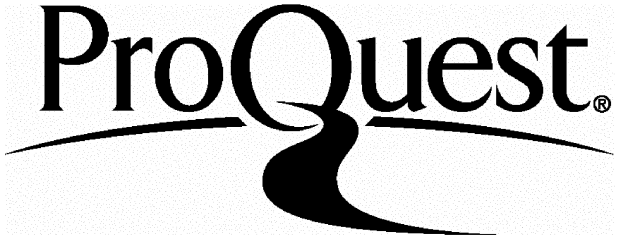
THE GENETICS OF JUVENILE POLYPOSIS SYNDROME

Kelly Woodford-Richens

Thesis submitted for the degree of
Doctor of Philosophy
in the
University of London

Imperial Cancer Research Fund

September 2001



ABSTRACT

Germline mutations in the *SMAD4* gene were described in juvenile polyposis patients just prior to the commencement of this project. Assessment of the true contribution of *SMAD4* mutations in the large JPS cohort available was undertaken using a variety of techniques, with immunohistochemistry found to be a reliable indicator of the presence of a germline *SMAD4* mutation. Morphological analysis of juvenile polyps uncovered subtle differences that may aid the segregation of juvenile polyps according to their *SMAD4* mutation status. The contribution of *SMAD4* inactivation to colorectal cancer was investigated and found to be at a higher frequency (39% in microsatellite stable tumours) and to probably occur earlier in tumorigenesis than otherwise reported, with mutations occurring after divergence of MSI+ and MSI- pathways, but before aneuploidy/polyploidy. Juvenile polyps were shown to be clonal lesions, by detection of loss of heterozygosity at the *SMAD4* locus in polyps from individuals with a germline *SMAD4* mutation. Furthermore, the 'landscaper' model for tumorigenesis was disproved in relation to JPS as the cells targeted for deletion were shown to be the epithelial cells rather than the suspected stromal cells. Several candidate genes, mainly those belonging to the same gene family as *SMAD4*, were screened for germline mutations in JPS patients without *SMAD4* mutations. No pathogenic changes were identified. A subsequent genome wide linkage and comparative genomic hybridisation search did not reveal any area of the genome with convincing evidence of a new JPS gene. The reason for this failure is almost certainly considerable remaining genetic heterogeneity in JPS. This was evidenced by the finding of germline mutations in BMPR1A/ALK3 in 32% of JPS patients. Another genome screen was performed for the related tumour-development disorder hereditary mixed polyposis syndrome (HMPS) and linkage found to 15q13-14. Identification of these genes is likely to rely on the ascertainment of large JPS families and/or candidate gene screening, combined with LOH analysis.

ACKNOWLEDGEMENTS

My grateful thanks go to my supervisor Ian Tomlinson for his help throughout my studies. He has been an invaluable source of guidance, encouragement and support. My thanks also go to my collaborators Lauri Aaltonen and Richard Houlston for their help and intellectual contribution, and also to colleagues at St. Mark's and Guy's Hospitals.

I am indebted to Nick Wright, Richard Poulsom, George Elia and colleagues from the ICRF Histopathology Laboratory for practical help and advice. There are also clinicians too numerous to mention, that I would like to thank for supplying patient details and samples.

A big thank you to all my colleagues in the Molecular and Population Genetics Laboratory, in particular Andrew Rowan for his constant help, Jill Williamson for her F.I.S.H. expertise, and also Pat Gorman, Emma Jaeger, Rebecca Roylance, Afrina Alam, Elly Sawyer and Hanan Lamlum. Thanks must also go to the ICRF Equipment Park for all their help.

Lastly, I would especially like to thank Paul and Amy for their love and support.

ABBREVIATIONS

APC- adenomatous polyposis coli

bp-base pairs

CD- Cowden disease

CGH- comparative genomic hybridisation

cDNA – copy DNA

dH_20 – distilled water

DNA - deoxyribonucleic acid

FAP- familial adenomatous polyposis

GI- gastrointestinal

GS - Gorlin Syndrome

HMPS – hereditary mixed polyposis syndrome

JPS –juvenile polyposis syndrome

Kb – kilo base pairs

LOH – loss of heterozygosity

Mb – mega base pairs

nt- nucleotides

PCR – polymerase chain reaction

PJS - Peutz-Jeghers Syndrome

PTT – protein truncation test

RNA- ribonucleic acid

All genes have been italicised throughout, whereas protein and phenotypic symbols are shown in plain text.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

CHAPTER ONE

INTRODUCTION

CANCER AS A GENETIC DISEASE

Cancer is the most common disease arising as a result of acquired genetic defects. It has considerable consequences for morbidity and mortality in humans - one in three people develop and one in five people die from cancer. Although other diseases such as heart disease may have an even higher impact on mortality, cancer is fascinating from a molecular biology point of view as it is a disruption of the normal cell machinery which usually makes multi-cellular organisms function so perfectly. Research into what actually causes cancer has led to a fundamental understanding of how normal cells control their division and growth - cancer represents anarchy against these normal interactions and constraints.

A normal somatic cell in the body is usually committed to terminally differentiate and ultimately to die, this 'altruism' being possible due to the presence of germ cells which will propagate an identical set of genes into the next generation. Usually, there is no selection pressure between the somatic cells in an individual, cells are co-ordinated to ensure that each tissue is of the right size, shape and structure for the body's requirements; selection is at the level of the organism, not at the cell or gene. A shift from this close collaboration would mean destruction of the organism and loss of the potential to pass on genes. Cancer occurs when somatic cells selfishly propagate their own genes, instead of relying on the germ cells, and this will be at ultimate expense to the organism. How can this selfish propagation and escape from normal constraints be

successful? In much the same way that the organism as a whole undergoes evolution through selection to maximise fitness by adapting to the environment, the development of cancer can be regarded as evolution on a cellular scale. Cancer cells undergo selection for multiple, successive advantageous mutations in genes which in a normal cell are responsible for such things as programmed cell death (apoptosis) or cell-cycle regulation etc. Whilst each of these mutations alone may only provide a small selective advantage, the cumulative mutations will provide a great growth and survival advantage over the normal surrounding tissue.

Generally, cancer cannot be blamed on a single selected event. It is most likely that there is an accumulation of several independent changes in the genes, each giving a selective advantage and allowing further growth. Of course, cells also acquire mutations which will not provide any advantage, and these cells will ultimately behave like normal cells and usually die off. In addition, any mutation has to bypass the cells in-built quality control machinery that detects and repairs such mistakes. Mutations in the genes, whether ultimately causing a selective advantage or not, may be due to a variety of well established carcinogenic agents in somatic cells (either chemical carcinogens such as nicotine, physical agents such as ionising radiation or virus particles which disrupt genes by incorporating their own genome into them) or due to an inherited defect in the germ cells which is passed from generation to generation. The genes targeted for disruption in cancer, occurring both sporadically and in an inherited fashion, generally code for those whose products stimulate cell proliferation and those whose products inhibit cell proliferation. Accordingly these two main types of genes which lead

to a selective advantage for the cell can be broadly divided into two categories – the proto-oncogenes and the tumour suppressor genes.

THE FUNCTION OF PROTO-ONCOGENES

Proto-oncogenes in the normal cell usually have a stimulatory role in cell growth and proliferation. A proto-oncogene which has undergone mutation, via a number of possible routes, has a hyperactive stimulatory effect on cell proliferation and is then known as an oncogene, derived from the Greek work 'oncos' meaning tumour. Oncogenes are therefore classified as gain-of-function genes. Only one of the two cellular copies of a proto-oncogene need be mutated into an oncogene to give the cell an advantage over its normal counterparts.

More than 60 proto-oncogenes or oncogenes have been described (Bishop, 1991), many of which were first reported through work with RNA tumour viruses such as *Rous sarcoma virus*, (src), isolated from chickens. These identified proto-oncogenes can be broadly divided into four groups including (i) growth factors (e.g. insulin-like growth factor 2 (*IGF2*) in Wilm's tumour), (ii) growth factor receptors (e.g. *erb-B2* in breast cancer), (iii) signal transducers (e.g. transforming growth factor-beta (*TGFβ)* in colon cancer) and transcription factors and nuclear proteins (e.g. c-*jun* and *c-myc*). The results of oncogene activation may therefore be an overproduction of growth factors, flooding of the cell with replication signals, uncontrolled stimulation through the signal transduction pathways, and/or unrestrained cell growth driven by increased levels of transcription factors.

Growth factors such as platelet derived growth factors (PDGF) exert their oncogenic properties by over-expression of the gene product. By binding to the cell's PDGF receptors the extra protein constantly stimulates the cell to proliferate. Likewise, *eRB1-B2* is the receptor for epidermal growth factor, and increased levels of *eRB1-B2* product via amplification of the gene leads the cell to act as though the ligand were bound, thereby stimulating the cell in an inappropriate manner. Signal transducers such as the *ras*-proteins may act to promote oncogenesis by acquiring activating mutations that make them constitutively active. Finally, nuclear proteins such as c-*myc* are oncogenic because they normally help to maintain a cell in G_0 phase, that is, resting. Levels of *c-myc* are only detectable when the cell is cycling and thus over-expression may lead to proliferation. A summary of some of the known oncogenes and the cancers that may be promoted by them is shown in Table 1.1. The proto-oncogenes together probably play a role in a tightly controlled network of interactions which ensure that no single oncogenic mutation is enough to cause cancer, a belt-and-braces protection against disruption. In other words, more than one mutation is usually required to bypass the safety net of the cells' constraints on growth.

There are several mechanisms that can activate a proto-oncogene into an oncogene. These include amplification of the copy number of a gene that leads to increased levels of product (e.g. *erbB-2* in breast cancer). Amplified genes can often be detected by light microscopy as 'double minutes' (extra-chromosomal material) or homogenously stained regions (HSRs) which appear as long

segments of identical chromosome material. Both double minutes and HSRs represent thousands of copies of a cellular oncogene. Translocation of chromosomes can activate oncogenes by moving a proto-oncogene to a new chromosomal location. This may result in either over-expression of the oncogene under the control of a nearby enhancer (e.g. *c-myc* in Burkitt's lymphoma) or expression of a novel fusion protein (e.g. *bcr-abl* in chronic myeloid leukaemia). Finally, point mutations or insertion/deletions of one or more bases may activate an oncogene by altering the function of the gene product (e.g. *k-ras* in colon cancer).

Although retroviral activation of oncogenes helped identify many of the known proto-oncogenes and is observed in other animals such as chickens and rodents, they are not highly correlated with cancer progression in man. In other animals, DNA copies of the RNA viruses can be incorporated into the host genome and may cause insertional mutagenesis, that is, disruption of a proto-oncogene. Alternatively, viruses can activate oncogenes by incorporating them into the viral genome which results in either the gene being transcribed under the control of a viral enhancer, leading to over-expression or the oncogene fusing with a viral gene making an abnormal or truncated product. DNA tumour viruses have been associated with tumour progression in man, though not via the activation of oncogenes but through the inactivation of tumour suppressor genes.

Oncogene	*Mechanism of Activation*	*Function*	*Associated cancer*
bcr-abl	*Translocation*	*Fusion protein non-receptor*	*Acute myeloid and lymphoblastic leukaemias*
β-cat	*Point mutation*	*Transcriptional co-activator*	*Colorectal, melanoma*
bcl-2	*Translocation*	*Anti-apoptosis protein*	*B-cell lymphoma*
cycd1	*Amplification Translocation*	*Cyclin D, cell cycle control*	*Breast, B-cell lymphoma Lymphomas, parathyroid adenomas*
cdk4	*Amplification/point mutation*	*Cyclin-dependent kinase*	*Sarcoma, familial melanoma*
erb-b	*Amplification*	*Epidermal growth factor receptor*	*Squamous and other ca., astrocytoma*
erb-b2	*Amplification*	*Epidermal growth factor receptor*	*Breast, ovarian, gastric, other ca.*
gli	*Amplification*	*Transcription factor*	*Sarcoma, glioma*
hst	*Amplification*	*FGF-like growth factor*	*Gastric*
mdm2	*Amplification*	*P53-binding protein*	*Sarcoma*
met	*Point mutation*	*HGF-receptor tyrosine kinase*	*Hereditary renal ca., papillary type*
C-myc	*Translocation Amplification*	*Transcription factor*	*Burkitt's lymphoma, SCLC, other ca.*
L-myc	*Amplification*	*Transcription factor*	*Breast, cervix, lung*
N-myc	*Amplification*	*Transcription factor*	*SCLC*
H-ras	*Point mutation*	*p21 GTPase*	*Neuroblastoma, SCLC*
K-ras	*Point mutation*	*p21 GTPase*	*Bladder*
N-ras	*Point mutation*	*p21 GTPase*	*Colorectal, pancreatic, endometrial, lung*
ret	*Translocation*	*GNDF-receptor tyrosine kinase*	*Acute myeloid and lymphoblastic leukaemias. Thyroid ca, hereditary Multiple Endocrine Neoplasia type 2 (MEN2)*
smo	*Point mutations*	*Transmembrane signalling molecule in sonic hedgehog pathway*	*Basal cell - skin*

Table 1.1 A subset of oncogenes and their mutations in human cancers.

ca. = carcinoma, FGF = fibroblast growth factor, GTPase = guanine trinucleotide repeat phosphatase, SCLC = small cell carcinoma of the lung, GDNF = glial derive neutropic factor, HGF=hepatocyte growth factor.

THE FUNCTION OF TUMOUR SUPPRESSOR GENES

Opposing the dominant stimulatory nature of the proto-oncogenes, tumour suppressor genes' role in normal cells is inhibitory, keeping tight control of proliferation. Loss-of–function of tumour suppressor genes leads to tumour promotion. Both copies of a tumour suppressor gene in a single cell must be inactivated in order to lose the constraints encoded for by the gene products. The recessive nature and realisation of the presence of tumour suppressor genes were shown by experiments in which malignant cells fused with non-malignant cells showed a non-malignant phenotype (Harris *et al.*, 1969). These experiments implied that chromosomes derived from the non-malignant cell had restored the normal cell division control. The malignant phenotype only became re-apparent when chromosomes bearing these 'restoring' properties were lost, these chromosomes being derived from the original 'normal' cell. The genes responsible for retaining the normal phenotype in the hybrid cells were thus tumour suppressor genes. Kinzler and Vogelstein coined the term 'gatekeepers' to describe tumour suppressor genes, given their role in controlling cellular proliferation (Kinzler and Vogelstein, 1997).

The most widely used paradigm for tumour suppressor genes is that of retinoblastoma, caused by the *RB1* gene, in which tumours develop from neural precursor cells in the immature retina. The disease is rare, affecting about 1 in 100000 individuals. Forty per cent of cases of retinoblastoma are hereditary, being transmitted in an autosomal dominant nature, where disease is usually

bilateral with an age of onset normally before five years old. The 60% of cases that are sporadic, that is without an inherited component, are usually unilateral with a much later age of onset. This disparity can be readily explained by the two-hit model of tumorigenesis (Knudson, 1971). In the inherited cases, one copy of the *RB1* gene is already inactivated in every cell of the body (the first 'hit'), and a mutation of the other *RB1* allele (second 'hit') is all that is required to leave a cell without any functional copies of the *RB1* product. In the sporadic cases however, a single cell has to undergo two independent mutations of the *RB1* gene to render it inactive, thus explaining why the age of onset is much higher because mutations, or 'hits', take time to acquire (Figure 1.1).

The retinoblastoma gene was first localised through the identification of retinoblastoma patients having a constitutional deletion of chromosome band 13q14, with the classification of *RB1* as a tumour suppressor confirmed when somatic loss of heterozygosity (representing the second 'hit' at *RB1*) was demonstrated in the tumours. The gene product has since been shown to be a cell-cycle regulatory protein, thus loss of the *RB1* leads to loss of the inhibitory effects of this protein.

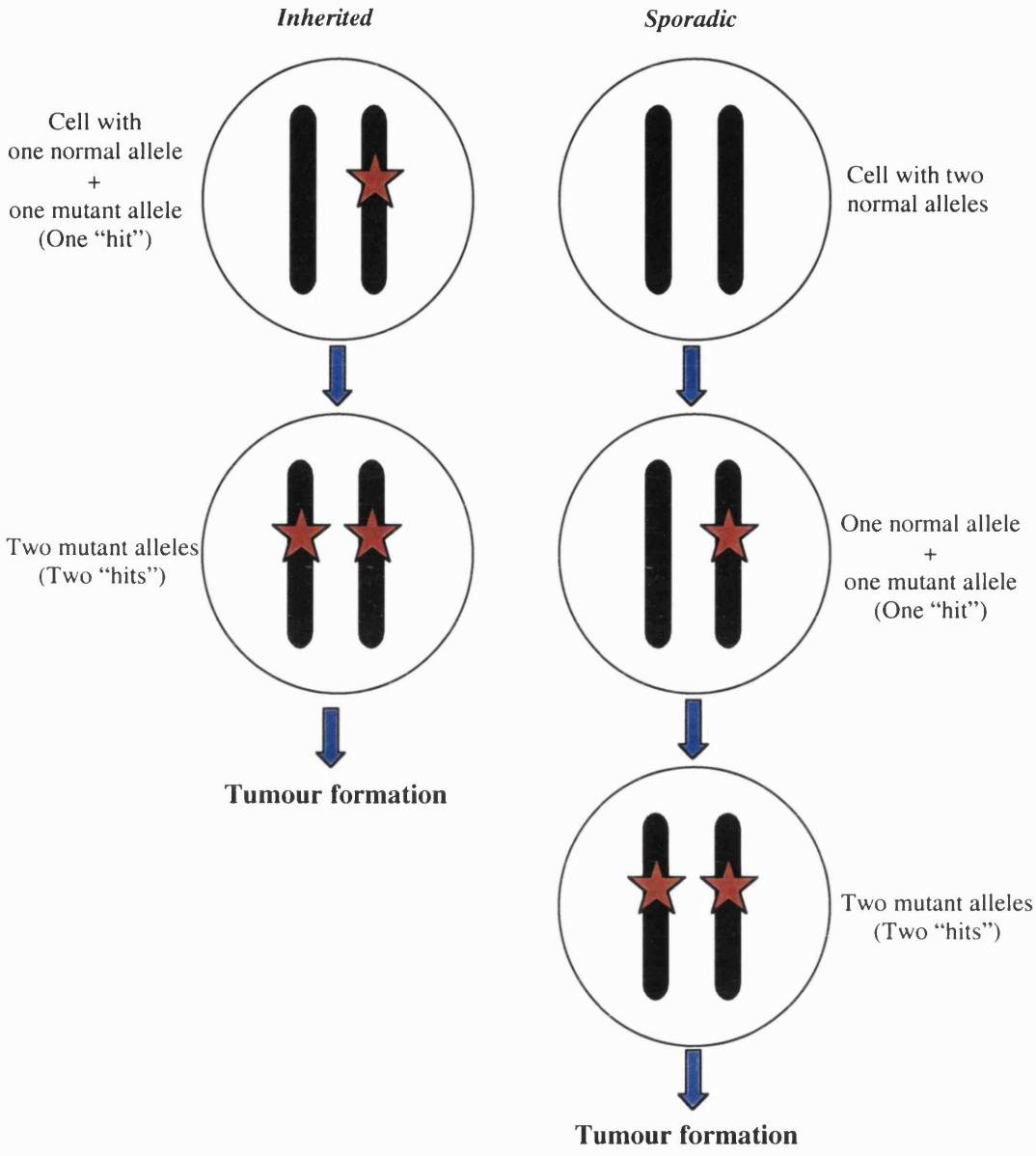


Figure 1.1. Genetic mechanisms underlying retinoblastoma.
In the inherited form, every cell in the body contains one inactivated Rb gene, either through deletion, point mutation or other mechanism such as insertion/deletion. It is highly likely that a cell in the body will acquire a somatic mutation, thus leading to tumour formation at an early age. In the sporadic, that is non-hereditary, form, a single cell must coincidentally inactivate both copies of the tumour suppressor gene to lead to tumour formation. This is far less likely and it takes longer for the cell to accumulate the two mutations, accounting for the higher age of onset and unilateral disease

Several mechanisms to explain how the loss of heterozygosity or other second hit observed in the tumours might arise were proposed (Cavenee *et al.*, 1983) (Figure 1.2)

These include chromosome loss with or without re-duplication, deletion, mitotic recombination, gene conversion and point mutation, the endpoint of which is always a cell with no functional tumour suppressor gene present. There is now evidence from the *APC* tumour suppressor gene (which causes familial adenomatous polyposis) that the two 'hits' may not be totally independent, that is, the nature of the first hit determines the type of second hit (discussed in detail later) (Lamlum *et al.*, 1999).

The mechanisms of biallelic inactivation (i.e. two hits) can be extrapolated for any of a number of the now identified tumour suppressor genes that cause both sporadic and inherited cancers. A summary of a subset of tumour suppressor genes, their function and their role in both inherited and sporadic cancers is given in Table 1.2

GENOME STABILITY BY MISMATCH REPAIR GENES

Mismatch repair genes have a more indirect role in cancer growth than oncogenes or classical tumour suppressor genes, but given that it is inactivation of these genes that begins a cell on the slippery slope toward cancer, they can be

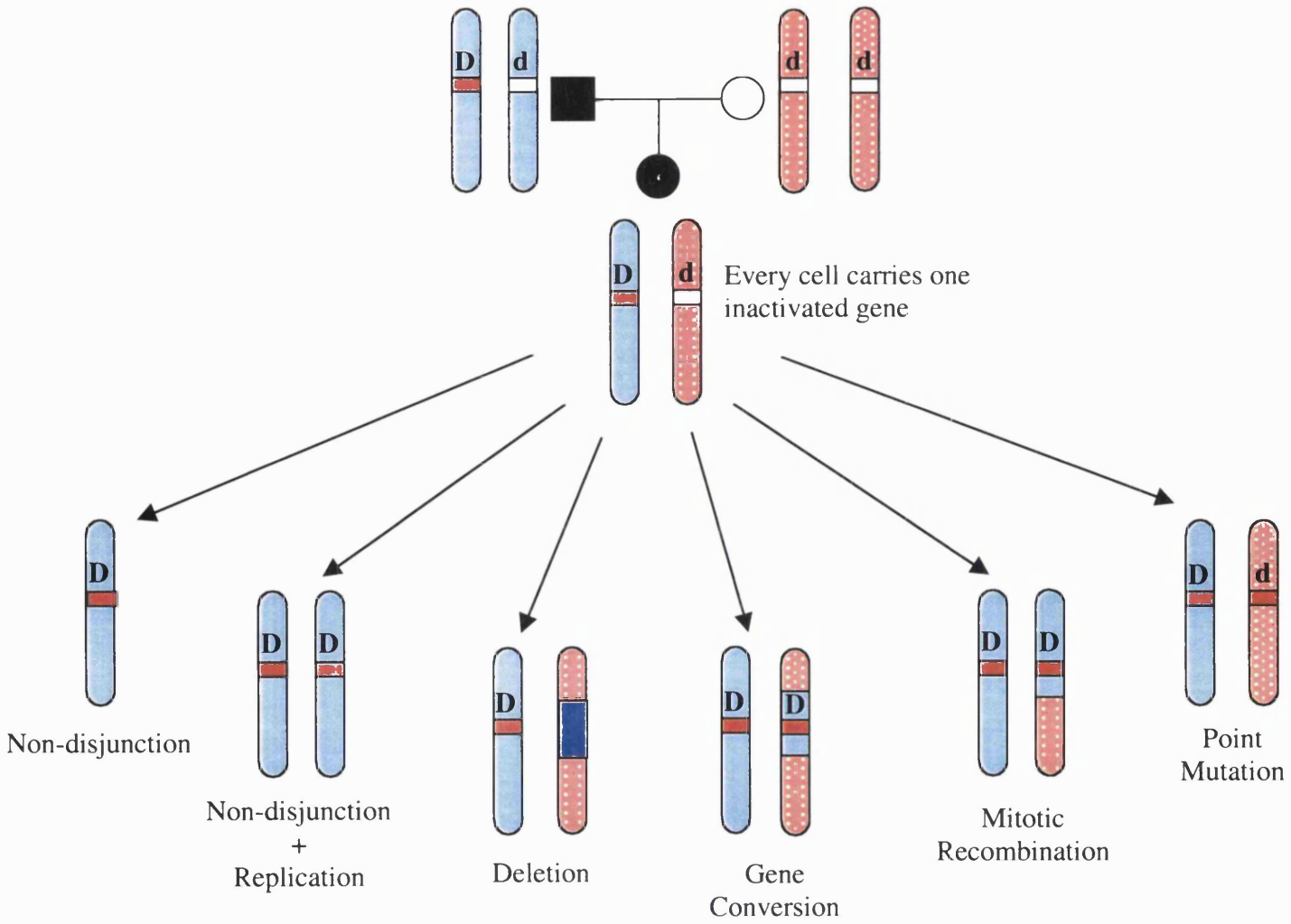


Figure 1.2 Mechanisms for loss of tumour suppressor genes.
A mutation in one copy of a tumour suppressor gene e.g. Rb is inherited from a parent. Loss of the second copy, that copy derived from the unaffected parent, can arise via a number of mechanisms, thereby leading progression of cancer (Based on (Cavenee et al., 1983)).

classified as tumour suppressor genes. The gene products of mismatch repair genes code for, as the name suggests, enzymes that detect and repair mistakes in the DNA. In other words they maintain the integrity of the cell, (Kinzler and Vogelstein coined the term 'caretaker' to describe the mismatch repair genes (Kinzler and Vogelstein, 1997)) particularly when faced with DNA damaging agents such as ionising radiation.

The inactivation of mismatch repair genes, or MMRs, appears to set the stage for the development of mutations in other genes that directly control cellular proliferation and growth control, giving rise to what is known as 'the mutator phenotype'. The most well studied of the MMR genes are those that cause Hereditary Non-Polyposis Colorectal Cancer (HNPCC, discussed in detail later), namely *MLH1*, *MSH2*, *PMS1* and *PMS2*. As with other tumour suppressor genes, the MMR genes require two hits to inactivate them, either one inherited and one somatic (as in HNPCC) or two somatic.

Gene /Locus	*Location*	*Function*	*Familial cancers / syndromes*	*Cancers with somatic mutations*
APC	*5q21-22*	*Regulates β-catenin, ?microtubule binding*	*Familial adenomatous polyposis (FAP), Gardner's syndrome, Turcot's syndrome, desmoid disease,*	*Colorectal*
BRCA1	*17q21*	*Transcriptional regulation, ?DNA repair*	*Breast and ovarian*	*Ovarian*
BRCA2	*13q12.3*	*DNA repair, binds to Rad51*	*Breast (male and female), pancreatic*	*Not known*
CDH1	*16q22.1*	*Cadherin, cell-cell adhesion*	*Gastric*	*Lobular breast*
CDKN2A	*9p21*	*Cyclin-dependent kinase inhibitor 2A (p16)*	*Melanoma, pancreatic*	*Many types inc. breast, lung, pancreatic etc*
CYLD	*16q12*	*?Cylindromatosis gene*	*Cylindromatosis*	*Not known*
DCC	*18q21.1*	*?Transmembrane netrin receptor*	*Not known*	*Colorectal*
E-CAD	*16q*	*Transmembrane cell-cell adhesion molecule*	*Not known*	*Diffuse type gastric and lobular breast*
LKB1	*19p13.3*	*Serine/threonine kinase*	*Peutz-Jeghers Syndrome – ovarian, pancreatic, jejunal hamartomas*	*Ovarian, pancreatic, testicular*
MEN1	*11q13*	*Unknown*	*Multiple endocrine neoplasia type 1 (MEN1)*	*Islet-cell, parathyroid / pituitary adenoma*
MLH1	*3p21.3*	*DNA-mismatch repair*	*Hereditary non-polyposis colorectal cancer (HNPCC), type 2 – colorectal, endometrial, gastric*	*Colorectal, gastric endometrial*
MSH2	*2p16*	*DNA-mismatch repair*	*Hereditary non-polyposis colorectal cancer (HNPCC), type 1 – colorectal, endometrial, gastric*	*Colorectal, gastric endometrial,*
NF1	*17q11*	*GTPase activator (RAS signalling)*	*Neurofibromatosis type 1 – neurofibromas*	*Melanoma, neuroblastoma*
NF2	*22q12*	*Juxta-membrane link to cytoskeleton*	*Neurofibromatosis type 1 – Schwann cell tumours, meningiomas*	*Schwannomas, meningiomas*

Table 1.2 A subset of tumour suppressor genes and their role in inherited and sporadic cancers.

Table continued on next page.

Gene /Locus	*Location*	*Function*	*Familial cancers / syndromes*	*Cancers with somatic mutations*
PRKAR1A	*17q23*	*?*	*Carney complex – myxoma and endocrine tumours*	*Unknown*
PTCH	*9q22.3*	*Transmembrane receptor for sonic hedgehog, negative regulator of smo protein*	*Gorlin's syndrome – basal cell carcimoma*	*Basal cell carcimoma, medulloblastoma*
PTEN	*10q23.3*	*Phosphatase and tensin homolog*	*Cowden Syndrome – thyroid and breast*	*prostate, gliomas, endometrial, breast*
RB1	*13q14*	*Cell cycle inhibitor*	*Retinoblastoma – retinas, osteosarcomas*	*Retinoblastomas, bone, bladder, small cell lung, breast*
SMAD4	*18q21.1*	*Downstream regulator of TGFβ signalling*	*Juvenile Polyposis Syndrome – colorectal, gastric*	*Pancreatic, colorectal*
TGFβIIR	*3p?*	*Transmembrane receptor for TGFβ*	*Not known*	*Colorectal, gastric*
TP53	*17p13.1*	*Regulates cell division and induces apoptosis*	*Li-Fraumeni Syndrome – multiple sites including breast, brain etc*	*Inactivated in >50% of all tumours e.g. prostate, colorectal*
TSC1	*9q34*	*Codes for hamartin, function unknown???*	*Tuberous Sclerosis Type 1- hamartomas, renal cell carcinoma*	*Unknown?*
TSC2	*16p13.3*	*Codes for tuberin, function unknown???*	*Tuberous Sclerosis Type 2- hamartomas, renal cell carcinoma*	*Unknown?*
VHL	*3p26*	*Down-regulates expression of hypoxia induced growth factors*	*von Hippie-Lindau syndrome – renal cell, haemangiomas, pheochromocytomas*	*renal cell, haemangiomas, pheochromocytomas*
WT1	*11p13*	*Zinc-finger transcription factor*	*WAGR and Denys Drash Syndromes – kidney (nephroblastomas)*	*Wilm's tumour – kidney (nephroblastomas)*

Table 1.2 continued.

Cells without functional MMRs then acquire mutations mainly in non-coding repeats (e.g. CA_n), but often in genes with repeat tracts that are liable to slippage. This slippage is known as MicroSatellite Instability (MSI) or Replication Error (RER) and can be detected via the presence of extra microsatellite bands in the tumours when compared to normal tissue from that patient. One of the best known is the Big Adenine Tract (BAT) in exon 3 of the *TGFβII-R* gene that consists of a run of 10 adenines. Mutations, usually deletions of one or two bases, in this adenine tract lead to a premature stop codon which has been shown to abrogate TGFβ signalling in colorectal cancer (Markowitz *et al.*, 1995; Parsons *et al.*, 1995), although it has been reported that a minor subset of cancers may still respond to TGFβ (Ilyas *et al.*, 1999). Thus, mutations in MMR genes lead to mutations in classical tumour suppressor or proto-oncogenes.

EPIGENETIC AND VIRAL INACTIVATION OF TUMOUR SUPPRESSOR GENES

In addition to genetic mechanisms which lead to inactivation of tumour suppressor genes, more complicated epigenetic mechanisms of gene silencing have been demonstrated. For example, 84% of colorectal tumours displaying the 'mutator phenotype' have promoter methylation of *MLH1*. Likewise, 9% of *RB1* inactivation in retinoblastoma and 33% of *VHL* inactivation in von Hippel Lindau disease has been shown to be as a result of promoter methylation (Esteller, 2000). GC rich areas, better known as CpG islands, tend to occur near gene promoters and are about 1-2kb in length and usually unmethylated. Every

cell in a woman normally has one of the two X-chromosomes randomly inactivated via methylation so that there is only the one copy necessary of each X-linked gene. When CpG islands of tumour suppressor genes become methylated the gene becomes silenced and this leads to loss of the gene product, and thus tumour promotion.

Another deviation from the classical two hit model is when genes are inactivated via imprinting of parent specific alleles. Genomic imprinting represents epigenetic silencing of genes depending on the parental origin, and may play a role in sporadic Wilm's tumour, embryonal rhabdomyosarcoma and sporadic osteosarcoma (Scrable *et al.*, 1989). In these cases, the first hit is imprinting, or silencing, of the paternal alleles, and the second hit involves loss (via the classical second hit mechanisms) of the maternal alleles. Although imprinting and other epigenetic mechanisms are an alternative to the usual first hit of mutation (point mutation, frameshift, deletion etc.) they can still be covered by Knudson's 'two hit' umbrella.

As mentioned earlier, DNA viruses can promote cancer in humans via the inactivation of tumour suppressor genes. DNA viruses such as human papilloma virus (HPV), hepatitis B (HepB) and Epstein Barr Virus (EBV) have had their role in human cancer well established (cervical cancer, hepatocarcinoma and Burkitt's lymphoma respectively). HPV coded proteins E6 and E7 are consistently expressed in cervical carcinomas, and been shown bind and inactivate *RB1* and *TP53* respectively (Tommasino and Crawford, 1995). Work

with other animal viruses has shown that there is a convergence of the viral disruption at these two genes. For example, the large T protein of SV40 in monkeys and the E1A protein of adenovirus in rodents (as well as HPV in man) have been shown to associate with *RB1* and result in its inactivation (Chellappan *et al.*, 1992). Similarly, the large T protein of SV40 inhibits binding of p53 to DNA target sites, and adenovirus E1B interferes with p53 transactivation both leading to p53 inactivation. The targets of HepB in hepatocarcinoma and EBV in Burkitt's lymphoma are less clear, but presumably they too target disruption and inactivation of tumour suppressor genes.

HEREDITARY PREDISPOSITION TO CANCER

Hereditary predisposition to cancer represents a minority of the total number of cases, probably about 10%. Studying the hereditary cases, however, is invaluable because by resolving the genes which causes them there is the corollary that sporadic cancer aetiology may also be revealed. Another advantage of understanding the genetics of hereditary cancers is that it may negate the rather unpleasant screening, anxiety and prophylactic measures taken by a subset of at-risk family members who indeed have not inherited the predisposition, and clearer management of those family members who have inherited the faulty gene. Most commonly for the disorders inherited in a Mendelian fashion, predisposition to cancer is a result of a mutation in a tumour suppressor gene (Table 1.2), although less frequently mutations of proto-oncogenes can also be transmitted in the germline e.g. the *met* and *ret* genes (Table 1.1). In 1991 the

adenomatous polyposis coli (*APC*) gene was shown to cause Familial Adenomatous Polyposis (FAP) (Groden *et al.*, 1991; Kinzler *et al.*, 1991a), and subsequently found to be mutated in the majority of sporadic colorectal cancers (Kinzler and Vogelstein, 1996). Prior to the localisation and unearthing of *APC* as the causative gene in FAP, children of affected individuals normally began annual sigmoidoscopy and colonoscopies at the age of 11 until the age of at least 40 when their risk fell to below 1 per cent. Now it is possible to pinpoint the mutation in *APC* in an affected proband, subsequent generations of at-risk individuals need only undergo testing for the mutation rather than the invasive screening techniques, which is probably preferable. Obviously this is an over-simplification of a difficult situation, counselling and support in the light of genetic test results are implicit (for both positive and negative results), but perhaps the uncertainty may be less bearable than the actual knowledge. Whereas in the case of FAP and familial retinoblastoma there is a well defined population for genetic screening, each with considerable advantages that the knowledge may provide, the population appropriate for *BRCA1* and *BRCA2* mutations is less well defined given the high incidence of sporadic breast cancer. Of those individuals who are tested and test positive for mutations in these genes, there is a considerably higher risk of developing breast, and in the case of *BRCA1*, ovarian cancer. However, the mutations are not wholly penetrant, and early detection of malignancy with mammography is not always conclusive. Even those mutation carriers who choose to undertake drastic prophylactic surgery do not completely eradicate their risk of disease. The genetic tests in these cases are not the complete answer, and further investigation into ways of preventing and diagnosing malignancy are needed.

What is remarkable about the genes that are mutated in both inherited and sporadic cancers is that they show different tissue specificities. For example, mutation of the *RB1* gene when transmitted through the germline leads to a small spectrum of disease - retinoblastomas and osteosarcomas. However, the *RB1* gene may be found to be somatically inactivated in a much broader spectrum of sporadic cancers e.g. bone, bladder, breast and small cell lung carcinomas, as well as retinoblastomas (Hodgson and Maher, 1999). What leads to this disparity i.e. why is it that germline carriers of *RB1* mutations do not develop breast, bladder or bone disease at a higher incidence than the normal population? It is possible that in the retina, the *RB1* gene product is the primary protein regulating growth control, whereas in other tissues, such as lung or breast, the epithelial cells rely less heavily on *RB1* to control proliferation. In these tissues, prior inactivation of other tumour suppressor genes (or activation of oncogenes) occurs before *RB1* is able to promote malignancy. In support of this, Kinzler and Vogelstein proposed that each cell type had it own unique 'gatekeeper (s)'; in the case of retinas the gatekeeper would be *RB1* (Kinzler and Vogelstein, 1997), thus leading a cancer spectrum specific to each tissue. An alternative explanation for the disparity may be that somatic *RB1* mutations in lung or breast epithelial cells are incompatible with cell viability, and therefore these cells undergo apoptosis rather than tumour progression.

HEREDITARY PREDISPOSITION TO GASTROINTESTINAL TUMOURS

Disorders pre-disposing to gastrointestinal malignancy can broadly divided two ways; those arising as a result of polyposis and those occurring in the absence of polyposis. These categories can be further sub-divided. Polyposis syndromes include familial adenomatous polyposis (FAP) and Turcot's syndrome are characterised by malignancies arising in an adenomatous precursor lesion whereas juvenile polyposis syndrome (JPS) (to be the main focus of the project) and Peutz-Jeghers syndrome (PJS) have hamartomatous lesions as their pre-malignant polyp. Adenomas are true neoplasms with the proliferating epithelial tubules packed closely together, and an epithelium with crowded nuclei, hyperchromatism and an increased number of mitotic figures. Hamartomas are considered benign lesions rather than neoplasms, consisting of differentiated but disorganised cells, which are the same types as the tissue of origin. The hereditary mixed polyposis syndrome (HMPS) is defined by both adenomatous and hamartomatous polyps. Hamartomatous polyps are also a feature of Cowden disease (CD), Bannayan-Zonana Syndrome (BZS), Gorlin's syndrome (GS) and McCune-Albright syndrome (each discussed in detail later) but in these pre-cancerous syndromes there are other organ-specific malignancies than gastrointestinal. Intestinal polyps have also been described in Cronkhite-Canada syndrome and in tuberous sclerosis, but their clinical significance is less certain.

GI carcinoma arises in hereditary non-polyposis colorectal cancer (HNPCC), and although the precursor lesion may be an adenoma, there is not the florid polyposis observed in FAP. With inflammatory bowel disease (IBD) and ulcerative colitis (UC) the cancer usually develops from flat dysplastic lesions or dysplastic associated lesion or mass.

NON-POLYPOSIS DISORDERS PREDISPOSING TO GASTROINTESTINAL MALIGNANCY

Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

HNPCC (HNPCC; MIM 114500) was originally described as two distinct autosomal dominant disorders (Lynch Syndromes I and II), but are now considered a single entity. HNPCC probably accounts for 5% of colorectal cancer cases, and as such is the most common hereditary condition predisposing to colorectal malignancy. In addition to the colorectal cancers seen in HNPCC families, there is an increased frequency of tumours in other organs, including the uterus, and more rarely the ovaries, stomach, small bowel, ureter and renal pelvis. As shown in Table 1.2, the genes responsible for HNPCC are the mismatch repair genes (*MLH1*, *MSH2*, *PMS1*, *PMS2*, and *MSH6*), tumour suppressor genes that when mutated do not directly cause cancer but lead to the rapid accumulation of un-repaired mutations throughout the genome. These mutations are known to include tumour suppressors such as the transforming growth factor beta II receptor, a gene normally involved in inhibition of the cell in response to other growth factors (discussed earlier). *MSH2* is responsible for

identifying the mistakes, and other MMRs such as *MLH1* are recruited to form complexes at the site of the mismatch. Segments of DNA upstream and downstream of the mistake are then excised and correct nucleotides inserted into place, with the help of the MMR genes. About half of HNPCC cases are due to mutations in *MSH2*, located on chromosome 2 (Fishel *et al.*, 1993), 30-40% due to mutations in *MLH1* (chromosome 3) (Bronner *et al.*, 1994), and the remainder accounted for by mutations in the *PMS1* (chromosome 2), and *PMS2* (chromosome 7) genes (Nicolaides *et al.*, 1994). Mutations of the mismatch repair genes have also been found in sporadic tumours and these tumours also display the characteristic microsatellite instability phenotype observed in HNPCC tumours (discussed earlier). Germline mutations have been described in another MMR gene, *MSH6*, although these families did not fulfil the Amsterdam criteria (Trojan *et al.*, 2000; Wagner *et al.*, 2001). The Amsterdam criteria for the diagnosing of HPNCC was drawn up to allow clearer distinction from other diseases. These criteria are; three or more family members with colorectal cancer, with one a first-degree relative of the other two, colorectal cancer which extends over two or more generations, one or more affected before the age of 45 years and finally, the exclusion of FAP.

In contrast with FAP, there is not an abundance of the precursor lesion, the adenoma, in HNPCC although cancer is highly likely to occur in an individual who carries a germline mutation in one of the mismatch repair genes before the age of 45 (compared to an average of 65-70 years in the general population). Only a single or few adenomas (usually right-sided) or metaplastic polyps are necessary to ensure this, each with a very high malignant potential - once an

adenoma has developed (often as a result of mutation in *APC*), progression through the adenoma-carcinoma sequence is rapid and further mutations in other genes potentiating this acquire readily (Jass, 1995).

Management of HNPCC in mutation carriers includes either prophylactic removal of the colon, or lifetime surveillance using colonoscopies and upper endoscopy if there is evidence of gastric or small bowel cancer in the family. In addition, annual endometrial biopsy and/or ultrasound, transvaginal ultrasound with Doppler examination and serum CA125 for the detection of ovarian tumours are sometimes performed.

Inflammatory bowel disease

Inflammatory bowel disease (IBD) has been associated with an increased risk of gastrointestinal cancer and can be broadly divided into Crohn's disease (IBD1; MIM 266600) and ulcerative colitis (UC) (UC; MIM 191390). UC and Crohn's disease are characterised by different tissue damage, the former by crypt abscess formation that is limited to the mucosa and the latter by transmural granulomatous inflammation that leads to fibrostenotic lesions and fistula formation (Lawrance *et al.*, 2001). Rather than originating in pre-malignant polyps, cancers in UC arise from a region of flat dysplasia or a dysplastic associated lesion or mass (DALM). The cancers associated with UC usually arise earlier than sporadic colorectal cancers and are typified by an anaplastic and mucinous phenotype and an even distribution throughout the colon. Although

mutations have been described in the genes normally associated with colorectal cancer progression (*APC*, *TP53* and *K-RAS*), these are found at a lower level than in sporadic colorectal cancer (Chaubert *et al.*, 1994; Itzkowitz, 1997) This suggests that IBD-associated cancers develop along an alternative, albeit overlapping, pathway. Some studies have suggested that *TP53* mutations are an early rather than late event in UC-associated cancer, allowing an acceleration to carcinoma due to the loss of the DNA damaged apoptotic pathway (Brentnall *et al.*, 1994; Yin *et al.*, 1993).

Although there is clearly an inherited component in both Crohn's disease and UC, the genetics is complicated and probably confounded by environmental factors which interact to mediate the immune and non-immune responses controlling inflammation. Many chromosomal regions have been described which contain putative susceptibility loci, some of which have been confirmed (*IBD1* (16p12-q13), *IBD2* (12p13.2-q24.1) *IBD3* (the major histocompatibility complex on chromosome 6) and *IBD4* (14q11-12)) and others that await confirmation (e.g. 1p36, 3q, 4q, 5q, 7q, 14q and 19p) (Lawrance *et al.*, 2001).

POLYPOSIS DISORDERS PRE-DISPOSING TO GI MALIGNANCY

Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) (APC; MIM 175100)is an autosomal dominant disorder which affects about 1 individual in 13500, accounting for about 1% of all colorectal cancer. The disease is characterised by 100-1000s of adenomatous polyps throughout the colorectum, resulting in a carpet of polyps under which it may be difficult to see any normal mucosa. Although the risk of any one adenoma progressing to carcinoma may be small (probably no higher than the risk associated with an adenoma occurring in the normal population) the sheer numbers of adenomas mean that one will almost certainly develop into carcinoma. Polyps usually first appear in the teens, with penetrance almost complete by the age of 40 years (compared to an age of onset of 65-70 years for sporadic colorectal cancer). Polyps can also occur in the small bowel and in the stomach, which carry an increased chance of malignancy at these sites.

Extra-colonic characteristics of FAP include epidermoid cysts, osteomas, and exostoses of the skull, digits and long bones, dentigenous cysts, impacted and supernumerary teeth, and congenital hypertrophy of the retinal pigment epithelium (CHRPE). CHRPEs represent the most common extra-colonic manifestation of FAP, appearing in up to three-quarters of mutation carriers, and as they are rare in the general population, may be indicative of FAP in an at-risk individual. A further complication of FAP is the presence of desmoid tumours,

which are comprised of vascular, fibrous tissue and occur in the small bowel mesentery, peritoneum or abdominal wall. Desmoids are more common in females and often develop after pregnancy or surgery, and are one of the three most common causes of death in FAP because they are able to infiltrate surrounding tissue. Other malignancies associated with FAP include papillary carcinoma of the thyroid, astrocytomas, medulloblastomas and hepatoblastomas (Hodgson and Maher, 1999).

Gardner's syndrome and Turcot's syndrome are two pre-malignant disorders which were thought to be distinct from FAP, but have since been shown to be, at least in a proportion of cases, allelic to FAP, and even to occur within the same family as FAP individuals. Gardner's syndrome is characterised by sebaceous cysts and osteomas, whereas Turcot's syndrome manifests as multiple polyposis of the colon (though less severe than in classical FAP) and malignant tumours of the central nervous system (medulloblastomas and glioblastomas).

Mutations of the *APC* gene on 5q21 were shown in 1991 to cause FAP (Groden *et al.*, 1991; Kinzler *et al.*, 1991a) and have made predictive testing of at-risk individuals feasible in the majority of cases. In accordance with Knudson's two-hit hypothesis, adenomas and carcinomas from FAP patients show inactivation of the second copy of *APC*. The *APC* gene is extremely large (2843 amino acids), comprising 15 exons (exon 15 being the largest known exon described in any known gene) and is responsible for the negative regulation of β-catenin protein levels. β-catenin is an abundant cell protein which complexes with the

cytoplasmic domain of the E-Cadherin cell-cell adhesion protein and α-catenin and is also involved in the Wnt-1 signalling pathway. When there is loss of the *APC* gene product (via two genetic hits), β-catenin is able to accumulate and there is increased binding of β-catenin to the transcription factor Tcf-4 (T-cell factor 4). This results in increased transcriptional activation of Tcf-4 target genes which include the proto-oncogenes *c-myc* and *cyclin-D1*, with tumorigenic consequences.

The *APC* gene can be broadly divided into segments, with each segment having a particular function in the mature protein. At the 5' end, the gene contains a number of coiled-coil heptad repeats which allow oligomerisation and the armadillo repeat region. The middle of the gene contains the β-catenin binding domains comprised of an imperfect 15 amino acids and a 20 amino acids repeat region, which are involved in β-catenin regulation as described above, these repeats being interspersed with S-A-M-P amino acid repeats which mediate axin binding. Toward the C-terminus are domains which are involved in nuclear export of the protein and microtubule binding and finally at the 3' end there is the EB1 binding domain (a microtubule associated protein).

Many correlations have been made between the location of the germline mutation in APC and the phenotype in FAP. Those mutations falling in the mutation cluster region (e.g. the common 1309 mutation) are associated with a severe disease; mutations before exon 9 are associated with a lack of CHRPE (Caspari *et al.*, 1995) and mutations at the 5' end of the gene are associated with

a milder 'attenuated' phenotype, which is characterised by multiple adenomas (<100), late onset of carcinoma and often with the absence of extracolonic features (Soravia *et al.*, 1998).

The majority of sporadic colorectal cancers and adenomas also have mutations in the *APC* gene (Ichii *et al.*, 1993; Kinzler and Vogelstein, 1996) underlining its importance as a tumour suppressor gene. As mentioned earlier however, the two inactivating hits at *APC* are not entirely independent. In the case of APC, the greatest selective advantage to the cell is a mutation within the 'mutation cluster region' (MCR; codons 1284 to 1580 (Nagase and Nakamura, 1993)). Thus, if the first germline hit is in the MCR then the second hit is usually allele loss (e.g. chromosome loss), but if the first germline hit is outside of the MCR, then the second hit is usually within the MCR (e.g. point mutation). This has been found to be true in both FAP (Lamlum *et al.*, 1999) and in sporadic colorectal cancer (Rowan *et al.*, 2000).

Peutz-Jeghers Syndrome

Peutz-Jeghers syndrome (PJS) (PJS, MIM 175200), is one of the hamartomatous polyposis syndromes and is characterised by the association of gastrointestinal polyposis and mucocutaneous pigmentation (Jeghers *et al.*, 1949), occurring at an approximate frequency of 1/25000-1/50000 individuals. PJS polyps are most prevalent in the small intestine (jejunum, ileum and duodenum) but do occur

throughout the GI tract, carrying an increased risk of malignancy at these sites. The PJS hamartomas are usually large glandular polyps with a tree-like structure and a smooth muscle core. As well as the increased risk of gastrointestinal malignancy, PJS patients also have an increased risk of cancer at other sites including the pancreas, breast, ovary, testes, and the cervix. The hyperpigmentation in PJS usually begins before the age of five as dark blue or brown macules around the lips, the perianal area, the hands and feet and the buccal mucosa. These melanin spots may fade as the individual enters puberty and adulthood. Females are at risk of sex cord tumours with annular tubules (SCTAT), a benign neoplasm of the ovaries, and males occasionally develop Sertoli cell tumours of the testes.

About half of PJS patients have inherited the disorder (in an autosomal dominant manner) and about half are sporadic. The susceptibility locus of PJS was mapped in 1997 to 19p13.3 (Hemminki *et al.*, 1997) and was shown the following year to encode a serine/threonine kinase (Hemminki *et al.*, 1998), the first enzyme of this type to cause cancer predisposition. *STK11* (for serine-threonine kinase, 11), or *LKB1* as the gene is otherwise known, has been shown to be mutated in the large proportion (50%) of the inherited cases and a substantial proportion of sporadic cases (30-60%). Many familial cases who have no detectable germline mutation of *LKB1* are positive for linkage to 19p13.3, perhaps indicating that either sequencing/mutation detection sensitivity is not 100%, that there is epigenetic inactivation of *LKB1*, or that there exists another causative gene in close proximity to *LKB1*.

The function of the *LKB1* gene, as mentioned previously, is as a serine/threonine kinase, responsible for phosphorylation of further proteins at serine or (less commonly) threonine amino acid residues. Phosphorylation within the cell is used to regulate the activity of enzymes, turning them on or off, so it is probable that inactivation of *LKB1* leads to altered activity of its normal protein targets (which are unknown). Inactivation of *LKB1* has also been described outside the setting of PJS. Rare mutations and inactivation by methylation have been reported in sporadic melanoma (Rowan *et al.*, 1999) as well as pancreatic, ovarian and testicular cancers (Avizienyte *et al.*, 1999; Rowan *et al.*, 1999; Su *et al.*, 1999).

Juvenile Polyposis Syndrome (JPS)

Juvenile polyposis coli (JPS, MIM 174900), to be the main focus of this project, is a rare disorder that either occurs sporadically or is inherited in an autosomal dominant fashion where is it known as Juvenile Polyposis Syndrome (JPS). JPS is characterised by hamartomatous polyps throughout the gastrointestinal tract, including the stomach and small bowel and although for many years the juvenile polyps (and indeed the hamartomas in PJS) were thought to be without malignant potential, it is now widely accepted that they do carry an increased risk of malignancy (estimated as 12-fold) (Jarvinen and Franssila, 1984; Jass *et al.*, 1988). Single juvenile polyps are not uncommon in the general population but are not associated with an increased risk of cancer. In JPS, however, hundreds of

polyps may be present throughout the GI tract, and in much the same way as a single FAP adenoma has a high chance of progressing further, so too it is likely that a single juvenile polyp in JPS will become cancerous.

The polyps in JPS tend to have a smooth surface rather than the lobulated appearance of an adenoma, and due to the absence of a smooth muscle core are easily removed from their stalks. The classical juvenile polyp consists of large-mucin filled cysts, in which the epithelium has become flattened, separated by a heavily inflamed stroma. The epithelium has been described as normal, with no evidence of neoplasia or hyperplasia (Veale *et al.*, 1966). Together, these characteristics have led to the juvenile polyp being described as a 'stromal' lesion. The polyps can lead to bleeding and anaemia, which are often the presenting symptoms. Other characteristics which have associated with JPS include mental retardation, macrocephaly with hypertelorism, heart defects and polydactyly.

The genetics of juvenile polyposis has proved slightly more contentious and difficult than for the other hamartoma syndromes because of the overlap of clinical spectrum between JPS and in particular Cowden disease (CD). After the description of *PTEN* on 10q23 as the causative gene in CD (discussed shortly), and evidence that there was a high frequency of somatic allele loss on 10q22-23 in juvenile polyps (Jacoby *et al.*, 1997) the suspicion arose that CD and JPS may be allelic. However, the general consensus has become that *PTEN* mutations are not responsible for JPS (Eng and Ji, 1998; Marsh *et al.*, 1997b) despite a few

PTEN mutations reported in supposed JP patients (Olschwang *et al.*, 1998b). It is likely that these were probably CD cases in whom the classical CD manifestations were not present due to the young ages.

An extremely large JPS kindred (consisting of 117 family members including 29 affected members) facilitated the localisation of the first JPS gene, *SMAD4* on chromosome 18q21.1 (Howe *et al.*, 1998a; Howe *et al.*, 1998b). *SMAD4* (the homolog of Drosophila mothers against decapentaplegic), also known as *DPC4* (deleted in pancreatic cancer, locus 4), is one of a family of *SMAD* members who are responsible for transducing TGFβ (transforming growth factor-beta) signals from the cell membrane to the nucleus, with SMAD4 playing a pivotal role. The screening of JPS cohorts for mutations in *SMAD4* has identified that it is the causative gene in up to 30% of cases (Friedl *et al.*, 1999; Houlston *et al.*, 1998; Kim *et al.*, 2000; Roth *et al.*, 1999), showing there existed genetic heterogeneity even amongst pure JPS cases.

Recently, the second JP locus has been identified. Using four JPS kindreds whose disease was not attributable to *SMAD4* (or *PTEN*) mutations (consisting 57 individuals of whom 27 were affected) (Howe *et al.*, 2001) linkage was demonstrated to 10q22-23, and the causative gene subsequently identified as *BMPR1A* (bone morphogenetic protein receptor type 1A). This gene encodes a serine-threonine kinase receptor, which, like SMAD4, belongs to the TGFβ superfamily of signalling molecules, and also has the SMAD genes transducing signals to the nucleus. The contribution of *BMPR1A* to the total number of JPS

cases remains to be resolved but there is strong evidence that there remains at least one causative JPS gene to be found.

Hereditary mixed polyposis syndrome

As the name suggests Hereditary Mixed Polyposis Syndrome (HMPS) (HMPS; MIM 601200) is characterised by mixed polyps. These are either juvenile, mixed juvenile/hyperplastic or adenomatous polyps, and carry a markedly increased risk of malignancy of the colorectum. Like JPS and FAP, HMPS is inherited in a Mendelian autosomal dominant nature. One extremely large family of Ashkenazi origin (consisting of 45 members of which 22 were affected) was used to assess linkage analysis in a genome screen. Linkage was reported for chromosome 6q (Thomas *et al.*, 1996), although this linkage assumed that an affected person who did not carry the affected haplotype to be a phenocopy. Subsequent attempts to fine map and clone the causative gene on 6q have not been fruitful.

Cowden's Disease/BZS

Cowden disease (CD) (CS, MIM 158350), also known as the multiple hamartoma syndrome, is an autosomal dominant disorder characterised by hamartomas at multiple sites (GI tract, breast, thyroid, skin) and increased risk of cancer of the breast and thyroid. One of the diagnostic features of CD is the facial and oral mucosal papules. In addition, there is a wide spectrum of benign and malignant lesions including brain tumours, trichilemmomas, and

macrocephaly which are also features of CD. Bannayan-Zonana Syndrome (BZS), also known as Ruvalcaba-Myhre-Smith Syndrome (BRR, MIM 153480), has a clinical spectrum which overlaps that of CD - including macrocephaly, GI hamartomas, lipomas and café-au-lait spots on the penis. As BZS gastrointestinal hamartomas are strikingly like the juvenile polyps observed in JPS, it may be assumed that they too carry an increased risk of malignancy (though this association is not proven due to the rarity of BZS) (Hodgson and Maher, 1999).

The susceptibility locus for CD was mapped in 1996 to chromosome 10q22-23 (Nelen *et al.*, 1996), a region which contained the candidate tumour suppressor gene *PTEN* (phosphatase and tensin homolog deleted on chromosome 10), without evidence of heterogeneity. Mutations of *PTEN* had previously been found in sporadic brain, breast and prostate cancer (Li *et al.*, 1997), and in 1997 was shown to be the causative gene in CD (Liaw *et al.*, 1997). The *PTEN*, also known as *MMAC1*, gene product is a dual specificity phosphatase. On the evidence of the clinical overlap between CD and BZS (Fargnoli *et al.*, 1996) the *PTEN* gene was screened in BZS patients and found to harbour germline mutations (Arch *et al.*, 1997) (Marsh *et al.*, 1997a), proving that the two syndromes were indeed allelic. Due to the rarity of the two syndromes however, it is difficult to speculate whether there are solid genotype-phenotype correlations (i.e. whether the position of the *PTEN* mutation determines the clinical spectrum as is the case for *APC* mutations in FAP).

Gorlin syndrome (GS) and McCune-Albright syndrome Syndrome

Gorlin syndrome (BCNS, MIM 109400), also known as nevoid basal cell carcinoma syndrome, is an autosomal dominant disorder that predisposes primarily to basal cell carcinomas (BCCs) of the skin, ovarian fibromas, and medulloblastomas. In addition, there are many developmental defects associated with GS including hypertelorism, odontogenic keratocysts and various skeletal abnormalities (Hodgson and Maher, 1999). Hamartomatous polyposis of the gastrointestinal tract has also been shown to occur, though is a less common feature than some of the other characteristics. Germline mutations in the *PTCH* gene, the human homologue of Drosophila *patched*, have been shown to cause GS (Hahn *et al.*, 1996a) and somatic mutations have been found in sporadic BCCs (Unden *et al.*, 1996). The *PTCH* gene was subsequently shown to be a tumour suppressor gene (Unden *et al.*, 1996).

Hamartomas of the GI tract are one of the features of McCune-Albright syndrome (MAS; MIM 174800), also known as polyostotic fibrous dysplasia. The disease is characterised by fibrous tissue proliferation which leads to the destruction of bone in turn giving rise to fractures (Hodgson and Maher, 1999). Mutations of the stimulatory G protein GS alpha have been shown to cause McCune-Albright syndrome (Weinstein and Shenker, 1993) possibly occurring somatically early in embryogenesis giving rise to a mosaic population of cells.

PATHWAYS OF TUMORIGENESIS

THE ADENOMA – CARCINOMA SEQUENCE

As discussed earlier, cancer is the accumulation of successive mutations in different genes, each providing the cells with selective growth advantages over the surrounding normal tissue. In each of the Mendelian disorders which predispose to GI malignancy, there is usually a precursor lesion. In FAP, HNPCC, and also Gardner's and Turcot Syndromes, the precursor lesion is the adenoma. Adenomas in FAP and the majority of sporadic colorectal adenomas, as well as in a subset of HNPCC adenomas, carry inactivating mutations of the *APC* gene. This is considered to be the first necessary step in what is now the paradigm for carcinogenesis, with a step-wise progression from adenoma to carcinoma (Fearon and Vogelstein, 1990) (Figure 1.3).

With loss of APC function being sufficient to initiate the growth of an adenoma, mutations in further genes are necessary for the adenoma to grow larger and become malignant. Without further genetic changes, the adenoma, which is benign (but dysplastic and therefore premalignant) will not continue to grow and may even regress. The next step considered to occur in the adenoma on its way to carcinoma is global hypomethylation, which may play a role in altering gene expression or causing mitotic non-disjunction with its associated chromosomal irregularities (Counts and Goodman, 1995; Gama-Sosa *et al.*, 1983). Mutations in critical oncogenes (such as *K-ras*) have been found in 50% of colorectal cancers, yet only 10% of adenomas less than 1cm in diameter (Bos *et al.*, 1987;

Forrester *et al.*, 1987) indicating ~~loss~~ gain of K-ras is critical for tumour advancement. During neoplastic mitosis, the faithful copying of the genome becomes less reliable and the cancer usually becomes chromosomally unstable, with loss of whole or part of chromosomes. There is strong selection for loss of particular genetic regions such as 17p (containing *TP53*) and 18q (containing *SMAD4* and *DCC*) i.e. it is loss of these tumour suppressor genes that is associated with increasing adenoma size and the transition to an invasive carcinoma (Figure 1.3).

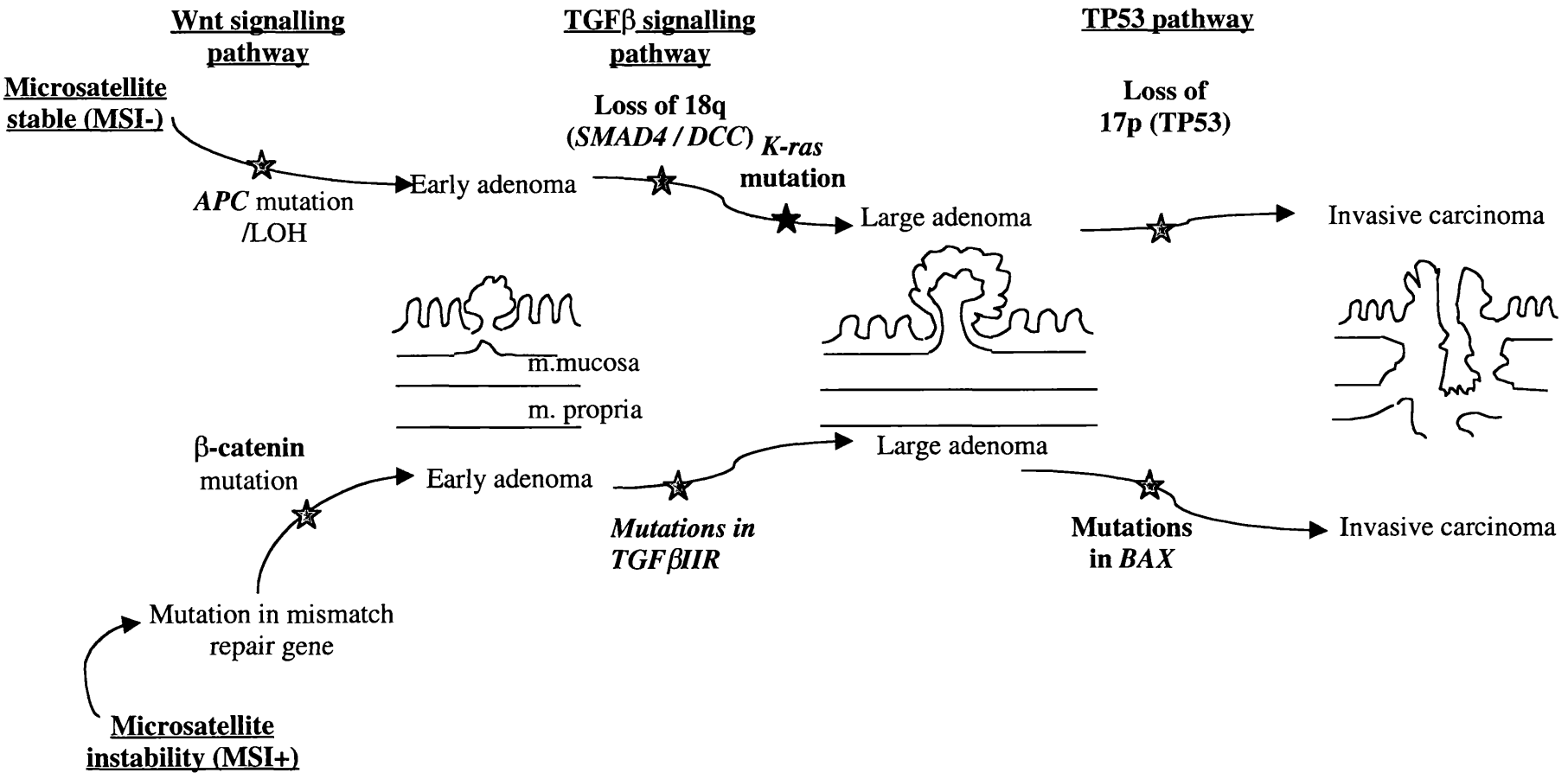


Figure 1.3 Pathways of colorectal tumorigenesis.
In the majority of tumours (FAP and sporadic) the tumours are microsatellite stable but chromosomally unstable. Preferential loss of genetic regions such as 5q, 18q, and 17p causes loss of tumour suppressor genes. A minority of tumours are microsatellite unstable but chromosomally stable. This instability causes mutations in genes from the same pathways as the microsatellite stable tumours.

Although the adenoma is often the premalignant lesion in HNPCC, and may often have *APC* mutation or loss, from here the pathways diverge (Figure 1.3). As discussed previously, tumours in HNPCC (and about 10% of sporadic colorectal cancers) have a 'mutator phenotype' and acquire mutations at a much increased rate. Although genes other than *K-ras, TP53* and *SMAD4* or *DCC* appear to be the targets of mutation the same pathways are ultimately disrupted (Figure 1.3). Tumours following this pathway (known as MSI+ or RER+) do not become chromosomally unstable, rather they tend to be diploid and have instability at the level of the DNA sequence. Tumour suppressor genes are then inactivated by small insertions/deletions and not disrupted via mutation plus LOH (loss of heterozygosity)(discussed previously).

MSI+ tumours frequently have stabilising mutations of β-catenin gene (Murata *et al.*, 2000), preventing β-catenin degradation, this disruption of the Wnt signalling pathway mirroring the inactivating *APC* mutations seen in MSI- (i.e. microsatellite stable) tumours. In a similar fashion, the TGFβ-signalling pathway is disrupted in MSI+ tumours via inactivating mutations of the *TGFβ2R* gene (Markowitz *et al.*, 1995; Parsons *et al.*, 1995), whereas this pathway is disrupted via *SMAD4* mutations in the MSI- tumours (Figure 1.3). Finally, there is convergence of the TP53 pathway with *TP53* mutations in MSI- tumours but *BAX* (BCL-2-associated X protein) mutations in MSI+ tumours (Mori *et al.*, 2001). Whilst the same genetic pathways may be inactivated in different ways in the different tumour types, tumours which are MSI+ tend to proceed through the adenoma-carcinoma sequence more rapidly (Jass, 1995).

THE HAMARTOMA – CARCINOMA SEQUENCE

For many years the increased cancer incidence in the hamartomatous polyposis syndromes (JPS, PJS, CD) was not thought to be of significance. It is now well established however that hamartomas do indeed have premalignant potential, although they are not perhaps as neoplastic as true adenomas. Peutz-Jeghers polyps are classically rather more epithelial than JPS hamartomas, so it is perhaps easier to explain how an epithelial malignancy might arise from these tumours. Wang *et al* demonstrated loss of heterozygosity (LOH) targeting the *LKB1* locus on 19p13.3 in hamartomas from PJS patients with a germline *LKB1* mutations (Wang *et al.*, 1999). The loss of the remaining *LKB1* allele was therefore sufficient to initiate growth of the PJS hamartoma. LOH at *LKB1* was also demonstrated in carcinomas derived from PJS patients indicating indeed there is a hamartoma-carcinoma sequence but involving genes outside of the classical adenoma-carcinoma sequence (*APC*, *K-ras* etc.). Gruber *et al* similarly demonstrated LOH at *LKB1* in PJS hamartomas but also showed that carcinomas derived from PJS patients had LOH at 17p (presumably targeting *TP53*) and 18q (presumably targeting *SMAD4* or *DCC*) (Gruber *et al.*, 1998). Interestingly, LOH of 5q (around *APC*) was not observed in these PJS adenocarcinomas, indicating that the *LKB1* mutation plus LOH apparently replaces *APC* as the initiating mutation. Thus, the PJS hamartoma – carcinoma sequence is probably similar to 'classical' colorectal tumorigenesis except inactivation of *LKB1* rather than *APC* is the first step. The pathways underlying JPS progression are discussed as part of this project.

AIMS OF THIS RESEARCH

At the beginning of the project, mutations of the *SMAD4* gene had just been shown to cause a subset of Juvenile Polyposis Syndrome cases. This gene was first described as a tumour suppressor gene deleted in pancreatic cancer (*DPC4*), however it was not known whether loss of SMAD4 function transmitted in the germline had tumour suppressing properties. Juvenile polyps are classically described as 'stromal' lesions. The contribution of SMAD4 loss and mutation to colorectal cancer overall was muddied by the fact that it is in close proximity to the *DCC* gene on 18q21.1, a gene which was thought to be the target of the high frequency of 18q loss observed in colorectal cancer. The gene or genes causing the remainder of the juvenile polyposis cases were yet to be discovered. In addition, reported linkage to 6q for Hereditary mixed polyposis syndrome was not possible in the light of new affection statuses of some family members. Thus, the aims of this project were;

1. To assess the contribution of *SMAD4* to juvenile polyposis. Once this was addressed I aimed to study the clinical and pathological features of patients according to their *SMAD4*-mutation status.
2. To characterise the role of *SMAD4* and assess the clonality of juvenile polyps. This was an attempt to clarify whether juvenile polyps were truly stromal lesions and, if so, how a stromal lesion could give rise to an epithelial malignancy.

3. To assess the true contribution of and timing of *SMAD4* mutations in the progression of colorectal cancers outside JPS. This aimed to ~~disseminate~~ determine the targets of 18q loss (*SMAD4* or *DCC*).

4. To screen candidate genes for germline mutations in JPS. Candidate genes were chosen from the same pathway as *SMAD4* under the hypothesis that they might too disrupt TGFβ signalling.

5. To search for new JPS loci using molecular (linkage analysis) and cytogenetic (comparative genomic hybridisation and loss of heterozygosity) techniques.

6. To assess the contribution of *BMPR1A* mutations to JPS.

7. Finally, to use linkage analysis to identify the correct HMPS locus.

CHAPTER TWO

MATERIALS AND METHODS

MATERIALS AND METHODS

2.1 DNA EXTRACTION

2.1.1 DNA EXTRACTION FROM BLOOD

DNA extraction from fresh or frozen blood was performed using either a DNA extraction kit (Nucleon), according to the manufacturer's instructions, or using the ammonium acetate method set out as follows. The first steps of this method acted to break down the cell wall to allow access to the nucleus. 9ml frozen blood samples were thawed and the blood transferred to a 50ml conical bottom Falcon tube (Greiner). Ice cold water wad added to the tubes to tubes to give a final volume of 50mls, then the tube inverted to mix and lyse the red blood cells. The tubes were then centrifuged at 2300rpm for 25 minutes at 4°C in a swing out rotor centrifuge (CR412 Jouan). The supernatant was discarded by inverting the tube gently, being careful not to disturb the pellet. The tube was inverted and placed on a clean paper towel to remove the last traces of supernatant. The pellet was then washed with 25 ml 0.1% NP-40 (Sigma) and centrifuged at 2300rpm for 20 minutes at 4°C. The wash was repeated if necessary. The supernatant was discarded and the tube inverted over a paper towel. To lyse the nuclei, 3ml nuclei lysis buffer was added to the pellets and the tube vortexed to resuspend the pellet completely. 200µl 10% SDS and 600µl proteinase K solution was then added to the tube to degrade any protein. The solutions were mixed by inversion and incubated at 60°C for 11/2-2 hours or overnight at 37°C. Following proteinase K

digestion, 1ml of saturated ammonium acetate solution (148g NH_4Ac (BDH) in 50mls distilled water) was added and the tube vortexed vigorously for 15 seconds. The tube was left to stand at room temperature for 20 minutes, then centrifuged at 2300rpm for 20mins at room temperature. The DNA was then in the supernatant and ready to be precipitated. The supernatant was transferred to a clean tube 50-ml falcon tube, and two volumes of ice-cold ethanol added. The contents were mixed by gentle inversion, then the DNA spooled out using either a fine glass rod or a fine plastic sterile loop. The spooled DNA was dipped into an eppendorf containing 70% ethanol (to wash the DNA and to remove any salts). The DNA was transferred to a labelled screw capped eppendorf, left to dry and then re-suspended in 1ml distilled water. To assess quantity and quality, an aliquot was diluted 1:50 and analysed by spectrophotometry at 260 and 280nm.

2.1.2 DNA EXTRACTION FROM CELL LINES

DNA was extracted from cell lines using a high salt method, avoiding the use of phenol. To pellet the cells, 50ml Falcon (Greiner) tubes were spun at 1000rpm for 10 minutes. The supernatant was removed and the cells washed twice in phospate buffered saline (PBS), before finally removing the supernatant. The cells (approximately $5x10^7$) were resuspended in 15mls SE buffer, and 50µl of 10mg/ml RNase A (Advanced Biotechnologies) added to the tube before incubation at 37°C for 1 hour to degrade any RNA. Proteinase K was then added to a final concentration of 200µg/ml and the tubes left overnight at 55°C to

degrade the protein. 4.5mls of pre-warmed 5M NaCl was added to the tubes to give a final concentration of 1.5M NaCl. 20ml chloroform (Merck) was then mixed in by rotation for 30 minutes before centrifugation at 2000rpm for 10 minutes. The aqueous layer was transferred to a new tube and an equal volume of isopropanol (BDH) mixed in for 5 minutes to allow the DNA to precipitate. After spooling out the DNA, it was washed in 70% ethanol for a minimum of one hour to remove any remaining salt from the DNA. The DNA was then re-suspended in distilled water and subjected to spectrophotometry to assess quality and quantity.

