrioid | 0.89 (0.49-1.6) | 0.69 |
| | | | Mucinous | 0.89 (0.46-1.75) | 0.746 |
| | | | Clear cell | 1.06 (0.54-2.07) | 0.861 |

SNP order in haplotypes is 5' to 3' of *RUVBL1*: rs9860614, rs13063604, rs3732402, rs7650365, rs4857836, rs9821568.

## Appendix IV-Q: Genotype specific ratios of *STAG3* tSNPs on ovarian cancer susceptibility

| Gene | tSNP | MAF | No. controls | No. cases | Histology | HetOR (95% CI) | HomOR (95% CI) | P-trend |
|---|---|---|---|---|---|---|---|---|
| *STAG3* | rs11762932 | 0.22 | 2965 | 1787 | All | 1.03 (0.90-1.17) | 1.06 (0.80-1.40) | 0.6327 |
| | | | | 846 | Serous | 1.03 (0.88-1.22) | 1.06 (0.74-1.52) | 0.6639 |
| | | | | 279 | Endometrioid | 0.98 (0.75-1.28) | 1.22 (0.71-2.11) | 0.695 |
| | | | | 194 | Mucinous | 0.83 (0.60-1.15) | 0.79 (0.37-1.67) | 0.2541 |
| | | | | 164 | Clear cell | 1.29 (0.93-1.79) | 1.01 (0.46-2.23) | 0.2781 |
| *STAG3* | rs2246713 | 0.47 | 1765 | 1295 | All | 0.97 (0.81-1.15) | 0.96 (0.78-1.19) | 0.6593 |
| | | | | 549 | Serous | 0.95 (0.75-1.20) | 0.87 (0.66-1.16) | 0.2436 |
| | | | | 212 | Endometrioid | 0.98 (0.69-1.40) | 1.02 (0.67-1.54) | 0.8107 |
| | | | | 143 | Mucinous | 1.11 (0.71-1.73) | 0.90 (0.52-1.55) | 0.9446 |
| | | | | 130 | Clear cell | 1.02 (0.66-1.58) | 1.16 (0.70-1.92) | 0.5676 |
| ***STAG3*** | **rs1637001** | 0.26 | 2967 | 1784 | All | 0.86 (0.76-0.98) | 0.92 (0.73-1.16) | 0.0692 |
| | | | | 843 | **Serous** | **0.84 (0.71-0.99)** | 0.77 (0.56-1.05) | **0.0177** |
| | | | | 278 | Endometrioid | 0.79 (0.61-1.04) | 0.92 (0.57-1.48) | 0.2456 |
| | | | | 194 | Mucinous | 1.23 (0.90-1.68) | 1.27 (0.74-2.18) | 0.1363 |
| | | | | 165 | Clear cell | 1.02 (0.73-1.41) | 0.61 (0.29-1.28) | 0.4137 |

MAF – minor allele frequency; Het – heterozygous; Hom – homozygous; OR – odds ratio; CI – confidence interval; § compared with common homozygous; Emboldened tSNP and histology names, and P-values are statistically associated with susceptibility; emboldened OR are statistically significant or do not cross 1.

## Appendix IV-R: Haplotype-specific risks of *STAG3* on ovarian cancer susceptibility

| Gene | Haplotype | Freq (%) | Histology | OR (95% CI) | P-value |
|---|---|---|---|---|---|
| ***STAG3*** | **h000** | 50.7 | All | 1.06 (0.96-1.16) | 0.257 |
| | | | **Serous** | **1.12 (1.01-1.25)** | **0.039** |
| | | | Endometrioid | 1.06 (0.89-1.27) | 0.523 |
| | | | Mucinous | 0.97 (0.79-1.2) | 0.805 |
| | | | Clear cell | 1 (0.8-1.25) | 0.996 |
| ***STAG3*** | **h011** | 26.8 | All | 0.94 (0.85-1.05) | 0.29 |
| | | | **Serous** | **0.88 (0.78-1)** | **0.046** |
| | | | Endometrioid | 0.89 (0.73-1.09) | 0.251 |
| | | | Mucinous | 1.22 (0.97-1.53) | 0.084 |
| | | | Clear cell | 0.89 (0.69-1.16) | 0.401 |
| *STAG3* | h110 | 20.5 | All | 1.03 (0.92-1.16) | 0.627 |
| | | | Serous | 1.05 (0.92-1.2) | 0.48 |
| | | | Endometrioid | 1.03 (0.83-1.28) | 0.765 |
| | | | Mucinous | 0.87 (0.66-1.13) | 0.297 |
| | | | Clear cell | 1.16 (0.89-1.51) | 0.285 |