2.1.3 DNA EXTRACTION FROM PARAFFIN EMBEDDED TISSUE

Paraffin blocks were cut to give 5x10μm sections on non-coated slides. The slides were de-waxed in xylene (BDH) for 10 minutes, followed by two washes of 10 minutes each in 100% ethanol (BDH). Using a haemotoxylin and eosin stained slide as a guide for the area to be micro-dissected, the slides were scraped with a needle into an appropriate amount of digestion buffer (4.45ml dH_20, 500μl 10x magnesium-free PCR buffer (Perkin-Elmer), 50μl 20mg/ml Proteinase K (Merck)), depending on the size of the lesion. The tubes were vortexed and placed at 55°C for 1-3 days, with intermediate vortexing. The Proteinase K was de-activated by heating the tube to 95°C for 10 minutes and the tubes spun for 15 minutes at 13000rpm in a microfuge. The supernatant containing the DNA was transferred to a fresh tube and ready to be used.

2.1.4 DNA EXTRACTION FROM FRESH FROZEN TISSUE

DNA was extracted from fresh-frozen tissue using the QIAamp tissue kit (Qiagen) which is specified to allow up to 40μg of DNA to be extracted from 25mg of soft tissue. The manufacturer's protocol was followed. Briefly, the tissue was cut into small pieces, lysed, precipitated with ethanol and added to a spin column to which the DNA bound. After several washes, the DNA was eluted from the column with distilled water.

2.1.5 DNA EXTRACTION FROM CLONES

Isolation of DNA from P1 artificial chromosomes (PACs) was achieved using the Plasmid Midi Kit (Qiagen), following the protocol designed for the isolation of BAC (Bacterial Artificial Chromosomes) DNA. Stabs were streaked onto agar containing 25μg/ml Kanamycin and grown overnight at 37°C. Single colonies were inoculated into starter cultures of 5ml LB medium containing 0.5μl 25mg/ml Kanamycin. 0.5ml of the starter culture was then used to inoculate 100ml selective LB medium with 10μl 25mg/ml Kanamycin, and left to grow overnight with vigorous shaking (~250rpm). The manufacturers instructions for the kit were then followed.

2.2 RNA EXTRACTION

2.2.1 EXTRACTION OF mRNA FROM CELL LINES

The Fast-track kit (Invitrogen) was used for the extraction of mRNA from cell lines, according to the manufacturer's protocol. Briefly, cultured cells were first spun down in 50ml Falcon tubes at 1000rpm for 10 minutes and washed twice in PBS. The tubes were then centrifuged at 2000rpm to pellet the cells, the supernatant removed and the pellets either snap frozen in liquid nitrogen and stored at -70°C for later extraction or re-suspended and incubated in the Fast-track lysis buffer to digest proteins and ribonucleases. After adding sodium chloride to the lysate to a final concentration of 0.5M, any remaining DNA was sheared by passing the lysate several times through a syringe fitted with an 18-gauge needle. The mRNA was then separated out via the polyadenylated tail by mixing the lysate with oligo(dT) cellulose. The mRNA was then bound to a column, washed and eluted. After ethanol precipitation the sample was re-suspended in DEPC-treated (diethylpyrocarbonate treated to rid of RNases (Gibco)) water and stored at -70°C.

2.2.2 RNA EXTRACTION FROM FRESH FROZEN TISSUE

TRIzol reagent (Life Technologies) was used to extract total RNA from fresh frozen tissue, according to the manufacturer's instructions. The tissue was homogenised in 1ml TRIzol reagent per 100mg tissue and incubated at room

temperature for 5 minutes to allow the dissociation of nucleoprotein complexes. 0.2ml chloroform/1ml TRIzol was added and the tubes shaken vigorously before a further 5 minute incubation at room temperature. After centrifugation at 12000g for 15 minutes at 4°C, the aqueous phase was transferred to a clean tube and precipitated with 0.5ml isopropanol/1ml TRIzol reagent. Following a further spin at 12000g for 10 minutes at 4°C, the RNA pellet was washed with 75% ethanol. The pellet was briefly dried before re-suspending in RNase-free distilled water.

2.3 cDNA SYNTHESIS

cDNA synthesis was performed using the First Strand cDNA Synthesis kit (Amersham). 5µl of RNA was gently mixed with 10µl of DEPC-treated water, incubated at 65°C for 10 minutes then plunged onto ice. Meanwhile, 11µl of First Strand mix was mixed with 1µl DTT and 1µl of the random primer. This mix was added to the chilled RNA preparation and incubated at 37°C for one hour to allow cDNA synthesis. The cDNA was then ready for use.

2.4 POLYMERASE CHAIN REACTION

The polymerase chain reaction (PCR) is used to amplify regions of target DNA, and can be used provided as least part of the target nucleotide sequence is

known. Portions of the sequence which flank the desired target are used to design two synthetic DNA oligonucleotides, one complementary to each strand of the DNA double helix. These oligonucleotides serve as primers for *in vitro* DNA synthesis, which is catalysed by a thermostable DNA polymerase, with the primers determining the ends of the amplified DNA fragment. The Primer3 program was used to design primers (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). PCRs were usually performed in either 25µl or 50µl volumes, although volumes could be scaled up and down as necessary. A typical 25µl PCR reaction would be:

2.5µl 10x Mg^{2+}-free PCR buffer (Promega)

1.5µl Mg^{2+} (@ 25mM, giving final concentration of 1.5mM) (Promega)

1µl dNTPs (deoxyribonucloside triphosphates @ 2.5mM, giving final concentration of 0.1mM)(Pharmacia)

0.5µl Forward Primers (@ 20mM)

0.5µl Forward Primers (@ 20mM)

0.25 Taq DNA Polymerase (made in-house)

10-50ng DNA

Volume made up to 25µl with sterile dH_20

The DNA was aliquoted separately into microtitre plates (Advanced Biotechnologies), then the PCR master mix was made up with the remaining components, vortexed briefly and added to the plate containing the DNA. The

plate was then sealed with a heat plate to prevent evaporation during thermocycling. Three main stages comprised the PCR - first denaturation into single stranded DNA, followed by annealing, allowing the primers to find and anneal to the target sequence, then extension of the primers along the target sequence. A typical PCR reaction consisted of an initial denaturation of 94°C for 5 minutes, then 30-35 cycles each of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, then a final extension step of 72°C for 10 minutes, usually performed on Tetrad PCR machines (MJResearch). The annealing temperature was optimised according to the T_m of the primers. When standard conditions failed to amplify the target DNA, a touchdown technique was employed, with the annealing temperature of 70°C decreasing by 0.5°C each cycle, for 19 cycles, then kept at 52°C but increasing the length of the cycle by 1 second per cycle for 19 cycles.

2.5 AGAROSE GEL ELECTROPHORESIS

Agarose gel electrophoresis separates DNA molecules according to their size and was most frequently used for checking for the presence of PCR products. Agarose was prepared by boiling a mixture of agarose powder (Gibco BRL) in 1XTBE, at concentrations of 1-3% depending on the size of the DNA fragments. When cooled to approximately 50°C, ethidium bromide (10mg/ml) (Pierce) was added and mixed in to a final concentration of 0.25μg/ml. The ethidium bromide intercalates with the DNA and fluoresced under ultraviolet light, allowing

visualisation of the DNA fragments. Molten agarose was poured into a gel-casting tray with a comb in position and left to set. The comb was then removed, the gel was placed in a running tank and then covered with a running buffer of 1xTBE. 5µl of DNA was combined with 2µl of tracking dye and loaded into each well. 10µl of 1Kb ladder was loaded in the final well to allow sizing of the fragments and the gel electrophoresed at 100-130V for 10-30 minutes. Visualisation of the DNA was then performed by placing the gel on a UV transilluminator (260nm) and photographs taken using a UV products camera.

2.6 PURIFICATION OF PCR PRODUCTS

Prior to sequencing, PCR products were purified using the Qiaquick PCR purification spin columns or the Qiaquick 96-well format columns (Qiagen) according to the manufacturer's instructions. The purification separated the target DNA from excess dNTPs and primers that may have interfered with subsequent downstream experiments. For example, if too much unbound primer were present in purified PCR products, the chances of primer-dimer formation increased. The sequencing reactions performed on such templates would mainly consist of primer sequences and not the PCR product.

2.7 SEQUENCING PROTOCOLS

2.7.1 DIRECT SEQUENCING OF PCR PRODUCTS

Sequencing of PCR products was performed using the Big Dye Terminator sequencing mix (Applied Biosystems) which incorporates base-specific fluorescent nucleotides, utilising the dideoxy chain-termination method (Sanger, 1981). This method is based upon the enzymatic incorporation of dideoxyribonucleoside triphosphates in which the deoxyribose 3'-OH normally present is missing. When these modified nucleotides are incorporated the addition of subsequent nucleotides is blocked, which leads to fluorescent DNA 'ladders' of differing lengths which can then be separated on polyacrylamide gels. Sequencing reactions were made as follows:

8µl Big Dye Terminator mix (PE Applied Biosystems)

0.5µl primer (either forward or reverse, as used in the PCR)

6.5µl dH_20

5µl purified PCR product

Cycle-sequencing was performed in a PCR machine with an initial denaturation at 94°C for 4 minutes, followed by 25 cycles of 94°C for 30 seconds, 50°C for 10 seconds and 60°C for 4 minutes, with a final extension of 60°C for 7 minutes. The sequencing products were cleaned up to remove excess salts and big dyes using the Qiaquick columns as described in 2.6, and spun down in a vacuum

centrifuge until dry. The products were then re-suspended in 3µl of microSTOP loading buffer (Perkin-Elmer), denatured at 94°C for 4 minutes, and run on an ABI377 Sequencer (Applied Biosystems) on 5% Severn Super Sequencing mix (Severn) polyacrylamide gels.

2.7.2 SEQUENCE ANALYSIS

Sequences were analysed using Semi-adaptive base calling and Sequencing Analysis Version 2.1 (Applied Biosystems). Database searching using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) was performed to ensure the correct origin of the sequenced fragments. Alignment of sequences was performed using the Clustal V method based on a distance matrix (Higgins and Sharp, 1989) included in the MegAlign software module (DNASTAR). In addition, all sequences were examined by eye to look for heterozygous peaks that might not be detected by alignment tools.

2.8 RESTRICTION ENZYME DIGESTION OF DNA

2.8.1 RESTRICTION ENZYME DIGESTION OF PCR PRODUCTS

PCR-Restriction Fragment Length Polymorphism (RFLP) analysis was used to assess the frequency of point mutations that change the recognition sequence of specific restriction endonucleases. Restriction enzyme digestion of PCR products

was performed in 15µl volumes containing 10µl of the PCR product, 1.5µl of 10Xbuffer, an appropriate amount of BSA if required, 15U of the relevant enzyme and distilled water to make up the volume. Digestion was performed overnight for completion at the recommended temperature, then the fragments resolved on 2-3% 1XTBE agarose gels, depending on the size of the fragments being separated.

2.8.2 RESTRICTION ENZYME DIGESTION OF GENOMIC DNA

For Southern blotting, 11µg of DNA for each digest was required in ~10µl of dH_20. If the DNA was in a larger volume, it was ethanol precipitated and re-suspended in 10µl of dH_20. It was then aliquoted into a 96 well plate. The digest mix was then typically made up as follows:

2µl of enzyme buffer (*Eco*R1 buffer or NEBuffer 2 or 4)

2µl of appropriate enzyme (*Eco*R1,*Hind*III or *Sau*3A1 (NEB))

6µl dH_20

10µl of the digest mix was added to each well and put on PCR machine overnight at 37°C. 5µl of the restriction enzyme digest was then run on a 1% agarose gel containing ethidium bromide to check digestion progress with visualisation under UV light. If necessary, a further 2µl of the respective enzymes were added to the 96 well plate and put back at 37°C for further digestion. Once clean smears of DNA were achieved, all of the digest combined

with 5µl of loading dye was loaded onto 0.8% agarose gel (3.2g /400mls, without ethidium bromide) with 15-20µl of 1Kb ladder loaded into the final well to allow sizing of the fragments. The gel was run at 30V overnight, visualised by adding ethidium bromide to the running buffer and photographed. The gel was then ready for Southern blotting

2.9 SOUTHERN BLOTTING

Southern Blotting involves the transfer of DNA which has been electrophoresed in agarose gels to nylon membranes, (Southern, 1975) where the DNA will be immobilised. The DNA in the gel was denatured by submerging in denaturation buffer for 30 minutes with gentle rocking, then neutralised by submerging the gel in neutralisation buffer for 30 minutes, again with gentle rocking. The blotting apparatus were assembled as follows: A sheet of Hybond (Amersham) nylon membrane was cut to a similar size as the gel. A large gel tank was filled with transfer buffer, and a large glass plate put across the tank to make a platform. A wick was made using 3MM (Whatman) paper which had been saturated in transfer buffer. The treated gel was placed on top of the wick platform and any air bubbles rolled out with a glass pipette. The nylon membrane was placed onto the gel, avoiding the trapping of air bubbles. Three pieces of 3MM paper were cut to size, saturated in transfer buffer and placed on top of the membrane, again avoiding air bubbles. parafilm was put around the edges of the gel to make sure capillary action was only through the gel. A 5-10cM stack of absorbent paper

towels was placed on the 3MM paper and the assembly then covered with another glass plate and small weight if necessary. The gel was left to transfer overnight, after which the towels were removed. The 3MM paper, membrane and gel were turned over and the wells marked onto the membrane with a pencil to allow orientation. The gel was then discarded. The membrane was then fixed by placing it for 20 minutes DNA side down on three pieces of 3MM which had been soaked in fixing solution (0.4M NaOH, 16g in 1L dH_20). The membrane was then washed twice in 2XSSC (1/10 dilution of 20xSSC) and put between dry 3MM to dry until ready for use.

2.10 HYBRIDISATION OF SOUTHERN BLOTS WITH ^{32}P LABELLED OLIGONUCLEOTIDES

2.10.1 PREPARATION OF THE PROBES

PCR (as previously described) was performed on a control cDNA to provide the probe for Southern blotting. The probe was gel purified from a 1% low melting point agarose excised gel slice using the Geneclean II (Bio 101) kit, following the manufacturer's instructions, and could be frozen at -20°C until ready for use. To label the probe, the probe needed to be single-stranded; 40µl of dH_20 was added to 5µl of the purified probe, mixed gently and then denatured at 100°C for 3 minutes before being plunged straight onto ice. In the hood, the $\alpha^{32}P$ dCTP (6000Ci/mmol) (Amersham) was carefully opened and 5µl added to the denatured probe. The mix was then added onto the 'Ready to go' (Amersham)

DNA labelling beads, placed in a pre-warmed perspex box and put at 37°C for 20 minutes. The beads contain the necessary buffer, dATP, dGTP, dTTP, FPL*pure*™ Klenow Fragment and random oligodeoxyribonucleotides (9mers). To allow hybridisation of the DNA probe to the membrane-bound single-stranded DNA, it was also required to be single-stranded. The probe was therefore denatured again at 100°C for 3 minutes before being plunged straight onto ice, and was added to the membrane which had been pre-hybridised.

2.10.2 PREHYBRIDISATION AND HYBRIDISATION

In preparation, the pre-hybridisation and hybridisation solutions were pre-warmed in a 65°C water bath. The membranes to be hybridised were wet in 2XSSC and rolled in mesh (Hybaid) which allowed good contact between the solution and the membrane. The roll was placed inside glass Hybaid cylinders and 75mls of warm pre-hybridisation solution added. The cylinders were then incubated 65°C for 30 minutes whilst rotating. The pre-hybridisation solution was then discarded and 50mls of hybridisation solution added to the cylinder. 50µl of the labelled, denatured probe was then added to the cylinder, the lid tightly screwed on then the cylinder was left to rotate at 65°C overnight.

2.10.3 WASHING THE MEMBRANES

The low and high stringency wash solutions were pre-warmed to 65°C in a water bath. In the hood, the membrane was removed from the glass cylinder and the hybridisation solution carefully disposed of. The mesh was removed from the membrane and discarded of appropriately. The membrane was covered with 400mls of low stringency wash solution in a plastic tub, sealed with a lid, and put on a rocking platform at 65°C for 15 minutes. A second low stringency wash was performed for a further 15 minutes. The labelling was monitored with a Geiger counter so see how efficient the wash was. Two high stringency washes were then performed at 65°C for 15 minutes, with the wash solutions being carefully disposed of between washes. The membrane was removed from the tub, dabbed dry with tissue before being wrapped in saran wrap and taped into a X-ray cassette, avoiding entrapment of air bubbles. In the dark room, X-ray film (Kodak) was laid onto the membrane, then the cassette placed at -70°C for 3-4 days before developing the film in an automated developer.

2.10.4 STRIPPING THE FILTERS

It was possible to re-probe the membrane by stripping off the probe. This was achieved by washing the membrane in 1 litre of boiling 0.1% SDS solution, until the solution had cooled to room temperature. The membrane was then blotted dry and ready to be re-hybridised.

2.11 MUTATION DETECTION TECHNIQUES

2.11.1 SINGLE STRANDED CONFORMATIONAL POLYMORPHISM ANALYSIS

The rate of migration of single stranded DNA under non-denaturing conditions through a polyacrylamide gel is sensitive to secondary structure, and this structure in turn depends on the nucleotide sequence..If there are sequence differences between different strands of DNA, the secondary structure may be altered and this will be detected as a band of altered mobility on Single stranded conformational polymorphism (SSCP) analysis. SSCP was used to search for mutations such as point mutations (missense or nonsense), insertions and deletions in PCR products of up to 350 base pairs in length. Two main SSCP methods were employed – polyacrylamide gel electrophoresis (PAGE) combined with silver staining or capillary separation using an ABI310 (Applied Biosystems).

2.11.1.1 SSCP USING PAGE AND SILVER STAINING

PAGE was performed using either the Phast minigel system (Pharmacia) or using large self-poured plates (Bio-rad Protean II SSCP system). For the Phast system, 2µl of the PCR product was combined with 2µl of SSCP loading buffer (98% formamide (Amersham) containing 0.05% bromophenol blue (Sigma) and 0.05% xylene cyanol (Sigma), and denatured at 95°C for 5 minutes. Larger fragments

were run on pre-cast 12.5% gels for 130-150 volt-hours, and smaller fragments separated on pre-cast 20% gels for 140-150 volt hours, with the running temperature generally 10°C or 15°C. Silver-staining was used to detect the bands according to the manufacturer's instructions. For the self-cast large gels, the following mix was prepared and poured between two clean plates assembled with spacers:

19.8mls dH_20

8.2ml acrylamide:bis 35% (39:1)

8ml 5XTBE

2ml Glycerol (BDH)

The gel mix was polymerised with 300µl 10% w/v ammonium persulphate (APS)(Sigma) and 50µl TEMED (NNN'N'-tetramethylethylenediamine) (BDH). 5µl of PCR product was combined with 5µl SSCP loading buffer before being denatured and loaded on the set gels. To size the separated products, 5µl of 100 base pair ladder (Gibco) was combined with 5µl of SSCP loading buffer and loaded into the final well. The gels were then run at 150Mamps at room temperature for ~4 hours or overnight in the cold room (4°C). The gels were separated from the plates and then silver stained (shown in Solutions section). Any samples showing aberrant migration were re-amplified from the source DNA, purified using Qiaquick columns (Qiagen) and directly sequenced using the ABI Big Dye Terminator kit (Applied Biosystems).

2.11.1.2 SSCP USING THE ABI310

The second technique for SSCP analysis utilised a capillary based method. PCR products were amplified with the forward and reverse primers both fluorescently dye-labelled (FAM, TET or HEX). 5µl of diluted (1/50) PCR products were combined with 0.5µl internal size standard (Tamra 350, Applied Biosystems) and 11.5µl 310 loading buffer. The samples were denatured at 95°C for five minutes, plunged onto ice and then run on an ABI310 sequencer (Applied Biosystems), using 2% Genescan polymer containing glycerol (Applied Biosystems). SSCP was performed under two different temperature conditions (20°C and 35°C), according to manufacturer's instructions. Sequence changes within the fragments presented as a different pattern or altered size when compared to the size standard. Fragments showing both aberrant and normal migration were re-amplified using non-fluorescently labelled primers, purified using Qiaquick columns (Qiagen) and then sequenced in both forward and reverse orientations using the ABI Big Dye Terminator kit (Applied Biosystems).This 96-well format high throughput system allowed three fragments to be assessed simultaneously in the same well.

2.11.2 PROTEIN TRUNCATION TEST

2.11.2.1 *IN-VITRO* TRANSLATION OF PCR PRODUCTS

The protein truncation test (PTT) detects nonsense mutations – point mutations which result in a stop codon, or frameshifts which results in a downstream aberrant stop signal. PCR primers were designed with MYC, T7 RNA-polymerase binding site and RBS (ribosomal binding site) tags and an in-frame start codon on the forward primer. These latter tags enabled transcription and translation respectively. The tag added to the forward primers was as follows:

5'GGATCCTAATACGACTCACTATAGGAACAGACCACCATGGAACAAA
AATTAATATCGGAAGAGGATTTGAAT

PCRs were performed on cDNA using standard conditions to amplify the entire coding region with overlapping fragments, ensuring each fragment started in-frame. The products were then kept at -20°C until ready for use. To allow *in vitro*-coupled transcription and translation (IVTT), 15µl of the tagged PCR products were combined with the following reagents which had been carefully mixed together:

8µl Rabbit Reticulocyte Lysate
0.66µl TnT buffer
0.34µl amino acids minus methionine
0.34µl T7 polymerase
0.66µl ^{35}Smethionine

0.33µl RNase inhibitor
4.67µl dH_20

The mix was then incubated for 1or 2 hours at 30°C on a PCR machine, with the plate covered by 3MM paper to stop any isotope vapour escaping.

2.11.2.2 ELECTROPHORESIS OF TRANSLATED PRODUCTS

For the electrophoresis of the resulting 'proteins', two polyacrylamide gel mixes were required – one to stack the proteins and one to resolve the proteins. Aberrant stops were detected as a truncated 'protein' band when subjected to PAGE. To prepare enough 12% resolving gel mix for two gels, 8ml acrylamide (30%, 37.5:1)(Severn Biotech) was mixed with 7ml distilled water and 5ml lower buffer mix.. Two clean plates were assembled with a gasket acting as a spacer, and bulldog clips holding the assembly together. 100µl of 20% w/v ammonium persulphate (APS) and 20µl of TEMED were added to the lower gel mix, swirled gently and then poured into the glass plate assembly. 500µl of water was added on top of the resolving gel to ensure a straight edge was obtained. Whilst the resolving gel polymerised, the 1.125% upper (stacking) gel mix was made by mixing 2.5ml acrylamide (30% 37.5:1) with 8.5ml distilled water and 2.5ml upper buffer. The water was poured away from the lower gel, and 500µl of the un-polymerised upper gel mix added in it's place to allow good contact between the two gels. A 12-well sharks-tooth comb was added at an angle to allow the pouring of the upper gel. The upper gel mix was poured away from the

lower gel, and 20µl TEMED and 100µl 20% APS added to the remaining upper mix before pouring it onto the lower gel. The comb was straightened and excess gel wiped away. 1.5-2 inches of PTT running buffer was added to the running tank. Once the gels had set (10-20 minutes), the clips and gasket and comb were carefully removed. The wells were rinsed and straightened using a syringe filled with running buffer. The two plates were clamped into the running tank and the reservoir filled with running buffer. 10µl of sample buffer (9µl bromophenol blue plus 1µl 1M DTT) was added to each well, the plate was sealed and covered with 3MM to stop any vapour escaping and then denatured at 95°C for 5 minutes. 15 µl of each sample was carefully loaded into wells 1-11, avoiding air bubbles, then 7.5µl of multicoloured protein marker (NEN) was added to the final well for orientation. The gel was run at 60m amps for 1-1.5 hours or until the loading buffer was visible near the bottom of the gel.

2.11.2.3 FIXING, DRYING AND EXPOSURE OF THE GEL

Following electrophoresis, the running buffer was carefully disposed of in the hood and the plate assemblies were placed on paper towels to ensure that no isotope was left on the surface. The plates were separated using forceps and the gels placed into fixing solution until the bromophenol blue in the sample buffer had turned green. In the meantime, the gel dryer was warmed to 80°C with a piece of 3MM paper in. The gels were then transferred onto a piece of 3MM on the bench added to the 3MM in the dryer. The gels were dried under vacuum for

1 hour and then taped into an X-ray cassette. In the dark room, Kodak film was laid over the gels and left overnight before developing in an automatic developer

2.11.3 WESTERN BLOTTING

Western blotting is useful for the detection of specific proteins using antibodies as probes, and was used to look for altered or reduced protein expression of particular genes. Cell proteins are separated using PAGE, transferred to a membrane to immobilise them, and after blocking unspecific protein sequences, exposed to the primary antibody. Exposure to a secondary antibody labelled with an HRP (horse radish peroxidase) conjugate is followed by detection with the light emitting ECL (enzyme chemical luminescence) reagents (Amersham).

2.11.3.1 PREPARATION OF CELL LYSATES FOR ELECTROPHORESIS

Cultured cells (either lymphoblastoid or epithelial lines) were spun down in 20ml volumes in Falcon tubes at 2000rpm for 5 minutes. The supernatant was removed and 10mls of fresh PBS added to the tubes, pipetting up and down to re-suspend the pellet completely. A further 10mls PBS was added and the tubes spun at 2000rpm for 5 minutes, then the supernatant removed. Judging by the pellet size, an appropriate amount of PBS was added and the re-suspended cells aliquoted into 1.5ml microfuge tubes, so each tube contained $5x10^6$–$5x10^7$ cells. The microfuge tubes were then spun down at maximum speed for two minutes in a

microfuge before removing the supernatant and storing the pellets at -70°C until required for Western blotting. Cell pellets were re-suspended in 100µl of cell lysis buffer and vortexed. The tube lids were pierced and the tubes boiled at 100°C for 10 minutes. After vortexing again, the lysates were ready for PAGE.

2.11.3.2 SEPARATION AND TRANSFER OF PROTEINS

The gels were prepared as for the separation of Protein Truncation Test products (2.11.2.2), except that the lower gel was 15% (15ml 30% acrylamide, 7.5ml lower buffer, and 10ml dH_20). 30µl of the prepared cell lysates were loaded into each well (the remainder could be stored at -20°C for later use), with the final well loaded with 15µl of the multicoloured protein marker (NEN). The gels were then run at 40-50Mamps for 1-1.5 hours. For the transfer of the proteins from the gels to the PVDf (polyvinyl)membrane (Millipore), the membranes (10x10cm) were first 'activated' by placing them in 200ml methanol (BDH) in a plastic tub. The Western transfer buffer was then added to the tub to make a final transfer buffer of 2X running buffer with 20% methanol. 4 pieces (per gel) of 10x10cm 3MM paper were also added to the tub to act as buffer reservoirs during transfer. Two pieces of 3MM paper and the wet PVDf membrane were placed in the SemiPhor™ semi-dry transfer equipment (Hoefer), and a little transfer buffer added. The stacking gel was removed with a scalpel, and the resolving gel placed on top of the membrane. The edges were trimmed to make them neat and 2 more pieces of 3MM added to the 'sandwich'. Air bubbles were removed by gently

rolling over the sandwich with a marker pen. After adding a little more transfer buffer, the transfer was run at 150mA for 1 hour, or until the transfer could be seen to be complete by the coloured marker in the membrane. The membrane was then blocked in 5% skimmed milk powder for 1 hour or overnight, on a rocking platform.

2.11.3.3 EXPOSURE AND DETECTION OF ANTIBODIES

The membranes were exposed to the primary antibody diluted in 3% skimmed milk powder and left for 1 hour at room temperature or 4°C overnight. They were then rinsed three times and washed three times for 30 minutes in 0.2%Tween (Sigma) in PBS. The membranes were then exposed for one hour to the secondary antibody (with HRP-conjugate) diluted in 3% skimmed milk powder, before being washed as they were for the primary antibody. 25ml of the two ECL reagents (Amersham) were mixed together and added to the membranes for one minute. The membranes were dried by dabbing with tissue and quickly taped into an X-ray cassette before being exposed to Hyper-film (Amersham) for 1 minute, and 5 minutes. The films were developed in an automatic developer.

2.11.3.4 STRIPPING OF WESTERN BLOTS

Western blots could be stripped and probed with an alternative antibody. The membranes were placed in stripping solution at room temperature for 30 minutes, then washed in PBS for 30 minutes, both with agitation. The membrane could then be re-blocked with 5% skimmed milk powder and exposed to the new antibody as described for the first.

2.12 STERILE LYMPHOCYTE SEPARATIONS

Separations were performed so that the lymphocytes could be established as permanent cell lines by Epstein-Barr Virus transformation. 25ml blood was collected in Falcon tubes containing 25ml sodium citrate medium. The contents of the Falcon were poured into a 250ml flask, and the tube rinsed with 4ml filtered RPMI which had been brought to room temperature. Approximately 25 sterile glass beads (BDH) and 0.6ml 1M $CaCl_2$ (BDH) were added to the flask and defibrination started immediately for 15 minutes at 250rpm on a gyratory shaker. 20ml RPMI was added to the flask and the defibrinated blood divided into two tubes, layering carefully over 15ml lymphoprep (Robbins Scientific). The tubes were spun at 1800 rpm for 20 minutes in a centrifuge with a swing out centrifuge rotor, with the speed carefully brought up and down. The interface was then transferred to a new Falcon and spun at 2300rpm for 10 minutes. The pellet was washed in 20ml RPMI then re-spun at 2300rpm for 10 minutes. The

pellets were re-suspended in 2ml freeze mix, divided into two labelled cryotubes (Corning) and placed at -80°C overnight. The lymphocytes were then stored in liquid nitrogen until ready for transformation.

2.13 TISSUE CULTURE

2.13.1 FEEDING CELL LINES

Once the lymphocytes had been transformed and returned as a growing culture, the cells would either need feeding or spinning down for DNA/RNA/protein extraction as described above. The cells were examined under a microscope to determine the viability and density. If the cells were to be left to continue to grow but were confluent, an equal volume of growth medium was added to the flask and left at 37°C overnight in 10% $C0_2$. The cells were then re-examined and either left to grow or spun down for required protocol.

2.13.2 FREEZING DOWN CELLS TO REPLACE STOCKS

To ensure the cell lines remain a permanent resource, an aliquot of the growing cultures was always frozen down. 50ml of the growing culture was spun at 2000rpm for 5 minutes, the supernatant removed and the tube inverted to dry. The pellet was then fully re-suspended in 2ml freeze mix, and aliquoted into

sterile cryotubes labelled with the cell line name, date, volume spun down and initials. The tubes were frozen at -70°C overnight then stored in liquid nitrogen.

2.14 FLUORESCENT IN SITU HYBRIDISATION

Fluorescent *in situ* hybridisation (FISH) is most frequently used to either map a specific probe to a particular chromosome by hybridising the probe to metaphase spreads, or to study the copy number of a particular gene by hybridising the probe to interphase nuclei. Rather than these conventional uses of FISH, the technique was here used on paraffin embedded sections to look for the copy number of a particular probe in specific cell types. In the normal cell, two copies of the probe would be present, but if a cell had lost a copy of the gene being probed then only a single copy of the probe would be detected.

2.14.1 DIRECT LABELLING OF PROBES

PAC 224_j_22 was obtained from ~~was obtained from~~ Human Genome Mapping Project Resources (http://www.hgmp.mrc.ac.uk/) and the DNA extracted as described in 2.1.5. The DNA was labelled with biotin using the Bionick kit (Life Technologies) as follows. 1µg PAC DNA was combined with 5µl dNTP mix containing BIO-14-dATP, 10 µL DNA polymerase I/DNase mix, 1µL DNA polymerase I and made up to 50µl with distilled water, before being left to label

at 15°C for one to three hours. 5µl was then run on a 1% agarose gel to check the progress of the nick translation. When the fragments were between 100 and 500 base pairs in length the reaction was stopped by adding 5µl of 0.5M EDTA.

2.14.2 PRECIPITATION OF DIRECTLY LABELLED PROBES

The labelled probes were precipitated and hybridised to metaphase spreads to ensure they mapped to the region of interest. After confirming the origin of the probe, the probe was hybridised to tissue sections to ascertain the copy number. To precipitate the probe, 15µl of the labelled DNA was combined with 8µl Cot1 competitor DNA (GibcoBRL), 2µl salmon sperm DNA (GibcoBRL), and 100µl 100% ethanol. The mix was put at -70°c for 1 hour or overnight at -20°c, and then centrifuged at maximum speed in a microfuge for 15 minutes. The ethanol was then removed and the probe left to air dry. After re-suspending in 10µl of hybridisation mix, the probe was denatured at 85°C for 5 minutes and ready to be hybridised to the metaphase spreads or tissue sections which had also been denatured.

2.14.3 MAPPING OF PROBES ON METAPHASE SPREADS

Metaphase spreads slides were made from phytohaemagglutinin (PHA) stimulated lymphocytes, cultured for 48-72 hours before thymidine was added to

synchronise the cells. This gave chromosomes lengths of 400-500 bands. The metaphase spread slides were denatured under 70% formamide on a 72°C hotplate for 1.5 to 2 minutes. The coverslip was flicked off and the slides put straight into ice cold 70% ethanol. The slides were then dehydrated in 95% and absolute ethanol for 1 minute each and air dried. The denatured probe was added to the denatured slide. A glass coverslip was added and sealed with rubber cement (Weldtite) and the slides placed in a humidified chamber at 37°C overnight. The slides were washed three times in 50% formamide/2 X SSC, and three times in 2 X SSC, with all washes performed at 42°C for five minutes each. The slides were rinsed in SSCT (SSCTween), blocked for 10 minutes with 5% skimmed milk powder in SSCT (SSCTM) and rinsed again in SSCT. The specific antibody for the detection of the probe (avidin/FITC for biotin labelled probes) was diluted 1/500 in SCCTM and added to the slide for 10 minutes. The slides were then washed again in SSCT for 3 minutes, then twice in PBS with agitation at room temperature. After dehydrating through an ethanol series of 70%, 95% and 100% ethanol for 2 minutes each, 20µl DAPI counter-stain was added and the slides viewed using a cooled coupled device camera at -25°C (Quantix Photometrix) attached to a microscope (Applied Imaging). Images were captured using Quips software (Vysis).

2.14.4 HYBRIDISATION OF PROBES TO TISSUE SECTIONS

10μm sections of paraffin embedded tissue were cut onto coated slides and de-waxed as described previously. The slides were then subjected to a protein digestion by proteases using the Tissue Kit (Oncor), a pre-treatment not being necessary. The slides were digested for 20 minutes at 45°C, and rinsed in 2xSSC for 10 seconds. The slides were then dehydrated through a 70%, 95% and 100% ethanol series and air dried. 20μl of propidium iodide was added to the slide and the digestion evaluated, according to the protocol. If the tissue was appropriately digested without loss of morphology, the slide was denatured on a 67°C hotplate for 5 minutes, the denatured probe added and the slide put to hybridise at 37°C overnight in a moist chamber. If the probe was commercial, 1.5μl was combined with 15μl hybridisation mix and denatured simultaneously with the slide.

2.14.5 DETECTION OF PROBE SIGNAL

40ml of the appropriate SSC wash solution (1xSSC for centromeres and 2xSSC for unique sequence probes) was pre-warmed at 72°C. The coverslips were removed from the slides and the slides immersed for 5 minutes without agitation. Slides were then put in 1xSSCT for 2 minutes at room temperature before being incubated with 60μl of the appropriate detection reagent for 5 minutes at 37°C. The slides were washed three time in SSCT for 2 minutes each, before being counter-stained with anti-fade DAPI, and visualised as described in 2.12.3.

The FISH experiments were not without technical difficulty. Often the sections would simply float off the slide, possibly because the sections were thick-cut. Many slides from different polyps of Family 20 were used in the experiments before a result was achieved, prior to this most of the slides would have no signal at all for either the centromere probe or the PAC probe. This was possibly a feature of the method of fixing the tissue after surgery (perhaps the sections were left in formalin too long). Finally, several different digestion times were used on the slides before 20 minutes was indeed found to be the optimum length of time (the slides were either over- or under-digested).

2.15 COMPARATIVE GENOMIC HYBRIDISATION

Comparative genomic hybridisation (CGH) is a useful tool for performing a genome wide scan of chromosomal loss and gains in a tumour (Kallioniemi *et al.*, 1992). A mixture of DNA from malignant and normal cells are differentially labelled with red or green fluorochromes and then hybridised onto metaphase spreads. Images of 5-10 metaphases are captured and quantification of the fluorescence ratios performed using a digital image analysis system. The relative ratios of red and green are then compared. Regions of genetic material which have been lost during tumour progression will show as red, and regions which have been gained will show as green. The thresholds of detection for CGH are

regions greater than 10Mb for loss of genetic material, and gains of 2Mb or more if the region is amplified five times.

2.15.1 NICK TRANSLATION AND PRECIPITATION OF PROBE

1μg of tumour DNA or test DNA were labelled with FITC–12-dUTP (Vysis) or Texas Red–5-dUTP (Vysis) respectively. This was achieved by mixing the DNA, 1μl of the relevant fluorochrome, 5μl dNTPs, 10μl DNA polymerase I/DNase mix, and 1μl DNA polymerase I and distilled water to make a total volume of 50μl. The mix was incubated at 15°C for 2 hours, then left on ice while 5μl was run on a 1% agarose gel. Probe fragments forming a smear ranging in size between 500 and 2000 base pairs were the best length for smooth hybridisations. The reactions were stopped by adding 5μl of 0.5M EDTA and could be stored at -20°C until ready for use. The tumour and test DNAs were then combined in an Eppendorf tube and mixed with 50μl of human Cot-1 DNA, 0.1 volume 3M sodium acetate, and 2 volumes cold absolute ethanol. The DNA was precipitated on dry ice for 30 minutes or overnight at -20°C. After spinning at 15000rpm, the supernatant was carefully removed and the pellet left to air dry. The pellet was then re-suspended in 10μl of hybridisation mix and denatured at 75°C for 5 minutes. The DNA was left to pre-anneal at 37°C for 30 minutes to 1 hour whilst the slides were prepared for hybridisation .

2.15.2 DOP-PCR LABELLING OF TUMOUR DNA FOR CGH

Degenerate oligonucleotide primed polymerase chain reaction (DOP-PCR) labelling of tumour DNA was performed when there was not sufficient or good quality DNA to nick translate, usually when the DNA was extracted from paraffin embedded tissue. The PCR-labeling used a degenerate primer (sequence 5' ccgactcgagnnnnnnatgtgg 3') and had two stages, initial low stringency cycles, where the specific bases at the 3' end of the oligonucleotide theoretically primed every 4 kb along the template DNA, and then an increased number of cycles with high stringency, whereby the oligonucleotide 'tailed' DNA from the initial cycles was amplified. Further DOP-PCR, with differential fluorescent nucleotides incorporated into the PCR reaction was then performed to label the tumour and normal DNA with their respective fluorochromes. Labeled DNA was then precipitated and hybridised to metaphase spreads, using the same protocols as those for nick-translated DNA. The first round DOP-PCR reactions were set up as follows;

2.5µl 10 X Mg^{2+} free PCR buffer

2.5µl dNTPs (@ 2 mM, giving final concentration of 200µm)

0.5µl DOP 6MW primer (at 100µm giving concentration of 2.0µ M)

4µl Mg^{2+} (@25mM giving concentration of 4mM)

0.5µl Taq polymerase

15µl tumour or normal DNA

Volume made up to 25µl with dH_20

The DOP-PCR conditions were as follows: 1 cycle of 9 mins @ 94°C, 8 cycles of (1 min @ 94°C, 1.5 mins @ 30°C, 3 mins @ 72°C) then 25 cycles of (1 min @ 94°C. 1 min @ 62°C, 1.5 min @ 72°C) followed by a final extension of 8 mins at 72°C.

The second round labeling DOP-PCR experiments were set up as follows;

5µl 10 X Mg^{2+} free PCR buffer

5µl labeling dNTPs (2 mM each of dATP, dCTP and dGTP, 0.5mM dTTP, giving final concentration of 200µm)

0.5µl DOP 6MW primer (at 100µm giving concentration of 2.0µ M)

8µl Mg^{2+} (@25mM giving concentration of 4mM)

1µl Taq polymerase

10µl of first round DOP-PCR tumour or normal DNA

Volume made up to 50µl with dH_20

The labeling DOP-PCR conditions were as follows: 1 cycle of 4 mins @ 94°C, then 25 cycles of (1 min @ 94°C. 1 min @ 62°C, 1.5 min @ 72°C) followed by a final extension of 8 mins at 72°C.

2.15.3 DENATURATION OF SLIDE AND HYBRIDISATION

Each batch of slides, whether commercial (Vysis) or made in-house, had a pre-determined optimal denaturation time. Prior to denaturation, slides were examined under the microscope to ensure they were suitable for hybridisation, with many, distinct metaphases free of cytoplasm being ideal. The slides were then denatured on a hotplate at 73°C with denaturation solution under a 22 mm X 50 mm coverslip. The coverslip was then flicked off and the slides placed in ice cold 70% ethanol for 3 minutes, then dehydrated through an ethanol series for 3 minutes each. The slides were air dried and then ready for use. For the hybridisation, the slides were placed on a hotplate at 37°C and 10µl of a denatured probe added to each half of the slide. Each probe was covered with a 22 X 22mm coverslip, sealed with rubber cement and sealed in a moist chamber for 48-72 hours at 37°C.

2.15.4 POST-HYBRIDISATION WASHES OF THE SLIDES

Following hybridisation, the coverslip was removed from the slides and the slides subjected to 3 X 5 minute washes in 50% formamide/2XSSC at 42°C, then 3 X 5 minutes in 2XSSC, again at 42°C. The slides were then subjected to a 5 minute wash at room temperature in SSCT, whilst shaking gently, before dehydrating through an ethanol series (70%, 95%, 100%) and being left to air dry. The slides were then mounted in DAPI (approximately 20µl under a 22 X

50mm coverslip), and either stored in a cardboard folder at 4 °C or captured immediately.

2.15.5 IMAGE ACQUISITION AND ANALYSIS

5-10 metaphases per experiment were captured using an epifluorescence microscope (Applied Imaging) equipped with a triple-color epifluorescence filter set (selective for the fluorochromes DAPI, FITC, and rhodamine) in combination with a cooled CCD camera (Quantix Photometrix). Images were captured using Quips software (Vysis). The metaphases were karyotyped using the digitally inverted DAPI image which gave a G-banded pattern. After karyotyping the relative intensities of the red and green signals were analysed and an average obtained for multiple metaphases. CGH experiments were considered successful if enough fluorochrome had been incorporated to give smooth intense color that was not granular in appearance.

2.16 SOLUTIONS AND MEDIA FOR MOLECULAR TECHNIQUES

1M $CaCl_2$

14.7g of Calcium Chloride (BDH) made up to 100ml with dH_20

Cell lysis buffer

1ml buffer (2x bromophenol blue, with SDS+sucrose)

800µl dH20

200µl 1M DTT () (makes 0.1M DTT)

Denaturing solution:(for 2 gels)

262.93g sodium chloride

60g sodium hydroxide pellets (BDH) - made up to 3L with dH_20.

310 dilution and running buffer

12.5ml 20xTBE

25ml glycerol (BDH)

made up to 250ml with dH_20 and stored at 4°C.

100XDenhardt's solution

10g ficoll 400 (Pharmacia)

10g polyvinlypyrrolidine (BDH)

10g bovine serum albumin (Sigma)

Made up to 500ml with dH_20, stored at -20°C.

0.5M Ethylenediamine tetraacetate (EDTA)

93g EDTA (BDH)

10g sodium hydroxide pellets (BDH)

400ml dH_20

pH adjusted to 8.0, made up to 500ml with dH_20 and autoclaved.

Ethidium bromide (10mg/ml)

0.1g of ethidium bromide (Pierce) dissolved in 10ml dH_20. Stored in dark.

Fixing solution :10% ethanol/5% acetic acid

100mls 100% ethanol (BDH)

50mls acetic acid (BDH)

850mls dH_20

Freeze mix (90%FCS/10%DMSO)

9ml foetal calf serum (GibcoBRL))

1ml Dimethyl sulphoxide (BDH)

5% stock Genescan polymer

7.14ml GS polymer (Applied Biosystems)

2.86ml 310 dilution buffer

2% Genescan polymer

2ml 5% stock GS polymer

3ml 310 dilution buffer

Growth Medium 10% FCS/RPMI

180mls RPMI (one bottle)

20mls FCS (one vial)

High stringency wash (0.2XSSC/0.1%SDS)

10mls 20XSSC

10mls 20% SDS - up to 1L with dH_20

Hybridisation solution (500ml)

211.25mls dH_20

150mls 20XSSC

100mls 50% Dextran Sulphate

12.5mls 10% SDS

25mls 100X Denhardts solution

1.25mls 10mg/ml Salmon Sperm DNA (Amersham)

310 Loading buffer

1100µl deionized formamide (Amersham)

50µl 0.3N (0.6g/50ml) sodium hydroxide

Low stringency wash (2XSSC/0.1% SDS)

100mls 20XSSC

10mls 20% SDS - up to 1L with dH_20

Lower buffer mix (pH 8.8)

90.75g Tris (1.5M)

20ml 10%SDS (0.4%)

made up to 500ml with dH_20 and pH adjusted to 8.8 with HCL.

Luria Broth (LB)

10g sodium chloride

5g bacto yeast extract

10g bacto-trytone

900ml dH_20

pH adjusted to 7.0 with sodium hydroxide, made up to 1L with with dH_20, and autoclaved.

LB agar plates

LB medium prepared as above, and 15g/Litre bacto-agar added before autoclaving.

Nuclei lysis buffer(50ml)

10mM Tris (500µl 1M)

400mM NaCl (4mls 5M)

2mM EDTA (200µl 0.5M) (BDH)

45.3mls dH_20

Neutralising solution:(for 2 gels)

262.93g sodium chloride (Sigma)

181.5g Trizma base (Sigma) - made up to 3L with dH_20.

Prehybridisation solution: (500ml)

311.25mls dH_20

150mls 20XSSC

12.5mls 10% SDS

25mls 100X Denhardts solution

1.25mls 10mg/ml Salmon Sperm DNA (Amersham)

Proteinase K buffer (stored at 4°C)

2mM Na-EDTA (200µl 0.5M solution)

1% SDS (5ml 10%SDS)

44.8mls dH_20

Proteinase K solution

2mg Proteinase K (Merck) in 1ml EDTA/SDS buffer.

PTT running buffer

3g Tris (Sigma)

14.4g glycine (BDH)

10ml 10% SDS, made up to 1L with distilled water

3% w/v skimmed milk powder

12g skimmed milk powder (Premier Brands) in 400mls PBS

5% w/v skimmed milk powder

20g skimmed milk powder (Premier Brands) in 400mls PBS

SE buffer

75mM NaCl (7.5mls 5M)

25mM EDTA pH 8.0 (2.5mls 0.5M)

1% SDS (50mls 10%)

made up to 500ml with dH_20 and sterilised through 0.2 μ pore filter.

Silver Staining solutions/protocol

2 x 3 minutes in **fixing** solution: 10% Ethanol(100ml/L), 0.5% acetic acid (5ml/L)

1 x 15 minutes in **staining** solution: 0.1% (1g/L) silver nitrate ($AgNO_4$) (Sigma)

2 x 1 minute in dH_20

1 x 20 minute in **developing** solution: 1.5% (9.374g/L)sodium hydroxide (BDH),

0.1% (1ml/L)Formaldehyde (BDH), prepared immediately before use.

1 x 10 minutes in **stop** solution:0.75% (7.5g/L) sodium carbonate (NA_3CO_4) (BDH).

Gentle agitation required with all solutions.

3M Sodium acetate

61.52g sodium acetate (BDH)

200ml dH_20

pH adjusted to 6.0, made up to 250ml with dH_20, and autoclaved.

5M sodium chloride (NaCl)

73.1g sodium chloride (BDH)

made up to 250ml with dH_20 and autoclaved.

Sodium dodecyl sulphate (SDS)

10% w/v volume SDS (BDH) in sterile dH_20.

5xTBE

54g Tris base (Sigma)

27.5g Boric acid (BDH)

20ml 0.5M EDTA pH 8.0

Made up to 1 litre with dH_20 and autoclaved.

Transfer buffer(20XSSC)

88.23g Tri-sodium citrate (BDH)

175.32g sodium chloride (Sigma)- made up to 1L with dH_20 (final pH7-8).

1M Tris

60.55g Tris base (Sigma)

400ml distilled water, pH adjusted to 8.0, then made up to 500ml dH_20 and autoclaved.

Upper buffer mix (pH 6.8)

30.25g Tris (0.5M)

20ml 10%SDS (0.4%)

Made up to 500ml with dH_20, pH adjusted to 6.8 with HCL.

Western transfer buffer

6g Tris

28.8g glycine

made up to 800ml with dH_20, then added to 200ml methanol used to activate the membrane.

Western Wash solution (0.2% Tween/PBS)

2ml Tween (Sigma) in 1L PBS

2.17 SOLUTIONS FOR CYTOGENETIC TECHNIQUES

Denaturing solution for slides

70% formamide (700µl)

2XSSC (100µl 20XSSC)

200µl dH_20

Detection reagents

For Biotin labelled probes – avidin-FITC (Vector labs) diluted 1/500 with SSCT

For digoxygenin labelled probes – anti-Dig-Rhodamine (Boehringer Mannheim) diluted 1/100 in SSCT.

Ethanol series for dehydration

70% - 700ml absolute ethanol and 300ml dH_20

95% - 950ml absolute ethanol plus 50ml dH_20

100% - absolute ethanol

Formamide wash solution

50% formamide (250ml)

2XSSC (50ml 20XSSC)

made up to 500ml with dH_20

Hybridisation mix

1ml 20 x SSC (=2XSSC)

5ml formamide

1g dextran sulphate (=10%)

1ml Tween (=10%)(Sigma) – made up to 10ml with dH_20, aliquoted and stored at -20°C.

2XSSC (10% v/v 20XSSC)

50ml 20XSSC

450ml dH_20

SSCT

4X SSC (100ml 20XSSC)

0.05% Tween-20 (250µl)

pH 7.0, made up to 500ml with dH_20

SSCTM

5% skimmed milk powder (25g)

made up to 500ml with SSCT

CHAPTER THREE

THE CONTRIBUTION OF SMAD4 TO JUVENILE POLYPOSIS SYNDROME AND THE CLINICAL FEATURES ASSOCIATED WITH *SMAD4* MUTATION STATUS

THE CONTRIBUTION OF SMAD4 TO JUVENILE POLYPOSIS SYNDROME AND THE CLINICAL FEATURES ASSOCIATED WITH *SMAD4* MUTATION STATUS

3.1 INTRODUCTION

Hahn *et al* in 1996 described a region on 18q21.1 which was deleted in a large proportion of pancreatic cancers, and termed the putative tumour suppressor locus *DPC4* (Deleted in Pancreatic Cancer locus 4) (Hahn *et al.*, 1996b). The gene was subsequently characterised and found to have homology to the Drosophila gene Mad (mothers against decapentaplegia, and thus the gene acquired its second, and more commonly used, name *SMAD4/MADH4* (Hahn *et al.*, 1996c). About 80% of pancreatic cancers and 60% of colorectal cancers have allele loss of 18q21.1, and a certain amount of this loss has been shown to target *SMAD4* (Hahn *et al.*, 1996b; Thiagalingam *et al.*, 1996). Hahn *et al* demonstrated that of 25 of 84 pancreatic carcinomas had *SMAD4* homozygously deleted, and a further six mutations in *SMAD4* were identified in twenty seven tumours where *SMAD4* was present (Hahn

et al., 1996c) thus explaining a significant amount of the allele loss observed at 18q21.1.

Linkage analysis for juvenile polyposis syndrome had previously been limited to exclusion of 5q as the region containing the causative gene, thus eliminating the *APC* and *MCC* (mutated in colorectal cancer) genes (Leggett *et al.*, 1993). These two genes are mutated or subjected to allele loss in Familial Adenomatous Polyposis (another disease with polyposis and increased risk of gastrointestinal cancer) and sporadic colorectal cancer (Groden *et al.*, 1991; Kinzler *et al.*, 1991b). In an attempt to identify JPS loci, Howe *et al* in 1998 performed targeted linkage analysis to candidate tumour suppressor regions. For this analysis, a five generation American kindred were used consisting of 43 individuals, of whom 13 were affected with familial juvenile polyposis. Overall the family consisted of 117 members of whom 29 were affected. The loci that were tested for linkage in this JPS family were *MSH2, MLH1, MCC, APC, HMPS, CDKM2A, JP1, PTEN, KRAS2, TP53, LKB1* and *DCC/SMAD4*, and a maximum LOD score of 5.0 was obtained with markers mapping to close to *DCC/SMAD4* (Howe *et al.*, 1998a). Subsequently, both *SMAD4* and *DCC* were screened for germline mutations and a 4 base pair deletion in exon 9 of *SMAD4* was identified that co-segregated with disease (Howe *et al.*, 1998b). In this study, a further two familial cases and two sporadic cases of JP were also found to have *SMAD4* mutations, and two familial and two sporadic cases were found not

to have *SMAD4* mutations, indicating early on that *SMAD4* mutations could not account for all JPS cases.

The aim of this chapter was to establish the true frequency of *SMAD4* mutations in juvenile polyposis syndrome, by studying the protein as well as the DNA. Whilst undertaking this (particularly the immunohistochemistry) it became apparent that the polyps of *SMAD4* mutation carriers were subtly different to those of non-*SMAD4* mutation carriers, and therefore a blinded analysis was performed in an attempt to segregate the polyps according to their *SMAD4* mutation status. Finally, with the confident exclusion of *SMAD4* as the causative gene, the clinical features of the mutation carriers *versus* non-mutation carriers were compared in order to determine the clinical spectrum conferred by a germline *SMAD4* mutation.

3.2 THE CONTRIBUTION OF SMAD4 MUTATIONS TO JPS

Subsequent to the demonstration that germline mutations could cause *SMAD4* in JPS, several reports were published detailing the findings in different groups of patients (summarised in Tables 3.2.1 and 3.2.2). Overall 78 separate families or cases of juvenile polyposis have been reported, and of these 21 have been shown to be due to mutations of the *SMAD4* gene (27%).

My own work on juvenile polyposis started immediately after the discovery that *SMAD4* mutations were responsible for some cases of the disease. Germline *SMAD4* mutation screening has been ongoing since this time, owing to the continual collection of patients and the need to exclude *SMAD4* as the cause of their disease. For inclusion as juvenile polyposis syndrome, the patients were classified as having five or more juvenile polyps of the gastrointestinal tract, or one or more juvenile polyps and a family history of juvenile polyposis, according to criteria suggested by Jass *et al* (Jass *et al.*, 1988). Pathology reports and medical notes were used to confirm the disease status. In addition, patients were assessed for features of Cowden disease (CD), Bannayan-Zonana syndrome (BZS) or Gorlin syndrome and not included if they met any of the diagnostic criteria associated with these respective syndromes (assessed using questionnaires, medical records and pathology reports). To ensure the patients were indeed not CD or BZS, they were screened for germline mutations in *PTEN*, which typically causes 80% of CD and 50% of BZS, but no mutations were found (Marsh *et al.*, 1997b). With the growing realisation that reports of *PTEN* mutations in JPS (Olschwang *et al.*, 1998b) were probably incorrect (Eng and Ji, 1998), *PTEN* screening was discontinued and cases were classified as JPS on clinicopathological grounds only.

Group, Journal	Number	Family/case	Mutation and its predicted effect	Clinical information
Howe et al, Science,1998	*5 out 7***	*1 13 **	*4 bp deletion exon 9, premature stop at codon*	*upper GI and colonic polyposis, crc and st ca. Pa ca*
		*M-1**	*4 bp deletion exon 9, premature stop at codon*	*not detailed*
		JP5/1	*4 bp deletion exon 9, premature stop at codon*	*not detailed*
		JP11/1	*2 bp deletion exon 8, premature stop*	*colonic and gastric jps*
		JP 10/1	*1 bp insertion exon 5, premature stop*	*30-40 colonic polyps aged 6*
Houlston et al, Hum Mol Gen	***1 out of 21***	***AF****	***Arg-Cys codon 361 exon 8, missense***	***4 colonic and gastric jps aged 16, father crc.***
Friedl et al, Gen Chr Can, 1999	*3 out of 11*	*FJP-4**	*4 bp deletion exon 9, premature stop*	*multiple colon and gastric jps, aged 35-40*
		*FJP-12**	*4 bp deletion exon 9, premature stop*	*multiple colonic jps aged 4-5, 'healthy' father*
		FJP-15	*2 bp deletion exon 6, premature stop*	*>50 jps aged 12, polyposis of stomach aged 28*
*Roth et al, Gen Chr Can**,1999*	*3 out of 5*	*FAM1*, aka 4/1*	*tyr396ser, missense*	*10-50 colonic jps and TAs*
		*FAM3**	*4 bp deletion exon 9, premature stop at codon*	*>50 colonic, sb and gastric jps*
		S2, aka 6/1	*Ser177X, premature stop codon*	*10-50 colonic, sb and gastric jps*
Kim et al, Int J Ca, 2000	*3 out of 5*	*IJU-JP1*	*gln-stop codon 388 exon 9 , premature stop*	*many jps stomach, colorectum aged 20*
		SNU-JP2	*glu-lys, codon 390 exon 9, missense*	*28 colorectal jps aged 16*
		*USU-JP1**	*arg-his, codon 361, exon 8, missense*	*>20 colorectal jps aged 16*
K W-Richens et al,Gut, 2000***	***4 out of 13***	***17***	***2 bp ins nt1564 exon 11, premature stop***	***>100 gastric jps, 8 colorectal jps***
		20	***189-197del9ins52 exon 2, premature stop***	***extensive colon and gastic jps colectomies aged 21,***
		BL	***Q445X exon 10, premature stop***	***>100 jps, colonic ad, exocrine pancreatic***
		SV	***11bp del nt 516 exon 4, premature stop***	***50+ sigmoid and rectal polyps, colon ca aged 48,***
K W-Richens et al,Am J Path,	***2 out of 11***	***21***	***g-a +1 intron 2 splice site, aberrant splicing***	***>50 colonic and gastric jps, bowel cancer aged 31,***
		MTW	***Q180X exon 4, premature stop***	***Multiple polyps***

Table 3.2.1 Summary of published SMAD4 mutations in juvenile polyposis syndrome.

*Bold type shows the ICRF patients analysed for the work presented in this thesis,. 'Number'= no of SMAD4 mutants in cohort studied. *= familial, **= overlap of patients between Howe et al and Roth et al, ***= patients reported more than once, but only counted once. ad=adenoma, ca=cancer, crc=colorectal ca. jps=juvenile polyps, sb=small bowel, st=stomach, TA=tubular adenoma, del=deletion,ins=insertiion*

Group, Journal	*Family/case*	Clinical information
Howe et al, Science	*Jp6/1 aka*	*no mutation found in this study but see Roth et al*
	JP 4/1, aka fam1*	*no mutation found in this study but see Roth et al*
	JP 1/1, aka SP1	*10-50 colonic jps, colon ca*
	JP 2/13, aka fam4*	*10-50 colonic and gastric jps*
Houlston et al, Hum Mol Gem	***1****	***sb, colonic and gastric jps and ad, ca ileum aged 49***
	3*	***colonic, jejeunal, gastric polyps, ca ileum, ca stomach***
	5*	***multiple jps, colorectal ca***
	6*	***colonic, jejeunal, gastric polyps, ca ileum, ca stomach***
	8*	***50+ colonic and sb polyps, colonic polyps, colorectal ca***
	9*	***colonic polyps aged 17, jps and ad aged 19 , colonic polyps aged 4***
	10*	***ileum and colon jps, ca colon aged 41***
	FT*/11	***colonic polyps aged 18 and 30, ca colon***
	12*	***florid polyposis aged 4, 7 and 8 colon and sb, ca jejunum x3 age 27, 37, ~35***
	14*	***multiple colonic jps, ad and TA, colorectal ca***
	16*	***20+ colonic and stomach polyps aged 14 and 20, , ca colon***
	SM96*	***juvenile polyps, mixed juvenile/adenomatous polyps, ca colon***
	SM397*	***?***
	SM524*	***>15 polyps***
	SM106	***70+polyps***
	1204	***?Cronkhite-Canada***
	c2	***Multiple colonic, ileal and duodenal polyps aged 4***

Table 3.2.2 Summary of patients found not to have germline SMAD4 mutations, and their gastrointestinal clinical features.

*Bold type shows the ICRF patients, *= familial, **= overlap of patients between Howe et al and Roth et al, ***= patients reported more than once, but only counted once, jps=juvenile polyps, ca=cancer, ad=adenoma, sb=small bowel, TA=tubular adenoma, del=deletion, ins=insertion . Patients highlighted red have subsequently been found to harbour mutations in BMPR1A (see chapter nine). Table continued on next page.*

Group. Journal	*Family/case*	
	KS	***multiple colonic jps, mixed Ad and JPs age 13***
	1262	***109 colorectal polyps***
	1469	***Multiple colonic and ileal polyps***
Friedl et al, Gen Chr Can	*FJP-1*	*multiple colonic jps aged 28*
	FJP-2	*3 colonic jps aged 3*
	FJP-3	*>30 colonic jps aged 32*
	FJP-8	*>6 jps aged 4*
	*FJP-9**	*>50 jps aged 10*
	*FJP-10**	*>200 jps aged 12*
	*FJP-11**	*15 colonic polyps aged 7*
Roth et al	*FAM2*, aka 7/1*	*>50 colonic and gastric jps, adenomas*
	sp3	*>10 colonic and sb polyps, 6 jps, 2 hyperplastic, 7 Tas*
Kim et al, Int J cancer	*SNU-JP1*	*>300 jps of stomach, small bowel and colorectum aged 16*
	SNU-JP3	*7 jps of rectum and rectal cancer aged 67*
K W-Richens et al,Gut***	***15****	***colorectal jps aged 14+, colorectal cancer aged 38, 61***
	WN*	***caecal ca aged 47, jejunal JPs and ad, and TA, ca colon aged 47, 100+ jps***
	SR	***50+ caecal polyps***
	SS	***50+ jps, Ca colon aged 25, ca pancreas, ca stomach***
	SH	***multiple colonic jps***
	SCA	***multiple sigmoid polyps aged 27***
	SC	***30-50 colonic polyps aged 15***
	SD	***jejunal polyps aged 15***
	RV	***multiple jps aged 13***

Table 3.2.2 Summary of patients found not to have germline SMAD4 mutations, and their gastrointestinal clinical features, continued. *Table continues on next page.*

Group	*Patient/case*	*Clinical information*
*K W-Richens et al***,Am J Path*	*18**	*>170 colonic jps aged 18 and 21, rectal cancer*
	*19**	*?*
	*22**	*Multiple jps and adenomas aged 7, 100+ jps*
	MD	*Mulitple JPs, througout colon, adenomas, colon cancer aged 21*
	*YC**	*?*
	WH	*Multiple polyps, colon ca.*
	LB	*80+ JPs aged 3, jejunal, ileal and colonic*
	DM	*Multiple caecal and colonic polyps, CRC*
	HR	*Multiple jps in colon, sb and stomach, rectal ca aged 35, colon ca aged 43*
	BN	*?*
	CV	*?jps 18q deletion*
	CWN	*?jps, 18q deletion*
	SM316	*<20 JPs*

Table 3.2.2 Summary of patients found not to have germline SMAD4 mutations, and their gastrointestinal clinical features, continued.

3.2.1 ASSESSMENT OF LINKAGE TO THE *SMAD4* REGION

Prior to the genome screen, eight JPS families had been assessed for linkage to the *SMAD4* region on chromosome 18 (Houlston *et al.*, 1998). Two families (12 and MD) provided good evidence against linkage to the *SMAD4* region, and a further two provided weaker evidence against linkage (6 and 14). Two-point linkage analysis for chromosome 18q markers in JPS families was undertaken using markers derived from the Weber8 (Research Genetics) set (analysis details in Chapter Seven). The two-point LOD scores for the chromosome 18 markers mapping near to *SMAD4* are shown in Table 3.2.1.1. Strong evidence against linkage was again provided by families MD and 12, where the LOD scores were consistently near to or less than –2, the figure considered to provide significant evidence against linkage. The two-point LOD scores were also negative for the *SMAD4* markers in Family 6 (but not 14) and it is therefore likely that *SMAD4* is not responsible for the JPS in Families 6, 12 or MD. The exclusion of *SMAD4* in these families was supported by the failure to share haplotypes for markers spanning the *SMAD4* region (Figure 3.2.1.1). Four families assessed (1, 5, 10 and C1) were compatible with linkage in the Houlston *et al* paper but in these four no mutations of *SMAD4* were found by screening of the gene (Houlston *et al.*, 1998). The genome wide search with alternative markers and the construction of haplotypes (Figure 3.2.1.1) for the *SMAD4* region confirmed that Families 1, 5 and C1 are compatible with linkage to 18q. In addition, Family 14 was compatible with linkage to 18q, but two unaffected

individuals shared the putative 'affected' haplotype and would thus be non-penetrant if an underlying *SMAD4* mutation were the cause of their JPS. Family 18 were also compatible with linkage to 18q. No *SMAD4* mutation has been identified in Families 1, 5, C1, 14 or 18 despite the comprehensive analysis detailed in Chapter Three. The LOD scores (and the haplotype sharing) may reflect the false positives which occur by chance in this type of analysis. This is certainly the case for Family 18, in whom a *BMPR1A* mutation has subsequently been identified (discussed in Chapter Nine), but who were still compatible with linkage to the *SMAD4* markers.

The two-point LOD scores for the 18 markers in Family 10, and additional to the Houlston *et al* paper, Family 19, were also negative, but did not reach –2 for markers mapping near to *SMAD4*. Haplotype construction (Figure 3.2.1.1) showed that the affected individuals (both sib-pairs) of Families 10 and 19 did share half of their 18q alleles, but without any parental DNA the phase could not be determined. It was therefore not determinable whether *SMAD4* was the causative JPS gene in these two families (but deemed unlikely given that no mutations were found despite the comprehensive screen detailed in Chapter Three, and indeed *BMPR1A* mutations have subsequently been identified in these two families). The two-point LOD scores of markers mapping to the *SMAD4* region in Family 15 were negative, and haplotype construction showed that this family showed evidence against linkage to 18q (Table 3.2.1.1 and Figure 3.2.1.1). Two Finnish families (2/13 and 7/1) were also assessed for linkage to *SMAD4* markers, where the two-point LOD scores

indicated no evidence of linkage to 18q markers, in keeping with the fact that no mutations were identified in these families by sequencing of the gene (Roth *et al.*, 1999).

Family	*D18S535 (64cM)*	*D18S851 (74cM)*	*D18S858 (79cM)*	*ATA7D07 (90cM)*	*GATA7E12 (102cM)*	*ATA82B02 (110cM)*
1	***0***	***0.3***	***0***	***0.18***	***0***	***-1.86***
5	***0.3***	***0.3***	***0.3***	***0***	***-1.86***	***0***
6	*-0.21*	*-0.66*	*-0.06*	*0*	*-0.8*	*-1.8*
*10**	*-0.08*	*-0.14*	*-0.11*	*0.26*	*0.25*	*0.24*
12	*-1.7*	*-2.04*	*-2.45*	*-2.14*	*-1.93*	*-0.08*
14	*-1.64*	***0.34***	***0.2***	*-0.05*	*0.51*	*-0.17*
15	*0.6*	*0.3*	*-1.87*	*-2.39*	*-1.92*	*-1.75*
*18**	*0.27*	*0.12*	*0.25*	*-1.89*	*-0.3*	*-0.03*
*19**	*0.24*	*-0.14*	*-0.08*	*0.17*	*-1.89*	*-0.08*
20 **	***0.78***	***0.34***	***0.65***	***0***	***0.69***	***0.45***
C1	*0.3*	*0*	*0*	*-1.8*	*-1.8*	*0.3*
MD	*-2*	*-2.02*	*-1.83*	*-1.86*	*0.16*	*-0.11*
*7/1**	*-0.22*	*0.59*	*-1.32*	*0*	*0*	*0*
2/13	*-3.87*	*-1.76*	*-1.44*	*0*	*0*	*0*

Table 3.2.1.1 Two-point LOD scores for chromosome 18q.
*Shown are the two-point scores for θ=0 for the markers mapping to chromosome 18 in the JPS families. Distances are shown in brackets after the marker name. SMAD4 lies between D18S851 and D18S858. *= families subsequently shown to harbour BMPR1A mutations, **= Family with a SMAD4 mutation. Bold type shows families who are compatible with linkage to the SMAD4 region but in whom no mutation has been identified.*

Family 17 have a *SMAD4* mutation (discussed in detail shortly), and haplotype construction confirmed the sharing of 18q alleles, as would be expected (Figure 3.2.1.1). Family 20 also have a *SMAD4* mutation and the two-point LOD scores for the markers mapping to chromosome 18q were positive, reflecting this (Table 3.2.1.1). The highest two-point LOD score achieved for Family 20 for 18q markers was 0.78, substantially lower than the maximum theoretical LOD score of 1.2 for this family.

In summary, Families 17 and 20 were compatible with linkage to 18q and subsequently found to have *SMAD4* mutations. Families 1, 5, C1, 14 and 18 were compatible with *SMAD4* linkage but no *SMAD4* mutations were identified. Linkage to 18q could not be disproved or proven in Families 10 and 19 (however, both have subsequently been found to harbour mutations in *BMPR1A*). Families 12, 15, MD, (and most likely Family 6), as well as the two Finnish families 7/1 and 2/13, were not compatible with linkage to markers mapping to 18q, the *SMAD4* region, and in these families it was highly likely a gene other than *SMAD4* was the causative gene.

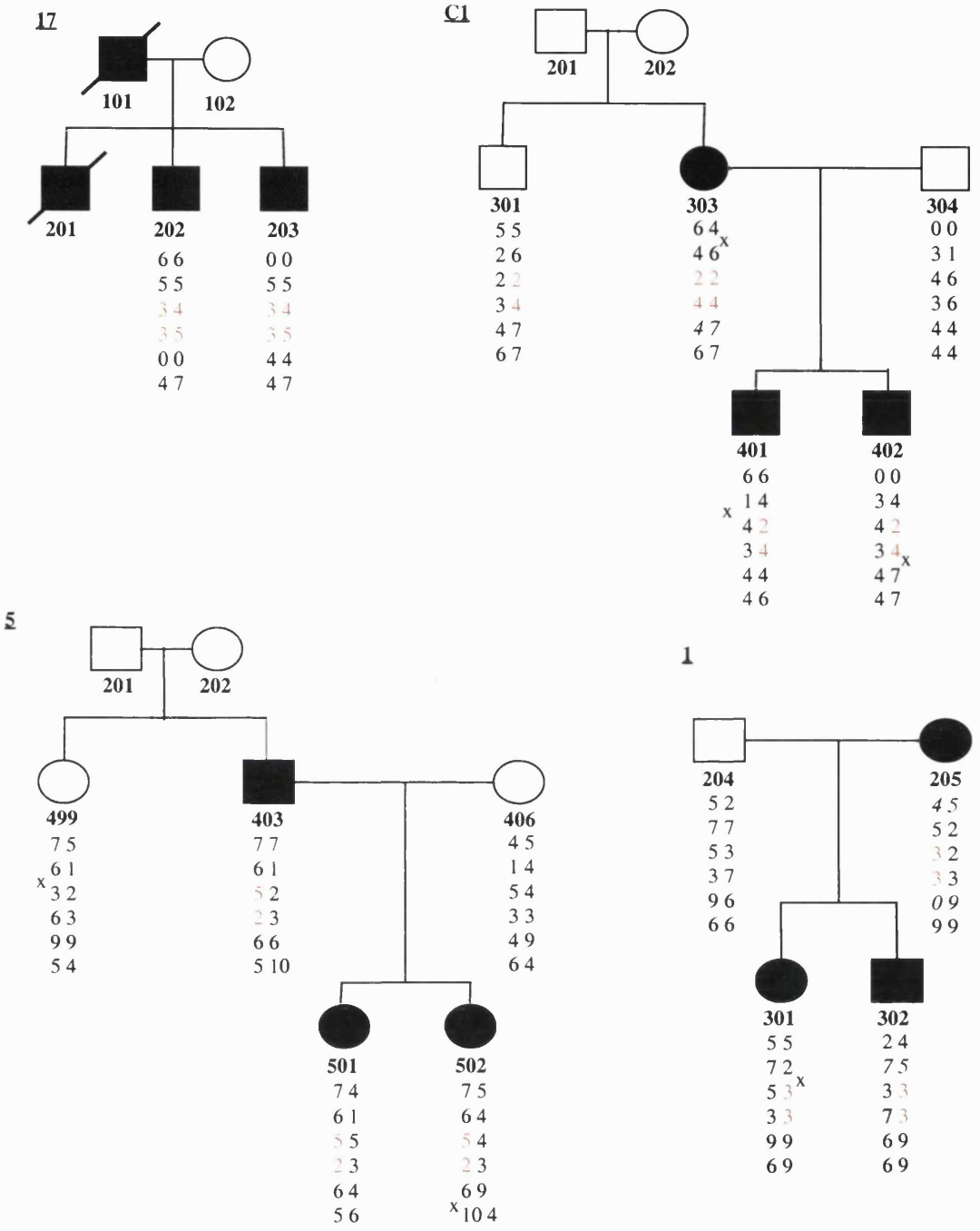


Figure 3.2.1.1. 18q haplotypes in juvenile polyposis syndrome families.

Shown are the haplotypes for 6 markers mapping to 18q (D18S877, D18S535, D18S851, D18S858, ATA7D07 and GATA7E12), from the genome wide linkage search (discussed in Chapter Seven). SMAD4 maps between D18S851 and D18S858, and these alleles are highlighted as red in affected individuals to aid the tracking of the affected haplotypes. Affected individuals are shown with filled symbols. Inferred alleles are shown in italics. Family 17 had a SMAD4 mutation and this was reflected in the sharing of alleles at 18q markers. Families 1, 5, and 18 were compatible with linkage to SMAD4 as all affected individuals shared the same 18q alleles. Figure continues on next page.

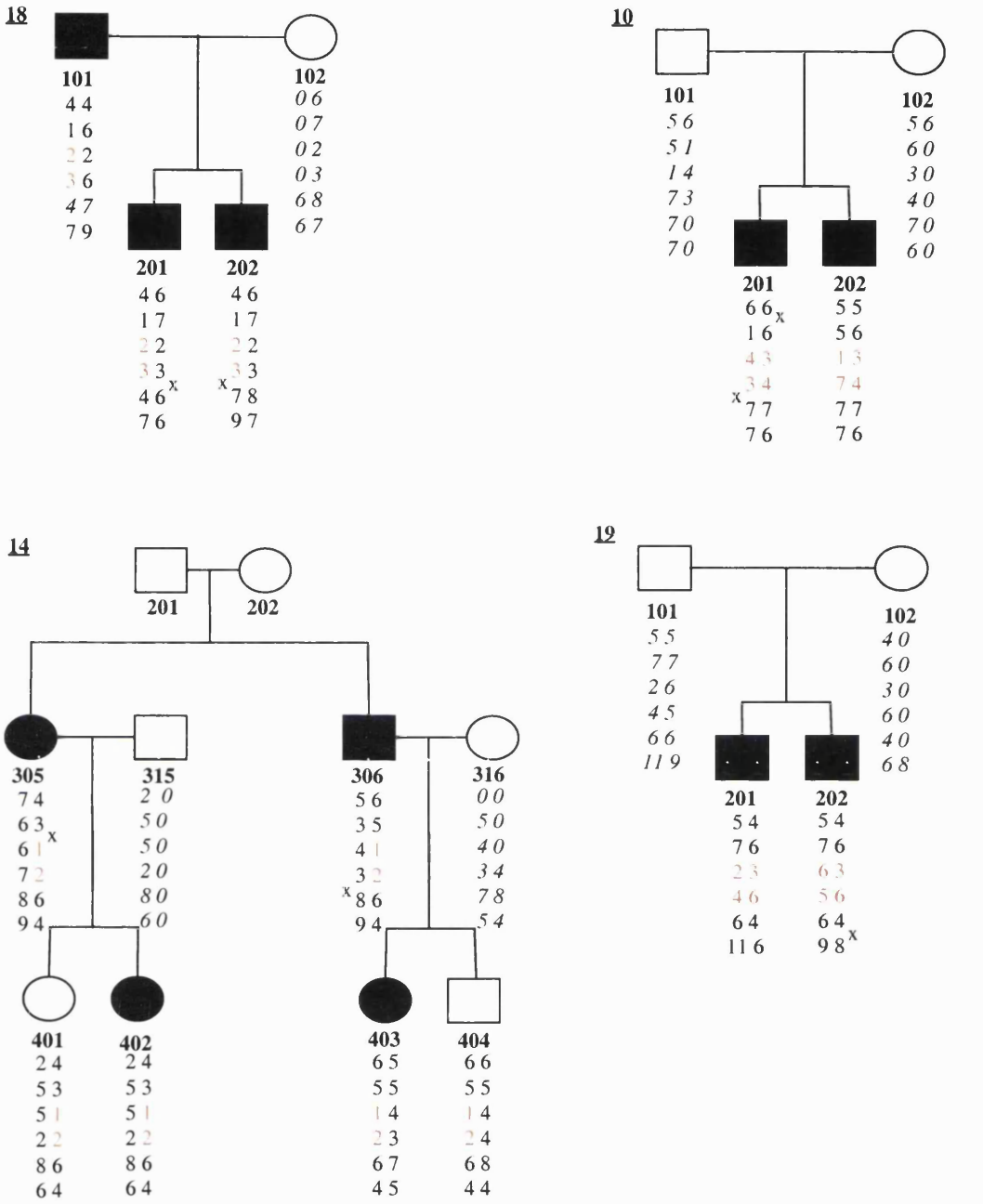


Figure 3.2.1.1. continued. *Linkage to SMAD4 in Family C1 can not be disproved as the affected siblings do share alleles but their affected parent is homozygote at these markers. Family 14 was compatible with linkage to 18q markers, but two unaffected individuals (401 and 404) also shared the 'affected' haplotype. Linkage to SMAD4 can not be disproved in families 10 and 19 who shared half the alleles at 18q. Due to the lack of parental DNA, it was not possible to determine the phase.* *Figure continued on next page*

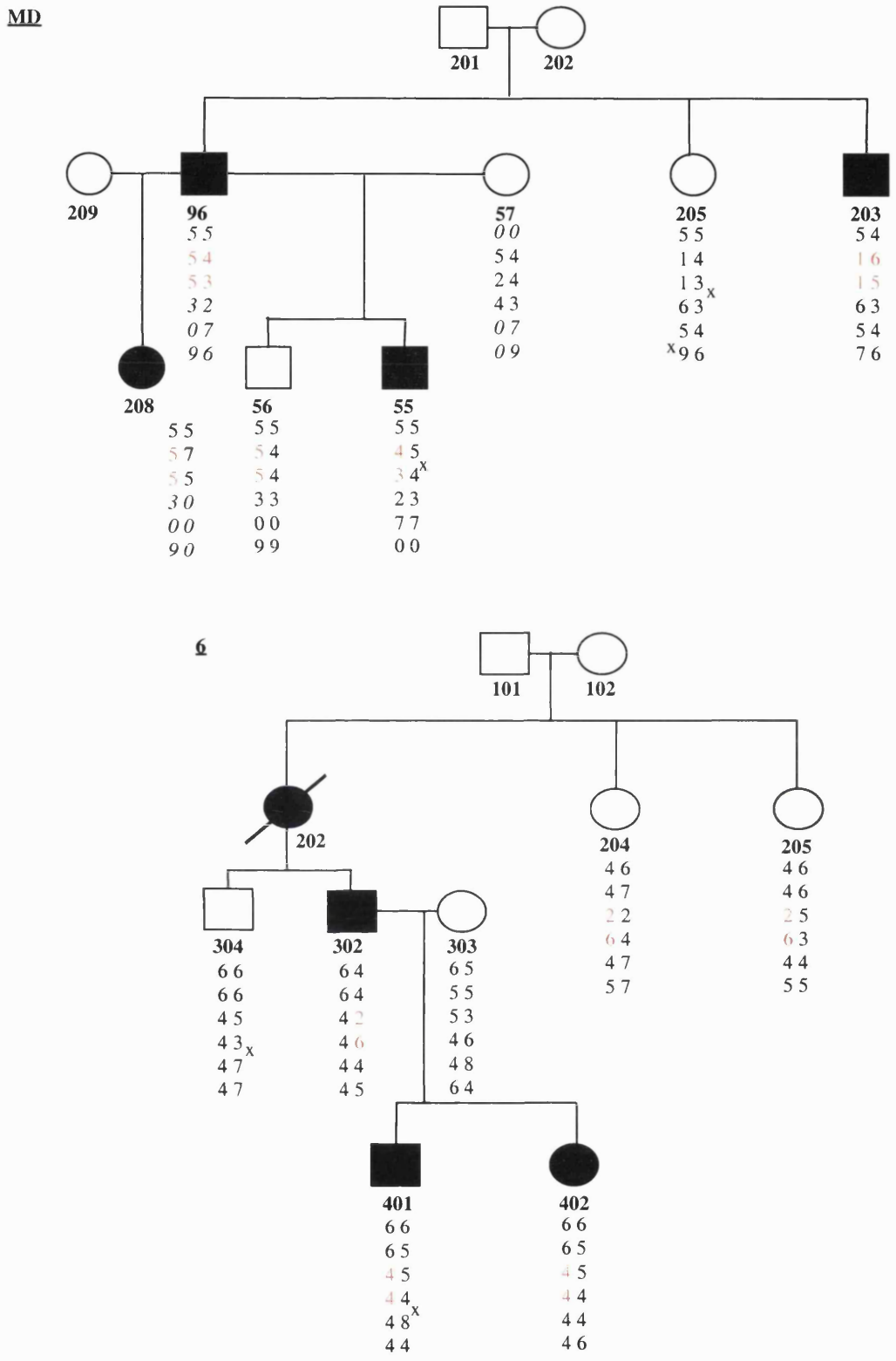


Figure 3.2.1.1. continued. *Figure continued on next page.*

12

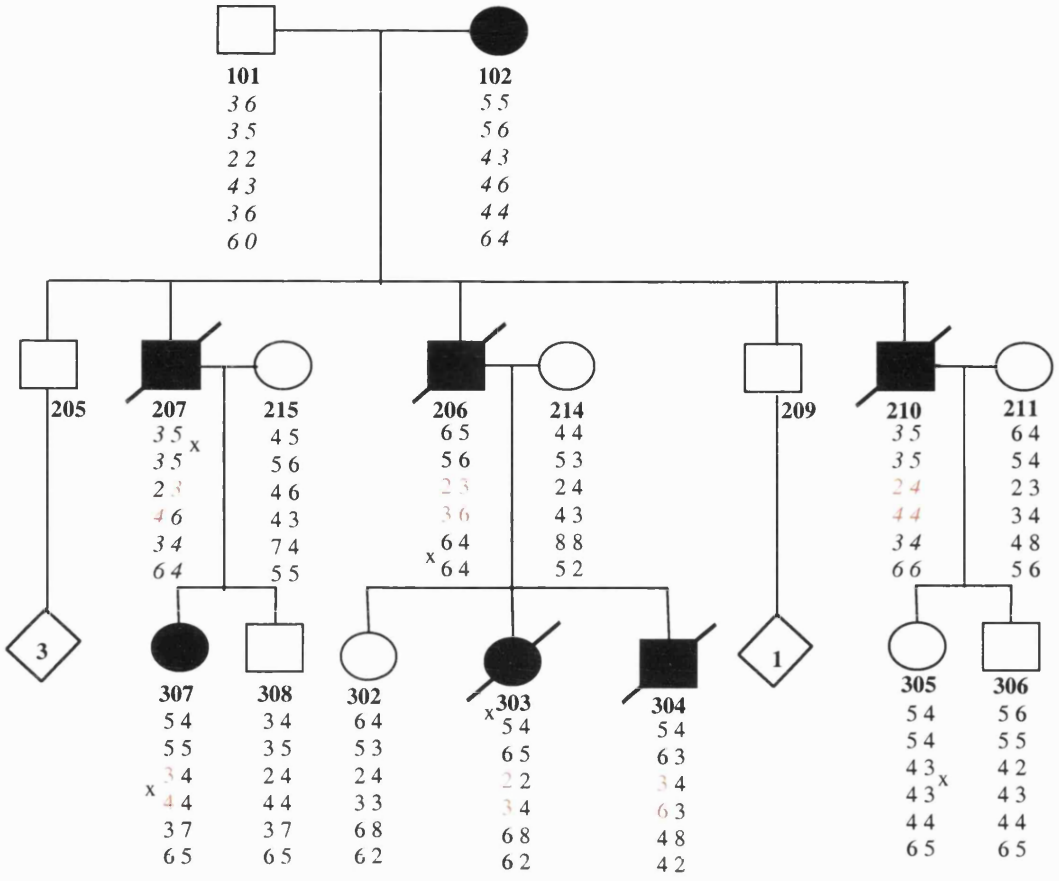


Figure 3.2.1.1. continued. *Family MD was not compatible with linkage to 18q as affected siblings 96 and 203 did not share any alleles at markers flanking SMAD4, and the affected offspring (208 and 55) of person 96 did not share paternal alleles. It is highly likely that Family 6 were also not compatible with linkage to SMAD4 as the two affected offspring of person 302 had most likely inherited the unaffected paternal grandfathers chromosome 18. This was indicated by the presence of the 4-4-2-6-4-5 haplotype in the two siblings (204 and 205) of the affected grandmother (202) and in the affected father (302) which was not transmitted to 401 and 402. Family 12 was not compatible with linkage to SMAD4, as affected siblings 303 and 304 had inherited different 18q alleles from their affected father, and there was a lack of sharing at 18q markers in affected siblings 206, 207 and 210. Figure continued on next page*

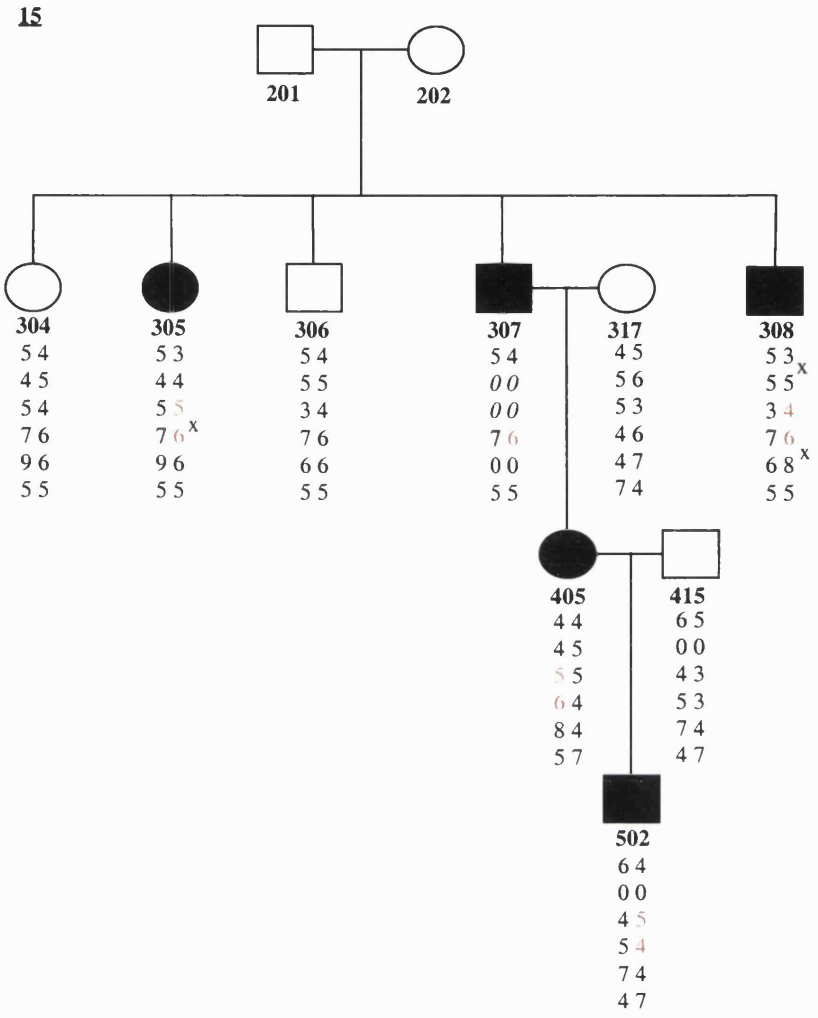


Figure 3.2.1.1. 18q haplotypes in juvenile polyposis syndrome families. *Family 15 was not compatible with linkage to 18q, indicated by the fact that person 502 had inherited his unaffected grandmothers (317) chromosome 18 from his mother (405) and not his affected grandfathers (307) chromosome.*

3.2.2 *SMAD4* MUTATION DETECTION

Patients were screened at different times owing to ongoing recruitment (summarised in Tables 3.2.1 and 3.2.2). Several methods were employed to assess the *SMAD4* mutation status in each patient, either PCR-based assays such as conformation specific gel electrophoresis (CSGE) and single stranded conformational polymorphism (SSCP) analysis, or techniques such as Western blotting and immunohistochemistry which detect the presence of protein. In addition the protein truncation test (PTT) was employed to search for truncating mutations, and Southern blotting was performed to search for large-scale anomalies such as deletions of whole or part of the *SMAD4* gene. A variety of techniques were considered necessary namely because one technique alone is not 100% sensitive and the detection of new JPS genes is rather dependent on the exclusion of *SMAD4* as the causative gene in the remaining patients.

PCR-based germline *SMAD4* mutation detection was performed in collaboration with Richard Houlston at the ICR. PCRs were performed using the primers detailed in Table 3.2.1.1 which cover all exons and exon/intron boundaries of *SMAD4*. PCR fragments were then subjected to CSGE and/or SSCP, and direct sequencing. Of twenty-one patients, only one (AF) was found to harbour a *SMAD4* mutation, a missense arginine to cysteine (cgc-tgc) change at codon 361 in exon 8. Subsequently with the arrival of newly recruited patients, four more *SMAD4* mutations were

identified using CSGE (Richard Houlston at the ICR). Family 17 was reported as having a 2 base pair deletion (but is in fact an insertion of 2 bp, correctly characterised by Ian Frayling, Cambridge) (CC) at nucleotide 1564 of exon 11 which creates a stop codon at nucleotide 1575. The mutation of Family 20 was reported as 189-197del, an in frame deletion of nine bases in exon 1. This mutation has been further characterised (by Ian Frayling for clinical genetics purposes) as a most unusual and complex change, which comprises a net 1bp deletion and 44bp insertion, resulting in a stop at codon 70 (Figure 3.2.2.1); Family BL have a CGA to TGA substitution causing a stop codon in exon 10 of *SMAD4*; and sporadic SV has an 11 base pair deletion at nucleotide 516 of exon 4, creating a stop codon at nucleotide 561) (summarised in Tables 3.2.1 and 3.2.2).

```
173
ACAGCTATAACTACAAATGGAGCTCATCCTAGTAAATGTGTTACCATACAGAGAACA
     T  A  I  T  T  N  G  A  H  P  S  K  C  V  T  I  Q  R  T    77

173
ACAGCTATAACTACTCTCCTAGTTATGTGTTCATAATGGAGCTCATCCTAGTAATGAGCTCATCATT
CATCCTAGTAAATG
     T  A  I  T  T  L  L  V  M  C  S  X
```

Figure 3.2.2.1 Details of sequence change in germline of patient 20.
Wild type cDNA sequence is shown above and mutant, below (based on Genbank U44378). The insertion of is underlined. The sequence in bold shows deletion of the 'A'. The boxed sequences show a region duplicated in the mutant. The flanking sequences of the insertion are shown in italics and have no known similarity to any gene or Alu sequence.

Exon	Sense primer 5'-3'	Antisense primer 5'-3'	Product size (bp)
1	TTGCTTCAGAAATTGGAGACA	GCTTGAAAGGAAACGTAGCAA	385
2	TGTATGACATGGCCAAGTTAG	CAATACTCGGTTTTAGCAGTC	530
3	CTGAATTGAAATGGTTCATGAAC	GCCCCTAACCTCAAAATCTAC	308
4	TTTTGCTGGTAAAGTAGTATGC	CTATGAAAGATAGTACAGTTAC	509
5+6	CATCTTTATAGTTGTGCATTATC	TAATGAAACAAAATCACAGGATG	557
7	TGAAAGTTTTAGCATTAGACAAC	TGTACTCATCTGAGAAGTGAC	224
8	GGATGTTCTTTCCCATTTAT	ACAATCAATACCTTGCTCTC	224
9	TATTAAGCATGCTATACAATCTG	CTTCCACCCAGATTTCAATTC	332
10	AGGCATTGGTTTTTAATGTATG	CTGCTCAAAGAAACTAATCAAC	293
11	CCAAAAGTGTGCAGCTTGTTG	ATTGTATTTTGTAGTCCACC	570

Table 3.2.2.1 SMAD4 primers used for mutation screening (CSGE and SSCP).

The identification of further JPS genes relies on the certainty that *SMAD4* is indeed not the causative gene. Whilst a combination of PCR-based assays should identify the majority of *SMAD4* mutants, it was considered conservative to use other methods to be sure of the lack of *SMAD4* contribution to the remaining patients' disease. To confidently exclude *SMAD4* as the causative JPS gene in the remainder of the cohort (comprising families 1, 3(a.k.a. 1868), 5, 6, 10, 11(a.k.a. FT), 12, 14 ,15, 18, 21, 22, MD, YC, GP, WN, SM397, MTW, SM524, HR, JP2/13 and JP7/1, and sporadics KS, WH, BN, CV, 1204, 1262, DM, SM316 (a.k.a HG), BW, RV, 1469, LB, CR1, FD, JP1/1 and JP8/1) a number of techniques were employed not only to study the gene, but the protein as well. Firstly, fluorescent SSCP analysis was used to screen the *SMAD4* gene in the remainder of the patients using the capillary based 310 prism at 25°C. The primers listed in Table 3.2.1.1 were used for the SSCP, but were labelled with 5' fluorescent tags (FAM, HEX or TET). Any

aberrant band was re-amplified using non-fluorescently tagged primers and subjected to direct sequencing. A new mutation was detected using this method, a G->A change at the first base of intron 2 in Family 21 (Figure 3.2.2.2). This mutation was seen in three affected individuals (two sisters and their maternal aunt, but not the unaffected father) and so was considered to be pathogenic as a result of aberrant splicing of exons 2 and 3. Unfortunately, no RNA material was available to look for novel mRNA species.

The protein truncation test (PTT) was used to identify nonsense, that is truncating, germline mutations potentially missed by CSGE and F-SSCP. PCR was performed on cDNA to provide a transcript using a forward primer tagged with T7 RNA-polymerase and ribosome binding sites, and an in-frame start codon. The resulting mRNAs are then translated *in vitro*, incorporating $\alpha 35^{s}$ labelled methionine, and any mutation which results in a truncation will be visualised as a shorter band than the controls. PTT was performed on 19 JPS individuals from whom cDNA was available (families 5, 6, 19, 22, MD, MTW, 1868 (aka 3), FT (aka 11), GP, WN, JP2/13, JP7/1, HR and sporadic cases CV, JP1/1, JP8/1, 1204, 1262 and 1469), plus controls. PCRs were performed using the iF/iiiR (to amplify codons 1-311) or iF/vR (to amplify codons 182-553) primer pairs shown in Table 3.2.2.2. One patient (MTW) had an extra PTT band with primer pair iF/iiiR, corresponding to a truncated protein (Figure 3.2.2.3). Sequencing of new products of MTW from exons 1-7 revealed a nonsense change Q180X in exon 4 (Figure 3.2.2.2). PTT was found to be

a reliable and useful technique for the identification of germline *SMAD4* mutations which may be missed in PCR-based assays. Although the method is dependent on having RNA available for every patient to be assessed, it does not frequently give false negatives.

Primer	Position	sense primer 5'-3'	Primer	Position	antisense primer 5'-3'	Product size
if	1	atggacaatatgtctattacga	iR	317	ttgtgaagatcaggccacct	316
ii F	256	ggtcggaaaggatttcctca	iiR	601	acagagctggggtgctgtat	345
iii F	547	cagcatccaccaagtaatcg	iiiR	931	ggaatgcaagctcattgtga	384
ivf	895	ggacattactggcctgttca	ivR	1260	acgcccagcttctctgtcta	365
vF	1207	agtgaccacgcggtctttg	vR	1659	aaggttgtgggtctgcaatc	452

Table 3.2.2.2. Primers for SMAD4 cDNA used for the protein truncation test and to prepare probes for Southern blotting.

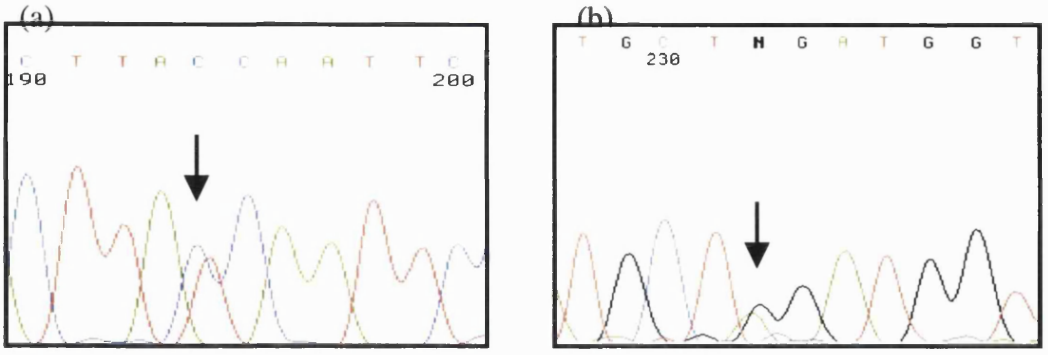


Figure 3.2.2.2. Sequence changes in patients (a) 21 and (b) MTW.
(a) SMAD4 exon 2 reverse sequence of Family 21 is shown with the +1 splice donor intron 2 c->t (g->a in forward) change arrowed. (b) SMAD4 exon 4 reverse sequence of MTW is shown with the change arrowed, g->a (c->t in forward).

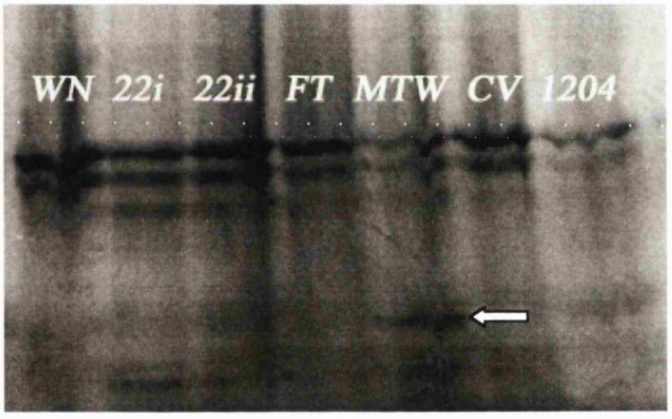


Figure 3.2.2.3. Protein truncation test results.
Shown are the PTT results using primers iF and iiiR (Table 3.2.2.2) covering exons 1-7 from 7 patients. The truncated protein in patient MTW is shown by an arrow.

Western blotting is a technique used to detect truncating mutations of various types, and may be useful as an additional mutation detection technique. Western blotting was performed using the anti-SMAD4 mouse monoclonal antibody B8 which recognises an epitope in exon 5 of *SMAD4* (64KDa)(The epitope maps to codons 68-108, M. Howell and C. Hill, personal communication). The blots were then incubated with an HRP-(horse radish peroxidase) conjugated secondary antibody and protein levels detected using enhanced chemiluminescence (ECL) reagents. To assess SMAD4 protein levels, two control antibodies were also used, one anti-MLH1 (92KDa SIZE)(Santa Cruz) and the anti-β-actin mouse monoclonal antibody (42KDa)(Sigma). Protein pellets derived from lymphoblastoid cell lines were available from thirteen patients (Families 17, 19, 22, MD, FT, WN, MTW, HR, JP2 and JP7, and sporadics CV, JP1 and JP8). No truncated bands were observed. Two of these families, MTW and 17, have germline *SMAD4* mutations identified by PTT and CSGE respectively. The mutation of MTW is in exon 4 which is N-terminal to the epitope in exon 5, and therefore the truncated protein would not be detectable by Western blotting.The mutation of family 17 is a 2 base pair insertion in exon 11 resulting in a truncated protein. As this was a previously known mutation, the expected result was a protein band of normal size (64Kda) plus a band corresponding to the truncated protein, given these were cell lines derived from lymphocytes and not polyp tissue. Instead, no bands at all for SMAD4 were observed in patient 17, not even the wild-type (Figure 3.2.2.4). This was unexpected and difficult to explain. Protein was clearly present with antibodies recognising the

control proteins MLH1 and beta-actin. Further work on colorectal cancer cell lines (discussed in the Chapter Four) showed that proteins derived from mutated *SMAD4* genes are unstable and degraded. This is upheld by previous work which has shown that SMAD4 proteins with N-terminal mutations are rapidly degraded via a ubiquitin-mediated pathway (Moren *et al.*, 2000; Xu and Attisano, 2000). This may explain why a band representing the mutated protein was not observed, but does not explain why the normal band was not seen in patient 17. It is well documented that SMAD4 forms complexes with SMADs 2 and 3, and forms SMAD4 homodimers (Kawabata *et al.*, 1998). One explanation for not observing the wild-type 64Kda SMAD4 band in patient 17 is that somehow the mutant protein binds to the wild-type protein, and these homodimers are degraded in much the same way as mutant proteins alone. RNA stability of *SMAD4* in patient 17 was established by amplifying across *SMAD4* exon 11 with fluorescently labelled primers and running on a genescan gel, where the two RNA species (the mutated band two base pairs longer than the wild-type) were clearly present. Alternative explanations for the complete absence of protein on the Western blot, such as the existence of a second, undetected mutation of *SMAD4* in this patient seem unlikely given the rarity of germline mutations in this gene. The affected brother of this patient and affected members of Family 20 (who also carry a germline *SMAD4* mutation) have been re-bled, the lymphocytes separated and sent off for transformation. It is hoped this will give insight into whether the phenomenon of no SMAD4 protein is unique to this member of Family 17, or whether it is common to other germline mutation carriers.

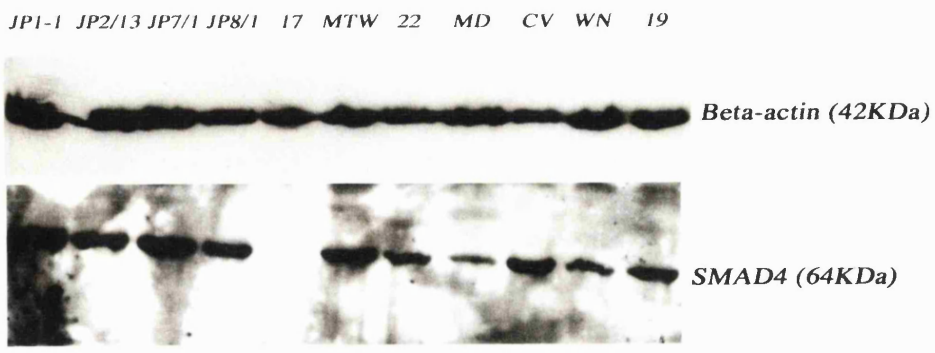


Figure 3.2.2.4.* SMAD4 *Western blot analysis.
Shown are results from 11 JPS patients and families using the anti-SMAD4 B8 antibody, and an anti-β-actin antibody exposed to the same blot, using protein derived from lymphoblastoid cell lines. Family 17 has a 2bp insertion in exon 11 of SMAD4, but no SMAD4 protein detectable (despite six repeats), but clearly has protein present for the control antibody.

Overall, Western blotting was therefore not found to be a particularly useful technique for detecting germline *SMAD4* mutations. Firstly, the B8 antibody recognises an epitope in exon 5 so any mutation occurring N-terminal to this will not be detected (as in MTW). Antibodies directed to the N-terminus are therefore much more useful for Western blotting, and whilst there is one available for SMAD4 (N16, Santa Cruz) the results obtained were messy and feint and therefore difficult to interpret. In addition, work with germline *SMAD4* mutation carriers such as family 17, with colorectal cancers (discussed in Chapter Four) and work published

by others (Moren *et al.*, 2000; Xu and Attisano, 2000), has shown that mutations that lead to a truncated protein leave the protein unstable and liable to proteosomal breakdown, and therefore not detectable by Western blotting. Additionally, the cell numbers loaded into each lane were approximately 5 $x10^7$, and it was expected that a heterozygous *SMAD4* mutation patient would give half the level of signal compared to homozygous wild-type patients. Comparison of the SMAD4 levels with the control proteins was performed using densitometry after scanning the gel images (Biorad GS-700 densitometer). This method, however, failed to detect a significant reduction of *SMAD4* copy number in the patient with a known *SMAD4* mutation (MTW) and was thus not a reliable indicator of the presence of a germline mutation. Other methods for detecting truncations such as the protein truncation test are therefore more robust and likely to detect mutations, without false negatives.

Whilst mutation screening techniques such as SSCP or PTT are useful for detecting small genetic changes such as point mutations and minor insertions and deletions, they are not able to detect large germline changes which may be pathogenic, such as deletions of whole or part of a gene. In addition, whilst PTT may provide a definitive mutation result, it is dependent on having RNA material available, which for most families was not the case. Southern blotting was therefore performed on 24 individuals from whom sufficient DNA was available, in order to detect large germline deletions (Families: 1, 5, 6, 10, 12, 14, 16, 17, 19, 21, 22, MD, FT, DM, HR, MTW, SM524, and WN; sporadics: BN, CV, KS, SM106, HG (aka SM316)

and 1469,). Five overlapping cDNA fragments were amplified using the primers detailed in Table 3.2.1.2. 10μg DNA was digested using the restriction enzymes *Hind*III, *Eco*RV (both four base cutters) and *Sau*3A1 (six base cutter). Only one aberrant band was observed in one individual, from Family 19, under *Hind*III digestion using probe IV (Figure 3.2.2.5). This change was not observed with any other restriction endonuclease or in the patient's affected brother and is therefore most unlikely to be pathogenic, but may be a polymorphism changing a restriction site. Southern blotting indicated that large deletions of part or all of the *SMAD4* gene are a not a likely cause of JPS.

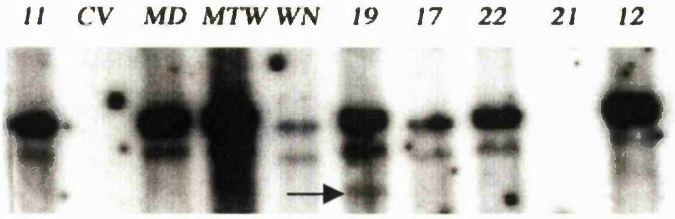


Figure 3.2.2.5. SMAD4 Southern blot in JPS patients.
Shown are the Southern blot results for SMAD4 probe III-V, hybridised to HindIV digested genomic DNA. Insufficient DNA was loaded for CV and patient 21. Family 19 showed an extra band (arrowed), but this band was not observed in another affected member of this family. No aberrant bands were observed with DNA digested with the EcoRV or Sau3A1, and therefore the extra band most likely represents a polymorphism changing a HindIV site.

3.2.3. SMAD4 IMMUNOHISTOCHEMISTRY

Immunohistochemistry was used as a means of identifying patients with germline *SMAD4* mutations on the basis that the gene is a tumour suppressor and that regardless of the type of second hit, be it loss of heterozygosity or a point mutation in the remaining wild-type allele, it should be detected via loss of the protein in JPS polyps and cancers. In addition, staining for the presence of protein should indicate the stage at which the loss occurred in the growth of the polyp, whether it be initiating the growth of the polyp or loss at the transition from polyp to a more aggressive phenotype. Using the B8 anti- SMAD4 antibody, detection levels have been shown to accurately mirror mutation status in pancreatic carcinomas with 91%

sensitivity and 94% specificity (Wilentz *et al.*, 2000). Immunohistochemistry was performed (with the assistance of Histopathology department, ICRF) on 5μm sections from all polyp and cancer tissue available using the B8 nuclear staining antibody at 1/100 dilution, after baking of the sections for 20 minutes. After counterstaining with haemotoxlyin, the slides were examined for SMAD4 expression, with scoring simply as absent or present. A total of 102 polyps and 10 cancers (from families 17, 20, 21, AF (all four with germline *SMAD4* mutations), and families MD, 6, 12, 15 and sporadics LB and WH (all without *SMAD4* mutations) were assessed for SMAD4 expression using immunohistochemistry with the B8 antibody. The results of the immunohistochemistry are summarised in Table 3.2.3.1. In total, 37/38 juvenile polyps and 8/9 cancers from 6 *SMAD4*-wild type families were positive for B8 staining, reflecting retention of SMAD4 expression. In stark contrast, only 1/64 polyps and 0/1 cancers from 4 *SMAD4*-mutant families were positive for B8, reflecting loss of SMAD4 expression in the great majority of tumours (Figure 3.2.3.1). In addition to showing that there was loss of the second copy of SMAD4 in the polyps, the immunohistochemistry also indicated that this loss probably initiated the growth of the polyp as there was a distinct border between the normal tissue expressing SMAD4 (stained brown) and the base of the polyp which did not express SMAD4 (stained blue) (Figure 3.2.3.1). The immunohistochemistry on polyps from Family 21 was performed after CSGE had failed to detect this family's germline *SMAD4* mutation, and thus the lack of SMAD4 expression in these polyps was a rather confusing result. It was only upon SSCP that the germline mutation in Family 21 was discovered, providing a full

explanation of why there was no SMAD4 protein detectable in the polyps, and indicating that the immunohistochemistry is a good marker of a germline mutation.

Family/ID	*Mutation (nucleotide no.)*	*Predicted effect*	*Method*	*SMAD4 expression*
*20**	*189-197del1ins44***	*stop codon 70*	*CSGE*	*0/16 polyps*
*17**	*1564-1565del*	*stop codon 525*	*CSGE*	*0/37 polyps*
SV	*516-527del*	*stop codon 187*	*CSGE*	-
BL	*c->a 1333*	*R445X*	*CSGE*	-
AF	*c->t 1083*	*R361C*	*CSGE*	*1/6 polyps , 0/1 cancer*
*21**	*+1 splice donor intron 2 g->a*	*Abrogation of splicing*	*F-SSCP*	*0/5 polyps*
*MTW**	*c->t 541*	*Q180X*	*PTT*	-
*MD**	*No*			*3/3 polyps*
LB	*No*			*7/8 polyps*
*12**	*No*			*3/3 polyps, 6/7 cancers*
*15**	*No*			*19/19 polyps*
*6**	*No*			*2/2 polyps 2/2 cancers*
WH	*No*			*3/3 polyps*

Table 3.2.3.1: Summary of germline SMAD4 mutations and B8 immunohistochemistry.
**= familial case. **= mutation previously reported as 189-197 deletion only. Those families not shown had both no mutation detected and no tumours analysed by immunohistochemistry. - = not done. Association between loss of SMAD4 expression and SMAD4 mutation is highly significant (Fisher's exact test, p~0.0)*

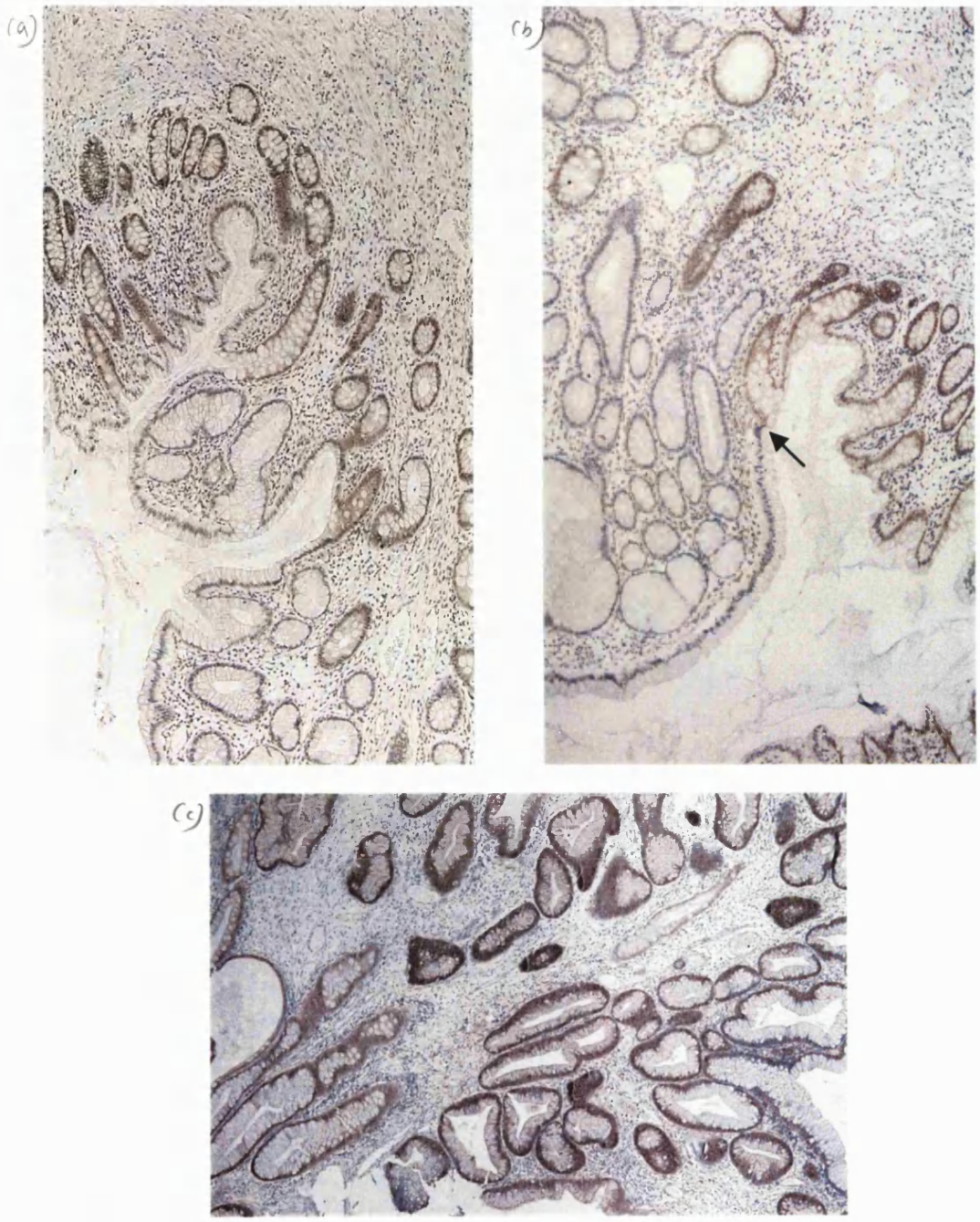


Figure 3.2.3.1 B8 immunohistochemistry.
(a) polyp (x20) from member of family 20 (with a SMAD4 germline mutation) showing no SMAD4 protein even in the smallest polyps (b) polyp (x5) from family 20 showing distinct border between where SMAD4 is expressed (brown) and where loss of expression begins (blue) (arrowed). (c) juvenile polyp (x5) from family 12 who do not have SMAD4 germline mutation showing strong SMAD4 expression.

Overall, there was excellent concordance between the presence of a germline mutation and the immunohistochemistry, indicating that the mutation screening had detected all the *SMAD4* mutations (confirmed in at least the people from who there was material available for immunohistochemistry). The results strongly suggest that disease in families without *SMAD4* mutations develops along a *SMAD4*-independent pathway, whereas the families who have a *SMAD4* germline mutation have lost the second copy of *SMAD4*, leading to growth of the polyp. These data corroborate results showing that *SMAD4* acts as a tumour suppressor gene in JPS (discussed in Chapter Five) and confirm that even missense changes (as in AF) are associated with loss of protein expression.

3.3 MORPHOLOGICAL REVIEW OF JPS POLYPS

In order to try and segregate the polyps by morphology according to *SMAD4* mutation status (as established by CSGE, linkage analysis, SSCP, Southern blotting, PTT and immunohistochemistry), a blinded analysis of haemotoxylin and eosin stained sections from all available polyps was performed by histopathologist Professor Nick Wright. The morphological review was performed with no prior knowledge of the *SMAD4* mutation status of the material. The slides were scored for several categories including (1) whether they resembled the 'classic' juvenile polyp, particularly the predominance of epithelium or stroma, (2) the amount of inflammation, (3) whether dysplastic features (for example, adenomatous regions) or

hyperplastic features were present in any region, (4) site of the polyp and (5) any extra features such as colitis or cryptitis. A total of 113 H&E-stained polyp sections were reviewed to look for potential differences between polyps derived from patients who possess a germline *SMAD4* mutation and polyps from patients who do not harbour *SMAD4* mutations.

A summary of the findings is shown in Table 3.3.1. Polyps from patients without *SMAD4* mutations were generally of the 'classical' morphology, with expanded cysts, predominant stroma and large numbers of inflammatory cells. Although many polyps from *SMAD4* mutation carriers had features of juvenile polyps, that is, expanded cysts and high levels of inflammation, polyps from mutation carriers were much more epithelial/non-classical, with many, long elongated crypts replacing the round cysts (Figure 3.3.1). Polyps from both mutation carriers and non-mutation carriers had similar frequencies of hyperplasia/dysplasia (Table 3.3.1). The overall 'epithelial content' was far more pronounced in the polyps of *SMAD4* mutation carriers than those without mutations, and consequently the number of classical juvenile polyps was significantly lower in the mutation carriers than in the non-mutation carriers (19/62 polyps were of classical morphology in the *SMAD4*-mutation carriers; 47/51 polyps were of classical morphology in the patients without *SMAD4* mutations (Fisher's exact test, $p<1x10^{-10}$)). The size of polyps was significantly greater in *SMAD4* mutation carriers versus non-carriers (mean 15.96mm versus 9.83mm, $t=4.98$, $v=102$, $p<0.001$).

Patient	SMAD4 mutation?	Classical JP polyps	Hyperplastic/dysplastic/ adenomatous areas	Non-classical JP polyps	Hyperplastic/dysplastic/ adenomatous areas	Notes
AF	yes	3/6	All 3 with dysplasia and hyperplasia	3/6	2 with hyperplasia, 1 without	Cryptitis in 3. All very epithelial.
20	yes	5/14	All 5 with areas of hyperplasia	9/14	All complex. 7/9 very dysplastic and/or adenomatous, 4/9 hyperplastic (inc. 2 which were not dysplastic)	Very elongated, dense crypts. Larger polyps very epithelial. Smooth muscle in 7.
17	yes	6/37	All 6 with hyperplasia	31/37	All very hyperplastic. 12/31 with dysplasia	Very elongated, dense crypts. Larger polyps very epithelial
21	yes	5/5	4/5 classical with hyperplasia; 1 with dysplasia and adenomatous region.	0/5	N/A	Small.
MD	no	3/3	2/3 had region of hyperplasia, one of these with small adenomatous region	0/3	N/A	Very inflamed, prominent stroma
LB	no	6/8	5/6 with region of hyperplasia, 2/6 with region of dysplasia	2/8	Note : both small bowel. 2/2 with region of hyperplasia	Very inflamed, prominent stroma, large cysts. Smooth muscle in 4.
WN	no	6/6	2 with hyperplasia and dysplasia, 3 with hyperplasia	0/6	N/A	Very inflamed, prominent stroma, large cysts. Granuloma in 2
12	no	10/10	Some regions hyperplastic-like	0/10	N/A	Very inflamed, prominent stroma, large cysts.
15	no	17/19	6/10 with dysplasia. 11 with region of hyperplasia, 2 with dysplasia	2/19	Very small polyps	Very inflamed, prominent stroma with large cysts.
6	no	2/2	No hyperplasia	0/2	N/A	Very inflamed, prominent stroma, large cysts
WH	no	3/3	No hyperplasia	0/3	N/A	Very inflamed, prominent stroma, large cysts

Table 3.3.1. Summary of morphology results.

All polyps were from colorectum unless stated otherwise. Polyps were categorised into 'classical' or 'non-classical' according to whether they fit the general description normally given for juvenile polyps (hypercellular stroma, expanded mucin-filled cysts, rounded edge etc.) Polyps of SMAD4 mutation carriers were much more epithelialthan those without mutations, and consequently the number of classical juvenile polyps was significantly lower in the mutation carriers than in the non-mutation carriers (Fisher's exact test, $p < 1x10^{-10}$).

There was only a borderline difference between the frequency of hyperplasia/dysplasia in polyps from mutation carriers and non-carriers (Fisher's exact test, p=0.06), although the spatial extent of hyperplasia and dysplasia appeared to be greater in *SMAD4* mutation carriers. The occurrence of gastrointestinal malignancy was common to all families, with the exception of LB who is less than ten years old so unlikely to have developed cancer. This suggests that although polyps from *SMAD4* mutation carriers appeared to be more aggressive than polyps from non-mutation carriers, that is rather more adenomatous and therefore pre-cancerous, essentially the end-result is the same i.e. morbidity due to cancer. It is therefore likely that JPS patients whose disease is not as a result of a *SMAD4* germline mutation almost certainly have an alternative tumour suppressor inactivated in the germline. *SMAD4*-independent polyps perhaps then evolve more rapidly into carcinoma, explaining the apparent shortage of adenoma-type lesions observed in *SMAD4*-independent patients. It was evident that there are morphological differences between polyps arising as a result of *SMAD4* loss, and those arising via a *SMAD4*-independent pathway, making it possible to segregate tumours according to *SMAD4* mutation status, although a single polyp from a given family would not be sufficient to reliably predict the presence of a germline *SMAD4* mutation..

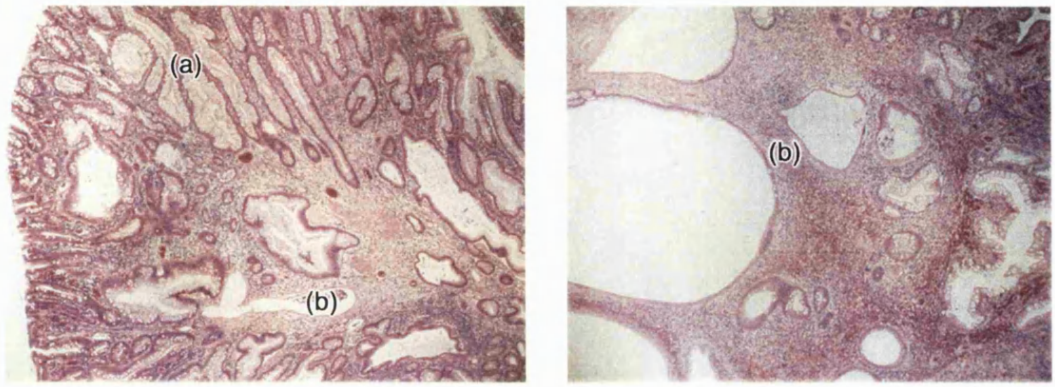


Figure 3.3.1. Haemotoxylin and eosin stained slides of Juvenile Polyps.
The left hand panel shows a juvenile polyp (x2.5) from a SMAD4 mutation carrier (Family 20). Note areas which look (a) hyperplastic and (b) areas of classical juvenile polyp morphology, with expanded cysts and normal epithelium. The right-hand panel shows a 'classical' juvenile polyp (x2.5) from a SMAD4 mutation-negative patient (Family MD) with morphology of type (b).

3.4 CLINICAL FEATURES OF SMAD4 MUTATION CARRIERS VERSUS NON-MUTATION CARRIERS

With apparent morphological differences between the polyps of *SMAD4* mutation carriers versus non-mutation carriers, the clinical features of these two groups were compared to look for similarities and differences between the two, other than the obvious existence of juvenile polyps in both. The reported clinical features of *SMAD4*-mutation carriers are summarised in Table 3.4.1, and the reported features of non-mutation carriers are summarised in Table 3.4.2. Obviously any noted differences between the two groups are inferred from the reported clinical features, and therefore may not be accurate if different manifestations are reported inconsistently. As mentioned above, the *SMAD4*-mutation carriers appeared to have a higher incidence of gastric polyps and gastric cancer, although the gastric involvement is not confined to the *SMAD4*-mutation carriers alone. This difference between the two groups does attain statistical significance with 11/18 (61%) *SMAD4* mutation carriers having stomach involvement *versus* 9/47 (19%) of the *SMAD4* independent patients having stomach involvement (Fishers exact test $p<0.0002$). There is no preponderance of colorectal cancer in the *SMAD4* mutation carriers, as might be expected from the more epithelial nature of their polyps, indicating that *SMAD4*-independent JPS cases are also as a result of a germline mutation in a tumour suppressor gene. No patients in either group had dermatological or skeletal phenotypes indicative of CS, BZS or GS, corroborated by the lack of mutations in the genes which cause these syndromes, *PTEN* (Marsh *et al.*, 1997b) and

PTCH (discussed in Chapter Six) respectively. There was some overlap in phenotype with CS of the non-*SMAD4* patients, such as the existence of seven patients with macrocephaly (15%), but no patient had a past history of thyroid cancer and only one had developed breast cancer (from Family 1). Furthermore, the increased risk of malignancy appeared to be confined to the gastrointestinal tract, in contradistinction to CS and GS. One patient with a *SMAD4* mutation also had macrocephaly (1/15, 7%).

Family /case	*GI malignancy*	*Skin*	*Skeletal*	*CNS*	*Cardiac*	*Breast*	*Thyroid*	*Other*
*1 13 **	*CRC*							*Pancreatic ca.*
*M-1**								
*JP 5/1**								
JP11/1								
JP 10/1								
AF*	***CRC***							
FAM1, aka 4/1*								*Pancreatic ca.*
*FAM3**	*CRC*							*Ad*
S2, aka 6/1								*PH*
17	***CRC***							***MIP***
20	***CRC***							***OVC, MIP***
BL SV	***CRC***							
21 MTW	***CRC***		***MC***	***DD***				***OVC T***

Table 3.4.1 Summary of extra-gastrointestinal features of SMAD4 mutation carriers.

*Bold type shows the ICRF patients, *= familial, **= overlap of patients between Howe et al and Roth et al, ***= patients reported more than once, but only counted once. Ad, adenoma; AR, aortic regurgitation; AS, aortic stenosis; BBD, benign breast disease; BK, bifid kidney; C, clubbing of hands; Ca, Cancer; CD, conduction defect; CLP, cleft palate; CL, clinodactylyl; CO, coarctation; CRC, colorectal cancer; DD, Developmental delay; FK, excessive freckling; HC; hydrocephaly; HT, hypertelorism; Hyper, hyperthyroidism; Hypo, hypothyroidism; MC, marcocephaly ; MIP, muscle in polyp; ML, multiple lipomas; OVC, ovarian cyst; P, poryphria; PC, Polycythemia; PD, polydactylyl; PH, panhypopituitarism; PN, palmal nodules; SAH, subarchnoid haemorrhage; Sb, small bowel; SR, Schatski ring; St, stomach; T, telangiectasia; TA, tubularadenoma; VSD, ventricular septal defect; WT, Wilm's tumour*

Family/case	GI malignancy	Skin	Skeletal	CNS	Cardiac	Breast	Thyroid	Other
JP 1/1, aka SP1								WT
JP 2/13*, aka fam4	CRC							
FAM2*, aka 7/1 sp3	CRC				AS, VSD			Ad Oslers
1*	**Sb**					**Ca, age 54**		**Ad**
3*	**Sb, St**	**ML**						
5*	**CRC**							**HT**
6*	**CRC**		**MC**					
8*	**CRC, St**							
9*	**CRC**				**AS, VO**			**Ad, BK**
10*	CRC							
FT*/11	**CRC**	**FK**						**HT, SR**
12*	**Sb, CRC**		**MC**	**SAH**				**OVC**
14*	**CRC**							
16*	CRC					**BBD**		
SM397*								
SM524*			**C**					**HT**
SM106					**VSD**			**HT**
1204								
1469								
c2	**N/A**	**ML**	**MC**					
1262		**PN**	**B, C**					
KS								**Ad**
15*	**CRC**		**CL**		**AR**		**Hypo, hyper**	**OVC**
MD	**CRC**		**MC, C**					**HT, T**
WN*	**CRC**					**BBD**	**Hyper**	
SR					**VSD**			
SS								
SH			**CLP, PD**					
SCA								
SC			**MC**					
SD					**CD**			
RV								**P**
18*	CRC			DD				
19*								
22*			**MC**					
YC*								
WH	**CRC**							
LB	**N/A**		**HC, C**	**DD**				**MIP**
BN								
CV								
DM	**CRC**					**BrCa, aged 54**		
HR	**CRC**							
CWN								
SM316			**MC**					**HT**

Table 3.4.2 Summary of extra-gastrointestinal features of SMAD4 negative patients.

*Bold type shows the ICRF patients, *= familial, **= overlap of patients between Howe et al and Roth et al. Abbreviations same as used for Table 3.4.1*

Arteriovenous and other cardiovascular anomalies were present in 6 patients (13%) in the *SMAD4*-negative group, but none of the *SMAD4* mutation carriers suggesting the association of heart defects with JPS is genuine and is independent of *SMAD4*. Hypertelorism and clubbing of the hands were another two clinical features which were confined to the *SMAD4*-independent patients, present in 5/47 (10.6%) and 4/47 (8.5%) respectively, but none of the germline *SMAD4* mutation carriers. Ovarian cysts were present in two *SMAD4*-mutation carriers and two SMAD4-independent patients (13% and 0.04%). Finally, as may be expected from a gene commonly inactivated in pancreatic cancer, two *SMAD4* mutation carriers had developed pancreatic cancers, whereas none of the 47 *SMAD4*-independent cases had reported pancreatic cancer. This was the only clinical feature between the two groups that was of borderline statistical significance ($P<0.055$).

3.5 CONCLUSIONS

Germline *SMAD4* mutations undoubtedly account for a minority of JPS cases. Most germline mutations are detectable by F-SSCP analysis or CSGE, and sequencing. After an initial screen with CSGE detected 5 mutations, F-SSCP detected one extra change, a novel splice site mutation at the +1 donor site of intron 2 in Family 21. One additional germline *SMAD4* mutation, was found using PTT in patient MTW. Southern analysis detected no large-scale mutations and Western analysis found no evidence of truncated proteins, although this was

not a reliable indicator alone of the absence of a mutation. Immunohistochemistry results suggest that cryptic *SMAD4* mutations were very rare and that the majority of *SMAD4* mutations have been identified with the comprehensive screening techniques. Just one of thirty-eight polyps from patients without a germline *SMAD4* mutations showed loss of protein expression, confirming the results of the mutation detection and showing that these tumours grow along a genetic pathway that does not involve *SMAD4*, at least in the early stages. By contrast, almost all polyps and cancers from our known *SMAD4* mutation carriers had absent protein expression. It appeared, therefore, that if the wild-type *SMAD4* allele is generally deleted as the 'second hit' which initiates the growth of JPS polyps, the remaining mutant protein is unstable. Whilst not unexpected for truncated proteins, it appears that even if the germline change is of a missense type, as in Family AF, and indicated by the Western analysis, protein instability generally results. This is upheld by data showing that missense mutations in the N-terminal MH1 region of *SMAD4* cause rapid degradation of the protein *in vitro* (Moren *et al.*, 2000; Xu and Attisano, 2000). The missense R361C mutation of family AF maps to the loop/helix domain in the C-terminal of *SMAD4* and has also been found in a sporadic colorectal cancer (Thiagalingam *et al.*, 1996). The functional effects of R361C have been well evaluated (Shi *et al.*, 1997) and it prevents both hetero- and homo-oligomerisation of *SMAD4*. The immunohistochemistry results for family AF also suggest that SMAD4 protein which is not bound into a complex is degraded or unstable *in vivo*.

Previous studies have found germline *SMAD4* mutations in about 25-60% of JPS cases, but one common mutation (4bp deletion, codons 414-416, stop at codon 434) accounted for many patients in some studies (Table 3.2.1). Howe *et al* used SSCP analysis and sequencing to find mutations in 5 of 9 patients studied (Howe *et al.*, 1998b). All of these were frameshift changes, including 3 examples of the 4bp deletion and two other mutations producing stop codons at 235 and 350. Friedl *et al* used direct sequencing in 11 cases to detect the common 4bp deletion in two patients and a codon 277 frameshift in one another (Friedl *et al.*, 1999). Roth *et al* used direct sequencing in 7 JPS cases to find one missense change (codon 353), one nonsense mutation (codon 177) and one patient with the common 4bp deletion (Roth *et al.*, 1999). Kim *et al* found 3 *SMAD4* mutations in 5 patients using SSCP analysis, comprising a nonsense change at codon 388 and two missense changes at codons 390 and 361 (the same codon as family AF in our cohort) (Kim *et al.*, 2000). We ourselves have found 7 mutations in 49 cases (summarised in Table 3.2.1 and 3.2.2). Thus, germline *SMAD4* mutations account for 21 out of 78 juvenile polyposis patients tested (27%) and appear to occur most commonly – but not exclusively - after codon 200, affecting the C-terminal of the gene which is involved in trimerisation of the *SMAD4* protein. Nonsense and frameshift changes predominate, but pathogenic missense mutations and splice variants can occur. Germline mutations in *SMAD4* also appear to confer an increased risk of gastric polyps and/or malignancy rather than confinement to the colorectum, in accord, perhaps, with the high frequency of *SMAD4* loss in sporadic pancreatic cancer. These data from JPS are consistent with the spectrum of somatic mutations found in colorectal and pancreatic

cancers, with the exception of the higher frequency of homozygous deletions found in the sporadic tumours (Hahn *et al.*, 1996b).

The morphology analysis has shown that polyps appear to be different in patients with and without germline *SMAD4* mutations, although no direct phenotype-genotype correlations within the *SMAD4*-mutation carriers (e.g. position and type of mutation) are apparent. *SMAD4* mutation carriers' polyps had less prominent stroma and a richer epithelial component than the 'classical' juvenile polyps of those patients without *SMAD4* mutations. Polyp morphology is, however, variable within the same individual and between patients from the same family, so that it cannot be used reliably for any one polyp as an indicator of the likelihood of a germline *SMAD4* mutation. Using immunohistochemistry to complement mutation screening and linkage analysis for confident exclusion of *SMAD4* as the causative gene ensures that the *SMAD4* mutation-negative cohort is as homogenous as possible. Without these 'false negatives' the identification of new JPS genes is, in theory, facilitated because the inclusion of families linked to *SMAD4* undoubtedly confounds the detection of linkage elsewhere. However, this is more applicable if a single gene is responsible for the remainder of the JPS patients which is probably not the case (see further chapters). A combination of mutation screening, immunohistochemistry and morphological assessment is reliable for identifying those families whose disease is clearly attributable to germline mutations in *SMAD4*, with clear implications for the testing of at-risk family members.

CHAPTER FOUR

THE TIMING AND FREQUENCY OF *SMAD4* MUTATIONS IN COLORECTAL CANCER

THE TIMING AND FREQUENCY OF *SMAD4* MUTATIONS IN COLORECTAL CANCER

4.1 INTRODUCTION

Allele loss at 18q21.1 is well established as a common step in the classical colorectal cancer progression pathway (Fearon and Vogelstein, 1990; Vogelstein *et al.*, 1988) (discussed in Chapter One) and as such has been demonstrated in up to 60% of colorectal cancers (CRCs) (Thiagalingam *et al.*, 1996). Mapping to this chromosomal band are *DCC* (Deleted in Colon Carcinoma), *DPC4/SMAD4* (Deleted in Pancreatic Cancer 4) and *SMAD2*. For many years the allele loss observed around 18q21.1 in colorectal cancer was believed to be targeting *DCC* (Fearon and Vogelstein, 1990; Hamilton, 1992). The gene has been described as a cell surface receptor for the ligand netrin, and has been implicated in axon guidance during the development of the nervous system (Kolodziej, 1997) and, more easily related to a role in tumorigenesis, as an inducer of apoptosis (Mehlen *et al.*, 1998). Cho *et al* first described the existence of mutations of the *DCC* gene in two colorectal cancers, but these mutations were of questionable significance (one intronic variant and one missense variant, both possible polymorphisms) (Cho *et al.*, 1994). Doubt over *DCCs* function was also raised when the *dcc* knockout mouse failed to develop intestinal tumours or other gastrointestinal phenotype (Fazeli *et al.*, 1997). Mutation screening of *DCC* gene is not often undertaken due to the very large size (29 exons

spanning more than 1300Kb), and therefore establishment of its involvement has mostly relied upon expression studies using Western blotting and immunohistochemistry. This has led to opposing results, with some authors finding that loss of the DCC protein is associated with a poor clinical outcome (Chen *et al.*,1999, Shibata *et al.*, 1996), but other authors (Goi *et al.*, 1998; Gotley *et al.*, 1996) failing to substantiate a major role for DCC in colorectal tumorigenesis. The role of the *DCC* gene in colorectal cancer is therefore very much under a cloud given these contradictory and unclear results. What is almost certain however is that the observed loss at 18q21.1 in colorectal cancer is not targeted to *DCC* the majority of the time. As discussed in detail in Chapter Six, the role of *SMAD2* in colorectal cancer tumorigenesis is probably real but very minor, with reports of infrequent, but functional, mutations of this gene (Eppert *et al.*, 1996; Prunier *et al.*, 2001; Xu and Attisano, 2000). Homozygous loss and inactivation of *SMAD4* were first identified in pancreatic cancer (Hahn *et al.*, 1996c) and it is now well established as a critical gene in the development of pancreatic cancer. In addition, by contrast to *DCC*, it is becoming apparent that changes in *SMAD4* are much more common in the pathogenesis and evolution of colorectal cancer (Miyaki *et al.*, 1999; Tagaki *et al.*, 1996), as well as in Juvenile Polyposis Syndrome in which colorectal cancer is a prominent feature (Houlston *et al.*, 1998; Howe *et al.*, 1998b). In particular, inactivation of *SMAD4* has been associated with late stage or metastatic colorectal cancer (Koyama *et al.*, 1999; Maitra *et al.*, 2000; Miyaki *et al.*, 1999). As a further confirmation of the role of *SMAD4* as a tumour suppressor, the transgenic mouse model for *Smad4* have been found to develop multiple polyps and gastrointestinal

malignancies (Takaku *et al.*, 1999), and compound *Apc/Smad4* knockout mice have a more aggressive malignant phenotype than the *Apc* mouse alone (Takaku *et al.*, 1998).

The *SMAD4* gene codes for a protein involved as a downstream regulator in the transforming growth factor beta-signal transduction pathway, facilitating the transcription of target genes (the pathway is discussed in detail in Chapter Six). These target genes include cyclin-dependent kinase inhibitors such as $p15^{(ink4B)}$ (Feng *et al.*, 2000) and the inhibitory *SMAD7* (von Gersdorff *et al.*, 2000), and there are doubtless many as yet unidentified genes which are under SMAD4 transcriptional control. It is possible to envisage how abrogation of SMAD4 function causes a breakdown in the signalling pathway, by loss of transcription of genes critical to cell-cycle control such as $p15^{(ink4B)}$ (de Caestecker *et al.*, 2000). Indeed, targeted deletion of *SMAD4* in experiments by Zhou *et al* resulted in loss of TGFβ and activin (where SMAD4 is also the common mediator) signalling (Zhou *et al.*, 1998). Loss of SMAD4 may lead cells to become TGFβ resistant and escape from TGFβ mediated growth control and apoptosis (Moren *et al.*, 2000).

This chapter aims to address the role of *SMAD4* in colorectal cancer, both in its frequency and the timing. By evaluating the relative contribution of *SMAD4* to CRC tumorigenesis and comparing this to the observed incidence of 18q21.1 loss, it is

hoped that it can be established whether the loss is targeted to *SMAD4*, *DCC* or another, as yet unidentified, gene.

4.2 ASSESSING THE FREQUENCY OF SMAD4 LOSS IN CRC CELL LINES

To attempt to resolve this issue of the *SMAD4* involvement in colorectal tumour progression, it was determined how many of 43 colorectal cancer cell lines had loss of the SMAD4 protein, then the cause of this loss was investigated by mutation analysis and assessment of allele loss. DNA and protein pellets were available from 43 established CRC cell lines (C10, C32, C70, C80, C84, C99, C106, C125, CAC02, COLO201/COLO205, COLO320, COLO678, COLO741, CX1/HT29, GP2D, HCA46, HCA7, HCT8, DLD1/HCT15, HCT116, HRA19, HT55, LIM1863, LOVO, LS174T, LS180, LS411, LS1034, PC/JW, SKC01, SW48, SW403, SW480, SW620, SW837, SW948, SW1116, SW1222, SW1417, VACO4A, VACO4S, VACO5, VACO10). The *APC* mutation status, β-catenin mutation status and microsatellite instability (MSI) status have been established for these cell lines (summarised in Rowan *et al.*, 2000). For detection of the SMAD4 protein, Western blotting using B8 and a control antibody (β-actin) was performed as described in Chapter Three. SMAD4 protein was classed simply as 'present' or 'absent'. To assess mutations of *SMAD4* in the CRC cell lines, F-SSCP (fluorescent single stranded conformational polymorphism analysis) and the PTT (protein truncation test) were performed as

described in Chapter Three, using the same primers and conditions. Repeated failure to amplify any segment of the *SMAD4* gene in the PCR, despite successful amplification in control PCRs (Rowan *et al.*, 2000) was taken to denote homozygous deletion of that segment. For the PTT, RNA was available from a subset of CRC cell lines (SW1222, HRA19, SKC01, SW948, JW, HCT8, COLO205, LS174T, SW48, LOVO, SW620, COLO320, GP2D, HT29, CACO2, HCA46). cDNA was synthesised using the First Strand Synthesis kit (Promega). For the assessment of 18q allelic loss, seven microsatellite markers were selected from the Location Database (http://cedar.genetics.soton.ac.uk/pub/chrom18/map.html); for *SMAD4* the markers were in the order, D18S479-1.45Mb-D18S474-0.01Mb-D18S46-0.35Mb-(*SMAD4*); and for *DCC* (about 1.5Mb telomeric of *SMAD4*) the order was D18S484-0.16Mb-D18S487-0.44Mb-DCC-0.19Mb-D18S35. The heterozygosity for each of the seven markers was reported to be 60%, 82%, 80%, 72%, 81%, 87% and 70% respectively. In addition, markers mapping to 18p (D18S481), the centromere (D18S877) and the telomeric region of 18q (D18S878 and D18S844) were included. The forward primer was fluorescently labelled with HEX, FAM or TET. Standard PCR conditions were used before running the amplimers on an ABI377 semi-automated sequencer and analysing the results with Genotyper™ software to assign peak sizes. Since no normal tissue was available, allelic loss was assumed to have occurred at *DCC* or *SMAD4* if all microsatellite markers close to that gene were hemi-/homozygous (corresponding to $p<0.01$).

Overall, 28% of the CRC cell lines showed loss of SMAD4 expression (Tables 4.2.1 and 4.2.2, Figure 4.2.1), comprising 12 of 31 (39%) MSI- (microsatellite stable or microsatellite instability-negative) lines and 0 of 12 MSI+ (microsatellite instability-positive), lines ($p<0.004$, Fisher's exact test). In no case was a truncated protein band observed, and several different types of mutation were found to account for the loss of expression (Table 4.2.1). Importantly, sequencing showed that all *SMAD4* mutations were present in the homozygous or hemizygous state (Figure 4.2.3), with no underlying wild type sequence. For one cell line (COLO678), the whole gene was homozygously deleted (plus the nearby marker D18S474), whereas other cell lines showed partial homozygous deletions (exons 1-4 in COLO201/COLO205, and exons 10-11 in SW403). All these mutations were accompanied by allelic loss at *SMAD4* (Table 4.2.1 and Table 4.2.3). Putative splicing mutations were detected in three cell lines, SW480 and SW620 (derived from a primary tumour and metastasis) and C10. All these lines showed absent SMAD4 protein and allelic loss at *SMAD4*. Of these three lines, SW620 was tested using the PTT and only normal length mRNA was detected (Figure 4.2.2). Thus, although the mutation in SW620 does not lead to detectable abnormal mRNA splicing, the possibility cannot be excluded that it has pathogenic effects through a reduction of normal mRNA levels. Nonsense *SMAD4* mutations, accompanied by allele loss and absent protein, were detected in two cell lines by PTT: HT29/CX1 (Q311X, exon 7); and VACO10MS (Q239X, exon 5) (Figure 4.2.3).

Cell line	MSI[a]	B8 Western[b]	SMAD4 mutation[c]	Predicted effect of mutation[d]	Loss at SMAD4[e]	Loss at DCC[f]	Karyotype[g]	18q21 status[h]	TGFBIIR[i]
COLO678	MSI-	Absent	EX1-11del	No protein	Yes	Yes	?		-
COLO201	MSI-	Absent	EX1-4del	No protein	Yes	Yes	78	2 copies	None
COLO205	MSI-	Absent	EX1-4del	No protein	Yes	Yes	68	2 copies	None
VACO10	MSI-	Absent	c.715C>T (Q239X), EX5	Nonsense	Yes	Yes	115	-	None
C10	MSI-	Absent	IVS6-1G>T	Splice disruption	Yes	Yes	49	-	-
HT29/CX1	MSI-	Absent	c.931C>T (Q311X), EX7	Nonsense	2/3 markers	Yes	71	2 copies	None
SW480	MSI-	Absent	IVS7+5G>C	Splice disruption	Yes	Yes	57	1 copy	None
SW403	MSI-	Absent	EX10-11del	Truncated protein	Yes	Yes	68	2 copies	-
SW620	MSI-	Absent	IVS7+5G>C	Splice disruption	Yes	Yes	50	1 copy	None
CACO2	MSI-	Present	c.1051G>C (D351H), EX8	Missense	Yes	Yes	96	-	None
C80	MSI-	Present	c.1051G>C (D351H), EX8	Missense	Yes	Yes	69	2 copies	-
SW948	MSI-	Present	c.1609G>T (D537Y), EX11	Missense	Yes	No	67	-	-
SW1222	MSI-	Present	c.1619T>G (L540R), EX11	Missense	Yes	Yes	?	18q loss	-
HT55	MSI-	Absent	None found	N/A	Yes	Yes	80	2 copies	-
SW1116	MSI-	Absent	None found	N/A	Yes	Yes	63	-	-
SW1417	MSI-	Absent	None found	N/A	Yes	No	70	2 copies	-
C106	MSI-	Absent	None found	N/A	Yes	Yes	79	2 copies	-
PC/JW	MSI-	Absent	None found	N/A	Yes	Yes	70	2 copies	-

Table 4.2.1 Summary of molecular and cytogenetic data for 18q21.1 status in MSI- colorectal cancer cell lines.

(a) MSI status; (b) SMAD4 protein expression as assessed using B8 and western blotting; (c) identified SMAD4 mutations and (d) their predicted effect; (e) allelic loss as inferred from homozygosity at microsatellite markers near SMAD4, and (f) at DCC; (g) modal chromosome number of cell line (? = not known); (h) 18q21.1 status determined by CGH, SKY and FISH (- = not done); (i) TGFBIIR mutation status (- = not done). COLO201/COLO205, HT29/CX1 and DLD1/HCT8/HCT15 are essentially identical cell lines. ***Continued on next page.***

Cell line	MSI[a]	B8 Western[b]	SMAD4 mutation[c]	Predicted effect of mutation[d]	Loss at SMAD4[e]	Loss at DCC[f]	Karyotype[g]	18q21 status[h]	TGFBIIR[i]
SW837	MSI-	Present	None found	N/A	Yes	Yes	40	1 copy	None
C99	MSI-	Present	None found	N/A	Yes	Yes	52	1 copy	-
C84	MSI-	Present	None found	N/A	2/3markers	Yes	56	1 copy	-
C125	MSI-	Present	None found	N/A	Yes	Yes	60	1 copy	-
VACO4A	MSI-	Present	None found	N/A	Yes	Yes	60	2 copies	-
HCA46	MSI-	Present	None found	N/A	Yes	Yes	71	2 copies	None
C32	MSI-	Present	None found	N/A	Yes	Yes	74	2 copies	-
LIM21-1863	MSI-	Present	None found	N/A	Yes	Yes	80	3 copies	-
C70	MSI-	Present	None found	N/A	Yes	Yes	127	4 copies	-
COLO320	MSI-	Present	None found	N/A	Yes	Yes	53	-	None
VACO4S	MSI-	Present	None found	N/A	Yes	Yes	64	-	-
LS1034	MSI-	Present	None found	N/A	Yes	Yes	77	-	-
HRA19	MSI-	Present	None found	N/A	Yes	Yes	?	-	-
SKC01	MSI-	Present	None found	N/A	No	Yes	hypertriploid	-	-

Table 4.2.1 continued.

Cell line	MSI[a]	B8 Western[b]	SMAD4 mutation[c]	Predicted effect of mutation[d]	Loss at SMAD4[e]	Loss at DCC[f]	Karyotype[g]	18q21 status[h]	TGFBIIR[i]
HCT8	MSI+	Present	None found	N/A	No	No	?	-	-
LS180	MSI+	Present	None found	N/A	No	No	45	-	-
SW48	MSI+	Present	None found	N/A	No	No	47	-	Mutant x 2
LS174T	MSI+	Present	None found	N/A	No	No	45	2 copies	Mutant x 2
DLD1/HCT15	MSI+	Present	None found	N/A	No	No	46	2 copies	Mutant x 2
LOVO	MSI+	Present	None found	N/A	No	No	49	2 copies	Mutant x 2
VACO5	MSI+	Present	None found	N/A	No	No	47	2 copies	Mutant x 2
HCA7	MSI+	Present	None found	N/A	No	No	43	2 copies	Mutant x 2
HCT116	MSI+	Present	None found	N/A	No	No	45	2 copies	Mutant x 2
LS411	MSI+	Present	None found	N/A	No	No	75	2 copies	Mutant x ?
GP2D	MSI+	Present	None found	N/A	No	No	46	2 copies	-
COLO741	MSI+	Present	None found	N/A	Yes	Yes	?	2 copies	-

Table 4.2.2 Summary of molecular and cytogenetic data for 18q21.1 status in MSI+ colorectal cancer cell lines.