SNP order in haplotypes is 5’ to 3’ of *STAG3*: rs11762932, rs2246713, rs1637001

## Appendix V: Genotype distributions of tagging SNPs in candidate genes analysed with AML test

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| BCAC | 2q22.1 | rs4954956 | GEOCS | 171 | 129 | 20 | **320** | 240 | 156 | 25 | **421** |
| BCAC | 2q22.1 | rs4954956 | MALOVA | 224 | 158 | 29 | **411** | 686 | 440 | 69 | **1195** |
| BCAC | 2q22.1 | rs4954956 | SEARCH | 414 | 262 | 39 | **715** | 466 | 323 | 64 | **853** |
| BCAC | 5q11.2 | rs889312 | GEOCS | 166 | 129 | 25 | **320** | 202 | 174 | 45 | **421** |
| BCAC | 5q11.2 | rs889312 | MALOVA | 225 | 178 | 31 | **434** | 616 | 485 | 86 | **1187** |
| BCAC | 5q11.2 | rs889312 | SEARCH | 378 | 286 | 55 | **719** | 447 | 341 | 64 | **852** |
| BCAC | 8q24.21 | rs13281615 | GEOCS | 97 | 167 | 57 | **321** | 138 | 204 | 79 | **421** |
| BCAC | 8q24.21 | rs13281615 | MALOVA | 162 | 193 | 67 | **422** | 462 | 557 | 170 | **1189** |
| BCAC | 8q24.21 | rs13281615 | SEARCH | 267 | 337 | 111 | **715** | 281 | 439 | 134 | **854** |
| BCAC | 8q24.21 | rs6983267 | GEOCS | 77 | 157 | 85 | **319** | 125 | 206 | 90 | **421** |
| BCAC | 8q24.21 | rs6983267 | MALOVA | 100 | 156 | 73 | **329** | 311 | 572 | 285 | **1168** |
| BCAC | 8q24.21 | rs6983267 | SEARCH | 210 | 370 | 136 | **716** | 224 | 423 | 203 | **850** |
| BCAC | 8q24.21 | rs9283954 | GEOCS | 250 | 65 | 6 | **321** | 340 | 79 | 2 | **421** |
| BCAC | 8q24.21 | rs9283954 | MALOVA | 290 | 80 | 4 | **374** | 927 | 209 | 13 | **1149** |
| BCAC | 8q24.21 | rs9283954 | SEARCH | 582 | 124 | 8 | **714** | 683 | 158 | 9 | **850** |
| BCAC | 11p15.5 | rs2107425 | GEOCS | 156 | 124 | 39 | **319** | 170 | 204 | 48 | **422** |
| BCAC | 11p15.5 | rs2107425 | MALOVA | 226 | 157 | 50 | **433** | 558 | 518 | 114 | **1190** |
| BCAC | 11p15.5 | rs2107425 | SEARCH | 385 | 263 | 60 | **708** | 390 | 376 | 85 | **851** |
| BCAC | 12p11.22 | rs7313833 | GEOCS | 138 | 147 | 35 | **320** | 194 | 174 | 53 | **421** |
| BCAC | 12p11.22 | rs7313833 | MALOVA | 187 | 191 | 45 | **423** | 527 | 505 | 152 | **1184** |
| BCAC | 12p11.22 | rs7313833 | SEARCH | 315 | 308 | 94 | **717** | 397 | 372 | 83 | **852** |
| BCAC | *FLJ41481* | rs4666451 | GEOCS | 107 | 161 | 49 | **317** | 146 | 209 | 65 | **420** |
| BCAC | *FLJ41481* | rs4666451 | MALOVA | 153 | 195 | 74 | **422** | 417 | 584 | 198 | **1199** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| BCAC | *FLJ41481* | rs4666451 | SEARCH | 260 | 336 | 119 | **715** | 306 | 409 | 135 | **850** |
| BCAC | *HCN1* | rs981782 | GEOCS | 103 | 150 | 67 | **320** | 136 | 204 | 80 | **420** |
| BCAC | *HCN1* | rs981782 | MALOVA | 98 | 192 | 93 | **383** | 327 | 577 | 258 | **1162** |
| BCAC | *HCN1* | rs981782 | SEARCH | 180 | 381 | 155 | **716** | 233 | 404 | 213 | **850** |
| BCAC | LOC100131885 | rs2981582 | GEOCS | 117 | 161 | 42 | **320** | 149 | 205 | 69 | **423** |
| BCAC | LOC100131885 | rs2981582 | MALOVA | 169 | 188 | 73 | **430** | 469 | 537 | 183 | **1189** |
| BCAC | LOC100131885 | rs2981582 | SEARCH | 296 | 314 | 103 | **713** | 302 | 420 | 132 | **854** |
| BCAC | *LSP1* | rs3817198 | GEOCS | 140 | 151 | 31 | **322** | 226 | 162 | 35 | **423** |
| BCAC | *LSP1* | rs3817198 | MALOVA | 195 | 183 | 48 | **426** | 606 | 460 | 101 | **1167** |
| BCAC | *LSP1* | rs3817198 | SEARCH | 313 | 306 | 90 | **709** | 380 | 381 | 84 | **845** |
| BCAC | *TOX3* | rs12443621 | GEOCS | 63 | 186 | 72 | **321** | 100 | 211 | 110 | **421** |
| BCAC | *TOX3* | rs12443621 | MALOVA | 156 | 190 | 69 | **415** | 395 | 561 | 232 | **1188** |
| BCAC | *TOX3* | rs12443621 | SEARCH | 214 | 352 | 143 | **709** | 237 | 425 | 186 | **848** |
| Cell cycle | *CCND1* | rs602652 | GEOCS | 93 | 153 | 70 | **316** | 121 | 217 | 85 | **423** |
| Cell cycle | *CCND1* | rs602652 | MALOVA | 108 | 206 | 123 | **437** | 364 | 590 | 263 | **1217** |
| Cell cycle | *CCND1* | rs602652 | SEARCH | 193 | 381 | 141 | **715** | 257 | 418 | 178 | **853** |
| Cell cycle | *CCND1* | rs3862792 | GEOCS | 302 | 19 | 0 | **321** | 402 | 19 | 0 | **421** |
| Cell cycle | *CCND1* | rs3862792 | MALOVA | 414 | 18 | 0 | **432** | 1130 | 64 | 0 | **1194** |
| Cell cycle | *CCND1* | rs3862792 | SEARCH | 688 | 30 | 0 | **718** | 803 | 50 | 0 | **853** |
| Cell cycle | *CCND1* | rs603965 | GEOCS | 97 | 151 | 73 | **321** | 129 | 217 | 79 | **425** |
| Cell cycle | *CCND1* | rs603965 | MALOVA | 117 | 202 | 118 | **437** | 365 | 578 | 242 | **1185** |
| Cell cycle | *CCND1* | rs603965 | SEARCH | 212 | 368 | 138 | **718** | 269 | 427 | 158 | **854** |
| Cell cycle | *CCND1* | rs3212879 | GEOCS | 87 | 144 | 85 | **316** | 101 | 211 | 108 | **420** |
| Cell cycle | *CCND1* | rs3212879 | MALOVA | 145 | 206 | 90 | **441** | 324 | 608 | 285 | **1217** |
| Cell cycle | *CCND1* | rs3212879 | SEARCH | 206 | 355 | 154 | **715** | 236 | 419 | 199 | **854** |
| Cell cycle | *CCND1* | rs3212891 | GEOCS | 91 | 151 | 79 | **321** | 118 | 211 | 95 | **424** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CCND1* | rs3212891 | MALOVA | 153 | 203 | 79 | **435** | 356 | 589 | 253 | **1198** |
| Cell cycle | *CCND1* | rs3212891 | SEARCH | 228 | 354 | 137 | **719** | 245 | 428 | 181 | **854** |
| Cell cycle | *CCND1* | rs678653 | GEOCS | 108 | 145 | 50 | **303** | 138 | 195 | 60 | **393** |
| Cell cycle | *CCND1* | rs678653 | MALOVA | 202 | 191 | 50 | **443** | 538 | 538 | 145 | **1221** |
| Cell cycle | *CCND1* | rs678653 | SEARCH | 302 | 321 | 96 | **719** | 341 | 390 | 123 | **854** |
| Cell cycle | *CCND1* | rs7178 | GEOCS | 279 | 41 | 1 | **321** | 369 | 49 | 5 | **423** |
| Cell cycle | *CCND1* | rs7178 | MALOVA | 359 | 78 | 5 | **442** | 1040 | 168 | 7 | **1215** |
| Cell cycle | *CCND1* | rs7178 | SEARCH | 596 | 119 | 2 | **717** | 732 | 120 | 1 | **853** |
| Cell cycle | *CCND2* | rs3217795 | GEOCS | 277 | 44 | 2 | **323** | 356 | 66 | 6 | **428** |
| Cell cycle | *CCND2* | rs3217795 | MALOVA | 365 | 76 | 2 | **443** | 1026 | 188 | 4 | **1218** |
| Cell cycle | *CCND2* | rs3217795 | SEARCH | 601 | 115 | 3 | **719** | 709 | 140 | 5 | **854** |
| Cell cycle | *CCND2* | rs3217805 | GEOCS | 120 | 158 | 44 | **322** | 171 | 183 | 73 | **427** |
| Cell cycle | *CCND2* | rs3217805 | MALOVA | 152 | 224 | 65 | **441** | 435 | 589 | 197 | **1221** |
| Cell cycle | *CCND2* | rs3217805 | SEARCH | 252 | 343 | 100 | **695** | 299 | 405 | 150 | **854** |
| Cell cycle | *CCND2* | rs3217820 | GEOCS | 127 | 154 | 41 | **322** | 181 | 185 | 59 | **425** |
| Cell cycle | *CCND2* | rs3217820 | MALOVA | 195 | 192 | 56 | **443** | 532 | 532 | 156 | **1220** |
| Cell cycle | *CCND2* | rs3217820 | SEARCH | 247 | 314 | 81 | **642** | 361 | 378 | 113 | **852** |
| Cell cycle | *CCND2* | rs3217852 | GEOCS | 194 | 113 | 14 | **321** | 250 | 147 | 23 | **420** |
| Cell cycle | *CCND2* | rs3217852 | MALOVA | 252 | 158 | 25 | **435** | 702 | 438 | 76 | **1216** |
| Cell cycle | *CCND2* | rs3217852 | SEARCH | 422 | 261 | 35 | **718** | 493 | 310 | 49 | **852** |
| Cell cycle | *CCND2* | rs3217862 | GEOCS | 224 | 93 | 3 | **320** | 308 | 104 | 12 | **424** |
| Cell cycle | *CCND2* | rs3217862 | MALOVA | 308 | 116 | 11 | **435** | 850 | 332 | 34 | **1216** |
| Cell cycle | *CCND2* | rs3217862 | SEARCH | 503 | 186 | 26 | **715** | 566 | 263 | 24 | **853** |
| Cell cycle | *CCND2* | rs3217863 | GEOCS | 267 | 47 | 6 | **320** | 331 | 85 | 6 | **422** |
| Cell cycle | *CCND2* | rs3217863 | MALOVA | 340 | 77 | 5 | **422** | 1018 | 176 | 5 | **1199** |
| Cell cycle | *CCND2* | rs3217863 | SEARCH | 615 | 96 | 6 | **717** | 725 | 119 | 7 | **851** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CCND2* | rs3217869 | GEOCS | 118 | 152 | 52 | **322** | 170 | 174 | 84 | **428** |
| Cell cycle | *CCND2* | rs3217869 | MALOVA | 152 | 215 | 76 | **443** | 446 | 564 | 206 | **1216** |
| Cell cycle | *CCND2* | rs3217869 | SEARCH | 268 | 322 | 128 | **718** | 298 | 430 | 126 | **854** |
| Cell cycle | *CCND2* | rs3217901 | GEOCS | 105 | 165 | 50 | **320** | 124 | 225 | 72 | **421** |
| Cell cycle | *CCND2* | rs3217901 | MALOVA | 136 | 208 | 97 | **441** | 386 | 590 | 240 | **1216** |
| Cell cycle | *CCND2* | rs3217901 | SEARCH | 214 | 382 | 122 | **718** | 294 | 394 | 164 | **852** |
| Cell cycle | *CCND2* | rs3217906 | GEOCS | 190 | 112 | 19 | **321** | 246 | 153 | 23 | **422** |
| Cell cycle | *CCND2* | rs3217906 | MALOVA | 227 | 170 | 31 | **428** | 661 | 457 | 79 | **1197** |
| Cell cycle | *CCND2* | rs3217906 | SEARCH | 392 | 283 | 43 | **718** | 472 | 319 | 63 | **854** |
| Cell cycle | *CCND2* | rs3217916 | GEOCS | 165 | 142 | 17 | **324** | 238 | 148 | 38 | **424** |
| Cell cycle | *CCND2* | rs3217916 | MALOVA | 249 | 163 | 21 | **433** | 602 | 494 | 103 | **1199** |
| Cell cycle | *CCND2* | rs3217916 | SEARCH | 361 | 297 | 61 | **719** | 440 | 342 | 72 | **854** |
| Cell cycle | *CCND2* | rs3217925 | GEOCS | 174 | 133 | 13 | **320** | 255 | 141 | 27 | **423** |
| Cell cycle | *CCND2* | rs3217925 | MALOVA | 265 | 154 | 16 | **435** | 628 | 475 | 84 | **1187** |
| Cell cycle | *CCND2* | rs3217925 | SEARCH | 390 | 277 | 48 | **715** | 468 | 323 | 60 | **851** |
| Cell cycle | *CCND2* | rs3217926 | GEOCS | 119 | 155 | 48 | **322** | 157 | 192 | 78 | **427** |
| Cell cycle | *CCND2* | rs3217926 | MALOVA | 174 | 204 | 66 | **444** | 477 | 569 | 174 | **1220** |
| Cell cycle | *CCND2* | rs3217926 | SEARCH | 274 | 342 | 102 | **718** | 332 | 375 | 146 | **853** |
| Cell cycle | *CCND2* | rs3217933 | GEOCS | 182 | 122 | 20 | **324** | 242 | 164 | 18 | **424** |
| Cell cycle | *CCND2* | rs3217933 | MALOVA | 221 | 180 | 34 | **435** | 662 | 458 | 76 | **1196** |
| Cell cycle | *CCND2* | rs3217933 | SEARCH | 389 | 294 | 36 | **719** | 479 | 309 | 66 | **854** |
| Cell cycle | *CCND2* | rs3217936 | GEOCS | 151 | 147 | 24 | **322** | 218 | 162 | 47 | **427** |
| Cell cycle | *CCND2* | rs3217936 | MALOVA | 210 | 197 | 32 | **439** | 512 | 559 | 139 | **1210** |
| Cell cycle | *CCND2* | rs3217936 | SEARCH | 320 | 322 | 75 | **717** | 386 | 372 | 94 | **852** |
| Cell cycle | *CCND3* | Rs1410492 | GEOCS | 181 | 125 | 17 | **323** | 251 | 142 | 34 | **427** |
| Cell cycle | *CCND3* | Rs1410492 | MALOVA | 245 | 173 | 24 | **442** | 708 | 423 | 86 | **1217** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CCND3* | Rs1410492 | SEARCH | 398 | 265 | 51 | **714** | 478 | 333 | 35 | **846** |
| Cell cycle | *CCND3* | rs3218092 | GEOCS | 212 | 96 | 13 | **321** | 261 | 130 | 32 | **423** |
| Cell cycle | *CCND3* | rs3218092 | MALOVA | 265 | 149 | 17 | **431** | 783 | 363 | 46 | **1192** |
| Cell cycle | *CCND3* | rs3218092 | SEARCH | 492 | 208 | 16 | **716** | 569 | 257 | 27 | **853** |
| Cell cycle | *CCND3* | Rs2479717 | GEOCS | 160 | 139 | 16 | **315** | 244 | 150 | 31 | **425** |
| Cell cycle | *CCND3* | Rs2479717 | MALOVA | 224 | 173 | 45 | **442** | 667 | 454 | 96 | **1217** |
| Cell cycle | *CCND3* | Rs2479717 | SEARCH | 382 | 276 | 57 | **715** | 414 | 352 | 81 | **847** |
| Cell cycle | *CCND3* | Rs1051130 | GEOCS | 81 | 173 | 65 | **319** | 112 | 203 | 110 | **425** |
| Cell cycle | *CCND3* | Rs1051130 | MALOVA | 118 | 215 | 110 | **443** | 373 | 540 | 298 | **1211** |
| Cell cycle | *CCND3* | Rs1051130 | SEARCH | 213 | 362 | 138 | **713** | 235 | 398 | 219 | **852** |
| Cell cycle | *CCND3* | rs9529 | GEOCS | 149 | 152 | 20 | **321** | 221 | 168 | 34 | **423** |
| Cell cycle | *CCND3* | rs9529 | MALOVA | 242 | 162 | 37 | **441** | 633 | 456 | 110 | **1199** |
| Cell cycle | *CCND3* | rs9529 | SEARCH | 381 | 281 | 57 | **719** | 423 | 340 | 90 | **853** |
| Cell cycle | *CCND3* | rs3218110 | GEOCS | 184 | 123 | 15 | **322** | 246 | 152 | 24 | **422** |
| Cell cycle | *CCND3* | rs3218110 | MALOVA | 250 | 157 | 31 | **438** | 684 | 443 | 79 | **1206** |
| Cell cycle | *CCND3* | rs3218110 | SEARCH | 382 | 295 | 42 | **719** | 503 | 308 | 43 | **854** |
| Cell cycle | *CCND3* | rs3218114 | GEOCS | 225 | 87 | 10 | **322** | 274 | 126 | 22 | **422** |
| Cell cycle | *CCND3* | rs3218114 | MALOVA | 279 | 147 | 17 | **443** | 812 | 356 | 48 | **1216** |
| Cell cycle | *CCND3* | rs3218114 | SEARCH | 496 | 207 | 16 | **719** | 576 | 252 | 26 | **854** |
| Cell cycle | *CCNE1* | rs997669 | GEOCS | 120 | 154 | 48 | **322** | 172 | 204 | 50 | **426** |
| Cell cycle | *CCNE1* | rs997669 | MALOVA | 151 | 218 | 73 | **442** | 436 | 601 | 180 | **1217** |
| Cell cycle | *CCNE1* | rs997669 | SEARCH | 259 | 349 | 108 | **716** | 331 | 389 | 134 | **854** |
| Cell cycle | *CCNE1* | rs3218036 | GEOCS | 147 | 146 | 26 | **319** | 207 | 183 | 33 | **423** |
| Cell cycle | *CCNE1* | rs3218036 | MALOVA | 192 | 193 | 54 | **439** | 545 | 546 | 114 | **1205** |
| Cell cycle | *CCNE1* | rs3218036 | SEARCH | 325 | 310 | 83 | **718** | 414 | 355 | 84 | **853** |
| Cell cycle | *CCNE1* | rs3218038 | GEOCS | 287 | 33 | 2 | **322** | 377 | 48 | 1 | **426** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CCNE1* | rs3218038 | MALOVA | 418 | 22 | 0 | **440** | 1160 | 58 | 0 | **1218** |
| Cell cycle | *CCNE1* | rs3218038 | SEARCH | 660 | 56 | 1 | **717** | 784 | 68 | 1 | **853** |
| Cell cycle | *CCNE1* | rs3218076 | GEOCS | 164 | 130 | 26 | **320** | 205 | 185 | 37 | **427** |
| Cell cycle | *CCNE1* | rs3218076 | MALOVA | 211 | 203 | 28 | **442** | 614 | 512 | 92 | **1218** |
| Cell cycle | *CCNE1* | rs3218076 | SEARCH | 370 | 291 | 57 | **718** | 445 | 352 | 55 | **852** |
| Cell cycle | *CDK2* | Rs2069408 | GEOCS | 167 | 128 | 27 | **322** | 194 | 190 | 43 | **427** |
| Cell cycle | *CDK2* | Rs2069408 | MALOVA | 189 | 202 | 46 | **437** | 522 | 549 | 145 | **1216** |
| Cell cycle | *CDK2* | Rs2069408 | SEARCH | 332 | 300 | 85 | **717** | 369 | 388 | 96 | **853** |
| Cell cycle | *CDK2* | Rs1045435 | GEOCS | 266 | 56 | 1 | **323** | 367 | 57 | 2 | **426** |
| Cell cycle | *CDK2* | Rs1045435 | MALOVA | 356 | 82 | 3 | **441** | 996 | 209 | 13 | **1218** |
| Cell cycle | *CDK2* | Rs1045435 | SEARCH | 583 | 123 | 12 | **718** | 708 | 144 | 1 | **853** |
| Cell cycle | *CDK4* | rs2270777 | GEOCS | 87 | 180 | 53 | **320** | 141 | 217 | 68 | **426** |
| Cell cycle | *CDK4* | rs2270777 | MALOVA | 138 | 231 | 70 | **439** | 379 | 626 | 207 | **1212** |
| Cell cycle | *CDK4* | rs2270777 | SEARCH | 238 | 347 | 133 | **718** | 289 | 419 | 145 | **853** |
| Cell cycle | *CDK4* | rs2069506 | GEOCS | 155 | 134 | 34 | **323** | 211 | 164 | 50 | **425** |
| Cell cycle | *CDK4* | rs2069506 | MALOVA | 214 | 178 | 51 | **443** | 507 | 556 | 144 | **1207** |
| Cell cycle | *CDK4* | rs2069506 | SEARCH | 321 | 305 | 88 | **714** | 396 | 363 | 91 | **850** |
| Cell cycle | *CDK6* | rs8179 | GEOCS | 200 | 106 | 15 | **321** | 265 | 139 | 14 | **418** |
| Cell cycle | *CDK6* | rs8179 | MALOVA | 264 | 150 | 22 | **436** | 721 | 420 | 56 | **1197** |
| Cell cycle | *CDK6* | rs8179 | SEARCH | 447 | 230 | 31 | **708** | 508 | 280 | 37 | **825** |
| Cell cycle | *CDK6* | rs2285332 | GEOCS | 180 | 117 | 22 | **319** | 208 | 177 | 37 | **422** |
| Cell cycle | *CDK6* | rs2285332 | MALOVA | 264 | 151 | 24 | **439** | 681 | 450 | 66 | **1197** |
| Cell cycle | *CDK6* | rs2285332 | SEARCH | 414 | 263 | 41 | **718** | 495 | 298 | 51 | **844** |
| Cell cycle | *CDK6* | rs42046 | GEOCS | 152 | 142 | 27 | **321** | 212 | 168 | 26 | **406** |
| Cell cycle | *CDK6* | rs42046 | MALOVA | 216 | 188 | 31 | **435** | 626 | 492 | 86 | **1204** |
| Cell cycle | *CDK6* | rs42046 | SEARCH | 383 | 279 | 48 | **710** | 448 | 328 | 60 | **836** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CDK6* | rs3731348 | GEOCS | 273 | 44 | 1 | **318** | 372 | 46 | 2 | **420** |
| Cell cycle | *CDK6* | rs8 | GEOCS | 203 | 102 | 9 | **314** | 277 | 127 | 10 | **414** |
| Cell cycle | *CDK6* | rs3731348 | MALOVA | 384 | 42 | 1 | **427** | 1049 | 139 | 4 | **1192** |
| Cell cycle | *CDK6* | rs8 | MALOVA | 255 | 162 | 26 | **443** | 736 | 428 | 55 | **1219** |
| Cell cycle | *CDK6* | rs3731348 | SEARCH | 631 | 86 | 1 | **718** | 761 | 86 | 6 | **853** |
| Cell cycle | *CDK6* | rs8 | SEARCH | 412 | 269 | 35 | **716** | 543 | 276 | 29 | **848** |
| Cell cycle | *CDK6* | rs2237570 | GEOCS | 264 | 54 | 2 | **320** | 329 | 87 | 6 | **422** |
| Cell cycle | *CDK6* | rs2237570 | MALOVA | 344 | 91 | 3 | **438** | 946 | 240 | 13 | **1199** |
| Cell cycle | *CDK6* | rs2237570 | SEARCH | 572 | 134 | 10 | **716** | 652 | 182 | 13 | **847** |
| Cell cycle | *CDK6* | rs3731343 | GEOCS | 111 | 153 | 59 | **323** | 156 | 191 | 77 | **424** |
| Cell cycle | *CDK6* | rs3731343 | MALOVA | 126 | 205 | 111 | **442** | 335 | 622 | 264 | **1221** |
| Cell cycle | *CDK6* | rs3731343 | SEARCH | 184 | 380 | 155 | **719** | 238 | 427 | 189 | **854** |
| Cell cycle | *CDK6* | rs3757823 | GEOCS | 242 | 77 | 5 | **324** | 341 | 84 | 3 | **428** |
| Cell cycle | *CDK6* | rs3757823 | MALOVA | 368 | 71 | 4 | **443** | 956 | 247 | 9 | **1212** |
| Cell cycle | *CDK6* | rs3757823 | SEARCH | 578 | 137 | 4 | **719** | 713 | 134 | 8 | **855** |
| Cell cycle | *CDK6* | rs2079147 | GEOCS | 69 | 165 | 90 | **324** | 95 | 224 | 108 | **427** |
| Cell cycle | *CDK6* | rs2079147 | MALOVA | 116 | 215 | 107 | **438** | 300 | 628 | 287 | **1215** |
| Cell cycle | *CDK6* | rs2079147 | SEARCH | 191 | 372 | 156 | **719** | 238 | 417 | 199 | **854** |
| Cell cycle | *CDK6* | rs2282991 | GEOCS | 274 | 46 | 1 | **321** | 330 | 84 | 4 | **418** |
| Cell cycle | *CDK6* | rs2282991 | MALOVA | 342 | 77 | 3 | **422** | 935 | 215 | 9 | **1159** |
| Cell cycle | *CDK6* | rs2282991 | SEARCH | 580 | 116 | 10 | **706** | 659 | 157 | 11 | **827** |
| Cell cycle | *CDK6* | rs4729049 | GEOCS | 263 | 59 | 3 | **325** | 351 | 75 | 2 | **428** |
| Cell cycle | *CDK6* | rs4729049 | MALOVA | 360 | 74 | 4 | **438** | 975 | 225 | 16 | **1216** |
| Cell cycle | *CDK6* | rs4729049 | SEARCH | 574 | 139 | 6 | **719** | 692 | 154 | 9 | **855** |
| Cell cycle | *CDK6* | rs445 | GEOCS | 235 | 78 | 7 | **320** | 331 | 87 | 8 | **426** |
| Cell cycle | *CDK6* | rs445 | MALOVA | 370 | 70 | 2 | **442** | 985 | 224 | 9 | **1218** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CDK6* | rs445 | SEARCH | 595 | 120 | 4 | **719** | 703 | 144 | 6 | **853** |
| Cell cycle | *CDK6* | rs992519 | GEOCS | 235 | 78 | 6 | **319** | 308 | 107 | 8 | **423** |
| Cell cycle | *CDK6* | rs992519 | MALOVA | 320 | 111 | 9 | **440** | 935 | 262 | 22 | **1219** |
| Cell cycle | *CDK6* | rs992519 | SEARCH | 533 | 171 | 14 | **718** | 625 | 210 | 19 | **854** |
| Cell cycle | *CDKN1A* | rs762624 | GEOCS | 149 | 142 | 29 | **320** | 206 | 180 | 40 | **426** |
| Cell cycle | *CDKN1A* | rs762624 | MALOVA | 230 | 171 | 36 | **437** | 673 | 463 | 70 | **1206** |
| Cell cycle | *CDKN1A* | rs762624 | SEARCH | 385 | 281 | 52 | **718** | 458 | 327 | 66 | **851** |
| Cell cycle | *CDKN1A* | rs2395655 | GEOCS | 103 | 163 | 52 | **318** | 164 | 192 | 66 | **422** |
| Cell cycle | *CDKN1A* | rs2395655 | MALOVA | 161 | 203 | 78 | **442** | 484 | 567 | 166 | **1217** |
| Cell cycle | *CDKN1A* | rs2395655 | SEARCH | 274 | 327 | 117 | **718** | 331 | 374 | 143 | **848** |
| Cell cycle | *CDKN1A* | rs3176331 | GEOCS | 241 | 77 | 4 | **322** | 324 | 95 | 5 | **424** |
| Cell cycle | *CDKN1A* | rs3176331 | MALOVA | 325 | 93 | 13 | **431** | 911 | 263 | 22 | **1196** |
| Cell cycle | *CDKN1A* | rs3176331 | SEARCH | 545 | 161 | 12 | **718** | 642 | 192 | 18 | **852** |
| Cell cycle | *CDKN1A* | rs3176336 | GEOCS | 100 | 154 | 66 | **320** | 141 | 199 | 82 | **422** |
| Cell cycle | *CDKN1A* | rs3176336 | MALOVA | 179 | 203 | 62 | **444** | 461 | 577 | 178 | **1216** |
| Cell cycle | *CDKN1A* | rs3176336 | SEARCH | 272 | 334 | 111 | **717** | 318 | 385 | 147 | **850** |
| Cell cycle | *CDKN1A* | rs3176343 | GEOCS | 279 | 43 | 0 | **322** | 380 | 41 | 1 | **422** |
| Cell cycle | *CDKN1A* | rs3176343 | MALOVA | 405 | 36 | 1 | **442** | 1101 | 113 | 1 | **1215** |
| Cell cycle | *CDKN1A* | rs3176343 | SEARCH | 639 | 79 | 1 | **719** | 758 | 94 | 2 | **854** |
| Cell cycle | *CDKN1A* | rs1801270 | GEOCS | 259 | 56 | 3 | **318** | 334 | 80 | 5 | **419** |
| Cell cycle | *CDKN1A* | rs1801270 | MALOVA | 394 | 49 | 0 | **443** | 1046 | 147 | 3 | **1196** |
| Cell cycle | *CDKN1A* | rs1801270 | SEARCH | 619 | 95 | 2 | **716** | 725 | 118 | 5 | **848** |
| Cell cycle | *CDKN1A* | rs3176352 | GEOCS | 158 | 131 | 32 | **321** | 213 | 173 | 37 | **423** |
| Cell cycle | *CDKN1A* | rs3176352 | MALOVA | 236 | 162 | 39 | **437** | 680 | 442 | 79 | **1201** |
| Cell cycle | *CDKN1A* | rs3176352 | SEARCH | 388 | 278 | 53 | **719** | 460 | 328 | 64 | **852** |
| Cell cycle | *CDKN1A* | rs1059234 | GEOCS | 263 | 55 | 4 | **322** | 338 | 82 | 5 | **425** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CDKN1A* | rs1059234 | MALOVA | 394 | 48 | 1 | **443** | 1053 | 147 | 4 | **1204** |
| Cell cycle | *CDKN1A* | rs1059234 | SEARCH | 624 | 92 | 2 | **718** | 733 | 114 | 5 | **852** |
| Cell cycle | *CDKN1A* | rs6457937 | GEOCS | 307 | 18 | 0 | **325** | 401 | 26 | 0 | **427** |
| Cell cycle | *CDKN1A* | rs6457937 | MALOVA | 414 | 25 | 0 | **439** | 1148 | 66 | 0 | **1214** |
| Cell cycle | *CDKN1A* | rs6457937 | SEARCH | 686 | 33 | 0 | **719** | 812 | 42 | 0 | **854** |
| Cell cycle | *CDKN1B* | rs3759217 | GEOCS | 225 | 88 | 9 | **322** | 314 | 101 | 8 | **423** |
| Cell cycle | *CDKN1B* | rs3759217 | MALOVA | 345 | 90 | 8 | **443** | 958 | 243 | 14 | **1215** |
| Cell cycle | *CDKN1B* | rs3759217 | SEARCH | 546 | 167 | 5 | **718** | 671 | 169 | 14 | **854** |
| Cell cycle | *CDKN1B* | rs34330 | GEOCS | 189 | 122 | 14 | **325** | 249 | 154 | 26 | **429** |
| Cell cycle | *CDKN1B* | rs34330 | MALOVA | 262 | 144 | 28 | **434** | 703 | 417 | 66 | **1186** |
| Cell cycle | *CDKN1B* | rs34330 | SEARCH | 413 | 259 | 37 | **709** | 470 | 330 | 52 | **852** |
| Cell cycle | *CDKN1B* | rs2066827 | GEOCS | 189 | 118 | 14 | **321** | 239 | 152 | 33 | **424** |
| Cell cycle | *CDKN1B* | rs2066827 | MALOVA | 249 | 165 | 28 | **442** | 660 | 468 | 79 | **1207** |
| Cell cycle | *CDKN1B* | rs2066827 | SEARCH | 449 | 235 | 34 | **718** | 476 | 314 | 63 | **853** |
| Cell cycle | *CDKN1B* | rs34329 | GEOCS | 140 | 149 | 33 | **322** | 179 | 186 | 57 | **422** |
| Cell cycle | *CDKN1B* | rs34329 | MALOVA | 212 | 191 | 37 | **440** | 590 | 513 | 112 | **1215** |
| Cell cycle | *CDKN1B* | rs34329 | SEARCH | 337 | 312 | 70 | **719** | 402 | 380 | 72 | **854** |
| Cell cycle | *CDKN1B* | rs3093736 | GEOCS | 305 | 17 | 0 | **322** | 406 | 17 | 0 | **423** |
| Cell cycle | *CDKN1B* | rs3093736 | MALOVA | 409 | 32 | 1 | **442** | 1119 | 94 | 3 | **1216** |
| Cell cycle | *CDKN1B* | rs3093736 | SEARCH | 675 | 43 | 1 | **719** | 794 | 60 | 0 | **854** |
| Cell cycle | *CDKN1B* | rs7330 | GEOCS | 112 | 157 | 54 | **323** | 154 | 194 | 78 | **426** |
| Cell cycle | *CDKN1B* | rs7330 | MALOVA | 177 | 194 | 71 | **442** | 468 | 562 | 190 | **1220** |
| Cell cycle | *CDKN1B* | rs7330 | SEARCH | 260 | 338 | 121 | **719** | 297 | 427 | 129 | **853** |
| Cell cycle | *CDKN1B* | rs1420023 | GEOCS | 271 | 48 | 3 | **322** | 342 | 77 | 2 | **421** |
| Cell cycle | *CDKN1B* | rs1420023 | MALOVA | 346 | 83 | 7 | **436** | 969 | 229 | 15 | **1213** |
| Cell cycle | *CDKN1B* | rs1420023 | SEARCH | 565 | 144 | 8 | **717** | 667 | 173 | 12 | **852** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CDKN2A* | Rs3217992 | GEOCS | 119 | 157 | 45 | **321** | 158 | 200 | 68 | **426** |
| Cell cycle | *CDKN2A* | Rs3217992 | MALOVA | 206 | 182 | 48 | **436** | 514 | 567 | 136 | **1217** |
| Cell cycle | *CDKN2A* | Rs3217992 | SEARCH | 282 | 351 | 84 | **717** | 343 | 399 | 104 | **846** |
| Cell cycle | *CDKN2A* | Rs3218005 | GEOCS | 252 | 59 | 5 | **316** | 345 | 69 | 3 | **417** |
| Cell cycle | *CDKN2A* | Rs3218005 | MALOVA | 346 | 79 | 5 | **430** | 979 | 215 | 11 | **1205** |
| Cell cycle | *CDKN2A* | Rs3218005 | SEARCH | 588 | 117 | 10 | **715** | 702 | 142 | 7 | **851** |
| Cell cycle | *CDKN2A* | Rs2811712 | GEOCS | 254 | 61 | 5 | **320** | 349 | 73 | 3 | **425** |
| Cell cycle | *CDKN2A* | Rs2811712 | MALOVA | 347 | 88 | 8 | **443** | 966 | 239 | 14 | **1219** |
| Cell cycle | *CDKN2A* | Rs2811712 | SEARCH | 584 | 119 | 13 | **716** | 681 | 162 | 7 | **850** |
| Cell cycle | *CDKN2A* | Rs3218020 | GEOCS | 123 | 154 | 38 | **315** | 166 | 178 | 56 | **400** |
| Cell cycle | *CDKN2A* | Rs3218020 | MALOVA | 226 | 175 | 39 | **440** | 565 | 539 | 110 | **1214** |
| Cell cycle | *CDKN2A* | Rs3218020 | SEARCH | 303 | 332 | 77 | **712** | 377 | 377 | 95 | **849** |
| Cell cycle | *CDKN2A* | rs3731197 | GEOCS | 110 | 157 | 52 | **319** | 147 | 192 | 83 | **422** |
| Cell cycle | *CDKN2A* | rs3731197 | MALOVA | 175 | 208 | 60 | **443** | 488 | 561 | 156 | **1205** |
| Cell cycle | *CDKN2A* | rs3731197 | SEARCH | 274 | 328 | 112 | **714** | 331 | 410 | 111 | **852** |
| Cell cycle | *CDKN2A* | Rs3731211 | GEOCS | 164 | 128 | 29 | **321** | 215 | 177 | 33 | **425** |
| Cell cycle | *CDKN2A* | Rs3731211 | MALOVA | 228 | 181 | 31 | **440** | 646 | 476 | 86 | **1208** |
| Cell cycle | *CDKN2A* | Rs3731211 | SEARCH | 395 | 269 | 51 | **715** | 443 | 353 | 55 | **851** |
| Cell cycle | *CDKN2A* | Rs3731222 | GEOCS | 235 | 84 | 2 | **321** | 303 | 114 | 8 | **425** |
| Cell cycle | *CDKN2A* | Rs3731222 | MALOVA | 343 | 92 | 8 | **443** | 944 | 252 | 23 | **1219** |
| Cell cycle | *CDKN2A* | Rs3731222 | SEARCH | 532 | 170 | 14 | **716** | 641 | 194 | 18 | **853** |
| Cell cycle | *CDKN2A* | rs4074785 | GEOCS | 258 | 55 | 8 | **321** | 331 | 82 | 9 | **422** |
| Cell cycle | *CDKN2A* | rs4074785 | MALOVA | 362 | 67 | 2 | **431** | 990 | 192 | 7 | **1189** |
| Cell cycle | *CDKN2A* | rs4074785 | SEARCH | 565 | 146 | 8 | **719** | 699 | 149 | 5 | **853** |
| Cell cycle | *CDKN2A* | rs3731239 | GEOCS | 139 | 152 | 32 | **323** | 201 | 171 | 49 | **421** |
| Cell cycle | *CDKN2A* | rs3731239 | MALOVA | 146 | 225 | 70 | **441** | 461 | 577 | 180 | **1218** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CDKN2A* | rs3731239 | SEARCH | 283 | 332 | 99 | **714** | 319 | 419 | 110 | **848** |
| Cell cycle | *CDKN2A* | rs3731249 | GEOCS | 301 | 18 | 2 | **321** | 397 | 22 | 1 | **420** |
| Cell cycle | *CDKN2A* | rs3731249 | MALOVA | 405 | 30 | 0 | **435** | 1094 | 98 | 1 | **1193** |
| Cell cycle | *CDKN2A* | rs3731249 | SEARCH | 685 | 22 | 1 | **708** | 804 | 33 | 2 | **839** |
| Cell cycle | *CDKN2A* | rs11515 | GEOCS | 227 | 82 | 12 | **321** | 316 | 104 | 6 | **426** |
| Cell cycle | *CDKN2A* | rs11515 | MALOVA | 304 | 127 | 9 | **440** | 870 | 321 | 25 | **1216** |
| Cell cycle | *CDKN2A* | rs11515 | SEARCH | 544 | 157 | 17 | **718** | 624 | 214 | 12 | **850** |
| Cell cycle | *CDKN2A* | rs3088440 | GEOCS | 261 | 55 | 6 | **322** | 329 | 89 | 9 | **427** |
| Cell cycle | *CDKN2A* | rs3088440 | MALOVA | 367 | 67 | 2 | **436** | 1024 | 186 | 5 | **1215** |
| Cell cycle | *CDKN2A* | rs3088440 | SEARCH | 567 | 143 | 7 | **717** | 695 | 150 | 3 | **848** |
| Cell cycle | *CDKN2A* | rs3731257 | GEOCS | 173 | 131 | 18 | **322** | 230 | 160 | 32 | **422** |
| Cell cycle | *CDKN2A* | rs3731257 | MALOVA | 266 | 154 | 20 | **440** | 684 | 449 | 72 | **1205** |
| Cell cycle | *CDKN2A* | rs3731257 | SEARCH | 413 | 265 | 40 | **718** | 451 | 350 | 48 | **849** |
| Cell cycle | *CDKN2B* | rs3217986 | GEOCS | 259 | 56 | 8 | **323** | 335 | 83 | 9 | **427** |
| Cell cycle | *CDKN2B* | rs3217986 | MALOVA | 367 | 71 | 2 | **440** | 1009 | 197 | 7 | **1213** |
| Cell cycle | *CDKN2B* | rs3217986 | SEARCH | 568 | 140 | 7 | **715** | 696 | 145 | 5 | **846** |
| Cell cycle | *CDKN2B* | rs1063192 | GEOCS | 120 | 146 | 53 | **319** | 165 | 189 | 68 | **422** |
| Cell cycle | *CDKN2B* | rs1063192 | MALOVA | 119 | 212 | 110 | **441** | 351 | 591 | 274 | **1216** |
| Cell cycle | *CDKN2B* | rs1063192 | SEARCH | 221 | 376 | 119 | **716** | 245 | 440 | 168 | **853** |
| Cell cycle | *CDKN2B* | rs3218009 | GEOCS | 255 | 51 | 5 | **311** | 324 | 79 | 4 | **407** |
| Cell cycle | *CDKN2B* | rs3218009 | MALOVA | 317 | 111 | 13 | **441** | 859 | 339 | 21 | **1219** |
| Cell cycle | *CDKN2B* | rs3218009 | SEARCH | 544 | 161 | 12 | **717** | 648 | 187 | 17 | **852** |
| Cell cycle | *CDKN2B* | rs3218012 | GEOCS | 89 | 156 | 75 | **320** | 114 | 198 | 110 | **422** |
| Cell cycle | *CDKN2B* | rs3218012 | MALOVA | 170 | 199 | 67 | **436** | 418 | 586 | 203 | **1207** |
| Cell cycle | *CDKN2B* | rs3218012 | SEARCH | 208 | 358 | 142 | **708** | 267 | 426 | 155 | **848** |
| Cell cycle | *CDKN2C* | Rs12855 | GEOCS | 254 | 64 | 4 | **322** | 354 | 72 | 2 | **428** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *CDKN2C* | Rs12855 | MALOVA | 353 | 84 | 5 | **442** | 967 | 234 | 16 | **1217** |
| Cell cycle | *CDKN2C* | Rs12855 | SEARCH | 597 | 114 | 7 | **718** | 697 | 152 | 5 | **854** |
| Cell cycle | *CDKN2C* | rs3176459 | GEOCS | 128 | 148 | 38 | **314** | 188 | 184 | 44 | **416** |
| Cell cycle | *CDKN2C* | rs3176459 | MALOVA | 201 | 194 | 42 | **437** | 590 | 489 | 134 | **1213** |
| Cell cycle | *CDKN2C* | rs3176459 | SEARCH | 321 | 312 | 82 | **715** | 362 | 375 | 114 | **851** |
| Cell cycle | *CDKN2D* | rs1465702 | GEOCS | 294 | 27 | 1 | **322** | 375 | 47 | 0 | **422** |
| Cell cycle | *CDKN2D* | rs1465702 | MALOVA | 406 | 36 | 0 | **442** | 1118 | 83 | 5 | **1206** |
| Cell cycle | *CDKN2D* | rs1465702 | SEARCH | 646 | 72 | 1 | **719** | 786 | 66 | 1 | **853** |
| Cell cycle | *CDKN2D* | rs3218222 | GEOCS | 185 | 113 | 23 | **321** | 207 | 174 | 34 | **415** |
| Cell cycle | *CDKN2D* | rs3218222 | MALOVA | 273 | 149 | 19 | **441** | 720 | 424 | 63 | **1207** |
| Cell cycle | *CDKN2D* | rs3218222 | SEARCH | 391 | 268 | 47 | **706** | 502 | 295 | 40 | **837** |
| Cell cycle | *RB1* | rs1981434 | GEOCS | 163 | 133 | 21 | **317** | 203 | 181 | 30 | **414** |
| Cell cycle | *RB1* | rs1981434 | MALOVA | 210 | 183 | 40 | **433** | 586 | 494 | 115 | **1195** |
| Cell cycle | *RB1* | rs1981434 | SEARCH | 370 | 282 | 61 | **713** | 438 | 349 | 64 | **851** |
| Cell cycle | *RB1* | rs2854345 | GEOCS | 206 | 104 | 11 | **321** | 280 | 135 | 8 | **423** |
| Cell cycle | *RB1* | rs2854345 | MALOVA | 288 | 133 | 19 | **440** | 757 | 396 | 61 | **1214** |
| Cell cycle | *RB1* | rs2854345 | SEARCH | 470 | 217 | 28 | **715** | 575 | 235 | 33 | **843** |
| Cell cycle | *RB1* | rs399413 | GEOCS | 169 | 130 | 23 | **322** | 210 | 188 | 29 | **427** |
| Cell cycle | *RB1* | rs399413 | MALOVA | 123 | 124 | 26 | **273** | 367 | 282 | 62 | **711** |
| Cell cycle | *RB1* | rs399413 | SEARCH | 380 | 280 | 59 | **719** | 452 | 340 | 59 | **851** |
| Cell cycle | *RB1* | rs4151540 | GEOCS | 173 | 124 | 21 | **318** | 216 | 179 | 24 | **419** |
| Cell cycle | *RB1* | rs4151540 | MALOVA | 228 | 164 | 42 | **434** | 572 | 502 | 114 | **1188** |
| Cell cycle | *RB1* | rs4151540 | SEARCH | 379 | 287 | 51 | **717** | 453 | 341 | 54 | **848** |
| Cell cycle | *RB1* | rs4151551 | GEOCS | 265 | 52 | 4 | **321** | 365 | 56 | 2 | **423** |
| Cell cycle | *RB1* | rs4151551 | MALOVA | 357 | 74 | 3 | **434** | 998 | 202 | 12 | **1212** |
| Cell cycle | *RB1* | rs4151551 | SEARCH | 590 | 120 | 7 | **717** | 707 | 138 | 7 | **852** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *RB1* | rs2854344 | GEOCS | 282 | 25 | 3 | **310** | 360 | 52 | 5 | **417** |
| Cell cycle | *RB1* | rs2854344 | MALOVA | 240 | 39 | 4 | **283** | 608 | 107 | 3 | **718** |
| Cell cycle | *RB1* | rs2854344 | SEARCH | 633 | 73 | 5 | **711** | 717 | 121 | 2 | **840** |
| Cell cycle | *RB1* | rs425834 | GEOCS | 299 | 20 | 2 | **321** | 380 | 42 | 0 | **422** |
| Cell cycle | *RB1* | rs425834 | MALOVA | 404 | 31 | 1 | **436** | 1136 | 59 | 3 | **1198** |
| Cell cycle | *RB1* | rs425834 | SEARCH | 668 | 44 | 3 | **715** | 800 | 48 | 1 | **849** |
| Cell cycle | *RB1* | rs4151611 | GEOCS | 296 | 23 | 2 | **321** | 375 | 46 | 2 | **423** |
| Cell cycle | *RB1* | rs4151611 | MALOVA | 385 | 42 | 3 | **430** | 1089 | 114 | 2 | **1205** |
| Cell cycle | *RB1* | rs4151611 | SEARCH | 657 | 55 | 0 | **712** | 770 | 74 | 4 | **848** |
| Cell cycle | *RB1* | rs4151620 | GEOCS | 231 | 79 | 0 | **310** | 311 | 83 | 5 | **399** |
| Cell cycle | *RB1* | rs4151620 | MALOVA | 215 | 62 | 2 | **279** | 547 | 157 | 11 | **715** |
| Cell cycle | *RB1* | rs4151620 | SEARCH | 527 | 182 | 2 | **711** | 635 | 196 | 13 | **844** |
| Cell cycle | *RB1* | rs3092904 | GEOCS | 186 | 114 | 20 | **320** | 229 | 171 | 20 | **420** |
| Cell cycle | *RB1* | rs3092904 | MALOVA | 140 | 113 | 27 | **280** | 373 | 285 | 64 | **722** |
| Cell cycle | *RB1* | rs3092904 | SEARCH | 399 | 273 | 45 | **717** | 472 | 327 | 52 | **851** |
| Cell cycle | *RB1* | rs4151636 | GEOCS | 294 | 20 | 1 | **315** | 372 | 47 | 2 | **421** |
| Cell cycle | *RB1* | rs4151636 | MALOVA | 399 | 37 | 0 | **436** | 1109 | 107 | 1 | **1217** |
| Cell cycle | *RB1* | rs4151636 | SEARCH | 650 | 64 | 2 | **716** | 772 | 77 | 2 | **851** |
| Cell cycle | *STK15* | rs732417 | GEOCS | 271 | 43 | 1 | **315** | 351 | 64 | 4 | **419** |
| Cell cycle | *STK15* | rs732417 | MALOVA | 315 | 52 | 3 | **370** | 770 | 108 | 2 | **880** |
| Cell cycle | *STK15* | rs732417 | SEARCH | 589 | 117 | 6 | **712** | 717 | 118 | 5 | **840** |
| Cell cycle | *STK15* | rs1047972 | GEOCS | 216 | 94 | 9 | **319** | 286 | 127 | 14 | **427** |
| Cell cycle | *STK15* | rs1047972 | MALOVA | 298 | 99 | 11 | **408** | 735 | 260 | 21 | **1016** |
| Cell cycle | *STK15* | rs1047972 | SEARCH | 485 | 207 | 20 | **712** | 566 | 246 | 31 | **843** |
| Cell cycle | *STK15* | rs2273535 | GEOCS | 185 | 99 | 20 | **304** | 271 | 112 | 15 | **398** |
| Cell cycle | *STK15* | rs2273535 | MALOVA | 167 | 126 | 22 | **315** | 391 | 239 | 46 | **676** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Cell cycle | *STK15* | rs2273535 | SEARCH | 436 | 250 | 28 | **714** | 524 | 285 | 34 | **843** |
| Cell cycle | *STK15* | rs8173 | GEOCS | 161 | 116 | 18 | **295** | 229 | 155 | 22 | **406** |
| Cell cycle | *STK15* | rs8173 | MALOVA | 209 | 128 | 24 | **361** | 509 | 339 | 37 | **885** |
| Cell cycle | *STK15* | rs8173 | SEARCH | 373 | 271 | 38 | **682** | 458 | 314 | 40 | **812** |
| DNA repair | *BRCA1* | rs799917 | GEOCS | 119 | 151 | 36 | **306** | 178 | 177 | 44 | **399** |
| DNA repair | *BRCA1* | rs799917 | MALOVA | 173 | 176 | 50 | **399** | 436 | 393 | 99 | **928** |
| DNA repair | *BRCA1* | rs799917 | SEARCH | 316 | 325 | 68 | **709** | 388 | 350 | 92 | **830** |
| DNA repair | *BRCA1* | rs1799950 | GEOCS | 276 | 41 | 1 | **318** | 376 | 50 | 0 | **426** |
| DNA repair | *BRCA1* | rs1799950 | SEARCH | 633 | 85 | 0 | **718** | 745 | 100 | 4 | **849** |
| DNA repair | *BRIP* | rs11871785 | GEOCS | 131 | 149 | 39 | **319** | 165 | 191 | 67 | **356** |
| DNA repair | *BRIP* | rs11871785 | MALOVA | 195 | 203 | 46 | **444** | 535 | 538 | 144 | **1073** |
| DNA repair | *BRIP* | rs11871785 | SEARCH | 274 | 343 | 98 | **715** | 346 | 396 | 106 | **742** |
| DNA repair | *BRIP* | rs1557720 | GEOCS | 120 | 139 | 63 | **322** | 175 | 193 | 56 | **368** |
| DNA repair | *BRIP* | rs1557720 | MALOVA | 89 | 135 | 52 | **276** | 245 | 347 | 118 | **592** |
| DNA repair | *BRIP* | rs1557720 | SEARCH | 259 | 352 | 108 | **719** | 315 | 403 | 135 | **718** |
| DNA repair | *BRIP* | rs11652980 | GEOCS | 292 | 25 | 2 | **319** | 374 | 49 | 0 | **423** |
| DNA repair | *BRIP* | rs11652980 | MALOVA | 404 | 38 | 1 | **443** | 1120 | 98 | 1 | **1218** |
| DNA repair | *BRIP* | rs11652980 | SEARCH | 637 | 76 | 1 | **714** | 773 | 76 | 4 | **849** |
| DNA repair | *BRIP* | rs2191249 | GEOCS | 191 | 108 | 20 | **319** | 239 | 155 | 33 | **394** |
| DNA repair | *BRIP* | rs2191249 | MALOVA | 158 | 98 | 19 | **275** | 398 | 251 | 51 | **649** |
| DNA repair | *BRIP* | rs2191249 | SEARCH | 413 | 269 | 37 | **719** | 463 | 318 | 62 | **781** |
| DNA repair | *BRIP* | rs16945628 | GEOCS | 147 | 140 | 34 | **321** | 191 | 177 | 55 | **368** |
| DNA repair | *BRIP* | rs16945628 | MALOVA | 195 | 186 | 60 | **441** | 534 | 546 | 131 | **1080** |
| DNA repair | *BRIP* | rs16945628 | SEARCH | 348 | 302 | 69 | **719** | 377 | 372 | 105 | **749** |
| DNA repair | *BRIP* | rs2191248 | GEOCS | 138 | 141 | 41 | **320** | 202 | 174 | 45 | **376** |
| DNA repair | *BRIP* | rs2191248 | MALOVA | 183 | 205 | 47 | **435** | 453 | 602 | 155 | **1055** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| DNA repair | *BRIP* | rs2191248 | SEARCH | 295 | 317 | 102 | **714** | 381 | 392 | 78 | **773** |
| DNA repair | *BRIP* | rs16945643 | GEOCS | 266 | 47 | 6 | **319** | 357 | 62 | 4 | **419** |
| DNA repair | *BRIP* | rs16945643 | MALOVA | 360 | 77 | 3 | **440** | 1019 | 189 | 4 | **1208** |
| DNA repair | *BRIP* | rs16945643 | SEARCH | 617 | 93 | 2 | **712** | 716 | 133 | 2 | **849** |
| DNA repair | *BRIP* | rs6504074 | GEOCS | 170 | 115 | 37 | **322** | 225 | 156 | 40 | **381** |
| DNA repair | *BRIP* | rs6504074 | MALOVA | 137 | 103 | 19 | **259** | 340 | 258 | 52 | **598** |
| DNA repair | *BRIP* | rs6504074 | SEARCH | 412 | 264 | 37 | **713** | 457 | 325 | 65 | **782** |
| DNA repair | *BRIP* | rs2378908 | GEOCS | 241 | 70 | 7 | **318** | 318 | 91 | 10 | **409** |
| DNA repair | *BRIP* | rs2378908 | MALOVA | 198 | 73 | 3 | **274** | 564 | 137 | 10 | **701** |
| DNA repair | *BRIP* | rs2378908 | SEARCH | 509 | 170 | 16 | **695** | 625 | 214 | 13 | **839** |
| DNA repair | *BRIP* | rs4988344 | GEOCS | 227 | 85 | 9 | **321** | 309 | 103 | 15 | **412** |
| DNA repair | *BRIP* | rs4988344 | MALOVA | 188 | 81 | 8 | **277** | 526 | 174 | 12 | **700** |
| DNA repair | *BRIP* | rs4988344 | SEARCH | 491 | 198 | 28 | **717** | 592 | 241 | 20 | **833** |
| DNA repair | *BRIP* | rs9908659 | GEOCS | 131 | 146 | 43 | **320** | 191 | 172 | 59 | **363** |
| DNA repair | *BRIP* | rs9908659 | MALOVA | 169 | 203 | 67 | **439** | 405 | 615 | 192 | **1020** |
| DNA repair | *BRIP* | rs9908659 | SEARCH | 269 | 333 | 117 | **719** | 341 | 401 | 112 | **742** |
| DNA repair | *BRIP* | rs4968451 | GEOCS | 225 | 87 | 9 | **321** | 306 | 103 | 14 | **409** |
| DNA repair | *BRIP* | rs4968451 | MALOVA | 306 | 121 | 13 | **440** | 877 | 310 | 25 | **1187** |
| DNA repair | *BRIP* | rs4968451 | SEARCH | 489 | 201 | 26 | **716** | 588 | 247 | 19 | **835** |
| DNA repair | *BRIP* | rs2048718 | GEOCS | 91 | 155 | 74 | **320** | 131 | 195 | 95 | **326** |
| DNA repair | *BRIP* | rs2048718 | MALOVA | 118 | 215 | 96 | **429** | 352 | 632 | 225 | **984** |
| DNA repair | *BRIP* | rs2048718 | SEARCH | 234 | 333 | 143 | **710** | 247 | 428 | 177 | **675** |
| DNA repair | *KU70* | rs132788 | GEOCS | 144 | 140 | 30 | **314** | 211 | 173 | 37 | **421** |
| DNA repair | *KU70* | rs132788 | MALOVA | 153 | 169 | 33 | **355** | 356 | 375 | 91 | **822** |
| DNA repair | *KU70* | rs132788 | SEARCH | 306 | 302 | 105 | **713** | 364 | 369 | 106 | **839** |
| DNA repair | *NBS1* | rs1063045 | GEOCS | 145 | 137 | 39 | **321** | 169 | 201 | 54 | **424** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| DNA repair | *NBS1* | rs1063045 | MALOVA | 197 | 185 | 42 | **424** | 479 | 454 | 108 | **1041** |
| DNA repair | *NBS1* | rs1063045 | SEARCH | 295 | 334 | 79 | **708** | 365 | 373 | 102 | **840** |
| DNA repair | *NBS1* | rs1805794 | GEOCS | 140 | 131 | 33 | **304** | 163 | 174 | 52 | **389** |
| DNA repair | *NBS1* | rs1805794 | MALOVA | 196 | 176 | 39 | **411** | 484 | 437 | 109 | **1030** |
| DNA repair | *NBS1* | rs1805794 | SEARCH | 303 | 327 | 78 | **708** | 369 | 372 | 107 | **848** |
| DNA repair | *NBS1* | rs709816 | GEOCS | 128 | 123 | 54 | **305** | 157 | 165 | 71 | **393** |
| DNA repair | *NBS1* | rs709816 | MALOVA | 173 | 199 | 53 | **425** | 429 | 470 | 132 | **1031** |
| DNA repair | *NBS1* | rs709816 | SEARCH | 278 | 342 | 98 | **718** | 336 | 388 | 125 | **849** |
| DNA repair | *NBS1* | rs1061302 | GEOCS | 137 | 131 | 34 | **302** | 159 | 178 | 50 | **387** |
| DNA repair | *NBS1* | rs1061302 | MALOVA | 156 | 144 | 32 | **332** | 390 | 346 | 82 | **818** |
| DNA repair | *NBS1* | rs1061302 | SEARCH | 302 | 325 | 72 | **699** | 367 | 364 | 99 | **830** |
| DNA repair | *RAD51* | rs1801320 | GEOCS | 266 | 52 | 4 | **322** | 363 | 61 | 1 | **425** |
| DNA repair | *RAD51* | rs1801320 | MALOVA | 315 | 44 | 2 | **361** | 725 | 88 | 7 | **820** |
| DNA repair | *RAD51* | rs1801320 | SEARCH | 629 | 84 | 3 | **716** | 745 | 100 | 2 | **847** |
| DNA repair | *RAD51* | rs1801321 | GEOCS | 117 | 144 | 56 | **317** | 151 | 193 | 74 | **418** |
| DNA repair | *RAD51* | rs1801321 | MALOVA | 139 | 166 | 73 | **378** | 279 | 330 | 112 | **721** |
| DNA repair | *RAD51* | rs1801321 | SEARCH | 216 | 359 | 142 | **717** | 273 | 433 | 141 | **847** |
| DNA repair | *RAD52* | rs11226 | GEOCS | 75 | 169 | 75 | **319** | 122 | 210 | 89 | **421** |
| DNA repair | *RAD52* | rs11226 | MALOVA | 119 | 203 | 97 | **419** | 311 | 489 | 232 | **1032** |
| DNA repair | *RAD52* | rs11226 | SEARCH | 217 | 358 | 136 | **711** | 269 | 374 | 197 | **840** |
| DNA repair | *XRCC2* | UNKNOWN | GEOCS | 286 | 29 | 0 | **315** | 369 | 45 | 4 | **418** |
| DNA repair | *XRCC2* | UNKNOWN | MALOVA | 303 | 28 | 2 | **333** | 627 | 66 | 1 | **694** |
| DNA repair | *XRCC2* | UNKNOWN | SEARCH | 634 | 78 | 4 | **716** | 755 | 89 | 3 | **847** |
| DNA repair | *XRCC2* | rs3218384* | GEOCS | 204 | 103 | 12 | **319** | 277 | 134 | 14 | **425** |
| DNA repair | *XRCC2* | rs3218384* | MALOVA | 237 | 134 | 21 | **392** | 522 | 316 | 49 | **887** |
| DNA repair | *XRCC2* | rs3218384* | SEARCH | 426 | 258 | 31 | **715** | 509 | 301 | 38 | **848** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| DNA repair | *XRCC2* | rs3218536 | GEOCS | 257 | 53 | 1 | **311** | 334 | 70 | 5 | **409** |
| DNA repair | *XRCC2* | rs3218536 | MALOVA | 275 | 35 | 0 | **310** | 467 | 67 | 2 | **536** |
| DNA repair | *XRCC2* | rs3218536 | SEARCH | 620 | 94 | 2 | **716** | 704 | 129 | 9 | **842** |
| DNA repair | *XRCC3* | rs1799794 | GEOCS | 201 | 111 | 9 | **321** | 269 | 136 | 18 | **423** |
| DNA repair | *XRCC3* | rs1799794 | MALOVA | 285 | 131 | 8 | **424** | 684 | 316 | 42 | **1042** |
| DNA repair | *XRCC3* | rs1799794 | SEARCH | 454 | 242 | 20 | **716** | 552 | 261 | 29 | **842** |
| DNA repair | *XRCC3* | rs1799796 | GEOCS | 156 | 130 | 35 | **321** | 195 | 185 | 41 | **421** |
| DNA repair | *XRCC3* | rs1799796 | MALOVA | 200 | 164 | 60 | **424** | 459 | 440 | 127 | **1026** |
| DNA repair | *XRCC3* | rs1799796 | SEARCH | 320 | 314 | 82 | **716** | 386 | 381 | 85 | **852** |
| DNA repair | *XRCC3* | rs861539 | GEOCS | 123 | 112 | 31 | **266** | 131 | 177 | 40 | **348** |
| DNA repair | *XRCC3* | rs861539 | MALOVA | 138 | 166 | 49 | **353** | 335 | 377 | 134 | **846** |
| DNA repair | *XRCC3* | rs861539 | SEARCH | 284 | 334 | 95 | **713** | 318 | 404 | 108 | **830** |
| Mismatch repair | *MLH1* | rs1800734 | GEOCS | 178 | 114 | 21 | **313** | 246 | 145 | 28 | **419** |
| Mismatch repair | *MLH1* | rs1800734 | MALOVA | 182 | 77 | 9 | **268** | 449 | 211 | 27 | **687** |
| Mismatch repair | *MLH1* | rs1800734 | SEARCH | 457 | 225 | 36 | **718** | 532 | 285 | 34 | **851** |
| Mismatch repair | *MLH1* | rs1540354 | GEOCS | 213 | 94 | 12 | **319** | 286 | 124 | 12 | **422** |
| Mismatch repair | *MLH1* | rs1540354 | MALOVA | 172 | 91 | 12 | **275** | 444 | 222 | 30 | **696** |
| Mismatch repair | *MLH1* | rs1540354 | SEARCH | 469 | 215 | 27 | **711** | 566 | 245 | 31 | **842** |
| Mismatch repair | *MLH1* | rs1799977 | GEOCS | 167 | 114 | 30 | **311** | 209 | 147 | 47 | **403** |
| Mismatch repair | *MLH1* | rs1799977 | SEARCH | 340 | 300 | 66 | **706** | 424 | 333 | 78 | **835** |
| Mismatch repair | *MLH1* | rs2286939 | GEOCS | 107 | 149 | 62 | **318** | 127 | 205 | 86 | **418** |
| Mismatch repair | *MLH1* | rs2286939 | MALOVA | 82 | 139 | 60 | **281** | 260 | 359 | 160 | **779** |
| Mismatch repair | *MLH1* | rs2286939 | SEARCH | 208 | 369 | 142 | **719** | 251 | 421 | 182 | **854** |
| Mismatch repair | *MLH3* | rs7303 | GEOCS | 74 | 132 | 102 | **308** | 101 | 198 | 103 | **402** |
| Mismatch repair | *MLH3* | rs7303 | MALOVA | 87 | 119 | 58 | **264** | 189 | 322 | 154 | **665** |
| Mismatch repair | *MLH3* | rs7303 | SEARCH | 179 | 366 | 164 | **709** | 241 | 399 | 199 | **839** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Mismatch repair | *MLH3* | rs175080 | GEOCS | 120 | 132 | 58 | **310** | 129 | 202 | 81 | **412** |
| Mismatch repair | *MLH3* | rs175080 | MALOVA | 78 | 127 | 68 | **273** | 214 | 345 | 166 | **725** |
| Mismatch repair | *MLH3* | rs175080 | SEARCH | 206 | 373 | 134 | **713** | 249 | 406 | 189 | **844** |
| Mismatch repair | *MSH2* | rs4952887 | GEOCS | 262 | 48 | 5 | **315** | 347 | 72 | 3 | **422** |
| Mismatch repair | *MSH2* | rs4952887 | MALOVA | 227 | 43 | 3 | **273** | 574 | 118 | 7 | **699** |
| Mismatch repair | *MSH2* | rs4952887 | SEARCH | 611 | 95 | 8 | **714** | 699 | 146 | 9 | **854** |
| Mismatch repair | *MSH2* | rs13425206 | GEOCS | 284 | 35 | 2 | **321** | 394 | 30 | 1 | **425** |
| Mismatch repair | *MSH2* | rs13425206 | MALOVA | 417 | 21 | 1 | **439** | 1133 | 82 | 2 | **1217** |
| Mismatch repair | *MSH2* | rs13425206 | SEARCH | 667 | 46 | 2 | **715** | 781 | 64 | 4 | **849** |
| Mismatch repair | *MSH2* | rs3771274 | GEOCS | 121 | 143 | 50 | **314** | 160 | 194 | 70 | **424** |
| Mismatch repair | *MSH2* | rs3771274 | MALOVA | 85 | 140 | 46 | **271** | 244 | 349 | 108 | **701** |
| Mismatch repair | *MSH2* | rs3771274 | SEARCH | 261 | 348 | 110 | **719** | 306 | 411 | 136 | **853** |
| Mismatch repair | *MSH2* | rs1981928 | GEOCS | 177 | 120 | 24 | **321** | 224 | 159 | 42 | **425** |
| Mismatch repair | *MSH2* | rs1981928 | MALOVA | 138 | 115 | 22 | **275** | 384 | 281 | 60 | **725** |
| Mismatch repair | *MSH2* | rs1981928 | SEARCH | 393 | 274 | 51 | **718** | 455 | 330 | 66 | **851** |
| Mismatch repair | *MSH2* | rs2059520 | GEOCS | 151 | 134 | 34 | **319** | 186 | 181 | 58 | **425** |
| Mismatch repair | *MSH2* | rs2059520 | MALOVA | 110 | 125 | 40 | **275** | 311 | 314 | 85 | **710** |
| Mismatch repair | *MSH2* | rs2059520 | SEARCH | 304 | 326 | 88 | **718** | 364 | 377 | 105 | **846** |
| Mismatch repair | *MSH2* | rs2303428 | GEOCS | 266 | 48 | 1 | **315** | 330 | 72 | 4 | **406** |
| Mismatch repair | *MSH2* | rs2303428 | MALOVA | 231 | 46 | 6 | **283** | 626 | 154 | 7 | **787** |
| Mismatch repair | *MSH2* | rs2303428 | SEARCH | 592 | 124 | 3 | **719** | 692 | 147 | 15 | **854** |
| Mismatch repair | *MSH3* | rs6151662 | GEOCS | 290 | 30 | 1 | **321** | 383 | 43 | 0 | **426** |
| Mismatch repair | *MSH3* | rs6151662 | MALOVA | 391 | 43 | 1 | **435** | 1057 | 122 | 4 | **1183** |
| Mismatch repair | *MSH3* | rs6151662 | SEARCH | 630 | 87 | 1 | **718** | 747 | 93 | 12 | **852** |
| Mismatch repair | *MSH3* | rs40139 | GEOCS | 104 | 160 | 56 | **320** | 118 | 197 | 102 | **417** |
| Mismatch repair | *MSH3* | rs40139 | MALOVA | 95 | 136 | 46 | **277** | 270 | 343 | 143 | **756** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Mismatch repair | *MSH3* | rs40139 | SEARCH | 213 | 369 | 136 | **718** | 301 | 392 | 160 | **853** |
| Mismatch repair | *MSH3* | rs26282 | GEOCS | 167 | 121 | 27 | **315** | 237 | 141 | 28 | **406** |
| Mismatch repair | *MSH3* | rs26282 | MALOVA | 149 | 109 | 18 | **276** | 386 | 296 | 49 | **731** |
| Mismatch repair | *MSH3* | rs26282 | SEARCH | 380 | 284 | 53 | **717** | 444 | 338 | 69 | **851** |
| Mismatch repair | *MSH3* | rs26779 | GEOCS | 121 | 154 | 47 | **322** | 175 | 171 | 74 | **420** |
| Mismatch repair | *MSH3* | rs26779 | MALOVA | 161 | 209 | 69 | **439** | 442 | 569 | 196 | **1207** |
| Mismatch repair | *MSH3* | rs26779 | SEARCH | 259 | 343 | 112 | **714** | 287 | 429 | 123 | **839** |
| Mismatch repair | *MSH3* | rs33008 | GEOCS | 171 | 129 | 18 | **318** | 224 | 166 | 36 | **426** |
| Mismatch repair | *MSH3* | rs33008 | MALOVA | 145 | 112 | 20 | **277** | 365 | 285 | 59 | **709** |
| Mismatch repair | *MSH3* | rs33008 | SEARCH | 360 | 300 | 56 | **716** | 463 | 316 | 69 | **848** |
| Mismatch repair | *MSH3* | rs10079641 | GEOCS | 271 | 48 | 3 | **322** | 350 | 64 | 8 | **422** |
| Mismatch repair | *MSH3* | rs10079641 | MALOVA | 365 | 63 | 6 | **434** | 1009 | 180 | 8 | **1197** |
| Mismatch repair | *MSH3* | rs10079641 | SEARCH | 579 | 133 | 5 | **717** | 708 | 127 | 16 | **851** |
| Mismatch repair | *MSH3* | rs184967 | GEOCS | 219 | 93 | 6 | **318** | 319 | 91 | 10 | **420** |
| Mismatch repair | *MSH3* | rs184967 | MALOVA | 204 | 66 | 7 | **277** | 514 | 186 | 17 | **717** |
| Mismatch repair | *MSH3* | rs184967 | SEARCH | 505 | 186 | 19 | **710** | 585 | 241 | 20 | **846** |
| Mismatch repair | *MSH3* | rs2897298 | GEOCS | 232 | 82 | 7 | **321** | 328 | 84 | 9 | **421** |
| Mismatch repair | *MSH3* | rs2897298 | MALOVA | 342 | 85 | 2 | **429** | 945 | 227 | 7 | **1179** |
| Mismatch repair | *MSH3* | rs2897298 | SEARCH | 541 | 162 | 12 | **715** | 646 | 186 | 18 | **850** |
| Mismatch repair | *MSH3* | rs26279 | GEOCS | 142 | 151 | 28 | **321** | 242 | 151 | 34 | **427** |
| Mismatch repair | *MSH3* | rs26279 | MALOVA | 151 | 104 | 23 | **278** | 403 | 289 | 54 | **746** |
| Mismatch repair | *MSH3* | rs26279 | SEARCH | 364 | 268 | 70 | **702** | 417 | 336 | 82 | **835** |
| Mismatch repair | *MSH3* | rs2112416 | GEOCS | 225 | 82 | 9 | **316** | 289 | 122 | 12 | **423** |
| Mismatch repair | *MSH3* | rs2112416 | MALOVA | 195 | 75 | 8 | **278** | 525 | 169 | 15 | **709** |
| Mismatch repair | *MSH3* | rs2112416 | SEARCH | 550 | 151 | 15 | **716** | 649 | 188 | 16 | **853** |
| Mismatch repair | *MSH6* | rs3136245 | GEOCS | 209 | 97 | 16 | **322** | 266 | 140 | 21 | **427** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Mismatch repair | *MSH6* | rs3136245 | MALOVA | 181 | 84 | 5 | **270** | 435 | 211 | 37 | **683** |
| Mismatch repair | *MSH6* | rs3136245 | SEARCH | 454 | 241 | 23 | **718** | 538 | 275 | 37 | **850** |
| Mismatch repair | *MSH6* | rs3136272 | GEOCS | 136 | 152 | 34 | **322** | 196 | 184 | 43 | **423** |
| Mismatch repair | *MSH6* | rs3136272 | MALOVA | 171 | 177 | 65 | **413** | 445 | 557 | 156 | **1158** |
| Mismatch repair | *MSH6* | rs3136272 | SEARCH | 303 | 319 | 93 | **715** | 334 | 398 | 112 | **844** |
| Mismatch repair | *MSH6* | rs1800932 | GEOCS | 206 | 99 | 16 | **321** | 291 | 122 | 13 | **426** |
| Mismatch repair | *MSH6* | rs1800932 | MALOVA | 187 | 82 | 8 | **277** | 506 | 191 | 26 | **723** |
| Mismatch repair | *MSH6* | rs1800932 | SEARCH | 471 | 220 | 28 | **719** | 560 | 253 | 39 | **852** |
| Mismatch repair | *MSH6* | rs2348244 | GEOCS | 227 | 67 | 9 | **303** | 287 | 91 | 13 | **391** |
| Mismatch repair | *MSH6* | rs2348244 | MALOVA | 201 | 65 | 2 | **268** | 523 | 164 | 15 | **702** |
| Mismatch repair | *MSH6* | rs2348244 | SEARCH | 501 | 193 | 14 | **708** | 614 | 206 | 27 | **847** |
| Mismatch repair | *MSH6* | rs3136317 | GEOCS | 204 | 104 | 10 | **318** | 294 | 114 | 15 | **423** |
| Mismatch repair | *MSH6* | rs3136317 | MALOVA | 308 | 128 | 5 | **441** | 849 | 360 | 9 | **1218** |
| Mismatch repair | *MSH6* | rs3136317 | SEARCH | 475 | 220 | 22 | **717** | 581 | 243 | 25 | **849** |
| Mismatch repair | *MSH6* | rs1800935 | GEOCS | 163 | 131 | 27 | **321** | 228 | 171 | 27 | **426** |
| Mismatch repair | *MSH6* | rs1800935 | MALOVA | 135 | 119 | 23 | **277** | 384 | 291 | 72 | **747** |
| Mismatch repair | *MSH6* | rs1800935 | SEARCH | 366 | 289 | 63 | **718** | 427 | 350 | 76 | **853** |
| Mismatch repair | *MSH6* | rs2020911 | GEOCS | 111 | 154 | 54 | **319** | 158 | 195 | 71 | **424** |
| Mismatch repair | *MSH6* | rs2020911 | MALOVA | 118 | 120 | 34 | **272** | 271 | 310 | 97 | **678** |
| Mismatch repair | *MSH6* | rs2020911 | SEARCH | 259 | 355 | 101 | **715** | 315 | 392 | 141 | **848** |
| Mismatch repair | *PMS1* | rs3762545 | GEOCS | 194 | 109 | 14 | **317** | 257 | 146 | 19 | **422** |
| Mismatch repair | *PMS1* | rs3762545 | MALOVA | 167 | 98 | 8 | **273** | 474 | 237 | 34 | **745** |
| Mismatch repair | *PMS1* | rs3762545 | SEARCH | 488 | 206 | 24 | **718** | 540 | 278 | 33 | **851** |
| Mismatch repair | *PMS1* | rs5742981 | GEOCS | 299 | 23 | 0 | **322** | 387 | 32 | 2 | **421** |
| Mismatch repair | *PMS1* | rs5742981 | MALOVA | 233 | 28 | 1 | **262** | 924 | 101 | 3 | **1028** |
| Mismatch repair | *PMS1* | rs5742981 | SEARCH | 676 | 40 | 0 | **716** | 792 | 57 | 1 | **850** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Mismatch repair | *PMS1* | rs5741593 | GEOCS | 291 | 31 | 0 | **322** | 373 | 46 | 1 | **420** |
| Mismatch repair | *PMS1* | rs5741593 | MALOVA | 378 | 58 | 2 | **438** | 1037 | 170 | 8 | **1215** |
| Mismatch repair | *PMS1* | rs5741593 | SEARCH | 625 | 75 | 2 | **702** | 705 | 104 | 3 | **812** |
| Mismatch repair | *PMS1* | rs1233291 | GEOCS | 169 | 132 | 20 | **321** | 220 | 174 | 31 | **425** |
| Mismatch repair | *PMS1* | rs1233291 | MALOVA | 137 | 114 | 22 | **273** | 352 | 294 | 58 | **704** |
| Mismatch repair | *PMS1* | rs1233291 | SEARCH | 414 | 260 | 41 | **715** | 443 | 348 | 57 | **848** |
| Mismatch repair | *PMS1* | rs1233255 | GEOCS | 194 | 111 | 13 | **318** | 270 | 131 | 23 | **424** |
| Mismatch repair | *PMS1* | rs1233255 | MALOVA | 149 | 81 | 7 | **237** | 413 | 193 | 31 | **637** |
| Mismatch repair | *PMS1* | rs1233255 | SEARCH | 465 | 223 | 28 | **716** | 530 | 277 | 37 | **844** |
| Mismatch repair | *PMS1* | rs1233258 | GEOCS | 167 | 130 | 21 | **318** | 217 | 164 | 40 | **421** |
| Mismatch repair | *PMS1* | rs1233258 | MALOVA | 137 | 113 | 28 | **278** | 330 | 320 | 57 | **707** |
| Mismatch repair | *PMS1* | rs1233258 | SEARCH | 394 | 260 | 55 | **709** | 447 | 328 | 66 | **841** |
| Mismatch repair | *PMS1* | rs256571 | GEOCS | 284 | 36 | 1 | **321** | 371 | 49 | 2 | **422** |
| Mismatch repair | *PMS1* | rs256571 | MALOVA | 378 | 60 | 0 | **438** | 1074 | 134 | 4 | **1212** |
| Mismatch repair | *PMS1* | rs256571 | SEARCH | 636 | 79 | 1 | **716** | 741 | 107 | 2 | **850** |
| Mismatch repair | *PMS1* | rs256563 | GEOCS | 250 | 65 | 5 | **320** | 318 | 96 | 7 | **421** |
| Mismatch repair | *PMS1* | rs256563 | MALOVA | 335 | 86 | 6 | **427** | 916 | 265 | 15 | **1196** |
| Mismatch repair | *PMS1* | rs256563 | SEARCH | 578 | 136 | 4 | **718** | 650 | 183 | 12 | **845** |
| Mismatch repair | *PMS2* | rs7797466 | GEOCS | 186 | 117 | 13 | **316** | 285 | 115 | 14 | **414** |
| Mismatch repair | *PMS2* | rs7797466 | MALOVA | 161 | 94 | 15 | **270** | 470 | 202 | 28 | **700** |
| Mismatch repair | *PMS2* | rs7797466 | SEARCH | 490 | 201 | 28 | **719** | 580 | 248 | 26 | **854** |
| Mismatch repair | *PMS2* | rs2345060 | GEOCS | 191 | 112 | 20 | **323** | 243 | 154 | 29 | **426** |
| Mismatch repair | *PMS2* | rs2345060 | MALOVA | 159 | 110 | 10 | **279** | 411 | 256 | 43 | **710** |
| Mismatch repair | *PMS2* | rs2345060 | SEARCH | 403 | 278 | 36 | **717** | 479 | 324 | 48 | **851** |
| Mismatch repair | *PMS2* | rs2286680 | GEOCS | 227 | 84 | 8 | **319** | 320 | 97 | 8 | **425** |
| Mismatch repair | *PMS2* | rs2286680 | MALOVA | 198 | 77 | 3 | **278** | 577 | 158 | 15 | **750** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| Mismatch repair | *PMS2* | rs2286680 | SEARCH | 546 | 157 | 16 | **719** | 632 | 205 | 14 | **851** |
| Mismatch repair | *PMS2* | rs12112229 | GEOCS | 173 | 117 | 20 | **310** | 231 | 151 | 24 | **406** |
| Mismatch repair | *PMS2* | rs12112229 | MALOVA | 152 | 105 | 15 | **272** | 408 | 246 | 46 | **700** |
| Mismatch repair | *PMS2* | rs12112229 | SEARCH | 399 | 263 | 55 | **717** | 442 | 351 | 46 | **839** |
| Mismatch repair | *PMS2* | rs1805321 | GEOCS | 117 | 143 | 62 | **322** | 142 | 205 | 79 | **426** |
| Mismatch repair | *PMS2* | rs1805321 | MALOVA | 100 | 126 | 51 | **277** | 230 | 346 | 127 | **703** |
| Mismatch repair | *PMS2* | rs1805321 | SEARCH | 249 | 347 | 122 | **718** | 270 | 436 | 145 | **851** |
| Mismatch repair | *PMS2* | rs2228006 | GEOCS | 231 | 84 | 6 | **321** | 308 | 116 | 3 | **427** |
| Mismatch repair | *PMS2* | rs2228006 | MALOVA | 324 | 103 | 7 | **434** | 898 | 288 | 20 | **1206** |
| Mismatch repair | *PMS2* | rs2228006 | SEARCH | 509 | 184 | 19 | **712** | 615 | 211 | 19 | **845** |
| OCAC | *ABL1* | rs2855192 | GEOCS | 232 | 77 | 13 | **322** | 333 | 85 | 5 | **423** |
| OCAC | *ABL1* | rs2855192 | MALOVA | 328 | 98 | 5 | **431** | 929 | 249 | 21 | **1199** |
| OCAC | *ABL1* | rs2855192 | SEARCH | 550 | 151 | 16 | **717** | 624 | 217 | 7 | **848** |
| OCAC | *BRCA2* | rs144848 | GEOCS | 173 | 128 | 20 | **321** | 217 | 189 | 21 | **427** |
| OCAC | *BRCA2* | rs144848 | MALOVA | 227 | 158 | 35 | **420** | 540 | 399 | 81 | **1020** |
| OCAC | *BRCA2* | rs144848 | SEARCH | 379 | 283 | 56 | **718** | 443 | 337 | 67 | **847** |
| OCAC | *CDC2* | rs2448343 | GEOCS | 122 | 152 | 40 | **314** | 169 | 203 | 46 | **418** |
| OCAC | *CDC2* | rs2448343 | MALOVA | 181 | 184 | 54 | **419** | 522 | 479 | 158 | **1159** |
| OCAC | *CDC2* | rs2448343 | SEARCH | 297 | 329 | 89 | **715** | 339 | 381 | 123 | **843** |
| OCAC | *CDK7* | rs12656449 | GEOCS | 282 | 36 | 2 | **320** | 356 | 61 | 4 | **421** |
| OCAC | *CDK7* | rs12656449 | MALOVA | 356 | 75 | 3 | **434** | 988 | 206 | 11 | **1205** |
| OCAC | *CDK7* | rs12656449 | SEARCH | 581 | 132 | 4 | **717** | 705 | 139 | 6 | **850** |
| OCAC | CHR8-P3 | rs7000448 | GEOCS | 121 | 163 | 36 | **320** | 164 | 199 | 57 | **420** |
| OCAC | CHR8-P3 | rs7000448 | MALOVA | 167 | 208 | 58 | **433** | 458 | 546 | 176 | **1180** |
| OCAC | CHR8-P3 | rs7000448 | SEARCH | 284 | 344 | 91 | **719** | 359 | 387 | 107 | **853** |
| OCAC | DESP-1979 | rs16901979 | GEOCS | 295 | 26 | 1 | **322** | 385 | 37 | 1 | **423** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| OCAC | DESP-1979 | rs16901979 | MALOVA | 401 | 29 | 0 | **430** | 1073 | 81 | 1 | **1155** |
| OCAC | DESP-1979 | rs16901979 | SEARCH | 666 | 44 | 1 | **711** | 798 | 51 | 1 | **850** |
| OCAC | *E2F2* | rs760607 | GEOCS | 103 | 156 | 60 | **319** | 146 | 189 | 66 | **401** |
| OCAC | *E2F2* | rs760607 | MALOVA | 191 | 181 | 49 | **421** | 489 | 520 | 150 | **1159** |
| OCAC | *E2F2* | rs760607 | SEARCH | 270 | 341 | 105 | **716** | 325 | 407 | 118 | **850** |
| OCAC | *E2F3* | rs7760528 | GEOCS | 152 | 132 | 36 | **320** | 185 | 185 | 50 | **420** |
| OCAC | *E2F3* | rs7760528 | MALOVA | 194 | 186 | 46 | **426** | 572 | 502 | 126 | **1200** |
| OCAC | *E2F3* | rs7760528 | SEARCH | 340 | 311 | 65 | **716** | 404 | 348 | 93 | **845** |
| OCAC | *KU70* | rs132788 | GEOCS | 144 | 140 | 30 | **314** | 211 | 173 | 37 | **421** |
| OCAC | *KU70* | rs132788 | MALOVA | 153 | 169 | 33 | **355** | 356 | 375 | 91 | **822** |
| OCAC | *KU70* | rs132788 | SEARCH | 306 | 302 | 105 | **713** | 364 | 369 | 106 | **839** |
| OCAC | *LIG4* | rs1805386 | GEOCS | 207 | 96 | 13 | **316** | 296 | 108 | 11 | **415** |
| OCAC | *LIG4* | rs1805386 | MALOVA | 321 | 97 | 13 | **431** | 839 | 330 | 36 | **1205** |
| OCAC | *LIG4* | rs1805386 | SEARCH | 478 | 215 | 22 | **715** | 615 | 210 | 21 | **846** |
| OCAC | *PGR* | rs10895068 | GEOCS | 294 | 27 | 0 | **321** | 382 | 41 | 0 | **423** |
| OCAC | *PGR* | rs10895068 | MALOVA | 372 | 51 | 1 | **424** | 997 | 136 | 7 | **1140** |
| OCAC | *PGR* | rs10895068 | SEARCH | 638 | 77 | 2 | **717** | 760 | 86 | 4 | **850** |
| OCAC | *PGR* | rs1042838 | GEOCS | 220 | 98 | 4 | **322** | 306 | 102 | 15 | **423** |
| OCAC | *PGR* | rs1042838 | MALOVA | 295 | 92 | 10 | **397** | 884 | 233 | 21 | **1138** |
| OCAC | *PGR* | rs1042838 | SEARCH | 490 | 187 | 28 | **705** | 629 | 191 | 27 | **847** |
| OCAC | *PGR* | rs608995 | GEOCS | 170 | 136 | 15 | **321** | 228 | 165 | 31 | **424** |
| OCAC | *PGR* | rs608995 | MALOVA | 270 | 142 | 25 | **437** | 754 | 390 | 58 | **1202** |
| OCAC | *PGR* | rs608995 | SEARCH | 418 | 249 | 50 | **717** | 521 | 280 | 46 | **847** |
| OCAC | *TGFB1* | rs1982073 | GEOCS | 125 | 143 | 46 | **314** | 166 | 189 | 58 | **413** |
| OCAC | *TGFB1* | rs1982073 | SEARCH | 281 | 329 | 102 | **712** | 326 | 393 | 130 | **849** |
| OCAC | *TNRC9* | rs3803662 | GEOCS | 152 | 142 | 27 | **321** | 187 | 189 | 44 | **420** |

| Pathway | Gene/cytoband | SNP | Study | Number of cases | | | | Number of controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | Aa | aa | Total | AA | Aa | aa | Total |
| OCAC | *TNRC9* | rs3803662 | MALOVA | 244 | 160 | 22 | **426** | 649 | 427 | 81 | **1157** |
| OCAC | *TNRC9* | rs3803662 | SEARCH | 401 | 258 | 50 | **709** | 464 | 329 | 57 | **850** |
| OCAC | *TP53* | rs1042522 | GEOCS | 171 | 128 | 19 | **318** | 230 | 171 | 21 | **422** |
| OCAC | *TP53* | rs1042522 | MALOVA | 231 | 154 | 38 | **423** | 569 | 375 | 79 | **1023** |
| OCAC | *TP53* | rs1042522 | SEARCH | 406 | 263 | 49 | **718** | 463 | 328 | 55 | **846** |
| OCAC | *TP53* | rs1625895 | GEOCS | 242 | 75 | 2 | **319** | 308 | 106 | 3 | **417** |
| OCAC | *TP53* | rs1625895 | MALOVA | 303 | 82 | 3 | **388** | 720 | 147 | 14 | **881** |
| OCAC | *TP53* | rs1625895 | SEARCH | 537 | 151 | 14 | **702** | 639 | 183 | 16 | **838** |
| OCAC | *TP53* | rs9894946 | GEOCS | 219 | 96 | 5 | **320** | 285 | 133 | 4 | **422** |
| OCAC | *TP53* | rs9894946 | MALOVA | 312 | 116 | 9 | **437** | 900 | 288 | 26 | **1214** |
| OCAC | *TP53* | rs9894946 | SEARCH | 514 | 179 | 20 | **713** | 629 | 194 | 26 | **849** |
| OCAC | *VDR* | rs10735810 | GEOCS | 118 | 136 | 65 | **319** | 158 | 200 | 63 | **421** |
| OCAC | *VDR* | rs10735810 | MALOVA | 159 | 208 | 57 | **424** | 475 | 545 | 163 | **1183** |
| OCAC | *VDR* | rs10735810 | SEARCH | 274 | 341 | 103 | **718** | 332 | 383 | 136 | **851** |
| OCAC | *XRCC5* | rs16855489 | GEOCS | 121 | 153 | 48 | **322** | 152 | 209 | 63 | **424** |
| OCAC | *XRCC5* | rs16855489 | MALOVA | 148 | 218 | 69 | **435** | 428 | 586 | 195 | **1209** |
| OCAC | *XRCC5* | rs16855489 | SEARCH | 277 | 347 | 95 | **719** | 344 | 394 | 113 | **851** |

AA – common homozygous; Aa – heterozygous; aa – rare homozygous; the distributions for oncogenes and MMCT-18 candidates are tabulated in Appendices II-A and II-B, respectively.