(a) MSI status; (b) SMAD4 protein expression as assessed using B8 and western blotting; (c) identified SMAD4 mutations and (d) their predicted effect; (e) allelic loss as inferred from homozygosity at microsatellite markers near SMAD4, and (f) at DCC; (g) modal chromosome number of cell line (? = not known); (h) 18q21.1 status determined by CGH, SKY and FISH (- = not done); (i) TGFBIIR mutation status (- = not done). COLO201/COLO205, HT29/CX1 and DLD1/HCT8/HCT15 are essentially identical cell lines.

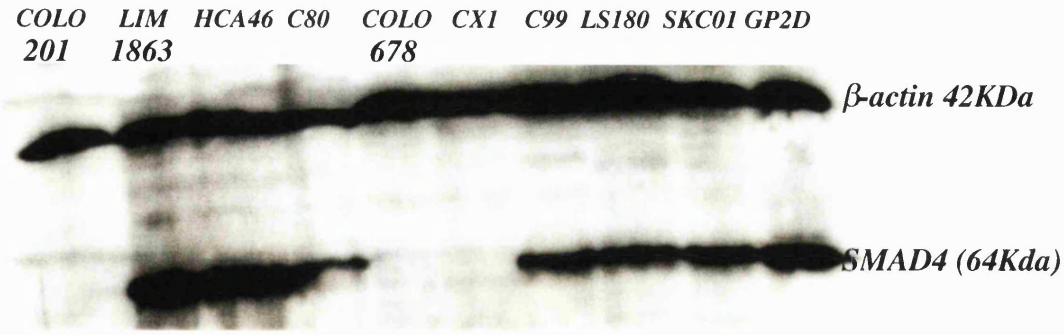


Figure 4.2.1: Western blot analysis of SMAD4 and β-actin in colorectal cancer cell lines.
Shown are the results of western blot analysis using the B8 antibody against SMAD4 (64Kda) , and a monoclonal anti-β-actin (42Kda) for ten colorectal cancer cell lines Note complete absence of SMAD4 expression in COLO201, COLO678 and CX1, and apparently reduced expressoin in C80..

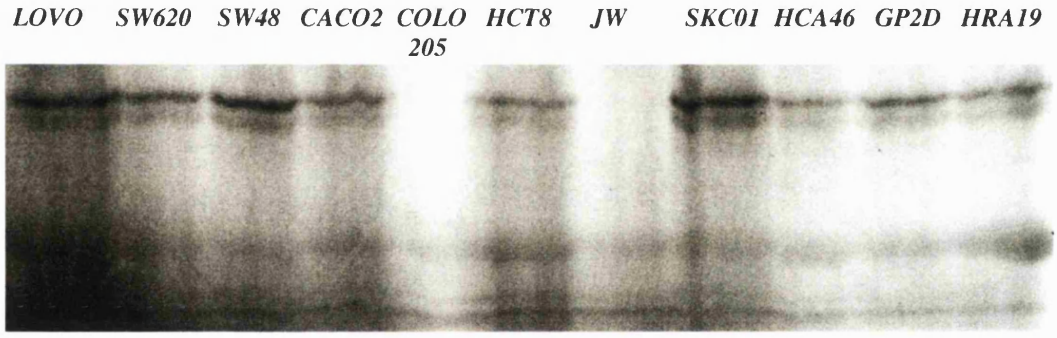


Figure 4.2.2 PTT results for SMAD4 from CRC cell lines.
Shown are the PTT results from 11 colorectal cancer cell lines using probe I/III. Note the complete absence of translated protein in COLO205 (deletion of exons 1-4) and JW (no mutation identified but no SMAD4 protein detectable on Western blotting)

In addition to mutations which led to the loss of SMAD4 protein, missense mutations, plus allelic loss, were found in four lines: CACO2 (D351H)(Figure 4.2.3); C80 (D351H); SW948 (D537Y); and SW1222 (L540R). SMAD4 protein was present in all of these four lines, although it did appear to be reduced when compared to other lanes probably as a result of protein degradation (Figure 4.2.1). Four MSI- cell lines (HT55, JW, SW1417, C106) had absent SMAD4 protein and allelic loss, but no detectable *SMAD4* mutation even by sequencing of all the exons (Table 4.2.1). One further cell line, SW1116, had no SMAD4 protein detected upon Western blotting but the DNA was of poor quality, resulting in failure to sequence the *SMAD4* gene in this line. The genetic defect underlying the lack of SMAD4 protein was therefore not resolved in this cell line either, but it is doubtless, as it is with HT55, JW, SW1417 and C106, that the defects exist. It is possible that cryptic *SMAD4* mutations or some other means of inactivating SMAD4 may have occurred, although the obvious epigenetic mechanism of inactivation has been reported not occur at *SMAD4* in colorectal cancers (Roth *et al.*, 2000). Consistent with this finding, we detected *SMAD4* mRNA in PC/JW which had absent protein but no detectable mutation. It is also possible that changes in a protein upstream of SMAD4 can sometimes lead to loss of SMAD4 expression as suggested by Salovaara *et al* (in press)

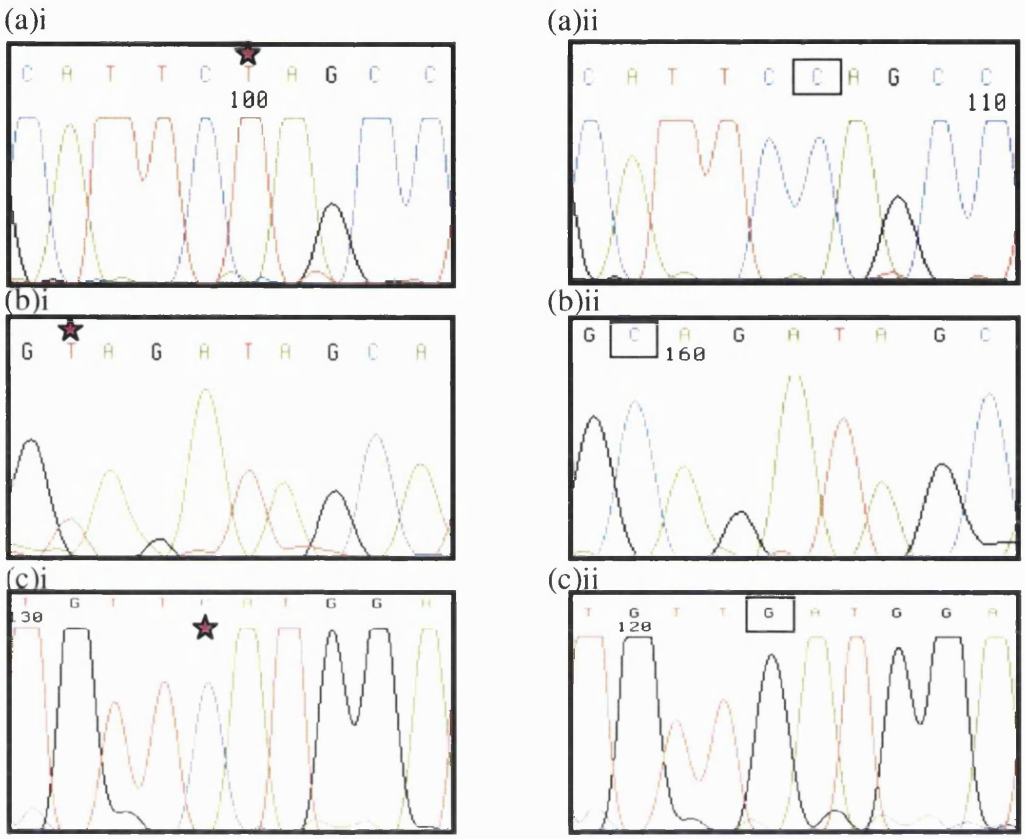


Figure 4.2.3: Sequencing results of SMAD4 for HT29/CX1, VACO10 and CACO2.

(a)i shows the mutated cell line HT29/CX1 (c.931C>T transition causing Q311X amino acid truncation) compared to (a)ii wild-type. (b)i SMAD4 exon 5 of VACO10 showing the c.715C>T transition causing Q239X truncation, compared to (b)ii wild-type. (c)i SMAD4 exon 8 of CACO2 showing c.1051G>C nucleotide change which results in a D351H missense change, compared to (c)ii wild-type. Note lack of wild-type sequence underlying mutated (starred) sequence. Corresponding wild-type base is shown in a box.

None of the 12 MSI+ cell lines possessed a pathogenic *SMAD4* mutation (or loss of the SMAD4 protein), compared to 16 of 31 (52%) (known or deduced from the lack of protein) MSI- lines, a highly significant difference ($p<0.005$, Fisher's exact test). A known *SMAD4* polymorphism in intron 7 was detected (MacGrogan *et al.*, 1997) in MSI- (CACO2, C106 and SW403 all apparent homozygotes for the non-wild-type allele) and MSI+ lines (GP2D, GP5D and HCT116, all heterozygotes).

Overall, there was a striking level of allele loss at *SMAD4* and *DCC* in all the MSI- lines, with 28/31 (90%) showing loss at *SMAD4* and 29/31 (94%) with loss at *DCC* (Table 4.2.3). In accordance with the sequencing data, all lines with allelic loss showed complete absence of one microsatellite allele; thus, all copies of 18q21 in each cell were derived from the same chromosome 18 homologue, even where the cancer was polyploid and had more than one copy of 18q. In contrast, 11/12 (92%) MSI+ lines showed heterozygosity for at least one marker each at *SMAD4* and *DCC*, indicating that loss of 18q material is not critical for the development and progression of these tumours.

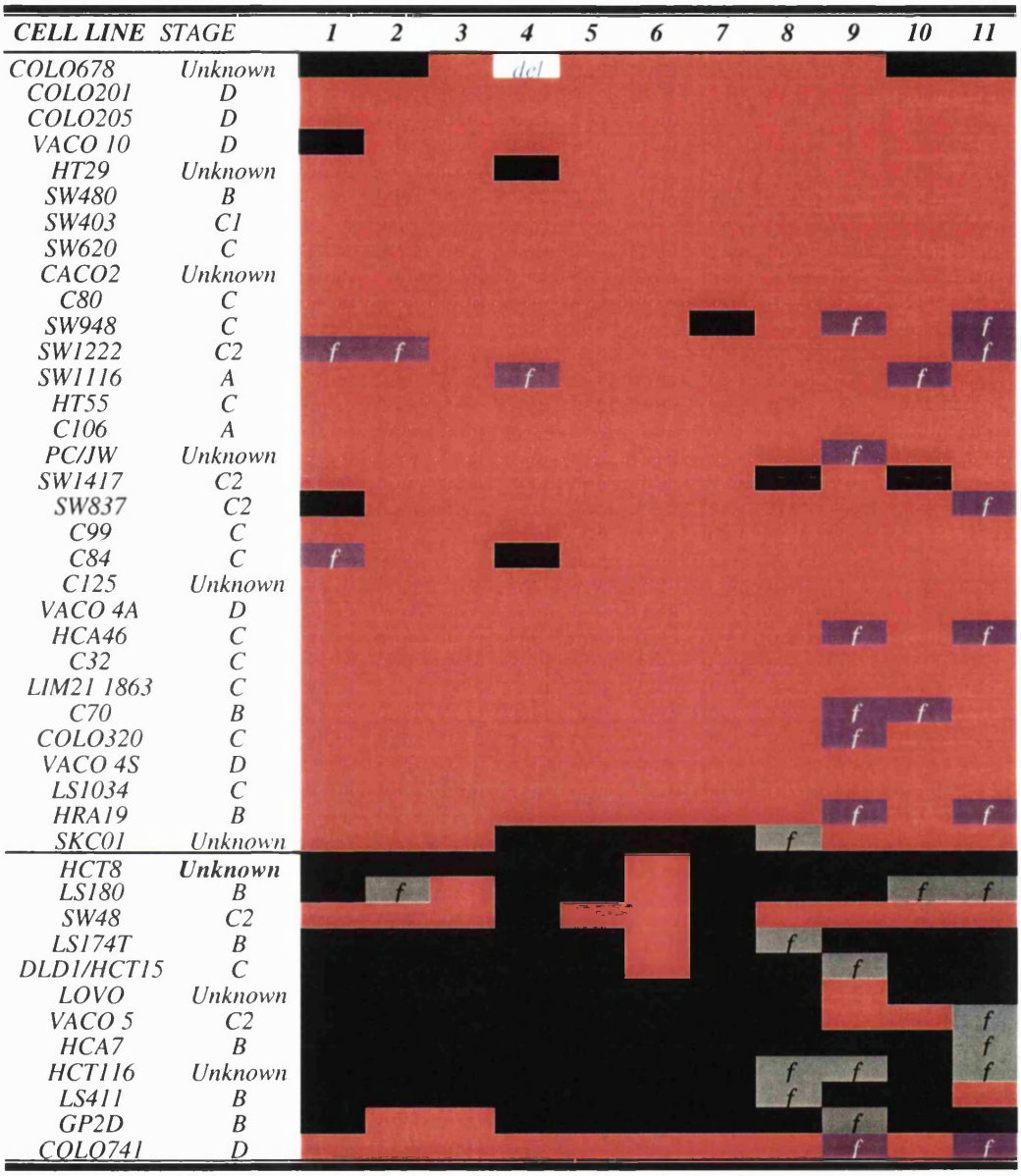


Table 4.2.3 LOH summary of chromosome 18 in CRC cell lines.

Red boxes=homozygote, Black boxes=heterozygote, Pink boxes=failed but probably homozygote, Grey boxes=failed, Del=deleted, 1=D18S481(p-arm), 2=D18S877 (centromere), 3, 4, 5 flank SMAD4 (=D18S479, D18S474, D18S46), 6, 7, 8, 9 flank DCC (=D18S487, D18S484, dcc, D18S35), 10=D18S878 , 11=D18S844 (telomere).

4.3 THE TIMING OF SMAD4 LOSS IN COLORECTAL CANCER

Loss of 18q and the genes contained within is a well documented feature of colorectal cancer. The finding of 18q loss in all of the MSI- lines was rather unexpected and, combined with the fact that all *SMAD4* mutations were present in the homozygous state, indicated that these lines only had one chromosome 18 homologue (and therefore two hits at *SMAD4* in the lines with *SMAD4* mutations). As can be seen from Table 4.2.1, the great majority of MSI- cell lines ~~have a~~ are triploid or near-triploid. It was decided to investigate the copy number of *SMAD4*, using molecular cytogenetic techniques, and relate this to the karyotype of the cell line. Copy number of 18q21 was assessed by a combination of comparative genomic hybridisation (CGH) (Weiss *et al.*, 1999) and locus-specific fluorescence *in situ* hybridisation (FISH) (performed on the cell lines by Pat Gorman, ICRF) using PAC 224_j_22, to which *SMAD4* is known to map (described in detail in Chapter Five). In addition spectral karyotyping (SKY) results for a subset of lines had recently been published (Abdel-Rahman *et al.*, 2001) and this data is also included in Table 4.2.1. The molecular cytogenetic and karyotype results are summarised in Table 4.2.1. Importantly, in the lines with 18q21 loss, the amount of chromosome 18q21 material was always less than the overall ploidy. For example MSI- cell lines with a modal chromosome number of less than 60 (C84, SW837, C99, C125, SW480 and SW620) only had one copy of chromosome 18q21, whereas cell lines with near-triploid karyotypes (SW403, HT29/CX1, HT55, COLO201, COLO205, C32, C106, C80, PC/JW, SW1417, VACO4A, HCA46) had two copies of chromosome 18q21. Two lines

with hypotetraploid (LIM1863) and hypohexaploid (C70) karyotypes had 3 and 4 copies of chromosome 18q21 respectively.

In most cases, the cytogenetic analysis showed the whole 18q arm to be deleted, but there were exceptions. SW1417 was homozygous for microsatellites at *SMAD4* and showed loss of SMAD4 protein, but was heterozygous for markers near *DCC*. The microsatellite data for SW1417 were consistent with an interstitial deletion which targeted *SMAD4*, but left *DCC* intact. CGH data for SW1417 also showed a deletion around *SMAD4* rather than loss of the whole arm. Conversely, C84 and SKC01 were heterozygous for microsatellites at *SMAD4*, but homozygous for microsatellites around *DCC*. Again this was substantiated by CGH data which showed loss of chromosome 18 distal to 18q21. In these lines, SMAD4 expression was retained and it appeared that *DCC* or another tumour suppressor was being targeted, leaving *SMAD4* intact. The line HT29/CX1 had two apparently identical copies of chromosome 18 with *SMAD4* mutations, plus one deleted 18q. This line was heterozygous at D18S474, but homozygous at the rest of the microsatellites and for the *SMAD4* mutation, showing that the breakpoint for the deletion almost certainly lay just centromeric to *SMAD4*. Finally, cell line COLO678 (with homozygous deletion of *SMAD4* and the nearby marker D18S474) was found to be homozygous for markers mapping to both the *SMAD4* and *DCC* regions but heterozygous for markers on the p-arm and 18q telomeric region. Cytogenetic data was not available for COLO678 but presumably it contains an interstitial deletion targeting *SMAD4* and *DCC* without loss of the whole q-arm.

The frequency of 18q loss in colorectal tumorigenesis is generally considered to increase with increasing tumour grade indicating that loss of *SMAD4* is a late event in colorectal tumorigenesis. Maitra *et al* found 0% of adenomas or stage I tumours had SMAD4 loss (assessed using B8 immunohistochemistry), whereas 8% of stage II, 6% of stage III and 22% of stage IV (with distant metastasis) had loss of SMAD4 expression (Maitra *et al.*, 2000). Although the data presented here indicates little role for SMAD4 in MSI+ tumours, the cancers of Maitra *et al* were not stratified into MSI+ or MSI-. MSI+ tumours are well documented to present clinically at a lower grade than MSI- tumours. It is likely therefore that rather than an increasing role for SMAD4 the higher the grade of tumour, there were probably just fewer MSI+ tumours in the high grade group to dilute out the real SMAD4 effect. This is borne out with the CRC cell line data. Looking at the MSI- cell lines alone (as they are the only tumours that have loss of SMAD4), 3/5 grade A or B tumours and 9/20 grade C or D have loss of SMAD4 (Fishers exact >0.1) (Table 4.3.1). These figures do not substantiate Maitra *et al*'s observations of increasing frequency of *SMAD4* inactivation with the progression of cancer, and the frequencies of loss of *SMAD4* overall are much higher in the colorectal cancer cell lines. Likewise, similar frequencies to Maitra *et al* of SMAD4 loss in the different tumour grades was found by Miyaki et al (Miyaki *et al.*, 1999), but here again tumours were not classified as MSI-positive or negative and therefore the comparison may not be legitimate. In keeping with the data obtained here, higher frequencies were also observed by Salovaara *et al.*, (in press) who showed 20/53 (38%) MSI- colorectal adenocarcinomas had total loss

of the SMAD4 protein, and 15/53 (28%) had markedly reduced SMAD4 protein (<5% of cells staining positive), equating overall to 66% of cancers with *SMAD4* targeted. Twenty six out of 27 MSI+ cancers of Salovaara *et al* were positive for SMAD4 staining, also substantiating the CRC cell line observations.

Dukes Stage	*MSI- lines without functional SMAD4*	*MSI- lines with functional SMAD4*	*MSI+ lines (all with functional SMAD4)*
A	*2*	*0*	*0*
B	*1*	*2*	*5*
C	*6*	*8*	*3*
D	*3*	*2*	*1*
Not done	*4*	*2*	*3*

Table 4.3.1 Dukes stage of the colorectal cancer cell lines.
A=non-invasive, B=invasive but contained in the bowel, C= local metastasis, D= distant metastasis.

4.4 HOW SMAD4 MUTATIONS FIT IN THE SEQUENCE OF EVENTS

The data has shown that all lines with SMAD4 abnormalities are uniformly hemi- or homozygous at both *SMAD4* and *DCC* microsatellites and for their *SMAD4* mutations (where applicable), showing that all copies of 18q were derived from the same chromosome 18 homologue, even where more than one copy of 18q21 was present in the cell. However, the dosage of 18q21 was always decreased relative to the overall level of ploidy. No cancer had more than one independent intragenic mutation in *SMAD4*. The simplest model to explain these data (Figure 4.4.1) is one in which *SMAD4* mutations occur before colorectal tumours acquire chromosomal instability (CIN) and polyploidy/aneuploidy, and

thus *SMAD4* mutations are not late changes, contrary to previous suggestions (Maitra *et al.*, 2000; Miyaki *et al.*, 1999). Alternative explanations to this model are less plausible, because they would require several independent mutations to inactivate all copies of *SMAD4* and/or frequent, independent occurrence of identical *SMAD4* mutations in the same cell. However, a recent article by Fodde *et al* has suggested that tetraploidy is a very early event in *APC*-mutant colorectal tumours (Fodde *et al.*, 2001), contrary to the model presented here, and if confirmed, it would suggest that some CRCs had lost three copies of 18q, acquired an intragenic *SMAD4* mutation and had reduplicated the remaining *SMAD4* mutant chromosome.

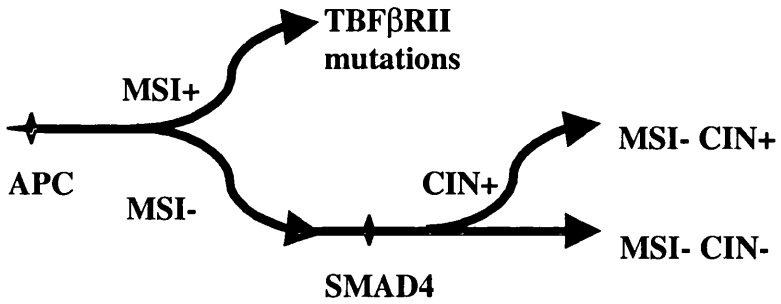


Figure 4.4.1: Pathway showing possible sequence of events in tumorigenesis. *Following mutations in the APC gene, a subset of tumours have inactivation of mismatch repair genes and so diverge along a pathway that includes mutations of the TGFβRII gene, and is characterised by microsatellite instability (MSI+). A subset of the remaining tumours acquire mutations in the SMAD4 gene, accompanied by loss of the wild-type chromosome 18, and this either precedes or causes another subset of tumours to diverge along a chromosomal instability (CIN+) pathway with aneuploidy/polyploidy. The remaining tumours are MSI-/CIN-.*

Given that none of 12 MSI+ cell lines had *SMAD4* mutations or loss of *SMAD4* protein, it is most likely that the MSI+ pathway had diverged before *SMAD4* inactivation occurred (Figure 4.4.1). Eight MSI+ lines have been previously studied and reported to have *TGFBIIR* mutations (Parsons, *et al.*, 1995;

Markowitz, *et al* 1995) compared with none of ten MSI- lines studied (Ilyas *et al.*, 1999) (Fisher's exact test, $p<0.000003$). Although inactivation of *SMAD4* and *TGFBIIR* may well not be functionally equivalent, these data certainly suggest that mutations of *TGFBIIR* occur after the divergence of the MSI-/+ lineages. Thus, our model shows initial divergence of the MSI+ pathway, followed by 18q loss and *SMAD4* mutation in the MSI- cancers. Given that limited data suggest a high frequency of 18q loss in MSI-CIN- CRCs ([Georgiades, 1999 #1291]), divergence of the CIN- and CIN+ pathways probably occurs after 18q loss and *SMAD4* mutation.

4.5 CONCLUSIONS

Loss of SMAD4 expression and/or mutation were found in about half of the MSI- lines but in no MSI+ lines. A broad spectrum of *SMAD4* mutations was seen, although, for unknown reasons, there were no frameshift changes. Mutations appeared to occur more frequently in the C-terminal MH2 domain of *SMAD4*, but two cell lines (VACO10 and HT29/CX1) had mutations (both truncating) in the less conserved 'linker' portion of the gene. Missense mutations in the N-terminal MH1 domain have been shown to cause *in vitro* protein instability (Xu and Attisano, 2000), Four cell lines with retained protein expression had missense *SMAD4* mutations in the MH2 region (Table 5.2.1). The missense mutations of C80 and CACO2 (D351H) occurred in the three loop/helix of the MH2 and have known pathogenic effects (Shi *et al.*, 1997). The

nearby R361C mutation, found not only in sporadic cancer (Koyama *et al.*, 1999), but also in the germline of JPS patients, is associated with undetectable levels of protein in JPS polyps, as assessed using B8 immunohistochemistry. Mutations in the 3 helix bundle of the MH2, such as R537Y (SW948) and L540R (SW1222) appear not to lead to unstable protein, but, like D351H, are predicted to abrogate protein function (Eppert *et al.*, 1996; Schutte *et al.*, 1996; Shi *et al.*, 1997) by impairing the ability of SMAD4 to act as a homo- or hetero-oligomer and so interfering with signal transduction. Early experiments with these SMAD4 missense cell lines in the Developmental Signalling Laboratory (ICRF) have suggested that signal transduction is indeed impaired.

Almost all MSI- cell lines showed allelic loss and even in MSI- lines without defects in *SMAD4*, the dosage of 18q21 was always decreased relative to the overall level of ploidy; the question therefore remains as to the cause of 18q loss in lines with no evidence of *SMAD4* inactivation. Loss of 18q21 is seen in 80% of pancreatic cancers, with inactivation of *SMAD4* shown to result from homozygous deletions in 30% of cases and point mutations/small alterations in 20% of cases (Hahn *et al.*, 1996c). Of the 60% of colorectal cancers showing 18q loss in the study by Thiagalingam *et al*, mutations of *SMAD4* were demonstrated in about one third (Thiagalingam *et al.*, 1996). Several authors have reported a lower frequency of SMAD4 mutations than this - overall approximately 14% of primary colorectal tumours have been shown to harbor *SMAD4* alterations (Maitra *et al.*, 2000). This is probably a great underestimation however due to multiple factors. Firstly, as the work shown here demonstrates, MSI+ tumours do

not have *SMAD4* inactivation as part of their genetic fingerprint, but most authors fail to differentiate between the two types of tumour when reporting the frequency of *SMAD4* alterations (Barbera *et al.*, 2000; MacGrogan *et al.*, 1997; Maitra *et al.*, 2000; Miyaki *et al.*, 1999). Next, several authors fail to find alterations in the *SMAD4* gene using SSCP analysis or using the assumption that if a subset of exons are amplifiable then the protein is intact, which is not 100% sensitive or reliable (Hoque *et al.*, 1997; Koyama *et al.*, 1999; Lei *et al.*, 1996; MacGrogan *et al.*, 1997; Takagi *et al.*, 1996). Thirdly, as shown with the CRC cell lines, the genetic alteration leading to a loss of the protein are not always identifiable - lack of a genetic alteration is not a definitive indication that there is an intact and functional protein. Finally, any screening technique with primary tumours is likely to lead to an underestimation of the frequency of mutations due to contaminating normal tissue, whereas cell lines are totally tumour enabling more sensitive detection of mutations.

Although *SMAD2* remains an unlikely target, haplo-insufficiency of *SMAD4* is possible and *DCC* changes, though unlikely, cannot be excluded with certainty. What is certainly true is that a subset of cancers have 18q loss which is not targeted to *SMAD4* (MacGrogan *et al.*, 1997; Tarafa *et al.*, 2000), and indeed appear to be targeting *DCC*. Any other target genes on 18q must, like *SMAD4*, be altered after the MSI+ pathway had diverged (although mutations in these gene(s) might, of course, be common to the MSI+ and MSI- pathways). With the advent of the draft human genome sequence being available, it may finally be possible to unravel the complete targets of 18q loss in colorectal cancer (apart

from *SMAD4*) and determine whether there does exist another as yet unidentified gene

CHAPTER FIVE

ALLELIC LOSS AT SMAD4 IN JUVENILE POLYPS, AND THE CLONAL ORIGIN OF JUVENILE POLYP EPITHELIUM

ALLELIC LOSS AT SMAD4 IN JUVENILE POLYPS, AND THE CLONAL ORIGIN OF JUVENILE POLYP EPITHELIUM

5.1 INTRODUCTION

Gastrointestinal hamartomas are a feature of multiple cancer-predisposing syndromes including Juvenile Polyposis Syndrome (JPS) and Peutz-Jeghers syndrome (PJS), where they are associated with an increased risk of gastrointestinal malignancy. In addition there are Cowden disease (CD), Bannayan-Zonana syndrome (BZS), Gorlin Syndrome (GS) and tuberous sclerosis (TSC), not to mention PJS, where the lesions (hamartomas included) are not confined to the GI tract but occur through the different organ systems where they are associated with syndrome-specific malignancies. Hamartomas are considered to be an overgrowth of normal tissue comprising of cells derived from the tissue of origin, but despite this normal appearance, the malignant potential of hamartomas in each of these syndromes is being increasingly recognised. PJS predisposes to cancers at various sites including the GI tract, the breast and cervix, whereas GS is primarily associated with basal cell carcinomas and CD with breast and thyroid cancers. The most serious morbidity in TSC is caused by central nervous system hamartomas which give rise

to epilepsy and mental retardation, but are only infrequently associated with malignant progression, with the most common site for malignancy in TSC being the kidney (Hodgson and Maher, 1999).

Loss of genetic material, in particular tumour suppressor genes, is normally associated with the development of 'true' neoplasms such as the adenomas which occur in HNPCC or FAP. The apparent non-neoplastic nature of the polyps associated with the hamartoma syndromes would perhaps imply that hamartomas are not as a result of the loss of tumour suppressor genes. The identification of the PJS locus was, however, by virtue of the loss of 19p (containing the *LKB1* locus) observed in PJS polyps using comparative genomic hybridisation (Hemminki *et al.*, 1997). This demonstrated that PJS hamartomas arose as a result of the inactivation of two copies of a tumour suppressor gene via classical 'two-hit' mechanisms. Confirmation of the monoclonal origin of the epithelium of PJS polyps was provided by the finding of allelic loss at the *LKB1* locus (Gruber *et al.*, 1998; Wang *et al.*, 1999). PJS associated cancers also show loss of the wild-type *LKB1* allele in patients with a germline *LKB1* mutations, strongly suggesting the progression from hamartoma to carcinoma (Wang *et al.*, 1999). Hamartomas, adenomas and carcinomas from CD patients have been shown to have LOH and/or reduced RNA levels at the *PTEN* locus on chromosome 10q23.3, again suggesting that this gene is acting as a tumour suppressor, with loss of the wild-type gene initiating hamartoma growth and leading to subsequent progression to carcinomas (Marsh *et al.*, 1998).

GS-associated juvenile polyps are an infrequent manifestation and so allelic loss has not been investigated in these, although basal cell carcinomas in GS have been shown to have allelic loss at *PTCH* (Unden *et al.*, 1996). Finally, allelic loss at both the *TSC1* locus on 9q34 and *TSC2* locus on 16p13.3 has been demonstrated in TSC hamartomas (Sepp *et al.*, 1996). In each of these syndromes (PJS, CD, GS and TSC) the data are consistent with each of the respective genes having a tumour suppressive function, with the first hit in the germline and the second hit occurring somatically.

In addition to historically being described as 'stromal' lesions, two pieces of evidence suggested that the hamartomas observed in JPS may somehow not fall into this classical tumour suppressor-inactivation pathway. The identification of germline mutations in *SMAD4* as a cause of a subset of JPS cases (Howe *et al.*, 1998b) was accompanied by allele loss studies. Only one of 11 polyps showed loss of the wild-type allele in a fragment spanning the germline 4 base pair deletion in exon 9 of *SMAD4*. Secondly, Jacoby *et al* in 1997 demonstrated that juvenile polyps showed loss of a putative locus (termed JP1) on 10q22, but the cells targeted for loss were in the stroma rather than the epithelium (Jacoby *et al.*, 1997). To explain how stromal defects, either at JP1 or *SMAD4*, may predispose to epithelial malignancy, Kinzler and Vogelstein proposed the 'landscaper hypothesis' (Kinzler and Vogelstein, 1998). Whereas classical tumour suppressor genes (such as *APC*) have been termed 'gatekeepers', preventing runaway growth, and DNA repair genes (such as *MSH2*)

have been dubbed the 'caretakers' of the genome (Kinzler and Vogelstein, 1997), genes causing JPS (and indeed ulcerative colitis) were considered to be acting via an indirect route with the stroma dictating epithelial growth via a 'landscaper' effect. Under this model, neoplastic progression is driven by initial genetic changes within the stromal cells of the hamartoma (classically the clonal component of these lesions) and the increased risk of cancer is as a result of an altered terrain for epithelial cell growth.

The aims of this chapter were to investigate whether the landscaper hypothesis was indeed true for juvenile polyposis syndrome. This would be achieved by looking for allele loss at *SMAD4* in JPS polyps to ascertain whether *SMAD4* does indeed act as a tumour suppressor in JPS (as it does in sporadic pancreatic and colon cancer), and then to determine which cells were targeted for deletion, the stromal cells or the epithelial cells.

5.2 ALLELIC LOSS AT THE SMAD4 LOCUS IN JUVENILE POLYPS

DNA was extracted from paraffin embedded blocks of juvenile polyps and cancers originating from Families 1, 12, 17, 18, 20, and 21 and sporadics LB and 1262. In addition, included in this study were polyps from 5 Korean sporadic cases (1-1, 4-1, 7-1, 9-1, 13-1) from whom no detailed clinical information was available. Families

17, 20 and 21 all have ~~all had~~ germline mutations in *SMAD4* identified and there were a total of 19 juvenile polyps available from 6 patients from these families. In addition there were 19 polyps and 8 carcinomas from 12 individuals who (a) had had *SMAD4* excluded as the cause of their JPS via linkage analysis (Families 1 and 12, discussed in Chapter Three), (b) had no detectable germline *SMAD4* mutation (Families 1, 12 and 18, sporadics LB and 1262), or (c) had not been tested for germline *SMAD4* mutations (the five Korean samples). For the assessment of allele loss, five 10μm slides were crudely microdissected, after de-waxing of the slide when necessary, using a needle to remove the polyp tissue from its stalk and any surrounding normal tissue. No attempt was made to separate the stromal and epithelial components in the juvenile polyps. The microdissected material was scraped into a tube containing an appropriate amount of proteinase K digestion buffer (depending on the size of the lesion e.g. 50μl per 0.5 x 0.5cm) and incubated at 55°C for three days. Initially, seven microsatellites were selected for the assessment of allele loss (D18S877, D18S851, D18S474, D18S878, ATA7D07, ATA82B02 and GATA177C03), spaced at approximately 10cM intervals along chromosome 18 (Figure 5.2.1).

SMAD4 lies between D18S851 and D18S878 and is closest to D18S474 (approximately 0.1Mb distal). The forward primers for each marker had a 5' fluorescent tag (either FAM, HEX or TET). An additional marker mapping to 18p (D18S542) was also used to determine whether the whole chromosome was lost or whether there were particular regions targeted for deletion.

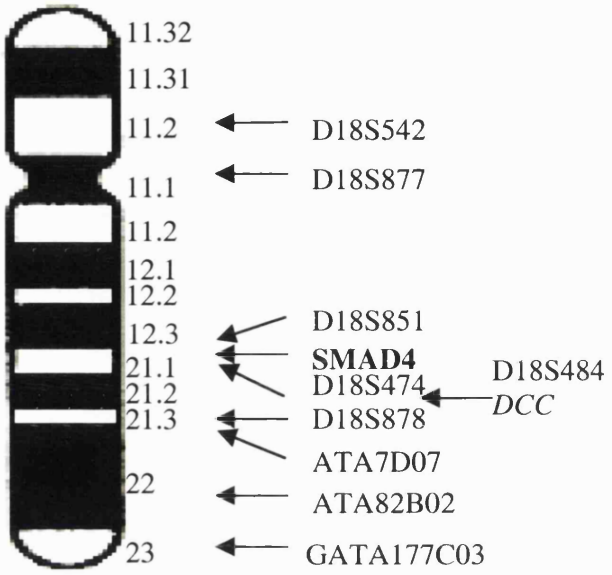


Figure 5.2.1 Positions of each of the microsatellite markers on chromosome 18 used for the LOH analysis.

The position of SMAD4 is shown in bold.

After the initial allele loss screen, additional markers mapping just distal to *SMAD4* were also assessed for loss of heterozygosity (D18S484 and *DCC*) and additional polyp material was available from Families 20 and 21. Standard conditions (1.5mM Mg^{2+}, 55°C annealing for 35 cycles) were used to amplify the tumour/normal pairs, then 0.2µl of the resulting PCRs were combined with Tamra500 size standard and formamide loading buffer before running for 2 hours on 5% Genescan gels (ABI377 prism). Results were analysed using Genotyper software to assign peak areas to all main and stutter bands. Allelic loss was considered if a person was heterozygous for

a marker and the relative ratio of normal:tumour areas was less than 0.5 or greater than 2, given by the formula (N1/N2) / (T1/T2) where N1= sum of normal peak areas (including stutter peaks) for one normal allele, N2 = sum of the peaks for the second normal allele, T1= sum of the tumour peaks for one allele and T2= sum of the tumour peaks for the second allele. The cut off point of less than 0.5 or greater than two allowed for any contaminating normal tissue within the microdissection and did not distinguish between true loss or gain of an allele (but any significant change was considered to represent a loss in this case).

For the initial loss of heterozygosity screen, a total of 46 polyps and cancers from 18 individuals with JPS were studied (Table 5.2.1). Six of the individuals (17, 20.1, 20.2, 20.3, 21.1 and 21.2) all had known constitutional *SMAD4* mutations. Allelic loss of microsatellites near the *SMAD4* locus was detected in multiple polyps from these three families, whereas markers proximal and distal to the SMAD4 locus did not show allele loss. In all cases where a distinction could be made between the wild-type and shared mutant allele, it was the wild-type allele which was lost (see Figures 5.2.2 and 5.2.4). The member of family 17 whose polyps were studied was informative for six out of the eight markers, and in all six juvenile polyps assessed, loss was apparent for one or more microsatellites (Figure 5.2.2). The marker which showed the most frequent loss in juvenile polyps from this patient was D18S474, the marker closest to *SMAD4*. In one polyp (17d) only D18S474 was lost, whereas other polyps showed loss of all the informative markers extended over the long arm of chromosome 18q.

Patient	Polyp no	Site	D18S542	D18S877	D18S851	D18S474	D18S878	ata7D07	ata82B02	Gata177c03
LB*	A	ileal p	○	○	●	■	■	○	○	■
	B	colon p	○	○	○	■	■	○	○	■
	C	colon p	○	○	○	■	■	○	○	■
	D	jp	○	○	○	■	■	○	○	■
	E	jp	○	○	○	■	■	○	○	■
	F	jp	○	○	○	■	■	○	○	■
	G	jejunal jp	○	○	○	■	■	○	○	■
17**	A	jp	○	■	●	●	●	■	●	●
	B	jp	○	■	●	X	X	■	●	○
	C	jp	○	■	●	●	X	■	●	○
	D	jp	○	■	○	●	○	■	○	○
	E	jp	○	■	●	●	X	■	●	○
	F	jp	●	■	○	●	●	■	●	○
12.1***	A	jp	○	○	○	■	○	○	○	○
	B	sb ca	○	○	●	■	X	○	○	○
	C	sb ca	○	○	●	■	X	○	○	○
	D	jejunal ca	X	○	○	■	●	○	○	○
	E	jejuanl ca	X	○	●	■	●	○	○	○
	F	colon ca	X	○	○	■	○	○	○	○
1*	A	ileal ca	○	○	○	○	■	○	○	■
	B	ileal ca	○	○	●	●	■	○	○	■
20.1**	A	rectal p	■	■	●	■	●	■	●	○
	B	colon jp	■	■	●	■	●	■	●	○
	C	colon jp	■	■	○	■	○	■	X	○
	D	colon jp	■	■	●	■	●	■	●	○
	E	colon jp	■	■	●	■	●	■	●	○
	F	colon jp	■	■	○	■	○	■	●	○
	G	colon jp	■	■	●	■	●	■	●	○
	H	colon jp	■	■	○	■	○	■	○	●
	I	colon jp	■	■	○	■	○	■	○	○
20.2**	A	rectal p	■	■	■	○	○	■	○	○
20.3**	A	jp	○	○	■	■	■	○	■	○
21.1**	A	jp	X	●	●	●	○	○	○	○
21.2**	A	jp	■	○	●	X	X	●	○	○
K1-1****	A	jp	○	○	■	■	○	○	■	○
K4-1****	A	jp	○	○	●	X	X	○	○	○
K7-1****	A	jp	○	○	○	○	○	●	○	X
K9-1****	A	jp	■	○	○	■	■	○	○	X
K13-1****	A	jp	■	○	●	X	X	■	○	X
18.1*	A	jp	■	○	■	X	X	○	■	○
18.2*	A	jp	○	○	■	■	■	■	■	X
1262*	A	jp	X	○	■	○	■	○	○	○
	B	jp	X	○	■	○	■	○	○	○
	C	jp	X	○	■	○	■	○	○	○
12.2***	A	sb ca	X	■	■	■	■	■	●	X
12.3***	A	sb ca	X	○	○	●	●	●	■	X

Table 5.2.1: LOH analysis of eight chromosome 18 polymorphic microsatellites, encompassing the SMAD4 locus.

*Shown are results of 46 polyps/cancers from 18 individuals. Open circles indicate retention of heterozygosity, grey circles show uninformative markers and black circles show loss of heterozygosity. Crosses indicate pcr failure. The SMAD4 locus lies between D18sS51 and D18S878.*SMAD4 excluded via screening of gene.**Mutations of SMAD4 responsible for JPS.***SMAD4 excluded via linkage analysis.****SMAD4 mutation status unknown. jp=juvenile polyp (site unspecified unless indicated), p=polyp, sb ca=carcinoma of small bowel (site unknown), ca=carcinoma*

Polyp 17a (patient 17) showed lost all informative markers on 18q, a major event, causing loss of the whole chromosome 18q, led to the loss of the second functional copy of *SMAD4;* this is interesting to observe in a benign lesion, previously considered to be non-neoplastic and without malignant potential. This patient was uninformative for markers just distal to *SMAD4* which might have resolved the breakpoints of the deletions or mitotic recombination etc., particularly in polyps 17b, 17c and 17d (Table 5.2.1). It is apparent that different regions of the chromosome have been lost in different polyps from this patient, indicating that different mechanisms may be involved in the inactivation of the second copy of *SMAD4*.

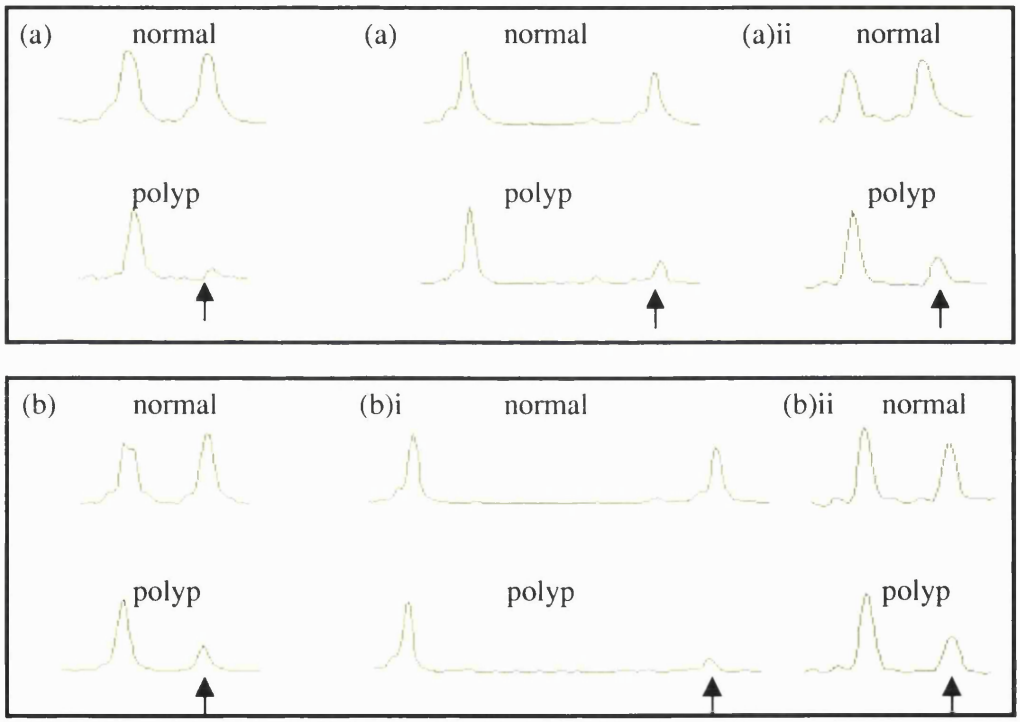


Figure 5.2.2 - Allele loss in Juvenile Polyps.
(a) Electropherogram of microsatellites from patient 17 for (i)D18S851, (ii)D18S878 and (iii) ATA82B02. (b) Electropherogram of microsatellites from patient 20.1 for markers (i)D18S851, (ii)D18S878 and (iii)ATA82b02. The SMAD4 locus lies between D18S851 and D18S878. Allele loss is arrowed.

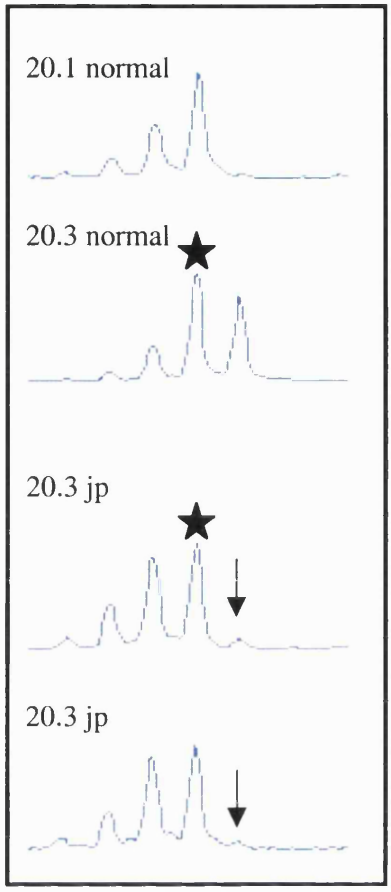


Figure 5.2.3. LOH near SMAD4 (marker D18S484) in polyps from Family 20.
Patient 20.1 (top) was uninformative at D18S484 but two polyps from her sister (20.3) showed LOH (arrowed) at this marker. The shared mutant allele is starred, indicating that the LOH has targeted the remaining wild-type allele.

Family 20 also has a germline *SMAD4* mutation, and patient 20.1 showed LOH of one or more markers in seven of nine juvenile polyps studied. The two markers that encompass *SMAD4* (D18S851 and D18S878) showed loss in five of seven polyps, and where ascertainable, it was the wild-type allele which was subjected to loss. Unfortunately the marker closest to *SMAD4* (D18S474) was uninformative in this individual. The other two polyps which showed loss (20.1f and 20.1h) showed loss

of just one marker each, neither of which was close to *SMAD4*, (these may be false positives). It is probable that these two polyps, plus the two which did not show loss at any other marker (20.1c and 20.1i) have had their second copy of *SMAD4* inactivated by a more subtle mechanism such as point mutation. This would be resolved by screening the polyp DNA by SSCP analysis or direct sequencing of the 11 *SMAD4* exons, but this was considered cumbersome and unnecessary given that one mechanism of 'second hit' had been already demonstrated.

Alternatively, there is the possibility that with only crude microdissection, there was contaminating normal tissue confounding the detection of loss in a fashion which was inconsistent between different markers. For example, some markers have a better amplification efficiency for the smaller of the two alleles in a heterozygote. If it is the smaller allele which represents the wild-type allele, and should therefore be the one subjected to loss, the amplification efficiency may still be better for any contaminating normal tissue and thus true LOH will not be detected. Again, like patient 17, patient 20.1 was uninformative for the two extra markers (D18S484 and *dcc*), but these markers were flanked by two showing loss (D18S851 and D18S878) so would not have helped to resolve any deletion breakpoints. Patients 20.2 and 20.3 from family 20 showed no loss in either of two polyps included in the initial seven-marker LOH screen, but both were poorly informative at many markers. Patient 20.3 was informative at one of the two additional markers (D18S484) and showed loss in

two juvenile polyps studied (Figure 5.2.3), again with the loss targeting the putative wild-type allele.

The third family with a known *SMAD4* germline mutation was Family 21. Initially, only one polyp from each of the two family members was included, and both of these showed loss of heterozygosity. The polyp from 21.1 (21.1A) showed loss of the three markers studied near to *SMAD4* – D18S877, D18S851 and D18S474 (Figure 5.2.4). The polyp from her sister, polyp 21.2.A, showed loss at D18S851 but the PCR repeatedly failed for D18S474 (a less reliable PCR generally). It may be assumed that D18S474 was in fact lost due to the fact that ATA7D07, which lay distal to D18S474, did show allele loss (Figure 5.2.4). Note, however that a polyp from 21.1 did not show loss at ATA7D07, although the peak does seem to be reduced with respect to the shared mutant allele. In addition, when the two extra markers were studied, polyps from both 21.1 and 21.2 showed allele loss at D18S484 which lies between D18S474 and D18S878, and this loss targeted the wild-type allele (figure 5.2.4). Polyps from families 17, 20, 21 who each have a germline *SMAD4* mutation, thus showed consistent allele loss around *SMAD4* in their juvenile polyps.

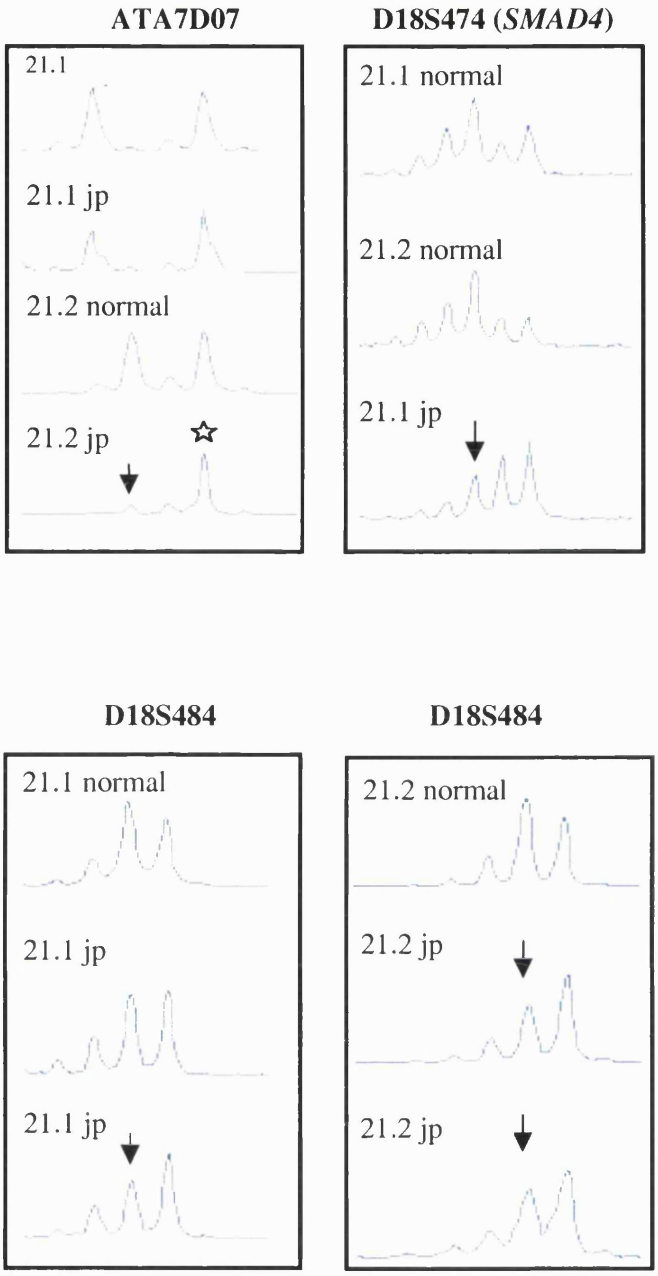


Figure 5.2.4. LOH results from family 2.
Shown are the results from two family 21 individuals - 21.1 and 21.2. Box 1= ATA7D07, note the reduction in height of 21.1 jp and the arrowed LOH of the wild-type allele in 21.2, the shared mutant allele is starred. Box 2= D18S474, LOH is arrowed, the wild-type is not distinguishable. Boxes 3 and 4= LOH results for 21.1 and 21.2 respectively. Note the reduction in one 21.1 jp and arrowed LOH in the other 21.1 jp and arrowed LOH in both 21.2 jps. Also note that the same, smaller allele is lost in the polyps from both individuals, indicating that this is the wild-type.

Eight patients (1, 12.1, 12.2, 12.3, 18.1, 18.2, LB and 1262) had no detectable germline *SMAD4* mutation, as assessed by SSCP, immunohistochemistry, and linkage analysis where appropriate (discussed in Chapter Three). Of 13 JPS hamartomas from these patients, 12 showed no allele loss at any marker mapping to 18q. Loss was observed in one polyp, an ileal polyp from patient LB (LBa), at marker D18S851 (Table 5.2.1), but the threshold of 0.5 was only just crossed and visual inspection of the electropherogram (Figure 5.2.5) suggested this to be a false positive. The patient was uninformative for D18S474 to see whether the loss was really targeting *SMAD4*, but no other marker for this polyp or any other polyp from this patient, showed LOH (Table 5.2.1). Furthermore, the two extra markers just distal to *SMAD4* did not show loss of heterozygosity in any polyp from LB. In addition to the 13 harmartomas from the *SMAD4*-negative group, 8 carcinomas were studied for LOH at *SMAD4*. Of these eight, six lost alleles at or near to *SMAD4* (12.1b, 12.1c, 12.1d, 12.1e, 1b, 12.2a, 12,3a) and two retained all chromosome 18 alleles (12.1f and 1a).

This high frequency of loss at *SMAD4* observed in these cancers most likely reflects the loss of 18q that occurs in a large proportion of colorectal cancers outside JPS (as discussed in Chapter Four) (MacGrogan *et al.*, 1997; Takagi *et al.*, 1996; Thiagalingam *et al.*, 1996). Figure 5.2.6 illustrates this; three affected members of Family 12 (in whom a high loss of 18q was observed in their cancers) did not share any alleles at D18S484 which lay just distal to *SMAD4*, indicating that germline

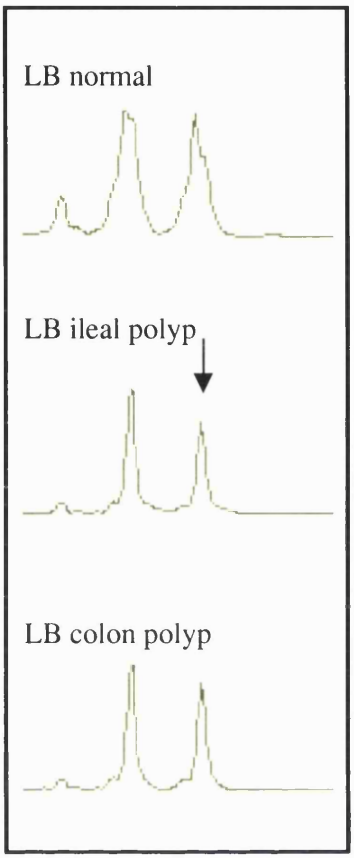


Figure 5.2.5 probable false positive LOH at SMAD4.
Shown are two polyps from LB for marker D18S851, the first of which the ratio of normal:tumour peaks is just less than 0.5, usually an indication of LOH. However, no other markers flanking D18S851, or any other polyp from this patient, showed LOH around SMAD4.

changes at *SMAD4* were not responsible for their JPS. In addition, two separate cancers from patient 12.1 lost the maternal and paternal allele respectively, confirming this was random loss of 18q and not loss of any 'wild-type' allele

(supported by the fact that this family are not linked to 18q and have had no *SMAD4* mutation detected using various methods (discussed in Chapter Three)).

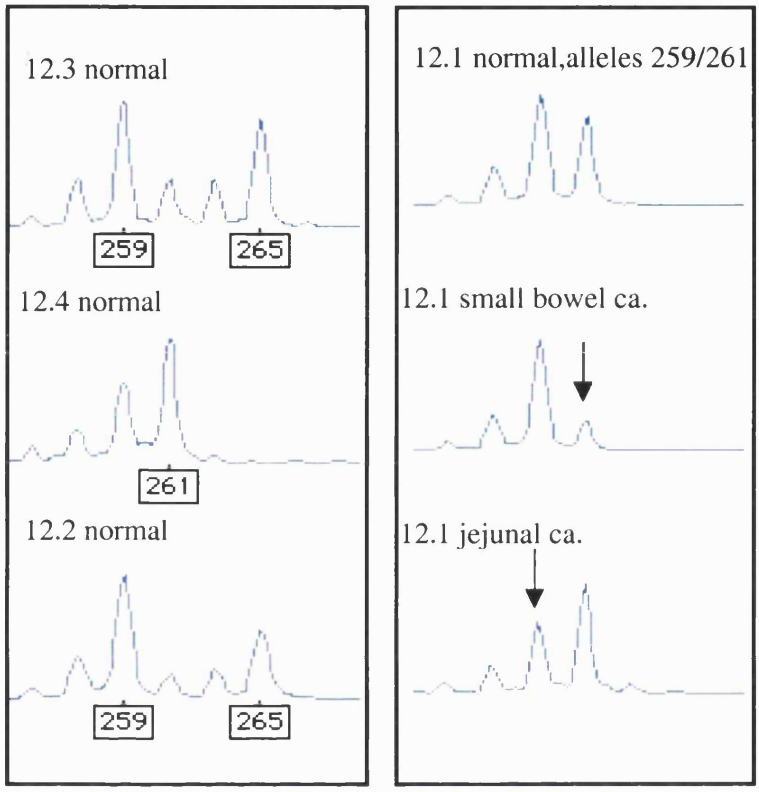


Figure 5.2.6 Marker D18S484 (just distal to SMAD4) in Family 12.
The left hand panel shows the normal results from three affected members of Family 12 (12.2 is the Father of 12.3 and the brother of 12.4 (12.4 was not included in the LOH studies). 12.2 and 12.3 share no alleles with 12.4 indicating that this family is not linked to the SMAD4 region. The right hand panel shows LOH for 12.1 (son of 12.2). Note that two different cancers from this patient show loss of the two different chromosome 18 alleles, once again highlighting that this LOH represents the LOH observed at SMAD4 in a high proportion of colorectal cancers.

Overall therefore, in the patients in whom a germline *SMAD4* mutation has not been detected, plus in some families who have *SMAD4* excluded because of lack of linkage to this chromosomal region, there appears to be little or no *SMAD4* LOH in the juvenile polyps, and a high level of *SMAD4* loss in the carcinomas. Of the five patients who have not had *SMAD4* screened for germline mutations (Table 5.2.1) two showed allele loss near *SMAD4* in their polyps (K4-1 and K13-1). Unfortunately in both cases the marker closest to *SMAD4*, D18S474, failed to amplify, so the region of possible deletion cannot be ascertained. It is quite likely however that the loss observed at D18S851 does reflect a second hit at *SMAD4* and this could be established by screening the constitutional DNA for a germline *SMAD4* mutation in these patients.

This data confirms that 'second hits' at *SMAD4* are not seen consistently in juvenile polyps when there is an absence of a 'first hit' at *SMAD4* in the germline, and that *SMAD4* is targeted for deletion in JPS carcinomas in much the same way as it is lost in sporadic carcinomas (discussed in Chapter four). When there is a known germline mutation at *SMAD4*, however, variable regions of chromosome 18 are lost in the polyps, comprising the 'second' necessary hit to induce the formation of the polyp. The occurrence of second hits at *SMAD4* in the polyps derived from *SMAD4* mutation carriers was confirmed using immunohistochemistry, as discussed in Chapter Three, and no SMAD4 protein was observed. This confirms that loss of the second copy of *SMAD4* does indeed initiate growth of the hamartoma and that

SMAD4 acts as a tumour suppressor gene in JPS, as in does in sporadic colorectal and pancreatic cancers (Hahn *et al.*, 1996c; MacGrogan *et al.*, 1997; Tagaki *et al.*, 1996; Thiagalingam *et al.*, 1996). These data also show that JPS polyps do indeed have a clonal component, and are not simply non-neoplastic lesions.

5.3 WHICH CELLS IN THE JUVENILE POLYP HAVE LOSS OF SMAD4?

With the demonstration of allele loss at *SMAD4*, the next question to address was which cells in the juvenile polyp were actually losing *SMAD4*, that is, which cells comprise the clonal component of the polyp? To answer this, fluorescent *in situ* hybridisation (FISH) was performed on slides cut from the paraffin blocks used for the LOH studies using a probe to which *SMAD4* is known to map. The PAC probe 224_j_22 was originally used by Hahn *et al* to map deletions in pancreatic cancers, with the finding that the region targeted for deletion was indeed *SMAD4* (Hahn *et al.*, 1996b). PAC 224_j_22 was ordered from Human Genome Mapping resources (http://www.hgmp.mrc.ac.uk), the DNA was extracted and then labelled with biotin. After confirming that the labelled products were of suitable size, they were precipitated in the presence of Cot1 competitor DNA and salmon sperm DNA. Before assessing the paraffin slides for *SMAD4* loss, it was first confirmed that the probe did indeed map to 18q21.1 (by hybridising to metaphase spreads) and that it did contain the *SMAD4* gene (by amplifying *SMAD4* exons 2 and 11 (Table 3.2.3, Chapter Three)) directly from picked colonies. After validating the probe target,

labelled and precipitated PAC DNA was hybridised to 9μm paraffin sections which had been protease digested. 9μm thick sections were used in order to ensure that a high proportion of the nuclei under scrutiny were intact, as Thompson *et al* have demonstrated that sections less than 6μm contain almost no nuclei that are uncut (Thompson *et al.*, 1994). The sections used were those from individuals who have a known constitutional *SMAD4* mutation, Families AF, 17, and 20. After appropriate SSC washes, the signal was detected using avidin-FITC and visualised using a cooled charge-coupled device camera and IPlab software to capture images. At least fifty epithelial and inflammatory cells, and thirty stromal cells were counted in each polyp and scored for zero, one, two or three signals. To control for hybridising efficiency, an α-satellite centromere probe was also hybridised to the polyp sections to ensure that the lost region was indeed *SMAD4*. To confirm that the PAC 224_j_22 probe also hybridised efficiently, it was hybridised to normal tissue (appendix and normal colon) sections derived from an individual without juvenile polyposis syndrome.

Using the PAC 224_j_22 probe, only a single probe signal was observed in the epithelial cells of juvenile polyps from *SMAD4* germline mutation carriers. Polyps from Family 17, 20 and AF showed a single signal in 90%, 95% and 82% of the epithelial cells respectively (Figure 5.3.1 *a-c*). Infiltrating stromal lymphocytes showed two signals in over 90% of nuclei from all three polyps (Figure 5.3.1 *f*). An intriguing finding was that stromal fibroblasts and pericryptal myofibroblasts from

each polyp also showed just one signal in between 83% and 90% of cells respectively (Figure 5.3.1 *d* and *e*). The 18 centromere probe showed two signals in the great majority of cells of both stromal (87%) and epithelial origin (85%) in the juvenile polyps. Combined with the observation of two PAC signals in the lymphocytes, this indicated that the loss observed with the PAC probe was indeed a true result. When the PAC was hybridised to the normal colon and appendix sections from an unrelated individual without JPS, two signals were observed in the vast majority of stromal and epithelial cells (94% and 90% respectively, Figure 5.3.1 *h*), again substantiating that the PAC hybridised efficiently and the single signals observed in the juvenile polyp slides were a true reflection of cell content. To ensure that the cells counted were of the origin indicated by their morphology (stromal, epithelial or inflammatory infiltrate), Giemsa staining of the sections was performed and the slides examined by light microscopy. Furthermore, antibodies particular to these cell types were hybridised to serial sections of the JPS polyps (performed by Histopathology Dept., ICRF). All antibodies (AuA1, MNf116, SmAct, Desmin, CD45 and CD31) confirmed the cells to be of the expected origin.

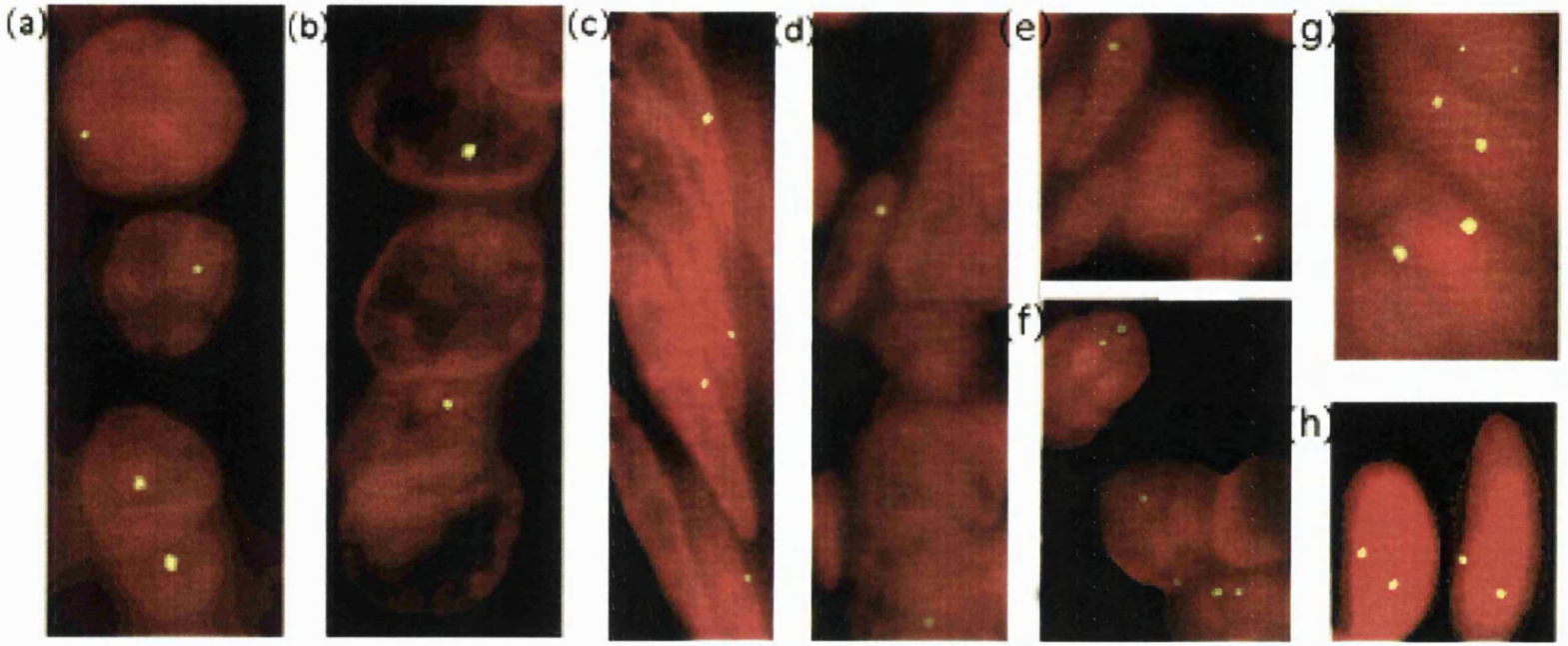


Figure 5.3.1. Results of FISH studies.

PAC 224_j_22 was used as a probe on juvenile polyps. A single probe signal was observed in: (a) epithelium of patient AF (61/64 cells counted); (b) epithelium of patient 17 (72/80 counted); (c) epithelium of patient 20 (41/50 counted); (d) myofibroblasts from juvenile polyps, shown is patient 17 (25/30 counted); and (e) stromal fibroblasts of juvenile polyps, shown is patient AF (27/30 counted). Two PAC probe signals were observed in infiltrating lymphocytes of the juvenile polyp, (f) patient 17 (46/50 counted). Two 18cen probe signals were observed in epithelial cells of juvenile polyps, (g) patient AF (41/48 counted). Two PAC signals (88/98 cells counted) were observed in normal colon epithelium from an unaffected individual, (h). The figure cannot show signals in all cells because of the different focal planes.

Contrary to previous studies (Jacoby *et al.*, 1997; Kinzler and Vogelstein, 1998; Veale *et al.*, 1966), the LOH and FISH results show that the juvenile polyps in JPS are not simply stromal lesions with an overgrowth of connective tissue or infiltration of lymphocytes. Using FISH, loss of *SMAD4* was observed in the epithelial cells from JPS polyps in patients with known germline *SMAD4* mutations. FISH found no loss of *SMAD4* in the stromal lymphocytes but did find loss in the less numerous stromal fibroblasts and peri-cryptal myofibroblasts. This result shows that the epithelium of JPS polyps is clonal and that part of the stroma is also derived from the same clone, contrary to accepted histological theory. Loss of *SMAD4* in both epithelium and stroma can also explain how microsatellite analysis can readily and consistently detect allele loss at multiple markers on 18q in crudely microdissected polyps, despite the presence of contaminating inflammatory infiltrate. It cannot entirely be excluded that the stromal and epithelial cells in JPS polyps arise from different clones that have independently lost *SMAD4*, which then results in growth of the polyp.

Evidence from other diseases however, suggests that lesions containing multiple lineages (i.e. epithelial and mesenchymal elements) may also be clonal. For example, Green *et al* showed that hamartomas (including renal angiomyolipomas, fibromas and skin hamartomas amongst others) from TSC patients had skewed X-chromosome inactivation patterns (Green *et al.*, 1996), whereas all control tissues studied had random X-inactivation patterns. These experiments supported the finding of LOH in TSC hamartomas around the *TSC* genes, proving that the *TSC* genes act as classical tumour suppressor genes (Sepp

et al., 1996), and showed that despite being comprised of different cell types, TSC hamartomas are clonal lesions. Abeln *et al* investigated the clonality of MMMTs (malignant mixed Müllerian tumours), which occur in the female genital tract (Abeln *et al.*, 1997). By studying 74 polymorphic markers spread over 19 different chromosomes, LOH was demonstrated in five of six MMMTs. Importantly, for these experiments the stromal and epithelial cells were microdissected separately and in all cases loss of the same allele was observed in both tumour components, indicating a monoclonal origin. Neo-differentiation of melanoma cells into stromal vascular channels has also been demonstrated (Maniotis *et al.*, 1999), suggesting that these malignant cells somehow reverse their terminal differentiation and once again become pluripotent. Taken together, these experiments indicate that the precursors of the juvenile polyps (or TSC hamartoma etc.) are either laid down very early in development before epithelial:mesenchymal differentiation and then undergo clonal expansion, or the second somatic mutation occurs later on in a stem cell with a degree of plasticity that allows differentiation into more than one cell type.

What is certain is that juvenile polyps are clonal lesions and the clonal component certainly includes the epithelium, making the paradox of how a stromal lesion can give rise to an epithelial malignancy, redundant.

Previously, after detecting a germ-line deletion of 10q in a patient who had juvenile polyps and multiple congenital abnormalities, Jacoby *et al* used allele loss and FISH analysis to show somatic deletion of 10q22 in juvenile polyps

(Jacoby *et al.*, 1997). Most of the patients in this study were less than 10 years of age, and given that the onset of Cowden disease (caused by the *PTEN* gene on 10q) manifestations do not usually occur before the second decade there was the possibility that a proportion of these patients were in fact CD and not JPS. Whilst there may have been doubt about the diagnosis in a subset of the JPS patients in Jacoby's study, it is now probable that some of the patients had germline mutations in *BMPR1A* and the observed LOH represents the second hit at this locus. This gene has just been shown to cause another subset of JPS cases (in addition to *SMAD4*) (Howe *et al.*, 2001) (discussed in Chapter Nine). Although the LOH data of Jacoby *et al* now has a reasonable standing, the FISH analysis that described deletions in the lymphocytes is still questionable (Jacoby *et al.*, 1997). To suggest that the inflammatory lymphocytes are the clonal component of the juvenile polyp implies that these lesions are lymphomatous neoplasms. Explaining how an epithelial malignancy may arise in this type of lesion is difficult, even for the 'landscaper hypothesis'. What is more likely is that there were methodological problems that gave rise to these results. As mentioned earlier, it has been demonstrated that when using FISH analysis, sections of less than 6μm contain almost no uncut nuclei (Thompson *et al.*, 1994), therefore it may be inappropriate to conclude loss has occurred in the 5μm sections of Jacoby *et al*, especially if the cells comprising the tumour vary in size from normal tissue and/or if stromal and epithelial cells are of different sizes. Furthermore, the 10q FISH probe was not hybridised to control sections to assess the hybridisation efficiency, and a chromosome 21 control probe was used on the juvenile polyps that may itself have undergone changes in the polyps or be

subject to important differences from the 10q probe. Doubt must therefore be cast on the conclusion that the lymphocytes are the proliferating component of the polyps.

5.4 CONCLUSIONS

The LOH and FISH experiments targeting *SMAD4* in juvenile polyps have indicated that loss of the wild-type copy of *SMAD4* initiates tumorigenesis, and the epithelium is intimately involved in the formation of the hamartoma and its subsequent progression to cancer. The classical categorisation of juvenile polyps as stromal lesions is therefore probably incorrect. In addition to the loss of *SMAD4* observed in the epithelial cells, stromal fibroblasts and peri-cryptal myofibroblasts also showed only a single copy of *SMAD4*. This findings are even more surprising and suggest that a single cell is able to give rise to more than one cell type in the JPS polyps, but this does also appear to be the case for other tumours (Green *et al.*, 1996) and, indeed, normal tissue (N. A. Wright *et al.*, personal communication). The results certainly show that there is no need to invoke the 'landscaper' hypothesis in juvenile polyposis to explain how 'stromal' lesions develop into epithelial malignancies. Juvenile polyps are truly pre-malignant clonal lesions, with the epithelium playing a central role in the development of the epithelial malignancies observed in JPS. If the role of *SMAD4* gene were to be categorised according to the Kinzler and Vogelstein's

classifications (Kinzler and Vogelstein, 1997), it would certainly be a 'gatekeeper' in both the epithelium of JPS polyps and sporadic cancers.

CHAPTER SIX

THE ASSESSMENT OF CANDIDATE GENES IN JUVENILE POLYPOSIS SYNDROME

THE ASSESSMENT OF CANDIDATE GENES IN JUVENILE POLYPOSIS SYNDROME

6.1 INTRODUCTION

The importance of the transforming growth factor-beta (TGFβ)-signalling pathway in colorectal cancer has previously been well established, with TGFβ itself having both tumour suppressor and tumour promoting properties (reviewed in (Gold, 1999)). In the normal epithelium and early in tumour progression, TGFβ is responsible for maintaining homeostasis by exerting anti-proliferative effects, thereby having tumour suppressing activity. This is via modulation of the transcription of target genes that determine the cell phenotype (Hata *et al.*, 1998). Disruption of the TGFβ-signalling pathway by loss of one of the pathway components can lead to escape from TGFβ mediated growth control (de Caestecker *et al.*, 2000). Conversely, TGFβ has oncogenic tumour promoting properties in late-stage or advanced tumours, indicated by high levels of expression of TGFβ when there is a downstream defect in the pathway (Massague *et al.*, 2000). Alternatively, if the pathway is intact, oncogenic Ras mutations have been shown to disrupt TGFβ signalling by inhibiting the TGFβ induced nuclear accumulation of downstream targets (SMAD2 and SMAD3) (Kretzschmar *et al.*, 1999). Altered levels of type 1

and type II receptor levels can also compromise the tumour suppressor activity of TGFβ, enabling the ligand to act on accessory cells or even the tumour cells themselves to promote tumorigenesis and metastasis (Reiss, 1999). In keeping with this, a large proportion of sporadic colorectal cancers with microsatellite instability and many cancers seen in HNPCC have TGFβIIR mutations which are predicted to result in a truncated protein (Markowitz *et al.*, 1995; Parsons *et al.*, 1995). Elucidating the TGFβ pathway components and interactions have shown that there exist many molecules that are important, and others which may indeed be important, in tumour progression.