# Appendix VI: Genotype-specific risks of all SNPs analysed with AML method

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| BCAC§ | 2q22.1 | rs4954956 | 98 | 1446 | 2469 | 1.02 (0.89-1.18) | 0.94 (0.71-1.25) | 0.9634 | 0.9651 |
| BCAC§ | 5q11.2 | rs889312 | 98 | 1473 | 2460 | 0.98 (0.85-1.12) | 0.92 (0.71-1.18) | 0.5136 | 0.533 |
| BCAC§ | 8q24.21 | rs10505477 | 98 | 1456 | 2459 | 0.97 (0.83-1.14) | 0.86 (0.71-1.04) | 0.1166 | 0.1342 |
| BCAC§ | 8q24.21 | rs10808556 | 98 | 1462 | 2453 | 1.15 (0.99-1.33) | 1.20 (0.99-1.46) | **0.0446** | 0.0552 |
| BCAC§ | 8q24.21 | rs13254738 | 97 | 1458 | 2451 | 0.98 (0.85-1.13) | 0.91 (0.73-1.13) | 0.4202 | 0.4416 |
| BCAC§ | 8q24.21 | rs13281615 | 98 | 1458 | 2464 | 0.93 (0.81-1.08) | 0.98 (0.80-1.20) | 0.6699 | 0.6841 |
| BCAC§ | 8q24.21 | rs6983267 | 95 | 1364 | 2439 | 0.96 (0.82-1.13) | 0.88 (0.72-1.07) | 0.1965 | 0.2171 |
| BCAC§ | 8q24.21 | rs6983561 | 97 | 1444 | 2456 | 0.97 (0.75-1.27) | 1.86 (0.31-11.15) | 0.9962 | 0.9964 |
| BCAC§ | 8q24.21 | rs9283954 | 95 | 1409 | 2420 | 1.06 (0.89-1.26) | 1.33 (0.71-2.50) | 0.3346 | 0.3565 |
| **BCAC§** | **11p15.5** | **rs2107425** | 98 | 1460 | 2463 | **0.71 (0.62-0.82)** | 0.88 (0.70-1.10) | **0.0012** | **0.002** |
| BCAC§ | 12p11.22 | rs7313833 | 98 | 1460 | 2457 | 1.08 (0.94-1.24) | 1.08 (0.87-1.34) | 0.3209 | 0.3434 |
| BCAC§ | *FLJ41481* | rs4666451 | 98 | 1454 | 2469 | 0.96 (0.83-1.11) | 1.03 (0.85-1.25) | 0.926 | 0.9294 |
| BCAC§ | *HCN1* | rs981782 | 96 | 1419 | 2432 | 1.12 (0.96-1.32) | 1.05 (0.87-1.27) | 0.5399 | 0.5582 |
| BCAC§ | *LOC100131885* | rs2981582 | 98 | 1463 | 2466 | 0.88 (0.76-1.02) | 0.89 (0.73-1.09) | 0.1381 | 0.157 |
| **BCAC§** | ***LSP1*** | **rs3817198** | 97 | 1457 | 2435 | **1.16 (1.01-1.34)** | **1.40 (1.11-1.75)** | **0.0016** | **0.0026** |
| BCAC§ | *TOX3* | rs12443621 | 97 | 1445 | 2457 | 0.96 (0.82-1.12) | 0.83 (0.69-1.01) | 0.0752 | 0.0894 |
| **Cell cycle** | ***CCND1*** | **rs602652** | 99 | 1468 | 2493 | 1.13 (0.97-1.32) | **1.24 (1.03-1.49)** | **0.0235** | **0.0307** |
| Cell cycle | *CCND1* | rs3862792 | 98 | 1471 | 2468 | 0.83 (0.61-1.13) | 0.91 (0.76-1.08) | 0.2311 | 0.2531 |
| **Cell cycle** | ***CCND1*** | **rs603965** | 98 | 1476 | 2464 | 1.06 (0.91-1.23) | **1.28 (1.06-1.55)** | **0.013** | **0.0178** |
| **Cell cycle** | ***CCND1*** | **rs3212879** | 99 | 1472 | 2491 | **0.85 (0.73-0.99)** | **0.82 (0.68-0.99)** | **0.0321** | **0.0409** |
| **Cell cycle** | ***CCND1*** | **rs3212891** | 98 | 1475 | 2476 | **0.86 (0.74-1.00)** | **0.83 (0.69-1.00)** | **0.0376** | **0.0472** |
| Cell cycle | *CCND1* | rs678653 | 98 | 1465 | 2468 | 0.94 (0.81-1.08) | 0.93 (0.76-1.15) | 0.3791 | 0.4012 |
| **Cell cycle** | ***CCND1*** | **rs7178** | 99 | 1480 | 2491 | **1.24 (1.04-1.49)** | 1.24 (0.50-3.04) | **0.0211** | **0.0278** |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Cell cycle | *CCND2* | rs3217795 | 99 | 1485 | 2500 | 1.00 (0.84-1.20) | 0.71 (0.28-1.76) | 0.819 | 0.8272 |
| Cell cycle | *CCND2* | rs3217805 | 99 | 1458 | 2502 | 1.08 (0.93-1.25) | 0.86 (0.70-1.05) | 0.3599 | 0.3824 |
| Cell cycle | *CCND2* | rs3217820 | 97 | 1407 | 2497 | 1.12 (0.97-1.29) | 1.01 (0.82-1.25) | 0.4954 | 0.5152 |
| Cell cycle | *CCND2* | rs3217852 | 99 | 1474 | 2488 | 0.99 (0.86-1.14) | 0.86 (0.64-1.15) | 0.4711 | 0.4917 |
| Cell cycle | *CCND2* | rs3217862 | 99 | 1470 | 2493 | 0.93 (0.80-1.08) | 0.94 (0.62-1.40) | 0.3481 | 0.3708 |
| Cell cycle | *CCND2* | rs3217863 | 98 | 1459 | 2472 | 0.99 (0.83-1.20) | 1.49 (0.75-2.93) | 0.6464 | 0.6615 |
| Cell cycle | *CCND2* | rs3217869 | 99 | 1483 | 2498 | 1.01 (0.87-1.16) | 1.05 (0.87-1.28) | 0.6306 | 0.6465 |
| Cell cycle | *CCND2* | rs3217901 | 99 | 1479 | 2489 | 1.10 (0.95-1.28) | 1.03 (0.85-1.25) | 0.5991 | 0.616 |
| Cell cycle | *CCND2* | rs3217906 | 98 | 1467 | 2473 | 1.05 (0.91-1.20) | 0.97 (0.74-1.28) | 0.7711 | 0.7813 |
| Cell cycle | *CCND2* | rs3217916 | 98 | 1476 | 2477 | 1.01 (0.88-1.16) | 0.75 (0.58-0.97) | 0.1541 | 0.1739 |
| Cell cycle | *CCND2* | rs3217925 | 98 | 1470 | 2461 | 0.98 (0.86-1.13) | 0.72 (0.54-0.96) | 0.1078 | 0.1248 |
| Cell cycle | *CCND2* | rs3217926 | 99 | 1484 | 2500 | 1.05 (0.91-1.21) | 0.90 (0.74-1.10) | 0.5268 | 0.5461 |
| Cell cycle | *CCND2* | rs3217933 | 98 | 1478 | 2474 | 1.13 (0.99-1.30) | 1.01 (0.76-1.33) | 0.241 | 0.2632 |
| Cell cycle | *CCND2* | rs3217936 | 99 | 1478 | 2489 | 1.02 (0.89-1.17) | **0.77 (0.61-0.97)** | 0.1303 | 0.1489 |
| Cell cycle | *CCND3* | rs1410492 | 99 | 1479 | 2490 | 1.08 (0.94-1.24) | 1.06 (0.80-1.40) | 0.3187 | 0.3415 |
| Cell cycle | *CCND3* | rs3218092 | 98 | 1468 | 2468 | 1.02 (0.89-1.18) | 0.74 (0.51-1.06) | 0.4626 | 0.4833 |
| Cell cycle | *CCND3* | rs2479717 | 99 | 1472 | 2489 | 1.05 (0.91-1.20) | 0.98 (0.76-1.25) | 0.7947 | 0.804 |
| Cell cycle | *CCND3* | rs1051130 | 99 | 1475 | 2488 | 1.13 (0.96-1.31) | 0.87 (0.72-1.05) | 0.2 | 0.2215 |
| Cell cycle | *CCND3* | rs9529 | 99 | 1481 | 2475 | 1.00 (0.87-1.15) | 0.79 (0.62-1.02) | 0.1898 | 0.2109 |
| Cell cycle | *CCND3* | rs3218110 | 99 | 1479 | 2482 | 1.11 (0.97-1.28) | 1.11 (0.83-1.47) | 0.145 | 0.1644 |
| Cell cycle | *CCND3* | rs3218114 | 99 | 1484 | 2492 | 1.01 (0.88-1.17) | 0.78 (0.53-1.13) | 0.5375 | 0.5565 |
| Cell cycle | *CCNE1* | rs997669 | 99 | 1480 | 2497 | 1.10 (0.95-1.27) | 1.14 (0.93-1.39) | 0.1414 | 0.1607 |
| Cell cycle | *CCNE1* | rs3218036 | 99 | 1476 | 2481 | 1.07 (0.93-1.23) | 1.27 (1.01-1.59) | **0.0458** | 0.0567 |
| Cell cycle | *CCNE1* | rs3218038 | 99 | 1479 | 2497 | 0.97 (0.76-1.25) | 1.90 (0.32-11.42) | 0.9863 | 0.987 |
| Cell cycle | *CCNE1* | rs3218076 | 99 | 1480 | 2497 | 1.02 (0.89-1.17) | 1.02 (0.79-1.33) | 0.7375 | 0.7492 |
| Cell cycle | *CDK2* | rs2069408 | 99 | 1476 | 2496 | 0.90 (0.78-1.03) | 0.89 (0.71-1.11) | 0.1327 | 0.1515 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Cell cycle | *CDK2* | rs1045435 | 99 | 1482 | 2497 | 1.11 (0.94-1.33) | 1.87 (0.92-3.82) | 0.0735 | 0.0879 |
| Cell cycle | *CDK4* | rs2270777 | 99 | 1477 | 2491 | 1.07 (0.92-1.24) | 1.07 (0.88-1.30) | 0.4203 | 0.442 |
| Cell cycle | *CDK4* | rs2069506 | 99 | 1480 | 2482 | 0.94 (0.82-1.08) | 1.00 (0.81-1.24) | 0.6775 | 0.6915 |
| Cell cycle | *CDK6* | rs8179 | 97 | 1465 | 2440 | 0.96 (0.84-1.11) | 1.07 (0.78-1.48) | 0.9164 | 0.9202 |
| Cell cycle | *CDK6* | rs2285332 | 98 | 1476 | 2463 | 0.92 (0.80-1.06) | 0.88 (0.66-1.17) | 0.1799 | 0.2006 |
| Cell cycle | *CDK6* | rs42046 | 97 | 1466 | 2446 | 1.07 (0.93-1.23) | 1.07 (0.82-1.39) | 0.3659 | 0.3881 |
| Cell cycle | *CDK6* | rs3731348 | 98 | 1463 | 2465 | 1.08 (0.88-1.32) | 0.39 (0.11-1.39) | 0.9009 | 0.9054 |
| **Cell cycle** | ***CDK6*** | **rs8** | 99 | 1473 | 2481 | **1.17 (1.02-1.35)** | **1.44 (1.04-1.99)** | **0.0039** | **0.0059** |
| Cell cycle | *CDK6* | rs2237570 | 98 | 1474 | 2468 | 0.89 (0.76-1.06) | 0.71 (0.38-1.33) | 0.1015 | 0.1182 |
| Cell cycle | *CDK6* | rs3731343 | 99 | 1484 | 2499 | 1.04 (0.89-1.21) | 1.09 (0.90-1.31) | 0.3637 | 0.3863 |
| Cell cycle | *CDK6* | rs3757823 | 99 | 1486 | 2495 | 1.05 (0.89-1.24) | 1.09 (0.53-2.23) | 0.534 | 0.553 |
| Cell cycle | *CDK6* | rs2079147 | 99 | 1481 | 2496 | 1.01 (0.86-1.18) | 1.01 (0.84-1.22) | 0.9246 | 0.928 |
| Cell cycle | *CDK6* | rs2282991 | 96 | 1449 | 2404 | 0.85 (0.71-1.01) | 0.86 (0.44-1.69) | 0.0626 | 0.0753 |
| Cell cycle | *CDK6* | rs4729049 | 99 | 1482 | 2499 | 1.01 (0.85-1.19) | 0.86 (0.44-1.69) | 0.9082 | 0.9125 |
| Cell cycle | *CDK6* | rs445 | 99 | 1481 | 2497 | 0.99 (0.83-1.17) | 0.89 (0.44-1.78) | 0.7706 | 0.7809 |
| Cell cycle | *CDK6* | rs992519 | 99 | 1477 | 2496 | 1.05 (0.90-1.22) | 0.99 (0.62-1.59) | 0.6382 | 0.6538 |
| Cell cycle | *CDKN1A* | rs762624 | 99 | 1475 | 2483 | 1.06 (0.92-1.21) | 1.11 (0.86-1.44) | 0.3019 | 0.3247 |
| Cell cycle | *CDKN1A* | rs2395655 | 99 | 1478 | 2487 | 1.12 (0.97-1.29) | 1.18 (0.97-1.43) | 0.0692 | 0.083 |
| Cell cycle | *CDKN1A* | rs3176331 | 98 | 1471 | 2472 | 1.01 (0.86-1.18) | 1.14 (0.70-1.85) | 0.7129 | 0.7255 |
| Cell cycle | *CDKN1A* | rs3176336 | 99 | 1481 | 2488 | 0.98 (0.85-1.14) | 0.94 (0.77-1.14) | 0.5351 | 0.5541 |
| Cell cycle | *CDKN1A* | rs3176343 | 99 | 1483 | 2491 | 1.03 (0.83-1.28) | 0.78 (0.14-4.37) | 0.8189 | 0.8272 |
| Cell cycle | *CDKN1A* | rs1801270 | 98 | 1477 | 2463 | 0.91 (0.75-1.11) | 0.53 (0.19-1.50) | 0.1987 | 0.22 |
| Cell cycle | *CDKN1A* | rs3176352 | 98 | 1477 | 2476 | 1.03 (0.89-1.18) | 1.17 (0.91-1.50) | 0.2978 | 0.3205 |
| Cell cycle | *CDKN1A* | rs1059234 | 99 | 1483 | 2481 | 0.90 (0.74-1.09) | 0.73 (0.29-1.83) | 0.2182 | 0.2401 |
| Cell cycle | *CDKN1A* | rs6457937 | 99 | 1483 | 2495 | 0.97 (0.72-1.30) | 0.90 (0.76-1.08) | 0.8263 | 0.8342 |
| Cell cycle | *CDKN1B* | rs3759217 | 99 | 1483 | 2492 | 1.15 (0.98-1.34) | 1.04 (0.60-1.79) | 0.1301 | 0.1487 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Cell cycle | *CDKN1B* | rs34330 | 98 | 1468 | 2467 | 0.94 (0.81-1.08) | 0.90 (0.67-1.20) | 0.2774 | 0.3 |
| **Cell cycle** | ***CDKN1B*** | **rs2066827** | 99 | 1481 | 2484 | 0.88 (0.77-1.01) | **0.68 (0.51-0.90)** | **0.0035** | **0.0053** |
| Cell cycle | *CDKN1B* | rs34329 | 99 | 1481 | 2491 | 1.01 (0.88-1.16) | 0.96 (0.76-1.21) | 0.8616 | 0.868 |
| Cell cycle | *CDKN1B* | rs3093736 | 99 | 1483 | 2493 | 0.94 (0.72-1.23) | 1.53 (0.25-9.44) | 0.7519 | 0.763 |
| Cell cycle | *CDKN1B* | rs7330 | 99 | 1484 | 2499 | 0.95 (0.82-1.09) | 1.01 (0.83-1.23) | 0.8981 | 0.9028 |
| Cell cycle | *CDKN1B* | rs1420023 | 99 | 1475 | 2486 | 0.95 (0.81-1.13) | 1.08 (0.59-1.97) | 0.7146 | 0.7272 |
| Cell cycle | *CDKN2A* | rs3217992 | 99 | 1474 | 2489 | 0.96 (0.83-1.10) | 0.91 (0.74-1.13) | 0.3764 | 0.3987 |
| Cell cycle | *CDKN2A* | rs3218005 | 98 | 1461 | 2473 | 1.04 (0.87-1.24) | 1.63 (0.87-3.05) | 0.2853 | 0.3079 |
| Cell cycle | *CDKN2A* | rs2811712 | 99 | 1479 | 2494 | 0.97 (0.82-1.15) | **1.91 (1.08-3.37)** | 0.4497 | 0.4709 |
| Cell cycle | *CDKN2A* | rs3218020 | 98 | 1467 | 2463 | 0.99 (0.86-1.14) | 0.93 (0.74-1.17) | 0.6253 | 0.6412 |
| Cell cycle | *CDKN2A* | rs3731197 | 99 | 1476 | 2479 | 1.02 (0.88-1.17) | 1.07 (0.87-1.30) | 0.5644 | 0.5824 |
| Cell cycle | *CDKN2A* | rs3731211 | 99 | 1476 | 2484 | 0.95 (0.83-1.09) | 1.06 (0.82-1.37) | 0.8613 | 0.8676 |
| Cell cycle | *CDKN2A* | rs3731222 | 99 | 1480 | 2497 | 1.01 (0.87-1.19) | 0.82 (0.50-1.36) | 0.8169 | 0.8252 |
| Cell cycle | *CDKN2A* | rs4074785 | 98 | 1471 | 2464 | 1.04 (0.88-1.24) | 1.30 (0.68-2.47) | 0.4466 | 0.4677 |
| Cell cycle | *CDKN2A* | rs3731239 | 99 | 1478 | 2487 | 1.08 (0.94-1.24) | 1.08 (0.88-1.33) | 0.324 | 0.3468 |
| Cell cycle | *CDKN2A* | rs3731249 | 98 | 1464 | 2452 | 0.86 (0.64-1.16) | 1.03 (0.23-4.67) | 0.3584 | 0.3806 |
| Cell cycle | *CDKN2A* | rs11515 | 99 | 1479 | 2492 | 1.00 (0.86-1.16) | 1.53 (0.98-2.41) | 0.3809 | 0.4032 |
| Cell cycle | *CDKN2A* | rs3088440 | 99 | 1475 | 2490 | 1.02 (0.86-1.21) | 1.33 (0.65-2.70) | 0.6122 | 0.6287 |
| Cell cycle | *CDKN2A* | rs3731257 | 99 | 1480 | 2476 | 0.89 (0.78-1.03) | 0.80 (0.60-1.07) | **0.0452** | 0.056 |
| Cell cycle | *CDKN2B* | rs3217986 | 99 | 1478 | 2486 | 1.04 (0.88-1.24) | 1.23 (0.64-2.37) | 0.4832 | 0.5035 |
| Cell cycle | *CDKN2B* | rs1063192 | 99 | 1476 | 2491 | 1.01 (0.87-1.17) | 0.98 (0.81-1.19) | 0.8927 | 0.8976 |
| Cell cycle | *CDKN2B* | rs3218009 | 98 | 1469 | 2478 | 0.93 (0.80-1.09) | 1.25 (0.77-2.02) | 0.7664 | 0.7768 |
| Cell cycle | *CDKN2B* | rs3218012 | 98 | 1464 | 2477 | 0.96 (0.83-1.12) | 0.96 (0.79-1.16) | 0.6371 | 0.6526 |
| Cell cycle | *CDKN2C* | rs12855 | 99 | 1482 | 2499 | 0.98 (0.83-1.17) | 1.30 (0.68-2.51) | 0.8688 | 0.8748 |
| Cell cycle | *CDKN2C* | rs3176459 | 98 | 1466 | 2480 | 1.06 (0.92-1.22) | 0.93 (0.75-1.16) | 0.9162 | 0.92 |
| Cell cycle | *CDKN2D* | rs1465702 | 99 | 1483 | 2481 | 1.12 (0.89-1.41) | 0.65 (0.13-3.33) | 0.4518 | 0.4729 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Cell cycle | *CDKN2D* | rs3218222 | 98 | 1468 | 2459 | 0.97 (0.85-1.12) | 1.04 (0.78-1.38) | 0.9412 | 0.9439 |
| Cell cycle | *RB1* | rs1981434 | 98 | 1463 | 2460 | 0.97 (0.85-1.12) | 1.01 (0.79-1.30) | 0.8937 | 0.8985 |
| Cell cycle | *RB1* | rs2854345 | 99 | 1476 | 2480 | 1.02 (0.88-1.17) | 1.02 (0.73-1.43) | 0.8221 | 0.8301 |
| Cell cycle | *RB1* | rs399413 | 82 | 1314 | 1989 | 1.02 (0.88-1.19) | 1.15 (0.88-1.51) | 0.3906 | 0.4093 |
| Cell cycle | *RB1* | rs4151540 | 98 | 1469 | 2455 | 0.91 (0.79-1.04) | 1.03 (0.80-1.33) | 0.5165 | 0.5358 |
| Cell cycle | *RB1* | rs4151551 | 99 | 1472 | 2487 | 1.08 (0.90-1.29) | 1.16 (0.58-2.31) | 0.3718 | 0.3941 |
| Cell cycle | *RB1* | rs2854344 | 82 | 1304 | 1975 | **0.73 (0.59-0.91)** | 1.84 (0.78-4.32) | 0.0552 | 0.0653 |
| Cell cycle | *RB1* | rs425834 | 98 | 1472 | 2469 | 1.04 (0.80-1.37) | 2.59 (0.71-9.43) | 0.3967 | 0.4186 |
| Cell cycle | *RB1* | rs4151611 | 98 | 1463 | 2476 | 0.87 (0.69-1.10) | 1.03 (0.33-3.23) | 0.2988 | 0.3214 |
| Cell cycle | *RB1* | rs4151620 | 81 | 1300 | 1958 | 1.12 (0.95-1.33) | **0.22 (0.08-0.62)** | 0.9699 | 0.9711 |
| Cell cycle | *RB1* | rs3092904 | 82 | 1317 | 1993 | 0.96 (0.83-1.12) | 1.09 (0.82-1.45) | 0.953 | 0.9549 |
| Cell cycle | *RB1* | rs4151636 | 99 | 1467 | 2489 | 0.87 (0.69-1.10) | 0.81 (0.19-3.43) | 0.2213 | 0.2432 |
| Cell cycle | *STK15* | rs732417 | 88 | 1397 | 2139 | 1.12 (0.92-1.35) | 1.32 (0.55-3.15) | 0.2128 | 0.2324 |
| Cell cycle | *STK15* | rs1047972 | 93 | 1439 | 2286 | 0.97 (0.83-1.13) | 0.90 (0.60-1.35) | 0.5405 | 0.5583 |
| Cell cycle | *STK15* | rs2273535 | 81 | 1333 | 1917 | 1.15 (0.99-1.34) | 1.20 (0.87-1.66) | 0.0515 | 0.0611 |
| Cell cycle | *STK15* | rs8173 | 86 | 1338 | 2103 | 1.01 (0.88-1.17) | 1.29 (0.94-1.76) | 0.2747 | 0.2945 |
| DNA repair | *BRCA1* | rs799917 | 89 | 1414 | 2157 | **1.16 (1.01-1.34)** | 1.09 (0.87-1.37) | 0.1299 | 0.1467 |
| DNA repair | *BRCA1* | rs1799950 | 58 | 1036 | 1275 | 1.04 (0.81-1.34) | 0.30 (0.03-2.72) | 0.9818 | 0.9823 |
| DNA repair | *BRIP* | rs11871785 | 99 | 1478 | 2488 | 1.05 (0.91-1.21) | 0.96 (0.78-1.18) | 0.9737 | 0.9749 |
| DNA repair | *BRIP* | rs1557720 | 82 | 1317 | 1987 | 1.06 (0.91-1.24) | 1.17 (0.95-1.45) | 0.1434 | 0.1596 |
| DNA repair | *BRIP* | rs11652980 | 99 | 1476 | 2495 | 1.02 (0.82-1.28) | 1.17 (0.31-4.44) | 0.7972 | 0.8064 |
| DNA repair | *BRIP* | rs2191249 | 82 | 1313 | 1970 | 0.94 (0.81-1.09) | 0.76 (0.56-1.02) | 0.0798 | 0.0921 |
| DNA repair | *BRIP* | rs16945628 | 99 | 1481 | 2488 | 0.93 (0.81-1.06) | 0.90 (0.73-1.12) | 0.2309 | 0.2531 |
| DNA repair | *BRIP* | rs2191248 | 98 | 1469 | 2482 | 1.00 (0.87-1.15) | 1.20 (0.97-1.49) | 0.2147 | 0.2364 |
| DNA repair | *BRIP* | rs16945643 | 99 | 1471 | 2486 | 0.97 (0.81-1.17) | 1.83 (0.76-4.38) | 0.8358 | 0.8433 |
| DNA repair | *BRIP* | rs6504074 | 80 | 1294 | 1918 | 0.94 (0.81-1.09) | 0.85 (0.65-1.12) | 0.208 | 0.2257 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| DNA repair | *BRIP* | rs2378908 | 81 | 1287 | 1982 | 1.10 (0.93-1.30) | 1.18 (0.70-2.00) | 0.2122 | 0.2303 |
| DNA repair | *BRIP* | rs4988344 | 82 | 1315 | 1992 | 1.09 (0.93-1.28) | 1.43 (0.94-2.18) | 0.0777 | 0.09 |
| DNA repair | *BRIP* | rs9908659 | 99 | 1478 | 2488 | 0.98 (0.85-1.14) | 1.08 (0.88-1.32) | 0.6005 | 0.6174 |
| DNA repair | *BRIP* | rs4968451 | 99 | 1477 | 2489 | 1.06 (0.91-1.23) | 1.38 (0.93-2.06) | 0.1459 | 0.1653 |
| DNA repair | *BRIP* | rs2048718 | 98 | 1459 | 2482 | 0.94 (0.81-1.10) | 1.04 (0.86-1.25) | 0.8081 | 0.8167 |
| DNA repair | *NBS1* | rs1063045 | 94 | 1453 | 2305 | 0.99 (0.86-1.15) | 0.93 (0.74-1.16) | 0.5945 | 0.6107 |
| DNA repair | *NBS1* | rs1805794 | 92 | 1423 | 2267 | 1.00 (0.87-1.16) | 0.85 (0.68-1.07) | 0.3049 | 0.3262 |
| DNA repair | *NBS1* | rs709816 | 93 | 1448 | 2273 | 1.03 (0.89-1.19) | 0.96 (0.78-1.18) | 0.8746 | 0.88 |
| DNA repair | *NBS1* | rs1061302 | 84 | 1333 | 2035 | 1.02 (0.88-1.18) | 0.89 (0.70-1.12) | 0.5305 | 0.5469 |
| DNA repair | *RAD51* | rs1801320 | 87 | 1399 | 2092 | 1.08 (0.88-1.33) | 1.51 (0.60-3.76) | 0.3156 | 0.3357 |
| DNA repair | *RAD51* | rs1801321 | 85 | 1412 | 1986 | 1.02 (0.87-1.18) | 1.21 (0.99-1.48) | 0.0985 | 0.1127 |
| DNA repair | *RAD52* | rs11226 | 93 | 1449 | 2293 | **1.17 (1.00-1.37)** | 1.03 (0.85-1.24) | 0.6166 | 0.6321 |
| DNA repair | *XRCC2* | XRCC212 | 83 | 1364 | 1959 | 0.95 (0.75-1.20) | 1.03 (0.35-2.98) | 0.6961 | 0.7074 |
| DNA repair | *XRCC2* | rs3218384 | 89 | 1426 | 2160 | 1.00 (0.86-1.15) | 0.99 (0.72-1.38) | 0.9722 | 0.9734 |
| **DNA repair** | ***XRCC2*** | **rs3218536** | 78 | 1337 | 1787 | 0.88 (0.72-1.08) | **0.23 (0.07-0.79)** | **0.0364** | **0.0439** |
| DNA repair | *XRCC3* | rs1799794 | 94 | 1461 | 2307 | 1.07 (0.93-1.24) | **0.66 (0.44-0.98)** | 0.6996 | 0.7122 |
| DNA repair | *XRCC3* | rs1799796 | 94 | 1461 | 2299 | 0.92 (0.80-1.06) | 1.12 (0.90-1.39) | 0.8358 | 0.8429 |
| DNA repair | *XRCC3* | rs861539 | 84 | 1332 | 2024 | 0.91 (0.78-1.06) | 0.91 (0.73-1.14) | 0.2705 | 0.2898 |
| Mismatch | *MLH1* | rs1800734 | 81 | 1299 | 1957 | 0.95 (0.82-1.11) | 1.07 (0.76-1.49) | 0.8455 | 0.8513 |
| Mismatch | *MLH1* | rs1540354 | 81 | 1305 | 1960 | 1.05 (0.90-1.23) | 1.10 (0.76-1.60) | 0.4616 | 0.4789 |
| Mismatch | *MLH1* | rs1799977 | 56 | 1017 | 1238 | 1.07 (0.90-1.28) | 0.96 (0.72-1.28) | 0.8334 | 0.8379 |
| Mismatch | *MLH1* | rs2286939 | 84 | 1318 | 2051 | 1.05 (0.89-1.23) | 0.98 (0.80-1.20) | 0.9276 | 0.9304 |
| Mismatch | *MLH3* | rs7303 | 79 | 1281 | 1906 | 1.03 (0.87-1.22) | 1.08 (0.89-1.32) | 0.4363 | 0.4537 |
| Mismatch | *MLH3* | rs175080 | 82 | 1296 | 1981 | 0.97 (0.82-1.14) | 0.89 (0.73-1.10) | 0.2971 | 0.3161 |
| Mismatch | *MSH2* | rs4952887 | 82 | 1302 | 1975 | 0.82 (0.67-1.00) | 1.24 (0.63-2.44) | 0.1534 | 0.1698 |
| Mismatch | *MSH2* | rs13425206 | 99 | 1475 | 2491 | 0.94 (0.73-1.22) | 1.06 (0.33-3.39) | 0.698 | 0.7113 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Mismatch | *MSH2* | rs3771274 | 82 | 1304 | 1978 | 1.02 (0.88-1.20) | 1.01 (0.81-1.25) | 0.8708 | 0.8757 |
| Mismatch | *MSH2* | rs1981928 | 83 | 1314 | 2001 | 1.00 (0.86-1.16) | 0.88 (0.67-1.15) | 0.5225 | 0.5388 |
| Mismatch | *MSH2* | rs2059520 | 82 | 1312 | 1981 | 1.03 (0.88-1.19) | 1.00 (0.80-1.26) | 0.884 | 0.8885 |
| Mismatch | *MSH2* | rs2303428 | 84 | 1317 | 2047 | 0.90 (0.75-1.08) | 0.59 (0.28-1.24) | 0.1071 | 0.1217 |
| Mismatch | *MSH3* | rs6151662 | 98 | 1474 | 2461 | 1.02 (0.82-1.26) | **0.27 (0.08-0.92)** | 0.4541 | 0.475 |
| Mismatch | *MSH3* | rs40139 | 83 | 1315 | 2026 | **1.17 (1.00-1.38)** | 0.96 (0.78-1.17) | 0.9751 | 0.976 |
| Mismatch | *MSH3* | rs26282 | 82 | 1308 | 1988 | 1.02 (0.88-1.19) | 1.01 (0.76-1.33) | 0.8299 | 0.8364 |
| Mismatch | *MSH3* | rs26779 | 98 | 1475 | 2466 | 1.00 (0.87-1.16) | 0.98 (0.80-1.19) | 0.8493 | 0.8561 |
| Mismatch | *MSH3* | rs33008 | 82 | 1227 | 1829 | 0.77 (0.32-1.87) | 0.85 (0.35-2.07) | 0.8039 | 0.8113 |
| Mismatch | *MSH3* | rs10079641 | 98 | 1473 | 2470 | 1.10 (0.92-1.32) | 0.67 (0.36-1.28) | 0.7589 | 0.7696 |
| Mismatch | *MSH3* | rs184967 | 92 | 1305 | 1983 | 1.01 (0.86-1.18) | 1.04 (0.65-1.65) | 0.8937 | 0.8978 |
| Mismatch | *MSH3* | rs2897298 | 98 | 1465 | 2450 | 1.10 (0.94-1.29) | 0.87 (0.50-1.52) | 0.4274 | 0.4487 |
| Mismatch | *MSH3* | rs26279 | 82 | 1301 | 2008 | 1.08 (0.93-1.25) | 1.10 (0.85-1.42) | 0.2922 | 0.3114 |
| Mismatch | *MSH3* | rs2112416 | 82 | 1310 | 1985 | 0.99 (0.83-1.16) | 1.14 (0.71-1.83) | 0.8772 | 0.8819 |
| Mismatch | *MSH6* | rs3136245 | 81 | 1310 | 1960 | 0.98 (0.84-1.14) | **0.68 (0.47-0.99)** | 0.1573 | 0.1739 |
| Mismatch | *MSH6* | rs3136272 | 97 | 1450 | 2425 | 0.92 (0.80-1.06) | 1.02 (0.83-1.26) | 0.7715 | 0.7815 |
| Mismatch | *MSH6* | rs1800932 | 83 | 1317 | 2001 | 1.09 (0.93-1.28) | 1.01 (0.70-1.45) | 0.4088 | 0.4273 |
| Mismatch | *MSH6* | rs2348244 | 80 | 1279 | 1940 | 1.07 (0.90-1.26) | 0.65 (0.40-1.05) | 0.727 | 0.7369 |
| Mismatch | *MSH6* | rs3136317 | 99 | 1476 | 2490 | 1.10 (0.95-1.27) | 1.08 (0.70-1.68) | 0.227 | 0.2491 |
| Mismatch | *MSH6* | rs1800935 | 83 | 1316 | 2026 | 1.04 (0.90-1.20) | 1.03 (0.79-1.33) | 0.6797 | 0.6916 |
| Mismatch | *MSH6* | rs2020911 | 81 | 1306 | 1950 | 1.05 (0.90-1.22) | 0.90 (0.72-1.12) | 0.5504 | 0.5658 |
| Mismatch | *PMS1* | rs3762545 | 83 | 1308 | 2018 | 0.94 (0.81-1.10) | 0.81 (0.56-1.18) | 0.2317 | 0.2505 |
| Mismatch | *PMS1* | rs5742981 | 90 | 1300 | 2299 | 0.94 (0.72-1.23) | 0.36 (0.04-3.10) | 0.4592 | 0.4783 |
| Mismatch | *PMS1* | rs5741593 | 97 | 1462 | 2447 | 0.87 (0.71-1.07) | 0.63 (0.20-2.00) | 0.1275 | 0.1457 |
| Mismatch | *PMS1* | rs1233291 | 82 | 1309 | 1977 | 0.89 (0.77-1.03) | 0.84 (0.63-1.12) | 0.08 | 0.0924 |
| Mismatch | *PMS1* | rs1233255 | 79 | 1271 | 1905 | 1.03 (0.88-1.21) | 0.79 (0.55-1.13) | 0.6249 | 0.6378 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Mismatch | *PMS1* | rs1233258 | 82 | 1305 | 1969 | 0.91 (0.79-1.06) | 0.94 (0.72-1.22) | 0.3118 | 0.3307 |
| Mismatch | *PMS1* | rs256571 | 99 | 1475 | 2484 | 1.02 (0.83-1.25) | 0.41 (0.09-1.97) | 0.8635 | 0.8697 |
| Mismatch | *PMS1* | rs256563 | 98 | 1456 | 2446 | 2.50 (0.99-6.33) | 2.15 (0.84-5.48) | **0.04** | 0.05 |
| **Mismatch** | ***PMS2*** | **rs7797466** | 82 | 1305 | 1968 | **1.18 (1.01-1.38)** | 1.38 (0.96-2.00) | **0.0108** | **0.0142** |
| Mismatch | *PMS2* | rs2345060 | 82 | 1319 | 1987 | 1.02 (0.88-1.18) | 0.82 (0.59-1.12) | 0.5412 | 0.557 |
| Mismatch | *PMS2* | rs2286680 | 83 | 1316 | 2026 | 1.09 (0.92-1.28) | 1.13 (0.68-1.88) | 0.2972 | 0.3165 |
| Mismatch | *PMS2* | rs12112229 | 81 | 1299 | 1945 | 0.95 (0.82-1.10) | 1.15 (0.86-1.54) | 0.8827 | 0.8872 |
| Mismatch | *PMS2* | rs1805321 | 82 | 1317 | 1980 | 0.85 (0.73-1.00) | 0.93 (0.75-1.14) | 0.2586 | 0.2775 |
| Mismatch | *PMS2* | rs2228006 | 98 | 1467 | 2478 | 1.01 (0.87-1.18) | 1.24 (0.77-2.00) | 0.5728 | 0.5905 |
| MMCT-18* | *AIFM2* | rs2394655 | 75 | 1037 | 1983 | 1.11 (0.83-1.49) | 0.52 (0.06-4.50) | 0.6144 | 0.6282 |
| MMCT-18* | *AIFM2* | rs7908957 | 74 | 1017 | 1939 | 0.84 (0.70-1.02) | 1.16 (0.67-2.02) | 0.2381 | 0.2529 |
| MMCT-18* | *AIFM2* | rs1053495 | 67 | 958 | 1741 | 0.91 (0.72-1.16) | 0.44 (0.14-1.32) | 0.1832 | 0.1945 |
| MMCT-18* | *AIFM2* | rs2894111 | 74 | 1046 | 1925 | 0.96 (0.81-1.12) | 0.75 (0.56-1.01) | 0.1055 | 0.1173 |
| MMCT-18* | *AIFM2* | rs2394656 | 39 | 440 | 1134 | 0.88 (0.69-1.12) | 0.80 (0.41-1.55) | 0.2354 | 0.2415 |
| MMCT-18* | *AIFM2* | rs6480440 | 39 | 422 | 1140 | 1.01 (0.79-1.29) | 0.90 (0.56-1.44) | 0.7992 | 0.8026 |
| MMCT-18* | *AIFM2* | rs2280201 | 76 | 1041 | 1996 | 1.04 (0.87-1.26) | 0.81 (0.44-1.50) | 0.96 | 0.9614 |
| MMCT-18* | *AIFM2* | rs10999147 | 76 | 1049 | 2017 | 1.10 (0.89-1.36) | 0.73 (0.19-2.82) | 0.4693 | 0.4866 |
| MMCT-18* | *AIFM2* | rs3750772 | 75 | 1027 | 1995 | 1.04 (0.81-1.33) | 0.99 (0.25-3.90) | 0.7791 | 0.787 |
| MMCT-18* | *AIFM2* | rs4295944 | 76 | 1047 | 2008 | 0.99 (0.84-1.17) | 1.08 (0.87-1.34) | 0.5842 | 0.5978 |
| MMCT-18* | *AIFM2* | rs2394644 | 34 | 600 | 745 | 1.05 (0.81-1.37) | 0.78 (0.35-1.74) | 0.9997 | 0.9997 |
| MMCT-18* | *AIFM2* | rs10999152 | 64 | 908 | 1672 | 1.17 (0.98-1.41) | 1.11 (0.70-1.75) | 0.1151 | 0.1279 |
| MMCT-18* | *AKTIP* | rs9931702 | 40 | 441 | 1150 | 0.95 (0.74-1.22) | 0.90 (0.66-1.24) | 0.5203 | 0.528 |
| MMCT-18* | *AKTIP* | rs17801966 | 31 | 348 | 898 | 0.92 (0.69-1.24) | 0.90 (0.32-2.53) | 0.5794 | 0.5836 |
| MMCT-18* | *AKTIP* | rs7189819 | 75 | 1034 | 1980 | 0.89 (0.76-1.04) | 0.82 (0.63-1.06) | 0.066 | 0.0754 |
| MMCT-18* | *AKTIP* | rs3743772 | 38 | 413 | 1093 | 0.88 (0.60-1.28) | 1.74 (0.29-10.48) | 0.6778 | 0.6817 |
| MMCT-18* | *AXIN2* | rs11868547 | 65 | 847 | 1764 | 0.95 (0.78-1.16) | 0.92 (0.73-1.16) | 0.463 | 0.4768 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| MMCT-18* | *AXIN2* | rs7591 | 74 | 1041 | 1921 | 1.12 (0.95-1.33) | 1.04 (0.82-1.32) | 0.461 | 0.4769 |
| MMCT-18* | *AXIN2* | rs4074947 | 74 | 1043 | 1946 | 1.09 (0.92-1.28) | 0.98 (0.66-1.45) | 0.5141 | 0.5294 |
| MMCT-18* | *AXIN2* | rs7210356 | 76 | 1047 | 2019 | 1.11 (0.92-1.34) | 0.84 (0.41-1.71) | 0.443 | 0.4606 |
| MMCT-18* | *AXIN2* | rs11655966 | 36 | 594 | 832 | 0.98 (0.79-1.23) | 1.03 (0.68-1.57) | 0.9882 | 0.9884 |
| MMCT-18* | *AXIN2* | rs4541111 | 35 | 594 | 830 | 1.24 (0.96-1.60) | 1.10 (0.82-1.49) | 0.5038 | 0.511 |
| MMCT-18* | *AXIN2* | rs4791171 | 67 | 950 | 1735 | 1.08 (0.91-1.28) | 1.20 (0.89-1.60) | 0.1827 | 0.1976 |
| MMCT-18* | *AXIN2* | rs11079571 | 36 | 603 | 831 | 1.08 (0.86-1.37) | 1.68 (0.87-3.25) | 0.1748 | 0.1831 |
| MMCT-18* | *AXIN2* | rs3923087 | 75 | 1041 | 1961 | 1.03 (0.87-1.21) | 1.37 (0.95-1.98) | 0.2081 | 0.2255 |
| MMCT-18* | *AXIN2* | rs3923086 | 75 | 1039 | 1990 | 1.01 (0.86-1.20) | 1.13 (0.91-1.42) | 0.329 | 0.3467 |
| **MMCT-18*** | ***CASP5*** | **rs518604** | 76 | 1041 | 2029 | 1.11 (0.93-1.33) | **1.27 (1.02-1.58)** | **0.032** | **0.0387** |
| MMCT-18* | *CASP5* | rs523104 | 75 | 1017 | 2006 | 0.92 (0.77-1.10) | 0.89 (0.71-1.11) | 0.2633 | 0.2803 |
| MMCT-18* | *CASP5* | rs3181328 | 75 | 1039 | 1981 | 0.98 (0.79-1.21) | 0.63 (0.32-1.26) | 0.3921 | 0.4072 |
| MMCT-18* | *CASP5* | rs17446518 | 34 | 574 | 805 | 0.82 (0.62-1.08) | 1.55 (0.56-4.30) | 0.3641 | 0.3707 |
| MMCT-18* | *CASP5* | rs9651713 | 74 | 1029 | 1959 | 1.02 (0.84-1.24) | 1.39 (0.67-2.90) | 0.5982 | 0.6117 |
| MMCT-18* | *CASP5* | rs3181175 | 76 | 1046 | 2001 | 0.91 (0.77-1.07) | 1.14 (0.74-1.76) | 0.5699 | 0.5832 |
| MMCT-18* | *CASP5* | rs3181174 | 76 | 1044 | 2006 | 0.85 (0.68-1.06) | 0.93 (0.35-2.47) | 0.1687 | 0.1814 |
| MMCT-18* | *CASP5* | rs2282657 | 67 | 958 | 1739 | 0.97 (0.82-1.15) | 1.01 (0.78-1.32) | 0.9472 | 0.9488 |
| MMCT-18* | *CASP5* | rs507879 | 75 | 1038 | 1957 | 1.04 (0.85-1.27) | 1.05 (0.85-1.31) | 0.6488 | 0.6603 |
| MMCT-18* | *FILIP1* | rs796977 | 40 | 437 | 1166 | 1.11 (0.88-1.40) | 1.28 (0.91-1.81) | 0.1458 | 0.1548 |
| MMCT-18* | *FILIP1* | rs793477 | 66 | 944 | 1705 | 0.92 (0.76-1.12) | 0.72 (0.37-1.40) | 0.2396 | 0.2528 |
| MMCT-18* | *FILIP1* | rs793446 | 76 | 1045 | 1997 | 1.06 (0.90-1.26) | 1.10 (0.88-1.38) | 0.359 | 0.3766 |
| MMCT-18* | *FILIP1* | rs3921767 | 75 | 1040 | 1956 | 1.10 (0.88-1.37) | 0.74 (0.26-2.13) | 0.5993 | 0.613 |
| MMCT-18* | *FILIP1* | rs17338680 | 77 | 1046 | 2027 | 1.05 (0.87-1.28) | 0.70 (0.37-1.32) | 0.8847 | 0.8887 |
| MMCT-18* | *FILIP1* | rs9864437 | 76 | 1047 | 2009 | 0.95 (0.80-1.11) | 1.15 (0.83-1.60) | 0.9395 | 0.9417 |
| MMCT-18* | *FILIP1* | rs6788750 | 57 | 711 | 1593 | 0.99 (0.82-1.21) | 0.91 (0.70-1.18) | 0.523 | 0.5341 |
| MMCT-18* | *FILIP1* | rs12494994 | 75 | 1038 | 1971 | 1.13 (0.95-1.33) | 0.98 (0.63-1.52) | 0.3168 | 0.3352 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| MMCT-18* | *RBBP8* | rs7239066 | 76 | 1040 | 1993 | **0.81 (0.67-0.98)** | 1.20 (0.62-2.32) | 0.1062 | 0.117 |
| MMCT-18* | *RBBP8* | rs11082221 | 76 | 1038 | 1991 | 1.00 (0.74-1.34) | 1.23 (0.29-5.24) | 0.9388 | 0.942 |
| MMCT-18* | *RBBP8* | rs4474794 | 74 | 1033 | 1947 | 0.93 (0.79-1.10) | 0.95 (0.75-1.21) | 0.5173 | 0.5318 |
| MMCT-18* | *RBBP8* | rs9304261 | 31 | 346 | 888 | 1.08 (0.82-1.40) | 0.66 (0.38-1.15) | 0.5163 | 0.5212 |
| MMCT-18* | *RGC32* | rs10467472 | 74 | 1045 | 1943 | 1.04 (0.87-1.26) | 1.24 (0.70-2.17) | 0.4527 | 0.4696 |
| MMCT-18* | *RGC32* | rs3783194 | 68 | 956 | 1762 | 1.00 (0.82-1.22) | 1.75 (0.85-3.59) | 0.4929 | 0.5077 |
| MMCT-18* | *RGC32* | rs11618371 | 76 | 1047 | 2009 | 1.04 (0.85-1.26) | 1.40 (0.70-2.80) | 0.4719 | 0.489 |
| MMCT-18* | *RGC32* | rs9532824 | 75 | 1051 | 1949 | 0.88 (0.71-1.10) | 0.42 (0.14-1.24) | 0.0912 | 0.1003 |
| MMCT-18* | *RGC32* | rs995845 | 76 | 1040 | 1995 | 1.01 (0.86-1.19) | 1.16 (0.85-1.58) | 0.4767 | 0.493 |
| MMCT-18* | *RGC32* | rs9594551 | 74 | 1043 | 1929 | 0.95 (0.80-1.14) | 1.20 (0.70-2.06) | 0.9182 | 0.921 |
| MMCT-18* | *RGC32* | rs975590 | 75 | 1033 | 1994 | 1.00 (0.85-1.18) | 0.76 (0.53-1.08) | 0.3487 | 0.3652 |
| MMCT-18* | *RUVBL1* | rs9860614 | 76 | 1039 | 2012 | 1.04 (0.86-1.25) | 1.43 (0.83-2.45) | 0.3153 | 0.3344 |
| **MMCT-18*** | ***RUVBL1*** | **rs13063604** | 34 | 564 | 785 | 1.23 (0.98-1.56) | **1.54 (1.00-2.39)** | **0.016** | **0.0181** |
| MMCT-18* | *RUVBL1* | rs3732402 | 76 | 1045 | 2005 | **1.24 (1.05-1.46)** | 1.17 (0.93-1.48) | 0.0554 | 0.0649 |
| MMCT-18* | *RUVBL1* | rs7650365 | 75 | 1009 | 2010 | 1.17 (0.97-1.40) | 0.87 (0.70-1.09) | 0.2912 | 0.3085 |
| MMCT-18* | *RUVBL1* | rs4857836 | 77 | 1049 | 2031 | 1.08 (0.93-1.27) | 1.13 (0.83-1.54) | 0.2526 | 0.2708 |
| MMCT-18* | *RUVBL1* | rs9821568 | 75 | 1028 | 1971 | 1.02 (0.86-1.22) | 1.10 (0.65-1.85) | 0.6988 | 0.7092 |
| MMCT-18* | *STAG3* | rs11762932 | 76 | 1047 | 2007 | 0.96 (0.82-1.13) | 1.00 (0.71-1.41) | 0.7446 | 0.7533 |
| MMCT-18* | *STAG3* | rs2246713 | 35 | 591 | 824 | 0.88 (0.68-1.14) | 0.89 (0.66-1.20) | 0.4251 | 0.4329 |
| MMCT-18* | *STAG3* | rs1637001 | 76 | 1047 | 2005 | 0.83 (0.71-0.97) | 0.88 (0.65-1.18) | 0.0512 | 0.0592 |
| OCAC§ | *11q13.2* | rs7931342 | 98 | 1466 | 2474 | 0.94 (0.80-1.11) | 0.93 (0.77-1.12) | 0.4153 | 0.4369 |
| OCAC§ | *17q22* | rs7501993 | 97 | 1454 | 2456 | 0.90 (0.78-1.05) | 1.02 (0.84-1.24) | 0.8645 | 0.8706 |
| OCAC§ | *17q24.3* | rs1859962 | 97 | 1446 | 2464 | 1.01 (0.86-1.19) | 0.98 (0.81-1.18) | 0.8063 | 0.815 |
| OCAC§ | *3p12.1* | rs2660753 | 96 | 1437 | 2428 | 1.14 (0.96-1.36) | 1.53 (0.80-2.94) | 0.0628 | 0.0755 |
| OCAC§ | *8q24.21* | rs7000448 | 98 | 1472 | 2453 | 1.09 (0.95-1.26) | 0.96 (0.78-1.19) | 0.8229 | 0.8308 |
| OCAC§ | *Xp11.22* | rs5945619 | 98 | 1457 | 2479 | 1.00 (0.87-1.15) | 1.08 (0.88-1.33) | 0.56 | 0.5781 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| OCAC§ | *ABL1* | rs2855192 | 98 | 1470 | 2470 | 0.98 (0.84-1.15) | **1.76 (1.07-2.88)** | 0.3501 | 0.3726 |
| OCAC§ | *BRCA2* | rs144848 | 93 | 1459 | 2294 | 0.94 (0.82-1.08) | 1.03 (0.79-1.33) | 0.6796 | 0.6929 |
| OCAC§ | *CDC2* | rs2448343 | 96 | 1448 | 2420 | 1.04 (0.90-1.20) | 0.95 (0.77-1.17) | 0.839 | 0.8462 |
| OCAC§ | *CDK7* | rs12656449 | 98 | 1471 | 2476 | 1.02 (0.85-1.22) | 0.75 (0.34-1.66) | 0.9306 | 0.9338 |
| OCAC§ | *CRCAC* | rs10795668 | 96 | 1428 | 2417 | 1.02 (0.89-1.17) | 0.92 (0.73-1.16) | 0.7112 | 0.7235 |
| OCAC§ | *CRCAC* | rs16892766 | 96 | 1439 | 2425 | 1.17 (0.98-1.41) | 0.91 (0.42-1.95) | 0.1575 | 0.177 |
| OCAC§ | *CTBP2* | rs12769019 | 56 | 996 | 1259 | 0.94 (0.79-1.12) | 1.24 (0.91-1.69) | 0.6017 | 0.6116 |
| OCAC§ | *DESP-1979* | rs16901979 | 97 | 1463 | 2428 | 0.98 (0.75-1.27) | 0.96 (0.16-5.86) | 0.8539 | 0.8605 |
| OCAC§ | *DNMT3A* | rs13420827 | 98 | 1462 | 2473 | 0.96 (0.83-1.11) | 1.13 (0.82-1.56) | 0.9359 | 0.9388 |
| OCAC§ | *DPYD* | rs1801265 | 97 | 1464 | 2444 | 1.00 (0.87-1.15) | 1.05 (0.76-1.45) | 0.8433 | 0.8503 |
| OCAC§ | *E2F2* | rs760607 | 96 | 1251 | 2109 | **2.94 (1.38-6.23)** | **2.91 (1.37-6.15)** | 0.8924 | 0.8972 |
| OCAC§ | *E2F3* | rs7760528 | 98 | 1462 | 2465 | 1.03 (0.90-1.18) | 0.92 (0.74-1.16) | 0.7598 | 0.7704 |
| OCAC§ | *EHBP1* | rs2710646 | 95 | 1454 | 2355 | 0.94 (0.81-1.08) | 1.32 (0.95-1.83) | 0.7223 | 0.7342 |
| OCAC§ | *ESR1* | rs712221 | 97 | 1445 | 2454 | 1.08 (0.93-1.25) | 1.12 (0.92-1.37) | 0.2115 | 0.233 |
| **OCAC§** | ***ESR1*** | **rs9322336** | 98 | 1453 | 2464 | **0.81 (0.70-0.93)** | 0.73 (0.52-1.02) | **0.0013** | **0.0021** |
| OCAC§ | *FANCE* | rs2395626 | 98 | 1460 | 2480 | 1.06 (0.91-1.23) | 0.98 (0.81-1.18) | 0.971 | 0.9723 |
| OCAC§ | *5q35.3* | SNP4 | 95 | 1461 | 2370 | 1.04 (0.90-1.20) | 1.02 (0.83-1.25) | 0.7217 | 0.7336 |
| OCAC§ | *5q35.3* | SNP3 | 95 | 1462 | 2371 | 0.96 (0.82-1.12) | 0.90 (0.74-1.09) | 0.2835 | 0.3055 |
| OCAC§ | *5q35.3* | SNP2 | 95 | 1467 | 2359 | 0.96 (0.84-1.11) | 0.75 (0.56-1.00) | 0.1085 | 0.1251 |
| OCAC§ | *5q35.3* | SNP1 | 97 | 1475 | 2411 | 0.92 (0.80-1.06) | 0.82 (0.58-1.14) | 0.1207 | 0.1384 |
| OCAC§ | *GALNTL2* | rs2271077 | 98 | 1460 | 2479 | 0.96 (0.77-1.21) | 0.97 (0.24-3.97) | 0.7433 | 0.7546 |
| OCAC§ | *11p15.5* | SNP3 | 96 | 1457 | 2396 | 0.87 (0.75-1.01) | 0.85 (0.61-1.17) | 0.0503 | 0.0616 |
| OCAC§ | *11p15.5* | SNP2 | 96 | 1459 | 2407 | **1.20 (1.02-1.40)** | **1.20 (1.00-1.44)** | **0.0473** | 0.0581 |
| **OCAC§** | ***11p15.5*** | **SNP1** | 97 | 1473 | 2402 | **0.84 (0.73-0.97)** | 0.86 (0.60-1.22) | **0.0243** | **0.0314** |
| **OCAC§** | ***IL18*** | **rs1834481** | 97 | 1449 | 2435 | 0.89 (0.77-1.02) | 0.77 (0.59-1.01) | **0.0227** | **0.0295** |
| OCAC§ | *JAZF1* | rs10486567 | 96 | 1456 | 2408 | 1.00 (0.87-1.14) | 1.09 (0.83-1.45) | 0.7233 | 0.7352 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| OCAC§ | *KLK3* | rs2735839 | 98 | 1459 | 2481 | 0.92 (0.79-1.08) | 1.38 (0.90-2.11) | 0.9525 | 0.9547 |
| OCAC§ | *KU70* | rs132788 | 86 | 1382 | 2082 | 1.04 (0.90-1.21) | 1.08 (0.87-1.36) | 0.4333 | 0.4522 |
| OCAC§ | *LIG4* | rs1805386 | 98 | 1462 | 2466 | 1.08 (0.93-1.26) | 1.24 (0.85-1.83) | 0.1498 | 0.1693 |
| OCAC§ | *LMTK2* | rs6465657 | 97 | 1439 | 2454 | 0.94 (0.80-1.09) | 1.00 (0.83-1.21) | 0.9184 | 0.9221 |
| OCAC§ | *MSMB* | rs10993994 | 98 | 1454 | 2473 | 1.03 (0.89-1.19) | 1.08 (0.88-1.32) | 0.4583 | 0.479 |
| OCAC§ | *MTHFD1* | rs1950902 | 97 | 1451 | 2458 | 0.97 (0.84-1.12) | 0.89 (0.62-1.27) | 0.4952 | 0.515 |
| OCAC§ | *MTHFS* | rs17284990 | 96 | 1418 | 2416 | 0.98 (0.85-1.13) | 1.06 (0.79-1.44) | 0.9714 | 0.9727 |
| OCAC§ | *NRIP1* | rs2822986 | 98 | 1458 | 2458 | 1.03 (0.89-1.18) | 0.89 (0.71-1.10) | 0.5123 | 0.5317 |
| OCAC§ | *PGR* | rs10895068 | 97 | 1462 | 2413 | 1.00 (0.81-1.24) | 0.49 (0.13-1.80) | 0.7231 | 0.7351 |
| **OCAC§** | ***PGR*** | **rs1042838** | 95 | 1424 | 2408 | **1.25 (1.07-1.46)** | 1.09 (0.73-1.64) | **0.0161** | **0.0215** |
| OCAC§ | *PGR* | rs608995 | 98 | 1475 | 2473 | 1.08 (0.94-1.24) | 1.12 (0.84-1.48) | 0.2427 | 0.265 |
| OCAC§ | *SLC22A3* | rs9364554 | 96 | 1449 | 2390 | 1.07 (0.93-1.24) | 1.19 (0.94-1.51) | 0.1223 | 0.1399 |
| OCAC§ | *TGFB1* | rs1982073 | 57 | 1026 | 1262 | 0.98 (0.82-1.17) | 0.95 (0.74-1.23) | 0.7069 | 0.7146 |
| OCAC§ | *TNRC9* | rs3803662 | 97 | 1456 | 2427 | 0.94 (0.82-1.08) | 0.85 (0.65-1.11) | 0.1849 | 0.2055 |
| OCAC§ | *TP53* | rs2287498 | 98 | 1459 | 2471 | 1.03 (0.86-1.24) | 2.20 (0.95-5.08) | 0.3145 | 0.3371 |
| OCAC§ | *TP53* | rs1042522 | 93 | 1459 | 2291 | 0.97 (0.84-1.11) | 1.11 (0.85-1.45) | 0.8221 | 0.8297 |
| OCAC§ | *TP53* | rs12951053 | 96 | 1428 | 2430 | 1.03 (0.86-1.24) | 1.48 (0.71-3.11) | 0.4635 | 0.4838 |
| OCAC§ | *TP53* | rs1625895 | 88 | 1409 | 2136 | 1.05 (0.89-1.24) | 0.85 (0.48-1.51) | 0.8096 | 0.8173 |
| OCAC§ | *TP53* | rs9894946 | 99 | 1470 | 2485 | 1.09 (0.94-1.27) | 1.02 (0.65-1.58) | 0.3293 | 0.3519 |
| OCAC§ | *TYMS* | rs495139 | 97 | 1433 | 2469 | 1.09 (0.94-1.26) | 0.93 (0.76-1.14) | 0.8128 | 0.8212 |
| OCAC§ | *12q13.11* | SNP1 | 98 | 1461 | 2455 | 1.06 (0.92-1.23) | 1.06 (0.86-1.29) | 0.4679 | 0.4884 |
| OCAC§ | *6p21.1* | SNP1 | 97 | 1475 | 2418 | 1.07 (0.91-1.25) | 1.17 (0.97-1.41) | 0.0943 | 0.1102 |
| OCAC§ | *XRCC5* | rs16855489 | 99 | 1476 | 2484 | 1.05 (0.91-1.21) | 1.02 (0.83-1.25) | 0.7089 | 0.7217 |
| Oncogene | *BRAF* | rs10487888 | 99 | 1480 | 2481 | 1.02 (0.87-1.19) | 0.88 (0.73-1.07) | 0.2443 | 0.2666 |
| Oncogene | *BRAF* | rs1733832 | 80 | 1149 | 2043 | 1.07 (0.85-1.34) | 3.40 (0.97-11.92) | 0.2404 | 0.2585 |
| Oncogene | *BRAF* | rs1267622 | 97 | 1458 | 2443 | 0.96 (0.83-1.10) | 0.83 (0.62-1.12) | 0.2461 | 0.2682 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Oncogene | *BRAF* | rs13241719 | 83 | 1304 | 2032 | 1.05 (0.91-1.22) | 0.89 (0.69-1.14) | 0.7355 | 0.7455 |
| Oncogene | *BRAF* | rs17695623 | 96 | 1441 | 2434 | 1.24 (0.61-2.51) | 1.28 (0.62-2.64) | 0.8847 | 0.8898 |
| Oncogene | *BRAF* | rs17161747 | 98 | 1473 | 2467 | 1.19 (0.96-1.47) | 1.38 (0.52-3.62) | 0.094 | 0.11 |
| Oncogene | *BRAF* | rs17623382 | 97 | 1465 | 2439 | 0.98 (0.83-1.15) | 1.01 (0.59-1.74) | 0.8527 | 0.8593 |
| Oncogene | *BRAF* | rs6944385 | 98 | 1471 | 2464 | 1.15 (0.99-1.35) | 0.98 (0.61-1.57) | 0.1463 | 0.1657 |
| Oncogene | *ERBB2* | rs2952155 | 98 | 1463 | 2462 | 1.01 (0.87-1.15) | 1.13 (0.84-1.52) | 0.5785 | 0.596 |
| Oncogene | *ERBB2* | rs2952156 | 98 | 1470 | 2473 | 1.00 (0.87-1.14) | 1.17 (0.90-1.52) | 0.4481 | 0.4691 |
| Oncogene | *ERBB2* | rs1801200 | 97 | 1455 | 2434 | 1.06 (0.92-1.22) | 1.07 (0.80-1.42) | 0.4236 | 0.4449 |
| Oncogene | *KRAS* | rs12305513 | 99 | 1483 | 2472 | 0.89 (0.75-1.06) | 0.83 (0.42-1.62) | 0.1668 | 0.1872 |
| Oncogene | *KRAS* | rs12822857 | 97 | 1465 | 2446 | 1.01 (0.86-1.18) | 0.99 (0.82-1.19) | 0.9441 | 0.9467 |
| Oncogene | *KRAS* | rs10842508 | 98 | 1476 | 2469 | 0.97 (0.84-1.11) | 0.99 (0.75-1.31) | 0.7587 | 0.7694 |
| Oncogene | *KRAS* | rs12579073 | 97 | 1460 | 2442 | 0.96 (0.82-1.12) | 0.94 (0.78-1.14) | 0.5369 | 0.5555 |
| Oncogene | *KRAS* | rs10842513 | 97 | 1469 | 2421 | 0.98 (0.82-1.17) | 0.84 (0.40-1.75) | 0.6988 | 0.7118 |
| Oncogene | *KRAS* | rs4623993 | 79 | 1140 | 2030 | 1.05 (0.88-1.24) | 1.17 (0.73-1.89) | 0.4498 | 0.4669 |
| Oncogene | *KRAS* | rs6487464 | 98 | 1470 | 2448 | 1.03 (0.89-1.19) | 1.02 (0.84-1.24) | 0.7956 | 0.8047 |
| Oncogene | *KRAS* | rs10842514 | 98 | 1473 | 2442 | 0.96 (0.83-1.12) | 1.12 (0.93-1.35) | 0.3125 | 0.335 |
| Oncogene | *KRAS* | rs11047917 | 98 | 1465 | 2456 | 0.88 (0.71-1.08) | 1.22 (0.42-3.50) | 0.311 | 0.3335 |
| Oncogene | *NMI* | rs394884 | 95 | 1412 | 2390 | 1.07 (0.92-1.26) | 0.99 (0.53-1.86) | 0.4338 | 0.4544 |
| Oncogene | *NMI* | rs11551174 | 79 | 1150 | 2040 | 0.98 (0.77-1.24) | 1.24 (0.45-3.42) | 0.9861 | 0.9866 |
| Oncogene | *NMI* | rs289831 | 96 | 1422 | 2403 | 1.07 (0.87-1.33) | 0.84 (0.40-1.79) | 0.5856 | 0.6026 |
| Oncogene | *NMI* | rs3771886 | 98 | 1464 | 2465 | 1.00 (0.87-1.16) | 1.15 (0.95-1.40) | 0.2006 | 0.2219 |
| Oncogene | *NMI* | rs11683487 | 83 | 1172 | 2149 | **0.78 (0.66-0.93)** | 0.85 (0.70-1.04) | 0.0668 | 0.0782 |
| Oncogene | *NMI* | rs2113509 | 97 | 1459 | 2452 | 1.09 (0.93-1.28) | 1.05 (0.59-1.85) | 0.3195 | 0.342 |
| Oncogene | *PIK3CA* | rs2865084 | 80 | 1154 | 2046 | 1.14 (0.89-1.45) | 0.91 (0.26-3.17) | 0.356 | 0.3744 |
| Oncogene | *PIK3CA* | rs7621329 | 97 | 1466 | 2424 | 1.02 (0.88-1.18) | 1.18 (0.79-1.76) | 0.557 | 0.575 |
| Oncogene | *PIK3CA* | rs1517586 | 96 | 1429 | 2434 | 0.89 (0.74-1.06) | 0.77 (0.39-1.51) | 0.143 | 0.1618 |