The TGFβ-signalling at its simplest consists of a receptor complex that activates SMADs and these activated SMADs controlling the transcription of target genes (Massague, 1998). The *SMAD* genes, which were first identified through the Drosophila and *Caenorhabditis elegans* proteins MAD (Mothers Against Decapentagplegic) and SMA (Small body size), therefore act as signal transducers (Massague, 1996). The signalling cascade begins at the cell membrane where the TGFβ ligand binds to two related serine/threonine kinase receptors, the type I and type II receptors. Signalling is initiated when the ligand (TGFβ) binds to the type II receptor which then recruits and phosphorylates, thereby activating, the type I receptor in a heteromeric complex (Massague, 1996). The type I receptor is then able to target and phosphorylate the downstream substrates SMAD2 and SMAD3. These two SMADs (there are eight in total in humans) are known as the receptor-

regulated SMADs, or R-SMADs (Nakao *et al.*, 1997b) and are phosphorylated on the last two serines of a conserved C-terminal SSXS motif. SMAD2 and SMAD3 are normally anchored in the cytoplasm by the protein SARA (SMAD Anchor for Receptor Activation) which helps to present the substrates to the TGFβIR (Wrana and Attisano, 2000) and masks the nuclear import signal (Xu *et al.*, 2000). Phosphorylation at the carboxy-terminal residues of the R-SMADs reduces the affinity for SARA and increases the affinity for the common-mediator SMAD, SMAD4, which then forms heteromeric complexes with either SMAD2 or SMAD3 (Kawabata *et al.*, 1998; Souchelnytskyi *et al.*, 1997). With the nuclear import signal unmasked upon phosphorylation of the R-SMADs, the SMAD2/4 or SMAD3/4 heteromeric complexes are able to translocate to the nucleus. The complexes then associate with transcription factors such as FAST1 which facilitates the transcription of target genes (Liu *et al.*, 1997), elicited ultimately in response to TGFβ. (figure 6.1.1).

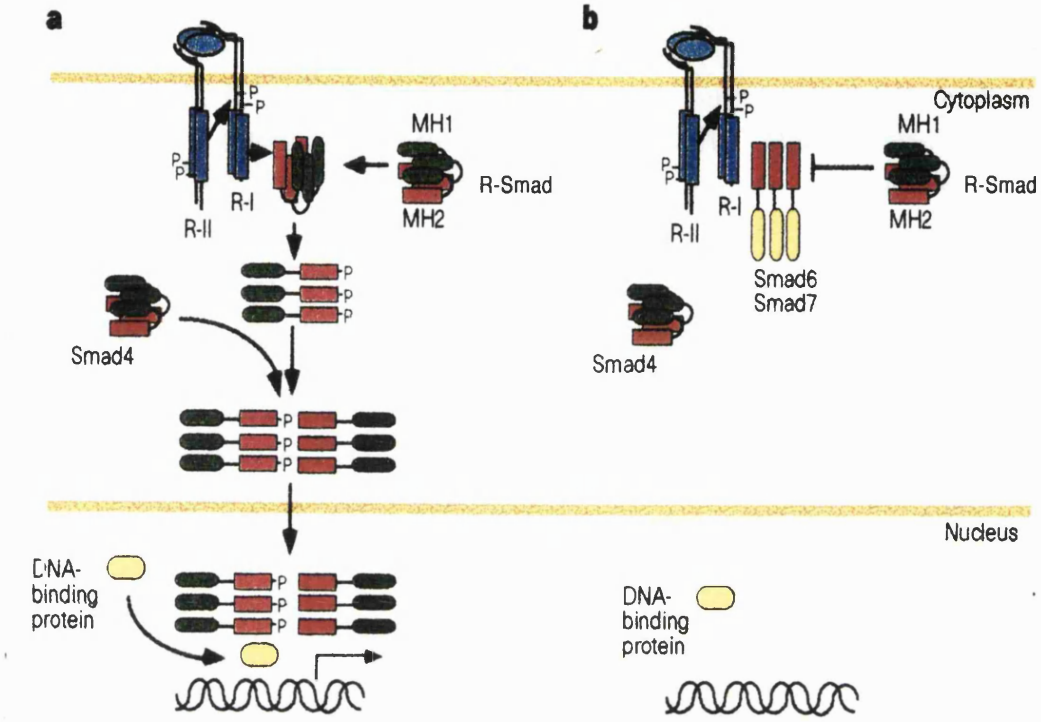


Figure 6.1.1 TGFβ/Activins/BMP signalling pathway.
(a) Binding of the ligand leads to the assembly of a heterotetrameric complex, where the type II receptor recruits and phosphorylates the type I receptor. The receptor regulated SMADs (also known as the R-SMADSs, 2 and 3 for TGFβ/activin, 1 and 5 for BMP) are then phosphorylated which allows them to form heteromeric complexes with the common mediator, SMAD4. This complex translocates to the nucleus, where it associates with DNA binding proteins or transcription factors to facilitate the transcription of target genes. (b) The inhibitory SMADs, (6 and 7) bind the receptors and prevent the phosphorylation of the receptor regulated SMADs. Alternatively they may compete with SMAD4 for the phosphorylated R-SMADs. (Figure based on (Heldin et al., 1997))

The inhibitory SMADs, SMAD6 and SMAD7 as the name suggests, acts as antagonists to TGFβ signal transduction by either blocking the phosphorylation of SMAD2 and SMAD3 by TGFβIR, or by competing with SMAD4 for the phosphorylated SMAD2 and SMAD3 (Imamura *et al.*, 1997; Nakao *et al.*, 1997a; Roberts, 1999). Transcription of SMAD7 has also been shown to be under the control of TGFβ, indicating a self-regulating negative feedback loop for TGFβ responses (Stopa *et al.*, 2000).

TGFβ is the founding member of a large superfamily of related growth and differentiation factors that include bone morphogenetic proteins (BMPs) and activins (Wrana, 1998). As in the TGFβ-signalling pathway, the BMPs and activins transduce their signal though pairs of transmembane serine/threonine kinase receptors (type I and type II) (Onichtchouk *et al.*, 1999). The SMADs also play a central role in transducing signals from the membrane to the nucleus (Itoh *et al.*, 2000) where, as in TGFβ signalling, they modulate the transcription of target genes. In all three pathways (TGFβ, Activin and BMP) SMAD4 acts as the common mediator by forming complexes with activated R-SMADs (Figure 6.1.1). Unlike TGFβ and activin however, where the R-SMADs are SMAD2 and SMAD3, the BMP pathway utilises SMADs 1, 5 and 8 as the R-SMADs (Yamamoto *et al.*, 1997; Zhu *et al.*, 1999) although there does appear to be some promiscuity as the activins also use SMAD1 (Chen and Massague, 1999).

6.2 DO GERMLINE MUTATIONS IN SMADS 1, 2, 3 OR 5 CAUSE JPS?

Somatic deletions of the *SMAD4* gene had already been found in a large proportion of pancreatic cancers (Hahn *et al.*, 1996b), when the genes importance was highlighted by the demonstration of germline mutations in juvenile polyposis (Howe *et al.*, 1998b) (discussed in detail earlier). This once again indicated the significance of TGFβ pathway disruption in tumorigenesis. Evidence that inactivation of *SMAD2* and *SMAD3* (which like SMAD4 may abrogate or affect the signalling pathway) were also significant was provided by several workers. Firstly, *SMAD2* is in close proximity to *SMAD4* on chromosome 18q21.1, and this region is found to be somatically lost in a large proportion of colorectal cancers (Vogelstein *et al.*, 1988). Furthermore, somatic *SMAD2* mutations have been found in colorectal and other cancers (Eppert, *et al.*, 1996; Thiagalingam, *et al.*, 1996), and mutations have been described in tumours which prevent SMAD2-SMAD4 interactions (Hata *et al.*, 1997), or increase the degradation of mutant SMAD2 protein through a ubiquitin-mediated pathway (Xu and Attisano, 2000), potentially affecting TGFβ signalling. *SMADs 2* and *3* are structurally similar, and evidence for a of SMAD3 role in tumorigenesis was provided by a knockout mouse model (Zhu *et al.*, 1998). The *Smad3* mutant mice, unlike their *Smad2* and *Smad4* mice counterparts, were viable in the homozygous state and developed metastatic colorectal adenocarcinomas at the age of 4-6 months.

Based on this evidence, *SMADs 2* and *3* appeared to good candidates for JPS. *SMADs 1* and *5* (the equivalent R-SMADs in the BMP pathway) being structurally similar and having a similar role were also considered to be candidates (albeit weaker) than SMADs 2 and 3. *SMADs 1, 2, 3* and *5* were therefore screened for germline mutations in all available JPS patients.

Genomic DNA was available or extracted from blood samples from 30 JPS patients (patients 1, 5, 6, 9, 10, 12, 14, 16, DM, FT, MD, SM524, 1868 who all had a family history of JPS, and BN, c2, c3, c4, CV, CWN, KS, MTW, SM106, SM316, SM397, WB, ZB, 374, 1204, 1262, and 1469 all of whom had no known relative affected with juvenile polyps, though some had relative(s) who had developed colon carcinoma). None of the patients had clinical features suggestive of Cowden, Gorlin, or Bannayan-Zonana syndromes. Germline mutations in *SMAD4* had previously been excluded in this group of patients by screening all exons and exon/intron boundaries using conformation specific gel electrophoresis (CSGE) (Houlston *et al.*, 1998).

For the mutation screen, oligonucleotides were designed to amplify each exon of the *SMAD2, SMAD3* and *SMAD5* genes (including exon/intron boundaries) using the polymerase chain reaction. Primer pairs and annealing conditions for each exon are shown in Table 6.2.1. The genomic structure of *SMAD1* was not available and therefore cDNA was prepared from the available RNA of 10 patients, extracted from

either fresh-frozen normal tissue using TRIzol reagent (patients SM524, 5, 1204, 1262, 1469) or from cell lines with the Fast track kit (patients CV, CWN, FT, MD, MTW). Oligonucleotides were designed for nested amplification of the *SMAD1* gene in four parts as shown in Table 6.2.1. *SMADs 1* and *3* were screened for mutations using a combination of single stranded conformation polymorphism analysis (SSCP) on large, self poured plates followed by direct sequencing of the PCR products. *SMADs 2* and *5* were screened by conformation specific gel electrophoresis (performed in collaboration with Dr. Richard Houlston of the ICR).

SMAD	Exon/region	Sense primer (5'-3')	Antisense primer (5'-3')	Size (bp)	Ta
1	67–901	tgctgactgggttacttttta	aagcaggtgggggcgtatca	835	55°C
	162–543	caaatctcttctgctgtcct	ttccagtggttttagttcat	382	53°C
	497–870	gtcatttactgccgtgtgtg	gaaagggcttcctgggtctg	374	55°C
	799–1714	ctctcccaatagcagttacc	gctcctttgtcagttctcaa	916	54°C
	838–1249	cagcagcacctaccctcact	cctcccctccaacataataa	412	55°C
	1229–1704	attggaaaaggagttcatct	cagttctcaatcgtgtctga	476	52°C
2	1 part 1	tccgctccctccgtcttccatac	cgcgcgcccgcagccctacc	340	67°C
	1 part 2	gacggcggccgggagtgttt	gcgggcgcccaggctttacc	168	63°C
	2	gtgaaggaagtattctgta	aatgctatgccttattttac	183	50°C
	3	tttacatcatggtattttg	attttacattaaggaaacat	282	47°C
	4	aatttagcccatttgactgc	gctattccaagaaacagata	473	48°C
	5	ttggatttcttgaacttttt	aacttgaatgcttatgaaca	225	49°C
	6	gctgtgcttgatttgtttta	atgcgtctcaacttctctaa	214	47°C
	7	tttttaaatccttttgtttt	ttatttggctattcattagg	199	46°C
	8	aatctatttttggcttgaat	aatgcctacattatgagtat	342	47°C
	9	ctcatttgtattttgtttca	gttgacatgataggtttatg	198	48°C
	10	atattctaaaacttgtaacc	agaatgcaatgaaacataat	290	47°C
	11	ctgcctgtggacttgaat	tcttgaacttttggatag	150	48°C
3	1	gtcgtccatcctgcctttca	tggtgatgcacttggtgttg	137	55°C
	2–3 part 1	atggccggttgcaggtgtcc	aggcaggccaggcagcatac	205	59°C
	2–3 part 2	cccccggacagttctacctc	tgccgcccacgtgcctacct	183	58°C
	4	gaccaccttccttctgattc	atgaccctgcatgactgacc	120	55°C
	5	tgtctcacctcgcaggttct	tgcacaaggagatactcacc	80	55°C
	6	gtagcccaccctctgtccac	agccacccataccgatgtg	250	58°C
	7	gaggcgtgcggctctactac	tgcctgtgcggctcgtttac	155	58°C
	8	gccctgtttctgtgtttttg	aggcagcacccataactgac	206	55°C
	9	cccacccctttccctatt	aagacacactggaacagc	150	58°C
5	1 part 1	tctccgaagatttgtgtcaa	ctaaagatctgggaatagtg	237	50°C
	1 part 2	aggacagccgagtaaatgtg	tttttccaaattcttctcag	260	50°C
	2	gacttttgatttttgttttt	tgggagctgaaatggacttc	274	51°C
	3	agattttaattattattttt	atgaagtgagtanttctctt	173	45°C
	4	tctgtgtctggtttgttcac	attaaatgtaggaaaatgac	272	47°C
	5	ttttaaggtgttcatctgta	tgttagaggtcacaactcac	284	50°C
	6	aagagggatttgtgatgata	ttaaaacaagtccactaaca	227	49°C

Table 6.2.1 Primers and PCR annealing temperature for SMAD amplifications. *Cycle conditions: four minutes 94°C initial denaturation; 94°C for 30 seconds, 0.5–1 minute at specified temperature (Ta), 72°C for 0.5–1 minute for 35 cycles; and a final extension step of seven minutes at 72°C. For SMAD1, the gene was amplified in two segments initially (67–901, 799–1714, GenBank U59912), followed by two nested PCRs within each of those regions as shown.*

No mutations in JPS patients were observed in *SMAD*s *1*, *2* or *5*. Two band shifts on SSCP were observed in *SMAD3*, both in the fragment for exons 2 and 3. Upon sequencing of new PCR products one of the aberrant shifts was found to be the previously reported G-A change at the third base of codon 103 (Arai *et al.*, 1998). This nucleotide change is synonymous (alanine-alanine) and therefore not considered to be of any significance with regard to disease status.

The second band shift was characterised as a previously unreported polymorphism G-A change at codon 89 in exon 2 of *SMAD3*, conferring an amino acid change of cysteine to tyrosine (Figure 6.2.1). Being an amino acid change, this polymorphism was considered to be of potential pathogenic importance and so investigated further. Presence of the polymorphism was confirmed by amplifying a 490 base pair fragment using the forward primers of exon2-3 part one, and the reverse primer of exon2-3 part 2, followed by restriction enzyme digested with Fnu4H1. This enzyme recognises the sequence GCnGC, therefore cut the wild-type allele but not the mutant allele. Protein analysis (performed by Mike Sternberg, ICRF) showed that cysteine 89 resides in a beta-pleated sheet and is therefore potentially involved in maintaining the protein secondary structure through a di-sulphide bridge. In addition, comparison of the protein sequences with the other human SMADs showed this cysteine to be conserved in SMAD1, 2, 3 and 5. Evolutionary conservation of this residue was also found in the SMAD3 homologues of the rat, mouse, Drosophila, and *C.elegans*.. Using Fnu4H1 restriction enzyme analysis, the

genotypes were obtained for all the JPS patients and a cohort of 25 control individuals. The genotypes for each JPS patient and the frequencies obtained for each genotype are shown in Tables 6.2.2 and 6.2.3 respectively. Where available, additional affected family members were typed to examine whether the mutant allele co-segregated with the disease. The frequencies of the mutant allele did not differ significantly between the JPS group and the control group (Fishers exact >0.5). In addition, only one of two affected siblings in Family 1 carried the mutant allele (Table 6.2.2), showing that this amino-acid change was not associated with the development of disease. However, given the potential functional significance of loss of a cysteine residue, this polymorphism in *SMAD3* cannot be entirely ruled out as a minor colorectal cancer predisposition allele. This is possible to address by genotyping the collection of colorectal cancer cell lines and / or sporadic colorectal cancers, and comparing the frequencies to a control group.

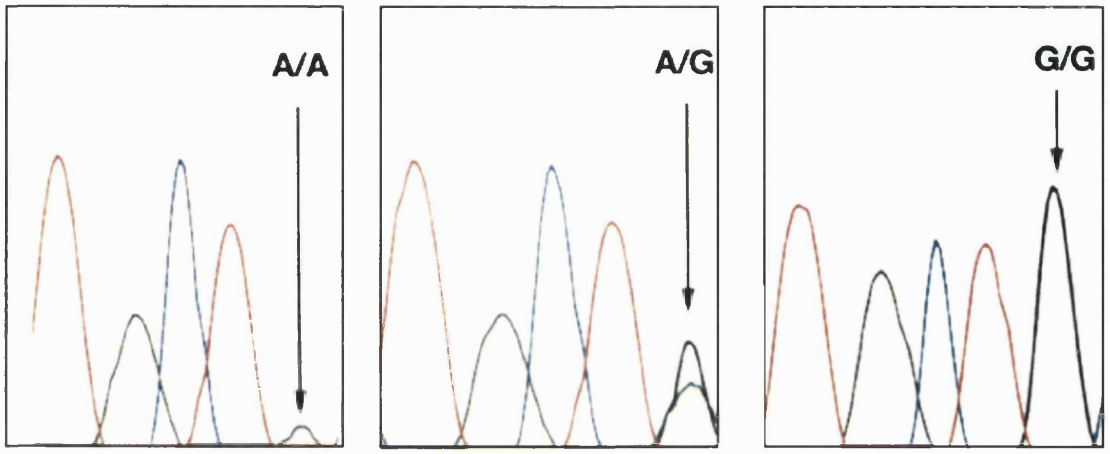


Figure 6.2.1 Codon 89 polymorphism of SMAD3.
Shown are the three possible genotypes, The mutant homozygote (a/a) is shown in the left hand panel, a heterozygote (g/a) in the middle panel and a wild-type homozygote (G/G) is shown in the right hand panel. The G-A nucleotide changes confers a missense amino acid change from cystein-tyrosine

Cohort	*G/G (frequency)*	*A/G (frequency)*	*A/A (frequency)*	*G allele (frequency)*	*A allele (frequency)*
JPS	*26/31 (0.83)*	*4/31 (0.13)*	*1/31 (0.03)*	*56 (0.90)*	*6 (0.10)*
Controls	*18/25 (0.72)*	*7/25 (0.28)*	*0/25 (0)*	*43 (0.86)*	*7 (0.14)*

Table 6.2.3 Genotype frequencies for SMAD3 exon 2 polymorphism.
There was no significant difference in allele frequencies between the JPS group and the control group, Fishers exact >0.5

Patient	*Genotype*	*Genotype of family member*
1	*2*	*3*
5	*3*	*3*
6	*3*	*3*
9	*3*	*3*
10	*3*	*3*
12	*3*	*3*
14	*3*	*3*
16	*2*	*n/a*
374	*3*	*n/a*
1204	*2*	*n/a*
1262	*3*	*n/a*
1469	*3*	*n/a*
1868	*2*	*2*
BN	*3*	*n/a*
c2	*3*	*n/a*
c3	*3*	*n/a*
c4	*3*	*n/a*
CV	*0*	*n/a*
CWN	*3*	*n/a*
DM	*3*	*3*
FT	*3*	*n/a*
KS	*2*	*n/a*
MD	*3*	*n/a*
MTW	*3*	*n/a*
sm106	*3*	*n/a*
sm316	*1*	*n/a*
sm397	*3*	*n/a*
sm524	*2*	*n/a*
sm96	*3*	*n/a*
TT	*3*	*n/a*
WB	*3*	*n/a*
ZB	*3*	*n/a*

Table 6.2.2 SMAD3 exon2 genotypes for codon 89 polymorphism.
The genotype is shown as 0. 1, 2 or 3 where 0= failed, 1=homozygous for mutant allele (a/a), 2= heterozygous for wild-type/mutant alleles (g/a) and 3= homozygous for wild-type allele. The third column shows the genotype of an affected family member where DNA was available.

Germline mutations in *SMADs 2, 3* and *5* were thus excluded as the causative gene in our juvenile polyposis syndrome cohort. There is also evidence that mutations in *SMAD1* are also not involved in the pathogenesis of JPS. Given that cDNA was only available from 10 patients, however, it cannot be decisively concluded that *SMAD1* mutations do not cause JPS. *SMADs 6* and *7* (and the recently characterised *SMAD8*) were not screened for mutations in this study. These genes have an inhibitory effect in the pathway (as discussed above) and were therefore not considered to be as good candidates. At the time of the study, genomic sequence of *SMADs 6* and *7* were not known. Primers were designed to amplify *SMAD6* from cDNA, but despite repeated attempts at optimisation, successful amplification was not achieved. The results obtained suggest that *SMAD4* is the only member of the SMAD family that appears to be mutated in juvenile polyposis.

There are several possibilities to explain why *SMAD4* appeared to be the only component of the TGFβ-pathway which is disrupted in JPS. Whilst there appears to be a second common mediator of TGFβ signal transduction in *Xenopus* (Xsmad4β), in humans SMAD4 is the only known mediator. It is thus conceivable that the presence of SMAD4 is critical for correct signalling. For the receptor regulated SMADs (2 and 3 for TGFβ) however, there may exist a level of redundancy which means that disrupting either one of the two is not sufficient to abolish signalling. This compensatory ability may also apply to SMADs 1 and 5 which are responsible for transducing the signals in the BMP pathway from the membrane to the nucleus

via binding to SMAD4, and for SMADs 6 and 7 which are both inhibitory molecules. The mouse knockout model for *Smad3* did develop colon carcinoma (Zhu *et al.*, 1998). However, the mouse was viable in the homozygous state, suggesting that functional *SMAD3* is not an absolute requirement for development. It is possible that the development of carcinomas in the *Smad3* knockout mouse was due to a gross imbalance in the TGFβ pathway where SMAD2 or other molecules were not able to fully compensate. This would potentially have downstream consequences on the transcription of target genes controlling homeostasis. Further evidence against a major role for SMAD3 in colorectal cancer has been provided by Arai *et al* who studied fifty colorectal cancers (sporadic and hereditary non-polyposis) and Miyaki *et al* who studied 176 colorectal tumours, with both groups failing to find any SMAD3 mutations (Arai *et al.*, 1998; Miyaki *et al.*, 1999).

Although there have been a few reports of *SMAD2* mutations in colorectal cancer (Eppert *et al.*, 1996), these are relatively infrequent compared to the number of *SMAD4* mutations. For example Miyaki *et al* screened 176 colorectal tumours of differing stages and found one *SMAD2* mutation and 21 *SMAD4* mutations (accompanied by loss of the second allele) (Miyaki *et al.*, 1999). In addition, no *SMAD6* or *7* mutations were found in this group of tumours. Tarafa *et al* similarly detected no *SMAD2* mutations and no loss of SMAD2 expression in twenty seven colorectal tumours (Tarafa *et al.*, 2000).

Mutations of *SMADs 2* and *3* (as well as *SMADs 6* and *7*) have also been excluded as important in pancreatic cancer, demonstrated by the existence of no mutations in any of these genes in 13 low passage cell lines from pancreatic carcinomas, (Jonson *et al.*, 1999). Of these 13 cancer lines, four had mutations in *SMAD4* and one had *SMAD4* homozygously deleted, a case similar to juvenile polyposis where *SMAD4* is mutated and critical in tumorigenesis. In support of our results, Roth *et al* also found no germline mutations in *SMADs 2*, *3* or *7* in their cohort of juvenile polyposis patients, consisting of four unrelated kindreds and three sporadic cases (Roth *et al.*, 1999). Overall, *SMAD2* mutations are found in less than ten percent of colorectal tumours (Barbera *et al.*, 2000), this itself probably being an overestimation. When *SMAD2* mutations do occur however, they have been shown to interfere with TGFβ-signalling. A *SMAD2.P445H* missense mutant detected in a colorectal tumour was found to be unable to dissociate from the type I receptor once phosphorylated, and blocked the accumulation of wild-type SMAD2 (Prunier *et al.*, 2001), thus inhibiting SMAD2-dependent transcription.

It would therefore seem that *SMAD2* and *SMAD3* (or *SMADs 1,5, 6* and *7*) are not mutated in the germline of juvenile polyposis patients and are rarely mutated in colorectal cancer. Although apparently pathogenic when they do occur, mutations in the *SMADs* other than *SMAD4* are probably weak targets for selection because of probable redundancy of the partner SMAD. *SMAD4* appears to be the only member of the SMAD family that is a target for deletion or mutation in juvenile polyposis

and colorectal cancer, probably because of its central and individual function in the signalling pathway.

6.3 DO MUTATIONS IN CDX2 ACCOUNT FOR JPS OR PJS?

The caudal-type homeobox gene *CDX2* on chromosome 13q12.3 (Genbank accession numbers AF00384/5/6) encodes a transcription factor whose expression is restricted to the intestine, with highest levels in the proximal colon and rectum (Drummond *et al.*, 1997). The protein is thought to play a role in differentiation and proliferation of intestinal epithelial cells, based on evidence that in the distal colon *CDX2* transcripts were found to be highest in the undifferentiated cells at the bottoms of the crypts, whereas protein levels were highest in the mature cells in the upper half of the crypts (Lorentz *et al.*, 1997). *Cdx2* knockout mice generated to investigate the phenotypic features associated with loss of this homeobox gene were not viable as homozygotes, dying between 3.5 and 5.5 days post coitum. The heterozygous *Cdx2* mutants showed a variable phenotype including tail abnormalities, stunted growth, homeotic shifts of the vertebrae and malformations of the ribs (Chawengsaksophak *et al.*, 1997). A feature common to 90% of the heterozygous mice was the development of multiple (1-10) intestinal hamartomatous polyps, particularly in the proximal colon (Chawengsaksophak *et al.*, 1997; Tamai *et al.*, 1999). The hamartomas in the heterozygous mice were found not to express *Cdx2*, though they did not have loss of heterozygosity of the second allele,

suggesting haploinsufficiency of *Cdx2* was adequate for polyp development. The polyps occasionally contained areas of metaplasia but were not associated with an increased risk of adenocarcinoma, maintaining their benign status, however they did have features often associated with the hamartomas seen in Peutz-Jeghers syndrome such as the existence of arborising bands of smooth muscle (Tamai *et al.*, 1999). The presence of hamartomas in the *Cdx2* knockout mouse indicated that loss of the human homologue *CDX2* might also be associated with the development of hamartomas. In the light of genetic heterogeneity in both JPS and PJS, *CDX2* was thus considered to be a candidate gene.

Further evidence for a role of CDX2 in colonic homeostasis was provided by the demonstration of somatic *CDX2* mutations in colorectal cancers. Mallo *et al* first identified human *CDX2* with differential screening of colorectal cancer mRNA *versus* normal mucosa, and subsequently found reduced CDX2 expression in a proportion of colorectal tumours (Mallo *et al.*, 1997). In addition, *CDX2* mRNA was not detectable in the colon cancer cell line LS174T. Wicking *et al* demonstrated mutations of both alleles of *CDX2* in a replication error cancer (Wicking *et al.*, 1998) and da Costa *et al* showed a cancer with normal APC/beta-catenin signalling to possess a *CDX2* mutation (da Costa *et al.*, 1999). Taken together, these findings suggested a putative role for *CDX2* as a tumour suppressor gene.

To investigate whether germline mutations in *CDX2* caused either PJS or JPS, the *CDX2* gene was screened in 10 PJS patients whose disease was not attributable to mutations in *LKB1* and 37 JPS patients whose disease was not attributable to mutations in *SMAD4*. In addition 49 colorectal cancer cell lines were screened for somatic *CDX2* mutations to evaluate the role of *CDX2* as a tumour suppressor gene involved in colorectal cancer tumorigenesis. JPS patients were selected as for the screening of the SMAD genes, but 7 new patients were also available (families 15, 17, 18, 19, 20, 21 and 22). The Peutz-Jeghers patients were (BC, BaC, 432, IO, JW, JS, EB, 5D. 5G amd 5A). The colorectal cancer cell lines screened for mutations were C10, C32, C70, C75, C80, C84, C99, C106, C125, CAC02, COLO201, COLO205, COLO206, COLO320, COLO678, COLO741, CX1, DLD1, GP2D, GP5D, HCA46, HCA7, HCT8, HCT15, HCT116, HRA19, HT29, HT55, LIM1863, LOVO, LS174T, LS180, LS411, LS1034, PC/JW, SKC01, SW48, SW403, SW480, SW620, SW837, SW948, SW1222, SW1417, T84, VACO4A, VACO4S, VACO5 and VACO10. Fifty control samples were derived from an unselected UK population with no known cancer predisposition, but were not matched for age or sex.

PCR primers were designed to amplify the three exons and exon-intron boundaries, both labeled with either FAM, TET, or HEX. PCRs were then performed using the conditions shown in Table 6.3.1. PCR products were then denatured and subjected to SSCP analysis using the Phast system at 10°C and/or the capillary based system on an ABI 310 genetic analyser under two different temperature conditions (20°C and

35°C). Fragments showing both normal and aberrant migration were re-amplified using non-fluorescently labelled primers, purified and sequenced in forward and reverse orientations.

Exon	*sense primer 5'-3'*	*antisense primer 5'-3'*	*Temp*	*Mg^{2+}(mM)*
1 part 1	*CAGCATGGTGAGGTCTGCT*	*GCGTAGCCATTCCAGTCCT*	*55*	*0.5*
1 part 2	*GGCAGCGAACTTGGACAG*	*GTTGAGCGTTTGCAGCAG*	*55*	*1*
1 part 3	*AGCCCCGCAGACTACCAT*	*CGCAGCCTCTGCTTACCTT*	*55*	*0.5*
2	*GCCCTCACTTCTCCTTCCTC*	*GTCCCCACCTGCCTCTCA*	*65*	*2.5*
3	*TTTTCTCCACCTTTCCATTTC*	*TCAGCCTGGAATTGCTCTG*	*55*	*2.5*

Table 6.3.1 Primers and PCR conditions for CDX2.
Exon 1 is divided into three parts so suitable fragments size for SSCP were obtained. 'Temp' indicates the annealing temperature of the PCR reaction and 'Mg 2+ ' shows Mg 2+ concentration required.

Patients	*Frequency of (t/t) homozygotes (%)*	*Frequency of (c/t) heterozygotes (%)*	*Frequency of (c/c) homozygotes (%)*
Juvenile polyposis	*68 (25/37)*	*30 (11/37*	*2 (1/37)*
Peutz–Jeghers	*80 (8/10)*	*20 (2/10)*	*0 (0/10)*
CRC cell lines	*72 (35/49)*	*22 (11/49)*	*6 (3/49)*
Controls	*78 (40/51)*	*22 (11/51)*	*0 (0/51)*

Table 6.3.2 Frequencies of the polymorphic CDX2 exon 2 alleles in JPS, PJS, colorectal cancer cell lines and a control cohort.
The observed frequencies of the respective alleles did not differ significantly between patients and controls (Fisher's exact test, P> 0.3).

No pathogenic mutations of *CDX2* were found in our 37 JPS patients or in the 49 colorectal cancer cell lines. One of the 10 PJS cases showed aberrant migration for the PCR fragment encompassing *CDX2* exon 3 using SSCP. Upon sequencing this anomaly was found to be an A to T base change at nucleotide 941. This changes lies in the 3' untranslated region of the *CDX2* gene upstream of the polyA signal, and is not conserved in the mouse *Cdx2*. This change was not seen in any of the JPS cases, the colon cancer cell lines or the fifty control subjects, but given its position is probably not of significance. One way for determining the role of the 941 variant in this patient would be to look for bi-allelic inactivation of CDX2, i.e. the 'second hit'. No tumour material was available however to assess whether there was loss of the second allele on 13q but given that haploinsufficiency of Cdx2 in the mouse appeared to be adequate to induce hamartoma formation, loss of heterozygosity may not be found even if the change were pathogenic.

A previously reported missense polymorphism was also detected in exon 3 of *CDX2*, a TCT to CCT transition at nucleotide 871 which introduces a serine to proline amino acid change at codon 291 (Yagi *et al.*, 1999). Although this may be a potential phosphorylation site, the mouse *Cdx2* has proline at this codon which is not known to be polymorphic and therefore the significance is in doubt. No significant difference was found between the serine and proline frequencies of the JPS or PJS patients, in the colon cancer cell lines or in the UK control cohort (Table 6.3.2), or from those previously reported (Yagi *et al.*, 1999). These data therefore support

previous suggestions that S291P is a polymorphism (Wicking *et al.*, 1998) which is not functionally significant for JPS or PJS , although some potential functional significance as a low-penetrance cancer predisposition allele cannot be entirely ruled out.

SSCP failed to detect a silent CCG to CCC polymorphism which had been previously reported (Yagi *et al.*, 1999). The amplimer flanking this polymorphism (exon 1 part 1) was only successfully sequenced in a small number of cell lines (10), despite repeated attempts. Every other fragment (exon 1 parts 2 and 3, plus exons 2 and 3) were sequenced in all of the patients and cell lines, so mutations in these exons can be confidently excluded. The inability of multiple SSCP conditions to detect the codon 61 polymorphism, and the failed attempts to sequence all the patients, does raise the remote possibility that further changes in this fragment would also have been missed in these people. It would be anticipated that most sequence variants would be detected under varying electrophoresis conditions, and therefore this part of exon 1 may also be excluded, with some caution.

Thus despite *CDX2* being a strong 'hamartoma' candidate, and previous reports of colorectal cancer mutations in this gene, it was concluded that *CDX2* was not responsible for JPS, PJS, and at best is infrequently mutated in colorectal cancer. Further investigations of the *Cdx2* mouse hamartomas indicated that rather than true polypoid lesions, the polyps were composed of heterotopic stomach and small

intestine mucosa (Beck *et al.*, 1999), and it was concluded that haploinsufficient levels of *Cdx2* in the developing intestine lead to transformation to a more endodermal phenotype i.e. such as forestomach epithelium that does not express *Cdx2* during normal development. This intercalary growth in a restricted space thus results in the formation of the polypoid lesions observed (Beck *et al.*, 1999). With this in mind, it is perhaps not surprising that no mutations were found in the JPS and PJS cohorts.

Although there have been reports of *CDX2* mutations in colorectal cancer, interestingly they are all in mismatch repair deficient (MSI+ or microsatellite instability positive) cancers. For example, Mallo *et al* reported the a lack of *CDX2* mRNA in the cell line LS174T which is well known as MSI+ (Mallo *et al.*, 1997), Wicking *et al* reported both alleles of *CDX2* to be mutated in a cancer which was characterised as MSI+ (Wicking *et al.*, 1998), and da Costa *et al* reported a *CDX2* mutation in the cell line RKO, which again is well known to be MSI+ (da Costa *et al.*, 1999). A recent report of *CDX2* mutations in four of fifty one sporadic colorectal cancers did not characterise the MSI status (Csivagnanasundaram et al., 2001). This frequency is lower than the ten per cent of sporadic tumours which will be MSI+, therefore it might not be unreasonable to suggest that these *CDX2* mutations were indeed in MSI+ tumours. Taken together it would seem that *CDX2* mutations do occur in colorectal cancer but perhaps exclusively in MSI+ tumours. In each case of Mallo *et al,* Wicking *et al,* and da Costa *et al,* the aberrations were all shown to be

functionally significant. SSCP and sequencing failed to detect any mutations in the MSI+ lines used in this study (including LS174T). There are two possibilities to explain this. Either any mutation(s) clustered in part 1 of exon 1 and were not detectable by multiple SSCP conditions, or, perhaps more likely, epigenetic inactivation of *CDX2* via promoter methylation has occurred and indeed CDX2 is not expressed in a subset of the MSI+ cancer cell lines.

It is probable therefore that *CDX2* mutations may represent part of the colorectal cancer pathway for mismatch repair deficient tumours which does not rely on the classical APC/Kras/SMAD4/TP53 pathway, and its clarification as such is warranted.

6.4 INVESTIGATING *PTCH* FOR GERMLINE MUTATIONS IN JPS

Juvenile polyps have been reported to occur as a manifestations of the dominantly inherited familial cancer syndromes including Cowden disease (CS), Bannayan-Zonana Syndrome (BZS) and to a lesser extent, Gorlin syndrome (GS) (each discussed in detail earlier). Germline mutations of the *PTEN* gene on chromosome 10q23 had been shown to cause CD and BZS, and the exclusion of this gene as the causative gene in JPS has been demonstrated (Marsh *et al.*, 1997b). GS results from germline mutations in the *PTCH* gene (homologue of Drosophila *patched*) on

chromosome 9q22.1 (Hahn *et al.*, 1996a). Juvenile polyps appear to comprise a relatively minor and infrequent component of GS, although few GS patients undergo gastrointestinal screening so the true number is difficult to estimate. Nevertheless, *PTCH* remains a candidate for JPS given that a different spectrum of mutations might cause JPS without the other features of GS.

DNA was extracted or available from 15 JPS families (1, 3, 6, 8, 9, 10, 11, 12, 15, 16, 17, 20, AF, BL, WN) and nine apparently sporadic cases (KS, RV, SC, SCA, SD, SH, SR, SS, SV). Published oligonucleotide sequences were used for the primer synthesis of each exon of the *PTCH* gene (Wicking *et al.*, 1998). Phast SSCP analysis on 12.5% and 20% gels was used to screen exons 1-15 of *PTCH* in the JPS patients at 10°C. (Exons 15-22 were screened by conformation specific gel electrophoresis analysis by Richard Houlston at the ICR). Where there was an aberrant shift on SSCP or CSGE, the PCR was re-amplified and subjected to direct sequencing.

No germline mutations of the *PTCH* gene were observed in the group of JPS patients studied. Due to the large size of the *PTCH* gene however, not every exon was directly sequenced in all of the patients, and therefore mutation detection is rather reliant on the sensitivity of the screening technique, in this case SSCP and CSGE. SSCP analysis is understood to have 80% sensitivity for detecting sequence differences, and under multiple running conditions the percentage of mutations

detected should be higher than this. If germline mutations of the *PTCH* gene accounted for a significant number of JPS cases, it would therefore be expected that our screening techniques would have identified a good proportion of these. If however germline mutations of the *PTCH* gene are rare and only occur in a minor subset of JPS cases, there is a higher chance they will have been missed with our screening techniques. This would perhaps only be likely if there was a misdiagnosis and the juvenile polyps were occurring as a feature of Gorlin syndrome rather that pure JPS. This is unlikely given that none of the patients had prototypical dermatological (such as basal cell carcinoma or palmar pits) or skeletal abnormalities, and malignancies in the cohort were confined to the gastrointestinal tract, in contradistinction to GS.

Germline mutations of the *PTCH* gene can be fairly confidently excluded as the cause of pure juvenile polyposis syndrome. Despite the presence of juvenile polyps as a feature of GS, it appears the situation is much like that of Cowden syndrome, where juvenile polyps are a characteristic ~~symptom~~ manifestation but the causative gene (*PTEN*) does not also cause pure JPS.

Although *PTCH* mutations may cause gastrointestinal hamartoma formation directly, it is possible that reports of juvenile polyps in GS either result from a chance association, or from a contiguous deletion of PTCH and at least one other nearby gene.

6.5 CONCLUSIONS

Screening of several candidate genes in juvenile polyposis revealed no new pathogenic mutations. Germline mutations of *SMAD4* have recently been shown to cause a subset of JPS cases, and despite being good candidates as they belong to the same family and are also involved in TGFβ-signalling, germline mutations of the other SMADs were not detected in our JPS cases. This is likely to be because SMAD4 is the only SMAD that does not have a known partner in the human to compensate when there is loss, and therefore is the strongest target to disrupt the pathway. Conversely, the other SMADs tend to have partners, e.g, SMADs 2 and 3 are both receptor-regulated SMADs, which may mean there is a level of redundancy that allows one to compensate if there is loss of the other, making them weaker targets for disrupting the signalling pathway.

Similarly no germline mutations of *CDX2* were detected in JPS patients without *SMAD4* mutations, or Peutz-Jeghers patients without *LKB1* mutations, or in a group of 49 colorectal cancer cell lines. This is probably because the reported hamartomas of the *Cdx2* knockout mice were in fact intercalations of gut tissue and not true polyps. Although there have been reports of *CDX2* mutations in colorectal cancer, these appear to be confined to MSI+ (microsatellite instability positive) cancers, where it does seem that CDX2 plays a true role in the tumorigenic pathway of these cancers.

Germline mutations of the *PTCH* gene, which have been shown to cause Gorlin syndrome, were excluded as the causative gene in our JPS patients. Juvenile polyps seem to occur as part of many disease spectrums but in each case the responsible genes are confined to a their particular disease and do not also cause pure JPS.

The genes causing the remainder of the JPS cases not caused by mutations of the SMAD4 gene remain elusive. It does appear that screening genes purely on good candidature is not a strong and fruitful approach, prior evidence such as linkage analysis or loss of heterozygosity to pinpoint relevant genomic regions might be a stronger approach and allow insight before candidates are chosen.

CHAPTER SEVEN

A GENOME WIDE SEARCH FOR NEW JPS GENES

A GENOME WIDE SEARCH FOR NEW JPS GENES

7.1 INTRODUCTION

Despite the screening of genes that were good candidates (discussed in the previous chapter), the genetic defects underlying the majority of JPS cases (those not caused by *SMAD4*) remained unresolved. What was clear was that a number of genes could predispose to juvenile polyps, either in a pure JPS setting (i.e. mutations in *SMAD4*), or as part of syndrome that affects multiple organ systems (e.g. *PTEN* mutations in Cowden disease, or *PTCH* mutations in Gorlin Syndrome). This clinical overlap may potentially confound the discovery of new JPS genes and therefore the first priority was to be as close as possible to certainty that the patients were indeed pure JPS and did not have phenotypic features associated with the other hamartoma syndromes (discussed in Chapter Three). The second priority was that *SMAD4* had been reliably excluded as the causative gene in the remaining cohort, and therefore that the likelihood of false negatives in subsequent studies was reduced (also detailed in Chapter Three). Candidate gene screening for JPS did not reveal any new pathogenic mutations (Chapter Six), and it was therefore decided to undertake this approach only after gaining some evidence that the causative gene mapped to a specific region. Genome wide strategies for mapping the remaining JPS gene or genes were thus required. Indeed, *SMAD4* itself was first identified as a JPS gene after targeted linkage analysis of candidate regions (Howe *et al.*, 1998a), and *PTEN* was shown

to cause Cowden disease after linkage analysis had homed in on 10q23, and the gene itself identified by homozygous deletion mapping (Liaw *et al.*, 1997). Alternative or complementary genome-wide approaches to linkage analysis for identifying tumour suppressing genes which cause inherited cancer-predisposing syndromes are comparative genomic hybridisation (CGH) (Kallioniemi *et al.*, 1992) and loss of heterozygosity studies. This strategy localised the Peutz-Jeghers locus to 19p13 (Hemminki *et al.*, 1997), and subsequent linkage analysis and screening of candidates mapping to this region identified *LKB1* as the causative gene (Hemminki *et al.*, 1998).

This chapter describes how a genome wide linkage search was performed on Juvenile Polyposis families in an attempt to discover the remaining underlying genes causing JPS. In addition, comparative genomic hybridisation was performed on polyps and cancers derived from JPS individuals in the belief that this may highlight regions showing loss of genetic material at JPS tumour suppressor susceptibility loci. Likewise, loss of heterozygosity analysis was undertaken in JPS polyp and cancer material in an attempt to identify lost genetic regions which may be pathogenic in JPS. The objective of these studies was to reveal, with the complementary techniques, specific areas of the genome which are involved in the pathogenesis of JPS. Whilst every attempt was made to exclude *SMAD4* mutant families from the linkage analysis, soon after the search was completed another group identified BMPR1A/ALK3 as a JPS gene which has led to a re-analysis of most of the data.

7.2 GENOME SCREEN FOR JUVENILE POLYPOSIS GENES

Families 1, 5, C1 and 14 were not included in the genome-wide linkage analysis as they were all compatible with linkage to *SMAD4*. No polyp material from these families was available for LOH analysis or immunohistochemistry and there therefore remained the possibility of unidentified *SMAD4* mutations, which would confound the linkage analysis. Although Family 18 were also compatible with linkage to *SMAD4*, two polyps available from this family did not show evidence of LOH around the *SMAD4* locus, and it was considered unlikely that *SMAD4* was the causative gene in this family. Thus, Family 18 were included in the linkage analysis. DNA was extracted from 45 individuals from 7 families (6, 10, 12, 15, 18, 19, and MD) suitable for linkage analysis. Within these families, 25 individuals were reported to be affected with JPS, ranging from 2 to 5 affected people per family. Families 6, 15, 19 and MD had been collected in England, Family 10 was from Israel, Family 18 was from Korea and Family 12 was of European origin from Australia. Affection status was assigned using medical records, questionnaires and histopathology reports. All other individuals were classified as 'unknown' for the purposes of the analysis. The maximum theoretical LOD score, assuming genetic homogeneity was 6.0. Families 12 and 15, each having five affected members, were the largest families.

PCR amplification (using standard conditions of 35 cycles with the annealing temperature 55°C and 1.5mM Mg^{2+}) of 387 microsatellite markers spaced at ~10cM intervals across the genome was performed on the 45 JPS individuals

using the Weber8 linkage set (Research Genetics, Huntsville, AL). The genome screen PCRs were performed jointly with Dr Richard Houlston at the ICR. JPS was modelled as a dominant trait for the linkage analysis (q=0.001) with nominal penetrances of AA=0.75, Aa=0.75 and aa=0.001. The value of 0.75 used for the 'affected' homozygotes and heterozygotes allowed for possible incomplete penetrance of JPS, whilst the 0.001 wild-type penetrance value allowed for the presence of phenocopies (single juvenile polyps without an inherited susceptibility are not uncommon). Two-point LOD scores were calculated for each marker using the subprogram MLINK (v5.1) of the LINKAGE program package (Lathrop *et al.*, 1984) as implemented in FASTLINK (v4.1) (Cottingham *et al.*, 1993). Multipoint analyses of regions which looked positive were undertaken using the VITESSE program (O'Connell and Weeks, 1995) Marker allele frequencies were taken from the Genome Database (http://www.gdb.org). Haplotype construction was undertaken either by hand or by using the Simwalk2 program.

Appendix One shows the genome wide two-point LOD scores for all families. No site in the genome gave a two-point LOD score of more than 3, (corresponding to a significance level of 5%), the figure generally accepted to provide significant evidence of linkage to a specific region. Eight sites in the genome gave a two-point LOD score of more than 1 (on chromosomes 1, 7, 10, 11, 12 and 13), and multipoint analyses were thus performed on markers flanking these regions. To confirm or refute linkage in these regions, haplotypes were constructed. In addition, haplotypes were constructed for the whole genome in

Families 12 and 15 in order to determine whether any regions had been overlooked in the two-point analyses. Probably reducing the power to detect the true JPS locus, certainly in the two larger families, was the presence of potential phenocopies. Person 307 in Family 12 has developed three small polyps (before the age of 40). The other affected members of this family, however, developed florid polyposis and small bowel carcinomas before the fifth decade, and it was therefore difficult to be sure that person 307 had inherited the same susceptibility gene. Conversely, an individual would be unlikely to develop three juvenile polyps at such a young age. For the purposes of the linkage analysis, person 307 in Family 12 was of uncertain affection status and was therefore classified as 'unknown'. In reality, it was suspected that this person was indeed affected and either had the phenotype modified by genetic or environmental means, or the susceptibility gene was not as penetrant as in other Family 12 individuals. In Family 15, person 308 from was classed as of 'unknown' affection status for the purposes of the linkage analysis. This individual had developed three small adenomas in his sixties. It was most likely however that this individual had inherited the susceptibility gene but this had not led to detected juvenile polyps, although had led to multiple adenomas, a recognised feature of JPS gene carriers.

7.2.1 ANALYSIS OF CHROMOSOME 1P32-33 IN JPS FAMILIES

A two-point LOD score of 1.09 was obtained with marker D1S1728 at 1p32-33. Two-point LOD scores for the region on chromosome 1p32-33 are shown in

Table 7.2.1.1 Haplotype constructions for this region are shown in Figure 7.2.1.1. Family 15 were compatible with linkage to markers on 1p32-33, reflected both in the two-point analysis and the haplotypes (Figure 7.2.1.1). Importantly, this was the only region in the entire genome that was compatible in this family where screened apparently unaffected family members did not share the putative affected haplotype, and where the adenomas of person 308 were not required to be due to a phenocopy. The two-point LOD score at D1S1728 for Family 12 was near to one (0.89) and indeed affected members did share a haplotype for the 1p32-33 markers (Figure 7.2.1.1). However, linkage to this region would rely on person 307 being a phenocopy as the putative 'affected' haplotype was not shared with this individual.

ID	D1S1665 (101cM)	D1S1728 (110cM)	D1S551 (113cM)	D1S551 (123cM)	D1S1631 (137cM)
	-0.61	1.09	-0.11	-0.11	-3.59
MD	-1.37	0.15	-1.08	-1.08	-1.37
6	0.26	-0.31	0	0	-1.2
10*	**0**	0.17	**0.14**	**0.14**	-0.06
12	0.21	0.87	0.43	0.43	-1.29
15	**0.29**	0.06	**0.16**	**0.16**	**0.16**
18*	**0**	0.15	**0.24**	**0.24**	0.17
19*	**0.16**	0.17	**0.17**	**0.17**	0

Table 7.2.1.1 Two-point LOD scores for chromosome 1 in JPS families.
*Shown are the two-point LOD scores at θ=0 for each family. The position of the markers are shown in brackets after the marker name. Multipoint analysis for D1S1665, D1S1728 and D1S551 gave a LOD score of 0.00. . *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who were compatible with linkage to this region.*

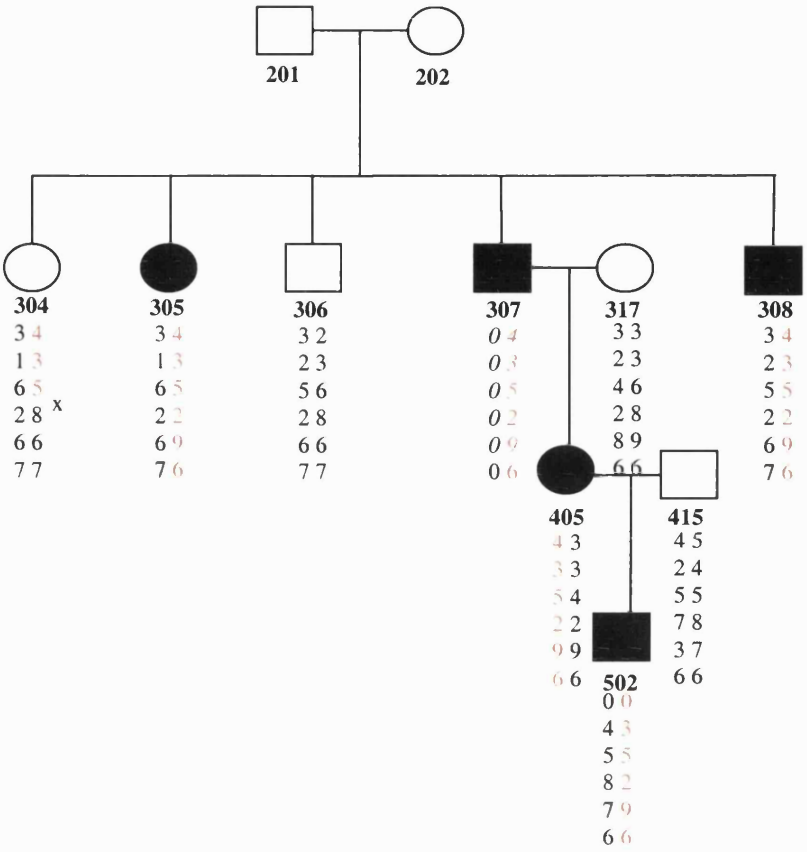


Figure 7.2.1.1 Haplotype construction for chromosome 1 in JPS families. *Shown are the genotypes for D1S1665, D1S1728, D1S551, D1S1588 D1S1631 and GATA176G01. Putative affected haplotypes are highlighted red to aid tracking, and inferred haplotypes are italicised. Family 15 were compatible with linkage to these markers as all affected members share a common haplotype.* *Figure continued on next page.*

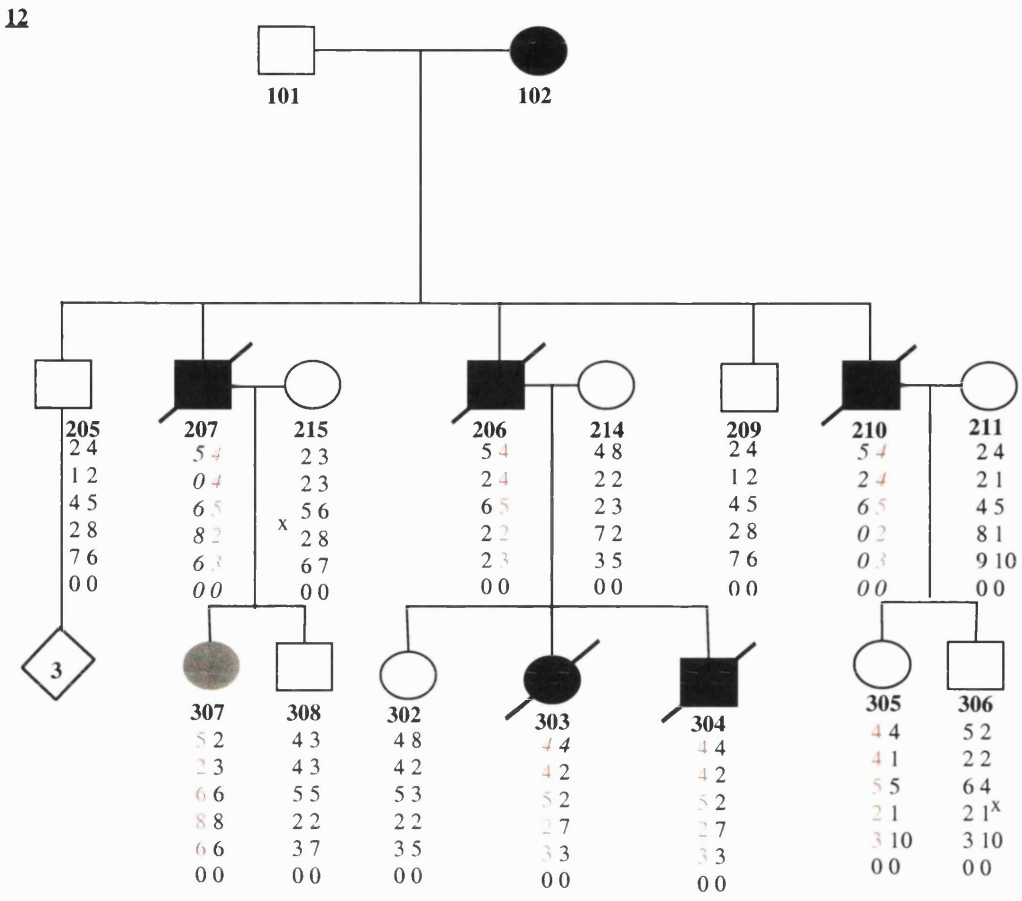


Figure 7.2..1.1 continued. *Family 12 were compatible with linkage to chromosome 1p32-33 markers, but only if person 307 was indeed a phenocopy - this individual did not share a haplotype with other affected family members. Persons 305 and 308 shared the haplotype with the affected individuals, and would therefore have to be non-penetrant for any disease susceptibility locus that mapped to these markers.*

In addition to the known affected members, individuals 305 and 308 also shared a haplotype for 1p32-33 markers. It was therefore considered unlikely that this region on chromosome 1 harboured a disease susceptibility locus for Family 12, as the polyps of person 307 are probably too rare in the general population to represent a true phenocopy, and two unaffected individuals shared a haplotype with the affected family members. Family MD was not compatible with linkage

to the 1p32-33 markers as affected siblings 208 and 55 had inherited different alleles from their affected father. This was reflected in the two-point LOD scores (Table 7.2.1). Family 6 was also unlikely to be compatible to the 1p32-33 markers as affected siblings 401 and 402 had most likely inherited their unaffected paternal grandfathers chromosome, rather than their affected grandmothers, and again this was reflected in the two-point LOD scores and multipoint analysis. Sib-pairs 10, 18 and 19 were identical at all alleles at 1p32-33 and therefore compatible with linkage to this region. However, germline mutations in *BMPR1A* have been subsequently demonstrated in these three families (10, 18 and 19) and therefore the linkage to the 1p32 markers was undoubtedly a false positive.

Overall therefore, only Family 15 was reliably compatible with linkage to the 1p32-33 markers. One gene mapping to the 1p32-33 interval is the TGFβ-type III receptor. This gene is *a priori* an excellent candidate, given it belongs to the same signalling pathway as *SMAD4* and that its inactivation may have effects similar to the loss of *SMAD4*.

7.2.2 ANALYSIS OF CHROMOSOME 7 LINKAGE COMPATIBILITY

A two-point LOD score of 1.28 was obtained with marker D7S3846 positioned at 7p11-12, and visual inspection of the two-point LODs for markers flanking D7S2846 indicated evidence for linkage to this region of chromosome 7, certainly in the largest family, 12 (Table 7.2.2.1).

ID	D7S1802 (33cM)	D7S1808 (41cM)	D7S817 (48cM)	D7S2846 (56cM)	D7S1818 (69cM)	GATA118G10 (77cM)	D7S2204 (90cM)
	-5.6	-3.5	-1.3	1.28	-6.9	-4.2	-1.2
MD	-1.68	-1.87	0	0.39	-1.97	-1.68	-1.83
6	-1.77	-0.45	-1.77	0.1	-1.77	-1.77	0
10*	-0.08	0	0.17	0	0.23	-0.08	-0.1
12	**0.16**	**0.35**	**0.41**	**0.6**	**0.9**	**0.89**	**0.3**
15	-1.83	0.37	-0.14	0.18	-2.68	-1.72	0.3
18*	-0.29	-0.11	0.17	0.21	0.21	0.23	0.25
19*	-0.14	-1.82	-0.11	-0.2	-1.82	-0.08	-0.1

Table 7.2.2.1 Two-point LOD scores for chromosome 7 in JPS families.
*Shown are the two-point scores for θ=0 for markers mapping to chromosome 7 in the JPS families. Distances are shown in brackets after the marker name. *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who were compatible with linkage to this region. Multipoint analyses gave a LOD score of 0.9 for Family 12 considered alone, and 0.35 for all families combined.*

Haplotype construction was therefore performed for markers spanning this region. Families 6, 19 and MD were not compatible with linkage to the 7p11-12 markers due to a failure of affected family members to share a haplotype. In Family 6, two affected siblings (401 and 402) had inherited different alleles from their affected father. The sib-pair Family 19 shared no alleles at the 7p11-12 markers. In Family MD, an affected uncle (203) did not share any 7p11-12 alleles with his affected nephew (55). Family 15 was not compatible with linkage to 7p11-12 due to the failure of affected person 305 to share alleles with other affected family members at D7S2846. The incompatibility of families 6, 15, 19 and MD to 7p11-12 markers was reflected in the two-point LOD scores (Table 7.2.2.1) and also in the overall multipoint LOD score of 0.35. Families 10 and 18 (comprising two-sib pairs) were again compatible with linkage to the 7p11-12 markers, as they were identical all alleles at all three markers. However, such sib-pairs have a 50% chance of sharing one allele, and a 25% chance of sharing

both alleles at a given marker. It was therefore not surprising that these small families were repeatedly compatible with linkage.

Family 12 were compatible with linkage to the 7p11-12 markers (D7S2846, D7S1818 and GATA118G10) and the multipoint analysis for this family considered alone at these markers gave a maximum LOD of 0.9 between D7S1818 and GATA118G10. Genotyping of additional markers mapping to this region (D7S555 (64cM), D7S634 (72cM) and D7S2242 (120cM)) and haplotype construction confirmed that all affected members of Family 12 shared a common region on chromosome 7 (Figure 7.2.2.1). For the genome screen markers, affected members of Family 12 shared a haplotype spanning from D7S1808 to D7S821, a region of 71cM. Person 307 shared the putative affected haplotype; there was therefore no need to invoke a phenocopy explanation for her three polyps. In addition, however, three individuals not thought to be affected also shared the haplotype (209, 302 and 306, Figure 7.2.2.1). Given the probable incomplete penetrance of the susceptibility locus (viz the mild phenotype of person 307), it is not inconceivable that these individuals also carried the same disease predisposition.

Loss of heterozygosity analysis was then performed, on the assumption that the putative gene in this shared region would be a tumour suppressor. No LOH was observed in 10 cancers from Family 12 at either D7S2846 or D7S2204. One of the ten cancers from Family 12 showed LOH at D7S1818. The relative infrequency of LOH at the chromosome 7 markers reflects several possibilities.

Firstly, the markers chosen for LOH may not have pinpointed accurately the area containing the putative gene - this was quite possible given the large size of the compatible region. Alternatively the microdissections contained too much contaminating normal tissue to detect LOH. Finally, there is the possibility that the shared region on chromosome 7p in Family 12 was a false positive.

A two-point LOD score of greater than 1 was also obtained at another chromosome 7 marker, D7S1824 at 7q32-36 (see Appendix One). Inspection of the genotypes for D7S1824 and flanking markers made it likely that this LOD was a false positive. Many individuals 'failed' at D7S1824.. This made phase determination impossible, and sharing of the alleles between the individuals in whom the PCR was successful made the two-point analysis positive. For example, individual 502 from Family 15 failed at D7S1824, but haplotype construction for the flanking markers showed that he had inherited his maternal grandmothers alleles, rather than his affected maternal grandfathers alleles at 7q32-36. Likewise, failure of some individuals at D7S1824 in family MD meant that a positive two-point was obtained at this marker. However, affected brothers 96 and 203 from Family MD shared no alleles at either of the two flanking markers and therefore the positive LOD at D7S1824 was a spurious result. The only family truly compatible with linkage to 7q32-36 was Family 6, where two affected siblings had inherited the same alleles from their affected father. The power to prove linkage in such a small family considered alone is very small, so confirmation that this region does indeed contain a susceptibility locus in just this family is difficult.

12

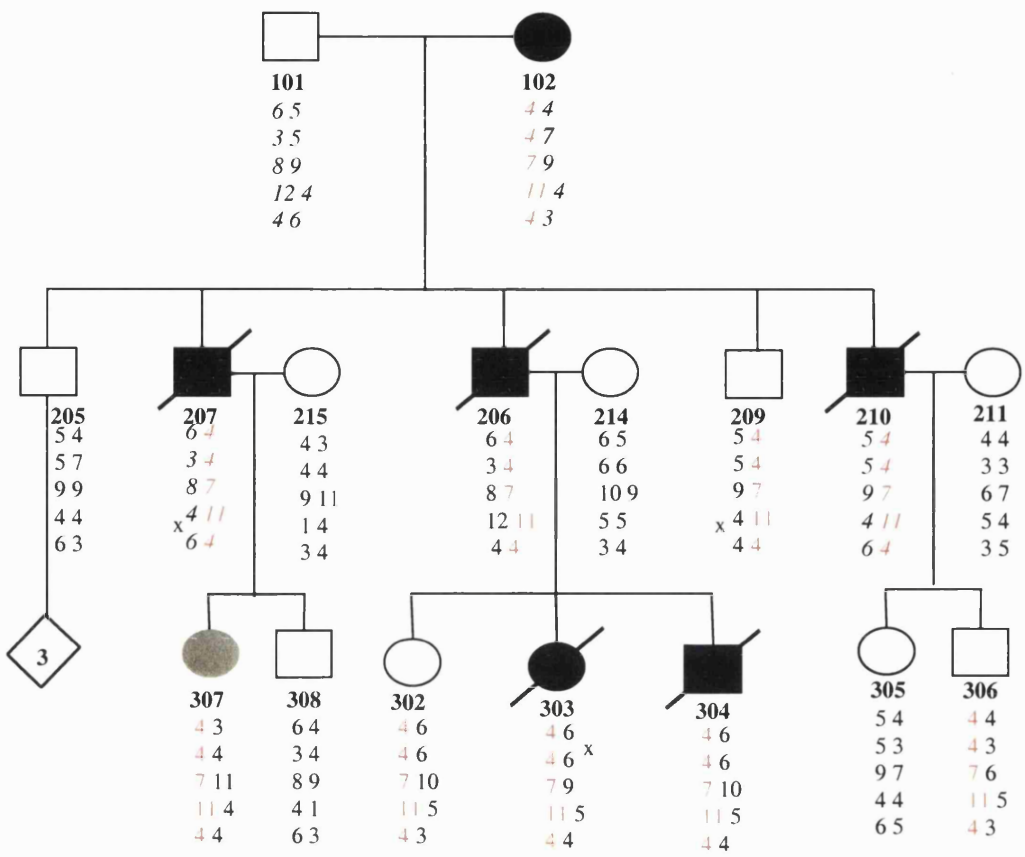


Figure 7.2.2.1. Chromosome 7 haplotype in Family 12.
Shown are the genotypes from D7S817, D7S2846, D7S1818, GATA118G10 and D7S2204. Shaded shapes represent affected individual, grey shading represents affected with a small number of polyps. The putative affected haplotype is shown in red. Family 12 shared a haplotype at these markers and were therefore compatible with linkage to 7p11-12, although 209 would have to be non-penetrant.

7.2.3 ANALYSIS OF CHROMOSOME 10 COMPATIBILITY IN JPS FAMILIES

ID	D10S1239 (126cM)	D10S1237 (137cM)	D10S1230 (150cM)	D10S1213 (155cM)	D10S1248 (175cM)	D10S212 (181cM)
	-2.28	1.77	-1.26	2.04	-1.68	-2.51
MD	**0.14**	0.53	**0.49**	0.3	**0**	**0.55**
6	**0.1**	0.42	**0**	0.56	**0**	**0.01**
10*	**0.21**	-0.08	**-0.14**	**-0.11**	**0**	**0.23**
12	-1.57	0.12	-1.67	0.88	0.12	-1.85
15	-1.64	0.28	**0.08**	0.39	-1.89	-1.66
18*	**0.28**	0.28	**0.12**	0.13	0.26	0
19*	**0.2**	0.22	**-0.14**	**-0.11**	-0.17	0.21

Table 7.2.3.1 Two-point LOD scores for chromosome 10q26 in JPS families. *Shown are the two-point scores for θ=0 for markers mapping to chromosome 10q26 in the JPS families. Distances are shown in brackets after the marker name. *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who were compatible with linkage to this region*

Two point LOD scores of greater than one were found at two markers on chromosome 10q26, D10S1237 and D10S1213. These markers lay ~40cM distal to the region subsequently shown to harbour the *BMPR1A* gene (10q22) in which mutations were identified in a subset of families (see Chapter Nine). The overall multipoint LOD score for D10S1237, D10S1230 and D10S1213 was 2.65, the highest achieved anywhere in the genome. However, inspection of the haplotypes for this region made it unlikely that this region harbours a disease susceptibility locus. For Family 12, the two-point LOD at D10S1230 was negative due to the failure of affected brothers 206 and 203 to share alleles with a third affected brother, 204. For the two flanking markers with positive two-point LOD scores, three unaffected individuals (305, 306 and 308) shared alleles with the affected individuals, yet 307, the possible phenocopy, did not. It was likely

that if incomplete penetrance explained the presence of sharing in the unaffected individuals, then person 307 had also inherited the susceptibility locus, rather than be a phenocopy. The multipoint analysis re-run with person 307 classed as 'affected' rather than 'unknown', reduced the multipoint LOD score to 1.36. Family 15 were compatible at the 10q26 markers (although they were not compatible at *BMPR1A* markers at 10q23). However, all unaffected members of Family 15 shared alleles with affected family members at D10S1237, D10S1230 and D10S1213, and would therefore have to be non-penetrant were an underlying genetic defect identified. Family 6 were also compatible at the 10q26 markers (though again not at *BMPR1A* markers, Chapter Nine) as both affected siblings (401 and 402) had inherited the same alleles from their affected father (302). Person 301 from Family 6 was classed as 'unknown' for the purposes of linkage analysis, but was not known to have developed any juvenile polyps. This individual shared alleles at the 10q26 with his affected brother, so one again would have to be a non-penetrant if the linkage was borne out. Once more, the new DNA from an affected person in Family 15 (406, son of 307), and also new DNA from Family 6 (affected mother of 302) should help confirm or refute the 10q26 region, (as well as other genome regions such as 1p32-33).

Families 10, 18, and 19 were compatible with linkage to D10S1237, D10S1230 and D10S1213, firstly because these families were small and therefore compatible at many loci, but also because they were subsequently shown to harbour mutations in *BMPR1A* on 10q22. Likewise, Family MD was compatible with linkage to BMPR1A, although no mutation has been identified as yet

(Chapter Nine), and it was therefore possible that they would also share alleles at markers distal to *BMPR1A*. Person 56 from Family MD would again have to be non-penetrant as he had inherited the same alleles as his affected brother. The inclusion of Families 10, 18, and 19 (and possibly MD) in the multipoint analysis for the 10q26 markers, most probably made the LOD score here higher than it would otherwise have been.

Overall, Families 6, 15 and MD were compatible with linkage to markers at 10q26 (D10S1237, D10S1230 and D10S1213), although unaffected individuals in each of these families would have to be non-penetrant due to sharing of alleles with affected family members. Finer mapping using new affected individuals (201 from Family 6 and 406 from Family 15) plus LOH analyses for tumours available from both these families, may help to confirm or refute the candidacy of this region in a subset of JPS families.

7.2.4 ANALYSIS OF CHROMOSOMES 11, 12 AND 13 COMPATIBILITY IN JPS FAMILIES

Table 7.2.4.1 shows the two-point LOD scores for markers mapping to 11q22-24. A LOD score of 1.01 was obtained for marker D11S1998, and a maximum multipoint LOD score of 1.5 was obtained for the region containing D11S2000, D11S1998 and D11S4464. Haplotype construction for Family 12 indicated that they were compatible with linkage to these markers.

ID	D11S2002 (82cM)	D11S2000 (94cM)	D11S1998 (100cM)	D11S4464 (110cM)	D11S912 (118cM)
	-4.26	0.69	1.01	-3.28	-5.01
MD	-1.71	0.17	0.52	0.12	-1.66
6	0	0	0.3	0	-1.77
10*	-0.11	0.25	0.21	0.2	-0.04
12	-0.24	0.08	-0.24	-1.94	-1.78
15	-1.8	0.19	0.3	-1.8	0
18*	-0.29	0	0	0.25	0.24
19*	-0.11	0	-0.08	-0.11	0

*Table 7.2.4.1 Two-point LOD scores for chromosome 11q22-24 in JPS families. Shown are the two-point scores for θ=0 for markers mapping to chromosome 11q22-24 in the JPS families. Distances are shown in brackets after the marker name. *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who were compatible with linkage to this region*

Once again, however, haplotype sharing was not observed in person 307 (the possible phenocopy) but was in her apparently unaffected brother (308). LOH with markers D11S2002 and D11S2000 was performed on two polyps and five cancers from Family 12. No LOH was observed in any tumour at either marker. These facts combined were probably indicative that Family 12's susceptibility locus lay elsewhere than 11q22-24.

Although the two-point LOD scores for Family 15 were weakly positive at D11S2000 and D11S1998 (Table 7.4.2.1), haplotype construction showed that individual 502 had not inherited his alleles from his affected grandfather but from his unaffected grandmother. Linkage to 11q22-24 in Family 15 was therefore refuted. The two-point LOD scores for Family 6 were weakly positive for 11q22-24, and this was most likely due to poor informativity at D11S2000, D11S1998 and D11S4464 in this family. All individuals, whether affected or thought to be unaffected, carried the same haplotype with the phases not being determinable. It was therefore unlikely that this region contained a disease

susceptibility locus for Family 6. Likewise, in Family MD, poor informativity at the 11q22-24 markers made the phase of inheritance indeterminable. However, inspection of the genotypes either side of this region (D11S2002 and D11S912) indicated that a haplotype was not shared between affected Family MD markers. This was reflected in the two-point LOD scores (Table 7.2.4.1). The sharing of alleles in sib-pairs 10, 18 and 19 was once again due to the high probability of this occurring in small families.

Overall therefore, it was likely that the LOD score of greater than one at D11S1998 was a false positive. Linkage to 11q22-24 could not be shown unequivocally in any JPS family, and thus this region is unlikely to contain a JPS susceptibility locus.

ID	GATA49D12 (11cM)	D12S391 (18cM)	D12S373 (26cM)	D12S1042 (38cM)	GATA91H06 (44cM)
	-4.65	-6.57		-4.73	-4.14
MD	-1.97	-3.77	0.46	0.3	0.6
6	-0.17	-0.55	0	-1.77	-1.77
10*	-0.14	-0.04	0.21	0.25	0.26
12	**0.13**	**0.55**	**0.28**	-1.77	-1.47
15	-2.68	-2.39	0.3	-1.87	-1.9
18*	0	-0.29	-0.05	-0.12	-0.08
19*	0.18	-0.08	-0.09	0.25	0.22

*Table 7.2.4.2 Two-point LOD scores for chromosome 12p12-13 in JPS families. Shown are the two-point scores for θ=0 for markers mapping to chromosome 12p12-13 in the JPS families. Distances are shown in brackets after the marker name. *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who were compatible with linkage to this region*

The two-point LOD scores for markers mapping to 12p12-13 are shown in Table 7.2.4.2, with a score >1 obtained at D12S373. The maximum multipoint score for

the region containing D12S391, D12S373 and D12s1042 was 0.00. The positive LOD obtained at D12S373 was most likely a false positive, indicated by the lack of haplotype sharing in Families MD, 6 and 15 (not shown). This was reflected in the two-point LOD scores either side of D12S373, which were negative in both Families 6 and 15. Known affected members of Family 12 did share a haplotype, except person 307 (the possible phenocopy) who did not share any alleles at D12S391, D12S373 or D12S1042. Once again two unaffected Family 12 individuals (209 and 302) also shared a haplotype with affected individuals, making it improbable that this region was related to disease susceptibility. Families 18 and 19 shared half of their alleles at the 12p12-13 markers, and Family 10 shared all alleles at these markers. This is reflected in negative two-point scores for Families 18 and 19, and positive two-point scores for Family 10 at D12S391, D12S373 and D12S1042. Compatibility to this region was once again viable for all three sib-pairs (but not a reflection of true linkage due to pathogenic mutations in the *BMPR1A* gene on 10q22 in these families).