| Pathway / Group | Gene/cytoband | SNP | Call rate (%)** | Cases | Controls | HetOR (95% CI) | HomOR (95% CI) | *P*-trend unadj | *P*-trend adj.‡ |
|---|---|---|---|---|---|---|---|---|---|
| Oncogene | *PIK3CA* | rs2699905 | 97 | 1445 | 2439 | 0.99 (0.86-1.14) | 1.04 (0.82-1.32) | 0.8746 | 0.8802 |
| Oncogene | *PIK3CA* | rs7641889 | 98 | 1478 | 2463 | 0.87 (0.71-1.07) | 1.22 (0.51-2.91) | 0.297 | 0.3196 |
| Oncogene | *PIK3CA* | rs7651265 | 97 | 1449 | 2435 | 0.91 (0.77-1.08) | 1.33 (0.73-2.42) | 0.5891 | 0.606 |
| Oncogene | *PIK3CA* | rs7640662 | 98 | 1475 | 2472 | 1.06 (0.91-1.23) | 0.88 (0.57-1.37) | 0.7612 | 0.7719 |
| Oncogene | *PIK3CA* | rs2677760 | 97 | 1451 | 2446 | 0.99 (0.84-1.16) | 1.08 (0.90-1.30) | 0.4223 | 0.4436 |

‡ Adjusted for population stratification
* GEOCS not done
**Call rate bases on total number of samples analysed.
§ candidate genes identified from the Breast Cancer Association Consortium (BCAC) and Ovarian Cancer Association Consortium (OCAC);
Emboldened P-values are significant at the 5% level. The P-trend looks for a trend between the OR and the heterozygous (Het); and rare homozygous (Hom) when compared with the common homozygous; the P-heterogeneity (P-het) does not assume a correlation with increasing number of rare allele.enboldened odds ratios (OR) do not cross 1.

## Appendix VII-A: Univariate and multivariate Cox regression results of *BRAF* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 1680 | All | 1 (0.89-1.12) | 0.972 | 0.94 (0.82-1.07) | 0.357 | 6 |
| | | | 804 | Serous | 1.05 (0.9-1.21) | 0.549 | 0.9 (0.8-1.1) | 0.468 | **14** |
| *BRAF* | rs10487888 | 0.46 | 251 | Endometrioid | 0.97 (0.67-1.39) | 0.852 | 0.87 (0.59-1.28) | 0.466 | 10 |
| | | | 180 | Mucinous | 1.1 (0.68-1.78) | 0.71 | 1.01 (0.65-1.59) | 0.95 | 8 |
| | | | 95 | Clear cell | 0.74 (0.43-1.28) | 0.277 | 1 (0.54-1.84) | 0.999 | **35** |
| | | | 1159 | All | 1.01 (0.79-1.29) | 0.936 | 1.28 (0.96-1.72) | 0.095 | **27** |
| | | | 525 | Serous | 0.99 (0.72-1.37) | 0.966 | 1.2 (0.8-1.8) | 0.291 | **21** |
| *BRAF* | rs1733832 | 0.07 | 182 | Endometrioid | 1.03 (0.51-2.08) | 0.94 | 1.67 (0.76-3.69) | 0.201 | **62** |
| | | | 135 | Mucinous | 0.93 (0.39-2.25) | 0.879 | 0.92 (0.38-2.22) | 0.856 | 1 |
| | | | 94 | Clear cell | 1.72 (0.66-4.46) | 0.264 | 1.67 (0.52-5.36) | 0.389 | 3 |
| | | | 1751 | **All** | 1.12 (0.99-1.27) | 0.077 | **1.19 (1.03-1.38)** | **0.02** | 6 |
| | | | 831 | Serous | 1.03 (0.87-1.22) | 0.727 | 1.2 (1-1.4) | 0.134 | **17** |
| ***BRAF*** | **rs1267622** | 0.23 | 268 | Endometrioid | 1.07 (0.72-1.59) | 0.733 | 1.1 (0.71-1.71) | 0.655 | 3 |
| | | | 187 | Mucinous | 1.08 (0.64-1.82) | 0.764 | 0.89 (0.54-1.49) | 0.663 | **18** |
| | | | 123 | Clear cell | 1.27 (0.7-2.3) | 0.429 | 1 (0.51-1.98) | 0.997 | **21** |
| | | | 1602 | **All** | 0.97 (0.85-1.1) | 0.606 | **0.79 (0.67-0.93)** | **0.004** | **19** |
| | | | 733 | **Serous** | 0.94 (0.79-1.12) | 0.507 | **0.8 (0.6-0.9)** | **0.006** | **15** |
| ***BRAF*** | **rs13241719** | 0.31 | 246 | Endometrioid | 0.77 (0.49-1.21) | 0.258 | 0.79 (0.48-1.29) | 0.339 | 3 |
| | | | 176 | Mucinous | 0.99 (0.58-1.69) | 0.98 | 1.05 (0.63-1.74) | 0.852 | 6 |
| | | | 135 | Clear cell | 1.12 (0.59-2.15) | 0.723 | 1.4 (0.64-3.06) | 0.404 | **25** |
| *BRAF* | rs17695623 | 0.07 | 1744 | All | 1.16 (0.94-1.43) | 0.174 | 0.98 (0.76-1.27) | 0.89 | **16** |
| | | | 829 | Serous | 0.93 (0.7-1.23) | 0.601 | 0.9 (0.6-1.2) | 0.472 | 3 |
| | | | 264 | Endometrioid | 1.23 (0.61-2.49) | 0.569 | 0.98 (0.43-2.22) | 0.955 | **20** |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 186 | Mucinous | 0.78 (0.21-2.97) | 0.721 | 0.29 (0.07-1.25) | 0.098 | **63** |
| | | | 132 | Clear cell | 1.77 (0.82-3.84) | 0.146 | 1.31 (0.57-3.04) | 0.523 | **26** |
| *BRAF* | rs17161747 | 0.06 | 1771 | All | 0.83 (0.66-1.05) | 0.119 | 0.84 (0.65-1.1) | 0.209 | 1 |
| | | | 847 | Serous | 0.77 (0.57-1.03) | 0.079 | 0.8 (0.6-1.1) | 0.146 | 4 |
| | | | 272 | Endometrioid | 0.97 (0.45-2.06) | 0.928 | 1.29 (0.61-2.7) | 0.506 | **33** |
| | | | 191 | Mucinous | 0.99 (0.38-2.56) | 0.981 | 1.64 (0.61-4.46) | 0.329 | **66** |
| | | | 134 | Clear cell | 0.83 (0.28-2.42) | 0.734 | 1.01 (0.29-3.51) | 0.986 | **22** |
| *BRAF* | rs17623382 | 0.12 | 1764 | All | 1.08 (0.92-1.28) | 0.333 | 0.98 (0.81-1.19) | 0.874 | 9 |
| | | | 841 | Serous | 1.09 (0.89-1.35) | 0.408 | 1 (0.8-1.2) | 0.935 | 8 |
| | | | 270 | Endometrioid | 0.73 (0.36-1.49) | 0.382 | 0.74 (0.35-1.53) | 0.413 | 1 |
| | | | 186 | Mucinous | 1.54 (0.83-2.87) | 0.17 | 1.86 (0.96-3.58) | 0.065 | **21** |
| | | | 133 | Clear cell | 0.74 (0.33-1.69) | 0.479 | 0.5 (0.17-1.45) | 0.201 | **32** |
| ***BRAF*** | **rs6944385** | **0.14** | 1758 | **All** | **1.19 (1.03-1.38)** | **0.021** | **1.25 (1.05-1.5)** | **0.013** | 5 |
| | | | 840 | Serous | 0.97 (0.79-1.2) | 0.804 | 1.1 (0.9-1.3) | 0.516 | **13** |
| | | | 268 | Endometrioid | 1.31 (0.84-2.07) | 0.235 | 1.43 (0.87-2.35) | 0.156 | 9 |
| | | | 187 | Mucinous | 0.83 (0.4-1.73) | 0.614 | 0.76 (0.36-1.62) | 0.477 | 8 |
| | | | 124 | **Clear cell** | **2.22 (1.18-4.17)** | **0.014** | 1.93 (0.95-3.92) | 0.07 | **13** |
| ***BRAF*** | **rs1267622, rs6944385; AA** | 73.3* | 1786 | **All** | 0.89 (0.79-1.01) | 0.076 | **0.84 (0.72-0.97)** | **0.018** | 6 |
| | | | 724 | Serous | 0.97 (0.82-1.14) | 0.708 | **0.9 (0.7-1)** | 0.115 | 7 |
| | | | 246 | Endometrioid | 0.9 (0.61-1.34) | 0.611 | 0.87 (0.57-1.34) | 0.532 | 3 |
| | | | 169 | Mucinous | 0.94 (0.56-1.58) | 0.82 | 1.18 (0.7-1.98) | 0.528 | **26** |
| | | | 126 | Clear cell | 0.79 (0.44-1.43) | 0.434 | 1 (0.51-1.98) | 0.999 | **27** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant. * Haplotype frequency

## Appendix VII-B: Univariate and multivariate Cox regression results of *BRAF* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *BRAF* | h10000000 | 22.9 | All | 0.98 (0.85-1.13) | 0.76 | 1.09 (0.92-1.3) | 0.335 | **11** |
| | | | Serous | 1.06 (0.89-1.28) | 0.502 | 0.4 (0.1-1.5) | 0.16 | **62** |
| | | | Endometrioid | 1.24 (0.79-1.95) | 0.358 | 1.06 (0.63-1.79) | 0.819 | **15** |
| | | | Mucinous | 1.09 (0.6-1.98) | 0.767 | 0.93 (0.49-1.76) | 0.812 | **15** |
| | | | Clear cell | 0.63 (0.31-1.31) | 0.217 | 0.87 (0.39-1.91) | 0.723 | **38** |
| ***BRAF*** | **h10010000** | 18.8 | **All** | 0.95 (0.81-1.1) | 0.493 | **0.8 (0.66-0.95)** | **0.014** | **16** |
| | | | Serous | 0.95 (0.79-1.16) | 0.633 | 1.1 (0.9-1.4) | 0.415 | **16** |
| | | | Endometrioid | 0.88 (0.54-1.42) | 0.591 | 0.88 (0.53-1.48) | 0.629 | 0 |
| | | | Mucinous | 0.63 (0.31-1.28) | 0.204 | 0.62 (0.32-1.23) | 0.173 | 2 |
| | | | Clear cell | 1.38 (0.61-3.08) | 0.438 | 2.38 (0.92-6.15) | 0.074 | **72** |
| ***BRAF*** | **h00000000** | 16.4 | All | 0.94 (0.81-1.1) | 0.443 | 1 (0.84-1.19) | 0.962 | 6 |
| | | | Serous | 0.98 (0.81-1.18) | 0.813 | 1 (0.8-1.3) | 0.944 | 2 |
| | | | Endometrioid | 0.99 (0.63-1.55) | 0.952 | 0.98 (0.61-1.56) | 0.917 | 1 |
| | | | Mucinous | 0.84 (0.43-1.65) | 0.618 | 1.31 (0.61-2.86) | 0.49 | **56** |
| | | | **Clear cell** | **1.24 (0.62-2.45)** | **0.542** | **0.93 (0.4-2.19)** | **0.874** | **25** |
| ***BRAF*** | **h10010010** | 12.2 | All | 1.1 (0.94-1.3) | 0.238 | 1 (0.83-1.21) | 0.96 | 9 |
| | | | **Serous** | 1.12 (0.91-1.37) | 0.296 | **0.8 (0.6-1)** | **0.037** | **29** |
| | | | Endometrioid | 0.74 (0.36-1.5) | 0.401 | 0.74 (0.36-1.54) | 0.425 | 0 |
| | | | Mucinous | 1.5 (0.81-2.8) | 0.199 | 1.91 (0.96-3.78) | 0.065 | **27** |
| | | | Clear cell | 0.79 (0.35-1.75) | 0.557 | 0.62 (0.23-1.7) | 0.351 | **22** |
| *BRAF* | h00100000 | 8.7 | All | 0.97 (0.79-1.2) | 0.804 | 1.09 (0.86-1.39) | 0.48 | **12** |
| | | | Serous | 1.11 (0.86-1.44) | 0.426 | 0.8 (0.6-1.1) | 0.161 | **28** |
| | | | Endometrioid | 0.86 (0.42-1.74) | 0.669 | 0.78 (0.36-1.65) | 0.511 | 9 |

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Mucinous | 1.4 (0.64-3.07) | 0.4 | 1.1 (0.46-2.61) | 0.833 | **21** |
| | | | Clear cell | 0.18 (0.02-1.32) | 0.092 | 0.19 (0.03-1.28) | 0.087 | 6 |
| | | | **All** | **1.21 (1-1.46)** | 0.055 | **1.43 (1.14-1.8)** | **0.002** | **18** |
| | | | Serous | 1.1 (0.85-1.42) | 0.483 | 0.9 (0.7-1.3) | 0.636 | **18** |
| ***BRAF*** | **h01100001** | 7.1 | Endometrioid | 1.3 (0.71-2.4) | 0.393 | 2.04 (1.05-3.99) | 0.036 | **57** |
| | | | Mucinous | 0.81 (0.33-1.99) | 0.652 | 0.9 (0.38-2.1) | 0.804 | **11** |
| | | | Clear cell | 1.86 (0.84-4.13) | 0.127 | 1.92 (0.74-4.96) | 0.179 | 3 |
| | | | All | 1.17 (0.95-1.44) | 0.147 | 1.01 (0.79-1.3) | 0.933 | **14** |
| | | | Serous | 0.94 (0.71-1.24) | 0.66 | 1.3 (1-1.7) | 0.099 | **38** |
| *BRAF* | h00101001 | 7 | Endometrioid | 1.2 (0.59-2.44) | 0.611 | 0.95 (0.42-2.16) | 0.896 | **21** |
| | | | Mucinous | 0.77 (0.23-2.61) | 0.672 | 0.37 (0.1-1.36) | 0.133 | **52** |
| | | | Clear cell | 1.72 (0.79-3.75) | 0.171 | 1.32 (0.57-3.05) | 0.523 | **23** |
| | | | All | 0.84 (0.67-1.05) | 0.128 | 0.85 (0.65-1.11) | 0.232 | 1 |
| | | | Serous | 0.77 (0.58-1.04) | 0.091 | 1 (0.8-1.3) | 0.807 | **30** |
| *BRAF* | h00000100 | 6.2 | Endometrioid | 0.97 (0.45-2.07) | 0.935 | 1.3 (0.62-2.73) | 0.486 | **34** |
| | | | Mucinous | 1.04 (0.41-2.59) | 0.941 | 1.66 (0.62-4.45) | 0.314 | **60** |
| | | | Clear cell | 0.82 (0.28-2.46) | 0.729 | 1.01 (0.29-3.57) | 0.985 | **23** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *BRAF*: rs10487888, rs1733832, rs1267622, rs13241719, rs17695623, rs17161747, rs17623382, rs6944385.

## Appendix VII-C: Univariate and multivariate Cox regression results of *ERBB2* tSNPs

| Gene | tSNP | MAF | Cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *ERBB2* | rs2952155 | 0.24 | 1667 | All | 1.05 (0.92-1.2) | 0.451 | 1.11 (0.95-1.3) | 0.184 | 6 |
| | | | 795 | Serous | 1.14 (0.96-1.35) | 0.127 | **1.2 (1-1.5)** | 0.061 | 5 |
| | | | 250 | Endometrioid | 1.22 (0.8-1.86) | 0.358 | 1.2 (0.79-1.84) | 0.396 | 2 |
| | | | 177 | Mucinous | 1.11 (0.64-1.92) | 0.705 | 0.83 (0.46-1.47) | 0.517 | **25** |
| | | | 135 | Clear cell | 1.11 (0.6-2.06) | 0.73 | 0.97 (0.45-2.06) | 0.933 | **13** |
| *ERBB2* | rs2952156 | 0.3 | 1766 | All | 1.06 (0.94-1.2) | 0.323 | 1.09 (0.94-1.26) | 0.235 | 3 |
| | | | 840 | Serous | 1.12 (0.96-1.31) | 0.154 | **1.1 (1-1.4)** | 0.118 | 2 |
| | | | 269 | Endometrioid | 1.27 (0.85-1.89) | 0.239 | 1.24 (0.8-1.91) | 0.337 | 2 |
| | | | 186 | Mucinous | 1.08 (0.64-1.84) | 0.77 | 0.81 (0.45-1.46) | 0.489 | **25** |
| | | | 134 | Clear cell | 1.04 (0.58-1.84) | 0.906 | 0.88 (0.42-1.85) | 0.731 | **15** |
| *ERBB2* | rs1801200 | 0.23 | 1766 | All | 0.92 (0.81-1.05) | 0.216 | 1.02 (0.87-1.19) | 0.818 | **11** |
| | | | 847 | Serous | 1.04 (0.88-1.22) | 0.635 | 1.1 (0.9-1.3) | 0.372 | 6 |
| | | | 263 | Endometrioid | 0.81 (0.53-1.22) | 0.309 | 0.74 (0.47-1.19) | 0.217 | 9 |
| | | | 188 | Mucinous | 1.26 (0.7-2.27) | 0.448 | 1.19 (0.64-2.21) | 0.586 | 6 |
| | | | 136 | Clear cell | 0.9 (0.5-1.64) | 0.738 | 1.41 (0.71-2.78) | 0.328 | **57** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VII-D: Univariate and multivariate Cox regression results of *ERBB2* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *ERBB2* | h000 | 51.2 | All | 1.02 (0.91-1.14) | 0.745 | 0.95 (0.83-1.09) | 0.458 | 7 |
| | | | Serous | 0.92 (0.79-1.06) | 0.224 | 0.9 (0.2-3.6) | 0.928 | 2 |
| | | | Endometrioid | 0.98 (0.68-1.4) | 0.892 | 1.06 (0.7-1.6) | 0.795 | 8 |
| | | | Mucinous | 0.81 (0.46-1.44) | 0.483 | 1.06 (0.58-1.93) | 0.85 | **31** |
| | | | Clear cell | 1.08 (0.63-1.85) | 0.789 | 0.92 (0.51-1.68) | 0.793 | **15** |
| *ERBB2* | h110 | 18.4 | All | 1.12 (0.95-1.31) | 0.171 | 1.14 (0.94-1.38) | 0.171 | 2 |
| | | | Serous | 1.18 (0.97-1.44) | 0.103 | 1 (0.7-1.3) | 0.767 | **15** |
| | | | Endometrioid | 1.55 (0.9-2.67) | 0.111 | 1.57 (0.9-2.73) | 0.111 | 1 |
| | | | Mucinous | 0.96 (0.5-1.86) | 0.911 | 0.71 (0.35-1.44) | 0.343 | **26** |
| | | | Clear cell | 1.16 (0.56-2.42) | 0.684 | 0.99 (0.38-2.55) | 0.978 | **15** |
| *ERBB2* | h001 | 17.7 | All | 0.9 (0.77-1.05) | 0.181 | 0.98 (0.81-1.18) | 0.835 | 9 |
| | | | Serous | 1.02 (0.84-1.24) | 0.853 | 1.2 (0.9-1.4) | 0.179 | **18** |
| | | | Endometrioid | 0.73 (0.44-1.21) | 0.222 | 0.62 (0.34-1.13) | 0.12 | **15** |
| | | | Mucinous | 1.21 (0.59-2.47) | 0.604 | 1.32 (0.61-2.82) | 0.479 | 9 |
| | | | Clear cell | 0.87 (0.43-1.74) | 0.686 | 1.45 (0.69-3.03) | 0.327 | **67** |
| *ERBB2* | h010 | 6.5 | All | 1 (0.8-1.25) | 0.982 | 0.95 (0.74-1.21) | 0.682 | 5 |
| | | | Serous | 0.94 (0.73-1.21) | 0.622 | 1.1 (0.9-1.3) | 0.562 | **17** |
| | | | Endometrioid | 1.14 (0.45-2.88) | 0.776 | 0.75 (0.26-2.13) | 0.588 | **34** |
| | | | Mucinous | 1.19 (0.4-3.57) | 0.753 | 1.24 (0.41-3.77) | 0.699 | 4 |
| | | | Clear cell | 0.8 (0.23-2.74) | 0.722 | 0.58 (0.13-2.56) | 0.47 | **28** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *ERBB2*: rs2952155, rs2952156, rs1801200.