The final genome region to give a LOD score of >1 was on chromosome 13, with a two-point LOD score of 2.02 at D13S779. The maximum multipoint LOD score for D13S793, D13S779 and D13S796 was 0.59. Haplotype construction showed that Families MD and 15 were compatible with linkage to this region (Figure 7.2.4.1), as affected family members shared haplotypes spanning these markers. Further linkage analysis with a new affected member, (406 in Family 15) combined with LOH analysis should help to determine whether the 13q31-32 region does indeed harbour a disease susceptibility locus for Families 15 and

MD. Person 306 would be non-penetrant were linkage confirmed to 13q31-32 as he also shared the putative 'affected' haplotype. PCR failure in Family MD for a proportion of individuals meant that sharing of haplotypes relied heavily on inferred alleles.

ID	D13S317	D13S793	D13S779	D13S796	D13S285
	-6.7	-2.8	2.02	-2.43	-3.52
MD	-1.68	**0.05**	**0.46**	**0.3**	0
6	-1.77	-1.77	0.55	0.3	0.25
10*	-1.82	0.27	-0.08	-0.04	0
12	-1.92	0.17	0.46	-3.78	-2.11
15	0.11	0.3	0.12	0.3	-1.84
18*	0.15	-1.82	0.28	0.24	-0.08
19*	0.23	0	0.23	0.25	0.26

*Table 7.2.4.3 Two-point LOD scores for chromosome 13q31-32 in JPS families. Shown are the two-point scores for θ=0 for markers mapping to chromosome 13q31-32-13 in the JPS families. Distances are shown in brackets after the marker name. *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who were compatible with linkage to this region.*

Haplotype construction showed Family 6 was only compatible with linkage at D13S793 (Figure 7.2.4.1), and not at flanking markers. Furthermore, three family members would have to be non-penetrant as they also shared an allele at D13S793 (Figure 7.2.4.1).

Family 12 was not compatible with linkage to 13q31-32 markers as affected siblings 303 and 304 inherited different alleles from their affected father. The two-point LOD scores at D13S779 and D13S793 were positive because the phase of these alleles was not determined in the two-point analysis. In support of this, LOH analyses was performed with D13S779 and D13S793 on 5 cancers and

two polyps from Family 12. No LOH was observed for either marker in any of the seven Family 12 tumours. Together this evidence suggests that the 13q31-32 region is unrelated to disease development in Family 12. Families 10 and 18 shared half of their alleles at D13S793 and D13S779, and Family 19 shared all alleles at these markers. These families were therefore compatible with linkage to 13q31-32.

Table 7.2.4.4 shows a summary of which families were compatible with particular regions where the two-point LOD score was >1. New affected members and tumour material will help to confirm or refute each particular region as susceptibility loci for JPS. Importantly, no region was convincingly compatible with linkage in both of the two larger families, 12 and 15. However, the phenotype of Family 12 is different to that of the other families, and this may explain why there is failure to find any area of the genome which is compatible in this and the other families who are not linked to *SMAD4* (or *BMPR1A*), namely Families 6 and 15. Although Family 12 certainly have juvenile polyps as part of their phenotype, the polyposis is florid and aggressive leading to a preponderance of small bowel carcinomas rather than the colorectal cancers observed in the majority of the JPS patients. It is therefore not inconceivable that Family 12 have a distinct locus causing their disease. Alternatively, the markers used for the genome screen were not polymorphic enough to reliably detect linkage.

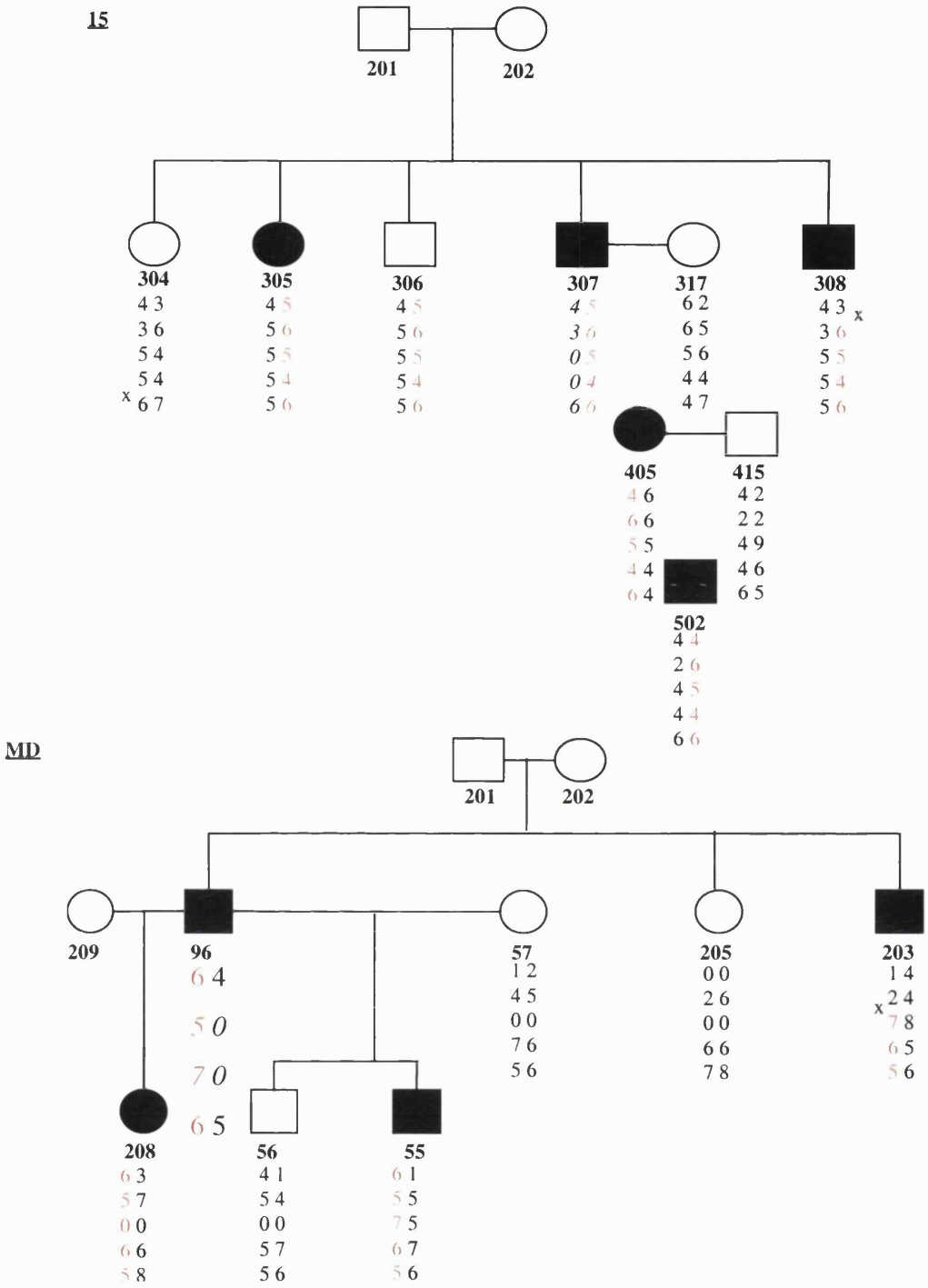


Figure 7.2.4.1 Chromosome 13q31-32 haplotypes in compatible JPS families. *Haplotypes shown were constructed from D13S800, D13S317, D13S793, D13S779 and D13S796. Shaded shapes represent affected individuals, and inferred haplotypes are italicised. Putative affected haplotypes are highlighted in red. Family 15 were compatible with linkage to 13q31-32 markers as affected individuals shared a haplotype. Individual 306 also shared the haplotype and would therefore be non-penetrant if the linkage were confirmed. Family MD were compatible with linkage to D13S793, D13S779 and D13S796. However, many individuals failed for D13S793 and therefore haplotypes have been inferred.* *Figure continued on next page.* 0 = not done.

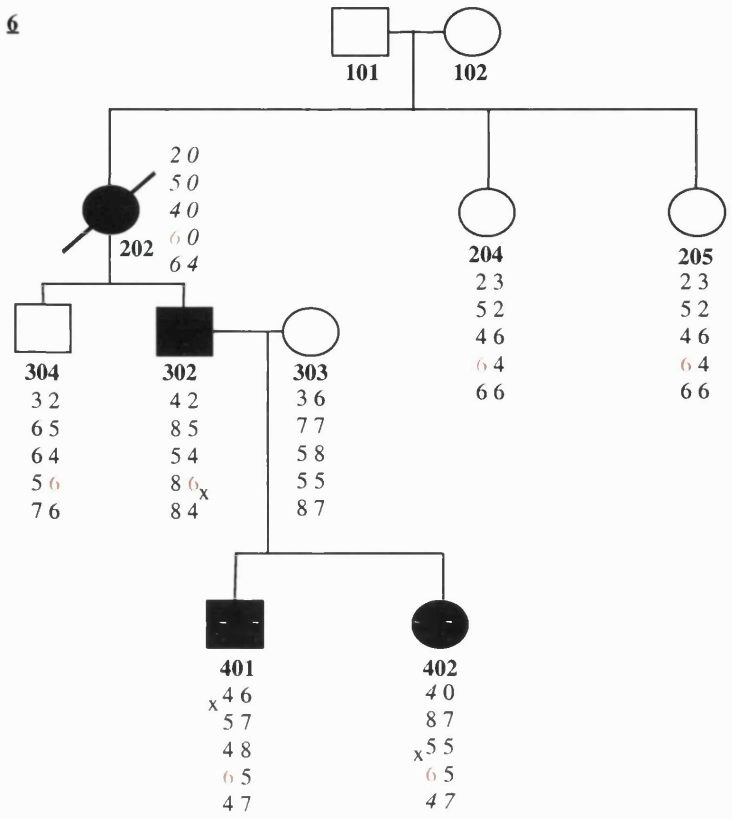


Figure 7.2.4.1. continued. *Family 6 were only compatible with linkage to D13S793, due to a crossover in 302. Individuals 204, 205 and 304 would also have to be non-penetrant were an underlying susceptibility locus confirmed.*

Family	*1q31-32*	*7p11-12*	*10q22-*	*11q22*	*12p12-*	*13q31-*
12	*Yes***	***YES****		*Yes***	*Yes***	
15	***YES***		*Yes**			*Yes**
MD			***YES****			*Yes*
6			***YES****			*Yes*
*10****	*Yes*	*Yes*	***Yes***	*Yes*	*Yes*	
*18****	*Yes*	*Yes*	***Yes***	*Yes*	*Yes*	
*19****	*Yes*		***Yes***	*yes*	*Yes*	

Table 7.2.4.4. Compatibility of JPS families to regions that gave a LOD score of >1 in the genome screen.
*Bold type shows the region that showed the most convincing evidence of linkage based on the number of unaffected individuals sharing the haplotype. *=unaffected individuals also shared the putative 'affected' haplotype. **person 307 in Family 12 would have to be a phenocopy if linkage were confirmed. ***=families subsequently found to have BMPR1A mutations.*

The *APC* gene has previously been excluded as the causative gene in this family (Leggett *et al.*, 1993), as their aggressive phenotype was considered to be not dissimilar to FAP. In addition, the two-point LOD scores for markers flanking the *APC* gene were significantly negative to exclude linkage to this gene (see Appendix One). Linkage to markers mapping to chromosome 7p11-12 in Family 12 was seemingly the most significant region of the genome for the marker set studied. DNA from individual 210 from Family 12 has recently become available, and this should increase the power to detect linkage. Further LOH analyses with new polymorphic markers performed on additional cancers from Family 12 may aid the confirmation or refutation of 7p11-12 as a candidate locus.

Although the genome screen did not identify new JPS genes, important lessons have been learned for the future that will aid the establishment of new JPS loci. Firstly, JPS is obviously a lot more heterogeneous than previously considered. The identification of *BMPR1A* mutations in a further subset of JPS families (discussed in Chapter Nine), has shown that there is certainly at least one more JPS gene accounting for those families not compatible with either *BMPR1A* or *SMAD4* (namely the large families 12 and 15). This has also shown that combining all the data for all the families may even obstruct the identification of JPS genes and each family, certainly the larger ones, should initially be considered separately (particularly in the light of phenotypic variance). Secondly, the inclusion of all family members (rather than exclusion of those who are considered unaffected) may increase the likelihood of identifying any real loci as

haplotypes can be constructed with confidence, and the more unaffected individuals who do not share alleles, the more likely that any candidate locus is real. Finally, the remaining gene or genes will almost certainly be tumour suppressors (as are *SMAD4* and *BMPR1A*) and inclusion of conservatively microdissected tumours at the time of a high density genome screen (in areas that look compatible) should help to pinpoint particular areas in the genome.

7.3 COMPARATIVE GENOMIC HYBRIDISATION

Comparative genomic hybridisation was performed using 1μg nick translated or DOP (degenerate oligonucleotide primed) labelled tumour DNA, all derived from paraffin embedded tissue (unless otherwise stated). A total of 32 tumours were analysed, comprising of 20 polyps and 12 carcinomas. None of the juvenile polyps analysed (from Families 12, 15 and MD, and sporadics LB, 1262 (fresh frozen material) and 1469 (fresh frozen material)) showed significant loss or gain of any genomic region (Figure 7.3.1). This may in part have been due to a number of important technical factors. Firstly, the sensitivity of CGH gives a level of detection of more than 10Mb for losses, and more than 2Mb for gains or amplifications. The genetic instability in pre-malignant polyps may not have reached these thresholds and would therefore not be detected in the polyps studied, or second hits may not have taken the form of deletions. Secondly, the sensitivity of the DOP experiment essentially relies on random priming.

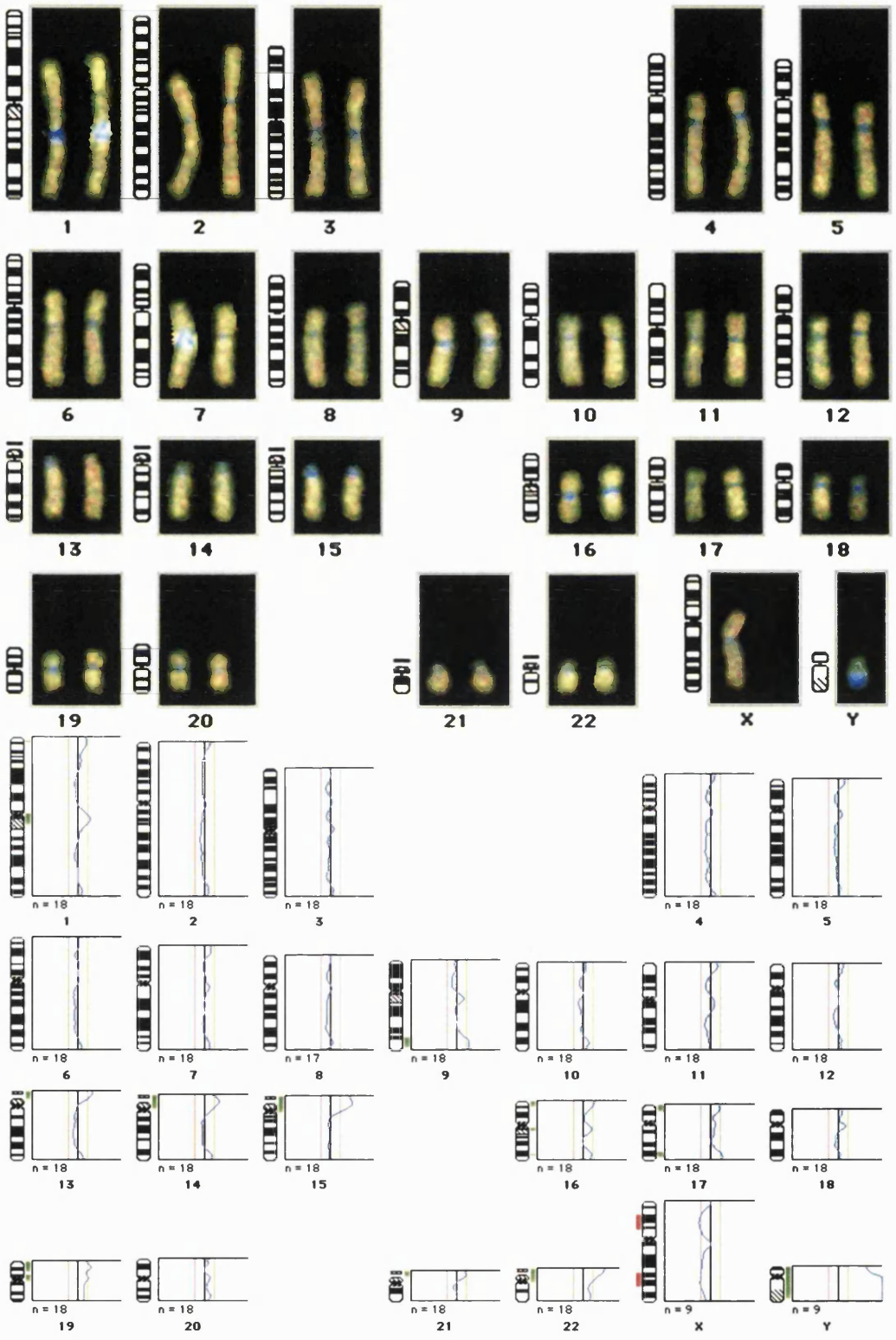


Figure 7.3.1. Comparative genomic hybridisation of a juvenile polyp.
Shown are CGH results for DOP-PCR-labelled juvenile polyp DNA from a Family 15 individual. No region showed significant loss of genetic material (red). Telomeric regions (e.g. chromosome 15) indicating apparent gain of material were spurious results common in DOP-PCR-labelled experiments.

It is quite likely that due to the poor quality of the paraffin embedded tissue DNA, the resultant DOP-PCR genomes were not representative of the tumour genomes. This would mean that even if the polyps did harbour deletions of tumour suppressor genes, they may not be detected.

Perhaps more surprisingly, none of 12 small bowel carcinomas from Family 12 showed significant loss or gain of any genomic region, and had essentially normal karyotypes (Figure 7.3.2). This was probably due to the factors described above, namely that the DNA was of poor quality (some of the cancers were removed fifty years ago). Alternatively, the microdissections may not have sufficiently excluded enough normal tissue, and this would have confounded the detection of regions of loss or gain. A third possibility is that small bowel cancers do not undergo the same high level of chromosomal losses and gains as classical colorectal tumours. There was no evidence from the LOH analysis that the Family 12 tumours were unstable at the microsatellite level (normally indicative of a defect in a mismatch repair gene, and consequently the presence of diploid rather than aneuploid tumours).

Prior to the identification of mutations in *BMPR1A* in a subset of JPS cases, a paraffin embedded tumour from the Finnish family 7/1 was nick-translated and CGH performed on the labelled DNA. Figure 7.3.3 shows the karyotype for this tumour and the associated CGH profile. A region just distal to the centromere of chromosome 16 was found to be lost in this tumour (a tubulovillous adenoma)

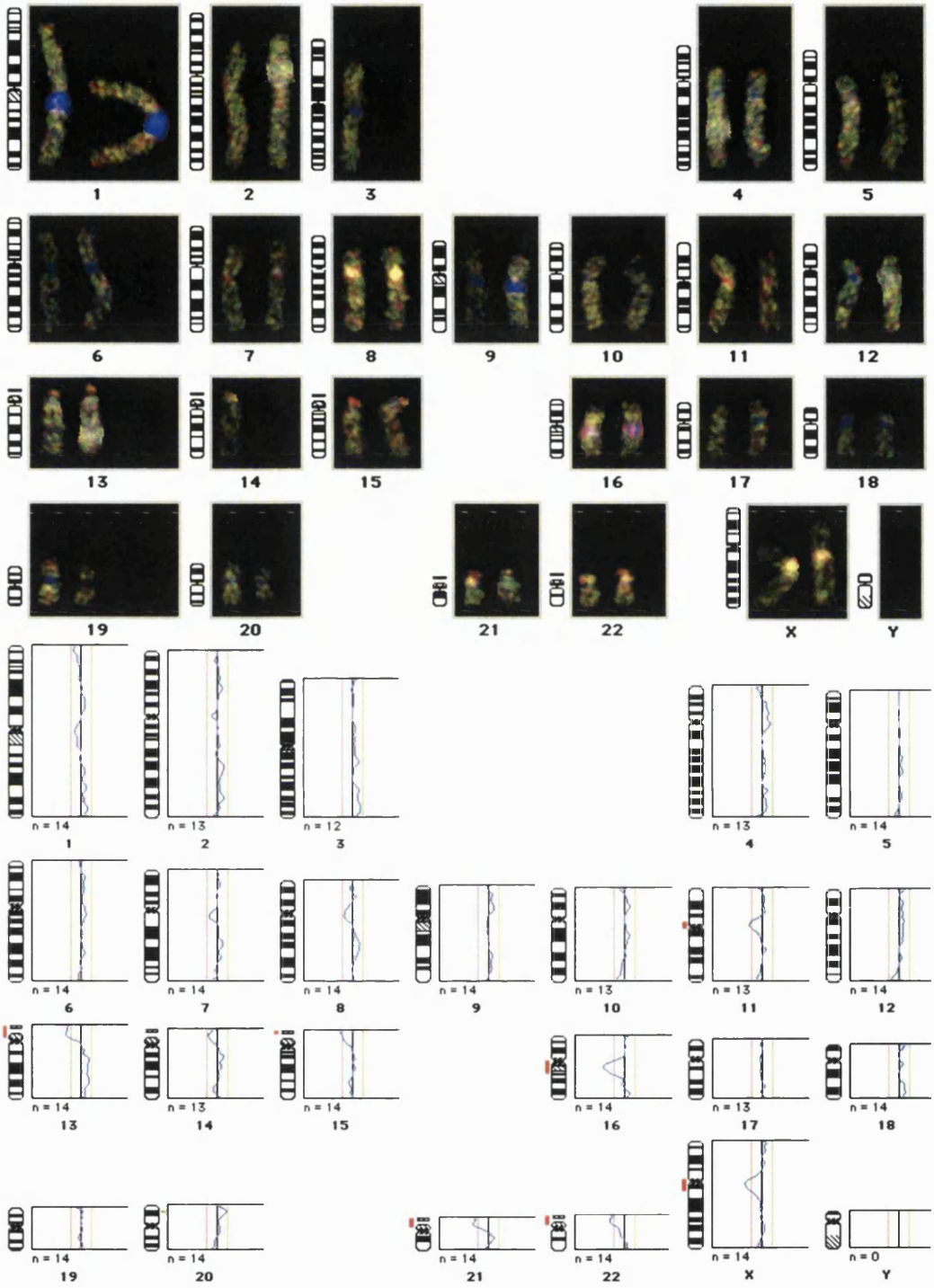


Figure 7.3.2. Comparative genomic hybridisation of a small bowel cancer.
Shown are CGH results(karyotype and profile) for DOP-PCR-labelled small bowel carcinoma DNA (from paraffin embedded tissue) from a Family 12 individual. The colours are poor, probably due to the old age of the DNA. No region showed significant loss of genetic material (red). Telomeric and centromeric regions (e.g. chromosome 13) indicating apparent loss of material were spurious results common in DOP-PCR-labelled experiments.

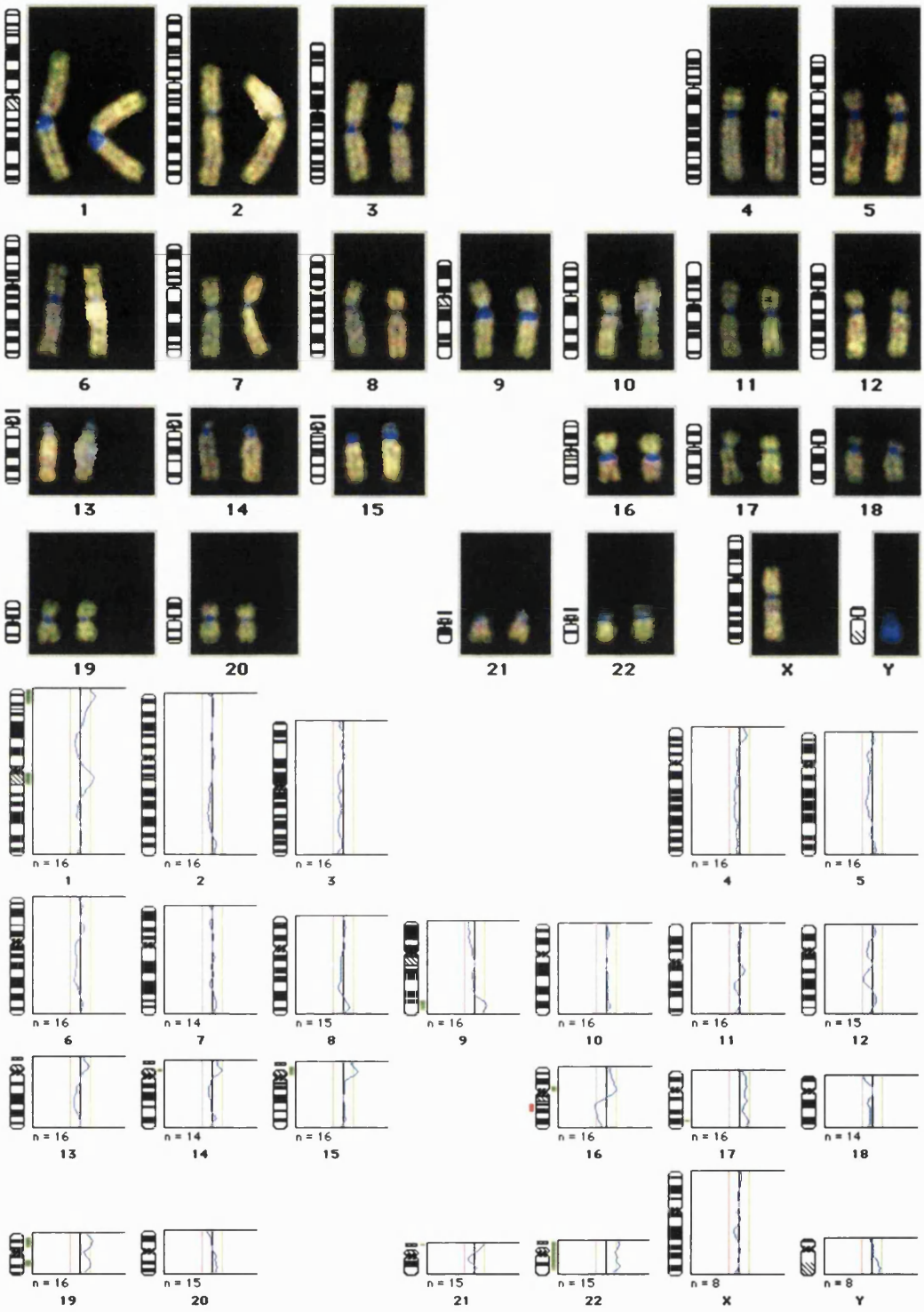


Figure 7.3.3. Comparative genomic hybridisation of a tubulovillous adenoma. *Shown are CGH results for nick translated labelled tubulovillous adenoma DNA (from paraffin embedded tissue) from a Family 7/1 individual. Significant loss was observed in all metaphases from band 16q21, indicated in the associated CGH profiles and apparent as a red region on the CGH karyotype.*

and therefore it was postulated that this region contained a tumour suppressor gene that was inactivated in the germline. The assessment of linkage analysis to 16q21 was performed with three markers taken from the genome screen set (ATA55A11, D16S3253 and GATA67G11), and this family (and indeed Family 15) were found to be compatible with linkage. LOH analysis was performed with these markers in tumours from Families 15 and 7/1 and 16q21 LOH was observed in one polyp from Family 15 (Figure 7.4.4), and the Family 7/1 tumour that showed loss on CGH.

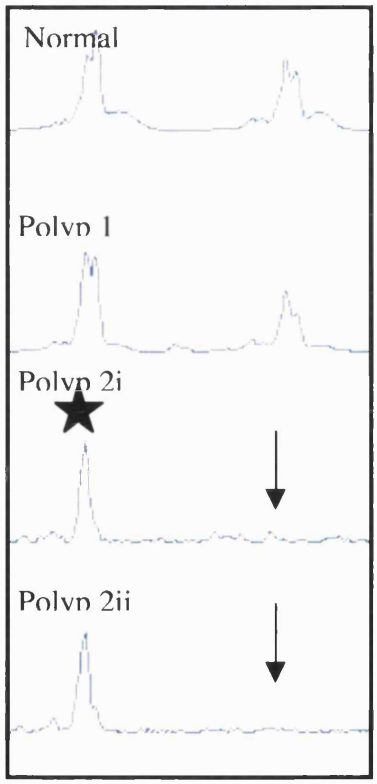


Figure 7.3.4 LOH analysis at 16q21 in a polyp from Family 15.
Shown are one polyp that did not show allele loss (polyp 1), plus two separate microdissections of a second polyp (polyp 2i and polyp 2ii) that showed allele loss at ATA55A11. The shared allele is shown with a star, and the lost putative wild-type allele is arrowed.

The significance of these results were short lived however, when a germline *BMPR1A* mutation was found to be the underlying genetic defect in Family 7/1, causing their JPS (see Chapter Nine). Interestingly, although all three of the component tumours from Family 7/1 were found to have LOH targeting *BMPR1A* on chromosome 10q (Chapter Nine), the 10q22 region did not show significant loss in the CGH analysis, confirming regions that undergo loss targeted to specific tumour suppressor genes may not necessarily be detectable by CGH analysis. The loss of 16q observed in the CGH analysis in the tumour from Family 7/1 may well reflect a real result, but rather than loss of a tumour suppressor inactivated in the germline, this represents the acquirement of genetic changes in a tumour as it progresses to malignancy. Secondly, further markers assessed in Family 15 (D16S527, D16S3110, D16S3039, D16S487 and D16S514, all mapping to 16q21), and the inclusion of a newly acquired DNA from an affected family member (son of 307) found this region to be no longer compatible with linkage. Explaining the loss observed at ATA55A11 in a polyp from this family is more problematic than in the 7/1 villous adenoma, but it may be related to tumorigenesis.

Results from the CGH analysis were therefore not found to be useful for identifying regions containing JPS tumour suppressor genes, as they were for the PJS gene (Hemminki *et al.*, 1997). Although juvenile polyps do show 'second hits' at loci mutated in the germline (i.e. *SMAD4* and *BMPR1A*, Chapters Three and Nine respectively), the majority of the time these losses are too small to be detectable by CGH. In addition, DOP-PCR labelling of genomes was probably

not sensitive enough for detecting losses and gains, and this may have been the result of poor quality DNA from the paraffin embedded tissue. LOH analyses were found to be much more sensitive for ascertaining loss of genetic material, but this generally relied on an *a priori* knowledge of a pathogenic germline mutation at a specific gene to target the LOH markers. Genome-wide LOH will be undertaken with tumour material from Families 12 and 15 in the future in the hope this will pinpoint regions harbouring JPS tumour suppressor genes. In addition, fresh frozen material is being collected in order to undertake expression analysis, using the latest chip-technology. Potentially, this will uncover differentially expressed genes between the polyp and normal tissue and does not rely on homing in on a genomic region or any prior knowledge of a genes candidacy.

7.4 CONCLUSIONS

A genome wide linkage analysis for new JPS genes did not uncover any proven genetic region associated with the disease. What has become clear is that there are probably at least two more JPS genes to be discovered, even after *BMPR1A*, and that the differences between the phenotypes of individual families may reflect this. For example, Family 12 whose phenotype is rather aggressive with a high frequency of small bowel carcinoma, probably do not share a JPS locus with the other large family, Family 15. Neither family is compatible with linkage to either *SMAD4* or *BMPR1A*. Overall, no region in the genome was compatible

linkage in both families. Several regions remain possible, for example 7p11-12 in Family 12 and 1p32-33 in Family 15. The confirmation or exclusion of these regions is anticipated with the aid of DNA from an affected member previously not available (person 210, a proportion only of whose alleles were previously deducible from his two children), and also new tumours that have been conservatively microdissected. With the new affected member of Family 15, and the potential to perform LOH in the tumours from this family, the identification of their causative locus should be assisted. In addition, a new large family (who have been screened and not found to harbour *SMAD4* or *BMPR1A* mutations) is being collected and this will obviously increase the power to detect new JPS loci. Contribution of the smaller families without *BMPR1A* mutations is somewhat smaller, and will largely depend on the identification of candidate regions in the larger families. The underlying genetic heterogeneity in JPS, and the lack of one large family who alone could provide sufficient power to detect linkage have made the identification of new loci, as yet, unfruitful in our family sets. With lessons learnt from the genome screen, however, I am hopeful that the remaining locus or loci will be uncovered.

CHAPTER EIGHT

IDENTIFICATION OF THE HEREDITARY MIXED POLYPOSIS SYNDROME LOCUS

IDENTIFICATION OF THE HEREDITARY MIXED POLYPOSIS SYNDROME LOCUS

8.1 INTRODUCTION

The Hereditary Mixed Polyposis Syndrome (HMPS, OMIM 601228) is characterised not only by 'typical' juvenile polyps, hyperplastic polyps with areas of dysplasia (serrated adenomas) and colonic adenomas, but also by the presence of single polyps that contain areas which are hyperplastic, areas which are adenomatous and also areas that are juvenile polyp-like (and hence the term 'mixed' polyp or 'atypical' juvenile polyp) (Whitelaw *et al.*, 1996) (Figure 8.1.1). The increased risk of cancer associated with these polyps is, like that in JPS, confined to the gastrointestinal tract, with a high frequency of colorectal malignancy. Most older individuals present with colorectal carcinoma whilst younger individuals tend to have polyps of either the atypical juvenile or hyperplastic type, strongly suggesting a natural history which entails progression from polyp to adenoma to carcinoma. One large family of Ashkenazi descent, SM96, was originally used to map the susceptibility locus for HMPS to 6q16-21 (Thomas *et al.*, 1996), although the causative gene was not identified for this autosomal dominant trait. Since the HMPS locus was mapped to chromosome 6, one individual from SM96 without the disease-

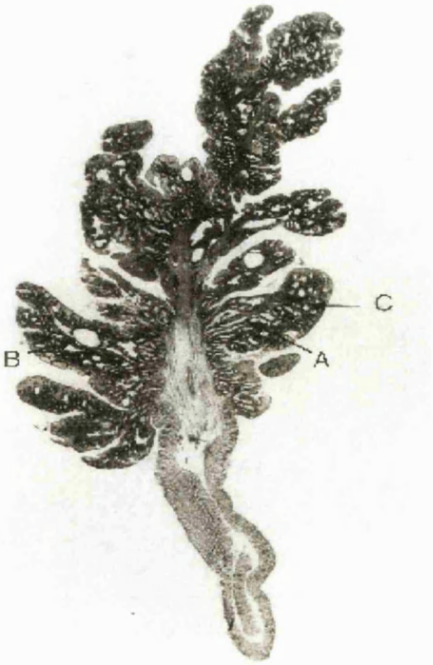


Figure 8.1.1. A mixed hyperplastic/juvenile/adenomatous polyp.
(A) hyperplastic area, (B) juvenile area, and (C) adenomatous area. (original magnification x7). Taken from (Whitelaw et al., 1996).

associated haplotype (patient 4.30) (Thomas *et al.*, 1996) has developed multiple colorectal adenomas, before the age of 40. At the time of the 6q linkage, this patient had only developed a single serrated adenoma (aged 29), and was thus considered to be a phenocopy. With the development of further adenomas, however, this cannot be upheld and strongly suggested that the reported location of the HMPS gene was incorrect. Dr Emma Jaeger (of Molecular and Population Genetics Laboratory, ICRF) confirmed this to be the case; SM96 was genotyped for three polymorphic markers not used in the original analysis (D6S1592, D6S1716, and D6S1580), spanning 7.4cM, which are located close to the reported site of the HMPS locus on 6q16-q21 and that were not available when the original linkage study had been performed. Two-point and multipoint LOD scores were uniformly negative and

haplotype construction confirmed that disease and 6q16-q21 alleles did not co-segregate (Figure 8.1.2). Individual 4.30 (the previous presumed phenocopy) was confirmed as not carrying the putative 'linked' haplotype. Furthermore, typing of additional markers revealed that one other individual (4.6) had developed adenomas without carrying the linked haplotype. This patient had almost certainly inherited the apparently identical-by-descent 6q markers from an individual who had married in to the family.

The following chapter describes how a new high density genome wide linkage search was performed (jointly with Dr Emma Jaeger) to identify the true HMPS locus. In addition, the identified predisposition locus was assessed for compatibility in juvenile polyposis families, given the clinical overlap between JPS and HMPS.

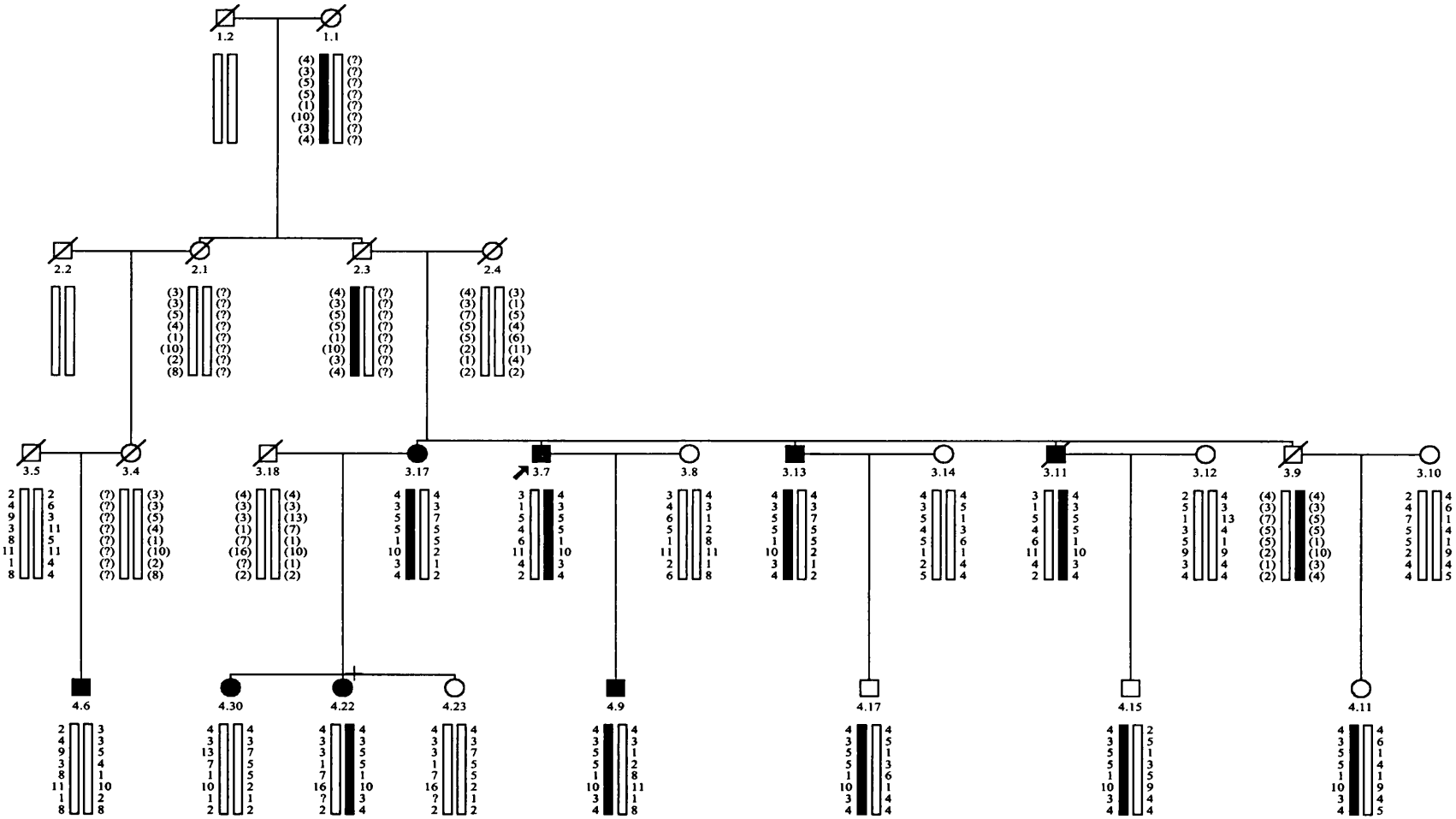


Figure 8.1.2 Evidence against linkage of the HMPS phenotype to chromosome 6q.
Haplotypes were constructed from the following markers; D6S1716, D6S468, D6S283, D6S434, D6S1580, D6S301, D6S1592 and D6S447. Black bars denote the haplotype that was previously thought to segrate with disease. Individuals 4.6 and 4.30 failed to share a haplotype with other affected individuals. Known affected individuals are indicated by a blackened symbol.

8.2 A GENOME WIDE SCAN FOR HMPS

In order to maximise the likelihood of identification of the correct HMPS locus, updated pedigree information was obtained from members of SM96, and the strictest criteria applied for the assignment of affection status. All clinicopathological data were re-verified from histology reports and unverified reports from patients were excluded. For linkage analysis, two affection statuses were applied to the patients and the data analysed separately. This was to ensure that there were not two separate diseases coincidentally occurring in the one large family, which would confound the detection of the true HMPS locus. Firstly, 'Q1 affected' patients were classified as those with three or more adenomas, or polyps with adenomatous areas (patients 4.6, 4.9, 3.11, 3.13, 3.19, 4.25, 3.17, 4.30, 4.22, 4.31 and 4.75) whereas 'HMPS affected' were classified with the more stringent criteria of the presence of mixed polyps (patients 3.7, 3.11, 3.13, 4.22 and 4.75). Spouses marrying into the family were classed as unaffected and all other individuals were classed as of 'unknown' affection status. Colorectal cancer, typical hyperplastic polyps, and extra-colonic tumours were all disregarded for the purpose of assigning affection status. Any family member who was known to harbour the missense I1307K variant in the APC gene (which has been associated with an increased risk of colorectal cancer (Frayling *et al.*, 1998)) were assigned as 'unknown' for the purposes of the genome screen.

DNA samples were available or extracted from established cell lines or blood for 57 SM96 family members who provided useful information for linkage analysis. PCR amplification (using standard conditions) of 387 microsatellite markers spaced at ~10cM intervals across the genome was performed on the 57 SM96 individuals using the Weber9 set (Research Genetics, Huntsville, AL). Both 'Q1' and 'HMPS' were modelled as dominant traits in the linkage analysis (q=0.001) with penetrances AA=0.75, Aa=0.75 and aa=0.001. The penetrance value of 0.75 used for the 'affected' homozygotes and heterozygotes allowed for possible incomplete penetrance of HMPS, whilst the 0.001 wild-type penetrance value allowed for the presence of phenocopies (more likely for the adenomas and carcinomas than the presence of the rare HMPS polyp). Two-point LOD scores were calculated for each marker using the subprogram MLINK (v5.1) of the LINKAGE program package (Lathrop *et al.*, 1984) as implemented in FASTLINK (v4.1) (Cottingham *et al.*, 1993). Multipoint analyses were undertaken using the VITESSE program (O'Connell and Weeks, 1995). Marker allele frequencies were taken from the Genome Database (http://www.gdb.org) or from the genotyping of pedigree founders.

Analysis of the new genome-wide screen data revealed that only one site in the genome provided good evidence of linkage for both 'Q1' and 'HMPS', thus confirming that indeed there was only one disease in SM96 that conferred slightly different phenotypes on different individuals. This region was on chromosome 15q14-q21, close to markers D15S165, ACTC, and D15S659 (Table 8.2.1).

Importantly, there was no evidence of linkage for the 'Q1', that is 'multiple adenoma', patients to 5q (the *APC* locus), or of 'HMPS' patients to 18q (the *SMAD4* locus). Furthermore, there was no evidence of linkage of either 'Q1' or 'HMPS' to 6q (the original putative HMPS locus). The genome wide two-point LOD scores for both 'Q1 and 'HMPS' are detailed in Appendix 2. A LOD score of >3.0 (corresponding to a significance level of 5%) is generally considered sufficient evidence of genetic linkage between the disease and test loci (and a LOD score of < -2.0 accepted as exclusion of linkage between the disease and test loci). A maximum two-point LOD score of 3.32 was found at ACTC and a maximum multipoint LOD score of 3.49 was also found at ACTC, providing good evidence for linkage of HMPS to 15q14.

Further markers mapping to this region were chosen to give a dense haplotype in an attempt to define the minimal region containing the disease gene (D15S1031, D15S1360, D15S1010, D15S144, D15S995, D15S1007, D15s1040, (ACTC), D15S971 and D15S118). Haplotype construction showed the minimal region containing the HMPS gene to lie between D15S1031 and D15S118, a 4.1cM interval (Figure 8.2.1). This shared region was shown to be highly penetrant with 18/20 individuals in SM96 sharing the haplotype and with confirmed symptoms of the disease.

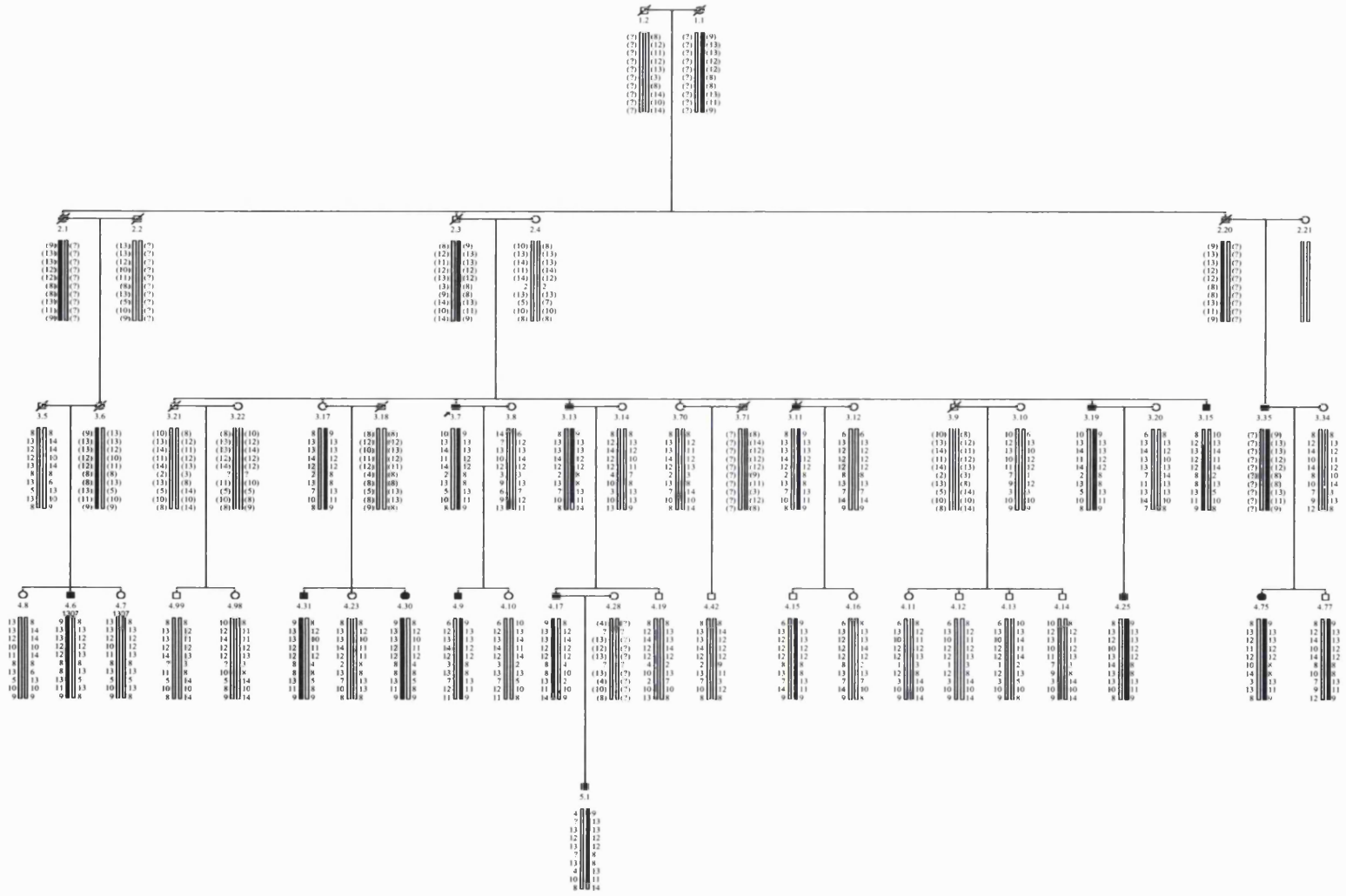


Figure 8.2.1 Pedigree of selected members of family SM96, showing haplotypes for the chromosome 15 markers:
D15S1031, D15S1360,D15S1010, D15S144, D15S995, D15S1007, D15S1040, ACTC, D15S971 and D15S118. Symbols are annotated as for Figure 8.1.2.

	Recombination Fraction					Approximate
Marker	*0.001*	*0.101*	*0.201*	*0.301*	*0.401*	*map position on Chr15*
D15S165	*1.00*	*1.51*	*1.32*	*0.93*	*0.45*	*27cM*
ACTC	***3.32***	*2.66*	*1.96*	*1.22*	*0.47*	*30cM*
D15S659	*1.34*	*1.08*	*0.82*	*0.55*	*0.28*	*42cM*

Table 8.2.1: Two-point LOD scores for chromosome 15q in SM96.
Three consecutive markers of the Weber 9a linkage analysis set showed positive two-point LOD scores for HMPS in SM96, with the highest LOD of 3.32 at marker ACTC.

Previously, a genome wide search performed on another Ashkenazi family (SM1311) mapped a new colorectal susceptibility gene, *CRAC1* (ColoRectal Adenomas and Carcinomas), to 15q14-q22 (Tomlinson *et al.*, 1999). The phenotype of this family includes large bowel adenomas of the tubular, villous, tubulovillous, and – notably - serrated histological types, as well as a high frequency of colorectal cancer. The linked 15q14-22 haplotype in SM1311 spanned a 40cM interval defined by D15S1031 and D15S153, a much larger region than that found for SM96. However, when CRAC1 and SM96 disease-associated haplotypes were compared, it was found that they were identical for markers shared within the HMPS region (D15S1031-D15S118).

Following this discovery, Dr. Emma Jaeger examined an additional Ashkenazi family with multiple colorectal adenomas (Family SM2952). No serrated adenomas or dysplastic hyperplastic polyps had been diagnosed in this family, although not all histopathologists use this classification. Typing of markers D15S1031 to D15S118 on chromosome 15q13-q14 in family SM2952 showed that all affected members shared the minimal HMPS region haplotype with affected members of SM96 and SM1311. The combined two-point LOD scores for the three families are shown in Table 8.2.2, with a highly significant LOD score of 5.49 obtained at ACTC. Furthermore, the multipoint LOD score of 7.44 was obtained near ACTC for the three families (Figure 8.2.2).

The initial genome wide screen for SM96 did not find evidence of linkage to 15q, and there are a number of possible explanations for this. Firstly, a few individuals previously classed as affected were re-classified for the new genome screen as 'unknown', because the original clinical data provided was not verified from hospital records (for example, patient 3.70) (Thomas *et al.*, 1996). In support of this prudent assignation of 'affected', these individuals do not share the new chromosome 15 disease-associated haplotype and so are indeed 'unaffected' as far as HMPS goes. Second, 30% of the general population will develop either a sporadic solitary adenoma, hyperplastic polyp or colon carcinoma by the age of 70, and for the previous genome screen SM96 family members who developed one of these were classified as affected. Patient 3.9 for example, is a highly plausible phenocopy,

developing colorectal carcinoma at age 63, but without any evidence of multiple adenomas or the characteristic HMPS phenotype.

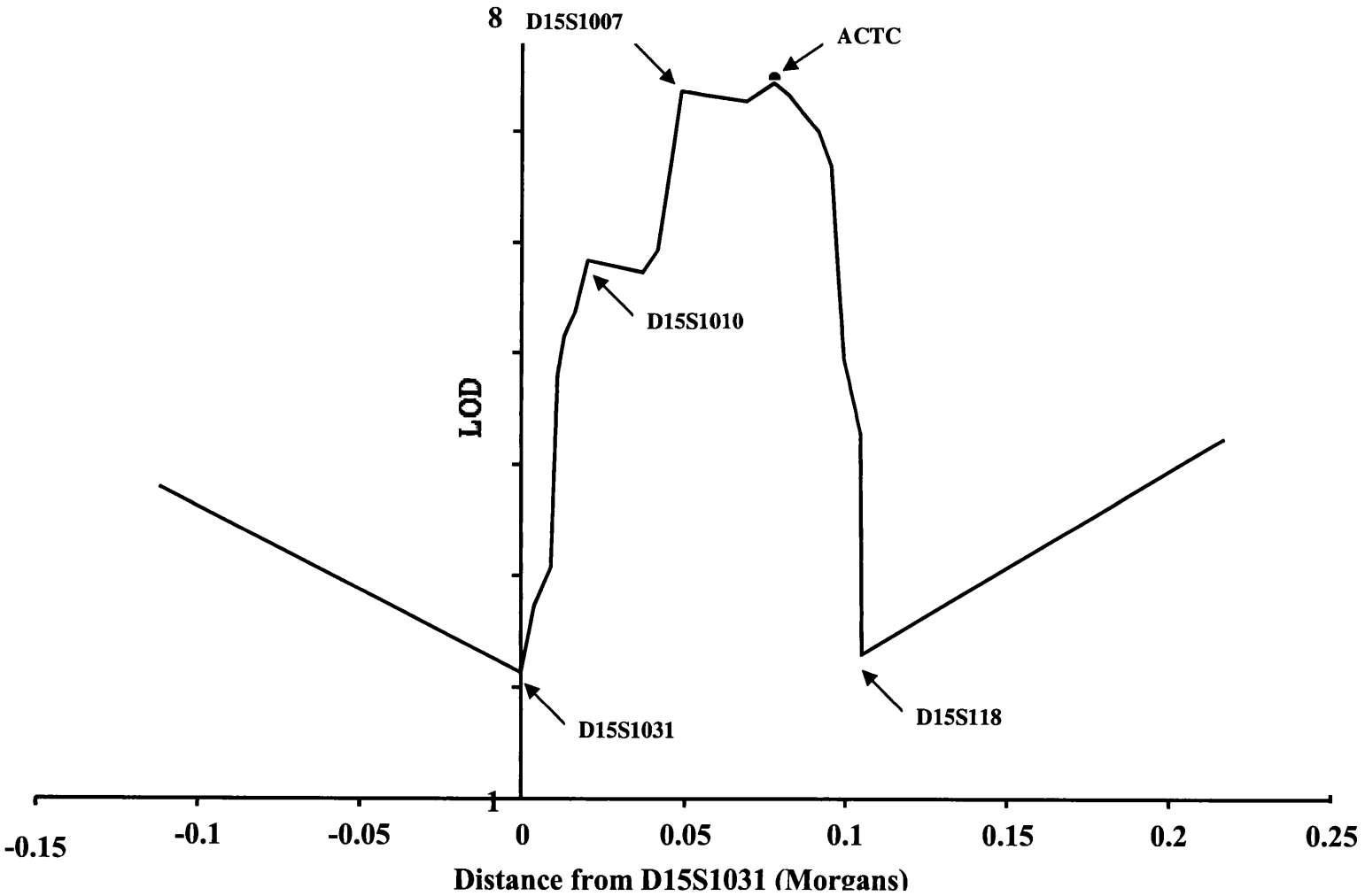


Figure 8.2.2: Multipoint HMPS linkage analysis between markers D15S1031 and D15S118.
Shown are the results of combined genotyping data from families SM96, SM1311 and SM2592.

Marker	*Recombination Fraction*					*Approximate map position on Chr15*
	0.001	*0.101*	*0.201*	*0.301*	*0.401*	
D15S1031	*1.45*	*2.16*	*1.68*	*1.04*	*0.40*	*26.0cM*
D15S1360	*1.10*	*0.82*	*0.51*	*0.24*	*0.07*	*26.9cM*
D15S1010	*3.65*	*2.77*	*1.90*	*1.07*	*0.38*	*28.2cM*
D15S144	*1.92*	*1.55*	*1.17*	*0.80*	*0.40*	*28.7cM*
D15S995	*3.26*	*2.42*	*1.57*	*0.77*	*0.17*	*28.8cM*
D15S1007	*1.68*	*1.19*	*1.21*	*0.87*	*0.35*	*28.9cM*
D15S1040	*3.26*	*2.48*	*1.76*	*1.13*	*0.54*	*29.3cM*
ACTC	***5.49***	*4.35*	*3.16*	*1.98*	*0.85*	*30.4cM*
D15S971	*1.72*	*2.32*	*1.73*	*1.02*	*0.32*	*30.7cM*
D15S118	*0.52*	*2.36*	*1.81*	*1.11*	*0.45*	*31.0cM*

Table 8.2.2: Two point LOD scores for 15q13-14 markers, using combined genotyping data from families SM96, SM1311, and SM2952.
The maximum LOD score obtained was 5.49 for ACTC.

It is most likely that the cancer of person 3.9 was sporadic and unrelated to an inherited susceptibility, substantiated by the fact his cancer was not at a particularly young age and none of his four children have been found to have colorectal adenomas despite regular screening. Whilst allowance was made for phenocopies in calculating LOD scores in the previous genome screen, this did not prevent incorrect chromosome 6q linkage for HMPS. The revised strategy of relying on individuals with a distinct phenotype to provide linkage information, and only including those who have confirmed affection status, has proved to be more prudent. Third, the development of more dense linkage maps allowed an apparently single 6q haplotype shared by affected individuals in the original

linkage study to be assigned as two different haplotypes, neither of which was disease associated. As mentioned above, Person 4.30 did not share the chromosome 6 disease haplotype in the previous analysis and was classified as a phenocopy to explain the presence of a serrated adenoma. With the development of further adenomas, this is less plausible, making it highly probable that she has inherited the susceptibility locus. Patient 4.30 does indeed carry the 15q disease-associated haplotype, and there is no need therefore to invoke a phenocopy explanation to justify the development of her adenomas.

On the assumption that a colorectal tumour predisposition gene may be a tumour suppressor gene (as are *APC* in FAP and *SMAD4* in JPS) component tumours from SM1311 and SM96 have been examined for loss of heterozygosity. Previously, only one of 23 adenomas from Family SM1311 showed consistent LOH for markers mapping to 15q14-22 (Tomlinson *et al.*, 1999), and preliminary LOH data for SM96 and for extra SM1311 tumours (performed by Dr Elinor Sawyer, MPGL, ICRF) also indicated that there is not a high frequency of LOH at the *HMPS/CRAC1* locus. This is unlikely to be attributable to contaminating stromal cells in the microdissection for the large number of tumours examined, and would therefore seem to indicate that either the causative gene is not a classical 'two-hit' tumour suppressor gene, or that the 'second' inactivating hit is via alternative mechanisms than loss of chromosomal material. Once the causative gene is identified and the germline defect in this gene is found, the role of the gene will be much easier to clarify. For example, immunohistochemistry should give an idea of whether the gene is a tumour

suppressor by examining the tumour tissue for loss of expression of the protein. Alternatively, loss of heterozygosity studies can be targeted directly to the causative gene on the assumption that the regions of deletions are rather small. Finally, the tumour DNA can be screened for 'second' hits at the causative gene, either conventional point mutations or small insertion/deletions, or the less common epigenetic inactivating mechanisms such as promoter methylation.

8.3 INVESTIGATION OF THE 15Q HMPS LOCUS IN JPS FAMILIES

Due to the overlap in phenotype between JPS and HMPS (namely the presence of juvenile polyps in HMPS) the 15q13-14 HMPS region was examined for linkage in the JPS families. Initial examination of the two-point LOD scores at the HMPS locus markers indicated that Families 1, 5, 6, 12 and 15 were compatible with linkage (Table 8.3.1). The HMPS phenotype appeared to be common families with Ashkenazi ancestry and given that Family 14 was from Israel, it was considered prudent to formally exclude the HMPS region in this family. Further markers (D15S144, D15S1007 and D15S1040) mapping near the minimal HMPS region were therefore genotyped in families 12, 14 and 15. Haplotype construction for these markers is shown in Figure 8.3.1. Families 12 and 14 were not compatible with linkage to the HMPS region as affected relatives failed to share a haplotype.

	D15S822	*D15S165*	*ACTC*	*D15S659*	*D15S643*
1	***0.3***	***0***	***0.26***	***0.3***	***0.3***
MD	*-1.54*	*0.23*	*-1.58*	***0.23***	***0.27***
5	***0.3***	***0.3***	***0.3***	***0.3***	***0***
6	*-1.77*	***0***	***0***	***0.57***	***0.48***
C1	*-1.77*	*-1.77*	*-1.77*	*-1.77*	*0*
*10**	*-0.04*	*-0.04*	*0.25*	*-0.08*	*-0.08*
12	***0.89***	***0.20***	***0.21***	*-1.72*	*-1.63*
14	*-1.77*	*0.08*	*-1.77*	*0.22*	*-1.77*
15	***0.9***	***0.42***	***0.19***	***0.07***	*-3.89*
*18**	*0*	*0.12*	*0*	*-0.04*	*0.28*
*19**	*-0.08*	*-0.08*	*0*	*-0.08*	*-0.08*
*7/1**	*0*	*0*	*0*	*-2.16*	*0*
2/13	*0*	*0*	*0*	*0.7*	*0*

Table 8.3.1. Two-point LOD scores for chromosome 15q, the HMPS locus. *Shown are the two-point scores for θ=0 for markers mapping to chromosome 15 in the JPS families. *= families subsequently shown to harbour BMPR1A mutations. Bold type shows families who are compatible with linkage to this region.*

Family 15 were compatible with the HMPS region, but only on the assumption that person 308, who had developed three small adenomas in his sixties, was a phenocopy. In addition, several members of this family classified as unknown also shared the putative 'affected' haplotype. There is the possibility that these individuals were indeed affected, given that endoscopies had not been performed on all family members. The polyps from Family 15 have been histologically examined (discussed in Chapter Three), and include true juvenile polyps, hyperplastic polyps and adenomatous polyps. These three types of tumours (as well as carcinomas) are found in the HMPS syndrome. If person 308 from Family 15 was indeed a phenocopy, mutations of the HMPS gene may not only be confined to families of Ashkenazi ancestry.

The search for the HMPS gene is underway. Once the gene is identified, it can obviously be screened in JPS families who were compatible with linkage to the region (1, 5, 6 and 15), and also in sporadic JPS cases. This will clarify with certainty whether there is genetic, as well as phenotypic, overlap between JPS and HMPS. The contribution of the HMPS gene to colorectal tumours (adenomas, juvenile polyps, hyperplastic polyps, as well as carcinoma) outside HMPS syndrome will obviously be important to establish, and it will be interesting to ascertain whether the gene is only mutated in those of Ashkenazi descent.

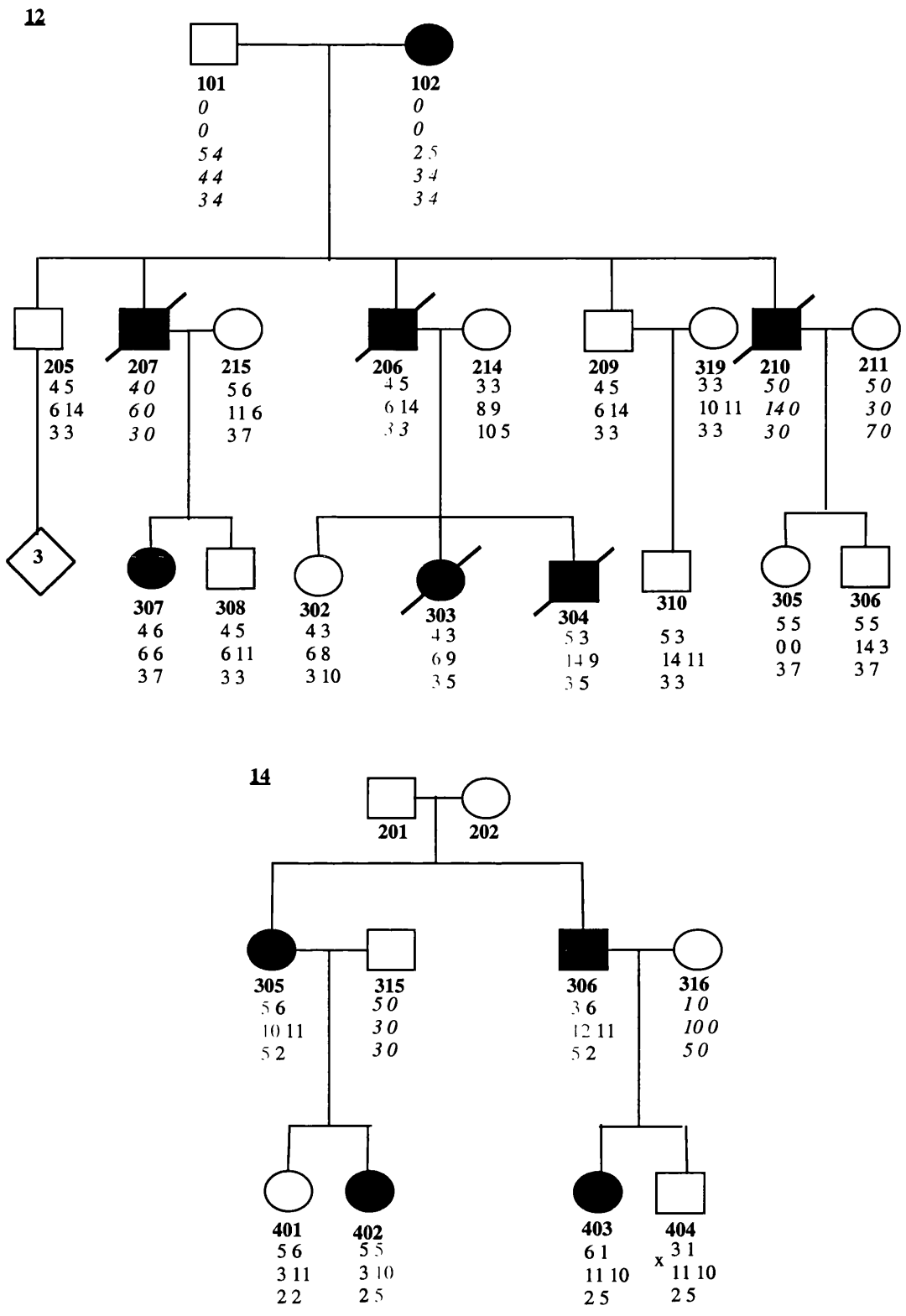


Figure 8.3 .1 15q haplotype in JPS families, continued.

Shown are the haplotypes constructed from three markers mapping to the HMPS region (D15S144, D15S1007 and D15S1040). Symbols representing affected individuals are shaded black. Inferred alleles are italicised. Possible phenocopies are shaded grey. Family 12 was not compatible with linkage to these markers, indicated by affected siblings 303 and 304 inheriting different alleles from their affected Father. Family 14 was not compatible with linkage to 15q markers, as affected persons 306 and 402 did not share a haplotype.

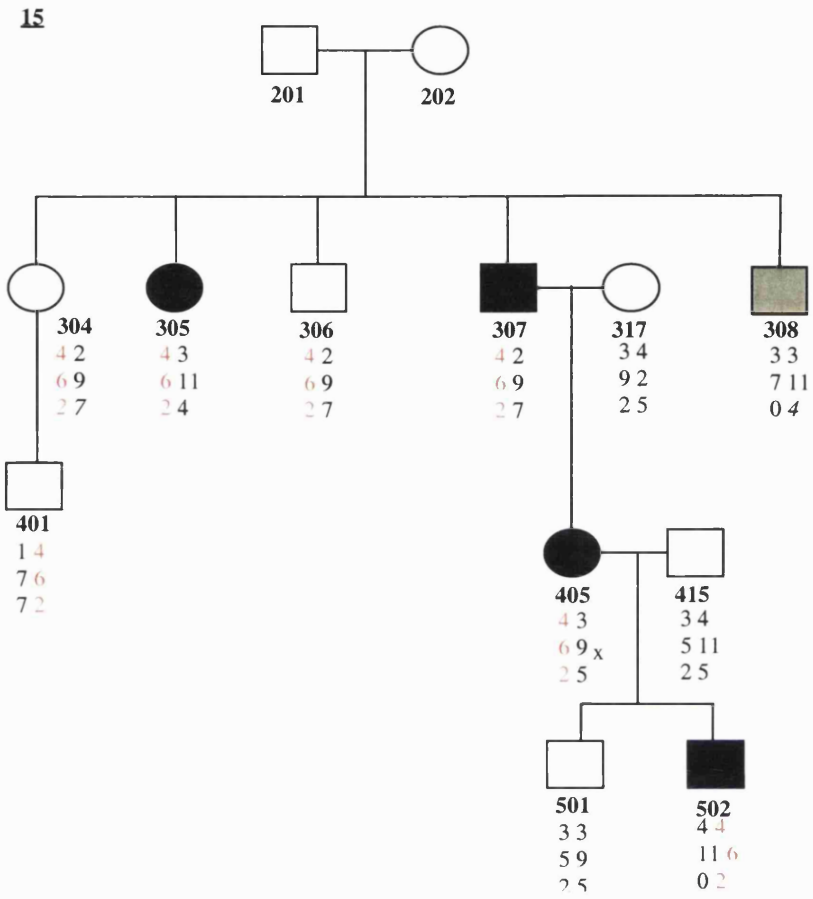


Figure 8.3.1 continued. *Family 15 was compatible with linkage to the 15q markers, but with the assumption that individual 308 (who has had three adenomas) was a phenocopy. Additionally, 304, 306 and 401 also shared the putative affected haplotypes.*

8.4 CONCLUSIONS

The new genome screen for HMPS has shown that the HMPS gene is not located on chromosome 6q16-21 as previously reported, but is located on chromosome 15q14-q21. Furthermore, families SM96, SM1311 and SM2952 share a haplotype which is common to all confirmed affected individuals in these families (details not shown). Taken together with the highly significant LOD scores obtained from two-point and multipoint analyses, this suggests that the region containing the HMPS gene overlaps with that of the CRAC1 gene, or more likely, that the CRAC1 gene may be identical to the HMPS gene. All three families share the phenotypic features of multiple colorectal adenomas and carcinomas. Although SM2952 have not been reported to have dysplastic polyps or serrated adenomas, both SM1311 and SM96 have had these lesions. In addition, a few members of SM96 have mixed or atypical juvenile polyps. The absence of these in SM2952 and SM1311 is more likely to be because they are rarer and may not be classified by all histopathologists rather than a true phenotypic difference between the families. Alternatively, the presence of the characteristic HMPS polyp may indeed be confined to SM96 and due to modifier loci. Given that SM96 individuals are spread throughout the world, it is unlikely that an environmental effect is influencing the development of the HMPS polyps, therefore the presence of these polyps is genetic.

Not only do affected individuals from SM96, SM1311 and MO all share the same haplotype (D15S1031 to D15S118), but recent genotyping (by Emma Jaeger) of individuals from two additional Ashkenazi families, one with a history of multiple colorectal adenomas and cancer (SMU) and one with mixed hyperplastic/adenomatous polyps (RF) indicates that they too carry the minimal 15q HMPS haplotype. This is compelling evidence for the presence of a founder mutation being carried on an ancestral haplotype. This as yet unknown gene is therefore a high-penetrance colorectal tumour predisposition gene, which may not only explain the increased prevalence of colorectal tumours in the Ashkenazi population, but may have a role in the development of sporadic cancer. Park *et al* studied 70 sporadic colorectal tumours (26 adenomas and 44 invasive carcinomas) for loss of heterozygosity around the *CRAC1* locus (Park *et al.*, 2000). None of 24 informative adenomas studied showed LOH around CRAC1, whereas 14/40 (35%) informative carcinomas showed convincing allele loss of 15q14-22. If the same gene is being targeted for loss in these sporadic tumours that is mutated in the germline of SM96 etc., this would seem to suggest that *HMPS/CRAC1* is indeed a tumour suppressor gene. Importantly, Park *et al* microdissected nests of carcinomas completely free of contaminating stromal tissue and this would aid the detection of true LOH (Park *et al.*, 2000). The sporadic adenomas, and the SM1311/SM96 adenomas which do not show loss may therefore be contaminated with too much normal tissue that would confound the detection of LOH. Further evidence for the role of 15q in sporadic colorectal tumorigenesis is provided by two comparative genomic hybridisation (CGH) studies which found >10% of cancers (De Angelis *et al.*, 1999; Paredes-Zaglul *et al.*, 1998) to have loss of 15q.

Although screening of candidate genes and ESTs (expressed sequence tags) mapping to the shared region is underway to determine the SM96/SM1311/SM2952 causative gene, no pathogenic mutation has yet been identified. What confounds the identification of the gene is that rather than a spectrum of mutations, a proportion of which would be detected via mutation screening, the mutation is a founder defect common to all the families. This means that a single change, possibly very minor and/or cryptic, may account for all the colorectal tumours in SM96, SM1311 and SM2952 and this may take time to elucidate.

CHAPTER NINE

MUTATIONS IN THE *BMPR1A/ALK3* GENE CAUSE A FURTHER SUBSET OF JPS CASES

MUTATIONS IN THE BMPR1A GENE CAUSE A FURTHER SUBSET OF JPS CASES

9.1 INTRODUCTION

Using linkage analysis, Howe *et al* recently assigned a new JPS susceptibility locus to chromosome 10q22 (Howe *et al.*, 2001). Four large families, comprising 57 individuals of whom 27 were known to be affected with JPS, gave a maximum LOD score of 2.33 at θ=0.10 with the marker D10S573. *PTEN* mutations had already been excluded as the causative defect. Finer mapping of the region gave a maximum LOD score of 4.74 with ALK3CA, which is situated just upstream of the bone morphogenetic protein type 1 receptor A (*BMPR1A*) gene, and subsequently, pathogenic mutations segregating with disease were found. The *BMPR1A*, also known as *ALK3*, gene maps to chromosome 10q22 between D10S2327 and GATA115E01 (Howe *et al.*, 2001) and encodes a serine-threonine kinase which belongs to the TGFB receptor – SMAD superfamily (Massague, 2000), acting as the BMP equivalent to TGFβR1.