## Appendix VII-E: Univariate and multivariate Cox regression results of *KRAS* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 1788 | All | 1.01 (0.84-1.22) | 0.879 | 0.98 (0.79-1.23) | 0.878 | 3 |
| | | | 852 | Serous | 0.92 (0.73-1.17) | 0.498 | 0.9 (0.7-1.2) | 0.662 | 2 |
| *KRAS* | rs12305513 | 0.09 | 272 | Endometrioid | 1.39 (0.77-2.52) | 0.272 | 1 (0.49-2.03) | 0.998 | **28** |
| | | | | Mucinous | 1.99 (0.92-4.28) | 0.079 | 1.39 (0.62-3.11) | 0.418 | **30** |
| | | | 132 | Clear cell | 0.9 (0.37-2.19) | 0.824 | 0.5 (0.18-1.36) | 0.175 | **44** |
| | | | 1751 | All | 1 (0.9-1.11) | 0.959 | 1.03 (0.91-1.17) | 0.622 | 3 |
| | | | 835 | Serous | 0.96 (0.83-1.1) | 0.517 | 1 (0.9-1.2) | 0.898 | 4 |
| *KRAS* | rs12822857 | 0.47 | 268 | Endometrioid | 0.96 (0.66-1.38) | 0.821 | 1.01 (0.67-1.5) | 0.976 | 5 |
| | | | 166 | Mucinous | 0.99 (0.65-1.52) | 0.976 | 0.94 (0.59-1.53) | 0.816 | 5 |
| | | | 136 | Clear cell | 0.93 (0.54-1.6) | 0.796 | 0.89 (0.47-1.7) | 0.726 | 4 |
| | | | 1776 | All | 0.97 (0.86-1.1) | 0.683 | 1 (0.86-1.17) | 0.966 | 3 |
| | | | 841 | Serous | 0.94 (0.8-1.1) | 0.429 | 1 (0.8-1.2) | 0.967 | 6 |
| *KRAS* | rs10842508 | 0.24 | 273 | Endometrioid | 1.03 (0.7-1.52) | 0.879 | 0.99 (0.64-1.54) | 0.965 | 4 |
| | | | 190 | Mucinous | 1.7 (0.96-3.02) | 0.071 | 1.26 (0.64-2.47) | 0.503 | **26** |
| | | | 135 | Clear cell | 0.97 (0.53-1.77) | 0.919 | 0.65 (0.27-1.54) | 0.329 | **33** |
| | | | 1765 | All | 0.96 (0.86-1.07) | 0.417 | 0.97 (0.85-1.09) | 0.583 | 1 |
| | | | 836 | Serous | 0.92 (0.8-1.06) | 0.253 | 0.9 (0.8-1.1) | 0.407 | 2 |
| *KRAS* | rs12579073 | 0.48 | 269 | Endometrioid | **0.7 (0.49-1)** | 0.053 | 0.74 (0.5-1.1) | 0.138 | 6 |
| | | | 190 | Mucinous | 0.99 (0.65-1.53) | 0.981 | 1.24 (0.77-2) | 0.367 | **25** |
| | | | 137 | Clear cell | 0.86 (0.51-1.43) | 0.554 | 0.89 (0.49-1.6) | 0.694 | 3 |
| ***KRAS*** | **rs10842513** | 0.09 | 1770 | **All** | 1.18 (0.98-1.42) | 0.08 | **1.25 (1.01-1.55)** | **0.039** | 6 |
| | | | 846 | **Serous** | **1.38 (1.09-1.75)** | **0.008** | 1.3 (1-1.6) | 0.091 | 6 |
| | | | 271 | Endometrioid | 1.19 (0.67-2.1) | 0.552 | 1.47 (0.79-2.74) | 0.227 | **24** |
| | | | 187 | Mucinous | 0.7 (0.29-1.69) | 0.432 | 0.74 (0.29-1.87) | 0.521 | 6 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 132 | Clear cell | **2.02 (1-4.1)** | 0.052 | 1.71 (0.81-3.58) | 0.156 | **15** |
| ***KRAS*** | **rs4623993** | 0.16 | 1748 | All | 0.93 (0.8-1.09) | 0.378 | 0.96 (0.8-1.16) | 0.676 | 3 |
| | | | 834 | Serous | 0.89 (0.73-1.08) | 0.236 | 0.9 (0.7-1.1) | 0.306 | 1 |
| | | | 242 | Endometrioid | 0.83 (0.51-1.36) | 0.463 | 0.93 (0.55-1.56) | 0.782 | **12** |
| | | | 187 | **Mucinous** | **1.79 (1.02-3.15)** | **0.044** | 1.56 (0.83-2.93) | 0.163 | **13** |
| | | | 136 | Clear cell | 0.93 (0.46-1.89) | 0.835 | 1.18 (0.46-3.01) | 0.732 | **27** |
| *KRAS* | rs6487464 | 0.39 | 1763 | All | 1 (0.9-1.12) | 0.933 | 1.03 (0.9-1.16) | 0.688 | 3 |
| | | | 836 | Serous | 0.97 (0.84-1.12) | 0.694 | 1 (0.8-1.1) | 0.795 | 3 |
| | | | 243 | Endometrioid | 0.77 (0.53-1.13) | 0.189 | 0.93 (0.61-1.41) | 0.74 | **21** |
| | | | 192 | Mucinous | 0.99 (0.65-1.52) | 0.978 | 0.92 (0.59-1.45) | 0.734 | 7 |
| | | | 134 | Clear cell | 0.94 (0.56-1.59) | 0.823 | 1.24 (0.68-2.27) | 0.484 | **32** |
| *KRAS* | rs10842514 | 0.45 | 1757 | All | 0.95 (0.85-1.06) | 0.339 | 0.93 (0.82-1.05) | 0.216 | 2 |
| | | | 835 | Serous | 0.96 (0.84-1.1) | 0.594 | 1 (0.9-1.1) | 0.82 | 4 |
| | | | 243 | Endometrioid | 1.09 (0.77-1.55) | 0.623 | 0.94 (0.64-1.37) | 0.748 | **14** |
| | | | 188 | Mucinous | 1.04 (0.66-1.64) | 0.88 | 1.18 (0.69-2.01) | 0.536 | **13** |
| | | | 118 | Clear cell | 0.71 (0.42-1.21) | 0.21 | 0.65 (0.35-1.19) | 0.162 | 8 |
| *KRAS* | rs11047917 | 0.06 | 1476 | All | 1 (0.79-1.25) | 0.982 | 1.03 (0.81-1.32) | 0.805 | 3 |
| | | | 685 | Serous | 0.88 (0.66-1.17) | 0.373 | 1 (0.7-1.3) | 0.832 | **14** |
| | | | 231 | Endometrioid | 0.64 (0.2-2.06) | 0.455 | 0.61 (0.19-1.99) | 0.409 | 5 |
| | | | 163 | Mucinous | 0.72 (0.18-2.82) | 0.634 | 0.77 (0.23-2.57) | 0.677 | 7 |
| | | | 117 | Clear cell | 1.95 (0.83-4.6) | 0.128 | 2.34 (0.89-6.17) | 0.085 | **20** |
| *KRAS* | rs4623993, rs12579073; TC | 11.6* | 1717 | All | 0.89 (0.71-1.11) | 0.289 | 0.86 (0.65-1.14) | 0.3 | 3 |
| | | | 698 | Serous | 0.86 (0.64-1.16) | 0.32 | 0.8 (0.6-1.1) | 0.233 | 7 |
| | | | 238 | Endometrioid | 0.76 (0.39-1.48) | 0.417 | 0.82 (0.4-1.7) | 0.592 | 8 |
| | | | 162 | Mucinous | 1.3 (0.54-3.14) | 0.556 | 0.98 (0.32-2.98) | 0.97 | **25** |
| | | | 119 | Clear cell | 0.73 (0.28-1.88) | 0.513 | 0.88 (0.22-3.5) | 0.856 | **21** |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *KRAS* | rs12822857, rs10842508; AC | 24.4* | 1730 | All | 1.03 (0.91-1.17) | 0.653 | 1.05 (0.91-1.22) | 0.488 | 2 |
| | | | 714 | Serous | 1.01 (0.86-1.18) | 0.904 | 1 (0.8-1.2) | 0.996 | 1 |
| | | | 243 | Endometrioid | 0.87 (0.56-1.38) | 0.563 | 1.01 (0.61-1.66) | 0.973 | **16** |
| | | | 164 | Mucinous | 0.61 (0.33-1.13) | 0.114 | 0.71 (0.36-1.4) | 0.322 | **16** |
| | | | 121 | Clear cell | 1.06 (0.61-1.86) | 0.826 | 1.32 (0.68-2.55) | 0.409 | **25** |
| *KRAS* | rs12822857, rs10842514; GT | 38.3* | 1715 | All | 0.95 (0.85-1.06) | 0.366 | 0.92 (0.81-1.04) | 0.194 | 3 |
| | | | 711 | Serous | 0.95 (0.83-1.1) | 0.494 | 1 (0.8-1.1) | 0.521 | 5 |
| | | | 239 | Endometrioid | 1.03 (0.72-1.48) | 0.868 | 0.88 (0.59-1.31) | 0.527 | 15 |
| | | | 164 | Mucinous | 1 (0.62-1.59) | 0.983 | 1.02 (0.59-1.77) | 0.934 | 2 |
| | | | 120 | Clear cell | 0.82 (0.45-1.48) | 0.508 | 0.96 (0.46-2.01) | 0.916 | **17** |
| *KRAS* | rs12822857, rs12579073, rs6487464; GAC | 44.5* | 1689 | All | 1.01 (0.89-1.15) | 0.824 | 0.97 (0.83-1.12) | 0.659 | 4 |
| | | | 695 | Serous | 1.13 (0.96-1.33) | 0.133 | 1.1 (0.9-1.3) | 0.331 | 3 |
| | | | 236 | Endometrioid | 1.32 (0.86-2.01) | 0.201 | 1.2 (0.76-1.91) | 0.436 | 9 |
| | | | 164 | Mucinous | 0.82 (0.48-1.43) | 0.49 | 0.65 (0.37-1.17) | 0.152 | **21** |
| | | | 129 | Clear cell | 0.97 (0.52-1.83) | 0.931 | 1.03 (0.46-2.29) | 0.942 | 6 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VII-F: Univariate and multivariate Cox regression results of *KRAS* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *KRAS* (haplotype block 1) | h100 | 52.7 | All | 1 (0.9-1.11) | 0.989 | 0.96 (0.85-1.09) | 0.542 | 4 |
| | | | Serous | 1.05 (0.91-1.2) | 0.519 | 0.3 (0-1.9) | 0.198 | **71** |
| | | | Endometrioid | 1.06 (0.74-1.52) | 0.755 | 1 (0.67-1.48) | 0.995 | 6 |
| | | | Mucinous | 1 (0.65-1.52) | 0.988 | 1.04 (0.65-1.68) | 0.857 | 4 |
| | | | Clear cell | 0.99 (0.58-1.69) | 0.98 | 1.03 (0.55-1.92) | 0.928 | 4 |
| *KRAS* (haplotype block 1) | h000 | 22.8 | All | 1.03 (0.91-1.17) | 0.591 | 1.07 (0.92-1.24) | 0.382 | 4 |
| | | | Serous | 1.03 (0.87-1.2) | 0.75 | 0.9 (0.7-1) | 0.084 | **13** |
| | | | Endometrioid | 0.88 (0.56-1.38) | 0.583 | 1.02 (0.62-1.67) | 0.95 | **16** |
| | | | Mucinous | 0.65 (0.37-1.16) | 0.147 | 0.77 (0.41-1.46) | 0.422 | **18** |
| | | | Clear cell | 1.04 (0.59-1.81) | 0.902 | 1.28 (0.67-2.44) | 0.463 | **23** |
| *KRAS* (haplotype block 1) | h001 | 15.5 | All | 0.95 (0.82-1.1) | 0.466 | 0.99 (0.83-1.18) | 0.889 | 4 |
| | | | Serous | 0.94 (0.78-1.13) | 0.504 | 1 (0.9-1.2) | 0.774 | 6 |
| | | | Endometrioid | 0.89 (0.56-1.43) | 0.631 | 0.99 (0.59-1.66) | 0.971 | **11** |
| | | | Mucinous | 1.29 (0.67-2.49) | 0.448 | 1.02 (0.46-2.24) | 0.964 | **21** |
| | | | Clear cell | 1.02 (0.5-2.11) | 0.949 | 1.22 (0.49-3.05) | 0.676 | **20** |
| *KRAS* (haplotype block 1) | h101 | 8.9 | All | 1.04 (0.86-1.26) | 0.692 | 1.02 (0.82-1.28) | 0.838 | 2 |
| | | | Serous | 0.94 (0.74-1.2) | 0.644 | 1 (0.8-1.2) | 0.944 | 6 |
| | | | Endometrioid | 1.43 (0.79-2.57) | 0.238 | 1.04 (0.51-2.11) | 0.916 | **27** |
| | | | Mucinous | 2.01 (0.94-4.3) | 0.07 | 1.38 (0.62-3.07) | 0.424 | **31** |
| | | | Clear cell | 0.9 (0.37-2.19) | 0.823 | 0.49 (0.18-1.37) | 0.174 | **46** |
| *KRAS* (haplotype block 2) | h000010 | 34.5 | All | 0.96 (0.84-1.09) | 0.495 | 0.92 (0.79-1.08) | 0.322 | 4 |
| | | | Serous | 1.03 (0.87-1.22) | 0.702 | 1 (0.7-1.4) | 0.988 | 3 |
| | | | Endometrioid | 1.23 (0.82-1.85) | 0.324 | 1.03 (0.65-1.64) | 0.883 | **16** |
| | | | Mucinous | 0.77 (0.43-1.39) | 0.387 | 0.66 (0.33-1.28) | 0.218 | **14** |
| | | | Clear cell | 0.65 (0.32-1.33) | 0.238 | 0.53 (0.22-1.26) | 0.15 | **18** |
| ***KRAS*** | **h100100** | 13.8 | All | 1.05 (0.87-1.25) | 0.617 | 1.06 (0.85-1.31) | 0.613 | 1 |

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Serous | 1.13 (0.9-1.43) | 0.293 | 0.9 (0.8-1.1) | 0.376 | **20** |
| | | | **Endometrioid** | **0.49 (0.24-1)** | 0.05 | 0.59 (0.27-1.32) | 0.2 | **20** |
| | | | Mucinous | 0.71 (0.25-2.04) | 0.521 | 1.06 (0.32-3.5) | 0.926 | **49** |
| | | | Clear cell | 0.32 (0.09-1.16) | 0.082 | 0.26 (0.04-1.81) | 0.172 | **19** |
| *KRAS* (haplotype block 2) | h101100 | 11.8 | All | 0.92 (0.75-1.11) | 0.375 | 0.91 (0.71-1.16) | 0.429 | 1 |
| | | | Serous | 0.89 (0.69-1.15) | 0.367 | 1 (0.6-1.7) | 0.908 | **12** |
| | | | Endometrioid | 0.83 (0.45-1.52) | 0.552 | 0.93 (0.49-1.77) | 0.821 | **12** |
| | | | Mucinous | 1.12 (0.49-2.57) | 0.787 | 0.91 (0.33-2.53) | 0.861 | **19** |
| | | | Clear cell | 0.74 (0.31-1.77) | 0.505 | 0.84 (0.25-2.79) | 0.771 | **14** |
| *KRAS* (haplotype block 2) | h100010 | 10.7 | All | 0.96 (0.83-1.12) | 0.625 | 0.95 (0.81-1.12) | 0.556 | 1 |
| | | | Serous | 0.9 (0.74-1.08) | 0.255 | 0.9 (0.6-1.3) | 0.479 | 0 |
| | | | Endometrioid | 0.86 (0.51-1.45) | 0.571 | 0.82 (0.47-1.43) | 0.481 | 5 |
| | | | Mucinous | 1.32 (0.79-2.22) | 0.288 | **1.79 (1.03-3.13)** | **0.04** | **36** |
| | | | Clear cell | 0.95 (0.47-1.91) | 0.881 | 0.87 (0.39-1.93) | 0.723 | 8 |
| ***KRAS* (haplotype block 2)** | **h010000** | 5.9 | All | 1.27 (0.99-1.62) | 0.056 | 1.24 (0.93-1.66) | 0.15 | 2 |
| | | | **Serous** | **1.69 (1.21-2.36)** | **0.002** | 0.9 (0.6-1.3) | 0.523 | **47** |
| | | | Endometrioid | 1.21 (0.59-2.48) | 0.599 | 1.53 (0.72-3.23) | 0.269 | **26** |
| | | | Mucinous | 0.66 (0.17-2.55) | 0.55 | 0.48 (0.11-2.12) | 0.33 | **27** |
| | | | Clear cell | 2.81 (0.95-8.33) | 0.062 | 2.29 (0.71-7.45) | 0.167 | **19** |
| *KRAS* (haplotype block 2) | h100000 | 4.9 | All | 0.89 (0.67-1.18) | 0.412 | 0.83 (0.59-1.17) | 0.283 | 7 |
| | | | Serous | 0.89 (0.62-1.28) | 0.536 | 1.3 (0.9-1.9) | 0.203 | **46** |
| | | | Endometrioid | 1.1 (0.41-2.94) | 0.848 | 0.72 (0.18-2.91) | 0.644 | **35** |
| | | | Mucinous | 0.84 (0.28-2.59) | 0.767 | 0.87 (0.27-2.77) | 0.814 | 4 |
| | | | Clear cell | 1.04 (0.33-3.28) | 0.95 | 0.91 (0.25-3.28) | 0.885 | **13** |
| *KRAS* (haplotype block 2) | h100101 | 4.1 | All | 0.96 (0.7-1.31) | 0.794 | 1.06 (0.73-1.54) | 0.759 | **10** |
| | | | Serous | 0.81 (0.53-1.25) | 0.352 | 1.1 (0.8-1.4) | 0.528 | **36** |
| | | | Endometrioid | 0.43 (0.09-2.11) | 0.296 | 0.52 (0.12-2.23) | 0.377 | **21** |
| | | | Mucinous | 0.06 (0-24.2) | 0.355 | 0.11 (0-32.59) | 0.451 | **83** |
| | | | Clear cell | 1.84 (0.67-5.03) | 0.236 | 2.37 (0.73-7.7) | 0.15 | **29** |

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *KRAS* (haplotype block 2) | h000100 | 3.7 | All | 1.03 (0.82-1.29) | 0.813 | 1.05 (0.83-1.34) | 0.669 | 2 |
| | | | Serous | 1.09 (0.81-1.45) | 0.577 | 1 (0.6-1.6) | 0.865 | 8 |
| | | | Endometrioid | 1.67 (0.85-3.28) | 0.134 | 1.48 (0.72-3.06) | 0.287 | **11** |
| | | | Mucinous | 0.43 (0.13-1.43) | 0.171 | 0.42 (0.14-1.31) | 0.134 | 2 |
| | | | Clear cell | 1.11 (0.46-2.66) | 0.811 | 1.21 (0.48-3.05) | 0.685 | 9 |
| ***KRAS* (haplotype block 2)** | **h001100** | 3.7 | All | 1.02 (0.79-1.31) | 0.902 | 1.05 (0.79-1.41) | 0.722 | 3 |
| | | | Serous | 0.87 (0.62-1.21) | 0.411 | 1.1 (0.8-1.5) | 0.631 | **26** |
| | | | Endometrioid | 0.99 (0.44-2.18) | 0.971 | 1.21 (0.52-2.82) | 0.652 | **22** |
| | | | **Mucinous** | **3.24 (1.55-6.74)** | **0.002** | **2.74 (1.27-5.9)** | **0.01** | **15** |
| | | | Clear cell | 2.42 (0.6-9.66) | 0.212 | 3.42 (0.65-18) | 0.146 | **41** |
| ***KRAS* (haplotype block 2)** | **h000000** | 2.6 | All | 1.26 (0.87-1.82) | 0.219 | 1.18 (0.78-1.77) | 0.44 | 6 |
| | | | Serous | 0.96 (0.61-1.53) | 0.872 | 1 (0.9-1.2) | 0.956 | 4 |
| | | | Endometrioid | 2.47 (0.84-7.23) | 0.099 | 1.55 (0.52-4.65) | 0.433 | **37** |
| | | | **Mucinous** | **6.59 (1.37-31.62)** | **0.018** | 2.57 (0.5-13.28) | 0.26 | **61** |
| | | | Clear cell | 1.53 (0.32-7.36) | 0.593 | 0.67 (0.09-4.97) | 0.698 | **56** |
| *KRAS* (haplotype block 2) | h110000 | 2.3 | All | 1.07 (0.77-1.5) | 0.686 | 1.28 (0.9-1.82) | 0.177 | **20** |
| | | | Serous | 1.07 (0.71-1.64) | 0.738 | 0.9 (0.7-1.2) | 0.427 | **16** |
| | | | Endometrioid | 1.87 (0.57-6.12) | 0.302 | 2.03 (0.56-7.32) | 0.282 | 9 |
| | | | Mucinous | 0.76 (0.22-2.63) | 0.663 | 1.05 (0.3-3.71) | 0.934 | **38** |
| | | | Clear cell | 1.84 (0.72-4.72) | 0.202 | 1.43 (0.56-3.65) | 0.459 | **22** |
| *KRAS* (haplotype block 2) | Rare | 1.7 | All | 1.1 (0.8-1.5) | 0.57 | 1.07 (0.79-1.46) | 0.66 | 3 |
| | | | Serous | 1.03 (0.72-1.47) | 0.884 | 1.2 (0.7-1.8) | 0.515 | **17** |
| | | | Endometrioid | 0.91 (0.23-3.62) | 0.889 | 1.29 (0.32-5.12) | 0.722 | **42** |
| | | | Mucinous | 1.82 (0.44-7.61) | 0.411 | 2.47 (0.57-10.6) | 0.225 | **36** |
| | | | Clear cell | 1.37 (0.26-7.1) | 0.708 | 2.22 (0.37-13.1) | 0.38 | **62** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *KRAS* -block 1: rs12305513, rs12822857, rs10842508; *KRAS* -block 2: rs12579073, rs10842513, rs4623993, rs6487464, rs10842514, rs11047917.

## Appendix VII-G: Univariate and multivariate Cox regression results of *NMI* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *NMI* | rs394884 | 0.14 | 1708 | All | 1.04 (0.88-1.22) | 0.655 | 1 (0.82-1.21) | 0.975 | 4 |
| | | | 809 | Serous | 0.97 (0.79-1.19) | 0.771 | 1 (0.8-1.2) | 0.934 | 3 |
| | | | 260 | Endometrioid | 1.09 (0.63-1.87) | 0.768 | 1.26 (0.73-2.15) | 0.404 | **16** |
| | | | 184 | Mucinous | 1.33 (0.64-2.76) | 0.443 | 0.73 (0.32-1.65) | 0.452 | **45** |
| | | | 95 | Clear cell | 0.37 (0.13-1.06) | 0.063 | 0.47 (0.13-1.63) | 0.232 | **27** |
| *NMI* | rs11551174 | 0.06 | 1159 | All | 0.9 (0.67-1.19) | 0.457 | 1.02 (0.73-1.43) | 0.901 | **13** |
| | | | 524 | Serous | 1.03 (0.71-1.49) | 0.887 | 1 (0.6-1.5) | 0.92 | 3 |
| | | | 185 | Endometrioid | 1.38 (0.67-2.85) | 0.382 | 0.84 (0.37-1.9) | 0.675 | **39** |
| | | | 133 | Mucinous | 1.03 (0.42-2.55) | 0.942 | 0.95 (0.33-2.7) | 0.918 | 8 |
| | | | 132 | Clear cell | 0.9 (0.26-3.09) | 0.867 | 1.26 (0.35-4.5) | 0.726 | **40** |
| *NMI* | rs289831 | 0.12 | 1665 | All | 1.02 (0.86-1.21) | 0.8 | 1.01 (0.83-1.23) | 0.933 | 1 |
| | | | 792 | Serous | 0.92 (0.74-1.13) | 0.425 | 1 (0.8-1.2) | 0.8 | 9 |
| | | | 258 | Endometrioid | 1.19 (0.67-2.12) | 0.562 | 1.28 (0.71-2.31) | 0.404 | 8 |
| | | | 176 | Mucinous | 0.9 (0.38-2.13) | 0.809 | 0.53 (0.18-1.55) | 0.246 | **41** |
| | | | 132 | Clear cell | 0.46 (0.16-1.32) | 0.149 | 0.61 (0.17-2.16) | 0.447 | **33** |
| *NMI* | rs3771886 | 0.43 | 1764 | All | 1.02 (0.91-1.13) | 0.749 | 1.02 (0.9-1.16) | 0.789 | 0 |
| | | | 843 | Serous | 1.02 (0.89-1.17) | 0.76 | 1.1 (0.9-1.2) | 0.42 | 8 |
| | | | 266 | Endometrioid | 0.94 (0.64-1.37) | 0.741 | 0.81 (0.54-1.23) | 0.332 | **14** |
| | | | 191 | Mucinous | 1.23 (0.79-1.91) | 0.353 | 1.08 (0.66-1.75) | 0.771 | **12** |
| | | | 107 | Clear cell | 1.07 (0.64-1.78) | 0.798 | 1.51 (0.81-2.82) | 0.194 | **41** |
| *NMI* | rs11683487 | 0.44 | 1464 | All | 0.98 (0.88-1.09) | 0.683 | 0.94 (0.83-1.08) | 0.397 | 4 |
| | | | 713 | Serous | 1 (0.87-1.16) | 0.971 | 0.9 (0.8-1.1) | 0.343 | 10 |
| | | | 227 | Endometrioid | 0.98 (0.69-1.4) | 0.91 | 1.06 (0.72-1.56) | 0.785 | 8 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 154 | Mucinous | 0.77 (0.49-1.22) | 0.265 | 1.12 (0.68-1.85) | 0.659 | **45** |
| | | | 138 | Clear cell | 1.15 (0.67-1.96) | 0.614 | 0.7 (0.36-1.37) | 0.304 | **39** |
| | | | 1776 | All | 1.03 (0.88-1.22) | 0.685 | 1 (0.82-1.21) | 0.986 | 3 |
| | | | 843 | Serous | 0.95 (0.78-1.17) | 0.65 | 1 (0.8-1.2) | 0.931 | 5 |
| *NMI* | rs2113509 | 0.13 | 272 | Endometrioid | 1.11 (0.63-1.96) | 0.722 | 1.17 (0.66-2.08) | 0.591 | 5 |
| | | | 190 | Mucinous | 1.1 (0.54-2.24) | 0.788 | 0.68 (0.3-1.54) | 0.36 | **38** |
| | | | 126 | Clear cell | 0.45 (0.16-1.31) | 0.146 | 0.6 (0.17-2.1) | 0.424 | **33** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VII-H: Univariate and multivariate Cox regression results of *NMI* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *NMI* | h00001 | 44.3 | All | 0.97 (0.87-1.08) | 0.573 | 0.98 (0.86-1.11) | 0.744 | 1 |
| | | | Serous | 1 (0.87-1.15) | 0.99 | 1.4 (0.9-2.3) | 0.131 | **40** |
| | | | Endometrioid | 1.01 (0.71-1.42) | 0.964 | 1.13 (0.77-1.66) | 0.532 | **12** |
| | | | Mucinous | 0.8 (0.51-1.25) | 0.328 | 1.12 (0.68-1.83) | 0.662 | **40** |
| | | | Clear cell | 1.11 (0.66-1.87) | 0.696 | 0.71 (0.37-1.36) | 0.301 | **36** |
| *NMI* | h00010 | 35 | All | 1.03 (0.92-1.16) | 0.64 | 1.01 (0.88-1.16) | 0.864 | 2 |
| | | | Serous | 1 (0.86-1.16) | 0.979 | 1.1 (0.9-1.3) | 0.55 | 10 |
| | | | Endometrioid | 0.87 (0.57-1.32) | 0.515 | 0.86 (0.54-1.36) | 0.524 | 1 |
| | | | Mucinous | 1.22 (0.76-1.96) | 0.418 | 1.1 (0.65-1.87) | 0.72 | 10 |
| | | | Clear cell | 1.27 (0.76-2.14) | 0.366 | 1.6 (0.87-2.95) | 0.133 | **26** |
| *NMI* | h10100 | 12.3 | All | 1.03 (0.87-1.22) | 0.755 | 1.01 (0.83-1.23) | 0.932 | 2 |
| | | | Serous | 0.96 (0.78-1.18) | 0.707 | 1 (0.7-1.5) | 0.924 | 4 |
| | | | Endometrioid | 1.05 (0.59-1.87) | 0.875 | 1.22 (0.68-2.2) | 0.5 | **16** |
| | | | Mucinous | 1.07 (0.52-2.19) | 0.852 | 0.66 (0.29-1.5) | 0.319 | **38** |
| | | | Clear cell | 0.45 (0.15-1.31) | 0.145 | 0.59 (0.17-2.1) | 0.419 | **31** |
| *NMI* | h01010 | 5.6 | All | 0.9 (0.68-1.19) | 0.44 | 1 (0.71-1.39) | 0.978 | **11** |
| | | | Serous | 1.07 (0.75-1.54) | 0.7 | 1 (0.9-1.2) | 0.664 | 7 |
| | | | Endometrioid | 1.22 (0.57-2.59) | 0.61 | 0.72 (0.31-1.71) | 0.46 | **41** |
| | | | Mucinous | 0.99 (0.4-2.49) | 0.991 | 0.94 (0.34-2.62) | 0.904 | 5 |
| | | | Clear cell | 1.05 (0.35-3.13) | 0.933 | 1.56 (0.52-4.73) | 0.43 | **49** |
| *NMI* | Rare | 1.9 | All | 1.16 (0.83-1.62) | 0.397 | 1.07 (0.74-1.56) | 0.712 | 8 |
| | | | Serous | 1.11 (0.69-1.76) | 0.674 | 1 (0.8-1.2) | 0.903 | 10 |
| | | | Endometrioid | 1.66 (0.51-5.37) | 0.4 | 0.8 (0.25-2.59) | 0.715 | **52** |
| | | | Mucinous | 1.21 (0.42-3.44) | 0.722 | 1.02 (0.34-3.03) | 0.968 | **16** |
| | | | Clear cell | 0.35 (0.04-2.93) | 0.333 | 0.39 (0.04-3.37) | 0.39 | **11** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *NMI*: rs394884, rs11551174, rs289831, rs3771886, rs11683487.

## Appendix VII-I: Univariate and multivariate Cox regression results of *PIK3CA* tSNPs

| Gene | tSNP | MAF | Cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *PIK3CA* | rs2865084 | 0.06 | 1164 | All | 1 (0.75-1.33) | 0.984 | 0.87 (0.61-1.24) | 0.449 | **13** |
| | | | 520 | Serous | 0.93 (0.64-1.34) | 0.69 | 0.7 (0.4-1.1) | 0.086 | **25** |
| | | | 183 | Endometrioid | 1.76 (0.87-3.55) | 0.117 | 1.29 (0.6-2.76) | 0.514 | **27** |
| | | | 99 | Mucinous | 0.81 (0.25-2.65) | 0.722 | 1.73 (0.36-8.28) | 0.495 | **114** |
| | | | 94 | Clear cell | 0.79 (0.1-5.98) | 0.819 | - | - | - |
| *PIK3CA* | rs7621329 | 0.17 | 1749 | All | 1.11 (0.96-1.27) | 0.159 | 1.05 (0.89-1.24) | 0.591 | 5 |
| | | | 834 | Serous | 1.05 (0.89-1.25) | 0.555 | 1 (0.8-1.2) | 0.807 | 5 |
| | | | 268 | Endometrioid | 1.39 (0.91-2.13) | 0.122 | 1.37 (0.86-2.17) | 0.186 | 1 |
| | | | 186 | Mucinous | 1.34 (0.7-2.57) | 0.384 | 1.23 (0.6-2.53) | 0.569 | 8 |
| | | | 134 | Clear cell | 1.68 (0.83-3.37) | 0.147 | 1.48 (0.64-3.43) | 0.358 | **12** |
| *PIK3CA* | rs1517586 | 0.1 | 1739 | All | 1.1 (0.92-1.32) | 0.311 | 0.96 (0.76-1.21) | 0.725 | **13** |
| | | | 827 | Serous | 1.13 (0.9-1.41) | 0.294 | 0.9 (0.7-1.2) | 0.691 | **20** |
| | | | 267 | Endometrioid | 0.67 (0.34-1.34) | 0.263 | 1.21 (0.56-2.59) | 0.628 | **81** |
| | | | 183 | Mucinous | 0.73 (0.31-1.71) | 0.463 | 0.7 (0.27-1.83) | 0.467 | 4 |
| | | | 135 | Clear cell | 1.29 (0.53-3.13) | 0.57 | 1.24 (0.37-4.16) | 0.727 | 4 |
| *PIK3CA* | rs2699905 | 0.26 | 1741 | All | 0.98 (0.87-1.1) | 0.703 | 0.92 (0.79-1.06) | 0.246 | 6 |
| | | | 825 | Serous | 1 (0.85-1.16) | 0.95 | **0.9 (0.7-1)** | 0.152 | **10** |
| | | | 266 | Endometrioid | 0.98 (0.68-1.42) | 0.917 | 1.09 (0.74-1.6) | 0.668 | **11** |
| | | | 184 | Mucinous | 0.73 (0.39-1.37) | 0.326 | 1.07 (0.53-2.15) | 0.857 | **47** |
| | | | 136 | Clear cell | 0.86 (0.49-1.51) | 0.598 | 0.99 (0.48-2.07) | 0.983 | **15** |
| *PIK3CA* | rs7641889 | 0.06 | 1779 | All | 1.14 (0.92-1.4) | 0.229 | 1.06 (0.83-1.35) | 0.639 | 7 |
| | | | 845 | Serous | 1.09 (0.85-1.4) | 0.497 | 1.1 (0.9-1.5) | 0.343 | 1 |
| | | | 273 | Endometrioid | 1.1 (0.53-2.28) | 0.791 | 1.42 (0.66-3.04) | 0.367 | **29** |

| Gene | tSNP | MAF | Cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 192 | Mucinous | 0.91 (0.26-3.1) | 0.876 | 0.61 (0.18-2.08) | 0.434 | **33** |
| | | | 136 | Clear cell | 2.03 (0.85-4.85) | 0.111 | 1.31 (0.47-3.64) | 0.607 | **35** |
| | | | 1794 | All | 1.07 (0.9-1.26) | 0.449 | 1.02 (0.83-1.24) | 0.868 | 5 |
| | | | 828 | Serous | 1.05 (0.85-1.29) | 0.678 | 1.1 (0.8-1.3) | 0.651 | 5 |
| ***PIK3CA*** | **rs7651265** | 0.1 | 267 | Endometrioid | 0.97 (0.56-1.67) | 0.913 | 0.98 (0.54-1.8) | 0.958 | 1 |
| | | | 189 | Mucinous | 1.66 (0.79-3.46) | 0.179 | 1.02 (0.47-2.21) | 0.952 | **39** |
| | | | 135 | **Clear cell** | **2.25 (1.06-4.79)** | **0.035** | 1.99 (0.85-4.7) | 0.115 | **12** |
| | | | 1765 | All | 0.88 (0.75-1.04) | 0.133 | 0.88 (0.73-1.06) | 0.169 | 0 |
| | | | 842 | Serous | 0.91 (0.74-1.12) | 0.374 | 0.9 (0.7-1.1) | 0.311 | 1 |
| *PIK3CA* | rs7640662 | 0.15 | 268 | Endometrioid | 1.1 (0.71-1.71) | 0.668 | 0.86 (0.55-1.35) | 0.516 | **22** |
| | | | 188 | Mucinous | 0.58 (0.24-1.37) | 0.215 | 0.95 (0.38-2.38) | 0.914 | **64** |
| | | | 134 | Clear cell | 0.7 (0.33-1.52) | 0.374 | 1.07 (0.42-2.71) | 0.889 | **53** |
| | | | 1762 | All | 0.97 (0.87-1.08) | 0.545 | 1 (0.88-1.14) | 0.973 | 3 |
| | | | 836 | Serous | 0.92 (0.81-1.05) | 0.23 | 1 (0.9-1.2) | 0.851 | 9 |
| *PIK3CA* | rs2677760 | 0.49 | 268 | Endometrioid | 0.98 (0.67-1.42) | 0.896 | 0.92 (0.63-1.35) | 0.676 | 6 |
| | | | 189 | Mucinous | 1.17 (0.67-2.05) | 0.572 | 1.11 (0.62-2) | 0.727 | 5 |
| | | | 125 | Clear cell | 1.1 (0.66-1.84) | 0.703 | 0.99 (0.51-1.9) | 0.964 | **10** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VII-J: Univariate and multivariate Cox regression results of *PIK3CA* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *PIK3CA* | h00000001 | 48.3 | All | 0.97 (0.87-1.08) | 0.539 | 1.01 (0.89-1.14) | 0.911 | 4 |
| | | | Serous | 0.95 (0.79-1.16) | 0.633 | 0.6 (0.4-1.1) | 0.094 | **37** |
| | | | Endometrioid | 1 (0.69-1.45) | 0.99 | 0.98 (0.67-1.41) | 0.894 | 2 |
| | | | Mucinous | 1.17 (0.67-2.04) | 0.582 | 1.1 (0.61-1.98) | 0.748 | 6 |
| | | | Clear cell | 1.1 (0.65-1.84) | 0.73 | 0.98 (0.5-1.89) | 0.942 | **11** |
| *PIK3CA* | h00010010 | 15.2 | All | 0.9 (0.77-1.06) | 0.206 | 0.91 (0.75-1.09) | 0.301 | 1 |
| | | | Serous | 0.93 (0.75-1.14) | 0.461 | 1.2 (1-1.6) | 0.104 | **29** |
| | | | Endometrioid | 1.18 (0.75-1.86) | 0.467 | 0.97 (0.6-1.57) | 0.911 | **18** |
| | | | Mucinous | 0.69 (0.31-1.56) | 0.376 | 1.12 (0.47-2.67) | 0.796 | **62** |
| | | | Clear cell | 0.68 (0.31-1.49) | 0.34 | 0.97 (0.38-2.48) | 0.945 | **43** |
| *PIK3CA* | h00000000 | 9.6 | All | 1.07 (0.88-1.29) | 0.501 | 1.16 (0.93-1.43) | 0.183 | 8 |
| | | | Serous | 1.24 (0.98-1.56) | 0.078 | 0.9 (0.8-1.1) | 0.438 | **27** |
| | | | Endometrioid | 0.72 (0.36-1.46) | 0.362 | 0.8 (0.41-1.56) | 0.511 | **11** |
| | | | Mucinous | 1.05 (0.46-2.4) | 0.911 | 0.74 (0.3-1.82) | 0.506 | **30** |
| | | | Clear cell | 0.57 (0.21-1.54) | 0.266 | 0.49 (0.13-1.79) | 0.281 | **14** |
| *PIK3CA* | h00110000 | 9.6 | All | 1.05 (0.87-1.26) | 0.609 | 0.95 (0.75-1.19) | 0.647 | 10 |
| | | | Serous | 1.08 (0.86-1.35) | 0.494 | 0.9 (0.7-1.1) | 0.416 | **17** |
| | | | Endometrioid | 0.64 (0.32-1.28) | 0.21 | 1.23 (0.57-2.63) | 0.6 | **92** |
| | | | Mucinous | 0.73 (0.31-1.73) | 0.479 | 0.78 (0.3-2.03) | 0.604 | 7 |
| | | | Clear cell | 1.17 (0.48-2.85) | 0.734 | 1.14 (0.34-3.84) | 0.838 | 3 |
| *PIK3CA* | h01001100 | 6.5 | All | 1.11 (0.9-1.36) | 0.332 | 1.05 (0.82-1.34) | 0.7 | 5 |
| | | | Serous | 1.06 (0.83-1.37) | 0.636 | 1 (0.7-1.5) | 0.978 | 6 |
| | | | Endometrioid | 1.02 (0.49-2.09) | 0.966 | 1.32 (0.62-2.81) | 0.466 | **29** |
| | | | Mucinous | 0.99 (0.3-3.31) | 0.991 | 0.66 (0.2-2.2) | 0.497 | **33** |

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Clear cell | 2.03 (0.85-4.85) | 0.112 | 1.31 (0.47-3.64) | 0.608 | **35** |
| ***PIK3CA*** | **h11000000** | 4.9 | All | 1.1 (0.86-1.42) | 0.444 | 1.12 (0.83-1.5) | 0.463 | 2 |
| | | | Serous | 1.06 (0.78-1.45) | 0.701 | 1.1 (0.9-1.5) | 0.365 | 4 |
| | | | **Endometrioid** | **2.19 (1.1-4.37)** | **0.026** | 1.7 (0.81-3.55) | 0.157 | **22** |
| | | | Mucinous | 0.76 (0.24-2.46) | 0.651 | 1.52 (0.33-7.08) | 0.59 | **100** |
| | | | Clear cell | 0.38 (0.04-3.48) | 0.394 | 0.44 (0.05-4.05) | 0.47 | **16** |
| *PIK3CA* | h01000100 | 4 | All | 1.01 (0.77-1.33) | 0.943 | 1.06 (0.76-1.46) | 0.739 | 5 |
| | | | Serous | 1.05 (0.73-1.51) | 0.782 | 0.9 (0.7-1.2) | 0.532 | **14** |
| | | | Endometrioid | 0.98 (0.43-2.23) | 0.967 | 0.67 (0.23-1.98) | 0.467 | **32** |
| | | | Mucinous | 2.25 (0.98-5.17) | 0.057 | 1.55 (0.65-3.68) | 0.323 | **31** |
| | | | Clear cell | 1.61 (0.55-4.69) | 0.387 | 2.74 (0.84-8.98) | 0.095 | **70** |
| *PIK3CA* | Rare | 1.6 | All | 1.01 (0.71-1.43) | 0.97 | 0.7 (0.44-1.12) | 0.133 | **31** |
| | | | Serous | 0.99 (0.65-1.52) | 0.971 | 1 (0.7-1.4) | 0.8 | 1 |
| | | | Endometrioid | 0.78 (0.21-2.81) | 0.699 | 0.65 (0.17-2.44) | 0.525 | **17** |
| | | | Mucinous | 0.08 (0-166.25) | 0.524 | 0.23 (0-282.55) | 0.687 | **188** |
| | | | Clear cell | 1.2 (0.13-10.66) | 0.872 | - | - | - |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *PIK3CA*: rs2865084, rs7621329, rs1517586, rs2699905, rs7641889, rs7651265, rs7640662, rs2677760.

## Appendix VIII-A: Univariate and multivariate Cox regression results of *AIFM2* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 1751 | All | 1.04 (0.78-1.39) | 0.788 | 1 (0.72-1.4) | 0.986 | 4 |
| | | | 827 | Serous | 0.97 (0.68-1.39) | 0.878 | 0.91 (0.61-1.36) | 0.652 | 6 |
| ***AIFM2*** | **rs2394655** | 0.04 | 269 | Endometrioid | 0.22 (0.03-1.56) | 0.129 | 0.32 (0.04-2.35) | 0.262 | **45** |
| | | | 189 | **Mucinous** | **4.88 (1.96-12.15)** | **0.001** | **3.05 (1.03-8.98)** | **0.043** | **38** |
| | | | 150 | Clear cell | 1.08 (0.55-2.11) | 0.824 | 1.27 (0.6-2.73) | 0.532 | **18** |
| | | | 1719 | All | 0.97 (0.81-1.15) | 0.731 | 0.92 (0.74-1.14) | 0.445 | 5 |
| | | | 817 | Serous | 0.83 (0.65-1.05) | 0.126 | 0.85 (0.64-1.13) | 0.276 | 2 |
| *AIFM2* | rs7908957 | 0.13 | 264 | Endometrioid | 0.67 (0.36-1.25) | 0.212 | 0.64 (0.32-1.27) | 0.203 | 4 |
| | | | 184 | Mucinous | 1.61 (0.96-2.73) | 0.073 | 1.2 (0.7-2.03) | 0.507 | **25** |
| | | | 159 | Clear cell | 1.24 (0.86-1.77) | 0.244 | 1.18 (0.74-1.88) | 0.482 | 5 |
| | | | 1697 | All | 1.03 (0.82-1.29) | 0.794 | 0.96 (0.73-1.27) | 0.799 | 7 |
| | | | 790 | Serous | 0.99 (0.75-1.32) | 0.968 | 0.94 (0.67-1.33) | 0.738 | 5 |
| ***AIFM2*** | **rs1053495** | 0.07 | 267 | Endometrioid | 0.54 (0.2-1.44) | 0.218 | 0.66 (0.24-1.78) | 0.408 | **22** |
| | | | 185 | **Mucinous** | **2.47 (1.25-4.9)** | **0.01** | 1.32 (0.59-2.99) | 0.5 | **47** |
| | | | 156 | Clear cell | 1.01 (0.63-1.64) | 0.956 | 1.03 (0.56-1.91) | 0.914 | 2 |
| | | | 1770 | All | 0.99 (0.87-1.13) | 0.914 | 0.97 (0.83-1.13) | 0.678 | 2 |
| | | | 835 | Serous | 0.88 (0.74-1.04) | 0.145 | 0.87 (0.72-1.07) | 0.186 | 1 |
| ***AIFM2*** | **rs2894111** | 0.28 | 276 | Endometrioid | 1.06 (0.72-1.56) | 0.766 | 1.2 (0.76-1.89) | 0.434 | **13** |
| | | | 192 | **Mucinous** | **1.59 (1.03-2.46)** | **0.036** | 1.37 (0.85-2.19) | 0.194 | **14** |
| | | | 164 | Clear cell | 1.04 (0.8-1.37) | 0.753 | 1.04 (0.76-1.43) | 0.8 | 0 |
| ***AIFM2*** | **rs2394656** | 0.19 | 913 | All | 1.01 (0.87-1.17) | 0.929 | 0.94 (0.79-1.14) | 0.542 | 7 |
| | | | 506 | Serous | 0.89 (0.73-1.08) | 0.244 | 0.96 (0.76-1.21) | 0.736 | 8 |
| | | | 136 | Endometrioid | 0.96 (0.61-1.53) | 0.88 | 0.9 (0.53-1.55) | 0.707 | 6 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 85 | **Mucinous** | **2.01 (1.19-3.4)** | **0.009** | 1.31 (0.76-2.26) | 0.326 | **35** |
| | | | 78 | Clear cell | 1.11 (0.82-1.51) | 0.509 | 1.06 (0.73-1.55) | 0.743 | 5 |
| | | | 422 | All | 0.99 (0.86-1.13) | 0.835 | 1 (0.85-1.17) | 0.973 | 1 |
| | | | 261 | Serous | 0.88 (0.73-1.04) | 0.14 | 0.87 (0.71-1.07) | 0.191 | 1 |
| *AIFM2* | rs6480440 | 0.24 | 254 | Endometrioid | 1 (0.66-1.51) | 0.998 | 1.23 (0.77-1.96) | 0.397 | **23** |
| | | | 180 | Mucinous | 1.41 (0.9-2.21) | 0.13 | 1.43 (0.88-2.32) | 0.145 | 1 |
| | | | 145 | Clear cell | 1.1 (0.81-1.49) | 0.546 | 1.15 (0.82-1.61) | 0.425 | 5 |
| | | | 1313 | All | 0.93 (0.77-1.11) | 0.392 | 0.95 (0.77-1.17) | 0.617 | 2 |
| | | | 556 | Serous | 0.81 (0.63-1.03) | 0.08 | 0.87 (0.66-1.13) | 0.296 | 7 |
| ***AIFM2*** | **rs2280201** | 0.12 | 216 | **Endometrioid** | 1.44 (0.84-2.45) | 0.182 | **2.03 (1.13-3.65)** | **0.018** | **41** |
| | | | 146 | Mucinous | 0.93 (0.49-1.78) | 0.833 | 2.02 (0.96-4.24) | 0.065 | **117** |
| | | | 150 | Clear cell | 1.01 (0.71-1.45) | 0.94 | 0.86 (0.56-1.33) | 0.496 | **15** |
| | | | 1285 | All | 0.95 (0.77-1.17) | 0.597 | 0.91 (0.71-1.18) | 0.477 | 4 |
| | | | 600 | Serous | 1.03 (0.79-1.33) | 0.842 | 0.93 (0.69-1.26) | 0.653 | 10 |
| *AIFM2* | rs10999147 | 0.09 | 260 | Endometrioid | 0.79 (0.37-1.66) | 0.532 | 0.65 (0.28-1.52) | 0.319 | **18** |
| | | | 183 | Mucinous | 0.63 (0.23-1.69) | 0.359 | 0.78 (0.29-2.14) | 0.636 | **24** |
| | | | 151 | Clear cell | 0.92 (0.59-1.43) | 0.699 | 0.89 (0.51-1.57) | 0.686 | 3 |
| | | | 1743 | All | 0.89 (0.70-1.14) | 0.363 | 1.01 (0.77-1.33) | 0.919 | **13** |
| | | | 831 | Serous | 0.78 (0.58-1.05) | 0.101 | 0.98 (0.71-1.36) | 0.922 | **26** |
| *AIFM2* | rs3750772 | 0.06 | 266 | Endometrioid | 1.14 (0.56-2.34) | 0.712 | 1.57 (0.77-3.21) | 0.215 | **38** |
| | | | 0 | Mucinous | 0.76 (0.18-3.18) | 0.71 | 0.62 (0.08-4.65) | 0.646 | **18** |
| | | | 0 | Clear cell | 1 (0.59-1.7) | 0.995 | 0.98 (0.53-1.8) | 0.938 | 2 |
| *AIFM2* | rs4295944 | 0.42 | 1335 | All | 0.93 (0.83-1.04) | 0.209 | 0.92 (0.8-1.05) | 0.198 | 1 |
| | | | 567 | Serous | 0.97 (0.84-1.13) | 0.708 | 0.9 (0.76-1.07) | 0.23 | 7 |
| | | | 220 | Endometrioid | 1.13 (0.8-1.6) | 0.485 | 1.05 (0.72-1.53) | 0.801 | 7 |
| | | | 149 | Mucinous | 0.93 (0.6-1.43) | 0.743 | 0.99 (0.56-1.75) | 0.964 | 6 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 133 | Clear cell | 0.78 (0.6-1.02) | 0.073 | 0.81 (0.6-1.08) | 0.145 | 4 |
| *AIFM2* | rs2394644 | 0.13 | 1324 | All | 1.10 (0.93-1.30) | 0.26 | 1.06 (0.87-1.29) | 0.581 | 4 |
| | | | 561 | Serous | 1.02 (0.83-1.26) | 0.843 | 1.09 (0.85-1.39) | 0.506 | 7 |
| | | | 218 | **Endometrioid** | **1.74 (1.07-2.82)** | **0.025** | 1.67 (0.98-2.87) | 0.061 | 4 |
| | | | 149 | Mucinous | 0.46 (0.17-1.25) | 0.129 | 0.68 (0.24-1.9) | 0.462 | **48** |
| | | | 133 | Clear cell | 1.12 (0.8-1.58) | 0.507 | 0.97 (0.64-1.46) | 0.883 | **13** |
| *AIFM2* | rs10999152 | 0.18 | 1618 | All | 1.05 (0.91-1.22) | 0.509 | 1.1 (0.93-1.31) | 0.276 | 5 |
| | | | 760 | Serous | 1.07 (0.89-1.29) | 0.451 | 1.18 (0.96-1.46) | 0.12 | **10** |
| | | | 251 | Endometrioid | 1.16 (0.74-1.82) | 0.521 | 1.19 (0.73-1.93) | 0.488 | 3 |
| | | | 170 | Mucinous | 0.48 (0.21-1.06) | 0.071 | 0.69 (0.29-1.66) | 0.408 | **44** |
| | | | 153 | Clear cell | 1.06 (0.77-1.45) | 0.719 | 1.08 (0.75-1.54) | 0.69 | 2 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-B: Univariate and multivariate Cox regression results of *AIFM2* haplotypes