This chapter describes how the JPS cohort were screened (in collaboration with Dr Charis Eng, Ohio State University) for germline *BMPR1A* mutations. Linkage analysis was also performed to assess the compatibility of *SMAD4*-negative families with linkage to the *BMPR1A* locus. In order to establish whether, like

SMAD4, this gene acts as a tumour suppressor in JPS, loss of heterozygosity analysis was performed on tumours from families found to harbour *BMPR1A* mutations.

9.2 SCREENING JPS FOR GERMLINE MUTATIONS IN BMPR1A

Exon-by-exon amplification (including exon/intron boundaries and flanking intronic sequences) of the coding exons of the *BMPR1A* gene, using the primers detailed in Table 9.2.1 was performed in all JPS families and sporadic cases who had no detectable germline *SMAD4* mutation. Standard PCR conditions were used (35 cycles with an annealing temperature of 55°C and 1.5.mM Mg^{2+}). families 1, 3(a.k.a. 1868), 5, 6, 10, 11(a.k.a. FT), 12, 14 ,15, 16, 18, 19, 22, MD, YC, WN, HR, JP2/13 and JP7/1, and sporadic cases KS, WH, BN, 1262, DM, SM316 (a.k.a HG), BW, RV, 1469, LB, CR1, FD, RH, JP1/1 and JP8/1). PCR products were then directly sequenced to search for germline mutations.

Eleven of 34 (32%) JPS patients were found to harbour pathogenic germline *BMPR1A* mutations. The *BMPR1A* mutations of families 10 ,11, 16, 18, 19 and 7/1 and sporadic cases JP8/1, RH, RV, SM316 and 1469 and their predicted effects are detailed in Table 9.2.2.

Exon	Sense primer 5'-3'	Antisense primer 5'-3'
1	5'-TCCAAAATTCAGTTGTATTCC-3'	5'-CACATACATTACTAAAATGAACACTG-3'
2	5'-GTCACGAAACAATGAGCTTT-3'	5'-TTAAGAAGGGCTGCATAAAA-3'
3	5'-CATTCAGACTCAAATTTCGTT-3'	5'-TCTCATGGGTCCCAAATTA-3'
4	5'-CCAAACCATTTCTAATTTTATCA-3'	5'-CATGCTCCGACTTTTCTC-3'
5	5'-CCAGGCTACCTAGAATTGAA-3'	5'-AACAGCGGTTGACATCTAAT-3'
6	5'-CCTCAAGGTTTTTCTTAGGG-3'	5'-TCAACACACCATTCATGTCT-3'
7	5'-TCATCAAGAGCTCAAACCTT-3'	5'-ACCTCACTAGCCTTGTCAAA-3'
8	5'-CCCTAGCCTATCTCTGATGA-3'	5'-AACAGTGGGGCAAAGAAC-3'
9	5'-TATTTTATTTTTGGCCCTCA-3'	5'-TGATGAGTAAATCAACATAATCAG-3'
10	5'-ATTTTTGTGCCCATGTTTT-3'	5'-AATCACTTCTTCAGGGGACT-3'
11	5'-ACTCAGTCCCCTGAAGAAGT-3'	5'-CTAGAGTTTCTCCTCCGATG-3'

Table 9.2.1 BMPR1A primer sequences

ID	BMPR1A exon (1-11)	BMPR1A mutation	Predicted effect
16*	7	c.826-7 del GA	Truncated protein
18*	2	S44X	Truncated protein
10*	8	R361X	Truncated protein
19*	7	R273X	Truncated protein
11*	6	c.665 ins T	Truncated protein
SM316	5	IVS5-1g/t	Skipping of exon 6
RV	7	c.784-805 del 22 bp	Truncated protein
RH	10	c.1469 ins T	Truncated protein
1469	4	C124R	Missense protein
JP7/1*	8	C376Y	Missense protein
JP8/1	1	CC,IVS3-3c/g	Skipping of exon 1

Table 9.2.2 BMPR1A mutations in JPS patients.
**familial. The exons are numbered 1-11, counting only the coding exons.*

Three of the mutations occur in the extra-cellular domain (JP8/1, 18 and 1469). The truncating mutation of Family 18 would be predicted to result in a very short peptide without a transmembrane domain. The missense change seen in 1469 was not seen in 100 normal control chromosomes (or the missense mutation

observed in JP7/1), and results in the loss of disulphide bond thus giving rise to conformational alterations (Kirsch *et al.*, 2000). The splice mutation IVS5-1g>t would be predicted to result in a receptor without a transmembrane domain. Family JP7/1 has a missense mutation in the middle of the kinase domain, C376Y. Residue 376 lies within the kinase domain and is highly conserved among species, from *C. elegans* to mouse and rat. Assuming that the mutant mRNAs and the truncated proteins were stable, the truncations in patients 10, 11, 16 ,19, JP7/1, RH, RV and SM316 all leave an intact transmembrane domain but are either lacking all or part of the kinase domain. Thus, all *BMPR1A* mutations resulted in a receptor whose function was either abrogated or impaired.

9.3 ASSESSMENT OF ALLELE LOSS AT BMPR1A IN JPS TUMOURS

On the assumption that *BMPR1A* may act as a tumour suppressor gene, in much the same way as *SMAD4* has been found to in JPS, loss of heterozygosity analysis was performed on tumours with three markers mapping close to *BMPR1A*. Firstly, D10S573 that lies just centromeric to *BMPR1A*, then ALK3GGAA that lies 76Kb upstream of *BMPR1A*, and finally ALK3CA which lies 49Kb upstream of *BMPR1A* exon 1.

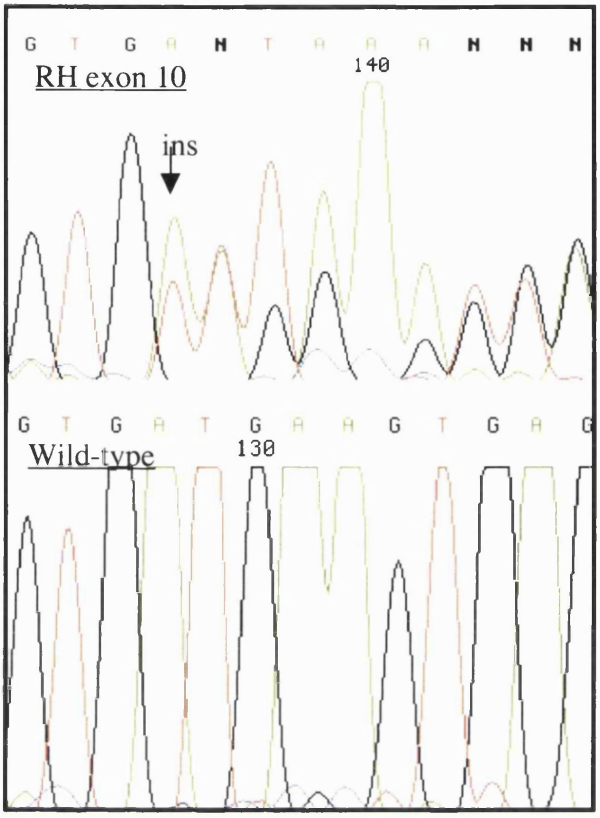


Figure 9.2.1 BMPR1A mutation in JPS patient RH.
Mutant sequence is shown above, and wild-type below. An insertion of a T at nucleotide 1469 in exon 10 results in a premature stop at codon 491.

In cases where these markers were uninformative, two markers flanking *BMPR1A* were used instead (D10S2327 and GATA115E01). Three tubular villous adenomas were available from Family 7/1 (who harbour a missense mutation) and two juvenile polyps were available from Family 18 (who possess a truncating mutation in exon 2). In addition, 17 polyps and cancers from Families 6, 12, 2/13, and sporadic cases WH and CR1 (in whom no *BMPR1A* mutation had been demonstrated) were also assessed for LOH at the five *BMPR1A* markers. The results of the LOH analysis are shown in Figure 9.3.1. and 9.3.2. All three tumours from Family 7/1 showed loss at all informative markers

mapping close to BMPR1A, and although one polyp from Family 18 was uninformative for these markers, the other juvenile polyp showed LOH at flanking markers. No LOH at *BMPR1A* was observed in the 17 polyps and cancers derived from patients without BMPR1A mutations. This indicates that the missense mutation observed in Family 7/1 is indeed pathogenic, and that *BMPR1A* acts as a tumour suppressor gene in JPS, as does the other JPS gene, *SMAD4*. Bi-allelic inactivation of *BMPR1A* presumably therefore initiates the growth of the polyp.

It is unclear how this high level of allele loss observed at 10q22 in *BMPR1A* germline mutation carriers is related to the high level of LOH reported at JP1 by Jacoby *et al* (Jacoby *et al.*, 1997). JP1 was first identified when an individual with juvenile polyps was found to have a germline interstitial deletion at 10q22-44. With the subsequent identification of the Cowden disease locus (*PTEN)* at 10q23, the juvenile polyps observed in this patient (and the LOH observed in 39/47 juvenile polyps from other patients) were believed to actually be from Cowden patients rather than true JPS patients. Methodological questions were also raised about the fluorescent *in situ* hybridisation techniques employed by Jacoby *et al* (discussed in Chapter Three), which indicated that the cells targeted for loss were the inflammatory lymphocytes, which was highly unlikely. However, now it has been shown that there does indeed exist a JPS susceptibility locus at 10q22, the cells targeted for deletion can be accurately investigated, and

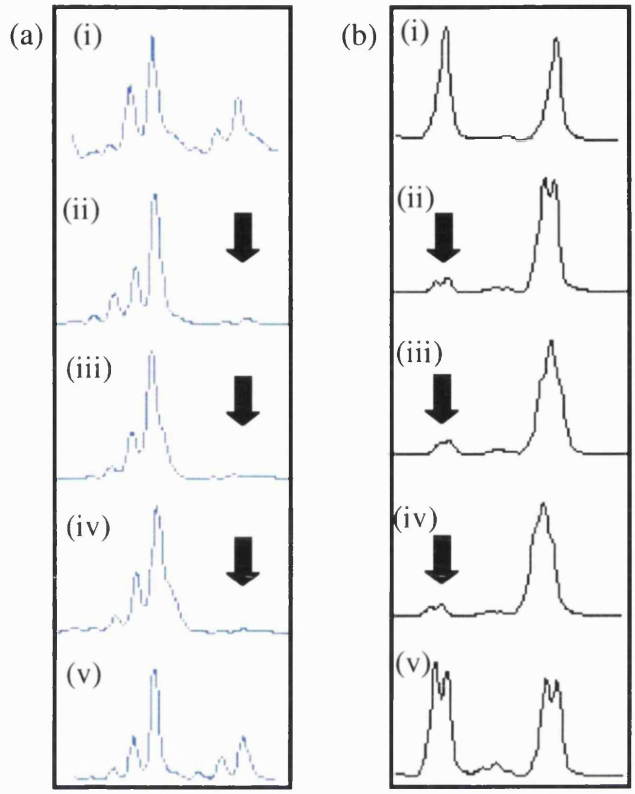


Figure 9.3.1. LOH analysis at BMPR1A in Family 7/1.
Shown are the LOH results from three tubular villous adenomas with markers (a) ALK3CA and (b) ALK3GGAA. (i) normal DNA extracted from blood. (ii)-(iv) tumours and (v) normal tissue extracted from the same slide as tumour (iv). The lost allele is arrowed.

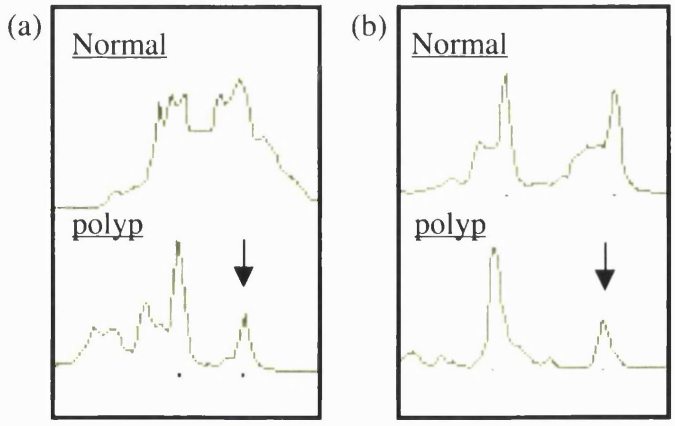


Figure 9.3.2 LOH analysis at BMPR1A in Family 18.
Shown is one polyp from Family 18 at markers D10s2327 and GATA115E01. Allele loss is arrowed. The patient was non-informative at all three markers mapping more closely to BMPR1A.

it is likely that they will be the epithelial cells of the juvenile polyps and other tumours arising as a result of *BMPR1A* germline mutations. The LOH analysis performed on Families 7/1 and 18 has confirmed that juvenile polyps, whether they arise as a result of a second hit at *SMAD4* or *BMPR1A*, are clonal lesions and as such are true neoplasms, and the clonal component almost certainly includes the epithelium rather than the inflammatory lymphocytes.

9.4 COMPATIBILITY OF JPS FAMILIES TO BMPR1A REGION

Although the JPS families were investigated for germline *BMPR1A* mutations via direct sequencing of the gene, there remained the possibility that undisclosed or cryptic mutations were responsible for more JPS cases than already established. Haplotype analysis using the three markers mapping closely to *BMPR1A* (D10S573, ALK3GGAA and ALK3CA) or markers flanking the BMPR1A gene (D10S2327, GATA115E01 and D10S677) was therefore performed in families 1, 5, 6, 12, 14, 15, MD and C1 in order to establish whether indeed the families were compatible with linkage to this region. In this way, *BMPR1A* could be screened in compatible families via other means, and confidently excluded in those families not compatible with linkage. Table 9.4.1 shows the two-point LOD scores obtained from the genome screen for chromosome 10 markers flanking *BMPR1A*.

Importantly, the power to detect linkage in the families shown to harbour *BMPR1A* mutations (Families 10, 18 and 19) in our genome screen was very small. Demonstrating this, the two-point LOD scores at D10S2327 were negative in Families 10 and 18 , and zero in Family 19. This was probably due to lack of informativity at this marker, or the inability to determine the phase (i.e. which parent each allele has been inherited from). The haplotype construction (Figure 9.4.1) for Family 18 (who harbour a *BMPR1A* mutation) clearly showed that two affected siblings shared alleles at D10S2327 and GATA115E01, but due to poor informativity, the phase was not determinable and thus negative two-point LOD scores were obtained. Multipoint analyses were performed with D10S1432, D10S2327, GATA115E01 and D10S677, with the highest score for all the families being 0.25, proximal to D10S1432. Exclusion of the families compatible with linkage to *SMAD4* (1, 5, 14 and C1) gave a multipoint score of 0.42, again proximal to D10S1432. The multipoint score for the actual position of *BMPR1A* was negative, largely because the two largest families, 12 and 15, were not compatible with linkage to this region, as demonstrated by haplotype construction. The haplotypes for the *BMPR1A* region for families 1, 5, 6, 12, 14, 15 , MD and C1 are shown in Figure 9.4.1.

Family	GATA121A08 (84cM)	D10S1432 (92cM)	D10S2327 (100cM)	GATA115E01 (113cM)	D10S677 (118cM)	D10S1239 (126cM)
1	**0.21**	**0**	**0**	**0**	-1.82	0
5	0	0	-1.82	0	-1.82	0
6	-0.38	0.31	-0.5	0.32	0	0.1
10*	0.17	0.22	-0.04	0	0.22	0.21
12	-0.43	0.26	-3.78	0.34	-0.06	-1.57
14	**0.36**	**0.5**	**0.22**	**0.24**	-1.63	-1.69
15	-0.11	-1.88	-1.9	0.3	-1.68	-1.64
18*	0.19	0.21	-0.07	-0.29	-0.29	0.28
19*	0.15	0	0	0.26	0.2	0.2
C1	**0.21**	**0.23**	**0**	**0.3**	-1.77	-1.77
MD	**0.12**	**0**	**0**	**0**	-0.03	0.14
Total	0.49	-0.15	-7.89	1.47	-8.68	-5.88

Table 9.4.1Two-point LOD scores for chromosome 10q.

*Shown are the two-point scores for θ=0 for the markers mapping to chromosome 10q in the JPS families. Families 1, 5, C1 and 14 are compatible with linkage to SMAD4, but were included in the two-point analysis. Distances are shown in brackets after the marker name. BMPR1A lies between D10S2327 and GATA115E01. *= families shown to harbour BMPR1A mutations. Bold-type shows families which are compatible with linkage to BMPR1A but in whom no mutation was identified.*

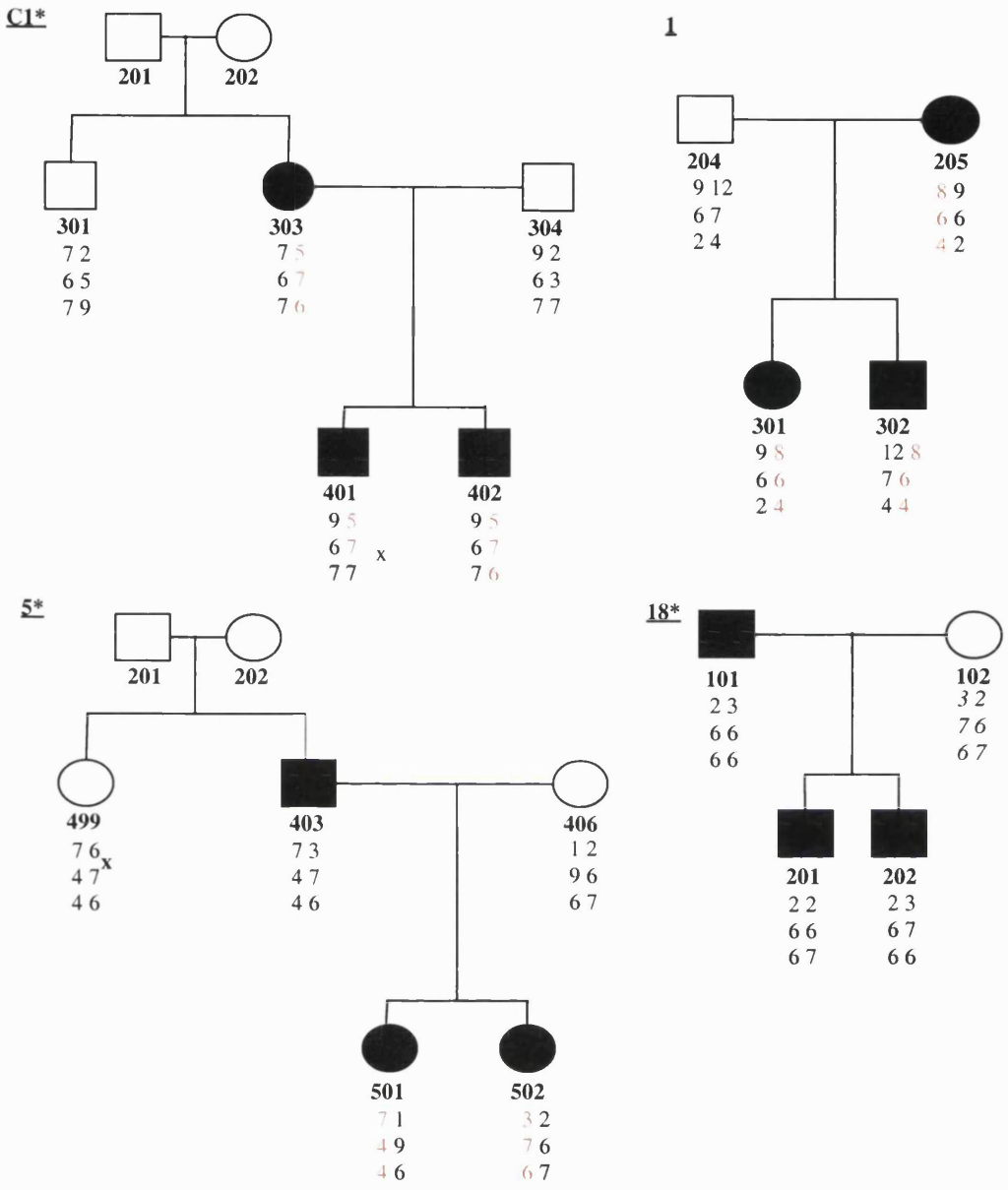
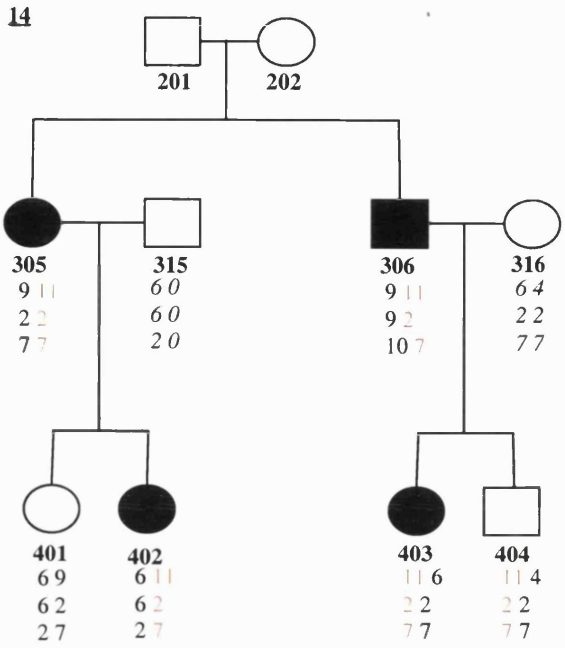


Figure 9.4.1. 10q haplotypes in juvenile polyposis syndrome families.
*Shown are the haplotypes for 3 markers mapping to the BMPR1A region (D10S537, ALK3GGAA and ALK3CA). BMPR1A maps 49Kb distal to ALK3CA. Putative 'affected' haplotypes are highlighted red Affected individuals are shown with filled symbols. Inferred alleles are shown in italics. *=haplotypes shown for D10S23327, GATA115E01 (which flank BMPR1A) and D10S677 due to poor informativity at the other markers. Families C1 and 1 were compatible with linkage to BMPR1A (as well as SMAD4) as affected individuals shared a haplotype. Family 5 was not compatible with linkage as 501 and 502 have inherited different alleles. Figure continues on next page. Family 18 ~~harbours~~ is compatible with linkage, and has a BMPR1A mutation.*



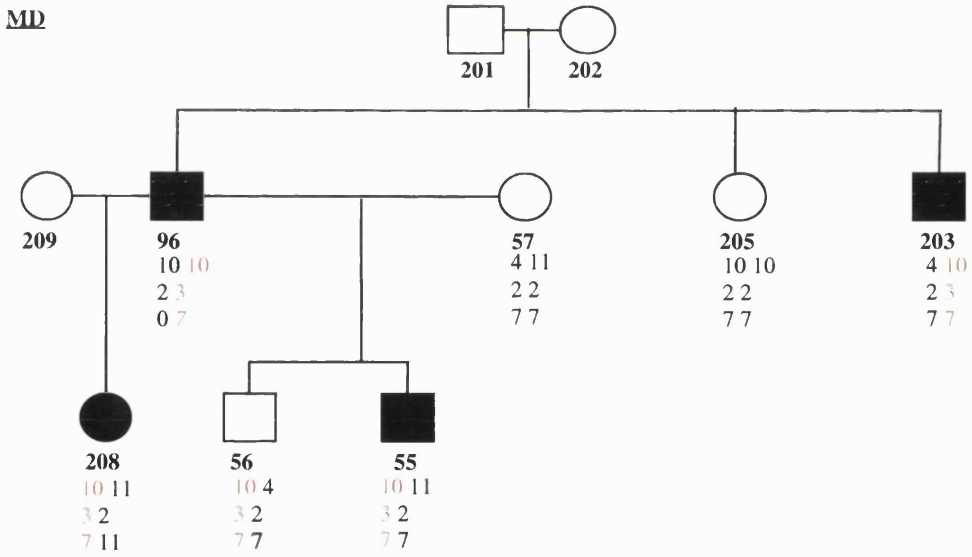


Figure 9.4.1. *Families 14 and MD were compatible with linkage to BMPR1A as all affected individuals shared a haplotype. However, individual 404 in Family 14, and individual 56 in Family MD, also shared the putative affected haplotype, and would have to be non-penetrant were an underlying BMPR1A mutations present in these families. Figure continued on next page.*

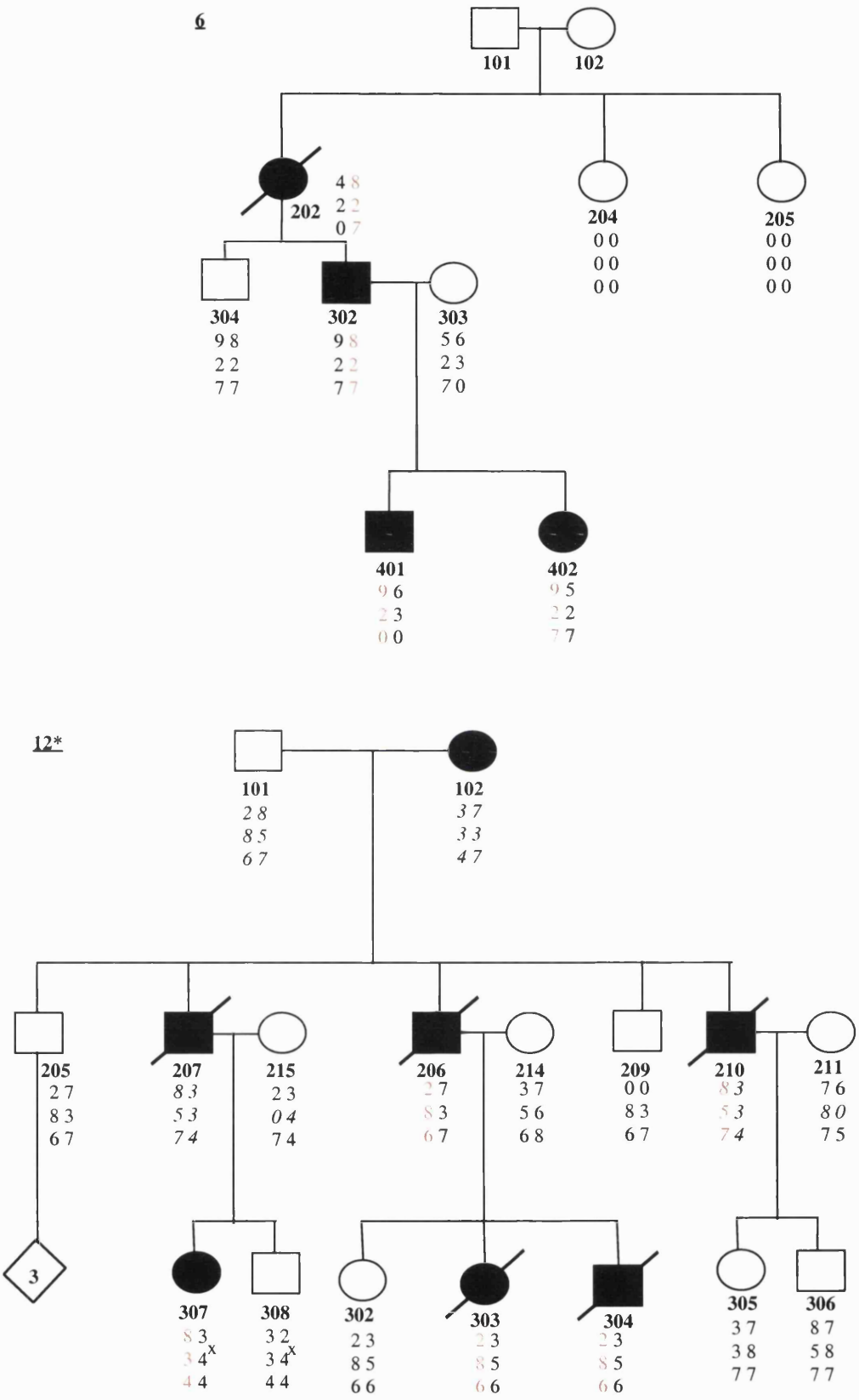


Figure 9.4.1. *Figure continued on next page.*

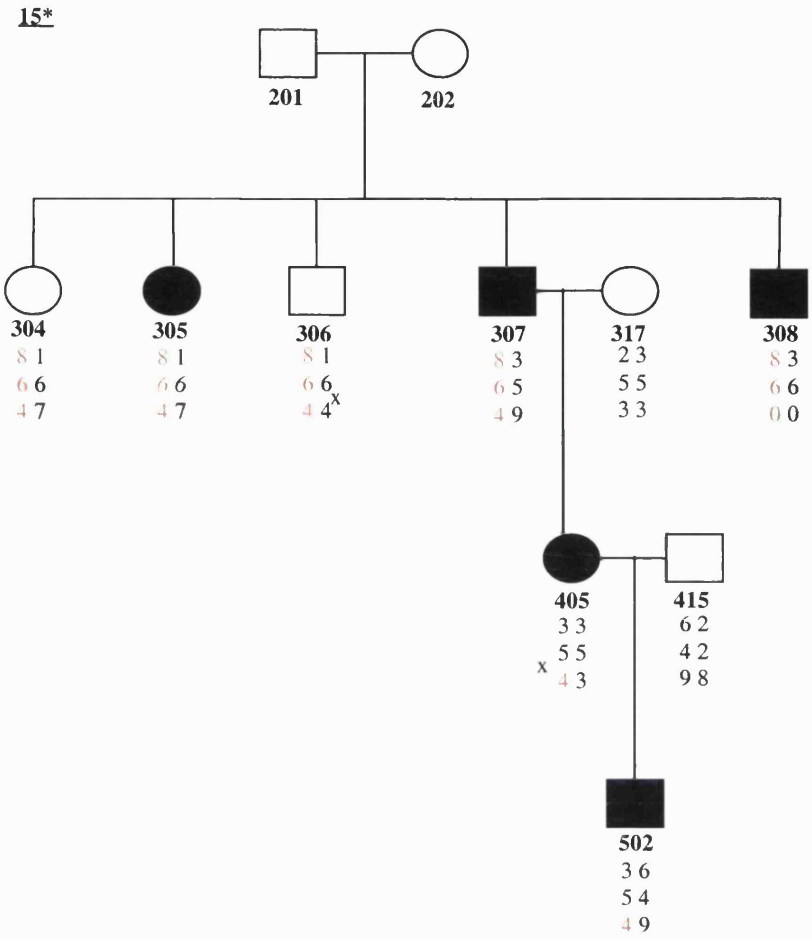


Figure 9.4.1. continued. *Family 6 demonstrates evidence against linkage as persons 401 and 402 did not inherited their affected grandmother's (202) allele at D10S573. Due to poor informativity at ALK3GGAA and ALK3CA, a crossover at BMPR1A cannot be entirely ruled out. Family 12 was not compatible with linkage to BMPR1A as affected siblings 303 and 304 did not share alleles at D10S2327 and GATA115E01 (or D10S573, not shown) with 210 and 307. Family 15 was not compatible with linkage as affected individuals 307, 405 and 502 did not share alleles at the two markers flanking BMPR1A (D10S2327 and GATA115E01) with other affected members of this family.*

Families 1, C1, MD and 14 were all compatible with linkage to *BMPR1A*. Of these Families, 1, C1 and 14 were also compatible with linkage to *SMAD4*, but no underlying germline mutation was identified in either gene. Families 5, 6, 12 and 15 were not compatible with linkage to *BMPR1A*, shown by the failure to share a haplotype at markers spanning the 10q22 region. Further methods are therefore required to confirm that mutations in *BMPR1A* do not exist in those families compatible with linkage, namely immunohistochemistry where tissue blocks are available (Family MD), the protein truncation test where RNA is available (MD), and alternative PCR-based methods for all compatible families (e.g. SSCP). In particular, immunohistochemistry can be used to establish whether there is absent BMPR1A protein in polyps from germline *BMPR1A* mutation carriers, and then can hopefully be used as a marker for the presence of a germline *BMPR1A* mutation, in much the same way as the anti-SMAD4 B8 antibody is able to reliably predict the presence of a germline *SMAD4* mutation via the absence of protein in the polyps. The disease-causative region in Families 6, 12 and 15 remains elusive, as these three families are not compatible with either *SMAD4* or *BMPR1A*. In addition, there are many sporadic JPS patients who have not had a *BMPR1A* or *SMAD4* mutation identified, although the power to identify new regions lies certainly with the familial cases, with subsequent screening of the sporadic cases.

9.5 CONCLUSIONS

Screening of the *BMPR1A* gene by direct sequencing in the JPS cohort, totalling 34 families and sporadic cases, identified 11 new mutations (32%). Unequivocal loss of heterozygosity in tumours derived from mutation carriers showed that *BMPR1A* acts as a tumour suppressor gene, with the second hit presumably leading to growth of the polyp. This was in contrast to Howe *et al*, who first described *BMPR1A* mutations in JPS, where no LOH was observed in six juvenile polyps studied (Howe *et al.*, 2001). Presumably the microdissections of Howe *et al* contained too much contaminating normal tissue that would confound the detection of LOH.

Linkage analysis and haplotype construction using markers that map closely to *BMPR1A* indicated that four families were compatible with linkage to this region (1, 14, MD and C1). No mutation has been identified as yet in these patients, despite repetition of the direct sequencing. Four families were not compatible with linkage to 10q22 (5, 6, 12 and 15). Three of these families (6, 12 and 15) are also not compatible with *SMAD4* linkage, indicating that there exists at least one more elusive JPS gene.

Importantly, BMPR1A is a receptor in the BMP pathway, acting as the equivalent of the TGFβR1 in the TGFβ signalling pathway (see Chapter Six). SMAD4 acts as the common mediator SMAD in the BMP pathway (as well as

the TGFβ- pathway), and thus there is convergence of two genes in the same signalling pathway whose inactivation leads to juvenile polyps and their associated cancers. SMADs 1 and 5, which act as the receptor-regulated SMADs in the BMP pathway, have already been screened for mutations, but none found (discussed in Chapter Six), although as only the cDNA sequence was available at the time, *SMAD1* has been screened in a subset of patients only. Further investigation into other genes converging on the BMP-signalling pathway may be good candidates. This would follow the planned high-density linkage analysis planned to ensure there is some evidence of linkage before screening vast numbers of genes. Disruption of the BMP-pathway by mutations in both *SMAD4* and *BMPR1A* presumably act through similar downstream targets to confer a similar phenotype. The elucidation of these downstream targets should give insight into the processes of tumour development, possibly not just in JPS but in sporadic colorectal cancer too.

CHAPTER TEN

GENERAL DISCUSSION AND CONCLUSIONS

GENERAL DISCUSSION AND CONCLUSIONS

Individuals with juvenile polyposis syndrome have polyps in the gastrointestinal tract that carry a high chance of developing malignancy, conferring considerable morbidity and premature death. Understanding the aetiology and the genetics of JPS is therefore important for potential gene carriers in affected families, for better understanding of disease development in those with JPS and for sporadic colorectal cancers. The first susceptibility locus for juvenile polyposis syndrome was shown to be the *SMAD4/DPC4* gene on chromosome 18q21.1 in 1998 (Howe *et al.*, 1998b). Germline mutations in *SMAD4* were subsequently reported in about a quarter of JPS families and sporadic cases (Friedl *et al.*, 1999; Houlston *et al.*, 1998; Kim *et al.*, 2000; Roth *et al.*, 1999). Work undertaken for this project has given insight into the development of polyps, both genetically and morphologically, in germline *SMAD4* mutation carriers and has indicated that *SMAD4* probably plays a more common role in sporadic colorectal tumorigenesis than previously thought. Together these data may give a better understanding of the evolution of colorectal cancer, occurring both in JPS and sporadically.

SMAD4 belongs to a family of 8 closely related SMAD genes, each of which is involved in the TGFβ-superfamily of signalling pathways (Massague and Chen, 2000). The *SMAD4* gene was previously known as a target for deletion and mutation

in pancreatic cancer (Hahn *et al.*, 1996c), where it was considered to be acting as a tumour suppressor gene. In addition, loss of the chromosomal band (18q21.1) containing both *SMAD4* and another putative tumour suppressor, *DCC*, was well documented in sporadic colorectal cancer (Thiagalingam *et al.*, 1996). Inactivation of *SMAD4* was generally considered to be a late event in colorectal neoplastic progression, and thought to be associated with tumour metastasis (Maitra *et al.*, 2000; Miyaki *et al.*, 1999). Work undertaken for this project has shown that *SMAD4* is frequently inactivated by mutation in microsatellite stable colorectal cancer cell lines, but was not found to be mutated in any colorectal cancer cell lines displaying microsatellite instability. In addition, loss of *SMAD4* was shown probably to occur earlier than previously suggested, probably because previous reports failed to distinguish between microsatellite stable and microsatellite unstable tumours. Loss of *SMAD4* most probably occurs before chromosomal instability, but after divergence of the microsatellite unstable tumours. The observed loss at 18q21.1 was shown to target *SMAD4* in a high proportion, but not all, of the colorectal cancer cell lines, indicating that there remains at least one other important gene in this region (which may or may not be *DCC*). The targets of this loss may soon be possible to elucidate with the advent of the draft human sequence.

Kinzler and Vogelstein proposed that in juvenile polyposis, loss of *SMAD4* did not directly alter epithelial cell growth as it does in sporadic colorectal cancer, but acted upon the stroma which in turn induced carcinoma formation through an altered

terrain for epithelial cell growth – the 'landscaper' hypothesis (Kinzler and Vogelstein, 1998). This study has shown that *SMAD4* does indeed act as a tumour suppressor gene, and not a 'landscaper', in juvenile polyposis syndrome. Loss of heterozygosity analysis also indicated that juvenile polyps are clonal lesions, and thus are true neoplasms. This concurs with other work in Peutz-Jeghers syndrome, tuberous sclerosis and Cowden syndrome where the pre-malignant hamartomas showed loss of heterozygosity at the site of the respective germline mutations (Marsh *et al.*, 1998; Sepp *et al.*, 1996; Wang *et al.*, 1999). Loss of *SMAD4* in juvenile polyposis is targeted to the epithelial cells, therefore negating the need for complicated explanations of how stromal cells can induce epithelial malignancy. In addition, stromal fibroblasts and peri-cryptal myofibroblasts in the juvenile polyps also showed deletion of *SMAD4*, indicating that the polyps probably arise from a stem cell with a greater degree of plasticity than generally considered. Although this is less in keeping with histological dogma, support to this data is given by the demonstration of the clonal origin of mesenchymal and epithelial components in malignant mixed Müllerian tumours (Abeln *et al.*, 1997), and the clonality of TSC hamartomas despite the mixture of mesencyhmal and epithelial elements (Green *et al.*, 1996). In addition, stem cell plasticity has recently been demonstrated with the neo-differentiation of bone marrow into liver cells .

Ensuring that *SMAD4* was not actually responsible for a higher proportion of JPS cases was critical for future experiments, as inclusion would run the risk of false

negatives. Immunohistochemistry using an antibody directed against SMAD4 had been shown to be sensitive and reliable for detecting loss of the SMAD4 protein in pancreatic cancers (Wilentz *et al.*, 2000) and proved to be highly indicative of a germline *SMAD4* mutation in this study when used on paraffin embedded tissue. The epitope for the *SMAD4* antibody lay within exon 5 (of 11) of the mature protein. This indicated (and was further supported by work with colorectal cancer cell lines) that regardless of the type and position of the genetic defect, mutant SMAD4 proteins were unstable and degraded *in vivo*. Degradation of both mutant SMAD4 and also SMAD2 has been reported previously (Xu and Attisano, 2000). This allowed the detection of loss of SMAD4 protein even when a mutation lay C-terminal to the antibody epitope, and as such was found to be a reliable predictor of a germline mutation.

Recently, germline mutations in the *BMPR1A* gene on chromosome 10q22 have been identified in a further subset of JPS cases (Howe *et al.*, 2001) and this has posed further questions for the pathogenesis of JPS. Overall *BMPR1A* mutations were found in 32% of our JPS cohort, indicating further genetic heterogeneity. One important issue to address is which cells are targeted for deletion in the polyps of germline *BMPR1A* mutation carriers. LOH analysis by Howe *et al* failed to detect any loss of the wild-type *BMPR1A* gene in polyps from germline mutation carriers (Howe *et al.*, 2001). LOH analysis performed for this study, however, showed unequivocal loss of the wild-type allele in these *BMPR1A*-polyps, suggesting that

these, like *SMAD4*-induced polyps, are clonal lesions. Jacoby *et al* previously reported loss of 10q22 in juvenile polyps, but fluorescent *in situ* hybridisation suggested that the inflammatory lymphocytes were the cells targeted for 10q22 deletion. (Jacoby *et al.*, 1997). Whilst this may lend support to the landscaper hypothesis, it is most likely due to experimental error and unlikely to be borne out. Loss of *BMPR1A* in a susceptible cell in the GI tract would once again be the most likely initiating event in the development of the GI polyps and malignancies in *BMPR1A* mutation carriers. An antibody directed against the C-terminal of BMPR1A is available, though it is not certain to work on paraffin embedded tissue. If successful however, immunohistochemistry should prove a reliable and quick indicator of a germline *BMPR1A* or a *SMAD4* mutation.

Whereas the majority of Peutz-Jeghers families are compatible with linkage to 19p13.3 (Olschwang *et al.*, 1998a), and there is no evidence of genetic heterogeneity in another hamartoma syndrome, Cowden disease (Nelen *et al.*, 1996), juvenile polyposis is more heterogeneous, with *SMAD4* and *BMPR1A* combined only accounting for ~50% of JPS cases thus far. Whilst subtle polyp morphological differences are distinct to *SMAD4* mutation carriers, discussed as part of this project, whether *BMPR1A* mutations also confer phenotypic differences remains to be determined. Tuberous sclerosis, another syndrome in which hamartomas develop, has been shown to be due to mutations at two distinct loci, 9q34 and 16p13.3. The phenotypic differences conferred by mutations at each locus appear to be limited to

the increased incidence of severe renal cystic disease in the 16p13.3 group when there is a contiguous deletion of the *APKD1* (polycystic kidney disease) gene (Hodgson and Maher, 1999). Importantly, the gene products for the two TSC loci are the closely associated proteins hamartin and tuberin, with loss of either protein leading to a similar phenotype. Disruption of a single pathway may likewise prove to be the case in juvenile polyposis syndrome. Loss of *SMAD4* was always believed to lead to abrogation of TGFβ-signalling as the SMAD4 protein plays a pivotal role in transducing signals from the membrane to the nucleus (Massague, 1998). However, SMAD4 also acts as a common mediator in the activin and BMP pathways (Itoh *et al.*, 2000). Germline inactivation of *BMPR1A*, upstream of *SMAD4* in the BMP signalling pathway also confers individuals to develop juvenile polyps and then associated malignancies. It may yet transpire that other members in the BMP-pathway are good candidates for JPS cases who are not explained by mutations in either *BMPR1A* or *SMAD4*, given the pathways role in development and remodelling. The *SMAD5* gene, a receptor-regulated SMAD in the BMP-pathway, was screened for germline mutations as part of this project, and no mutations identified. Likewise *SMAD1* was screened in a subset of individuals (those with cDNA available) and no mutations identified. These two genes were probably two of the best candidates in the BMP-pathway that could conceivably lead to a similar phenotype if disrupted, but it seems that new JPS loci are situated elsewhere. SMADs 2 and 3 (receptor-regulated SMADs in the TGFβ-signalling pathway) that were also good candidate JPS genes, were screened for germline JPS mutations and none identified. The TGFβ-superfamily is complicated and the pathways are

entwined so it may not be straightforward to pick out the future JPS loci. It is also possible that other JPS loci do not form part of the same pathways as *SMAD4* and *BMPR1A*. For example, mutations in *PTEN* cause Cowden disease and Bannayan-Zonana syndrome in which juvenile polyps also develop (Nelen *et al.*, 1997), but has no known association with the BMP-signalling pathway.

Genome wide linkage analysis in juvenile polyposis families did not reveal any new area compatible with linkage in everyone of the families who did not have mutations in either *SMAD4* or *BMPR1A*. A subset of families were not compatible with linkage to either locus, so there undoubtedly remains at least one further JPS gene (and probably two) to be identified. The identification of *SMAD4* as a JPS susceptibility locus by Howe *et al* was possible due to an extremely large family with 29 affected individuals (Howe *et al.*, 1998a), and the *BMPR1A* linkage analysis included two families with 11 affected individuals each (Howe *et al.*, 2001). The two largest families in our cohort each only had 5 affected members, with an obvious reduction in the power to detect, and then confirm, linkage, particularly in a genetically heterogeneous disease. Extra samples (both constitutional and from tumours) from these families should aid the confirmation or refutation of those regions compatible with linkage. In addition, expression analysis of fresh frozen polyp material using chip technology may help identify genes that are lost in these tumours, and is not reliant on candidacy or linkage analysis of a specific gene or region.

As loss of *BMPR1A* induces juvenile polyp and adenoma formation, and subsequent progression to colorectal and other GI cancer in JPS patients, it is readily conceivable that this gene will have a role in sporadic colorectal cancer, as does *SMAD4*. TGFβ-signalling has been shown to be both SMAD4-dependent, and also SMAD4-independent (Dai *et al.*, 1999). It is probable that loss of *SMAD4* in colorectal tumorigenesis also affects the activin and BMP-pathways. Investigation of BMPR1A expression in the colorectal cancer cell lines using Western blotting may provide new insights into genes important in colorectal cancer progression, and may distinguish which signalling pathway loss of *SMAD4* actually affects.

The susceptibility locus for Hereditary Mixed Polyposis Syndrome has been mapped to 15q13-14 as a part of this project. Furthermore, haplotype analysis has shown that three distinct Ashkenazi families with colorectal tumours shared a common ancestor. Although atypical juvenile polyps were a feature of this HMPS, two of the larger JPS families were not compatible with linkage to this area. Once the gene is identified, however, it will be screened in familial and sporadic JPS cases. The role of the HMPS gene again will be interesting to investigate in sporadic colorectal cancer.

In conclusion, juvenile polyposis syndrome has been revealed to be more heterogeneous than previously considered. Certainly more JPS loci remain to be identified in those patients without *SMAD4* or *BMPR1A* mutations but their

identification may rely on ascertainment of larger JPS families and/or candidate gene screening combined with LOH analysis. The availability of a large HMPS family allowed the identification of the HMPS locus on chromosome 15q13-14. Analysis of both *SMAD4* and *BMPR1A* in juvenile polyps from patients who do carry respective mutations has revealed that far from being stromal lesions, juvenile polyps are true neoplasms and are clonal. The risk of cancer in such polyps is therefore understandably very real, even though for a long time they were considered to be without malignant potential. Identifying the genetic changes observed in juvenile polyposis has and will give insight into the role of such tumour suppressor genes in sporadic colorectal cancer, and will hopefully increase our understanding of its aetiology.

PAPERS PUBLISHED AS A RESULT OF THE PROJECT

*=contributed equally

Bevan,S.*, **Woodford-Richens, K. L.***, Rozen, P., Eng, C., Young, J., Dunlop, M., Neale, K., Phillips, R., Markie, D., Rodriguez-Bigas, M., Leggett, B., Sheridan, E., Hodgson, S., Iwama, T., Eccles, D., Bodmer, W., Houlston, R., and Tomlinson, I., 1999, Screening SMAD1, SMAD2, SMAD3, and SMAD5 for germline mutations in juvenile polyposis syndrome, *Gut* **45**(3):406-408.

Woodford-Richens, K. L.*, Bevan, S.*, Churchman, M., Dowling, B., Jones, D., Norbury, C. G., Hodgson, S. V., Desai, D., Neale, K., Phillips, R. K., Young, J., Leggett, B., Dunlop, M., Rozen, P., Eng, C., Markie, D., Rodriguez-Bigas, M. A., Sheridan, E., Iwama, T., Eccles, D., Smith, G. T., Kim, J. C., Kim, K. M., Sampson, J. R., Evans, G., Tejpar, S., Bodmer, W. F., Tomlinson, I. P. M., and Houlston, R. S., 2000a, Analysis of genetic and phenotypic heterogeneity in juvenile polyposis, *Gut* **46**(5):656-60.

Woodford-Richens, K. L., Halford, S., Rowan, A., Bevan, S., Aaltonen, L. A., Wasan, H., Bicknell, D., Bodmer, W. F., Houlston, R. S., and Tomlinson, I. P. M., 2001a, CDX2 mutations do not account for juvenile polyposis or Peutz-Jeghers syndrome and occur infrequently in sporadic colorectal cancers, *Br J Cancer* **84**(10):1314-6.

Woodford-Richens, K. L., Williamson, J., Bevan, S., Young, J., Leggett, B., Frayling, I., Thway, Y., Hodgson, S., Kim, J. C., Iwama, T., Novelli, M., Sheer, D., Poulsom, R., Wright, N., Houlston, R., and Tomlinson, I. P. M., 2000b, Allelic loss at SMAD4 in polyps from juvenile polyposis patients and use of fluorescence in situ hybridization to demonstrate clonal origin of the epithelium, *Cancer Res* **60**(9):2477-82.

Woodford-Richens, K. L., Rowan, A., Poulsom, R., Bevan, S., Salovaara, R., Aaltonen, L. A., Houlston, R. S., Wright, N. A., and Tomlinson, I. P. M., 2001b, Comprehensive analysis of *SMAD4* mutations and protein expression in juvenile polyposis: evidence for a distinct genetic pathway and polyp morphology in *SMAD4* mutation carriers, *American Journal of Pathology* **in press.**

Woodford-Richens, K. L., Rowan, A. J., Gorman, P., Halford, S., Bicknell, D. C., Wasan, H. S., Roylance, R. R., Bodmer, W. F., and Tomlinson, I. P. M., 2001c, SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway, *Proc Natl Acad Sci U S A* **31**:31.

Zhou, X. P.*, **Woodford-Richens, K. L.***, Lehtonen, R., Kurose, K., Aldred, M., Hampel, H., Launonen, V., Virta, S., Pilarski, R., Salovaara, R., Bodmer, W. F., Conrad,

B. A., Dunlop, M., Hodgson, S. V., Iwama, T., Jarvinen, H., Kellokumpu, I., Kim, J. C., Leggett, B., Markie, D., Mecklin, J. P., Neale, K., Phillips, R., Piris, J., Rozen, P., Houlston, R. S., Aaltonen, L. A., Tomlinson, I. P.M., and Eng. C., 2001, Germline mutations in BMPR1A/ALK3 cause a subset of cases of juvenile polyposis syndrome and of Cowden and Bannayan-Riley-Ruvalcaba syndromes, *Am J Hum Genet* **69**(4):704-11

Jaeger, E. E. M*., **Woodford-Richens, K. L.*.,** Lockett, M. A. R., Sawyer, E. J., Heinimann, K., Rozen, Murday, V. A., Whitelaw, S. C., Ginsberg, A., Lynch, H. T., Southey, M. C., Eng, C., Bodmer, W. F., Talbot, I. C., Hodgson, S. V., Thomas, H. J. W., and Tomlinson, I. P. M., 2001, An Ashkenazi Founder Mutation At The HMPS/CRAC1 Locus Causes Hereditary Mixed Polyposis Syndrome., *Am J Hum Gen* (submitted).

REFERENCES

Abdel-Rahman, W. M., Katsura, K., Rens, W., Gorman, P. A., Sheer, D., Bicknell, D., Bodmer, W. F., Arends, M. J., Wyllie, A. H., and Edwards, P. A., 2001, Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement, *Proc Natl Acad Sci U S A* **98**(5):2538-2543.

Abeln, E. C., Smit, V. T., Wessels J. W., de Leeuw W. J., Cornelisse C. J., and Fleuren G. J., 1997, Molecular genetic evidence for the conversion hypothesis fo the origin of malignant mixed Mullerian tumours, *Journal of Pathology* **183**:424-431.

Arai, T., Akiyama, Y., Okabe, S., Ando, M., Endo, M., and Yuasa, Y., 1998, Genomic structure of the human Smad3 gene and its infrequent alterations in colorectal cancers, *Cancer Lett* **122**(1-2):157-63.

Arch, E. M., Goodman, B. K., Van Wesep, R. A., Liaw, D., Clarke, K., Parsons, R., McKusick, V. A., and Geraghty, M. T., 1997, Deletion of PTEN in a patient with Bannayan-Riley-Ruvalcaba syndrome suggests allelism with Cowden disease, *Am. J. Med. Genet* **71**(4):489-93 issn: 0148-7299.

Avizienyte, E., Loukola, A., Roth, S., Hemminki, A., Tarkkanen, M., Salovaara, R., Arola, J., Butzow, R., Husgafvel-Pursiainen, K., Kokkola, A., Jarvinen, H., and Aaltonen, L. A., 1999, LKB1 somatic mutations in sporadic tumors, *Am J Pathol* **154**(3):677-81.

Barbera, V. M., Martin, M., Marinoso, L., Munne, A., Carrato, A., Real, F. X., and Fabre, M., 2000, The 18q21 region in colorectal and pancreatic cancer: independent loss of DCC and DPC4 expression, *Biochim Biophys Acta* **1502**(2):283-296.

Beck, F., Chawengsaksophak, K., Waring, P., Playford, R. J., and Furness, J. B., 1999, Reprogramming of intestinal differentiation and intercalary regeneration in Cdx2 mutant mice, *Proc Natl Acad Sci U S A* **96**(13):7318-23.

Bishop, J. M., 1991, Molecular themes in oncogenesis, *Cell* **64**:235-248.

Bos, J. L., Fearon, E. R., Hamilton, S. R., Verlaan-de Vries, M., van Boom, J. H., van der Eb, A. J., and Vogelstein, B., 1987, Prevalence of ras gene mutations in human colorectal cancers, *Nature* **327**(6120):293-7.

Brentnall, T. A., Crispin, D. A., Rabinovitch, P. S., Haggitt, R. C., Rubin, C. E., Stevens, A. C., and Burmer, G. C., 1994, Mutations in the p53 gene: An early marker of neoplastic progression in ulcerative colitis, *Gastroenterology* **107**(2):369-78.

Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., and et, a., 1994, Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer, *Nature* **368**(6468):258-61.

Caspari, R., Olschwang, S., Friedl, W., Mandl, M., Boisson, C., Boker, T., Augustin, A., Kadmon, M., Moslein, G., Thomas, G., and et al., 1995, Familial adenomatous polyposis: desmoid tumours and lack of ophthalmic lesions (CHRPE) associated with APC mutations beyond codon 1444, *Hum Mol Genet* **4**(3):337-40.

Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C., and White, R. L., 1983, Expression of recessive alleles by chromosomal mechanisms in retinoblastoma, *Nature* **305**(5937):779-84.

Chaubert, P., Benhattar, J., Saraga, E., and Costa, J., 1994, K-ras mutations and p53 alterations in neoplastic and nonneoplastic lesions associated with longstanding ulcerative colitis, *Am J Pathol* **144**(4):767-75.

Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F., and Beck, F., 1997, Homeosis and intestinal tumours in Cdx2 mutant mice, *Nature* **386**(6620):84-7.

Chellappan, S., Kraus, V. B., Kroger, B., Munger, K., Howley, P. M., Phelps, W. C., and Nevins, J. R., 1992, Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product, *Proc Natl Acad Sci U S A* **89**(10):4549-53.

Chen, Y. G., and Massague, J., 1999, Smad1 recognition and activation by the ALK1 group of transforming growth factor-beta family receptors, *J Biol Chem* **274**(6):3672-7.

Cho, K. R., Oliner, J. D., Simons, J. W., Hedrick, L., Fearon, E. R., Preisinger, A. C., Hedge, P., Silverman, G. A., and Vogelstein, B., 1994, The DCC gene: structural analysis and mutations in colorectal carcinomas, *Genomics* **19**(3):525-31.

Cottingham, R. W., Jr., Idury, R. M., and Schaffer, A. A., 1993, Faster sequential genetic linkage computations, *Am J Hum Genet* **53**(1):252-63.

Counts, J., and Goodman, J., 1995, Alterations in DNA methylation may play a variety of roles in carcinogenesis, *Cell* **83:**13-15.

da Costa, L., He, T., Yu, J., Sparks, A., Morin, P., Polyak, K., Laken, S., Vogelstein, B., and Kinzler, K., 1999, CDX2 is mutated in a colorectal cancer with normal APC/beta-catenin signaling., *Oncogene* **18**(35):5010-5014.

Dai, J. L., Schutte, M., Bansal, R. K., Wilentz, R. E., Sugar, A. Y., and Kern, S. E., 1999, Transforming growth factor-beta responsiveness in DPC4/SMAD4-null cancer cells, *Mol Carcinog* **26**(1):37-43.

De Angelis, P. M., Clausen, O. P., Schjolberg, A., and Stokke, T., 1999, Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes, *Br J Cancer* **80**(3-4):526-35.

de Caestecker, M. P., Piek, E., and Roberts, A. B., 2000, Role of transforming growth factor-beta signaling in cancer, *J Natl Cancer Inst* **92**(17):1388-402.

Drummond, F., Putt, W., Fox, M., and Edwards, Y. H., 1997, Cloning and chromosome assignment of the human CDX2 gene, *Ann Hum Genet* **61**(Pt 5):393-400.

Eng, C., and Ji, H.-L., 1998, Molecular classification of the inherited hamartoma syndromes: clearing the muddied waters., *Amer J Hum Genet.* **62:**1020-1022.

Eppert, K., Scherer, S. W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L. C., Bapat, B., Gallinger, S., Andrulis, I. L., Thomsen, G. H., Wrana, J. L., and Attisano, L., 1996, MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma, *Cell* **86**(4):543-52.

Esteller, M., 2000, Epigenetic lesions causing genetic lesions in human cancer. promoter hypermethylation of DNA repair genes, *Eur J Cancer* **36**(18):2294-300.

Fargnoli, M. C., Orlow, S. J., Semel-Concepcion, J., and Bolognia, J. L., 1996, Clinicopathologic findings in the Bannayan-Riley-Ruvalcaba syndrome, *Arch Dermatol* **132**(10):1214-8.

Fazeli, A., Dickinson, S. L., Hermiston, M. L., Tighe, R. V., Steen, R. G., Small, C. G., Stoeckli, E. T., Keino-Masu, K., Masu, M., Rayburn, H., Simons, J., Bronson, R. T., Gordon, J. I., Tessier-Lavigne, M., and Weinberg, R. A., 1997, Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene, *Nature* **386**(6627):796-804.

Fearon, E. R., and Vogelstein, B., 1990, A genetic model for colorectal tumorigenesis, *Cell* **61**(5):759-67.

Feng, X. H., Lin, X., and Derynck, R., 2000, Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15(Ink4B) transcription in response to TGF-beta, *Embo J* **19**(19):5178-93.

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R., 1993, The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer, *Cell* **75**(5):1027-38.

Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., van Es, J. H., Breukel, C., Wiegant, J., Giles, R. H., and Clevers, H., 2001, Mutations in the APC tumour suppressor gene cause chromosomal instability, *Nat Cell Biol* **3**(4):433-8.

Forrester, K., Almoguera, C., Han, K., Grizzle, W. E., and Perucho, M., 1987, Detection of high incidence of K-ras oncogenes during human colon tumorigenesis, *Nature* **327**(6120):298-303.

Frayling, I. M., Beck, N. E., Ilyas, M., Dove-Edwin, I., Goodman, P., Pack, K., Bell, J. A., Williams, C. B., Hodgson, S. V., Thomas, H. J., Talbot, I. C., Bodmer, W. F., and Tomlinson, I. P., 1998, The APC variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history, *Proc Natl Acad Sci U S A* **95**(18):10722-7.

Friedl, W., Kruse, R., Uhlhaas, S., Stolte, M., Schartmann, B., Keller, K. M., Jungck, M., Stern, M., Loff, S., Back, W., Propping, P., and Jenne, D. E., 1999, Frequent 4-bp deletion in exon 9 of the SMAD4/MADH4 gene in familial juvenile polyposis patients, *Genes Chromosomes Cancer* **25**(4):403-6.

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W., and Ehrlich, M., 1983, The 5-methylcytosine content of DNA from human tumors, *Nucleic Acids Res* **11**(19):6883-94.

Georgiades, I. B., Curtis, L. J., Morris, R. M., Bird, C. C., and Wyllie, A. H., 1999, Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence for chromosomal or microsatellite instability, *Oncogene* **18**(56):7933-40.

Goi, T., Yamaguchi, A., Nakagawara, G., Urano, T., Shiku, H., and Furukawa, K., 1998, Reduced expression of deleted colorectal carcinoma (DCC) protein in established colon cancers, *Br J Cancer* **77**(3):466-71.

Gold, L. I., 1999, The role for transforming growth factorb (TGF-b) in human cancer., *Crit. Rev. Oncog.* **10**:303–360.

Gotley, D. C., Reeder, J. A., Fawcett, J., Walsh, M. D., Bates, P., Simmons, D. L., and Antalis, T. M., 1996, The deleted in colon cancer (DCC) gene is consistently expressed in colorectal cancers and metastases, *Oncogene* **13**(4):787-95.

Green, A. J., Sepp, T., and Yates, J. R., 1996, Clonality of tuberous sclerosis hamartomas shown by non-random X-chromosome inactivation, *Hum Genet* **97**:240-243.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., and et, a. l., 1991, Identification and characterization of the familial adenomatous polyposis coli gene, *Cell* **66**(3):589-600.

Gruber, S. B., Entius, M. M., Petersen, G. M., Laken, S. J., Longo, P. A., Boyer, R., Levin, A. M., Mujumdar, U. J., Trent, J. M., Kinzler, K. W., Vogelstein, B., Hamilton, S. R., Polymeropoulos, M. H., Offerhaus, G. J., and Giardiello, F. M.,

1998, Pathogenesis of adenocarcinoma in Peutz-Jeghers syndrome, *Cancer Res* **58**(23):5267-70.

Hahn, H., Wicking, C., Zaphiropoulos, P. G., Gailani, M. R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A. B., Gillies, S., Negus, K., Smyth, I., Pressman, C., Leffell, D. J., Gerrard, B., Goldstein, A. M., Dean, M., Toftgard, R., Wainwright, B., and Bale, A. E., 1996a, Mutations of the human homologue of Drosophila patched in the naevoid basal cell carcinoma syndrome., *Cell* **85**:841-851.

Hahn, S. A., Hoque, A. T., Moskaluk, C. A., da Costa, L. T., Schutte, M., Rozenblum, E., Seymour, A. B., Weinstein, C. L., Yeo, C. J., Hruban, R. H., and Kern, S. E., 1996b, Homozygous deletion map at 18q21.1 in pancreatic cancer, *Cancer Res* **56**:490-494.

Hahn, S. A., Schutte, M., Hoque, A. T., Moskaluk, C. A., da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., and Kern, S. E., 1996c, DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1, *Science* **271**(5247):350-3.

Hamilton, S. R., 1992, Molecular genetics of colorectal carcinoma, *Cancer* **70**:1216-1221.

Harris, H., Miller, O. J., Klein, G., Worst, P., and Tachibam, T., 1969, Suppression of malignancy by cell fusion, *Nature* **223**:363-368.

Hata, A., Lo, R. S., Wotton, D., Lagna, G., and Massague, J., 1997, Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4, *Nature* **388**(6637):82-7.

Hata, A., Shi, Y., and Massague, J., 1998, TGF-beta signaling and cancer: structural and functional consequences of mutations in Smads, *Mol Med Today* **4**(6):257-62.

Heldin, C., Miyazono, K., and ten Dijke, P., 1997, TGF-b signalling from cell membrane to nucleus through SMAD proteins, *Nature* **390:**465-471.

Hemminki, A., Markie, D., Tomlinson, I. P. M., Avizienyte, E., Roth, S., Loukola, A., Bignell, G., Warren, W., Järvinen, H., Aminoff, M., Höglund, P., Pelin, K., Ridanpää, M., Salovaara, R., Olschwang, S., Bodmer, W. F., Olsen, A., Stratton, M. R., de la Chapelle, A., and Aaltonen, L. A., 1998, A serine/threonine kinase gene defective in Peutz-Jeghers syndrome., *Nature* **391:**184-187.

Hemminki, A., Tomlinson, I., Markie, D., Jarvinen, H., Sistonen, P., Bjorkqvist, A. M., Knuutila, S., Salovaara, R., Bodmer, W., Shibata, D., de la Chapelle, A., and Aaltonen, L. A., 1997, Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis *Nat-Genet* **15**(1):87-90.

Higgins, D. G., and Sharp, P. M., 1989, Fast and sensitive multiple sequence alignments on a microcomputer, *Comput Appl Biosci* **5:**151-153.

Hodgson, S. V., and Maher, E. R., 1999, A practical guide to human cancer genetics.

Hoque, A. T., Hahn, S. A., Schutte, M., and Kern, S. E., 1997, DPC4 gene mutation in colitis associated neoplasia, *Gut* **40**(1):120-2.

Houlston, R., Bevan, S., Williams, A., Young, J., Dunlop, M., Rozen, P., Eng, C., Markie, D., Woodford-Richens, K., Rodriguez-Bigas, M., Leggett, B., Neale, K., Phillips, R., Sheridan, E., Hodgson, S., Iwama, T., Eccles, D., Bodmer, W., and Tomlinson, I., 1998, Mutations in DPC4 (SMAD4) cause juvenile polyposis syndrome, but only account for a minority of cases, *Hum Mol Genet* **7:**1907-12.

Howe, J. R., Bair, J. L., Sayed, M. G., Anderson, M. E., Mitros, F. A., Petersen, G. M., Velculescu, V. E., Traverso, G., and Vogelstein, B., 2001, Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis, *Nat Genet* **28**(2)**:**184-7.

Howe, J. R., Ringold, J. C., Summers, R. W., Mitros, F. A., Nishimura, D. Y., and Stone, E. M., 1998a, A gene for familial juvenile polyposis maps to chromosome 18q21.1., *Amer J Hum Genet* **62:**1129-1136.

Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Jarvinen, H., Sistonen, P., Tomlinson, I. P. M., Houlston, R. S., Bevan, S., Mitros, F. A., Stone, E. M., and Aaltonen, L. A., 1998b, Mutations in the SMAD4/DPC4 gene in juvenile polyposis., *Science* **280:**1086-1088.

Ichii, S., Takeda, S., Horii, A., Nakatsuru, S., Miyoshi, Y., Emi, M., Fujiwara, Y., Koyama, K., Furuyama, J., Utsunomiya, J., ~~and~~ et al.~~, e.~~, 1993, Detailed analysis of genetic alterations in colorectal tumors from patients with and without familial adenomatous polyposis (FAP), *Oncogene* **8**(9)**:**2399-405.

Ilyas, M., Efstathiou, J. A., Straub, J., Kim, H. C., and Bodmer, W. F., 1999, Transforming growth factor beta stimulation of colorectal cancer cell lines: type II

receptor bypass and changes in adhesion molecule expression, *Proc Natl Acad Sci U S A* **96**(6):3087-91.

Imamura, T., Takase, M., Nishishara, A., Oeda, E., Hanai, J., Kawabata, M., and Miyazono, K., 1997, Smad6 inhibits signalling by the TGF-beta superfamily., *Nature* **389**:622-626.

Itoh, S., Itoh, F., Goumans, M. J., and Ten Dijke, P., 2000, Signaling of transforming growth factor-beta family members through smad proteins, *Eur J Biochem* **267**(24):6954-67.

Itzkowitz, S. H., 1997, Inflammatory bowel disease and cancer., *Gastroenterol Clin North Am* **26**:129-139.

Jacoby, R. F., Schlack, S., Cole, C. E., Skarbek, M., Harris, C., and Meisner, L. F., 1997, A juvenile polyposis tumor suppressor locus at 10q22 is deleted from nonepithelial cells in the lamina propria, *Gastroenterology* **112**(4):1398-403.

Jarvinen, H., and Franssila, K. O., 1984, Familial juvenile polyposis coli; increased risk of colorectal cancer, *Gut* **25**(7):792-800.

Jass, J. R., 1995, Colorectal adenoma progression and genetic change: is there a link?, *Ann Med* **27**(3):301-6.

Jass, J. R., Williams, C. B., Bussey, H. J., and Morson, B. C., 1988, Juvenile polyposis--a precancerous condition, *Histopathology* **13**(6):619-30.

Jeghers, H., McCusick, V., and Katz, K., 1949, Generalised intestinal polyposis and melanin spots of the oral mucosa, lips and digits: a syndrome of diagnostic significance., *New Engl J Med* **241**:1031-1036.

Jonson, T., Gorunova, L., Dawiskiba, S., Andren-Sandberg, A., Stenman, G., ten Dijke, P., Johansson, B., and Hoglund, M., 1999, Molecular analyses of the 15q and 18q SMAD genes in pancreatic cancer, *Genes Chromosomes Cancer* **24**(1):62-71.

Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D., 1992, Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors, *Science* **258**(5083):818-21.

Kawabata, M., Inoue, H., Hanyu, A., Imamura, T., and Miyazono, K., 1998, Smad proteins exist as monomers in vivo and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors, *Embo J* **17**(14):4056-65.

Kim, I. J., Ku, J. L., Yoon, K. A., Heo, S. C., Jeong, S. Y., Choi, H. S., Hong, K. H., Yang, S. K., and Park, J. G., 2000, Germline mutations of the DPC4 gene in Korean juvenile polyposis patients, *Int J Cancer* **86**(4):529-32.

Kinzler, K. W., Nilbert, M. C., Su, L. K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., McKechnie, D., and et al., 1991a, Identification of FAP locus genes from chromosome 5q21, *Science* **253**(5020):661-5.

Kinzler, K. W., Nilbert, M. C., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hamilton, S. R., Hedge, P., Markham, A., and et al., 1991b, Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers, *Science* **251**(4999):1366-70.

Kinzler, K. W., and Vogelstein, B., 1996, Lessons from hereditary colorectal cancer, *Cell* **87**(2):159-70.

Kinzler, K. W., and Vogelstein, B., 1997, Cancer-susceptibility genes. Gatekeepers and caretakers, *Nature* **386**(6627):761, 763.

Kinzler, K. W., and Vogelstein, B., 1998, Landscaping the cancer terrain, *Science* **280:**1036-1037.

Kirsch, T., Sebald, W., and Dreyer, M. K., 2000, Crystal structure of the BMP-2-BRIA ectodomain complex, *Nat Struct Biol* **7**(6):492-6.

Knudson, A. G., 1971, Mutation and Cancer: statistical study of retinoblastoma, *Proc. Natl. Acad. Sci. USA* **68:**820-823.

Kolodziej, P. A., 1997, DCC's function takes shape in the nervous system, *Curr Opin Genet Dev* **7**(1):87-92.

Koyama, M., Ito, M., Nagai, H., Emi, M., and Moriyama, Y., 1999, Inactivation of both alleles of the DPC4/SMAD4 gene in advanced colorectal cancers: identification of seven novel somatic mutations in tumors from Japanese patients, *Mutat Res* **406**(2-4):71-7.

Kretzschmar, M., Doody, J., Timokhina, I., and Massague, J., 1999, A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras, *Genes Dev* **13**(7):804-16.

Lamlum, H., Ilyas, M., Rowan, A., Clark, S., Johnson, V., Bell, J., Frayling, I., Efstathiou, J., Pack, K., Payne, S., Roylance, R., Gorman, P., Sheer, D., Neale, K., Phillips, R., Talbot, I., Bodmer, W., and Tomlinson, I., 1999, The type of somatic mutation at *APC* in FAP is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis, *Nature Med* **5:**1071-1075.

Lathrop, G. M., Lalouel, J.-M., Julier, C., and Ott, J., 1984, Strategies for multilocus linkage analysis in humans., *Proc natl Acad Sci USA* **81:**3443-6.

Lawrance, I. C., Fiocchi, C., and Chakravarti, S., 2001, Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes, *Hum Mol Genet* **10**(5)**:**445-56.

Leggett, B. A., Thomas, L. R., Knight, N., Healey, S., Chenevix-Trench, G., and Searle, J., 1993, Exclusion of APC and MCC as the gene defect in one family with familial juvenile polyposis, *Gastroent.* **105:**1313-1316.

Lei, J., Zou, T. T., Shi, Y. Q., Zhou, X., Smolinski, K. N., Yin, J., Souza, R. F., Appel, R., Wang, S., Cymes, K., Chan, O., Abraham, J. M., Harpaz, N., and Meltzer, S. J., 1996, Infrequent DPC4 gene mutation in esophageal cancer, gastric cancer and ulcerative colitis-associated neoplasms, *Oncogene* **13**(11)**:**2459-62.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S. H., Giovanella, B. C., Ittmann, M., Tycko, B., Hibshoosh, H., Wigler, M. H., and Parsons, R., 1997, PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer., *Science* **275**(5308)**:**1943-7 issn: 0036-8075.

Liaw, D., Marsh, D. J., Li, J., Dahia, P. L., Wang, S. I., Zheng, Z., Bose, S., Call, K. M., Tsou, H. C., Peacocke, M., Eng, C., and Parsons, R., 1997, Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome, *Nat Genet* **16**(1)**:**1061-4036.

Liu, F., Pouponnot, C., and Massague, J., 1997, Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes, *Genes Dev* **11**(23):3157-67.

Lorentz, O., Duluc, I., Arcangelis, A. D., Simon-Assmann, P., Kedinger, M., and Freund, J. N., 1997, Key role of the Cdx2 homeobox gene in extracellular matrix-mediated intestinal cell differentiation, *J Cell Biol* **139**(6):1553-65.

MacGrogan, D., Pegram, M., Slamon, D., and Bookstein, R., 1997, Comparative mutational analysis of DPC4 (Smad4) in prostatic and colorectal carcinomas, *Oncogene* **15**(9):1111-4.

Maitra, A., Molberg, K., Albores-Saavedra, J., and Lindberg, G., 2000, Loss of dpc4 expression in colonic adenocarcinomas correlates with the presence of metastatic disease, *Am J Pathol* **157**(4):1105-11.

Mallo, G. V., Rechreche, H., Frigerio, J. M., Rocha, D., Zweibaum, A., Lacasa, M., Jordan, B. R., Dusetti, N. J., Dagorn, J. C., and Iovanna, J. L., 1997, Molecular cloning, sequencing and expression of the mRNA encoding human Cdx1 and Cdx2 homeobox. Down-regulation of Cdx1 and Cdx2 mRNA expression during colorectal carcinogenesis, *Int J Cancer* **74**(1):35-44.

Maniotis, A. J., Folberg, R., Hess, A., Seftor, E. A., Gardner, L. M. G., Pe'er, J., Trent, J. M., Meltzer, P. S., and Hendrix, M. J. C., 1999, Vascular channel formation by human melanoma cells *in vivo* and *in vitro*: Vasculogenic Mimicry, *Am J Path* **155**:739-752.

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., Vogelstein, B., and et, a., 1995, Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability, *Science* **268**(5215):1336-8.

Marsh, D., Dahia, P., Coulon, V., Zheng, Z., Dorion-Bonnet, F., Call, K., Little, R., Lin, A., Eeles, R., Goldstein, A., Hodgson, S., Richardson, A., Robinson, B., Weber, H., Longy, M., and C, E., 1998, Allelic imbalance, including deletion of PTEN/MMACI, at the Cowden disease locus on 10q22-23 in hamartomas from patients with Cowden syndrome and germline PTEN mutation, *Genes Chrom Cancer* **21**:61-69.