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| ***AIFM2* (haplotype block 1)** | **h0000000** | 69 | All | 0.98 (0.87-1.12) | 0.849 | 1.01 (0.87-1.17) | 0.905 | 3 |
| | | | Serous | 1.10 (0.93-1.30) | 0.255 | 1.07 (0.88-1.3) | 0.486 | 3 |
| | | | Endometrioid | 0.91 (0.62-1.33) | 0.614 | 0.8 (0.51-1.23) | 0.308 | **12** |
| | | | **Mucinous** | **0.62 (0.4-0.96)** | **0.032** | 0.76 (0.49-1.19) | 0.235 | **23** |
| | | | Clear cell | 0.95 (0.73-1.24) | 0.726 | 0.96 (0.71-1.32) | 0.822 | 1 |
| ***AIFM2* (haplotype block 1)** | **h0001011** | 7 | All | 0.90 (0.71-1.13) | 0.348 | 1 (0.77-1.3) | 0.98 | **11** |
| | | | Serous | 0.80 (0.60-1.08) | 0.149 | 0.86 (0.62-1.19) | 0.352 | 7 |
| | | | **Endometrioid** | 1.51 (0.81-2.8) | 0.191 | **2.76 (1.36-5.59)** | **0.005** | **83** |
| | | | Mucinous | 0.88 (0.37-2.12) | 0.776 | 1.98 (0.69-5.7) | 0.204 | **125** |
| | | | Clear cell | 0.87 (0.52-1.47) | 0.602 | 0.86 (0.48-1.56) | 0.626 | 1 |
| *AIFM2* (haplotype block 1) | h0001100 | 4 | All | 1.03 (0.75-1.41) | 0.855 | 0.92 (0.62-1.38) | 0.697 | **11** |
| | | | Serous | 0.99 (0.65-1.51) | 0.967 | 1.18 (0.72-1.93) | 0.514 | **19** |
| | | | Endometrioid | 1.33 (0.55-3.24) | 0.532 | 0.87 (0.29-2.66) | 0.811 | **35** |
| | | | Mucinous | 1.98 (0.66-5.92) | 0.222 | 0.65 (0.18-2.34) | 0.51 | **67** |
| | | | Clear cell | 0.88 (0.44-1.78) | 0.719 | 0.94 (0.42-2.07) | 0.871 | 7 |
| ***AIFM2* (haplotype block 1)** | **h1111110** | 4 | All | 1.02 (0.76-1.38) | 0.879 | 0.99 (0.7-1.41) | 0.957 | 3 |
| | | | Serous | 0.91 (0.62-1.32) | 0.612 | 0.82 (0.53-1.27) | 0.379 | 10 |
| | | | Endometrioid | 0.34 (0.07-1.71) | 0.191 | 0.5 (0.1-2.58) | 0.408 | **47** |
| | | | **Mucinous** | **4.87 (1.95-12.17)** | **0.001** | **3.02 (1.02-8.91)** | **0.045** | **38** |
| | | | Clear cell | 1.15 (0.58-2.27) | 0.69 | 1.57 (0.71-3.48) | 0.27 | **37** |
| *AIFM2* (haplotype block 1) | h0001010 | 3 | All | 1.19 (0.85-1.65) | 0.311 | 1.12 (0.75-1.68) | 0.58 | 6 |
| | | | Serous | 1.33 (0.90-1.98) | 0.153 | 1 (0.6-1.67) | 0.997 | **25** |
| | | | Endometrioid | 0.83 (0.27-2.53) | 0.741 | 0.86 (0.22-3.3) | 0.821 | 4 |

| Gene | Haplotype[†] | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Mucinous | 1.09 (0.29-4.05) | 0.901 | 0.93 (0.24-3.65) | 0.919 | **15** |
| | | | Clear cell | 0.96 (0.4-2.32) | 0.923 | 1.12 (0.46-2.73) | 0.802 | **17** |
| *AIFM2* (haplotype block 1) | h0101111 | 3 | All | 1.01 (0.72-1.40) | 0.973 | 0.91 (0.61-1.36) | 0.635 | 10 |
| | | | Serous | 0.91 (0.58-1.42) | 0.675 | 1.06 (0.64-1.78) | 0.814 | **16** |
| | | | Endometrioid | 0.71 (0.22-2.29) | 0.564 | 0.63 (0.19-2.06) | 0.445 | **11** |
| | | | Mucinous | 1.08 (0.35-3.33) | 0.898 | 2.12 (0.65-6.92) | 0.215 | **96** |
| | | | Clear cell | 1.46 (0.78-2.72) | 0.236 | 1 (0.44-2.24) | 0.994 | **32** |
| *AIFM2* (haplotype block 1) | h0111110 | 3 | All | 1.07 (0.75-1.52) | 0.715 | 1.11 (0.71-1.74) | 0.649 | 4 |
| | | | Serous | 1.25 (0.79-1.98) | 0.337 | 1.53 (0.87-2.69) | 0.142 | **22** |
| | | | Endometrioid | 0.91 (0.25-3.3) | 0.886 | 0.81 (0.22-2.98) | 0.756 | **11** |
| | | | Mucinous | 1.02 (0.31-3.35) | 0.973 | 0.55 (0.16-1.87) | 0.339 | **46** |
| | | | Clear cell | 0.92 (0.44-1.9) | 0.816 | 0.95 (0.36-2.51) | 0.925 | 3 |
| *AIFM2* (haplotype block 1) | Rare | 8 | All | 1.02 (0.81-1.30) | 0.853 | 0.97 (0.72-1.3) | 0.824 | 5 |
| | | | Serous | 0.81 (0.58-1.12) | 0.207 | 0.84 (0.56-1.26) | 0.4 | 4 |
| | | | Endometrioid | 1.39 (0.72-2.68) | 0.331 | 1.54 (0.76-3.13) | 0.231 | **11** |
| | | | Mucinous | 2.08 (0.89-4.86) | 0.091 | 1.78 (0.68-4.63) | 0.237 | **14** |
| | | | Clear cell | 1.15 (0.72-1.84) | 0.559 | 1.08 (0.62-1.89) | 0.788 | 6 |
| *AIFM2* (haplotype block 2) | h00100 | 38 | All | 0.95 (0.84-1.07) | 0.386 | 0.93 (0.81-1.07) | 0.322 | 2 |
| | | | Serous | 1.00 (0.86-1.16) | 0.989 | 0.93 (0.78-1.11) | 0.431 | 7 |
| | | | Endometrioid | 1.15 (0.79-1.67) | 0.473 | 1.11 (0.74-1.67) | 0.599 | 3 |
| | | | Mucinous | 1.01 (0.65-1.58) | 0.948 | 1.05 (0.59-1.89) | 0.862 | 4 |
| | | | Clear cell | 0.77 (0.58-1.02) | 0.066 | 0.79 (0.58-1.07) | 0.127 | 3 |
| ***AIFM2* (haplotype block 2)** | **h00000** | 35 | All | 1.03 (0.91-1.16) | 0.627 | 1.05 (0.91-1.2) | 0.539 | 2 |
| | | | Serous | 0.97 (0.83-1.13) | 0.696 | 0.99 (0.83-1.18) | 0.873 | 2 |
| | | | Endometrioid | 0.86 (0.6-1.25) | 0.431 | 0.94 (0.63-1.4) | 0.771 | 9 |
| | | | **Mucinous** | **1.72 (1.03-2.88)** | **0.038** | 1.25 (0.65-2.39) | 0.498 | **27** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Clear cell | 1.17 (0.89-1.53) | 0.251 | 1.17 (0.86-1.57) | 0.313 | 0 |
| *AIFM2* (haplotype block 2) | h00011 | 7 | All | 1.15 (0.93-1.42) | 0.207 | 1.05 (0.81-1.36) | 0.699 | 9 |
| | | | Serous | 1.18 (0.90-1.54) | 0.236 | 1.17 (0.85-1.6) | 0.336 | 1 |
| | | | Endometrioid | 1.61 (0.89-2.9) | 0.116 | 1.34 (0.68-2.65) | 0.401 | **17** |
| | | | Mucinous | 0.57 (0.19-1.76) | 0.331 | 0.76 (0.25-2.31) | 0.623 | **33** |
| | | | Clear cell | 0.96 (0.59-1.56) | 0.873 | 0.93 (0.52-1.66) | 0.817 | 3 |
| *AIFM2* (haplotype block 2) | h00001 | 4 | All | 1.07 (0.79-1.45) | 0.651 | 1.21 (0.86-1.71) | 0.279 | **13** |
| | | | Serous | 1.18 (0.79-1.75) | 0.414 | 1.37 (0.89-2.11) | 0.155 | **16** |
| | | | Endometrioid | 0.27 (0.05-1.34) | 0.11 | 0.33 (0.07-1.51) | 0.153 | **22** |
| | | | Mucinous | 0.47 (0.11-1.97) | 0.304 | 1.31 (0.29-5.91) | 0.725 | **179** |
| | | | Clear cell | **1.77 (1-3.12)** | 0.05 | **2.29 (1.23-4.28)** | **0.009** | **29** |
| *AIFM2* (haplotype block 2) | h10000 | 3 | All | 1.06 (0.74-1.53) | 0.735 | 1.05 (0.68-1.64) | 0.819 | 1 |
| | | | Serous | 1.13 (0.74-1.75) | 0.569 | 1.12 (0.68-1.85) | 0.645 | 1 |
| | | | Endometrioid | 0.58 (0.1-3.52) | 0.558 | 0.55 (0.1-3) | 0.49 | 5 |
| | | | Mucinous | 1.22 (0.3-4.92) | 0.778 | 0.93 (0.23-3.69) | 0.918 | **24** |
| | | | Clear cell | 0.71 (0.28-1.78) | 0.461 | 0.78 (0.23-2.65) | 0.694 | 10 |
| *AIFM2* (haplotype block 2) | h00010 | 2 | All | 1.14 (0.56-2.33) | 0.722 | 0.8 (0.31-2.07) | 0.642 | **30** |
| | | | Serous | 0.78 (0.26-2.33) | 0.652 | 1 (0.26-3.84) | 0.995 | **28** |
| | | | Endometrioid | 2.73 (0.35-21.44) | 0.34 | 1.01 (0-274.28) | 0.998 | **63** |
| | | | Mucinous | 0.13 (0-25.3) | 0.452 | 0.21 (0-137.26) | 0.638 | **62** |
| | | | Clear cell | 1.84 (0.59-5.67) | 0.291 | 0.89 (0.22-3.54) | 0.864 | **52** |
| ***AIFM2* (haplotype block 2)** | **h01011** | 2 | All | 0.93 (0.64-1.35) | 0.702 | 1.16 (0.77-1.74) | 0.476 | **25** |
| | | | Serous | 0.73 (0.45-1.17) | 0.188 | 1.04 (0.63-1.73) | 0.879 | **42** |
| | | | **Endometrioid** | **2.74 (1.07-7.04)** | **0.036** | **5.31 (2.04-13.8)** | **0.001** | **94** |
| | | | Mucinous | - | - | - | 1 | - |
| | | | Clear cell | 1 (0.43-2.36) | 0.992 | 0.62 (0.22-1.73) | 0.36 | **38** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *AIFM2* (haplotype block 2) | h10011 | 2 | All | 1.09 (0.70-1.70) | 0.702 | 0.92 (0.52-1.62) | 0.76 | **16** |
| | | | Serous | 1.16 (0.67-1.99) | 0.599 | 0.98 (0.5-1.91) | 0.945 | **16** |
| | | | Endometrioid | 0.73 (0.12-4.39) | 0.733 | 0.67 (0.11-3.95) | 0.66 | 8 |
| | | | Mucinous | - | - | - | - | - |
| | | | Clear cell | 1.36 (0.57-3.26) | 0.49 | 1.28 (0.39-4.24) | 0.688 | 6 |
| *AIFM2* (haplotype block 2) | h10100 | 2 | All | 0.78 (0.49-1.23) | 0.285 | 0.72 (0.41-1.24) | 0.233 | 8 |
| | | | Serous | 0.91 (0.49-1.70) | 0.77 | 0.63 (0.3-1.29) | 0.206 | **31** |
| | | | Endometrioid | 1.06 (0.35-3.22) | 0.921 | 0.64 (0.14-3) | 0.573 | **40** |
| | | | Mucinous | 0.25 (0.02-2.94) | 0.27 | 0.41 (0.03-5.59) | 0.507 | **64** |
| | | | Clear cell | 0.8 (0.31-2.05) | 0.645 | 0.62 (0.19-2.03) | 0.432 | **23** |
| *AIFM2* (haplotype block 2) | Rare | 4 | All | 0.89 (0.64-1.23) | 0.476 | 0.99 (0.67-1.46) | 0.96 | **11** |
| | | | Serous | 0.89 (0.59-1.34) | 0.577 | 1.06 (0.68-1.66) | 0.786 | **19** |
| | | | Endometrioid | 0.54 (0.16-1.83) | 0.322 | 0.65 (0.2-2.17) | 0.486 | **20** |
| | | | Mucinous | 1.13 (0.27-4.84) | 0.866 | 0.99 (0.13-7.52) | 0.989 | **12** |
| | | | Clear cell | 1.03 (0.49-2.15) | 0.937 | 1.57 (0.68-3.66) | 0.293 | **52** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *AIFM2* - block 1: rs2394655, rs7908957, rs1053495, rs2894111, rs2394656, rs6480440, rs2280201. *AIFM2* -block 2: rs10999147, rs3750772, rs4295944, rs2394644, rs10999152.

## Appendix VIII-C: Univariate and multivariate Cox regression results of *AKTIP* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *AKTIP* | rs9931702 | 0.44 | 917 | All | 1.07 (0.96-1.20) | 0.229 | 1 (0.88-1.14) | 0.981 | 7 |
| | | | 506 | Serous | 1.10 (0.95-1.26) | 0.213 | 0.98 (0.84-1.16) | 0.843 | **11** |
| | | | 137 | Endometrioid | 0.93 (0.65-1.32) | 0.675 | 0.99 (0.68-1.46) | 0.974 | 6 |
| | | | 86 | Mucinous | 0.99 (0.62-1.6) | 0.981 | 0.87 (0.5-1.5) | 0.605 | **12** |
| | | | 79 | Clear cell | 1.03 (0.81-1.3) | 0.821 | 1.07 (0.82-1.41) | 0.608 | 4 |
| *AKTIP* | rs17801966 | 0.15 | 828 | All | 0.97 (0.82-1.14) | 0.695 | 0.99 (0.82-1.2) | 0.91 | 2 |
| | | | 450 | Serous | 0.93 (0.75-1.14) | 0.477 | 0.92 (0.72-1.17) | 0.475 | 1 |
| | | | 125 | Endometrioid | 0.9 (0.52-1.54) | 0.692 | 0.96 (0.54-1.69) | 0.88 | 7 |
| | | | 0 | Mucinous | 1.42 (0.73-2.77) | 0.299 | 1.62 (0.75-3.51) | 0.222 | **14** |
| | | | 78 | Clear cell | 0.95 (0.67-1.33) | 0.762 | 1.04 (0.71-1.53) | 0.845 | 9 |
| ***AKTIP*** | **rs7189819** | 0.3 | 1745 | All | 1.12 (1.00-1.27) | 0.056 | 1.02 (0.88-1.18) | 0.764 | 9 |
| | | | 825 | **Serous** | **1.17 (1.01-1.36)** | **0.036** | 1.05 (0.88-1.26) | 0.552 | **10** |
| | | | 271 | Endometrioid | 0.92 (0.61-1.38) | 0.685 | 0.89 (0.56-1.41) | 0.618 | 3 |
| | | | 186 | Mucinous | 0.83 (0.49-1.43) | 0.51 | 0.72 (0.39-1.33) | 0.295 | **13** |
| | | | 163 | Clear cell | 1.13 (0.88-1.45) | 0.324 | 1.07 (0.8-1.43) | 0.665 | 5 |
| *AKTIP* | rs3743772 | 0.07 | 413 | All | 0.95 (0.74-1. 23) | 0.709 | 0.9 (0.67-1.23) | 0.52 | 5 |
| | | | 256 | Serous | 0.92 (0.67-1.27) | 0.63 | 0.94 (0.66-1.33) | 0.708 | 2 |
| | | | 199 | Endometrioid | 0.4 (0.1-1.6) | 0.195 | 0.27 (0.04-2.03) | 0.205 | **33** |
| | | | 143 | Mucinous | 0.62 (0.15-2.47) | 0.494 | 0.84 (0.2-3.58) | 0.816 | **35** |
| | | | 109 | Clear cell | 1.11 (0.68-1.79) | 0.678 | 1.07 (0.62-1.86) | 0.81 | 4 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-D: Univariate and multivariate Cox regression results of *AKTIP* haplotypes

| Gene | Haplotype[†] | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *AKTIP* | h0000 | 54 | All | 0.93 (0.83-1.04) | 0.196 | 1 (0.88-1.13) | 0.957 | 8 |
| | | | Serous | 0.91 (0.79-1.05) | 0.199 | 1 (0.85-1.18) | 0.954 | 10 |
| | | | Endometrioid | 1.05 (0.74-1.5) | 0.784 | 1 (0.68-1.47) | 0.993 | 5 |
| | | | Mucinous | 0.98 (0.61-1.56) | 0.927 | 1.07 (0.63-1.83) | 0.797 | 9 |
| | | | Clear cell | 1 (0.79-1.26) | 0.983 | 1.02 (0.79-1.33) | 0.866 | 2 |
| *AKTIP* | h1100 | 8 | All | 0.94 (0.75-1.18) | 0.59 | 0.97 (0.74-1.26) | 0.791 | 3 |
| | | | Serous | 0.85 (0.64-1.14) | 0.292 | 0.8 (0.57-1.12) | 0.195 | 6 |
| | | | Endometrioid | 1.36 (0.67-2.74) | 0.392 | 1.52 (0.75-3.1) | 0.247 | **12** |
| | | | Mucinous | 1.69 (0.78-3.65) | 0.183 | 1.77 (0.75-4.18) | 0.192 | 5 |
| | | | Clear cell | 0.8 (0.47-1.36) | 0.404 | 0.95 (0.5-1.77) | 0.863 | **19** |
| *AKTIP* | h1101 | 5 | All | 1.01 (0.78-1.30) | 0.965 | 1.01 (0.75-1.36) | 0.95 | 0 |
| | | | Serous | 1.00 (0.74-1.36) | 0.998 | 1.01 (0.72-1.41) | 0.97 | 1 |
| | | | Endometrioid | 0.41 (0.12-1.4) | 0.153 | 0.34 (0.07-1.56) | 0.165 | **17** |
| | | | Mucinous | 0.84 (0.22-3.15) | 0.795 | 1.09 (0.28-4.32) | 0.898 | **30** |
| | | | Clear cell | 1.17 (0.67-2.05) | 0.573 | 1.38 (0.76-2.5) | 0.294 | **18** |
| *AKTIP* | h1010 | 3 | All | 1.09 (0.97-1.24) | 0.15 | 1 (0.87-1.16) | 0.957 | 8 |
| | | | Serous | 1.15 (0.99-1.34) | 0.072 | 1.03 (0.86-1.23) | 0.773 | **10** |
| | | | Endometrioid | 0.94 (0.62-1.41) | 0.757 | 0.96 (0.61-1.51) | 0.863 | 2 |
| | | | Mucinous | 0.76 (0.43-1.35) | 0.352 | 0.62 (0.32-1.2) | 0.154 | **18** |
| | | | Clear cell | 1.07 (0.84-1.38) | 0.575 | 1.07 (0.79-1.43) | 0.671 | 0 |
| *AKTIP* | Rare | 2 | All | 1.12 (0.82-1.53) | 0.46 | 1.07 (0.73-1.57) | 0.732 | 4 |
| | | | Serous | 1.14 (0.75-1.74) | 0.528 | 1.34 (0.82-2.18) | 0.243 | **18** |
| | | | Endometrioid | 1.45 (0.55-3.84) | 0.453 | 1.28 (0.4-4.05) | 0.675 | **12** |
| | | | Mucinous | 1.42 (0.58-3.49) | 0.442 | 1.5 (0.5-4.51) | 0.473 | 6 |
| | | | **Clear cell** | 0.75 (0.38-1.51) | 0.425 | **0.31 (0.1-0.97)** | **0.043** | **59** |

SNP order in haplotypes, 5' to 3', *ATKIP*: rs9931702, rs17801966, rs7189819, rs3743772.

## Appendix VIII-E: Univariate and multivariate Cox regression results of *AXIN2* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *AXIN2* | rs11868547 | 0.48 | 919 | All | 0.97 (0.86-1.08) | 0.553 | 0.94 (0.83-1.08) | 0.381 | 3 |
| | | | 509 | Serous | 0.90 (0.78-1.04) | 0.171 | 0.9 (0.76-1.06) | 0.212 | 0 |
| | | | 136 | Endometrioid | 0.91 (0.64-1.29) | 0.602 | 0.89 (0.61-1.29) | 0.525 | 2 |
| | | | 85 | Mucinous | 1.18 (0.74-1.88) | 0.492 | 1.11 (0.71-1.73) | 0.655 | 6 |
| | | | 81 | Clear cell | 1.04 (0.81-1.34) | 0.772 | 1 (0.74-1.35) | 0.983 | 4 |
| *AXIN2* | rs7591 | 0.38 | 1779 | All | 1.06 (0.94-1.19) | 0.347 | 1.03 (0.9-1.18) | 0.706 | 3 |
| | | | 838 | Serous | 1.10 (0.95-1.28) | 0.215 | 1.06 (0.9-1.26) | 0.478 | 4 |
| | | | 277 | Endometrioid | 1.14 (0.8-1.62) | 0.463 | 1.2 (0.81-1.78) | 0.364 | 5 |
| | | | 195 | Mucinous | 1.12 (0.71-1.77) | 0.618 | 1 (0.64-1.55) | 0.985 | **11** |
| | | | 165 | Clear cell | 0.93 (0.73-1.2) | 0.596 | 1.03 (0.76-1.38) | 0.86 | **11** |
| *AXIN2* | rs4074947 | 0.22 | 1775 | All | 1.06 (0.92-1.22) | 0.417 | 1.05 (0.89-1.24) | 0.53 | 1 |
| | | | 840 | Serous | 1.07 (0.89-1.28) | 0.482 | 1.06 (0.86-1.31) | 0.586 | 1 |
| | | | 276 | Endometrioid | 1.06 (0.7-1.62) | 0.775 | 1.19 (0.74-1.93) | 0.47 | **12** |
| | | | 192 | Mucinous | 1.44 (0.91-2.28) | 0.124 | 1.39 (0.88-2.19) | 0.158 | 3 |
| | | | 163 | Clear cell | 1.01 (0.75-1.35) | 0.96 | 1.02 (0.72-1.44) | 0.915 | 1 |
| *AXIN2* | rs7210356 | 0.11 | 1777 | All | 0.99 (0.82-1.19) | 0.922 | 0.91 (0.73-1.14) | 0.409 | 8 |
| | | | 838 | Serous | 1.02 (0.81-1.29) | 0.836 | 0.92 (0.7-1.2) | 0.523 | 10 |
| | | | 277 | Endometrioid | 1.01 (0.56-1.84) | 0.962 | 1.15 (0.59-2.23) | 0.687 | **14** |
| | | | 193 | Mucinous | 0.61 (0.26-1.41) | 0.248 | 0.6 (0.24-1.53) | 0.287 | 2 |
| | | | 165 | Clear cell | 0.98 (0.65-1.47) | 0.91 | 1.06 (0.64-1.76) | 0.827 | 8 |
| *AXIN2* | rs11655966 | 0.27 | 1301 | All | 1.00 (0.88-1.13) | 0.971 | 1.02 (0.88-1.19) | 0.753 | 2 |
| | | | 552 | Serous | 0.98 (0.83-1.16) | 0.847 | 1 (0.83-1.21) | 0.998 | 2 |
| | | | 213 | Endometrioid | 1.25 (0.85-1.86) | 0.256 | 1.44 (0.94-2.19) | 0.091 | **15** |
| | | | 144 | Mucinous | 0.71 (0.42-1.2) | 0.204 | 0.72 (0.42-1.25) | 0.242 | 1 |
| | | | 130 | Clear cell | 0.98 (0.76-1.27) | 0.885 | 1.13 (0.84-1.52) | 0.41 | **15** |
| ***AXIN2*** | **rs4541111** | 0.48 | 1297 | All | 1.05 (0.94-1.17) | 0.423 | 1.08 (0.94-1.23) | 0.285 | 3 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 554 | **Serous** | **1.16 (1.00-1.34)** | **0.043** | 1.15 (0.97-1.35) | 0.108 | 1 |
| | | | 214 | Endometrioid | 1.07 (0.76-1.51) | 0.685 | 1.07 (0.74-1.55) | 0.712 | 0 |
| | | | 141 | Mucinous | 0.75 (0.45-1.24) | 0.261 | 0.96 (0.6-1.53) | 0.859 | **28** |
| | | | 128 | Clear cell | 0.91 (0.71-1.18) | 0.493 | 0.98 (0.73-1.32) | 0.917 | 8 |
| | | | 1185 | All | 1.06 (0.94-1.20) | 0.349 | 1 (0.86-1.16) | 0.988 | 6 |
| | | | 539 | Serous | 1.11 (0.95-1.30) | 0.176 | 1.04 (0.87-1.26) | 0.649 | 6 |
| *AXIN2* | rs4791171 | 0.3 | 180 | Endometrioid | 1.05 (0.72-1.52) | 0.8 | 1.1 (0.73-1.66) | 0.642 | 5 |
| | | | 133 | Mucinous | 0.94 (0.57-1.54) | 0.806 | 1.01 (0.61-1.67) | 0.973 | 7 |
| | | | 111 | Clear cell | 1.06 (0.8-1.39) | 0.686 | 1.07 (0.77-1.5) | 0.686 | 1 |
| | | | 839 | All | 1.05 (0.90-1.22) | 0.527 | 0.98 (0.81-1.18) | 0.804 | 7 |
| | | | 326 | Serous | 1.11 (0.91-1.34) | 0.301 | 1.07 (0.85-1.35) | 0.558 | 4 |
| *AXIN2* | rs11079571 | 0.17 | 137 | Endometrioid | 1.12 (0.71-1.77) | 0.612 | 1.1 (0.65-1.86) | 0.721 | 2 |
| | | | 104 | Mucinous | 1.12 (0.68-1.82) | 0.662 | 0.88 (0.51-1.51) | 0.634 | **21** |
| | | | 83 | Clear cell | 0.86 (0.61-1.23) | 0.42 | 0.98 (0.65-1.49) | 0.932 | **14** |
| | | | 1780 | All | 1.05 (0.92-1.21) | 0.477 | 1.05 (0.88-1.25) | 0.614 | 0 |
| | | | 843 | Serous | 1.11 (0.93-1.33) | 0.25 | 1.16 (0.93-1.44) | 0.181 | 5 |
| *AXIN2* | rs3923087 | 0.22 | 275 | Endometrioid | 1.2 (0.81-1.79) | 0.359 | 1.26 (0.8-1.97) | 0.314 | 5 |
| | | | 193 | Mucinous | 1.12 (0.67-1.89) | 0.659 | 1.12 (0.65-1.91) | 0.689 | 0 |
| | | | 164 | Clear cell | 0.94 (0.69-1.27) | 0.664 | 0.84 (0.57-1.23) | 0.368 | **11** |
| | | | 1753 | All | 0.98 (0.87-1.1) | 0.689 | 0.99 (0.86-1.14) | 0.896 | 1 |
| | | | 833 | Serous | 1.05 (0.91-1.21) | 0.535 | 1.09 (0.91-1.3) | 0.345 | 4 |
| *AXIN2* | rs3923086 | 0.42 | 267 | Endometrioid | 0.88 (0.61-1.26) | 0.474 | 0.92 (0.62-1.36) | 0.666 | 5 |
| | | | 188 | Mucinous | 0.79 (0.48-1.29) | 0.347 | 1.04 (0.64-1.7) | 0.859 | **32** |
| | | | 163 | Clear cell | 1.01 (0.79-1.3) | 0.92 | 0.87 (0.64-1.18) | 0.379 | **14** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-F: Univariate and multivariate Cox regression results of *AXIN2* haplotypes

| Gene | Haplotype[†] | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI)e | P-value | |
| *AXIN2* (haplotype block 1) | h100000 | 46 | All | 0.95 (0.85-1.07) | 0.407 | 0.93 (0.81-1.06) | 0.291 | 2 |
| | | | Serous | 0.90 (0.78-1.04) | 0.164 | 0.9 (0.76-1.06) | 0.205 | 0 |
| | | | Endometrioid | 0.89 (0.63-1.25) | 0.508 | 0.83 (0.57-1.19) | 0.313 | 7 |
| | | | Mucinous | 1.21 (0.75-1.95) | 0.431 | 1.11 (0.71-1.75) | 0.653 | 8 |
| | | | Clear cell | 1.02 (0.8-1.3) | 0.884 | 1 (0.75-1.34) | 0.985 | 2 |
| *AXIN2* (haplotype block 1) | h000001 | 14 | All | 0.95 (0.79-1.14) | 0.592 | 1.09 (0.89-1.34) | 0.406 | **15** |
| | | | Serous | 1.04 (0.82-1.33) | 0.755 | 1.1 (0.84-1.44) | 0.475 | 6 |
| | | | Endometrioid | 0.93 (0.55-1.57) | 0.785 | 0.93 (0.53-1.62) | 0.791 | 0 |
| | | | Mucinous | 0.61 (0.29-1.27) | 0.184 | 1.04 (0.45-2.38) | 0.928 | **70** |
| | | | Clear cell | 1 (0.68-1.47) | 0.998 | 1.02 (0.66-1.56) | 0.931 | 2 |
| *AXIN2* (haplotype block 1) | h011011 | 6 | All | 0.92 (0.71-1.19) | 0.537 | 0.95 (0.68-1.32) | 0.756 | 3 |
| | | | Serous | 0.99 (0.71-1.39) | 0.962 | 0.99 (0.66-1.49) | 0.971 | 0 |
| | | | Endometrioid | 1.02 (0.49-2.14) | 0.951 | 1.11 (0.47-2.6) | 0.812 | 9 |
| | | | Mucinous | 0.74 (0.26-2.09) | 0.57 | 1.06 (0.37-3.08) | 0.912 | **43** |
| | | | Clear cell | 0.87 (0.5-1.53) | 0.629 | 1.06 (0.53-2.11) | 0.874 | **22** |
| *AXIN2* (haplotype block 1) | h010011 | 5 | All | 1.09 (0.86-1.38) | 0.484 | 1.12 (0.85-1.47) | 0.414 | 3 |
| | | | Serous | 1.15 (0.86-1.55) | 0.344 | 1.23 (0.89-1.72) | 0.213 | 7 |
| | | | Endometrioid | 1.46 (0.78-2.73) | 0.235 | 1.24 (0.62-2.5) | 0.54 | **15** |
| | | | Mucinous | 0.58 (0.17-2) | 0.385 | 0.28 (0.06-1.23) | 0.091 | **52** |
| | | | Clear cell | 0.88 (0.5-1.53) | 0.648 | 1.3 (0.72-2.37) | 0.388 | **48** |
| *AXIN2* (haplotype block 1) | h011010 | 4 | All | 1.06 (0.77-1.44) | 0.728 | 1.21 (0.85-1.72) | 0.299 | **14** |
| | | | Serous | 0.84 (0.53-1.31) | 0.435 | 1.04 (0.64-1.69) | 0.867 | **24** |
| | | | Endometrioid | 1.24 (0.41-3.74) | 0.707 | 2.57 (0.82-8) | 0.104 | **107** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI)e | P-value | |
| | | | Mucinous | 1.8 (0.56-5.82) | 0.324 | 1.39 (0.35-5.63) | 0.64 | **23** |
| | | | Clear cell | 1.25 (0.75-2.1) | 0.389 | 1.39 (0.77-2.49) | 0.272 | **11** |
| *AXIN2* (haplotype block 1) | Rare | 6 | All | 1.00 (0.78-1.29) | 0.976 | 0.94 (0.71-1.25) | 0.681 | 6 |
| | | | Serous | 1.01 (0.74-1.39) | 0.934 | 0.97 (0.69-1.37) | 0.853 | 4 |
| | | | Endometrioid | 1.08 (0.43-2.69) | 0.874 | 1.47 (0.58-3.73) | 0.412 | **36** |
| | | | Mucinous | 1 (0.42-2.38) | 0.995 | 0.75 (0.29-1.94) | 0.55 | **25** |
| | | | Clear cell | 0.94 (0.56-1.55) | 0.801 | 0.6 (0.31-1.18) | 0.139 | **36** |
| *AXIN2* (haplotype block 2) | h0000 | 54 | All | 0.98 (0.87-1.10) | 0.697 | 0.99 (0.86-1.14) | 0.881 | 1 |
| | | | Serous | 0.91 (0.78-1.05) | 0.179 | 0.9 (0.76-1.08) | 0.258 | 1 |
| | | | Endometrioid | 1.1 (0.78-1.57) | 0.583 | 1.03 (0.7-1.51) | 0.893 | 6 |
| | | | Mucinous | 1.3 (0.83-2.06) | 0.256 | 1.21 (0.79-1.85) | 0.387 | 7 |
| | | | Clear cell | 0.95 (0.74-1.21) | 0.659 | 0.99 (0.73-1.33) | 0.933 | 4 |
| *AXIN2* (haplotype block 2) | h0001 | 14 | All | 0.97 (0.81-1.17) | 0.773 | 1.08 (0.87-1.33) | 0.491 | **11** |
| | | | Serous | 1.09 (0.86-1.37) | 0.475 | 1.22 (0.95-1.58) | 0.124 | **12** |
| | | | Endometrioid | 0.61 (0.31-1.19) | 0.149 | 0.61 (0.29-1.26) | 0.18 | 0 |
| | | | Mucinous | 0.72 (0.37-1.4) | 0.337 | 1.02 (0.48-2.16) | 0.963 | **42** |
| | | | Clear cell | 1.05 (0.72-1.54) | 0.795 | 0.84 (0.53-1.32) | 0.442 | **20** |
| *AXIN2* (haplotype block 2) | h1111 | 12 | All | 1.06 (0.90-1.26) | 0.466 | 0.99 (0.8-1.22) | 0.934 | 7 |
| | | | Serous | 1.13 (0.91-1.41) | 0.262 | 1.12 (0.87-1.45) | 0.389 | 1 |
| | | | Endometrioid | 1.13 (0.69-1.86) | 0.621 | 1.09 (0.61-1.93) | 0.771 | 4 |
| | | | Mucinous | 1.23 (0.69-2.19) | 0.485 | 1.04 (0.51-2.1) | 0.918 | **15** |
| | | | Clear cell | 0.86 (0.58-1.29) | 0.467 | 0.86 (0.53-1.41) | 0.555 | 0 |
| *AXIN2* (haplotype block 2) | h1001 | 7 | All | 0.89 (0.70-1.13) | 0.324 | 0.82 (0.62-1.09) | 0.177 | 8 |
| | | | Serous | 0.86 (0.63-1.16) | 0.32 | 0.77 (0.54-1.08) | 0.125 | **10** |
| | | | Endometrioid | 0.66 (0.3-1.47) | 0.31 | 0.74 (0.31-1.77) | 0.499 | **12** |
| | | | Mucinous | 0.57 (0.18-1.8) | 0.337 | 0.83 (0.22-3.18) | 0.784 | **46** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI)e | P-value | |
| | | | Clear cell | 1.35 (0.83-2.19) | 0.229 | 1.45 (0.79-2.65) | 0.226 | 7 |
| *AXIN2* (haplotype block 2) | h1011 | 7 | All | 0.99 (0.78-1.25) | 0.921 | 1.04 (0.78-1.38) | 0.781 | 5 |
| | | | Serous | 1.06 (0.78-1.43) | 0.717 | 1.06 (0.74-1.52) | 0.746 | 0 |
| | | | Endometrioid | 1.24 (0.67-2.32) | 0.49 | 1.35 (0.7-2.58) | 0.367 | 9 |
| | | | Mucinous | 0.89 (0.38-2.08) | 0.792 | 2.39 (0.95-6.02) | 0.065 | **169** |
| | | | Clear cell | 0.9 (0.53-1.52) | 0.69 | 0.7 (0.35-1.4) | 0.314 | **22** |
| *AXIN2* (haplotype block 2) | h1000 | 2 | All | 1.32 (0.92-1.91) | 0.137 | 1.06 (0.68-1.65) | 0.8 | **20** |
| | | | Serous | 1.39 (0.90-2.13) | 0.135 | 1.08 (0.65-1.8) | 0.76 | **22** |
| | | | Endometrioid | 0.98 (0.15-6.31) | 0.981 | 1.07 (0.16-7.29) | 0.942 | 9 |
| | | | Mucinous | 0.82 (0.14-4.76) | 0.822 | 0.22 (0.03-1.55) | 0.13 | **73** |
| | | | Clear cell | 1.39 (0.6-3.24) | 0.442 | 2.2 (0.92-5.28) | 0.077 | **58** |
| *AXIN2* (haplotype block 2) | Rare | 4 | All | 1.09 (0.82-1.45) | 0.548 | 1.13 (0.81-1.57) | 0.483 | 4 |
| | | | Serous | 1.05 (0.72-1.53) | 0.813 | 1.11 (0.74-1.67) | 0.619 | 6 |
| | | | Endometrioid | 1.37 (0.58-3.24) | 0.472 | 1.72 (0.71-4.16) | 0.231 | **26** |
| | | | Mucinous | 0.75 (0.24-2.37) | 0.623 | 0.37 (0.1-1.33) | 0.128 | **51** |
| | | | Clear cell | 1.16 (0.65-2.07) | 0.61 | 1.63 (0.83-3.19) | 0.156 | **41** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *AXIN2* - block 1: rs11868547, rs7591, rs4074947, rs7210356, rs11655966, rs4541111. *AXIN2* - block 2: rs4791171, rs11079571, rs3923087, rs3923086.

## Appendix VIII-G: Univariate and multivariate Cox regression results of *CASP5* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| ***CASP5*** | **rs518604** | 0.46 | 438 | **All** | **1.13 (1.00-1.27)** | **0.042** | 1.08 (0.95-1.24) | 0.243 | 4 |
| | | | 270 | **Serous** | **1.16 (1.00-1.34)** | **0.047** | 1.11 (0.93-1.31) | 0.257 | 4 |
| | | | 56 | Endometrioid | 1.14 (0.79-1.65) | 0.485 | 1.06 (0.74-1.53) | 0.74 | 7 |
| | | | 42 | Mucinous | 0.89 (0.58-1.39) | 0.612 | 0.7 (0.42-1.18) | 0.182 | **21** |
| | | | 32 | Clear cell | 1.14 (0.88-1.47) | 0.321 | 1.12 (0.84-1.49) | 0.455 | 2 |
| ***CASP5*** | **rs523104** | 0.46 | 824 | All | 0.90 (0.80-1.02) | 0.105 | 0.91 (0.78-1.04) | 0.173 | 1 |
| | | | 320 | **Serous** | **0.85 (0.73 – 1.00)** | **0.049** | 0.9 (0.75-1.08) | 0.254 | 6 |
| | | | 131 | Endometrioid | 0.82 (0.55-1.22) | 0.322 | 0.88 (0.59-1.32) | 0.529 | 7 |
| | | | 102 | Mucinous | 1.16 (0.73-1.85) | 0.531 | 1.18 (0.7-2) | 0.531 | 2 |
| | | | 81 | Clear cell | 1 (0.77-1.29) | 0.971 | 0.95 (0.71-1.28) | 0.742 | 5 |
| *CASP5* | rs3181328 | 0.09 | 829 | All | 1.08 (0.89-1.32) | 0.421 | 1.09 (0.86-1.38) | 0.497 | 1 |
| | | | 319 | Serous | 1.16 (0.91-1.49) | 0.232 | 1 (0.75-1.34) | 0.983 | **14** |
| | | | 255 | Endometrioid | 1.28 (0.66-2.51) | 0.465 | 1.32 (0.64-2.7) | 0.455 | 3 |
| | | | 102 | Mucinous | 0.62 (0.26-1.46) | 0.272 | 0.86 (0.35-2.11) | 0.745 | **39** |
| | | | 83 | Clear cell | 0.99 (0.65-1.51) | 0.971 | 1.29 (0.77-2.18) | 0.334 | **30** |
| *CASP5* | rs17446518 | 0.11 | 803 | All | 0.83 (0.68-1.01) | 0.067 | 0.96 (0.77-1.2) | 0.723 | **16** |
| | | | 311 | Serous | 0.92 (0.72-1.17) | 0.485 | 1 (0.76-1.3) | 0.977 | 9 |
| | | | 130 | Endometrioid | 0.72 (0.35-1.51) | 0.389 | 0.96 (0.45-2.04) | 0.921 | **33** |
| | | | 97 | Mucinous | 0.34 (0.11-1.08) | 0.068 | 0.58 (0.17-1.94) | 0.377 | **71** |
| | | | 147 | Clear cell | 0.73 (0.47-1.14) | 0.165 | 0.74 (0.45-1.22) | 0.237 | 1 |
| *CASP5* | rs9651713 | 0.11 | 1730 | All | 0.85 (0.70-1.04) | 0.111 | 0.88 (0.71-1.11) | 0.285 | 4 |
| | | | 819 | Serous | 0.96 (0.76-1.22) | 0.748 | 0.95 (0.72-1.26) | 0.728 | 1 |
| | | | 269 | Endometrioid | 0.93 (0.53-1.63) | 0.791 | 0.89 (0.48-1.67) | 0.715 | 4 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 183 | Mucinous | 0.37 (0.13-1.03) | 0.056 | 0.56 (0.2-1.6) | 0.282 | **51** |
| | | | 162 | Clear cell | 0.77 (0.49-1.23) | 0.275 | 0.81 (0.48-1.38) | 0.441 | 5 |
| | | | 1282 | All | 0.94 (0.81-1.09) | 0.393 | 0.93 (0.78-1.11) | 0.413 | 1 |
| | | | 597 | Serous | 1.06 (0.89-1.27) | 0.514 | 1.07 (0.87-1.32) | 0.495 | 1 |
| ***CASP5*** | **rs3181175** | 0.19 | 194 | Endometrioid | 0.78 (0.48-1.26) | 0.313 | 0.62 (0.35-1.09) | 0.095 | **21** |
| | | | 147 | **Mucinous** | **0.4 (0.19-0.86)** | **0.019** | 0.5 (0.22-1.15) | 0.102 | **25** |
| | | | 115 | Clear cell | 0.94 (0.68-1.32) | 0.732 | 0.92 (0.62-1.37) | 0.682 | 2 |
| | | | 1780 | All | 1.11 (0.90-1.37) | 0.316 | 1.02 (0.8-1.31) | 0.876 | 8 |
| | | | 840 | Serous | 1.20 (0.92-1.55) | 0.174 | 1.2 (0.9-1.61) | 0.209 | 0 |
| *CASP5* | rs3181174 | 0.07 | 278 | Endometrioid | 0.66 (0.29-1.47) | 0.306 | 0.34 (0.11-1.07) | 0.064 | **48** |
| | | | 183 | Mucinous | 0.65 (0.23-1.84) | 0.42 | 0.54 (0.17-1.79) | 0.318 | **17** |
| | | | 165 | Clear cell | 1.27 (0.83-1.94) | 0.268 | 1.24 (0.73-2.11) | 0.434 | 2 |
| | | | 852 | All | 0.94 (0.83-1.06) | 0.329 | 0.94 (0.81-1.09) | 0.442 | 0 |
| | | | 462 | Serous | 1.10 (0.95-1.28) | 0.216 | 1.12 (0.93-1.34) | 0.247 | 2 |
| ***CASP5*** | **rs2282657** | 0.35 | 128 | Endometrioid | 0.82 (0.55-1.22) | 0.327 | 0.78 (0.5-1.2) | 0.254 | 5 |
| | | | 80 | Mucinous | 0.75 (0.47-1.19) | 0.224 | 0.92 (0.57-1.48) | 0.735 | **23** |
| | | | 73 | **Clear cell** | **0.76 (0.57-1)** | **0.049** | **0.68 (0.48-0.96)** | **0.029** | **11** |
| | | | 1768 | All | 0.94 (0.83-1.06) | 0.298 | 0.94 (0.81-1.08) | 0.385 | 0 |
| | | | 835 | Serous | 1.07 (0.92-1.24) | 0.403 | 1.05 (0.87-1.25) | 0.621 | 2 |
| *CASP5* | rs507879 | 0.46 | 276 | Endometrioid | 0.78 (0.52-1.17) | 0.226 | 0.81 (0.52-1.26) | 0.351 | 4 |
| | | | 194 | Mucinous | 0.9 (0.59-1.38) | 0.627 | 1.04 (0.63-1.69) | 0.89 | **16** |
| | | | 164 | Clear cell | 0.83 (0.64-1.07) | 0.145 | 0.77 (0.57-1.03) | 0.079 | 7 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-H: Univariate and multivariate Cox regression results of *CASP5* haplotypes

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P -value | |
| ***CASP5* (haplotype block 1)** | **h010** | 45 | **All** | **0.86 (0.77-0.97)** | **0.014** | 0.89 (0.78-1.03) | 0.116 | 3 |
| | | | **Serous** | **0.81 (0.70-0.94)** | **0.006** | 0.89 (0.75-1.06) | 0.202 | 10 |
| | | | Endometrioid | 0.81 (0.55-1.19) | 0.278 | 0.86 (0.58-1.28) | 0.466 | 6 |
| | | | Mucinous | 1.29 (0.83-1.99) | 0.261 | 1.35 (0.81-2.25) | 0.248 | 5 |
| | | | Clear cell | 0.93 (0.72-1.2) | 0.573 | 0.9 (0.68-1.2) | 0.489 | 3 |
| ***CASP5* (haplotype block 1)** | **h100** | 42 | **All** | **1.12 (1.00-1.26)** | 0.05 | 1.08 (0.94-1.24) | 0.289 | 4 |
| | | | Serous | 1.14 (0.99-1.32) | 0.077 | 1.09 (0.92-1.29) | 0.339 | 4 |
| | | | Endometrioid | 1.14 (0.79-1.64) | 0.486 | 1.04 (0.72-1.49) | 0.847 | 9 |
| | | | Mucinous | 0.99 (0.62-1.59) | 0.965 | 0.79 (0.45-1.37) | 0.401 | **20** |
| | | | Clear cell | 1.11 (0.86-1.44) | 0.422 | 1.09 (0.81-1.46) | 0.561 | 2 |
| *CASP5* (haplotype block 1) | h001 | 9 | All | 1.04 (0.85-1.28) | 0.679 | 1.02 (0.8-1.31) | 0.863 | 2 |
| | | | Serous | 1.14 (0.88-1.46) | 0.324 | 0.96 (0.71-1.29) | 0.765 | **16** |
| | | | Endometrioid | 1.13 (0.56-2.28) | 0.741 | 1.13 (0.53-2.41) | 0.747 | 0 |
| | | | Mucinous | 0.62 (0.26-1.47) | 0.282 | 0.9 (0.37-2.19) | 0.811 | **45** |
| | | | Clear cell | 0.93 (0.6-1.45) | 0.762 | 1.16 (0.67-1.99) | 0.599 | **25** |
| *CASP5* (haplotype block 1) | h000 | 2 | All | 1.04 (0.63-1.73) | 0.871 | 1.19 (0.68-2.09) | 0.549 | **14** |
| | | | Serous | 1.26 (0.68-2.35) | 0.459 | 1.31 (0.68-2.53) | 0.413 | 4 |
| | | | Endometrioid | 0.94 (0.16-5.71) | 0.949 | 1.32 (0.21-8.15) | 0.764 | **40** |
| | | | Mucinous | 1.19 (0.28-5.06) | 0.817 | 3.31 (0.72-15.19) | 0.124 | **178** |
| | | | Clear cell | 0.42 (0.09-2.03) | 0.28 | 0.24 (0.03-2) | 0.186 | **43** |
| *CASP5* (haplotype block 1) | h110 | 2 | All | 1.05 (0.63-1.76) | 0.852 | 1.05 (0.45-2.42) | 0.914 | 0 |
| | | | Serous | 1.34 (0.69-2.59) | 0.387 | 1.94 (0.61-6.18) | 0.261 | **45** |
| | | | Endometrioid | 1.3 (0.2-8.49) | 0.783 | 2.74 (0.43-17.66) | 0.288 | **111** |
| | | | Mucinous | 0.29 (0.04-2.25) | 0.236 | 0.28 (0.04-2.15) | 0.223 | 3 |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P -value | |
| | | | Clear cell | 1.43 (0.47-4.39) | 0.533 | 1.34 (0.26-6.88) | 0.723 | 6 |
| ***CASP5* (haplotype block 1)** | **Rare** | 4 | **All** | **4.56 (1.71-12.12)** | **0.002** | **11.73 (4.14-33.28)** | **$8.85x10^{-}$** | **157** |
| | | | Serous | 2.80 (0.70-11.25) | 0.146 | **16.8 (3.39-83.24)** | **0.001** | **500** |
| | | | **Endometrioid** | **11.54 (1.26-105.63)** | **0.03** | 6.52 (0.77-54.93) | 0.085 | **44** |
| | | | Mucinous | - | - | - | - | - |
| | | | **Clear cell** | 3.82 (0.63-23.06) | 0.144 | **9.13 (1.57-52.94)** | **0.014** | **139** |
| *CASP5* (haplotype block 2) | h000000 | 47 | All | 1.08 (0.96-1.22) | 0.223 | 1.02 (0.79-1.31) | 0.896 | 6 |
| | | | Serous | 0.93 (0.80-1.09) | 0.385 | 0.94 (0.67-1.31) | 0.697 | 1 |
| | | | Endometrioid | 0.93 (0.49-1.76) | 0.816 | 1.15 (0.56-2.32) | 0.706 | **24** |
| | | | Mucinous | 1.42 (0.71-2.87) | 0.324 | 1.23 (0.59-2.57) | 0.588 | **13** |
| | | | Clear cell | 1.17 (0.79-1.74) | 0.431 | 1.16 (0.71-1.9) | 0.542 | 1 |
| ***CASP5* (haplotype block 2)** | **h000011** | 13 | All | 1.02 (0.87-1.20) | 0.821 | 1.01 (0.83-1.23) | 0.916 | 1 |
| | | | Serous | 1.11 (0.91-1.35) | 0.298 | 1.08 (0.86-1.36) | 0.493 | 3 |
| | | | Endometrioid | 1.05 (0.63-1.77) | 0.847 | 1.17 (0.66-2.07) | 0.583 | **11** |
| | | | Mucinous | 1.69 (0.92-3.1) | 0.089 | 1.64 (0.85-3.15) | 0.139 | 3 |
| | | | **Clear cell** | **0.62 (0.4-0.97)** | **0.034** | **0.57 (0.34-0.97)** | **0.037** | 8 |
| *CASP5* (haplotype block 2) | h001111 | 7 | All | 1.14 (0.91-1.43) | 0.251 | 1 (0.76-1.32) | 0.998 | **12** |
| | | | Serous | 1.16 (0.88-1.54) | 0.29 | 1.15 (0.83-1.58) | 0.408 | 1 |
| | | | Endometrioid | 0.74 (0.31-1.77) | 0.498 | 0.34 (0.1-1.13) | 0.079 | **54** |
| | | | Mucinous | 0.74 (0.25-2.17) | 0.588 | 0.56 (0.17-1.87) | 0.347 | **24** |
| | | | Clear cell | 1.3 (0.85-1.99) | 0.224 | 1.3 (0.76-2.22) | 0.343 | 0 |
| *CASP5* (haplotype block 2) | h100000 | 6 | All | 0.93 (0.69-1.25) | 0.615 | 1.13 (0.81-1.58) | 0.479 | **22** |
| | | | Serous | 0.91 (0.63-1.31) | 0.61 | 1.06 (0.7-1.61) | 0.786 | **16** |
| | | | Endometrioid | 1.18 (0.46-3.07) | 0.73 | 1.51 (0.59-3.86) | 0.384 | **28** |
| | | | Mucinous | 0.35 (0.05-2.55) | 0.297 | 0.46 (0.06-3.39) | 0.444 | **31** |
| | | | Clear cell | 0.84 (0.45-1.58) | 0.595 | 1.04 (0.52-2.08) | 0.922 | **24** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P -value | |
| *CASP5* (haplotype block 2) | h100011 | 3 | **All** | **0.5 (0.26-0.98)** | **0.042** | 0.73 (0.34-1.58) | 0.424 | **46** |
| | | | Serous | 0.69 (0.30-1.55) | 0.366 | 0.92 (0.37-2.28) | 0.856 | **33** |
| | | | Endometrioid | 0.26 (0.02-3.18) | 0.294 | 0.53 (0.04-6.45) | 0.619 | **104** |
| | | | Mucinous | - | - | - | - | - |
| | | | Clear cell | 0.64 (0.18-2.36) | 0.507 | 0.34 (0.07-1.78) | 0.202 | **47** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *CASP5* (block 1): rs518604, rs523104, rs3181328. *CASP5* (block 2): rs17446518, rs9651713, rs3181175, rs3181174, rs2282657, rs507879.

## Appendix VIII-I: Univariate and multivariate Cox regression results of *FILIP1L* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *FILIP1L* | rs796977 | 0.33 | 437 | All | 0.94 (0.84-1.06) | 0.328 | 0.94 (0.81-1.08) | 0.38 | 0 |
| | | | 269 | Serous | 1.03 (0.89-1.19) | 0.692 | 1.01 (0.84-1.2) | 0.948 | 2 |
| | | | 56 | Endometrioid | 1.1 (0.75-1.61) | 0.636 | 1.09 (0.74-1.62) | 0.654 | 1 |
| | | | 43 | Mucinous | 0.66 (0.41-1.06) | 0.084 | 0.63 (0.37-1.08) | 0.094 | 5 |
| | | | 31 | Clear cell | 0.84 (0.64-1.1) | 0.216 | 0.82 (0.6-1.11) | 0.196 | 2 |
| *FILIP1L* | rs793477 | 0.13 | 1653 | All | 0.97 (0.81-1.15) | 0.717 | 1.03 (0.84-1.26) | 0.774 | 6 |
| | | | 771 | Serous | 1.11 (0.90-1.36) | 0.337 | 1.1 (0.87-1.38) | 0.43 | 1 |
| | | | 257 | Endometrioid | 0.78 (0.41-1.46) | 0.432 | 0.9 (0.48-1.7) | 0.75 | **15** |
| | | | 175 | Mucinous | 1.01 (0.51-2.02) | 0.976 | 0.97 (0.4-2.36) | 0.955 | 4 |
| | | | 155 | **Clear cell** | **0.62 (0.39-0.99)** | **0.047** | 0.68 (0.4-1.17) | 0.162 | 10 |
| ***FILIP1L*** | **rs793446** | 0.41 | 1773 | All | 0.96 (0.85-1.07) | 0.461 | 1.04 (0.9-1.19) | 0.628 | 8 |
| | | | 838 | Serous | 1.02 (0.88-1.18) | 0.816 | 1.08 (0.91-1.28) | 0.399 | 6 |
| | | | 274 | Endometrioid | 1.08 (0.74-1.59) | 0.69 | 1.15 (0.77-1.71) | 0.501 | 6 |
| | | | 194 | **Mucinous** | 0.65 (0.41-1.03) | 0.065 | **0.57 (0.33-0.99)** | **0.046** | **12** |
| | | | 164 | Clear cell | 0.97 (0.76-1.25) | 0.83 | 0.99 (0.74-1.32) | 0.932 | 2 |
| ***FILIP1L*** | **rs3921767** | 0.07 | 1773 | **All** | 1.03 (0.83-1.29) | 0.786 | **1.39 (1.07-1.81)** | **0.014** | **35** |
| | | | 840 | Serous | 0.98 (0.71-1.34) | 0.895 | 1.28 (0.89-1.84) | 0.186 | **31** |
| | | | 276 | Endometrioid | 0.99 (0.5-1.93) | 0.967 | 1.23 (0.59-2.57) | 0.576 | **24** |
| | | | 191 | Mucinous | 1.03 (0.45-2.34) | 0.949 | 1.09 (0.44-2.73) | 0.849 | 6 |
| | | | 166 | Clear cell | 1.29 (0.88-1.89) | 0.196 | 1.59 (0.99-2.58) | 0.057 | **23** |
| *FILIP1L* | rs17338680 | 0.11 | 1786 | All | 0.96 (0.80-1.15) | 0.675 | 0.99 (0.81-1.22) | 0.932 | 3 |
| | | | 574 | Serous | 0.99 (0.78-1.27) | 0.944 | 0.98 (0.74-1.29) | 0.866 | 1 |
| | | | 221 | Endometrioid | 1.1 (0.69-1.76) | 0.688 | 1.21 (0.76-1.93) | 0.414 | 10 |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 184 | Mucinous | 1.82 (0.93-3.58) | 0.083 | 1.21 (0.58-2.54) | 0.606 | **34** |
| | | | 133 | Clear cell | 0.86 (0.57-1.3) | 0.479 | 0.82 (0.51-1.31) | 0.408 | 5 |
| | | | 1786 | All | 0.96 (0.84-1.09) | 0.515 | 0.93 (0.8-1.09) | 0.366 | 3 |
| | | | 843 | Serous | 1.03 (0.88-1.20) | 0.708 | 1 (0.84-1.2) | 0.964 | 3 |
| ***FILIP1L*** | **rs9864437** | 0.22 | 278 | Endometrioid | 1.07 (0.69-1.68) | 0.752 | 0.97 (0.6-1.57) | 0.892 | 9 |
| | | | 195 | **Mucinous** | **0.45 (0.25-0.82)** | **0.009** | **0.46 (0.23-0.91)** | **0.027** | 2 |
| | | | 165 | Clear cell | 0.89 (0.66-1.2) | 0.455 | 0.87 (0.63-1.2) | 0.396 | 2 |
| | | | 1414 | All | 1.00 (0.89-1.13) | 0.944 | 0.94 (0.81-1.08) | 0.363 | 6 |
| | | | 710 | Serous | 0.97 (0.84-1.13) | 0.727 | 0.92 (0.77-1.1) | 0.368 | 5 |
| *FILIP1L* | rs6788750 | 0.41 | 226 | Endometrioid | 1.05 (0.71-1.55) | 0.803 | 0.92 (0.62-1.39) | 0.703 | **12** |
| | | | 140 | Mucinous | 0.95 (0.58-1.56) | 0.843 | 0.86 (0.48-1.56) | 0.62 | 9 |
| | | | 131 | Clear cell | 0.96 (0.75-1.24) | 0.778 | 0.98 (0.73-1.32) | 0.889 | 2 |
| | | | 1273 | All | 0.96 (0.82-1.12) | 0.573 | 1.1 (0.92-1.32) | 0.277 | **15** |
| | | | 594 | Serous | 0.93 (0.75-1.15) | 0.514 | 1.05 (0.82-1.33) | 0.705 | **13** |
| *FILIP1L* | rs12494994 | 0.18 | 193 | Endometrioid | 1 (0.65-1.53) | 0.986 | 1.16 (0.75-1.8) | 0.512 | **16** |
| | | | 145 | Mucinous | 1.36 (0.75-2.5) | 0.313 | 1.16 (0.62-2.18) | 0.65 | **15** |
| | | | 113 | Clear cell | 1.1 (0.81-1.5) | 0.525 | 1.13 (0.8-1.61) | 0.486 | 3 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-J: Univariate and multivariate Cox regression results of *FILIP1L* haplotypes

| Gene | Haplotype[†] | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *FILIP1L* (haplotype block 1) | h00000 | 47 | All | 1.06 (0.94-1.19) | 0.333 | 0.95 (0.83-1.09) | 0.49 | **10** |
| | | | Serous | 0.95 (0.82-1.10) | 0.485 | 0.89 (0.75-1.06) | 0.211 | 6 |
| | | | Endometrioid | 0.98 (0.67-1.42) | 0.904 | 0.88 (0.59-1.32) | 0.542 | **10** |
| | | | Mucinous | 1.37 (0.91-2.07) | 0.129 | 1.58 (0.94-2.65) | 0.086 | **15** |
| | | | Clear cell | 1.21 (0.95-1.55) | 0.117 | 1.16 (0.87-1.54) | 0.316 | 4 |
| ***FILIP1L* (haplotype block 1)** | h10100 | 21 | All | 0.96 (0.84-1.09) | 0.52 | 0.94 (0.8-1.1) | 0.424 | 2 |
| | | | Serous | 1.03 (0.89-1.21) | 0.667 | 0.94 (0.67-1.31) | 0.697 | 9 |
| | | | Endometrioid | 1.12 (0.71-1.78) | 0.618 | 1 (0.61-1.63) | 0.998 | **11** |
| | | | **Mucinous** | **0.42 (0.22-0.78)** | **0.006** | **0.44 (0.21-0.9)** | **0.024** | 5 |
| | | | Clear cell | 0.88 (0.65-1.19) | 0.419 | 0.88 (0.64-1.22) | 0.447 | 0 |
| ***FILIP1L* (haplotype block 1)** | **h01000** | 12 | All | 0.97 (0.81-1.16) | 0.728 | 1.02 (0.83-1.25) | 0.839 | 5 |
| | | | Serous | 1.09 (0.88-1.34) | 0.427 | 1.08 (0.86-1.36) | 0.516 | 1 |
| | | | Endometrioid | 0.83 (0.44-1.56) | 0.562 | 0.92 (0.49-1.74) | 0.799 | **11** |
| | | | Mucinous | 0.97 (0.48-1.97) | 0.937 | 0.95 (0.39-2.32) | 0.911 | 2 |
| | | | **Clear cell** | **0.62 (0.38-0.99)** | **0.048** | 0.67 (0.39-1.16) | 0.156 | 8 |
| *FILIP1L* (haplotype block 1) | h10101 | 11 | All | 0.96 (0.80-1.15) | 0.651 | 0.99 (0.8-1.23) | 0.94 | 3 |
| | | | Serous | 0.99 (0.77-1.27) | 0.919 | 1.02 (0.85-1.22) | 0.834 | 3 |
| | | | Endometrioid | 1.1 (0.69-1.76) | 0.693 | 1.21 (0.76-1.93) | 0.418 | 10 |
| | | | Mucinous | 1.82 (0.93-3.59) | 0.082 | 1.22 (0.58-2.54) | 0.605 | **33** |
| | | | Clear cell | 0.87 (0.58-1.31) | 0.505 | 0.82 (0.51-1.32) | 0.415 | 6 |
| ***FILIP1L* (haplotype block 1)** | **h00110** | 7 | **All** | 1.02 (0.81-1.27) | 0.871 | **1.36 (1.04-1.77)** | **0.024** | **33** |
| | | | Serous | 0.95 (0.69-1.30) | 0.742 | 1.22 (0.85-1.76) | 0.283 | **28** |
| | | | Endometrioid | 1 (0.5-2) | 0.996 | 1.22 (0.58-2.55) | 0.604 | **22** |
| | | | Mucinous | 1.01 (0.44-2.31) | 0.977 | 1.08 (0.43-2.69) | 0.877 | 7 |
| | | | Clear cell | 1.29 (0.88-1.89) | 0.197 | 1.61 (0.99-2.6) | 0.053 | **25** |
| *FILIP1L* (haplotype | Rare | 2 | All | 1.00 (0.63-1.58) | 0.991 | 0.98 (0.54-1.79) | 0.955 | 2 |
| | | | Serous | 0.76 (0.36-1.62) | 0.479 | 0.98 (0.73-1.3) | 0.876 | **29** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Endometrioid | 0.65 (0.14-2.92) | 0.574 | 0.72 (0.15-3.54) | 0.691 | **11** |
| | | | Mucinous | 1.12 (0.17-7.37) | 0.903 | 0.81 (0.11-6.26) | 0.843 | **28** |
| | | | Clear cell | 1.57 (0.77-3.18) | 0.211 | 1.73 (0.78-3.83) | 0.177 | **10** |
| *FILIP1L* (haplotype block 2) | h010 | 41 | All | 1.01 (0.90-1.14) | 0.894 | 0.94 (0.81-1.08) | 0.385 | 7 |
| | | | Serous | 0.97 (0.84-1.13) | 0.719 | 1.06 (0.83-1.34) | 0.652 | 9 |
| | | | Endometrioid | 1.01 (0.68-1.49) | 0.97 | 0.9 (0.6-1.34) | 0.592 | **11** |
| | | | Mucinous | 0.96 (0.6-1.54) | 0.857 | 0.84 (0.47-1.51) | 0.571 | **13** |
| | | | Clear cell | 0.96 (0.75-1.24) | 0.781 | 1 (0.74-1.34) | 0.989 | 4 |
| ***FILIP1L* (haplotype block 2)** | **h100** | 22 | All | 0.96 (0.84-1.09) | 0.506 | 0.94 (0.8-1.09) | 0.42 | 2 |
| | | | Serous | 1.03 (0.89-1.21) | 0.668 | 0.92 (0.77-1.1) | 0.371 | **11** |
| | | | Endometrioid | 1.08 (0.69-1.69) | 0.727 | 0.97 (0.6-1.57) | 0.902 | **10** |
| | | | **Mucinous** | **0.45 (0.25-0.82)** | **0.009** | **0.46 (0.23-0.91)** | **0.026** | 2 |
| | | | Clear cell | 0.88 (0.65-1.18) | 0.385 | 0.87 (0.63-1.21) | 0.407 | 1 |
| ***FILIP1L* (haplotype block 2)** | **h000** | 19 | All | 1.10 (0.94-1.28) | 0.223 | 1.11 (0.92-1.33) | 0.292 | 1 |
| | | | Serous | 1.07 (0.87-1.31) | 0.506 | 0.64 (0.26-1.55) | 0.321 | **40** |
| | | | Endometrioid | 0.87 (0.53-1.44) | 0.594 | 0.98 (0.59-1.64) | 0.946 | **13** |
| | | | **Mucinous** | 1.72 (1.09-2.72) | 0.019 | **1.96 (1.15-3.33)** | **0.013** | **14** |
| | | | Clear cell | 1.13 (0.82-1.58) | 0.456 | 1.09 (0.74-1.6) | 0.664 | 4 |
| *FILIP1L* (haplotype block 2) | h001 | 17 | All | 0.96 (0.82-1.11) | 0.563 | 1.11 (0.93-1.33) | 0.24 | **16** |
| | | | Serous | 0.93 (0.75-1.15) | 0.494 | 1.09 (0.85-1.38) | 0.505 | **17** |
| | | | Endometrioid | 1.03 (0.68-1.57) | 0.89 | 1.19 (0.78-1.84) | 0.42 | **16** |
| | | | Mucinous | 1.47 (0.81-2.69) | 0.208 | 1.26 (0.66-2.41) | 0.476 | **14** |
| | | | Clear cell | 1.08 (0.79-1.46) | 0.627 | 1.12 (0.79-1.59) | 0.535 | 4 |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *FILIP1L* -block 1: rs796977, rs793477, rs793446, rs3921767, rs17338680. *FILIP1L* - block 2: rs9864437, rs6788750, rs12494994.

## Appendix VIII-K: Univariate and multivariate Cox regression results of *RBBP8* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RBBP8* | rs7239066 | 0.11 | 1272 | All | 1.14 (0.96-1.35) | 0.147 | 1.02 (0.83-1.24) | 0.875 | **11** |
| | | | 594 | Serous | 1.04 (0.83-1.30) | 0.743 | 0.93 (0.72-1.19) | 0.555 | **11** |
| | | | 188 | Endometrioid | 1.21 (0.67-2.19) | 0.524 | 1 (0.54-1.85) | 0.997 | **17** |
| | | | 144 | Mucinous | 0.95 (0.49-1.86) | 0.884 | 0.79 (0.36-1.71) | 0.543 | **17** |
| | | | 115 | Clear cell | 1.35 (0.95-1.93) | 0.098 | 1.35 (0.91-2) | 0.135 | 0 |
| ***RBBP8*** | **rs11082221** | 0.04 | 1748 | All | 1.29 (0.99-1.68) | 0.055 | 1.09 (0.8-1.46) | 0.593 | **16** |
| | | | 826 | Serous | 1.08 (0.77-1.51) | 0.664 | 0.92 (0.62-1.36) | 0.67 | **15** |
| | | | 253 | Endometrioid | 1.69 (0.72-3.95) | 0.228 | 1.3 (0.51-3.33) | 0.587 | **23** |
| | | | 176 | Mucinous | 1.23 (0.43-3.54) | 0.698 | 0.54 (0.16-1.84) | 0.327 | **56** |
| | | | 151 | **Clear cell** | **1.67 (1.01-2.75)** | **0.044** | 1.54 (0.93-2.54) | 0.092 | 8 |
| ***RBBP8*** | **rs4474794** | 0.36 | 1764 | **All** | **0.85 (0.75-0.95)** | **0.007** | **0.86 (0.74-0.99)** | **0.034** | 1 |
| | | | 829 | Serous | 0.88 (0.75-1.02) | 0.098 | 0.85 (0.71-1.01) | 0.065 | 3 |
| | | | 271 | Endometrioid | 0.8 (0.53-1.19) | 0.265 | 0.86 (0.56-1.31) | 0.479 | 7 |
| | | | 193 | Mucinous | 0.67 (0.42-1.05) | 0.079 | 0.83 (0.51-1.36) | 0.465 | **24** |
| | | | 165 | Clear cell | 0.91 (0.71-1.18) | 0.484 | 0.98 (0.73-1.32) | 0.899 | 8 |
| ***RBBP8*** | **rs9304261** | 0.22 | 346 | **All** | **0.83 (0.71-0.95)** | **0.009** | **0.83 (0.7-0.99)** | **0.038** | 0 |
| | | | 215 | Serous | 0.87 (0.72-1.05) | 0.143 | 0.82 (0.66-1.02) | 0.073 | 6 |
| | | | 44 | Endometrioid | 0.87 (0.56-1.35) | 0.536 | 0.99 (0.62-1.6) | 0.982 | **14** |
| | | | 33 | Mucinous | 0.61 (0.35-1.05) | 0.074 | 0.81 (0.44-1.49) | 0.497 | **33** |
| | | | 21 | Clear cell | 0.84 (0.62-1.13) | 0.242 | 0.86 (0.61-1.22) | 0.401 | 2 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-L: Univariate and multivariate Cox regression results of *RBBP8* haplotypes

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| ***RBBP8*** | **h0000** | 62 | **All** | **1.16 (1.03-1.31)** | **0.015** | **1.17 (1.01-1.34)** | **0.032** | 1 |
| | | | Serous | 1.12 (0.97-1.30) | 0.134 | 0.02 (0-27735) | 0.589 | **98** |
| | | | Endometrioid | 1.26 (0.85-1.86) | 0.257 | 1.15 (0.76-1.75) | 0.511 | 9 |
| | | | Mucinous | 1.57 (1-2.47) | 0.05 | 1.22 (0.75-2) | 0.422 | **22** |
| | | | Clear cell | 1.06 (0.83-1.35) | 0.659 | 1 (0.75-1.32) | 0.98 | 6 |
| ***RBBP8*** | **h0011** | 23 | **All** | **0.81 (0.71-0.94)** | **0.005** | **0.82 (0.69-0.98)** | **0.029** | 1 |
| | | | Serous | 0.84 (0.70-1.02) | 0.079 | 0.99 (0.53-1.85) | 0.976 | **18** |
| | | | Endometrioid | 0.83 (0.53-1.3) | 0.422 | 0.94 (0.58-1.52) | 0.789 | **13** |
| | | | Mucinous | 0.64 (0.37-1.08) | 0.096 | 0.86 (0.48-1.53) | 0.614 | **34** |
| | | | Clear cell | 0.85 (0.64-1.13) | 0.261 | 0.87 (0.63-1.22) | 0.429 | 2 |
| *RBBP8* | h1010 | 7 | All | 0.99 (0.78-1.26) | 0.941 | 0.93 (0.7-1.22) | 0.588 | 6 |
| | | | Serous | 1.00 (0.73-1.36) | 0.99 | **0.81 (0.65-1)** | 0.054 | **19** |
| | | | Endometrioid | 0.85 (0.38-1.88) | 0.689 | 0.78 (0.35-1.73) | 0.548 | 8 |
| | | | Mucinous | 0.87 (0.36-2.13) | 0.762 | 1.15 (0.45-2.93) | 0.771 | **32** |
| | | | Clear cell | 1.05 (0.63-1.75) | 0.841 | 1.03 (0.57-1.87) | 0.928 | 2 |
| *RBBP8* | h1110 | 4 | All | 1.26 (0.97-1.65) | 0.086 | 1.04 (0.77-1.42) | 0.789 | **17** |
| | | | Serous | 1.06 (0.76-1.49) | 0.725 | 0.92 (0.77-1.11) | 0.403 | **13** |
| | | | Endometrioid | 1.43 (0.56-3.63) | 0.457 | 1.02 (0.36-2.89) | 0.974 | **29** |
| | | | Mucinous | 1.27 (0.45-3.63) | 0.651 | 0.55 (0.16-1.85) | 0.333 | **57** |
| | | | Clear cell | 1.64 (0.99-2.72) | 0.057 | 1.51 (0.91-2.5) | 0.111 | 8 |
| ***RBBP8*** | **h0010** | 3 | **All** | **0.60 (0.39-0.93)** | **0.022** | 0.75 (0.45-1.25) | 0.275 | **25** |
| | | | **Serous** | 0.75 (0.41-1.36) | 0.347 | **1.2 (1.01-1.42)** | **0.041** | **60** |
| | | | Endometrioid | 0.35 (0.08-1.56) | 0.169 | 0.54 (0.12-2.46) | 0.429 | **54** |
| | | | Mucinous | 0.56 (0.13-2.31) | 0.419 | 1.03 (0.15-7.23) | 0.976 | **84** |

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Clear cell | 0.7 (0.31-1.58) | 0.388 | 0.78 (0.28-2.18) | 0.637 | **11** |
| ***RBBP8*** | Rare | 1 | All | 1.24 (0.72-2.15) | 0.436 | 1.34 (0.71-2.52) | 0.366 | 8 |
| | | | Serous | 1.10 (0.57-2.12) | 0.786 | 1.08 (0.86-1.36) | 0.493 | 2 |
| | | | **Endometrioid** | 2.48 (0.32-19.53) | 0.388 | **11.36 (1.27-101.18)** | **0.029** | **358** |
| | | | Mucinous | - | - | - | - | - |
| | | | Clear cell | 1.37 (0.44-4.31) | 0.587 | 1.5 (0.47-4.83) | 0.493 | 9 |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *RBBP8*: rs7239066, rs11082221, rs4474794, rs9304261.

## Appendix VIII-M: Univariate and multivariate Cox regression results of *RGC32* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RGC32* | rs10467472 | 0.13 | 1769 | All | 0.94 (0.79-1.12) | 0.486 | 0.99 (0.8-1.22) | 0.922 | 5 |
| | | | 839 | Serous | 0.91 (0.72-1.15) | 0.447 | 0.95 (0.73-1.24) | 0.693 | 4 |
| | | | 273 | Endometrioid | 1.08 (0.62-1.88) | 0.78 | 1.4 (0.78-2.52) | 0.259 | **30** |
| | | | 191 | Mucinous | 1.12 (0.64-1.97) | 0.688 | 1.19 (0.62-2.29) | 0.591 | 6 |
| | | | 164 | Clear cell | 0.97 (0.69-1.38) | 0.88 | 0.72 (0.45-1.13) | 0.151 | **26** |
| ***RGC32*** | **rs3783194** | 0.11 | 1690 | All | 0.95 (0.79-1.14) | 0.567 | 1.11 (0.9-1.36) | 0.342 | **17** |
| | | | 788 | **Serous** | 1.11 (0.88-1.41) | 0.359 | **1.44 (1.12-1.86)** | **0.005** | **30** |
| | | | 264 | Endometrioid | 1.09 (0.65-1.84) | 0.742 | 1.12 (0.61-2.05) | 0.713 | 3 |
| | | | 184 | Mucinous | 0.63 (0.29-1.35) | 0.232 | 0.64 (0.24-1.73) | 0.38 | 2 |
| | | | 155 | Clear cell | 0.81 (0.55-1.19) | 0.282 | 0.76 (0.5-1.16) | 0.202 | 6 |
| *RGC32* | rs11618371 | 0.11 | 1771 | All | 1.01 (0.84-1.20) | 0.931 | 0.96 (0.79-1.18) | 0.72 | 5 |
| | | | 835 | Serous | 1.00 (0.81-1.25) | 0.968 | 0.94 (0.73-1.21) | 0.63 | 6 |
| | | | 275 | Endometrioid | 0.88 (0.49-1.6) | 0.677 | 0.8 (0.41-1.57) | 0.522 | 9 |
| | | | 193 | Mucinous | 1.22 (0.58-2.56) | 0.598 | 0.93 (0.42-2.05) | 0.862 | **24** |
| | | | 164 | Clear cell | 0.98 (0.67-1.44) | 0.918 | 1.1 (0.72-1.69) | 0.65 | **12** |
| ***RGC32*** | **rs9532824** | 0.07 | 1782 | All | 1.18 (0.95-1.46) | 0.137 | 1.12 (0.86-1.46) | 0.386 | 5 |
| | | | 841 | Serous | 1.19 (0.90-1.57) | 0.233 | 1.06 (0.75-1.48) | 0.75 | **11** |
| | | | 276 | Endometrioid | 0.6 (0.27-1.3) | 0.193 | 0.45 (0.18-1.14) | 0.093 | **25** |
| | | | 196 | Mucinous | 1.62 (0.81-3.23) | 0.169 | 1.75 (0.82-3.76) | 0.149 | 8 |
| | | | 150 | **Clear cell** | **1.69 (1.07-2.67)** | **0.023** | 1.45 (0.86-2.42) | 0.162 | **14** |
| ***RGC32*** | **rs995845** | 0.2 | 1274 | All | 0.96 (0.85-1.08) | 0.488 | 1.13 (0.93-1.38) | 0.218 | **18** |
| | | | 595 | Serous | 1.03 (0.88-1.21) | 0.682 | 1.25 (0.97-1.61) | 0.082 | **21** |
| | | | 193 | **Endometrioid** | 1.55 (0.93-2.6) | 0.093 | **1.8 (1.03-3.14)** | **0.039** | **16** |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 146 | Mucinous | 1.09 (0.58-2.03) | 0.797 | 0.75 (0.36-1.54) | 0.43 | **31** |
| | | | 112 | Clear cell | 0.94 (0.66-1.34) | 0.716 | 1.04 (0.69-1.56) | 0.851 | **11** |
| | | | 1766 | All | 0.97 (0.83-1.13) | 0.697 | 0.92 (0.77-1.11) | 0.377 | 5 |
| | | | 833 | Serous | 1.02 (0.84-1.24) | 0.813 | 0.92 (0.74-1.15) | 0.488 | 10 |
| *RGC32* | rs9594551 | 0.15 | 275 | Endometrioid | 0.94 (0.55-1.59) | 0.811 | 0.74 (0.4-1.34) | 0.317 | **21** |
| | | | 193 | Mucinous | 1.23 (0.67-2.29) | 0.505 | 1.24 (0.68-2.28) | 0.482 | 1 |
| | | | 163 | Clear cell | 0.78 (0.55-1.11) | 0.167 | 0.94 (0.63-1.42) | 0.776 | **21** |
| | | | 1749 | All | 1.06 (0.92-1.21) | 0.407 | 0.98 (0.84-1.16) | 0.837 | 8 |
| | | | 828 | Serous | 1.05 (0.88-1.25) | 0.571 | 0.95 (0.78-1.15) | 0.574 | 10 |
| *RGC32* | rs975590 | 0.23 | 268 | Endometrioid | 0.77 (0.49-1.22) | 0.266 | 0.6 (0.35-1.01) | 0.056 | **22** |
| | | | 188 | Mucinous | 1.68 (0.99-2.87) | 0.056 | 1.62 (0.91-2.9) | 0.102 | 4 |
| | | | 161 | Clear cell | 1.13 (0.84-1.52) | 0.421 | 1.22 (0.86-1.73) | 0.262 | 8 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-N: Univariate and multivariate Cox regression results of *RGC32* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| RGC32 | h0000000 | 41.5 | All | 0.96 (0.84-1.1) | 0.56 | 0.92 (0.79-1.07) | 0.283 | 4 |
| | | | Serous | 0.94 (0.81-1.1) | 0.46 | 0.91 (0.76-1.08) | 0.265 | 3 |
| | | | Endometrioid | 1 (0.68-1.47) | 0.986 | 0.96 (0.62-1.47) | 0.838 | 4 |
| | | | Mucinous | 0.88 (0.56-1.38) | 0.586 | 0.83 (0.51-1.37) | 0.476 | 6 |
| | | | Clear cell | 0.65 (0.37-1.14) | 0.131 | 0.82 (0.45-1.52) | 0.534 | **26** |
| RGC32 | h0000011 | 4.9 | All | 1.1 (0.82-1.48) | 0.525 | 0.97 (0.69-1.38) | 0.873 | **12** |
| | | | Serous | 1.12 (0.79-1.57) | 0.522 | 1.04 (0.71-1.54) | 0.826 | 7 |
| | | | Endometrioid | 1.24 (0.55-2.8) | 0.604 | 0.73 (0.29-1.88) | 0.518 | **41** |
| | | | Mucinous | 1.69 (0.62-4.63) | 0.304 | 2.46 (0.88-6.89) | 0.086 | **46** |
| | | | Clear cell | 0.59 (0.14-2.59) | 0.485 | 0.8 (0.17-3.66) | 0.773 | **36** |
| RGC32 | h0000100 | 10.4 | All | 1.05 (0.86-1.29) | 0.611 | 1.03 (0.82-1.3) | 0.785 | 2 |
| | | | Serous | 1.01 (0.79-1.28) | 0.959 | 1 (0.76-1.3) | 0.975 | 1 |
| | | | Endometrioid | 1.36 (0.81-2.29) | 0.244 | 1.54 (0.9-2.64) | 0.118 | **13** |
| | | | Mucinous | 0.77 (0.32-1.82) | 0.546 | 0.74 (0.3-1.79) | 0.5 | 4 |
| | | | Clear cell | 1.42 (0.6-3.36) | 0.424 | 1.4 (0.45-4.37) | 0.561 | 1 |
| RGC32 | h0001001 | 6.7 | All | 1.11 (0.86-1.43) | 0.434 | 1.06 (0.78-1.43) | 0.73 | 5 |
| | | | Serous | 1.13 (0.83-1.53) | 0.442 | 1.06 (0.74-1.51) | 0.752 | 6 |
| | | | Endometrioid | 0.65 (0.29-1.47) | 0.303 | 0.51 (0.19-1.33) | 0.168 | **22** |
| | | | Mucinous | 1.6 (0.77-3.3) | 0.208 | 1.54 (0.69-3.42) | 0.289 | 4 |
| | | | Clear cell | 1.83 (0.63-5.28) | 0.267 | 1.02 (0.23-4.54) | 0.977 | **44** |
| RGC32 | h0010011 | 7.8 | All | 1.03 (0.84-1.26) | 0.791 | 0.94 (0.74-1.18) | 0.575 | 9 |
| | | | Serous | 1.01 (0.8-1.28) | 0.928 | 0.94 (0.73-1.22) | 0.646 | 7 |
| | | | Endometrioid | 0.93 (0.48-1.81) | 0.84 | 0.9 (0.43-1.88) | 0.775 | 3 |

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Mucinous | 1.06 (0.47-2.41) | 0.89 | 0.81 (0.35-1.89) | 0.627 | **24** |
| | | | Clear cell | 1.01 (0.47-2.17) | 0.989 | 1.23 (0.5-3.05) | 0.649 | **22** |
| RGC32 | h0100100 | 10.8 | All | 1.01 (0.83-1.23) | 0.93 | 1.29 (1.03-1.62) | 0.024 | **28** |
| | | | Serous | 1.13 (0.89-1.42) | 0.324 | 1.5 (1.16-1.94) | 0.002 | **33** |
| | | | Endometrioid | 1.23 (0.69-2.2) | 0.476 | 1.34 (0.72-2.49) | 0.364 | 9 |
| | | | Mucinous | 0.65 (0.3-1.4) | 0.271 | 0.64 (0.24-1.71) | 0.37 | 2 |
| | | | Clear cell | 0.94 (0.47-1.88) | 0.866 | 0.95 (0.45-2.01) | 0.884 | 1 |
| RGC32 | h1000000 | 8.1 | All | 0.98 (0.76-1.25) | 0.843 | 1.01 (0.76-1.34) | 0.935 | 3 |
| | | | Serous | 0.88 (0.65-1.2) | 0.418 | 0.89 (0.63-1.25) | 0.487 | 1 |
| | | | Endometrioid | 0.94 (0.46-1.93) | 0.875 | 1.2 (0.58-2.48) | 0.626 | **28** |
| | | | Mucinous | 1.45 (0.68-3.12) | 0.338 | 1.57 (0.64-3.82) | 0.322 | 8 |
| | | | Clear cell | 1.65 (0.77-3.56) | 0.201 | 1.59 (0.51-4.94) | 0.421 | 4 |
| RGC32 | h1000100 | 4.1 | All | 0.97 (0.69-1.36) | 0.863 | 1.46 (0.98-2.19) | 0.063 | **51** |
| | | | Serous | 1.06 (0.71-1.59) | 0.767 | 1.48 (0.92-2.39) | 0.109 | **40** |
| | | | Endometrioid | 1.42 (0.55-3.64) | 0.469 | 1.9 (0.72-5.04) | 0.196 | **34** |
| | | | Mucinous | 0.81 (0.29-2.23) | 0.686 | 0.75 (0.17-3.23) | 0.695 | 7 |
| | | | Clear cell | 0.61 (0.14-2.6) | 0.501 | 0.57 (0.14-2.23) | 0.415 | 7 |
| **RGC32** | **Rare** | | **All** | 0.67 (0.41-1.1) | 0.114 | **0.49 (0.27-0.9)** | **0.022** | **27** |
| | | | Serous | 0.72 (0.42-1.24) | 0.234 | 0.56 (0.29-1.06) | 0.073 | **22** |
| | | | Endometrioid | 0.35 (0.05-2.56) | 0.298 | 0.09 (0-9.19) | 0.312 | **74** |
| | | | Mucinous | 0.77 (0.1-6.16) | 0.809 | 1.31 (0.17-10.38) | 0.796 | **70** |
| | | | Clear cell | 0.82 (0.13-5.14) | 0.832 | 0.64 (0.1-4.06) | 0.639 | **22** |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *RGC32*: rs10467472, rs3783194, rs11618371, rs9532824, rs995845, rs9594551, rs975590.

## Appendix VIII-O: Univariate and multivariate Cox regression results of *RUVBL1* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RUVBL1* | rs9860614 | 0.12 | 1777 | All | 1.07 (0.91-1.27) | 0.404 | 0.97 (0.79-1.19) | 0.789 | 9 |
| | | | 839 | Serous | 1.03 (0.83-1.29) | 0.772 | 1.04 (0.8-1.34) | 0.789 | 1 |
| | | | 276 | Endometrioid | 1.19 (0.71-1.98) | 0.516 | 1.19 (0.71-2.01) | 0.504 | 0 |
| | | | 193 | Mucinous | 1.01 (0.53-1.93) | 0.981 | 0.64 (0.27-1.5) | 0.306 | **37** |
| | | | 162 | Clear cell | 1.11 (0.79-1.56) | 0.555 | 0.81 (0.53-1.24) | 0.335 | **27** |
| *RUVBL1* | rs13063604 | 0.25 | 1266 | All | 1.08 (0.95-1.23) | 0.248 | 0.99 (0.85-1.16) | 0.925 | 8 |
| | | | 537 | Serous | 1.04 (0.88-1.23) | 0.626 | 1.02 (0.84-1.24) | 0.862 | 2 |
| | | | 207 | Endometrioid | 1.3 (0.88-1.9) | 0.186 | 1.2 (0.78-1.85) | 0.397 | 8 |
| | | | 143 | Mucinous | 0.75 (0.44-1.28) | 0.286 | 0.75 (0.41-1.38) | 0.358 | 0 |
| | | | 124 | Clear cell | 1.08 (0.81-1.43) | 0.609 | 0.82 (0.58-1.14) | 0.24 | **24** |
| *RUVBL1* | rs3732402 | 0.4 | 1280 | All | 1.00 (0.89-1.13) | 0.977 | 0.92 (0.8-1.07) | 0.275 | 8 |
| | | | 596 | Serous | 0.97 (0.83-1.14) | 0.729 | 0.91 (0.76-1.1) | 0.326 | 6 |
| | | | 194 | Endometrioid | 0.98 (0.68-1.43) | 0.931 | 1.07 (0.71-1.62) | 0.743 | 9 |
| | | | 147 | Mucinous | 0.8 (0.51-1.26) | 0.344 | 0.87 (0.53-1.43) | 0.58 | 9 |
| | | | 114 | Clear cell | 1.06 (0.82-1.39) | 0.645 | 0.91 (0.67-1.24) | 0.556 | **14** |
| *RUVBL1* | rs7650365 | 0.46 | 1645 | All | 1.03 (0.92-1.15) | 0.64 | 1.06 (0.92-1.22) | 0.429 | 3 |
| | | | 769 | Serous | 1.07 (0.92-1.24) | 0.388 | 1.09 (0.91-1.3) | 0.341 | 2 |
| | | | 256 | Endometrioid | 0.9 (0.61-1.34) | 0.615 | 0.81 (0.52-1.28) | 0.377 | 10 |
| | | | 175 | Mucinous | 1.2 (0.79-1.83) | 0.393 | 0.87 (0.55-1.39) | 0.567 | **28** |
| | | | 155 | Clear cell | 1.03 (0.81-1.31) | 0.824 | 1.1 (0.84-1.44) | 0.486 | 7 |
| ***RUVBL1*** | **rs4857836** | 0.2 | 1787 | **All** | 0.98 (0.86-1.12) | 0.758 | **0.81 (0.67-0.98)** | **0.03** | **17** |
| | | | 845 | **Serous** | 1.01 (0.86-1.19) | 0.879 | **0.75 (0.59-0.95)** | **0.018** | **26** |
| | | | 278 | Endometrioid | 1.12 (0.67-1.88) | 0.66 | 0.92 (0.52-1.61) | 0.762 | **18** |

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate[§] | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | 195 | Mucinous | 0.8 (0.42-1.55) | 0.513 | 0.9 (0.42-1.95) | 0.797 | **13** |
| | | | 165 | Clear cell | 0.97 (0.68-1.38) | 0.863 | 0.84 (0.55-1.28) | 0.406 | **13** |
| | | | 1733 | All | 0.89 (0.76-1.05) | 0.179 | 0.91 (0.74-1.11) | 0.336 | 2 |
| | | | 820 | Serous | 0.92 (0.75-1.14) | 0.45 | 0.88 (0.68-1.12) | 0.292 | 4 |
| ***RUVBL1*** | **rs9821568** | 0.15 | 269 | **Endometrioid** | **0.49 (0.27-0.9)** | **0.021** | 0.65 (0.35-1.2) | 0.172 | **33** |
| | | | 186 | Mucinous | 0.64 (0.31-1.35) | 0.244 | 1.1 (0.47-2.6) | 0.824 | **72** |
| | | | 161 | Clear cell | 1.13 (0.82-1.55) | 0.447 | 1.37 (0.94-1.98) | 0.098 | **21** |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-P: Univariate and multivariate Cox regression results of *RUVBL1* haplotypes

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *RUVBL1* | h000000 | 13.3 | All | 0.95 (0.76-1.2) | 0.682 | 1.04 (0.81-1.34) | 0.751 | 9 |
| | | | Serous | 0.89 (0.68-1.17) | 0.405 | 0.97 (0.72-1.31) | 0.855 | 9 |
| | | | Endometrioid | 1.89 (0.97-3.68) | 0.061 | 1.65 (0.84-3.25) | 0.149 | **13** |
| | | | Mucinous | 0.95 (0.42-2.15) | 0.91 | 1.76 (0.75-4.15) | 0.195 | **85** |
| | | | Clear cell | 0.8 (0.31-2.06) | 0.646 | 0.86 (0.33-2.23) | 0.754 | 7 |
| *RUVBL1* | h000100 | 48 | All | 1.08 (0.92-1.26) | 0.367 | 1.11 (0.93-1.32) | 0.246 | 3 |
| | | | Serous | 1.11 (0.92-1.33) | 0.279 | 1.12 (0.92-1.38) | 0.259 | 1 |
| | | | Endometrioid | 0.76 (0.46-1.26) | 0.282 | 0.76 (0.44-1.32) | 0.329 | 0 |
| | | | Mucinous | 1.58 (0.93-2.7) | 0.09 | 1.05 (0.58-1.88) | 0.88 | **34** |
| | | | Clear cell | 1.23 (0.62-2.43) | 0.561 | 1.45 (0.73-2.91) | 0.292 | **18** |
| *RUVBL1* | h001011 | 14.5 | All | 0.99 (0.58-1.69) | 0.984 | 1.29 (0.62-2.69) | 0.499 | **30** |
| | | | Serous | 1.34 (0.72-2.48) | 0.351 | 1.72 (0.73-4.07) | 0.216 | **28** |
| | | | Endometrioid | 0.21 (0.04-1.09) | 0.063 | 0.47 (0.08-2.73) | 0.398 | **124** |
| | | | Mucinous | 2.34 (0.33-16.57) | 0.394 | 2.8 (0.24-32.22) | 0.408 | **20** |
| | | | Clear cell | 0.28 (0.02-4.97) | 0.382 | 1.18 (0.06-23.01) | 0.912 | **321** |
| *RUVBL1* | h011010 | 11.7 | All | 1.15 (0.91-1.46) | 0.234 | 1.05 (0.81-1.36) | 0.715 | 9 |
| | | | Serous | 1.11 (0.85-1.45) | 0.457 | 1.05 (0.78-1.41) | 0.738 | 5 |
| | | | Endometrioid | 1.93 (0.94-3.97) | 0.073 | 1.35 (0.61-2.96) | 0.456 | **30** |
| | | | Mucinous | 0.6 (0.23-1.58) | 0.302 | 0.77 (0.28-2.1) | 0.616 | **28** |
| | | | Clear cell | 1.35 (0.48-3.78) | 0.573 | 0.77 (0.22-2.66) | 0.676 | **43** |
| *RUVBL1* | h111000 | 9.8 | All | 1.02 (0.8-1.3) | 0.855 | 0.95 (0.73-1.25) | 0.724 | 7 |
| | | | Serous | 1.03 (0.77-1.36) | 0.862 | 1.02 (0.75-1.39) | 0.885 | 1 |

| Gene | Haplotype | Freq (%) | Histology | Univariate | | Multivariate[§] | | Diff HR |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| | | | Endometrioid | 1.07 (0.49-2.33) | 0.869 | 1.19 (0.53-2.7) | 0.676 | **11** |
| | | | Mucinous | 0.97 (0.42-2.2) | 0.934 | 0.63 (0.23-1.72) | 0.368 | **35** |
| | | | Clear cell | 0.69 (0.21-2.24) | 0.539 | 0.27 (0.04-1.74) | 0.17 | 61 |
| *RUVBL1* | Rare | | All | 0.82 (0.51-1.33) | 0.429 | 0.93 (0.56-1.55) | 0.789 | 13 |
| | | | Serous | 0.88 (0.51-1.53) | 0.655 | 0.96 (0.54-1.69) | 0.877 | 9 |
| | | | Endometrioid | 1.43 (0.44-4.66) | 0.553 | 1.47 (0.44-4.89) | 0.525 | 3 |
| | | | Mucinous | 0.05 (0-62.52) | 0.405 | 0.02 (0-2257451) | 0.674 | 60 |
| | | | Clear cell | 0.54 (0.07-4.06) | 0.549 | 1.37 (0.17-11.14) | 0.769 | 154 |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *RUVBL1*: rs9860614, rs13063604, rs3732402, rs7650365, rs4857836, rs9821568.

## Appendix VIII-Q: Univariate and multivariate Cox regression results of *STAG3* tSNPs

| Gene | tSNP | MAF | No. cases | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| *STAG3* | rs11762932 | 0.22 | 1787 | All | 0.95 (0.83-1.09) | 0.482 | 0.92 (0.78-1.09) | 0.342 | 3 |
| | | | 846 | Serous | 1.00 (0.84-1.18) | 0.985 | 0.94 (0.77-1.14) | 0.533 | 6 |
| | | | 279 | Endometrioid | 0.87 (0.56-1.34) | 0.52 | 0.86 (0.52-1.42) | 0.557 | 1 |
| | | | 194 | Mucinous | 0.8 (0.45-1.39) | 0.421 | 0.87 (0.44-1.75) | 0.702 | 9 |
| | | | 164 | Clear cell | 0.9 (0.66-1.21) | 0.473 | 0.98 (0.7-1.37) | 0.893 | 9 |
| ***STAG3*** | **rs2246713** | 0.47 | 1295 | All | 1.07 (0.95-1.19) | 0.268 | 1.06 (0.93-1.21) | 0.412 | 1 |
| | | | **549** | Serous | **1.17 (1.02-1.35)** | **0.03** | 1.11 (0.95-1.31) | 0.188 | 5 |
| | | | 212 | Endometrioid | 0.94 (0.65-1.37) | 0.749 | 0.96 (0.63-1.47) | 0.866 | 2 |
| | | | 143 | Mucinous | 0.87 (0.56-1.36) | 0.544 | 0.99 (0.61-1.61) | 0.972 | **14** |
| | | | 130 | Clear cell | 1 (0.79-1.27) | 0.992 | 1 (0.76-1.32) | 0.983 | 0 |
| ***STAG3*** | **rs1637001** | 0.26 | 1784 | All | 1.11 (0.97-1.26) | 0.119 | 1.1 (0.95-1.28) | 0.198 | 1 |
| | | | **843** | **Serous** | **1.20 (1.02-1.42)** | **0.029** | 1.15 (0.95-1.38) | 0.157 | 4 |
| | | | 278 | Endometrioid | 1.15 (0.79-1.68) | 0.472 | 1.15 (0.75-1.74) | 0.524 | 0 |
| | | | 194 | Mucinous | 0.97 (0.59-1.58) | 0.904 | 1.01 (0.64-1.61) | 0.966 | 4 |
| | | | 165 | Clear cell | 1.03 (0.78-1.35) | 0.852 | 0.96 (0.7-1.31) | 0.777 | 7 |

HR: Hazard ratio; CI: confidence interval; MAF- minor allele frequency; §: adjusted for clinical factors; emboldened histology names are statistically associated with survival; emboldened HR are statistically significant.

## Appendix VIII-R: Univariate and multivariate Cox regression results of *STAG3* haplotypes

| Gene | Haplotype† | Freq (%) | Histology | Univariate | | Multivariate§ | | Diff HR (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | P-value | HR (95% CI) | P-value | |
| ***STAG3*** | **h011** | 27 | All | 1.11 (0.98-1.26) | 0.1 | 0.88 (0.66-1.18) | 0.381 | **21** |
| | | | Serous | **1.21 (1.03-1.43)** | **0.021** | 0.76 (0.23-2.51) | 0.654 | **37** |
| | | | Endometrioid | 1.09 (0.74-1.6) | 0.678 | 1.07 (0.7-1.64) | 0.758 | 2 |
| | | | Mucinous | 0.98 (0.6-1.6) | 0.932 | 1.01 (0.64-1.61) | 0.953 | 3 |
| | | | Clear cell | 1.05 (0.8-1.38) | 0.708 | 0.98 (0.72-1.35) | 0.914 | 7 |
| *STAG3* | h110 | 21 | All | 0.96 (0.84-1.11) | 0.599 | 1 (0.77-1.28) | 0.977 | 4 |
| | | | Serous | 1.02 (0.86-1.21) | 0.846 | 1.04 (0.71-1.54) | 0.834 | 2 |
| | | | Endometrioid | 0.82 (0.52-1.3) | 0.405 | 0.83 (0.5-1.39) | 0.482 | 1 |
| | | | Mucinous | 0.75 (0.42-1.36) | 0.347 | 0.81 (0.39-1.69) | 0.573 | 8 |
| | | | Clear cell | 0.93 (0.69-1.25) | 0.624 | 1 (0.71-1.41) | 0.998 | 8 |
| *STAG3* | h000 | 5 | All | 0.96 (0.86-1.07) | 0.468 | 1.38 (0.96-1.99) | 0.085 | **44** |
| | | | Serous | 0.88 (0.76-1.01) | 0.077 | 0.9 (0.76-1.08) | 0.267 | 2 |
| | | | Endometrioid | 0.97 (0.68-1.39) | 0.872 | 0.95 (0.63-1.44) | 0.821 | 2 |
| | | | Mucinous | 1.2 (0.78-1.84) | 0.405 | 1.06 (0.66-1.71) | 0.798 | **12** |
| | | | Clear cell | 1.06 (0.83-1.35) | 0.646 | 1.06 (0.8-1.41) | 0.669 | 0 |
| ***STAG3*** | **Rare** | 2 | All | 0.76 (0.45-1.27) | 0.296 | 0.9 (0.46-1.77) | 0.762 | **18** |
| | | | Serous | 0.52 (0.21-1.29) | 0.159 | 0.95 (0.68-1.34) | 0.771 | **83** |
| | | | **Endometrioid** | 2.2 (0.88-5.52) | 0.093 | **2.83 (1.01-7.95)** | **0.048** | **29** |
| | | | Mucinous | 1 (0.14-7.16) | 0.998 | 1.45 (0.19-10.86) | 0.717 | **45** |
| | | | Clear cell | 0.55 (0.2-1.49) | 0.239 | 0.55 (0.17-1.75) | 0.31 | 0 |

†: '0'= common allele and '1'= rare allele; §: adjusted for clinical factors; SNP order in haplotypes, 5' to 3', *STAG3*: rs11762932, rs2246713, rs1637001

## Appendix IX-A: Call rates and concordance of WGA samples vs gDNA genotyped with TaqMan

| SNP | MAF | gDNA (n=95) | GenomePlex (n=90) | | GenomiPhi (n=95) | | PEP (n=95) | | REPLI-g (n=95) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Call rate | Call rate | Discord | Call rate | Discord | Call rate | Discord | Call rate | Discord |
| rs602652 | 0.44 | 99% | 99% | 0% | 90% | 0% | 100% | 0% | **83%** | 1% |
| rs3217805 | 0.42 | 100% | 99% | 0% | 93% | 0% | 100% | 0% | **87%** | 0% |
| rs3217869 | 0.41 | 95% | 96% | 0% | 93% | 0% | 100% | 0% | **79%** | 1% |
| rs2079147 | 0.48 | 94% | 99% | 0% | 96% | 0% | 99% | 0% | **80%** | 0% |
| rs10487888 | 0.47 | 98% | 94% | 0% | 100% | 0% | **88%** | 0% | 99% | **2%** |
| Average | | 97% | 97% | 0% | 94% | 0% | 97% | 0% | **82%** | 1% |

Discord: discordance rate, the proportion of genotypes which were the same where both gDNA (non-amplified genomic DNA) and WGA method had calls. Bold rates failed quality control of call rate >90%; discordance <2%. Average call rate of assays that passed.

## Appendix IX-B: Call rates of assays genotyped on iPLEX (by amplification method)

| SNP | MAF | Genomic | GenomePlex | GenomiPhi | PEP | REPLI-g |
|---|---|---|---|---|---|---|
| rs10487888 | 0.47 | 94 (98%) | 90 (100%) | 95 (100%) | **84 (88%)** | 95 (100%) |
| rs10842514 | 0.44 | 66 (96%) | **75 (83%)** | 78 (82.1%) | **2 (2%)** | **78 (82%)** |
| rs11047898 | 0.01 | 96 (100%) | 90 (100%) | 95 (100%) | **71 (75%)** | 95 (100%) |
| rs11047917 | 0.06 | 94 (98%) | 90 (100%) | 95 (100%) | **80 (84%)** | 94 (99%) |
| rs11551174 | 0.06 | 90 (94%) | 90 (100%) | 94 (98.9%) | 87 (92%) | 95 (100%) |
| rs12305513 | 0.1 | 94 (98%) | 90 (100%) | 94 (98.9%) | **39 (41%)** | 94 (99%) |
| rs12822857 | 0.48 | 89 (93%) | 89 (99%) | 95 (100%) | **3 (3%)** | 93 (98%) |
| rs17161747 | 0.05 | 94 (98%) | 90 (100%) | 95 (100%) | **37 (39%)** | 94 (99%) |
| rs17191185 | 0.04 | 90 (94%) | 90 (100%) | 95 (100%) | 94 (99%) | 95 (100%) |
| rs1733832 | 0.06 | 88 (92%) | 89 (99%) | 95 (100%) | 86 (90%) | 95 (100%) |
| rs17623382 | 0.12 | **82 (85%)** | 88 (98%) | 94 (98.9%) | **10 (11%)** | 94 (99%) |
| rs17695623 | 0.13 | 94 (98%) | 90 (100%) | 95 (100%) | **79 (83%)** | 94 (99%) |
| rs1801200 | 0.23 | 95 (99%) | 90 (100%) | 94 (98.9%) | 95 (100%) | 94 (99%) |
| rs2161841 | 0.28 | 95 (99%) | 90 (100%) | 93 (97.9%) | **80 (84%)** | 95 (100%) |
| rs2699905 | 0.26 | 94 (98%) | 89 (99%) | 95 (100%) | 90 (95%) | 95 (100%) |
| rs2865084 | 0.05 | 91 (95%) | 90 (100%) | 95 (100%) | 94 (99%) | 94 (99%) |
| rs2952155 | 0.26 | 95 (99%) | 90 (100%) | 95 (100%) | 95 (100%) | 93 (98%) |
| rs2952156 | 0.3 | 94 (98%) | 90 (100%) | 95 (100%) | 94 (99%) | 95 (100%) |
| rs3771882 | 0.43 | 94 (98%) | 90 (100%) | 94 (98.9%) | **84 (88%)** | 95 (100%) |
| rs3771886 | 0.41 | 92 (96%) | 89 (99%) | 95 (100%) | **32 (34%)** | 95 (100%) |
| rs3854012 | 0.28 | 94 (98%) | 90 (100%) | 95 (100%) | 95 (100%) | 95 (100%) |
| rs453226 | 0.15 | 94 (98%) | 90 (100%) | 95 (100%) | **59 (62%)** | 95 (100%) |
| rs4623993 | 0.16 | 90 (94%) | 90 (100%) | 95 (100%) | **56 (59%)** | 94 (99%) |
| rs6978734 | 0.01 | 95 (99%) | 90 (100%) | 95 (100%) | 94 (99%) | 95 (100%) |
| **Average** | | **84 (88%)** | 83 (92%) | 87 (91%) | **64 (67%)** | 86 (91%) |

N=95 (GenomePlex =90)

# Appendix IX-C: Discordance rates on iPLEX (by WGA method)

| Assay | GenomePlex | | Genomi-Phi | | PEP | | REPLI-g | |
|---|---|---|---|---|---|---|---|---|
| | **No. called** | **Discordance** | **No. called** | **Discordance** | **No. called** | **Discordance** | **No. called** | **Discordance** |
| rs10487888 | 88 | 0 (0%) | 94 | 0 (0%) | 82 | **11 (13.4%)** | 93 | **2 (2.2%)** |
| rs10842514 | 60 | 0 (0%) | 65 | 0 (0%) | Fail | | 63 | 0 (0%) |
| rs11047917 | 88 | 0 (0%) | 94 | 0 (0%) | 78 | 1 (1.3%) | 93 | **2 (2.2%)** |
| rs11551174 | 87 | **2 (2.3%)** | 89 | 1 (1.2%) | 83 | 0 (0%) | 90 | **2 (2.2%)** |
| rs12305513 | 88 | 0 (0%) | 94 | 0 (0%) | 37 | **1 (2.7%)** | 93 | 1 (1.1%) |
| rs12822857 | 82 | 0 (0%) | 89 | 0 (0%) | Fail | | 86 | 1 (1.2%) |
| rs17161747 | 88 | 0 (0%) | 94 | 0 (0%) | 35 | 0 (0%) | 92 | 0 (0%) |
| rs17191185 | 85 | 0 (0%) | 90 | 0 (0%) | 88 | 0 (0%) | 89 | 0 (0%) |
| rs1733832 | 81 | 0 (0%) | 88 | 0 (0%) | 77 | 0 (0%) | 87 | 0 (0%) |
| rs17623382 | 74 | 1 (1.4%) | 82 | 1 (1.2%) | Fail | | 80 | **2 (2.5%)** |
| rs1801200 | 89 | 1 (1.1%) | 94 | 0 (0%) | 94 | **2 (2.1%)** | 93 | 1 (1.08%) |
| rs2161841 | 89 | 0 (0%) | 93 | 0 (0%) | 79 | 1 (1.3%) | 94 | 0 (0%) |
| rs2699905 | 87 | 1 (1.2%) | 94 | 0 (0%) | 88 | 1 (1.1%) | 93 | **2 (2.2%)** |
| rs2865084 | 85 | 0 (0%) | 91 | 0 (0%) | 89 | 0 (0%) | 89 | 0 (0%) |
| rs2952155 | 89 | 0 (0%) | 95 | 0 (0%) | 94 | 0 (0%) | 93 | 1 (1.1%) |
| rs2952156 | 88 | 0 (0%) | 94 | 0 (0%) | 92 | 0 (0%) | 92 | 0 (0%) |
| rs3771882 | 89 | 0 (0%) | 93 | 0 (0%) | 82 | 0 (0%) | 93 | 1 (1.1%) |
| rs3771886 | 85 | 0 (0%) | 92 | 0 (0%) | 31 | **6 (19.4%)** | 91 | 1 (1.1%) |
| rs3854012 | 88 | 0 (0%) | 94 | 0 (0%) | 93 | 0 (0%) | 93 | **2 (2.2%)** |
| rs453226 | 89 | 0 (0%) | 94 | 0 (0%) | 57 | 0 (0%) | 93 | 0 (0%) |
| rs4623993 | 84 | 0 (0%) | 90 | 0 (0%) | 51 | 0 (0%) | 88 | **2 (2.3%)** |
| rs6978734 | 89 | 0 (0%) | 95 | 0 (0%) | 93 | 0 (0%) | 94 | 0 (0%) |

Discordance rate, the proportion of genotypes which were different where the sample had a genotype in both gDNA and the WGA method. Bold pass rates less than 90% or failed in only one method and discordance 2% or more

# Appendix IX-D: Call rates of SNPs genotyped on SNPlex (by WGA method)

| SNP | MAF | Genomic | GenomePlex | GenomiPhi | REPLI-g |
|---|---|---|---|---|---|
| rs1419755 | 0.08 | 91 (96.8%) | 87 (92.6%) | Fail | Fail |
| rs1569244 | 0.17 | Fail | Fail | Fail | Fail |
| rs354893 | 0.21 | 92 (97.9%) | 90 (95.7%) | Fail | 86 (91.5%) |
| rs729673 | 0.34 | 90 (95.7%) | Fail | **81 (86.2%)** | 85 (90.4%) |
| rs751340 | 0.41 | 92 (97.9%) | 87 (92.6%)* | Fail | 87 (92.6%) |
| rs1323001 | 0.33 | 91 (96.8%) | 89 (94.7%) | **80 (85.1%)** | 86 (91.5%) |
| rs1115261 | 0.44 | 92 (97.9%) | Fail | Fail | 85 (90.4%) |
| rs1323881 | 0.20 | Fail | Fail | Fail | 86 (91.5%) |
| rs1507213 | 0.46 | 91 (96.8%) | Fail | Fail | 86 (91.5%) |
| rs1425151 | 0.24 | 91 (96.8%) | Fail | **80 (85.1%)** | 86 (91.5%) |
| HCV2059319 | 0.31 | Fail | Fail | Fail | Fail |
| rs2286216 | 0.22 | 92 (97.9%) | 89 (94.7%)* | Fail | Fail |
| rs220860 | 0.22 | Fail | 88 (93.6%) | Fail | 86 (91.5%) |
| rs1016146 | 0.31 | 92 (97.9%) | 89 (94.7%) | Fail | 85 (90.4%) |
| rs1548543 | 0.34 | Fail | Fail | Fail | **84 (89.4%)** |
| rs1980408 | 0.23 | Fail | Fail | Fail | 86 (91.5%) |
| rs1861606 | 0.29 | 91 (96.8%) | Fail | Fail | 85 (90.4%) |
| HCV2962785 | 0.11 | Fail | Fail | Fail | Fail |
| rs705681 | 0.48 | 92 (97.9%) | Fail | Fail | Fail |
| rs1007106 | 0.34 | Fail | 90 (95.7%) | Fail | 87 (92.6%) |
| rs984071 | 0.33 | 90 (95.7%) | Fail | Fail | 85 (90.4%) |
| rs992690 | 0.16 | 92 (97.9%) | Fail | Fail | 86 (91.5%) |
| rs1520483 | 0.38 | 89 (94.7%) | Fail | Fail | 86 (91.5%) |
| rs1569125 | 0.32 | 91 (96.8%) | Fail | Fail | Fail |
| rs288423 | 0.34 | 92 (97.9%) | Fail | **79 (84%)** | 87 (92.6%) |
| rs963014 | 0.40 | Fail | Fail | Fail | Fail |
| rs238196 | 0.10 | Fail | Fail | Fail | 87 (92.6%) |
| rs961495 | 0.18 | 92 (97.9%) | Fail | **81 (86.2%)** | 86 (91.5%) |
| rs1388276 | 0.27 | 91 (96.8%) | Fail | **79 (84%)** | 87 (92.6%) |
| rs748573 | 0.22 | Fail | Fail | Fail | 86 (91.5%) |
| rs1457947 | 0.42 | Fail | Fail | Fail | Fail |
| rs927221 | 0.12 | 92 (97.9%) | 85 (90.4%) | **81 (86.2%)** | 87 (92.6%) |
| rs1378324 | 0.12 | 92 (97.9%) | Fail | Fail | **84 (89.4%)** |
| rs1570903 | 0.44 | Fail | Fail | Fail | Fail |
| rs893613 | 0.48 | 90 (95.7%) | Fail | Fail | 85 (90.4%) |
| rs1597695 | 0.39 | Fail | Fail | Fail | 87 (92.6%) |
| rs954779 | 0.18 | Fail | Fail | Fail | 79 (84%) |
| rs879253 | 0.47 | 92 (97.9%) | Fail | Fail | 87 (92.6%) |
| rs1156404 | 0.47 | 90 (95.7%) | Fail | Fail | 86 (91.5%) |
| rs1460239 | 0.46 | 90 (95.7%) | Fail | Fail | 86 (91.5%) |
| rs1334334 | 0.24 | Fail | 89 (94.7%) | Fail | 87 (92.6%) |
| HCV8879897 | 0.46 | Fail | Fail | Fail | Fail |
| rs1713423 | 0.47 | 91 (96.8%) | 86 (91.5%)* | Fail | 86 (91.5%) |

| SNP | MAF | Genomic | GenomePlex | GenomiPhi | REPLI-g |
|---|---|---|---|---|---|
| rs1035089 | 0.42 | Fail | Fail | Fail | Fail |
| rs794108 | 0.46 | 92 (97.9%) | Fail | **81 (86.2%)** | 87 (92.6%) |
| rs1925643 | 0.30 | Fail | 89 (94.7%) | Fail | Fail |
| rs995178 | 0.41 | 92 (97.9%) | 89 (94.7%) | 80 (85.1%) | 86 (91.5%) |
| rs1129167 | 0.31 | 92 (97.9%) | Fail | Fail | 87 (92.6%) |

Based on auto-called results; * Failed after checking clusters.

# Appendix IX-E: Discordance rates of WGA-DNA on SNPlex (auto-call genotypes)

| SNP | MAF | GenomePlex | | GenomiPhi | | REPLI-g | |
|---|---|---|---|---|---|---|---|
| | | No. called | Discord. | No. called | Discord. | No. called | Discord |
| rs1419755 | 0.08 | 86 | **4 (4.7%)** | Fail | | Fail | |
| rs354893 | 0.21 | 90 | 1 (1.1%) | Fail | | 85 | **3 (3.5%)** |
| rs729673 | 0.34 | Fail | | 78 | 1 (1.3%) | 82 | **3 (3.7%)** |
| rs751340 | 0.41 | 83 | **43 (52%)** | Fail | | 86 | **4 (4.7%)** |
| rs1323001 | 0.33 | 88 | 0 (0%) | 78 | 0 (0%) | 84 | **4 (4.8%)** |
| rs1115261 | 0.44 | Fail | | Fail | | 86 | **2 (2.3%)** |
| rs1507213 | 0.46 | Fail | | Fail | | 84 | **2 (2.4%)** |
| rs1425151 | 0.24 | Fail | | 78 | 0 (0%) | 84 | **4 (4.8%)** |
| rs2286216 | 0.22 | 89 | **78 (88%)** | Fail | | Fail | |
| rs1016146 | 0.31 | Fail | | Fail | | 84 | **3 (3.6%)** |
| rs1861606 | 0.29 | Fail | | Fail | | 83 | **2 (2.4%)** |
| rs705681 | 0.48 | Fail | | Fail | | Fail | |
| rs984071 | 0.33 | Fail | | Fail | | 82 | **5 (6.1%)** |
| rs992690 | 0.16 | Fail | | Fail | | 85 | **2 (2.4%)** |
| rs1520483 | 0.38 | Fail | | Fail | | 82 | **2 (2.4%)** |
| rs1569125 | 0.32 | Fail | | Fail | | Fail | |
| rs288423 | 0.34 | Fail | | 78 | **2 (2.6%)** | 86 | **2 (2.3%)** |
| rs961495 | 0.18 | Fail | | 80 | 1 (1.3%) | 85 | 1 (1.2%) |
| rs1388276 | 0.27 | Fail | | 77 | 0 (0%) | 85 | **3 (3.5%)** |
| rs927221 | 0.12 | 85 | **27 (32%)** | 80 | 1 (1.3%) | 86 | **5 (5.8%)** |
| rs1378324 | 0.12 | Fail | | Fail | | 83 | **4 (4.8%)** |
| rs893613 | 0.48 | Fail | | Fail | | 82 | **3 (3.7%)** |
| rs879253 | 0.47 | Fail | | Fail | | 86 | 1 (1.2%) |
| rs1156404 | 0.47 | Fail | | Fail | | 83 | **3 (3.6%)** |
| rs1460239 | 0.46 | Fail | | Fail | | 83 | 1 (1.2%) |
| rs1713423 | 0.47 | 85 | **62 (73%)** | Fail | | 84 | **5 (6%)** |
| rs794108 | 0.46 | Fail | | 80 | 0 (0%) | 86 | **3 (3.5%)** |
| rs995178 | 0.41 | 89 | **2 (2.3%)** | 79 | 0 (0%) | 85 | **4 (4.7%)** |
| rs1129167 | 0.31 | Fail | | Fail | | 86 | **5 (5.8%)** |

N=94 (90 for GenomePlex): Discordance rate, the proportion of genotypes which were different where the sample had a genotype in both gDNA and the WGA method. Bold pass rates less than 90% or failed in only one method and discordance 2% or more.

# Appendix IX-F: Types of discordances found with SNPlex (by WGA method)

| SNP | MAF | GenomePlex | | GenomiPhi | | REPLI-g | |
|---|---|---|---|---|---|---|---|
| | | **Dropout** | **Miscall** | **Dropout** | **Miscall** | **Dropout** | **Miscall** |
| rs1419755 | 0.08 | 0 | 4 | Fail | | Fail | |
| rs354893 | 0.21 | 1 | 0 | Fail | | 1 | 2 |
| rs729673 | 0.34 | Fail | | 1 | 0 | 3 | 0 |
| rs751340 | 0.41 | Fail | | Fail | | 2 | 2 |
| rs1323001 | 0.33 | 0 | 0 | 0 | 0 | 1 | 3 |
| rs1115261 | 0.44 | Fail | | Fail | | 1 | 1 |
| rs1507213 | 0.46 | Fail | | Fail | | 0 | 2 |
| rs1425151 | 0.24 | Fail | | 0 | 0 | 1 | 3 |
| rs2286216 | 0.22 | 0 | 0 | Fail | | 2 | 1 |
| rs1016146 | 0.31 | Fail | | Fail | | 1 | 1 |
| rs1861606 | 0.29 | Fail | | Fail | | 2 | 3 |
| rs705681 | 0.48 | Fail | | Fail | | 1 | 1 |
| rs984071 | 0.33 | Fail | | Fail | | 1 | 1 |
| rs992690 | 0.16 | Fail | | 2 | 0 | 2 | 0 |
| rs1520483 | 0.38 | Fail | | 1 | 0 | 1 | 0 |
| rs1569125 | 0.32 | Fail | | 0 | 0 | 0 | 3 |
| rs288423 | 0.34 | 0 | 0 | 1 | 0 | 2 | 3 |
| rs961495 | 0.18 | Fail | | Fail | | 1 | 3 |
| rs1388276 | 0.27 | Fail | | Fail | | 1 | 2 |
| rs927221 | 0.12 | Fail | | Fail | | 1 | 0 |
| rs1378324 | 0.12 | Fail | | Fail | | 1 | 2 |
| rs893613 | 0.48 | Fail | | Fail | | 0 | 1 |
| rs879253 | 0.47 | Fail | | Fail | | 0 | 5 |
| rs1156404 | 0.47 | Fail | | 0 | 0 | 0 | 3 |
| rs1460239 | 0.46 | 1 | 1 | 0 | 0 | 0 | 4 |
| rs1713423 | 0.47 | Fail | | Fail | | 0 | 5 |
| **Total** | | **2** | **5** | **5** | **0** | **25** | **51** |

Discord: discordance rate, the proportion of genotypes which were the same where both gDNA (non-amplified genomic DNA) and WGA method had calls. Bold rates failed quality control of call rate >90%; discordance <2%. Average call rate of assays that passed.

## Appendix IX-G: Call rates of polymorphisms genotyped on OpenArray (by WGA method)

| SNP | MAF | Genomic | GenomePlex | Genomi-Phi | REPLI-g |
|---|---|---|---|---|---|
| rs11209026 | 0.07 | 85 (91.4%) | 85 (91.4%) | **81 (87.1%)** | **78 (83.9%)** |
| rs28665122 | UK | **81 (87.1%)** | Fail | Fail | **59 (63.4%)** |
| rs6920220 | 0.18 | 90 (96.8%) | **79 (84.9%)** | Fail | **72 (77.4%)** |
| rs10499194 | 0.18 | 91 (97.8%) | **76 (81.7%)** | **82 (88.2%)** | **75 (80.6%)** |
| rs7517847 | 0.5 | 90 (96.8%) | **80 (86%)** | **83 (89.2%)** | **76 (81.7%)** |
| rs12722489 | 0.16 | 90 (96.8%) | 88 (94.6%) | 84 (90.3%) | **79 (84.9%)** |
| rs13119723 | 0.15 | 89 (95.7%) | Fail | **82 (88.2%)** | **72 (77.4%)** |
| rs6855911 | 0.31 | 91 (97.8%) | **82 (88.2%)** | 84 (90.3%) | **81 (87.1%)** |
| rs2241880 | 0.46 | 89 (95.7%) | Fail | **81 (87.1%)** | **71 (76.3%)** |
| rs12150220 | 0.47 | 91 (97.8%) | Fail | 84 (90.3%) | **72 (77.4%)** |
| rs2066845 | 0.02 | 84 (90.3%) | Fail | 85 (91.4%) | Fail |
| rs1048990 | 0.12 | 91 (97.8%) | 84 (90.3%) | 86 (92.5%) | **80 (86%)** |
| rs1063857 | 0.34 | 91 (97.8%) | **81 (87.1%)** | **83 (89.2%)** | **80 (86%)** |
| rs216320 | 0.09 | 92 (98.9%) | 87 (93.5%) | 86 (92.5%) | **80 (86%)** |
| rs8177374 | 0.14 | 90 (96.8%) | Fail | 85 (91.4%) | **79 (84.9%)** |
| rs2476601 | 0.14 | 92 (98.9%) | 86 (92.5%) | **80 (86%)** | **80 (86%)** |
| rs6822844 | 0.2 | 92 (98.9%) | **82 (88.2%)** | 86 (92.5%) | **79 (84.9%)** |
| rs3761847 | 0.48 | 91 (97.8%) | **78 (83.9%)** | **77 (82.8%)** | **66 (71%)** |
| rs2233406 | 0.27 | 89 (95.7%) | **82 (88.2%)** | **82 (88.2%)** | **74 (79.6%)** |
| rs3138053 | 0.26 | 91 (97.8%) | **80 (86%)** | 86 (92.5%) | **73 (78.5%)** |
| rs1050152 | 0.46 | 92 (98.9%) | 86 (92.5%) | 86 (92.5%) | **79 (84.9%)** |
| rs3087243 | 0.46 | 88 (94.6%) | **82 (88.2%)** | **82 (88.2%)** | **76 (81.7%)** |
| rs6897932 | 0.24 | 90 (96.8%) | Fail | 87 (93.5%) | **81 (87.1%)** |
| rs1990760 | 0.39 | 92 (98.9%) | 85 (91.4%) | **79 (84.9%)** | **76 (81.7%)** |
| rs2076756 | 0.35 | 91 (97.8%) | **80 (86%)** | **78 (83.9%)** | **77 (82.8%)** |
| rs6502867 | 0.28 | 92 (98.9%) | **80 (86%)** | 85 (91.4%) | **78 (83.9%)** |
| rs4965373 | UK | 92 (98.9%) | Fail | **80 (86%)** | **74 (79.6%)** |
| rs8025174 | UK | 90 (96.8%) | Fail | Fail | Fail |
| rs4790797 | UK | 89 (95.7%) | **76 (81.7%)** | **72 (77.4%)** | **78 (83.9%)** |
| rs41295061 | UK | 89 (95.7%) | 85 (91.4%) | 86 (92.5%) | **79 (84.9%)** |
| rs11597367 | UK | 89 (95.7%) | Fail | Fail | Fail |
| rs10487888 | 0.47 | 91 (97.8%) | **81 (87.1%)** | 85 (91.4%) | **79 (84.9%)** |

N=93; PEP amplified DNA failed for all the assays. UK is unknown

# Appendix IX-H: Discordance rates of polymorphisms genotyped on OpenArray (by WGA method)

| Assay | GenomePlex | | GenomiPhi | | REPLI-g | |
|---|---|---|---|---|---|---|
| | **No. called** | **Discord** | **No. called** | **Discord** | **No. called** | **Discord** |
| rs11209026 | 80 | 1 (1.3%) | 76 | 1 (1.3%) | 74 | **3 (4.1%)** |
| rs28665122 | Fail | | Fail | | 51 | **9 (17.7%)** |
| rs6920220 | 76 | **3 (4%)** | Fail | | 69 | **7 (10.1%)** |
| rs10499194 | 76 | **2 (2.6%)** | 81 | 1 (1.2%) | 74 | **4 (5.4%)** |
| rs7517847 | 79 | **4 (5.1%)** | 81 | **2 (2.5%)** | 74 | **7 (9.5%)** |
| rs12722489 | 86 | **3 (3.5%)** | 82 | **4 (4.9%)** | 78 | **6 (7.7%)** |
| rs13119723 | Fail | | 79 | 1 (1.3%) | 70 | **5 (7.1%)** |
| rs6855911 | 81 | **3 (3.7%)** | 83 | **4 (4.8%)** | 80 | **8 (10%)** |
| rs2241880 | Fail | | 78 | **2 (2.6%)** | 70 | **5 (7.1%)** |
| rs12150220 | Fail | | 83 | **2 (2.4%)** | 72 | **5 (6.9%)** |
| rs2066845 | Fail | | 77 | **13 (17%)** | Fail | |
| rs1048990 | 83 | 1 (1.2%) | 85 | 1 (1.2%) | 79 | **4 (5.1%)** |
| rs1063857 | 80 | 1 (1.3%) | 82 | **2 (2.4%)** | 79 | **4 (5.1%)** |
| rs216320 | 87 | 1 (1.2%) | 86 | 1 (1.2%) | 80 | 1 (1.3%) |
| rs8177374 | Fail | | 83 | 0 (0%) | 77 | **4 (5.2%)** |
| rs2476601 | Fail | | 80 | 1 (1.3%) | 80 | 1 (1.3%) |
| rs6822844 | 82 | 0 (0%) | 86 | 1 (1.2%) | 79 | **3 (3.8%)** |
| rs3761847 | 77 | **3 (3.9%)** | 76 | **3 (4%)** | 65 | **9 (14%)** |
| rs2233406 | 80 | **2 (2.5%)** | 79 | **2 (2.5%)** | 72 | **6 (8. 3%)** |
| rs3138053 | 80 | **3 (3.8%)** | 85 | 1 (1.2%) | 73 | **3 (4.1%)** |
| rs1050152 | 86 | **4 (4.7%)** | 86 | 1 (1.2%) | 79 | **4 (5.1%)** |
| rs3087243 | 80 | **4 (5%)** | 79 | **3 (3.8%)** | 74 | **5 (6.8%)** |
| rs6897932 | Fail | | 85 | **8 (9.4%)** | 79 | **9 (11%)** |
| rs1990760 | 85 | **3 (3.5%)** | 79 | 0 (0%) | 76 | **6 (7.9%)** |
| rs2076756 | 80 | **3 (3.8%)** | 78 | 1 (1.3%) | 76 | **4 (5.3%)** |
| rs6502867 | 80 | **2 (2.5%)** | 85 | 1 (1.2%) | 78 | **2 (2.6%)** |
| rs4965373 | Fail | | 80 | **4 (5%)** | 74 | **4 (5.4%)** |
| rs4790797 | 73 | 1 (1.4%) | 69 | **2 (2.9%)** | 75 | **22 (29%)** |
| rs41295061 | 82 | **2 (2.4%)** | 83 | **2 (2.4%)** | 76 | **2 (2.6%)** |
| rs10487888 | 80 | **3 (3.8%)** | 84 | **2 (2.4%)** | 78 | **5 (6.4%)** |

rs8025174 and rs11597367 failed for all the WGA methods. Discord: discordance rate, the proportion of genotypes which were the same where both gDNA (non-amplified genomic DNA) and WGA method had calls. Bold rates failed quality control of call rate >90%; discordance <2%. Average call rate of assays that passed.

# Appendix IX-I: Types of discordances on OpenArray

| Assay | GenomePlex | | GenomiPhi | | REPLI-g | |
|---|---|---|---|---|---|---|
| | **Dropout** | **Miscall** | **Dropout** | **Miscall** | **Dropout** | **Miscall** |
| rs11209026 | 0 | 1 | 0 | 1 | 2 | 1 |
| rs28665122 | Fail | | Fail | | 8 | 1 |
| rs6920220 | 1 | 3 | Fail | | 2 | 5 |
| rs10499194 | 0 | 2 | 0 | 1 | 3 | 1 |
| rs7517847 | 0 | 4 | 0 | 2 | 4 | 3 |
| rs12722489 | 1 | 2 | 0 | 4 | 3 | 3 |
| rs13119723 | Fail | | 0 | 1 | 2 | 3 |
| rs6855911 | 2 | 1 | 0 | 4 | 5 | 5 |
| rs2241880 | Fail | | 0 | 2 | 1 | 3 |
| rs12150220 | Fail | | 0 | 2 | 3 | 1 |
| rs2066845 | Fail | | 13 | 0 | Fail | |
| rs1048990 | 0 | 1 | 0 | 1 | 3 | 1 |
| rs1063857 | 0 | 1 | 1 | 1 | 1 | 4 |
| rs216320 | 0 | 1 | 0 | 1 | 0 | 1 |
| rs8177374 | Fail | | 0 | 0 | 2 | 2 |
| rs2476601 | Fail | | 0 | 1 | 0 | 1 |
| rs6822844 | 0 | 0 | 0 | 1 | 0 | 3 |
| rs3761847 | 1 | 2 | 1 | 2 | 7 | 2 |
| rs2233406 | 0 | 2 | 0 | 2 | 4 | 1 |
| rs3138053 | 2 | 1 | 0 | 1 | 2 | 1 |
| rs1050152 | 0 | 4 | 0 | 1 | 2 | 2 |
| rs3087243 | 1 | 3 | 0 | 3 | 0 | 5 |
| rs6897932 | Fail | | 0 | 8 | 4 | 5 |
| rs1990760 | 2 | 1 | 0 | 0 | 5 | 1 |
| rs2076756 | 0 | 3 | 0 | 1 | 2 | 2 |
| rs6502867 | 1 | 1 | 0 | 1 | 1 | 1 |
| rs4965373 | Fail | | 0 | 4 | 2 | 2 |
| rs4790797 | 0 | 1 | 0 | 2 | 1 | 21 |
| rs41295061 | 0 | 2 | 0 | 2 | 0 | 2 |
| rs10487888 | 1 | 2 | 0 | 2 | 3 | 2 |
| **Total** | **12** | **38** | **15** | **51** | **72** | **85** |

Dropout- allele dropout

## Appendix IX-J: Reproducibility of genotyping on OpenArray - comparison of duplicates

| SNP | Genomic | | GenomePlex | | Genomi-Phi | | REPLI-g | |
|---|---|---|---|---|---|---|---|---|
| | **Call rate** | **Discord** | **Call rate** | **Discord** | **Call rate** | **Discord** | **Call rate** | **Discord** |
| rs11209026 | **10 (83.3%)** | **1 (10%)** | 11 (91.7%) | 0 (0%) | **10 (83.3%)** | 0 (0%) | **10 (83.3%)** | **1 (10%)** |
| rs28665122 | **10 (83.3%)** | **1 (10%)** | Fail | | Fail | | Fail | |
| rs6920220 | **9 (75%)** | 0 (0%) | 11 (91.7%) | 0 (0%) | Fail | | **10 (83.3%)** | **1 (10%)** |
| rs10499194 | 12 (100%) | 0 (0%) | Fail | | 11 (91.7%) | 0 (0%) | 11 (91.7%) | 1 (9.1%) |
| rs7517847 | 11 (91.7%) | 0 (0%) | 11 (91.7%) | 0 (0%) | 12 (100%) | **1 (8.3%)** | **9 (75%)** | **1 (11.1%)** |
| rs12722489 | 12 (100%) | 0 (0%) | 11 (91.7%) | 0 (0%) | 12 (100%) | 0 (0%) | **9 (75%)** | **1 (11.1%)** |
| rs13119723 | 11 (91.7%) | 0 (0%) | Fail | | 12 (100%) | 0 (0%) | **8 (66.7%)** | 0 (0%) |
| rs6855911 | 12 (100%) | 0 (0%) | 11 (91.7%) | 0 (0%) | 11 (91.7%) | **1 (9.1%)** | 11 (91.7%) | 0 (0%) |
| rs2241880 | **9 (75%)** | 0 (0%) | Fail | | **10 (83.3%)** | 0 (0%) | **7 (58.3%)** | **1 (14.3%)** |
| rs12150220 | 11 (91.7%) | 0 (0%) | Fail | | 12 (100%) | 0 (0%) | 11 (91.7%) | **2 (18.2%)** |
| rs1048990 | 11 (91.7%) | 0 (0%) | 11 (91.7%) | 0 (0%) | 12 (100%) | 0 (0%) | **10 (83.3%)** | 0 (0%) |
| rs1063857 | 12 (100%) | 0 (0%) | 12 (100%) | 0 (0%) | 12 (100%) | 0 (0%) | 11 (91.7%) | **2 (18.2%)** |
| rs216320 | 11 (91.7%) | 0 (0%) | 12 (100%) | 0 (0%) | 12 (100%) | 0 (0%) | **10 (83.3%)** | 0 (0%) |
| rs8177374 | 12 (100%) | 0 (0%) | Fail | | 12 (100%) | 0 (0%) | **9 (75%)** | 0 (0%) |
| rs2476601 | 12 (100%) | 0 (0%) | 11 (91.7%) | 0 (0%) | **10 (83.3%)** | 0 (0%) | **10 (83.3%)** | 0 (0%) |
| rs6822844 | 12 (100%) | 0 (0%) | **9 (75%)** | 0 (0%) | 12 (100%) | 0 (0%) | **10 (83.3%)** | 0 (0%) |
| rs3761847 | 11 (91.7%) | 0 (0%) | **10 (83.3%)** | 0 (0%) | **9 (75%)** | 0 (0%) | **8 (66.7%)** | 0 (0%) |
| rs2233406 | 12 (100%) | **1 (8.3%)** | 12 (100%) | 0 (0%) | 11 (91.7%) | 0 (0%) | **9 (75%)** | **2 (22.2%)** |
| rs3138053 | **10 (83.3%)** | 0 (0%) | Fail | | 12 (100%) | 0 (0%) | Fail | |
| rs1050152 | 12 (100%) | 0 (0%) | 12 (100%) | 0 (0%) | 12 (100%) | 0 (0%) | 11 (91.7%) | 0 (0%) |

| SNP | Genomic | | GenomePlex | | Genomi-Phi | | REPLI-g | |
|---|---|---|---|---|---|---|---|---|
| | **Call rate** | **Discord** | **Call rate** | **Discord** | **Call rate** | **Discord** | **Call rate** | **Discord** |
| rs3087243 | **9 (75%)** | 0 (0%) | 12 (100%) | 0 (0%) | **10 (83.3%)** | 0 (0%) | **8 (66.7%)** | 0 (0%) |
| rs6897932 | 11 (91.7%) | **1 (9.1%)** | Fail | | 11 (91. 7%) | 0 (0%) | **10 (83.3%)** | **1 (10%)** |
| rs1990760 | 11 (91.7%) | 0 (0%) | **10 (83.3%)** | 0 (0%) | 11 (91. 7%) | 0 (0%) | 11 (91.7%) | 0 (0%) |
| rs2076756 | 11 (91.7%) | **1 (9.1%)** | **10 (83.3%)** | 0 (0%) | **9 (75%)** | 0 (0%) | Fail | |
| rs6502867 | 11 (91.7%) | 0 (0%) | **9 (75%)** | 0 (0%) | 11 (91. 7%) | 0 (0%) | **9 (75%)** | **1 (11.1%)** |
| rs4965373 | **10 (83.3%)** | 0 (0%) | Fail | | 11 (91. 7%) | 0 (0%) | **10 (83.3%)** | **1 (10%)** |
| rs4790797 | **9 (75%)** | 0 (0%) | Fail | | **8 (66. 7%)** | 0 (0%) | **9 (75%)** | **4 (44.4%)** |
| rs41295061 | 11 (91.7%) | 0 (0%) | 11 (91.7%) | 0 (0%) | 11 (91. 7%) | 0 (0%) | 11 (91.7%) | 0 (0%) |
| rs10487888 | 11 (91.7%) | 0 (0%) | **10 (83.3%)** | 0 (0%) | 11 (91. 7%) | 0 (0%) | 11 (91.7%) | **1 (9.1%)** |
| **Average** | **90.8%** | **5 (1.6%)** | **206** | **0 (0%)** | **299 (90%)** | **2 (0.67)** | **253 (81.1%)** | **20 (8.03%)** |

Discord: discordance rate, the proportion of genotypes which were the same where both gDNA (non-amplified genomic DNA) and WGA method had calls. Bold rates failed quality control of call rate >90%; discordance <2%. Average call rate of assays that passed.

# Appendix X-K: iPLEX gold vs Illumina genome-wide association data (based on SEARCH and UKOPS cases only)

| SNP | MAF | SEARCH | | UKOPS | |
|---|---|---|---|---|---|
| | | Total no. of genotypes | Discord. (%) | Total no. of genotypes | Discord. (%) |
| rs6788750 | 0.47 | 493 | 2 (0.4%) | 469 | **17 (3.6%)** |
| rs7650365 | 0.48 | 811 | 4 (0.5%) | 452 | **12 (2.7%)** |
| rs2280201 | 0.08 | 821 | 0 (0%) | 471 | 7 (1.5%) |
| rs2394644 | 0.13 | 823 | 1 (0.1%) | 479 | 7 (1.5%) |
| rs3181175 | 0.21 | 834 | 4 (0.5%) | Fail | |
| rs3783194 | 0.14 | 835 | 4 (0.5%) | 486 | 6 (1.2%) |
| rs3923086 | 0.39 | 818 | 3 (0.4%) | 473 | **20 (4.2%)** |
| rs793477 | 0.12 | 817 | 7 (0.9%) | 468 | **14 (3%)** |
| rs12494994 | 0.2 | 829 | 1 (0.1%) | Fail | |
| rs9860614 | 0.16 | 832 | 8 (1%) | 489 | 7 (1.4%) |
| rs10999147 | 0.08 | 835 | 2 (0.2%) | Fail | |
| rs3181328 | 0.06 | 820 | 1 (0.1%) | 464 | 7 (1.5%) |
| rs2282657 | 0.35 | 832 | 2 (0.2%) | 489 | **13 (2.7%)** |
| rs7189819 | 0.33 | 815 | 1 (0.1%) | 471 | **30 (6.4%)** |
| rs4541111 | 0.44 | 814 | 0 (0%) | 460 | **20 (4.4%)** |
| rs4791171 | 0.3 | 828 | 1 (0.1%) | Fail | |

Discord- discordance; bold: discordance rate >2%