Marsh, D. J., Dahia, P. L., Zheng, Z., Liaw, D., Parsons, R., Gorlin, R. J., Eng, C., Li, D. M., and Sun, H., 1997a, Germline mutations in PTEN are present in Bannayan-Zonana syndrome, *Nat-Genet* **16**(4):333-334.

Marsh, D. J., Roth, S., Lunetta, K. L., Sistonen, P., Dahia, P. L. M., Hemminki, A., Zheng, Z., Caron, S., van Orsouw, N. J., Bodmer, W. F., Cottrell, S. E., Dunlop, M. G., Eccles, D., Hodgson, S. V., Jarvinen, H., Kellokumpu, I., Markie, D., Neale, K., Phillips, R., Rosen, P., Syngal, S., Vijg, J., Tomlinson, I. P. M., Aaltonen, L. A., and Eng, C., 1997b, Exclusion of PTEN/MMAC1/TEP1 and 10q22-24 as the susceptibility locus for juvenile polyposis syndrome (JPS)., *Cancer Res* **57**:5017-5020.

Massague, J., 1996, TGFbeta signaling: receptors, transducers, and Mad proteins, *Cell* **85**(7):947-50.

Massague, J., 1998, TGF-beta signal transduction, *Annu Rev Biochem* **67**:753-91.

Massague, J., 2000, How cells read TGF-beta signals, *Nat Rev Mol Cell Biol* **1**(3):169-78.

Massague, J., Blain, S. W., and Lo, R. S., 2000, TGFbeta signaling in growth control, cancer, and heritable disorders, *Cell* **103**(2):295-309.

Massague, J., and Chen, Y. G., 2000, Controlling TGF-beta signaling, *Genes Dev* **14**(6):627-44.

Mehlen, P., Rabizadeh, S., Snipas, S. J., Assa-Munt, N., Salvesen, G. S., and Bredesen, D. E., 1998, The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis, *Nature* **395**(6704):801-4.

Miyaki, M., Iijima, T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., Iwama, T., Utsunomiya, J., Kuroki, T., and Mori, T., 1999, Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis, *Oncogene* **18**(20):3098-103.

Moren, A., Itoh, S., Moustakas, A., Dijke, P., and Heldin, C. H., 2000, Functional consequences of tumorigenic missense mutations in the amino-terminal domain of Smad4, *Oncogene* **19**(38):4396-404.

Mori, Y., Yin, J., Rashid, A., Leggett, B. A., Young, J., Simms, L., Kuehl, P. M., Langenberg, P., Meltzer, S. J., and Stine, O. C., 2001, Instabilotyping: comprehensive identification of frameshift mutations caused by coding region microsatellite instability, *Cancer Res* **61**(16):6046-9.

Murata, M., Iwao, K., Miyoshi, Y., Nagasawa, Y., Ohta, T., Shibata, K., Oda, K., Wada, H., Tominaga, S., Matsuda, Y., Ohsawa, M., Nakamura, Y., and Shimano, T., 2000, Molecular and biological analysis of carcinoma of the small intestine: beta-catenin gene mutation by interstitial deletion involving exon 3 and replication error phenotype, *Am J Gastroenterol* **95**(6):1576-80.

Nagase, H., and Nakamura, Y., 1993, Mutations of the APC (adenomatous polyposis coli) gene, *Hum Mutat* **2**(6):425-34.

Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J. L., Heuchel, R., Itoh, S., Kawabata, N., Heldin, N. E., Heldin, C. H., and ten Dijke, P., 1997a, Identification of Smad7, a TGF-beta-inducible antagonist of TGF-beta signalling., *Nature* **389**:631-635.

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C. H., Miyazono, K., and ten Dijke, P., 1997b, TGF-beta receptor mediated signalling through Smad2, Smad3 and Smad 4., *EMBO J.* **16**:5353-5362.

Nelen, M. R., Padberg, G. W., Peeters, E. A. J., Lin, A. Y., Vandenhelm, B., et al., ~~and~~ ~~e.,~~ 1996, Localization of the gene for Cowden disease to chromosome 10q22-23., *Nat Genet* **13**:114-116.

Nelen, M. R., van Staveren, W. C., Peeters, E. A., Hassel, M. B., Gorlin, R. J., Hamm, H., Lindboe, C. F., Fryns, J. P., Sijmons, R. H., Woods, D. G., Mariman, E. C., Padberg, G. W., and Kremer, H., 1997, Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease, *Hum Mol Genet* **6**(8):1383-7

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., and et, a. l., 1994, Mutations of two PMS homologues in hereditary nonpolyposis colon cancer, *Nature* **371**(6492):75-80.

O'Connell, J. R., and Weeks, D. E., 1995, The Vitesse algorithm for rapid, exact multilocus linkage analysis via genotype set recoding and fuzzy inheritance., *Nat Genet* **11**:402-8.

Olschwang, S., Markie, D., Seal, S., Neale, K., Phillips, R., Cottrell, S., Ellis, I., Hodgson, S., Zauber, P., Spigelman, A., Iwama, T., Loff, S., McKeown, C., Marchese, C., Sampson, J., Davies, S., Talbot, I. C., Wyke, J., Thomas, G., Bodmer, W. F., Hemminki, A., Avizienyte, E., de la Chapelle, A., Aaltonen, L. A., and Tomlinson, I. P. M., 1998a, Peutz-Jeghers disease: most, but not all, families are compatible with linkage to 19p13.3, *J. Med. Genet.* **35**:42-44.

Olschwang, S., Serova-Sinilnikova, O. M., Lenoir, G. M., and Thomas, G., 1998b, PTEN germ-line mutations in juvenile polyposis coli., *Nature Genetics* **18**:12-14.

Onichtchouk, D., Chen, Y. G., Dosch, R., Gawantka, V., Delius, H., Massague, J., and Niehrs, C., 1999, Silencing of TGF-beta signalling by the pseudoreceptor BAMBI, *Nature* **401**(6752):480-5.

Paredes-Zaglul, A., Kang, J. J., Essig, Y. P., Mao, W., Irby, R., Wloch, M., and Yeatman, T. J., 1998, Analysis of colorectal cancer by comparative genomic hybridization: evidence for induction of the metastatic phenotype by loss of tumor suppressor genes, *Clin Cancer Res* **4**(4):879-86.

Park, W. S., Park, J. Y., Oh, R. R., Yoo, N. J., Lee, S. H., Shin, M. S., Lee, H. K., Han, S., Yoon, S. K., Kim, S. Y., Choi, C., Kim, P. J., Oh, S. T., and Lee, J. Y., 2000, A distinct tumor suppressor gene locus on chromosome 15q21.1 in sporadic form of colorectal cancer, *Cancer Res* **60**(1):70-3.

Parsons, R., Myeroff, L. L., Liu, B., Willson, J. K., Markowitz, S. D., Kinzler, K. W., and Vogelstein, B., 1995, Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer, *Cancer Res* **55**(23):5548-50.

Prunier, C., Ferrand, N., Frottier, B., Pessah, M., and Atfi, A., 2001, Mechanism for mutational inactivation of the tumor suppressor Smad2, *Mol Cell Biol* **21**(10):3302-13.

Reiss, M., 1999, TGF-beta and cancer., *Microbes Infect* **1**:1327–47.

Roberts, A. B., 1999, TGF-beta signaling from receptors to the nucleus, *Microbes Infect* **1**(15):1265-73.

Roth, S., Laiho, P., Salovaara, R., Launonen, V., and Aaltonen, L. A., 2000, No SMAD4 hypermethylation in colorectal cancer, *Br J Cancer* **83**(8):1015-9.

Roth, S., Sistonen, P., Salovaara, R., Hemminki, A., Loukola, A., Johansson, M., Avizienyte, E., Cleary, K. A., Lynch, P., Amos, C. I., Kristo, P., Mecklin, J. P., Kellokumpu, I., Jarvinen, H., and Aaltonen, L. A., 1999, SMAD genes in juvenile polyposis, *Genes Chromosomes Cancer* **26**(1):54-61.

Rowan, A., Bataille, V., MacKie, R., Healy, E., Bicknell, D., Bodmer, W., and Tomlinson I. P. M., 1999, Somatic mutations in the Peutz-Jeghers (LKB1/STKII) gene in sporadic malignant melanomas, *J Invest Dermatol* **112:**509-511.

Rowan, A. J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W. F., and Tomlinson, I. P. M., 2000, APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits", *Proc Natl Acad Sci U S A* **97**(7)**:**3352-7.

Sanger, F., 1981, Determination of nucleotide sequences in DNA, *Science* **214:**1205-1210.

Schutte, M., Hruban, R. H., Hedrick, L., Cho, K. R., Nadasdy, G. M., Weinstein, C. L., Bova, G. S., Isaacs, W. B., Cairns, P., Nawroz, H., Sidransky, D., Casero, R. A., Jr., Meltzer, P. S., Hahn, S. A., and Kern, S. E., 1996, DPC4 gene in various tumor types, *Cancer Res* **56**(11)**:**2527-30.

Scrable, H., Cavenee, W., Ghavimi, F., Lovell, M., Morgan, K., and Sapienza, C., 1989, A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting, *Proc Natl Acad Sci U S A* **86**(19)**:**7480-4.

Sepp, T., Yates, J. R., and Green, A. J., 1996, Loss of heterozygosity in tuberous sclerosis hamartomas, *J Med Genet* **33**(11)**:**962-4.

Shi, Y., Hata, A., Lo, R. S., Massague, J., and Pavletich, N. P., 1997, A structural basis for mutational inactivation of the tumour suppressor Smad4, *Nature* **388**(6637)**:**87-93.

Shibata, D., Reale, M. A., Lavin, P., Silverman, M., Fearon, E. R., Steele, G., Jr., Jessup, J. M., Loda, M., and Summerhayes, I. C., 1996, The DCC protein and prognosis in colorectal cancer, *N Engl J Med* **335**(23):1727-32.

Soravia, C., Berk, T., Madlensky, L., Mitri, A., Cheng, H., Gallinger, S., Cohen, Z., and Bapat, B., 1998, Genotype-phenotype correlations in attenuated adenomatous polyposis coli, *Am J Hum Genet* **62**(6):1290-301.

Souchelnytskyi, S., Tamaki, K., Engstrom, U., Wernstedt, C., ten Dijke, P., and Heldin, C. H., 1997, Phosphorylation of Ser(465) and Ser(467) in the C terminus of Smad2 mediates interaction with Smad4 and is required for TGF-beta signalling., *J. Biol. Chem.* **272:**28107-28115.

Southern, E. M., 1975, Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J Mol Biol* **98:**503-517.

Stopa, M., Anhuf, D., Terstegen, L., Gatsios, P., Gressner, A. M., and Dooley, S., 2000, Participation of SMAD2, SMAD3, and SMAD4 in transforming growth factor beta (TGF-beta)-induced activation of SMAD7. *J Biol Chem* **275**(38):29308-17.

Su, G. H., Hruban, R. H., Bansal, R. K., Bova, G. S., Tang, D. J., Shekher, M. C., Westerman, A. M., Entius, M. M., Goggins, M., Yeo, C. J., and Kern, S. E., 1999, Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers, *Am J Pathol* **154**(6):1835-40.

Takagi, Y., Kohmura, H., Futamura, M., Kida, H., Tanemura, H., Shimokawa, K., and Saji, S., 1996, Somatic alterations of the DPC4 gene in human colorectal cancers in vivo, *Gastroenterology* **111**(5):1369-72.

Takaku, K., Miyoshi, H., Matsunaga, A., Oshima, M., Sasaki, N., and Taketo, M. M., 1999, Gastric and duodenal polyps in Smad4 (Dpc4) knockout mice, *Cancer Res* **59**(24):6113-7.

Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin, M. F., and Taketo, M. M., 1998, Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes, *Cell* **92**(5):645-56.

Tamai, Y., Nakajima, R., Ishikawa, T., Takaku, K., Seldin, M., and Taketo, M., 1999, Colonic hamartoma development by anomalous duplication in Cdx2 knockout mice., *Cancer Res* **59**(12):2965.

Tarafa, G., Villanueva, A., Farre, L., Rodriguez, J., Musulen, E., Reyes, G., Seminago, R., Olmedo, E., Paules, A. B., Peinado, M. A., Bachs, O., and Capella, G., 2000, DCC and SMAD4 alterations in human colorectal and pancreatic tumor dissemination, *Oncogene* **19**(4):546-55.

Thiagalingam, S., Lengauer, C., Leach, F. S., Schutte, M., Hahn, S. A., Overhauser, J., Willson, J. K., Markowitz, S., Hamilton, S. R., Kern, S. E., Kinzler, K. W., and Vogelstein, B., 1996, Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers, *Nat Genet* **13**(3):343-6.

Thomas, H. J. W., Whitelaw, S. C., Cottrell, S. E., Murday, V. A., Tomlinson, I. P. M., Markie, D., Jones, T., Bishop, D. T., Hodgson, S. V., Sheer, D., Northover, J. M. A., Talbot, I. C., Solomon, E., and Bodmer, W. F., 1996, Genetic mapping of the hereditary mixed polyposis syndrome to chromosome 6q, *Am J Hum Genet* **58**:770-776.

Thompson, C., Le Boit, P., Nederlof, P., and Gray, J., 1994, Thick-section fluorescence in situ hybridization on formalin-fixed, paraffin-embedded archival tissue provides a histogenetic profile., *Am J Pathol* **144:**237-243.

Tomlinson, I. P. M., Rahman, N., Frayling, I., Mangion, J., Barfoot, R., Hamoudi, R., Seal, S., Northover, J., Thomas, H. J., Neale, K., Hodgson, S., Talbot, I., Houlston, R., and Stratton, M. R., 1999, Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22, *Gastroenterology* **116**(4)**:**789-95.

Tommasino, M., and Crawford, L., 1995, Human papillomavirus E6 and E7: proteins which deregulate the cell cycle, *Bioessays* **17**(6)**:**509-18.

Trojan, J., Raedle, J., and Zeuzem, S., 2000, Germline mutations of the MSH6 gene in patients with an atypical hereditary nonpolyposis colorectal carcinoma (HNPCC), *Z Gastroenterol* **38**(7)**:**607-9.

Unden, A., Holmberg, E., Lundh-Rozell, B., Stahle-Backdahl, M., Zaphiropoulos, P., Toftgard, R., and Vorechovsky, I., 1996, Mutations in the human homologue of Drosophila patched (PTCH) in basal cell carcinomas and the Gorlin syndrome: different in vivo mechanisms of PTCH inactivation., *Cancer Res* **56**(20)**:**4562-5.

Veale, A., McColl, I., Bussey, H., and Morson, B., 1966, Juvenile polyposis coli, *J Med Genet* **3:**5-16.

Vogelstein, B. V., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. M., and Bos, J. L., 1988,

Genetic alterations during colorectal tumour development, *N Engl J Med* **319:**525-532.

von Gersdorff, G., Susztak, K., Rezvani, F., Bitzer, M., Liang, D., and Bottinger, E. P., 2000, Smad3 and Smad4 mediate transcriptional activation of the human Smad7 promoter by transforming growth factor beta, *J Biol Chem* **275**(15):11320-6.

Wagner, A., Hendriks, Y., Meijers-Heijboer, E. J., de Leeuw, W. J., Morreau, H., Hofstra, R., Tops, C., Bik, E., Brocker-Vriends, A. H., van Der Meer, C., Lindhout, D., Vasen, H. F., Breuning, M. H., Cornelisse, C. J., van Krimpen, C., Niermeijer, M. F., Zwinderman, A. H., Wijnen, J., and Fodde, R., 2001, Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree, *J Med Genet* **38**(5):318-22.

Wang, Z., Ellis, I., Zauber, P., Iwama, T., Marchese, C., Talbot, I., Xue, W., Yan, Z., and Tomlinson, I., 1999, Allelic imbalance at the LKB1 (STK11) locus in tumours from patients with Peutz-Jeghers' syndrome provides evidence for a hamartoma-(adenoma)-carcinoma sequence, *J Pathol* **188:**9-13.

Weinstein, L. S., and Shenker, A., 1993, G protein mutations in human disease, *Clin Biochem* **26**(5):333-8.

Weiss, M. M., Hermsen, M. A., Meijer, G. A., van Grieken, N. C., Baak, J. P., Kuipers, E. J., and van Diest, P. J., 1999, Comparative genomic hybridisation, *Mol Pathol* **52**(5):243-51.

Whitelaw, S. C., Murday, V. A., Tomlinson, I. P. M., Thomas, H. J. W., Cottrell, S. E., Ginsberg, A., Bukofzer, S., Hodgson, S. V., Skudowitz, R., Jass, J. R., Talbot, I.

C., Northover, J. M. A., Bodmer, W. F., and Solomon, E., 1996, Clinical and molecular features of the Hereditary Mixed Polyposis Syndrome., *Gastroenterology* **112:**327-334.

Wicking, C., Simms, L. A., Evans, T., Walsh, M., Chawengsaksophak, K., Beck, F., Chenevix-Trench, G., Young, J., Jass, J., Leggett, B., and Wainwright, B., 1998, CDX2, a human homologue of Drosophila caudal, is mutated in both alleles in a replication error positive colorectal cancer, *Oncogene* **17**(5)**:**657-9.

Wilentz, R. E., Su, G. H., Dai, J. L., Sparks, A. B., Argani, P., Sohn, T. A., Yeo, C. J., Kern, S. E., and Hruban, R. H., 2000, Immunohistochemical labeling for DPC4 mirrors genetic status in pancreatic adenocarcinomas : a new marker of DPC4 inactivation, *Am J Pathol* **156**(1)**:**37-43.

Wrana, J. L., 1998, TGF-beta receptors and signalling mechanisms, *Miner Electrolyte Metab* **24**(2-3)**:**120-30.

Wrana, J. L., and Attisano, L., 2000, The Smad pathway, *Cytokine Growth Factor Rev* **11**(1-2)**:**5-13.

Xu, J., and Attisano, L., 2000, Mutations in the tumor suppressors Smad2 and Smad4 inactivate transforming growth factor beta signaling by targeting Smads to the ubiquitin-proteasome pathway, *Proc Natl Acad Sci U S A* **97**(9)**:**4820-5.

Xu, L., Chen, Y. G., and Massague, J., 2000, The nuclear import function of Smad2 is masked by SARA and unmasked by TGFbeta-dependent phosphorylation, *Nat Cell Biol* **2**(8)**:**559-62.

Yagi, O., Akiyama, Y., and Yuasa, Y., 1999, Genomic structure and alterations of homeobox gene CDX2 in colorectal carcinomas., *Br J Cancer* **79**(3-4):440-444.

Yamamoto, N., Akiyama, S., Katagiri, T., Namiki, M., Kurokawa, T., and Suda, T., 1997, Smad1 and Smad5 act downstream of intracellular signallings of BMP-2 that inhibits myogenic differentiation and induces osteoblast differentiation in C2C12 myoblasts., *Bioch. Biophys. Res. Comm.* **238:**574-580.

Yin, J., Harpaz, N., Tong, Y., Huang, Y., Laurin, J., Greenwald, B. D., Hontanosas, M., Newkirk, C., and Meltzer, S. J., 1993, p53 Point mutations in dysplastic and cancerous ulcerative colitis lesions, *Gastroenterology* **104**(6):1633-9.

Zhou, S., Buckhaults, P., Zawel, L., Bunz, F., Riggins, G., Dai, J. L., Kern, S. E., Kinzler, K. W., and Vogelstein, B., 1998, Targeted deletion of Smad4 shows it is required for transforming growth factor beta and activin signaling in colorectal cancer cells, *Proc Natl Acad Sci U S A* **95**(5):2412-6.

Zhu, H., Kavsak, P., Abdollah, S., Wrana, J. L., and Thomsen, G. H., 1999, A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation, *Nature* **400**(6745):687-93.

Zhu, Y., Richardson, J., Parada, L., and Graff, J., 1998, Smad3 mutant mice develop metastatic colorectal cancer., *Cell* **94**(6):703-14.

APPENDIX ONE

TWO-POINT LOD SCORES FOR THE JPS GENOME SCREEN

TWO-POINT LOD SCORES FOR THE JPS GENOME SCREEN

The two-point LOD scores for each chromosome (p-arm through to q-arm) are shown from left to right for θ=0 only. The total two-point LOD score for all families combined are shown at the top of each column.

Chromosome 1

ID	D1S468	D1S1612	D1S1597	GATA29 A05	D1S552	GATA129 H04	D1S2134	GATA16 5C03	D1S1665
	-3.3	-0.42	-0.04	-4.08	-3.3	-1.43	-2.78	-2.24	-0.61
MD	0	0.17	0.14	-1.18	-1.18	-1.37	-2.65	0.16	-1.37
6	0	-0.51	-0.45	-0.25	0.03	-0.39	0.34	-0.01	0.26
10	-0.7	-0.08	0	0	-0.16	0	0.21	0.19	0
12	-0.91	0	-0.1	-1.16	-0.1	0.13	-0.91	-0.91	0.21
15	-1.53	0	0.25	-1.43	-1.19	0.2	0	-1.45	0.29
18	0	0	0.12	-0.06	-0.7	0	0.23	-0.22	0
19	-0.16	0	0	-0.7	0	-0.06	-0.7	0.19	0.16

ID	D1S1728	D1S551	D1S1631	GATA176 G01	D1S534	D1S1653	D1S1679	D1S1677	D1S1589
	1.09	-0.11	-3.59	-2.52	-2.18	-1.17	-1.42	-0.97	-0.09
MD	0.15	-1.08	-1.37	-1.3	-1.31	-1.05	-1.31	-1.37	0.05
6	-0.31	0	-1.2	-1.2	-1.2	-1.19	0	-1.19	0.46
10	0.17	0.14	-0.06	0.21	-0.08	0.18	-0.09	-0.06	0.12
12	0.87	0.43	-1.29	-0.91	0	0.44	0.7	0.8	0.24
15	0.06	0.16	0.16	0.23	0.25	0.29	-1.16	0.46	-1.16
18	0.15	0.24	0.17	0.22	0	0.16	0.23	0.22	0.08
19	0.17	0.17	0	0.23	0.16	0	0.21	0.17	0.12

ID	D1S518	D1S1660	D1S1678	GATA124 F08	D1S549	D1S3462	D1S235	D1S547	D1S1609
	-3.59	0.64	-1.37	-3.71	-3.26	-3.01	-2.12	-5.29	-2.13
MD	-1.31	0.11	0	0.11	-1.44	-1.04	-0.34	-1.18	0.29
6	-1.2	0	-0.44	0	0	-1.19	0	0	-1.19
10	-0.02	-0.08	0.16	0.16	-0.1	-0.05	-0.05	0	-0.11
12	0.36	0.24	0.23	-1.71	-1.67	-0.91	-0.42	-0.72	0.44
15	-1.17	0.29	-1.41	-1.43	0.06	0.15	-1.16	-2.36	-1.45
18	-0.22	0.22	-0.07	-0.7	0	0.19	0	-0.33	0
19	-0.03	-0.14	0.16	-0.14	-0.11	-0.16	-0.15	-0.7	-0.11

Chromosome 2

ID	GATA165 C07	GATA116 B01	D2S1400	D2S1360	D2S1788	D2S405	D2S1356	D2S2739	D2S441
	-0.33	-3.73	-2.08	-4.57	-0.69	-6.87	-3.25	-4.98	-1.63
MD	-0.03	-0.12	-1.9	0.3	0.13	-1.68	0.2	0.28	0.24
6	0.13	0	-0.16	-1.77	0	-1.77	-1.77	-1.77	0
10	-1.82	-1.82	-0.11	-1.82	0	-1.82	-0.04	-1.82	-0.08
12	0.6	-0.22	-0.17	-1.55	-1.72	0.22	-1.55	-1.57	0.27
15	0.6	-1.83	0	0.06	0.6	0	0.03	0	-1.9
18	0.19	0.26	0.1	0.25	0	-1.82	-0.04	-0.02	-0.08
19	0	0	0.16	-0.04	0.3	0	-0.08	-0.08	-0.08

ID	D2S1394	D2S1777	D2S1790	D2S410	D2S1328	D2S1334	D2S442	D2S1399	D2S1353
	-7.96	-6.26	-9.8	-7.86	-3.62	-4.18	-4.02	-0.6	-2.35
MD	-2.01	-1.72	-3.77	-1.98	0.12	0	0.12	0.19	0.47
6	0.55	-0.24	-1.88	0	-1.77	-1.77	-1.77	0	0
10	-0.2	-0.11	-0.2	-1.82	-0.2	-0.14	-0.04	-0.08	0
12	-1.92	-2.08	-2.04	-2.02	-0.17	-0.43	-0.17	0.89	0
15	-3.89	0	0.11	-1.9	-1.89	-1.82	-1.83	-1.68	-2.68
18	-0.29	-0.29	-0.2	0	0.29	0	-0.29	0.25	-0.04
19	-0.2	-1.82	-1.82	-0.14	0	-0.02	-0.04	-0.17	-0.1

ID	D2S1776	D2S1391	D2S1384	GATA30 E06	D2S434	D2S1363	D2S427	GATA178 G09	D2S125
	-3.32	-1.96	-3.16	-3.06	-3.83	-5.45	-0.65	-1.1	-0.45
MD	0.44	-0.39	0.23	0.3	-1.68	-1.84	-1.68	-0.07	0.45
6	-1.77	-1.77	0.11	0.51	0.1	-0.14	0.36	0.26	0.11
10	0.19	-0.14	0.23	-0.14	-0.04	-1.82	0.23	0.21	0.23
12	0.02	0.22	-1.48	-1.91	-1.99	0.41	-0.19	-1.63	-1.54
15	-2.06	0.12	-2.39	0	-0.13	-1.68	0.35	0.05	0.44
18	0	0	-0.07	0	-0.29	-0.29	0.1	-0.11	-0.05
19	-0.14	0	0.21	-1.82	0.2	-0.09	0.18	0.19	-0.09

Chromosome 3

ID	D3S2387	GATA164B0 8	D3S3038	D3S2432	D3S1768	D3S2409	D3S1766	GATA14E04
	-4.64	-2.08	-3.71	-3.31	-2.15	-3.37	-4.21	-6.52
MD	0.47	-0.07	-0.11	0.27	0	-1.74	0.12	-1.68
6	0.29	0	-1.77	0	0	0	-0.86	0.05
10	0.24	-0.06	0.24	0.2	-0.25	-0.11	-0.04	0.2
12	-2	-0.17	-0.47	-0.08	-0.26	0.43	-1.8	-1.5
15	-1.72	-1.84	-1.61	-3.89	0.01	-1.93	-1.85	-1.97
18	-0.1	0.16	0.18	0.25	-1.82	0.18	0	-1.82
19	-1.82	-0.1	-0.17	-0.06	0.17	-0.2	0.22	0.2

ID	D3S2406	GATA128C02	D3S2459	D3S3045	D3S2460	ATA34G06	D3S1764	D3S1744
	-2.8	-3.52	-3.54	-2.6	0.51	-1.42	-2.79	-5.62
MD	0.46	0.29	0.52	0.6	0.2	0.6	0.3	-1.62
6	-1.88	0	0	0.49	-0.14	-0.55	0.11	-0.43
10	0.27	-0.08	-0.1	-0.04	-0.11	-0.08	-1.82	-1.82
12	-0.19	-1.81	-1.97	-1.6	0.17	-1.64	-0.18	0
15	-1.72	-1.71	-1.94	-1.97	0.1	0	-1.64	-1.88
18	0.26	-0.07	0	-0.02	0.1	0.25	0.22	0.21
19	0	-0.14	-0.05	-0.06	0.19	0	0.22	-0.08

ID	D3S1763	D3S3053	D3S2427	D3S2398	D3S2418	D3S1311
	-7.09	-3.11	-2.86	-6.63	-3.81	0.54
MD	-1.91	-1.56	-1.83	0	0	0.3
6	-0.44	0.27	-0.33	-1.77	0	0
10	-1.82	-0.08	-1.82	-1.82	-0.04	0.3
12	-1.74	-1.95	0.49	-1.79	-0.16	-0.17
15	-1.67	0	0.4	0.36	-1.97	-0.19
18	0.27	0	0	-1.82	-1.82	0.3
19	0.22	0.21	0.23	0.21	0.18	0

Chromosome 4

ID	D4S2366	D4S403	D4S2639	D4S2397	D4S2632	D4S1627	D4S3248	D4S2367	D4S3243	D4S2361
	-1.18	-0.29	-4.64	-3.78	-7.1	-5.27	-4.07	-6.91	-6.08	0.19
MD	0.18	0.3	-1.77	-1.91	-1.77	-0.04	-0.14	-1.77	-1.97	0.44
6	0.57	0.54	-1.77	0	-1.77	-1.77	-1.77	-1.77	0.18	0.21
10	-0.14	-0.11	0.27	-0.11	0.22	0.22	-0.1	0.19	-0.17	0
12	0.02	-1.56	-1.59	-2.39	-2.04	-1.88	-1.81	-1.55	-1.55	-1.5
15	-1.77	0.27	0	0.16	-1.78	-1.92	-0.52	-2.08	-2.68	0.6
18	0.1	0	0.26	0.26	0.15	-0.08	0.27	0.13	0.28	0.2
19	-0.14	0.27	-0.04	0.21	-0.11	0.2	0	-0.06	-0.17	0.24

ID	D4S1647	D4S2623	D4S1625	D4S1629	D4S2368	D4S2431	D4S2417	D4S408	D4S1652
	-0.62	-4.93	0.31	-0.12	-5.23	-5.15	-3.2	0.82	-0.43
MD	0.47	-1.68	0.07	0.54	-1.97	-1.99	-1.94	0	0.15
6	0.12	-1.77	0	0.46	0.32	0.41	0	0.22	-0.47
10	0.21	-0.06	0	0.19	0	-0.08	0.21	0.3	-0.11
12	0.3	0.3	0.36	0.4	-1.79	-1.8	0.31	0	0
15	-1.64	-1.64	-0.24	-1.9	-1.9	-1.97	-1.91	0	0
18	0	0	-0.14	0	0.11	0.1	-0.08	0.3	0
19	-0.08	-0.08	0.26	0.19	0	0.18	0.21	0	0

Chromosome 5

APC is flanked by D5S2501 and D5S1505, highlighted blue.

ID	D5S2488	D5S2505	D5S807	D5S817	GATA134B03	GATA145D09	D5S1470	D5S1457	D5S2500	D5S1501	GATA808
	-4.07	-3.78	-5.5	-0.94	-3.8	-2.08	-4.52	-1.26	-3.08	-2.62	-5.81
MD	-1.97	-0.12	-1.84	-0.12	-1.87	-1.52	0.3	-2.06	0.6	-0.16	-1.85
6	0	0	-1.77	-1.77	0	-1.77	-1.77	0	-1.77	-1.77	-1.77
10	-0.2	-0.08	-0.14	0.18	-0.11	0.23	-0.02	-0.08	-0.08	0.23	-0.14
12	-2.06	-1.8	0.37	0.49	-0.3	0.49	-1.47	0.52	-0.11	-1.5	-2.02
15	0.46	0.08	-1.9	0.16	-1.97	0.02	-1.76	0.22	-1.9	0.3	0.19
18	-0.1	-0.04	-0.05	-0.08	0.24	0.26	0.28	0.28	0.26	0.1	-0.08
19	-0.2	-1.82	-0.17	0.2	0.21	0.21	-0.08	-0.14	-0.08	0.18	-0.14

ID	D5S1462	D5S1453	D5S2501	D5S1505	D5S816	D5S1480	D5S820	D5S1471	D5S1456	D5S211	D5S408
	-4.59	-9.34	-5.46	-5	-1.58	-4.65	-1	-5.14	-1.06	0.29	-0.93
MD	-1.83	-1.87	-1.94	-1.83	-1.67	-0.1	0	-1.97	-1.72	-0.1	-0.16
6	-1.77	-1.77	-1.77	-1.77	0	-1.77	0	-1.77	0	0	-1.77
10	0.21	0.23	-0.08	-0.11	0.1	0.24	0	-0.14	-0.14	0	-0.08
12	-1.88	-3.75	-2.01	-1.72	-0.12	-1.65	0.26	0.16	0.38	0.09	0.6
15	0.55	-2.35	0.2	0.6	0.01	0.49	0.38	-1.64	0.17	0.3	0.3
18	-0.08	-0.05	-0.07	-0.06	0	-1.82	-1.64	0	0	0	0
19	0.21	0.22	0.21	-0.11	0.1	-0.04	0	0.22	0.25	0	0.18

Chromosome 6

ID	F13A1	D6S1959	GGAA15B08	D6S1017	GATA11E02	D6S1053	D6S1031	D6S1056	D6S1021
	-3.17	-3.63	-2.51	-3.52	-2.82	-3.49	-1.01	-2.57	-4.66
MD	0.14	-1.97	0.51	-1.68	-1.7	0.13	-1.74	0	-1.68
6	-1.77	0.26	0	0	0.3	0	0.53	-1.77	-1.77
10	-1.82	-0.14	0.24	0.2	0.21	0.22	0	-0.04	0.23
12	-0.03	-0.23	-1.74	-1.96	-1.73	-2.05	0.2	-1.55	-1.81
15	0.3	-1.9	0.3	-0.12	-0.13	-1.64	0	0.3	0
18	0.15	0.12	0	0.12	0	-0.07	0	0.26	0.12
19	-0.14	0.23	-1.82	-0.08	0.23	-0.08	0	0.23	0.25

ID	D6S474	D6S1040	D6S1009	GATA184A08	GATA165G02	D6S305	D6S1277	D6S1027
	-0.94	-6.53	-0.76	-2.55	-4.74	-2.36	-4.57	-6.94
MD	0	-1.68	0.6	0.17	0.45	0.5	0.51	0.3
6	0	0.06	0.35	0	-1.77	0	-1.77	-1.77

10	0	0.21	0.18	0.19	0.21	-0.14	-0.04	-1.82
12	-1.68	-1.54	-1.38	-1.62	0.22	0.59	0.27	-1.84
15	0.36	-1.97	-0.11	0.3	-1.92	-1.64	-1.89	-1.6
18	0.2	-1.82	-0.29	-1.82	-1.82	0.15	0.15	-0.1
19	0.18	0.21	-0.11	0.23	-0.11	-1.82	-1.8	-0.11

Chromosome 7

ID	GATA2 4F03	D7S513	GATA 137H02	D7S1802	D7S1808	D7S817	D7S2846	D7S1818	GATA 118G10	D7S2204
	-4.89	-0.7	-3.4	-5.6	-3.5	-1.3	1.28	-6.9	-4.2	-1.2
MD	-1.68	0.49	-1.78	-1.68	-1.87	0	0.39	-1.97	-1.68	-1.83
6	0.21	-1.77	-1.77	-1.77	-0.45	-1.77	0.1	-1.77	-1.77	0
10	-0.14	-0.08	-0.14	-0.08	0	0.17	0	0.23	-0.08	-0.1
12	-1.81	0.43	0.14	0.16	0.35	0.41	0.6	0.9	0.89	0.3
15	-1.95	0	0.25	-1.83	0.37	-0.14	0.18	-2.68	-1.72	0.3
18	0.26	0.28	-0.1	-0.29	-0.11	0.17	0.21	0.21	0.23	0.25
19	0.22	0	0	-0.14	-1.82	-0.11	-0.2	-1.82	-0.08	-0.1

ID	D7S820	D7S821	D7S1799	GGAA 6D03	D7S1804	D7S1824	D7S2195	GATA 189C06	GATA 30D09	D7S559
	-4.73	-4.18	-3.09	-4.34	-4.78	1.55	-2.69	-4.5	-0.93	-1.79
MD	-1.64	-0.1	0.16	0	-1.68	0.03	-3.77	-1.68	0.3	0.3
6	0	-1.77	-1.77	-0.53	0	0.57	0.54	0.5	0.04	0
10	-1.82	-0.11	-0.08	0.25	0.25	0.23	-0.04	-0.11	-0.11	-1.82
12	0	-0.17	0.22	-2.01	-1.9	0	0	-3.75	0	0.87
15	-1.74	-1.81	0.2	-1.97	-1.97	0.48	0.3	0.3	-1.64	-1.64
18	0.26	-0.11	-1.82	-0.08	0.27	0	0.28	0	0.26	0.26
19	0.21	-0.11	0	0	0.25	0.24	0	0.24	0.22	0.24

Chromosome 8

ID	D8S264	D8S277	D8S1130	D8S1106	D8S1145	D8S136	D8S1477	D8S1110	D8S1113
	-1.9	-2.91	-1.91	-2.46	0.56	-1.15	-1.88	-2.22	-4.54
MD	0	0.17	0.3	0.22	-0.13	0.12	0.6	0.22	-1.84
6	0	0.18	0.3	0.4	0	0.49	0	0.57	0.6
10	-1.82	-0.08	-0.08	-0.08	-0.09	-0.04	-1.82	-1.82	-0.2
12	0	-1.66	0.14	-1.32	0.57	0.13	-1.6	0.4	0.32
15	0	-1.72	-2.68	-1.88	0	-1.81	0.44	0.23	-1.6
18	0	0.28	0.19	0.28	0.3	-0.29	0.28	0	0
19	-0.08	-0.08	-0.08	-0.08	-0.09	0.25	0.22	-1.82	-1.82

ID	D8S1136	GATA1 4E09	D8S1119	D8S1132	D8S592	D8S1179	D8S1128	D8S256	D8S373
	-0.93	-3.15	-3.75	-7.93	-5.98	-7.39	-4.94	-1.33	-0.5
MD	-0.11	-1.77	0.3	0.23	0.24	-0.11	0.3	0.3	0.4
6	0.25	0	-1.77	-1.77	-1.77	-1.77	-1.77	-1.77	0.19
10	0.21	-1.82	0.19	-1.82	-1.82	-0.14	-0.11	0	0.2
12	0.18	0.44	-0.74	-1.6	-0.17	-1.55	-1.66	0.16	-1.54
15	-1.64	0	-1.73	-2.6	-2.68	-1.89	-1.89	0	0.3
18	0.26	0	0	-0.29	0	-1.82	-0.05	-0.02	-0.29
19	-0.08	0	0	-0.08	0.22	-0.11	0.24	0	0.24

Chromosome 9

ID	GATA62F03	D9S925	D9S1121	D9S1118	D9S301	D9S1122	D9S922
	-3.95	-5.42	-2.83	-3.21	0.33	-1.68	-2.24
MD	-1.89	0.19	0.26	0	-0.07	-1.97	0.07
6	0.12	-0.54	0.44	0	0	-0.46	-0.7
10	0.2	0.22	0.23	0	0.25	0.21	0.21
12	-0.09	-1.77	-1.75	0.46	0	0.13	0.44
15	-2.58	-1.95	-1.68	-1.77	0	0.2	-1.97
18	0.11	-1.82	-0.29	-1.82	0.15	0	-0.29
19	0.18	0.25	-0.04	-0.08	0	0.21	0

ID	D9S257	D9S910	D9S930	D9S934	D9S282	D9S158
	-1.17	-1.53	-4.34	-4.68	-1	-3.74
MD	-1.9	0	-3.77	-1.91	-1.94	0
6	0.56	0	0.3	0.32	0.51	0
10	0.23	-0.11	-0.08	-0.11	-0.14	0
12	0.25	0.48	-0.32	-0.61	0.35	-1.97
15	0	0.06	0	-1.97	0	0.05
18	-0.31	-1.82	-0.36	-0.29	0.22	0
19	0	-0.14	-0.11	-0.11	0	-1.82

Chromosome 10

BMPR1A is flanked by D10S2327 and GATA115E01, shown in blue.

ID	D10S1435	D10S189	D10S1412	D10S1423	D10S1426	D10S1208	D10S1221	D10S1225	GATA1 21A08	D10S1432
	-5.74	-4.38	-3.11	-0.25	-1.48	-3.98	0.97	0.31	-2.93	-0.88
MD	0.04	-1.79	-1.77	0.21	-1.84	0	0.17	0.33	0.21	0
6	-1.77	0.45	0	0.56	0	-1.77	0	0.59	-0.33	0.31
10	0.23	0.21	0.22	-0.05	-0.05	-0.1	0.21	0.21	0.25	0.22
12	-1.72	-1.54	-0.25	0.26	-0.09	-0.57	-0.12	-1.61	-1.91	0.26
15	-2.63	-1.97	-1.83	-1.74	0.17	-1.76	0.59	0.37	-1.64	-1.88

18	-0.09	0.26	0.28	0.27	0.12	0	0.26	0.21	0.27	0.21
19	0.2	0	0.24	0.24	0.21	0.22	-0.14	0.21	0.22	0

ID	D10S2327	GATA115E01	D10S677	D10S1239	D10S1237	D10S1230	D10S1213	D10S1248	D10S212
	-6.29	0.93	-1.64	-2.28	1.77	-1.26	2.04	-1.68	-2.51
MD	0	0	-0.03	0.14	0.53	0.49	0.3	0	0.55
6	-0.5	0.32	0	0.1	0.42	0	0.56	0	0.01
10	-0.04	0	0.22	0.21	-0.08	-0.14	-0.11	0	0.23
12	-3.78	0.34	-0.06	-1.57	0.12	-1.67	0.88	0.12	-1.85
15	-1.9	0.3	-1.68	-1.64	0.28	0.08	0.39	-1.89	-1.66
18	-0.07	-0.29	-0.29	0.28	0.28	0.12	0.13	0.26	0
19	0	0.26	0.2	0.2	0.22	-0.14	-0.11	-0.17	0.21

Chromosome 11

ID	D11S1984	D11S2362	D11S1999	D11S1981	ATA34E08	D11S1392	D11S2371	D11S2002	D11S2000	D11S1998	D11S4464	D11S912	D11S2359
	-2.34	-4.44	-2.98	0.43	-1.29	-5.65	-0.35	-4.26	0.69	1.01	-3.28	-5.01	0.48
MD	-1.83	0.34	0.3	0	0	0.3	0	-1.71	0.17	0.52	0.12	-1.66	0.1
6	0.3	0	-1.77	0	-0.16	-1.77	0	0	0	0.3	0	-1.77	0
10	0.24	-1.82	-0.11	0	-1.82	-0.08	0	-0.11	0.25	0.21	0.2	-0.04	0.22
12	-1.55	-1.47	-1.77	0.35	0.44	-1.91	-0.47	-0.24	0.08	-0.24	-1.94	-1.78	0.16
15	0.3	-1.88	0	0.19	0.11	-1.97	0.34	-1.8	0.19	0.3	-1.8	0	0
18	-0.04	0.15	0.13	-0.11	0.28	-0.08	-0.14	-0.29	0	0	0.25	0.24	0
19	0.24	0.24	0.24	0	-0.14	-0.14	-0.08	-0.11	0	-0.08	-0.11	0	0

Chromosome 12

ID	D12S372	GATA49D12	D12S391	D12S373	D12S1042	GATA91H06	D12S398	D12S1294
	-0.32	-4.65	-6.57	1.11	-4.73	-4.14	0.39	-5.87
MD	0.22	-1.97	-3.77	0.46	0.3	0.6	0.08	0
6	0.26	-0.17	-0.55	0	-1.77	-1.77	0	-1.77
10	0.22	-0.14	-0.04	0.21	0.25	0.26	0.17	0.21
12	-1.45	0.13	0.55	0.28	-1.77	-1.47	-0.12	-2.76
15	0.23	-2.68	-2.39	0.3	-1.87	-1.9	0.15	-1.86
18	0	0	-0.29	-0.05	-0.12	-0.08	-0.06	0.1
19	0.2	0.18	-0.08	-0.09	0.25	0.22	0.17	0.21

ID	D12S1052	D12S1064	D12S1300	PAH	D12S2070	D12S395	D12S2078	D12S1045	D12S392
	-6.64	-1.2	-3.89	-5.04	-5.36	-1.66	-0.08	-5.94	-1.31
MD	-1.76	0.27	0.27	-1.73	-1.97	0	0.23	-1.97	0.46
6	-1.77	0	0	-1.77	-1.77	0	0	0	0.46
10	0.25	-1.82	-0.1	-0.14	-1.82	0	-0.08	-0.04	-0.08
12	-1.96	0.3	-2.1	-1.61	-0.3	0.27	-0.62	-1.98	-1.9

15	-1.89	0.13	-1.87	0.28	0.21	-1.97	0.18	-1.62	0.12
18	0.28	0	-0.29	-0.29	0.29	0.12	-0.04	-0.29	-0.29
19	0.21	-0.08	0.2	0.22	0	-0.08	0.25	-0.04	-0.08

Chromosome 13

ID	D13S787	D13S1493	D13S894	D13S788	D13S800	D13S317	D13S793	D13S779	D13S796	D13S285
	-2.66	-4.59	-1.21	-6.81	-4.89	-6.7	-2.8	2.02	-2.43	-3.52
MD	0.29	-0.08	-0.05	-1.68	-1.68	-1.68	0.05	0.46	0.3	0
6	-1.77	-0.9	0	0.11	-1.77	-1.77	-1.77	0.55	0.3	0.25
10	0.21	-1.82	-0.08	0	-0.12	-1.82	0.27	-0.08	-0.04	0
12	0.15	0	0.6	-3.81	-1.72	-1.92	0.17	0.46	-3.78	-2.11
15	-1.68	-1.64	-1.6	-1.64	0.44	0.11	0.3	0.12	0.3	-1.84
18	-0.07	-0.1	-0.29	-0.05	-0.27	0.15	-1.82	0.28	0.24	-0.08
19	0.21	-0.05	0.21	0.26	0.23	0.23	0	0.23	0.25	0.26

Chromosome 14

ID	D14S742	D14S1280	D14S608	D14S599	D14S306	D14S587	D14S592	D14S588	D14S53	D14S606	D14S617	GATA168F06	GATA136B01
	-3.48	-2.83	-6.15	-5.02	-5.49	-4.02	-3.68	-6.5	-5.56	-1.99	-4.81	-4.46	-4.26
MD	-1.97	-1.66	0.22	0.6	0.52	0.6	0	-1.59	-0.29	0.44	-1.76	0.22	0.4
6	0	-1.77	-1.77	-1.77	-1.77	-1.77	-1.77	-1.77	-1.77	0.4	0	-1.77	-1.21
10	0.23	-0.1	-0.1	-1.82	-0.04	-0.08	0	-0.1	-1.82	-0.11	0.24	0.16	0.26
12	0.15	0.78	-1.77	0.48	-0.23	0.37	0.6	0.6	0.11	-0.15	-1.55	-1.79	-1.74
15	-1.97	0.06	-2.68	-2.68	-3.89	-2.68	-2.51	-1.72	-1.9	-2.39	-1.64	-1.64	-1.68
18	-0.14	-0.06	-0.05	0.25	0	-0.46	0	-1.82	-0.1	-0.1	-0.02	0.18	-0.29
19	0.22	-0.08	0	-0.08	-0.08	0	0	-0.1	0.21	-0.08	-0.08	0.18	0

Chromosome 15

ID	D15S822	D15S165	ACTC	D15S659	D15S643	GATA151F03	D15S655	D15S652	D15S816	D15S657	D15S642
	-1.94	0.48	-0.93	-1.23	-4.65	-6.96	-6.21	-1.25	-1.94	-3.64	-4.08
MD	-1.54	0.23	-1.58	0.23	0.27	-1.68	-1.68	-1.77	-1.57	-1.87	-1.6
6	-1.77	0	0	0.57	0.48	0.3	0.47	0	0.41	0.15	-0.6
10	-0.04	-0.04	0.25	-0.08	-0.08	-0.14	0.19	0.21	-0.1	-0.14	0.21
12	0.89	0.13	0.21	-1.72	-1.63	-1.76	-3.73	0.09	-0.29	-2.32	-2.1
15	0.6	0.12	0.19	-0.11	-3.89	-2.05	-1.83	0	0	0.6	0.38
18	0	0.12	0	-0.04	0.28	0.19	0.17	0	-0.29	-0.29	-0.29
19	-0.08	-0.08	0	-0.08	-0.08	-1.82	0.2	0.22	-0.1	0.23	-0.08

Chromosome 16

ID	ATA41E04	D16S748	D16S403	D16S769	D16S753	ATA55A11	D16S3253	GATA67G11	D16S2624	D16S402	D16S539	D16S621
	-3.69	-2.5	-4.39	-6.84	-3.25	-2.18	-1.56	-0.64	-1.27	-3.23	-0.89	-2.1
MD	-1.83	0.52	0.14	-3.77	-1.68	-1.83	0.52	-1.69	-1.68	-1.84	0.22	0.3
6	0	0.3	0	-1.77	0	0	-1.77	0.11	-0.28	0	0.23	0.48
10	0.24	-0.08	-1.82	0.24	0.21	0.22	0.25	-0.08	0	0.18	0.2	0.24
12	-0.71	-1.62	-1.57	-1.59	0.17	-1.17	-1.54	0.27	0.19	0.22	-1.7	-1.47
15	-1.35	-1.82	-1.48	0	-1.73	0.37	0.8	0.37	0.37	-1.8	0.06	-1.81
18	0	-0.02	0.18	0.05	-0.08	0	-0.08	0.15	-0.08	0.11	0.26	0.24
19	-0.04	0.22	0.16	0	-0.14	0.23	0.26	0.23	0.21	-0.1	-0.16	-0.08

Chromosome 17

ID	D17S1308	D17S1298	D17S974	D17S1303	D17S1294	D17S1293	D17S1299	ATC6A06	D17S1290	ATA43A10	D17S1301	D17S784
	-1.27	-1.18	-5.1	-4.58	-1.65	-4.77	-3.42	-3.16	-3.52	-6.24	-3.34	-3.78
MD	0.36	0.12	0.24	0.15	-1.72	0	0	-1.73	0	0.34	0.22	-1.68
6	-0.35	0.51	-1.77	-1.77	0	-1.77	0.13	-1.77	-1.77	-1.77	-1.77	-1.77
10	0.17	0.19	0.21	0.21	-0.14	0.24	0	-0.25	0.22	-0.08	-0.14	-0.04
12	-1.68	0.08	-1.96	-1.84	0.03	0.4	-1.82	0.3	-0.13	-1.97	0.31	-0.45
15	0.23	-1.97	0	-1.77	0	-3.89	-1.87	0.06	-1.97	-2.68	-1.9	0.24
18	0	0	0	0.22	0.26	0	-0.08	0.08	-0.09	-0.29	-0.29	0
19	0	-0.11	-1.82	0.22	-0.08	0.25	0.22	0.15	0.22	0.21	0.23	-0.08

Chromosome 18

SMAD4 lies between D18S851 and D18S858, shown in blue.

ID	GATA178F11	D18S976	D18S843	D18S542	D18S877	D18S535	D18S851	D18S858	ATA7D07	GATA7E12	ATA82B02	GATA177C03	D18S844
	-7.28	-8.9	-3.89	-5.72	-2.57	-2.88	-4.58	-6.15	-7.85	-6.43	-3.61	-1.78	-5.78
MD	-1.75	-1.81	-1.7	0	-0.04	-2	-2.02	-1.83	-1.86	0.16	-0.11	0.17	-1.8
6	0	0.44	-0.46	-1.83	0.04	-0.21	-0.66	-0.06	0	-0.8	-1.8	0.27	0
10	-1.89	-1.89	-0.17	0	-1.89	-0.08	-0.14	-0.11	0.26	0.25	0.24	-0.14	0.24
12	0.29	-1.89	0.24	-1.89	0.9	-1.7	-2.04	-2.45	-2.14	-1.93	-0.08	0	-0.53
15	-3.96	-3.96	-2	-1.9	-1.92	0.6	0.3	-1.87	-2.39	-1.92	-1.75	-0.05	-1.7
18	0.14	0	0.2	-0.02	0.15	0.27	0.12	0.25	-1.89	-0.3	-0.03	-1.89	-1.89
19	-0.11	0.21	0	-0.08	0.19	0.24	-0.14	-0.08	0.17	-1.89	-0.08	-0.14	-0.1

Chromosome 19

ID	D19S591	D19S1034	D19S586	D19S714	D19S433	D19S245	D19S178	D19S246	D19S589	D19S254
	0.85	-2.82	-3.03	-2.71	-2.2	0.73	-2.49	-1.87	-0.06	-4.09

MD	0	-1.18	-2.65	0.44	-0.05	0.26	0.42	0	0.17	-1.2
6	-0.23	-0.43	0	-1.19	-1.2	0.58	-1.2	-1.2	-1.2	0
10	0.22	-0.06	-0.09	-0.06	-0.03	-0.7	-0.07	-0.02	-0.06	-0.09
12	0.68	0.3	-0.18	-1.58	-1.05	0.59	-0.07	0.39	0.46	0.48
15	0.24	-1.43	0.59	0.18	0.59	0	-1.45	-1.19	0.29	-2.68
18	-0.06	-0.22	0	0.2	0.24	0	0	0.24	0.1	0.1
19	0	0.2	-0.7	-0.7	-0.7	0	-0.12	-0.09	0.18	-0.7

Chromosome 20

ID	D20S103	D20S482	D20S604	D20S470	D20S477	D20S478	D20S481	D20S480	D20S171
	-1.29	-3.17	0.57	-0.24	-1.1	-1.04	-0.41	-1.72	-0.4
MD	-1.17	0.1	0.07	0	-1.18	-0.1	0	-1.34	0.23
6	0.18	-1.2	0	-1.2	0	-1.2	-1.19	-1.2	0.07
10	-0.06	-0.7	0.2	0	0.2	-0.11	0.19	0.19	0.2
12	0.02	0.13	0.29	0.47	-0.12	-0.08	0.29	0.32	0.52
15	0.05	-1.17	0.29	0.29	0	0.03	0.14	0.35	-1.31
18	-0.22	-0.22	-0.22	0	0	0.22	0	-0.22	0
19	-0.09	-0.11	-0.06	0.2	0	0.2	0.16	0.18	-0.11

Chromosome 21

ID	D21S1432	D21S1437	GATA129D11	D21S1440	GATA188F04	D21S1446
	0.26	-4.82	-5.95	-3.06	-2.85	-3.11
MD	0	-1.83	-1.83	0.25	0.6	0.33
6	0	-1.8	-1.8	-1.8	0	0.57
10	-0.14	-1.89	-0.11	-0.14	-0.04	-0.08
12	-0.07	-0.17	-0.17	0.44	-1.61	0.6
15	0.42	0.6	-2	-2.07	-2	-2.6
18	0.19	0.27	0	0.26	0.24	-1.89
19	-0.14	0	-0.04	0	-0.04	-0.04

Chromosome 22

ID	D22S345	D22S689	D22S685	D22S683
	-0.12	-0.37	-0.37	-2.1
MD	0	0	0.19	-0.94
6	0	0	-0.57	0.5
10	-0.09	0	-0.05	-0.7
12	0.2	0.16	0.04	-1.02
15	0.08	-0.2	0.29	0.15
18	-0.22	-0.22	-0.22	0
19	-0.09	-0.11	-0.05	-0.09

APPENDIX TWO

GENOME WIDE TWO POINT LOD SCORES

FOR HMPS

GENOME WIDE TWO POINT LOD SCORES FOR HMPS

Shown are the two point LOD scores for all genome screen markers (Weber9 set, Research Genetics) for both affection statuses (Q1 and HMPS) in descending order down the chromosome (p-arm through to q-arm). Two-point scores are shown for θ=0, 0.1 .0.2, 0.3 and 0.4. The only area of the genome that provided significant evidence of linkage was 15q13-14 (highlighted in red).

Chromosome	1	Q1				Chromosome	1	HMPS			
D1S1612	-2.48	-1.1	-0.53	-0.23	-0.07	D1S1612	-1.12	-0.62	-0.23	-0.06	0.00
D1S1597	0.59	0.43	0.42	0.36	0.22	D1S1597	-0.55	-0.42	-0.2	-0.07	-0.01
D1S552	-3.99	-1.37	-0.74	-0.3	-0.07	D1S552	-1.11	-0.72	-0.25	-0.03	0.04
GATA129H04	-1.58	-0.4	-0.02	0.09	0.06	GATA129H04	-0.57	0.08	0.15	0.12	0.06
D1S2134	-1.88	-0.55	-0.27	-0.09	0.01	D1S2134	-1.09	-0.69	-0.39	-0.17	-0.04
GATA165C03	-1.51	-0.31	-0.15	-0.08	-0.04	GATA165C03	0.08	0.05	0.02	0.01	0.00
D1S1665	-2.51	-0.87	-0.42	-0.17	-0.05	D1S1665	-1.32	-0.72	-0.35	-0.15	-0.04
D1S1728	-1.8	-0.53	-0.19	-0.02	0.03	D1S1728	-1.15	-0.57	-0.22	-0.06	0.00
D1S551	-1.69	-0.56	-0.18	-0.02	-0.00	D1S551	-1.41	-0.56	-0.22	-0.08	-0.04
D1S1588	-1.54	-0.41	-0.14	-0.02	0.02	D1S1588	0.69	0.49	0.32	0.19	0.09
GATA176G01	-1.97	-0.8	-0.39	-0.08	0.04	GATA176G01	-2.72	-0.84	-0.3	-0.08	0.00
D1S1653	-1.91	-0.89	-0.37	-0.12	-0.02	D1S1653	-0.85	0.14	0.19	0.13	0.06
D1S1679	-4.65	-0.96	-0.37	-0.12	-0.01	D1S1679	-2.64	-0.66	-0.28	-0.11	-0.03
D1S1677	-3.83	-0.84	-0.31	-0.11	-0.03	D1S1677	-1.37	-0.75	-0.31	-0.1	-0.01
D1S1589	-1.24	-0.5	-0.12	0.02	0.04	D1S1589	-0.11	-0.2	-0.17	-0.1	-0.03
D1S518	-2.04	-0.03	0.19	0.2	0.13	D1S518	-1.28	-0.41	-0.2	-0.09	-0.03
D1S1660	0.33	0.38	0.35	0.18	0.02	D1S1660	1.01	0.76	0.5	0.26	0.07
D1S1678	-0.04	-0.13	-0.04	0.05	0.06	D1S1678	-1.18	-0.16	-0.03	0	0.00
GATA124F08	-0.16	-0.27	-0.19	-0.1	-0.04	GATA124F08	-0.96	-0.33	-0.19	-0.11	-0.05
D1S2141	-1.66	-0.38	-0.07	0.03	0.04	D1S2141	-0.65	-0.23	0.03	0.11	0.09
D1S549	-3.89	-1.25	-0.55	-0.17	-0.01	D1S549	-1.36	-0.57	-0.27	-0.12	-0.03
D1S3462	-4.54	-1.57	-0.86	-0.46	-0.19	D1S3462	-1.5	-0.69	-0.45	-0.24	-0.08
D1S235	-4.79	-1.54	-0.72	-0.32	-0.11	D1S235	-1.19	-0.7	-0.35	-0.15	-0.04
D1S1609	-1	-0.66	-0.33	-0.1	0.00	D1S1609	-1.01	-0.43	-0.13	-0.01	0.02
Chromosome	**2**	**Q1**				**Chromosome**	**2**	**HMPS**			

GATA165C07	-3.91	-1.65	-0.83	-0.4	-0.14	GATA165C07	0.44	0.23	0.09	0.01	-0.01
GATA116B01	-5.64	-1.94	-0.89	-0.36	-0.11	GATA116B01	-1.85	-0.9	-0.66	-0.38	-0.15
D2S1400	0.42	0.27	0.18	0.05	-0.03	D2S1400	-0.18	0.11	0.1	0.04	0.00
D2S405	-1.21	-0.25	-0.12	-0.06	-0.02	D2S405	-0.09	-0.12	-0.08	-0.04	-0.00
D2S1356	-2.47	-1.03	-0.65	-0.38	-0.18	D2S1356	0.74	0.54	0.37	0.22	0.10
D2S1352	-1.15	-0.43	-0.18	-0.1	-0.06	D2S1352	0.55	0.4	0.25	0.12	0.03
D2S441	-0.62	-0.24	-0.09	-0.04	-0.01	D2S441	-4.33	-1.17	-0.68	-0.44	-0.21
D2S1394	1.21	0.84	0.47	0.15	-0.02	D2S1394	-1.95	-0.28	0.02	0.01	-0.05
D2S1777	-0.26	-0.1	-0.05	-0.04	-0.03	D2S1777	-2.66	-1.27	-0.74	-0.4	-0.17
D2S1790	-0.26	-0.33	-0.29	-0.18	-0.07	D2S1790	-0.32	-0.12	0.1	0.13	0.07
GATA176C01	0.14	0.05	0.03	0.03	0.02	GATA176C01	0.85	0.52	0.27	0.1	0.02
D2S410	-3.02	-1.02	-0.43	-0.16	-0.04	D2S410	-4.89	-1.53	-0.71	-0.3	-0.09
D2S1328	-0.92	-0.16	-0.01	0.03	0.03	D2S1328	-2.25	-0.98	-0.42	-0.15	-0.02
D2S442	-1.23	-0.22	-0.08	-0.03	-0.01	D2S442	-1.83	-0.04	0.11	0.07	0.01
D2S1399	-0.89	-0.31	-0.03	0.07	0.07	D2S1399	-3.83	-1.17	-0.47	-0.12	0.02
D2S1353	0.57	0.36	0.22	0.13	0.06	D2S1353	-1.76	-0.33	-0.03	0.02	0.01
D2S1776	-0.68	0.26	0.26	0.17	0.07	D2S1776	-3.78	-0.88	-0.27	0	0.06
D2S1391	0.42	0.3	0.19	0.09	0.02	D2S1391	0.81	0.53	0.29	0.11	0.01
D2S1384	-2.82	-0.67	-0.26	-0.08	-0.01	D2S1384	-2	-1.27	-0.69	-0.32	-0.11
GATA30E06	0.11	0.09	0.06	0.02	-0.00	GATA30E06	-2.32	-0.36	-0.12	-0.07	-0.05
D2S434	-2.89	-0.94	-0.44	-0.19	-0.06	D2S434	-2.15	-0.46	-0.1	-0.04	-0.06
GATA178G09	0.01	-0.01	0.01	0.03	0.02	GATA178G09	-2.85	-0.18	-0.04	-0.15	-0.18
D2S125	-1.26	-0.46	-0.12	0	0.03	D2S125	-3.56	-1.16	-0.81	-0.51	-0.22

Chromosome	**3**	**Q1**				**Chromosome**	**3**	**HMPS**			
D3S2387	-3	-1.34	-0.69	-0.29	-0.08	D3S2387	-1	-0.2	0.02	0.06	0.04
D3S1304	-1.24	-0.36	-0.02	0.03	0.00	D3S1304	-1.35	-0.65	-0.27	-0.1	-0.02
GATA164B08	-2.01	-1.26	-0.64	-0.29	-0.08	GATA164B08	-1.27	-0.52	-0.18	-0.05	-0.00
D3S1259	-2.55	-0.66	-0.1	0.1	0.11	D3S1259	-0.95	-0.65	-0.22	-0.01	0.04
D3S2432	-2.86	-0.7	-0.19	0.01	0.05	D3S2432	0.94	0.67	0.43	0.23	0.09
D3S1768	-0.85	0.01	0.05	-0.02	-0.08	D3S1768	0.62	0.42	0.26	0.14	0.05
D3S1766	-0.98	-0.54	-0.2	-0.05	-0.01	D3S1766	0.8	0.62	0.42	0.23	0.08
D3S2406	-4.64	-2.33	-1.45	-0.85	-0.37	D3S2406	-1.23	-0.39	-0.2	-0.1	-0.04
D3S2459	-5.25	-1.79	-0.94	-0.49	-0.18	D3S2459	-1.09	-0.05	0.1	0.11	0.06
D3S3045	-2.82	-1.14	-0.56	-0.25	-0.09	D3S3045	0.51	0.31	0.16	0.06	0.02
D3S2460	-3.62	-1.38	-0.64	-0.27	-0.09	D3S2460	-0.73	0.12	0.19	0.14	0.06
ATA34G06	-1.1	-0.02	0.03	0.01	0.00	ATA34G06	0.85	0.6	0.38	0.19	0.06
D3S1764	-3.87	-1.51	-0.68	-0.27	-0.08	D3S1764	-1.36	-0.26	-0.02	0.05	0.04
D3S1744	-5.38	-1.62	-0.61	-0.17	-0.02	D3S1744	-0.35	0.54	0.48	0.3	0.11
D3S1763	-3.51	-1.56	-0.79	-0.35	-0.11	D3S1763	-1.71	-0.74	-0.37	-0.19	-0.08
D3S2427	-4.34	-0.95	-0.18	0.09	0.11	D3S2427	-2.57	-0.39	-0.07	0.03	0.03
D3S1262	-1.88	-0.51	-0.2	-0.09	-0.04	D3S1262	-5.63	-1.08	-0.25	0.04	0.07
D3S2398	-1.43	-0.74	-0.37	-0.16	-0.04	D3S2398	-2.76	-1.16	-0.6	-0.3	-0.11
D3S2418	-0.62	0.37	0.34	0.21	0.07	D3S2418	-3.06	-1.14	-0.5	-0.2	-0.06
D3S1311	-1.12	-0.67	-0.33	-0.14	-0.04	D3S1311	-6.7	-2.3	-1.26	-0.59	-0.19

Chromosome	**4**	**Q1**	**Chromosome**	**4**	**HMPS**

D4S2366	-1.08	0.05	0.23	0.27	0.16	D4S2366	-1	-0.43	-0.1	0.03	0.04
D4S403	-2.53	-0.92	-0.41	-0.19	-0.08	D4S403	-0.43	0.45	0.43	0.27	0.08
D4S2639	-4.01	-1.35	-0.72	-0.3	-0.08	D4S2639	-1.04	-0.21	0.07	0.12	0.07
D4S2632	-4.01	-1.33	-0.78	-0.48	-0.23	D4S2632	-2.23	-0.14	0.12	0.12	0.03
D4S1627	-1.92	-0.94	-0.35	-0.08	-0.01	D4S1627	-1.88	-0.13	0.11	0.11	0.02
D4S3248	-1.96	-0.33	0.1	0.18	0.10	D4S3248	1.06	0.78	0.54	0.31	0.10
D4S2367	-0.53	-0.06	0.11	0.1	0.04	D4S2367	-0.45	0.11	0.16	0.1	0.03
D4S2361	-1.74	-0.53	-0.17	0.01	0.04	D4S2361	-1.3	-0.62	-0.24	-0.08	-0.02
D4S1647	-4.18	-1.42	-0.57	-0.16	-0.00	D4S1647	-2.73	-1.04	-0.47	-0.21	-0.08
D4S2623	-3.74	-1.02	-0.44	-0.17	-0.05	D4S2623	-0.46	0.55	0.49	0.32	0.13
D4S2394	-4.35	-1.04	-0.47	-0.21	-0.07	D4S2394	0.35	0.25	0.16	0.1	0.05
D4S1644	-0.33	0.06	0.09	0.06	0.03	D4S1644	0.13	0.07	0.03	0.01	-0.00
D4S1625	-1.68	-0.38	-0.11	0	0.02	D4S1625	-0.63	0.19	0.21	0.13	0.05
D4S1629	-0.67	-0.7	-0.48	-0.23	-0.08	D4S1629	-1.12	-0.57	-0.37	-0.2	-0.07
D4S2368	-2.1	-0.65	-0.3	-0.12	-0.03	D4S2368	-1.03	-0.24	-0.12	-0.07	-0.04
D4S408	-2.04	-0.79	-0.37	-0.12	-0.01	D4S408	-1.35	-0.5	-0.17	-0.04	0.01
D4S1652	-1.1	-0.71	-0.32	-0.12	-0.04	D4S1652	-1.42	-0.72	-0.34	-0.14	-0.04

Chromosome	**5**	**Q1**				**Chromosome**	**5**	**HMPS**			
D5S2488	-2.6	-0.35	0	0.06	0.02	D5S2488	-1.43	-0.32	-0.02	0.05	0.03
D5S2505	-0.75	-0.78	-0.38	-0.1	0.01	D5S2505	-2.77	-1	-0.43	-0.2	-0.09
D5S807	-0.4	0.04	0.16	0.13	0.06	D5S807	-0.18	0.09	0.11	0.08	0.03
GATA134B03	-2.71	-0.37	0.03	0.15	0.12	GATA134B03	-1.19	-0.26	-0.08	-0.02	-0.00
D5S1470	0.88	1.75	1.42	0.93	0.38	D5S1470	-0.33	0.53	0.49	0.32	0.12
D5S1457	-3.15	-1.22	-0.73	-0.39	-0.15	D5S1457	0.03	0.19	0.16	0.08	0.02
D5S2500	1.45	1.13	0.8	0.46	0.15	D5S2500	-2.17	-0.68	-0.26	-0.12	-0.07
D5S1501	-2.7	-1.12	-0.56	-0.25	-0.08	D5S1501	-4.5	-1.42	-1.03	-0.62	-0.27
GATA89G08	-1.04	-0.3	-0.11	-0.03	0.00	GATA89G08	-0.04	0.21	0.25	0.17	0.06
D5S1453	-1.52	-0.41	-0.32	-0.23	-0.10	D5S1453	-4.41	-1.4	-0.81	-0.48	-0.21
D5S1505	0.29	0.2	0.13	0.07	0.03	D5S1505	-1.15	-0.74	-0.4	-0.2	-0.07
D5S816	-2.55	-0.9	-0.44	-0.2	-0.06	D5S816	-4.07	-1.59	-1.05	-0.74	-0.37
D5S1480	-0.37	0.02	0.01	-0.04	-0.04	D5S1480	-1.63	-0.07	0.1	-0.01	-0.11
D5S820	-0.33	0.36	0.34	0.21	0.06	D5S820	-0.03	0.46	0.36	0.19	0.04
D5S1471	0.16	0.04	-0.06	-0.08	-0.04	D5S1471	-2.53	0.06	0.25	0.04	-0.17
D5S1456	-0.43	0.2	0.22	0.14	0.04	D5S1456	-2.02	-0.22	0.01	0.02	-0.02
D5S408	-2.19	-1.3	-0.73	-0.37	-0.14	D5S408	-5.78	-2.1	-1.25	-0.65	-0.23

Chromosome	**6**	**Q1**				**Chromosome**	**6**	**HMPS**			
F13A1	-2.34	-0.59	-0.21	-0.08	-0.04	F13A1	-1.25	-0.49	-0.2	-0.06	-0.01
D6S1959	-1.99	-0.66	-0.32	-0.11	-0.02	D6S1959	-1.55	-0.73	-0.3	-0.1	-0.02
GATA163B10	-1.57	-0.71	-0.37	-0.18	-0.07	GATA163B10	-0.76	0.32	0.3	0.18	0.04
GGAA15B08	-0.54	-0.65	-0.5	-0.23	-0.06	GGAA15B08	-0.93	0.04	0.09	0.05	0.01
D6S1017	-4.31	-1.68	-0.93	-0.48	-0.17	D6S1017	-1.74	-0.84	-0.44	-0.2	-0.07
D6S2410	-2.11	-0.65	-0.27	-0.11	-0.02	D6S2410	-1.04	-0.25	-0.12	-0.07	-0.03
D6S1053	-2.2	-0.65	-0.33	-0.17	-0.06	D6S1053	-1.32	-0.5	-0.28	-0.15	-0.07
D6S1031	-2.76	-1.13	-0.59	-0.28	-0.10	D6S1031	-1.42	-0.66	-0.33	-0.15	-0.06
D6S1056	-1.83	0.2	0.27	0.11	-0.03	D6S1056	0.82	0.52	0.24	0.05	-0.01

D6S1021	-2.4	-0.48	-0.45	-0.44	-0.24	D6S1021	-1.11	-0.79	-0.55	-0.3	-0.12
D6S474	0.55	1.17	0.99	0.61	0.19	D6S474	1.66	1.28	0.88	0.47	0.14
D6S1040	0.35	0.57	0.54	0.32	0.07	D6S1040	-1.02	0.03	0.03	-0.02	-0.03
D6S1009	-0.34	-0.08	0.01	0.02	0.00	D6S1009	-1.25	-0.52	-0.2	-0.07	-0.02
GATA184A08	-3.62	-0.94	-0.39	-0.1	0.02	GATA184A08	0.73	0.56	0.38	0.22	0.1
GATA165G02	-0.44	0.52	0.44	0.25	0.07	GATA165G02	-1.18	-0.04	0.05	0.04	0.01
D6S305	-0.24	0.01	0.08	0.05	0.03	D6S305	-1.2	-0.39	-0.2	-0.1	-0.04
D6S1277	-1.91	-0.5	0	0.12	0.09	D6S1277	-2.81	-0.68	-0.3	-0.14	-0.06
D6S1027	-3.47	-0.98	-0.41	-0.15	-0.03	D6S1027	-1.57	-0.76	-0.4	-0.21	-0.09
Chromosome	**7**	**Q1**				**Chromosome**	**7**	**HMPS**			
GATA24F03	-0.48	0.27	0.27	0.18	0.08	GATA24F03	0.33	0.33	0.27	0.16	0.05
GATA137H02	-0.78	-0.42	-0.21	-0.11	-0.07	GATA137H02	-0.08	-0.07	-0.05	-0.02	-0.01
D7S1802	-3.51	-1.11	-0.49	-0.19	-0.06	D7S1802	-3.32	-1.26	-0.59	-0.26	-0.09
D7S1808	-2.68	-0.86	-0.32	-0.11	-0.03	D7S1808	-1.95	-0.58	-0.21	-0.07	-0.01
D7S817	-4.03	-0.98	-0.37	-0.11	-0.01	D7S817	-1.34	-0.15	-0.07	-0.05	-0.02
D7S2846	0.13	0.39	0.27	0.14	0.05	D7S2846	-1.54	-0.31	-0.13	-0.06	-0.02
D7S1818	-0.73	-0.52	-0.24	-0.08	-0.02	D7S1818	-0.15	-0.12	-0.08	-0.04	-0.01
GATA118G10	-4.29	-1.53	-0.81	-0.39	-0.13	GATA118G10	-1.49	-0.5	-0.31	-0.17	-0.06
D7S820	-0.64	0.05	-0.01	-0.11	-0.10	D7S820	-1.68	-0.51	-0.23	-0.09	-0.02
D7S821	-2.17	-0.39	-0.02	0.07	0.04	D7S821	-1.1	0.11	0.15	0.1	0.04
D7S1799	-4.79	-1.66	-0.75	-0.3	-0.08	D7S1799	-1.77	-0.35	-0.03	0.05	0.04
GGAA6D03	-2.2	-0.77	-0.33	-0.11	-0.01	GGAA6D03	-1.19	0.04	0.11	0.08	0.03
D7S1804	-1.56	0.54	0.69	0.53	0.26	D7S1804	-0.77	0.43	0.39	0.25	0.11
D7S1824	-2.2	-0.26	-0.03	0.03	0.03	D7S1824	-1.49	-0.24	-0.05	0.02	0.03
GATA189C06	-1.53	-0.33	-0.01	0.17	0.16	GATA189C06	-1.79	-0.76	-0.24	-0.03	0.04
GATA30D09	-2.08	-0.66	-0.26	-0.06	0.02	GATA30D09	-1.82	-0.7	-0.3	-0.11	-0.02
D7S559	-2.34	-0.94	-0.47	-0.21	-0.06	D7S559	-1.48	-0.67	-0.4	-0.2	-0.07
Chromosome	**8**	**Q1**				**Chromosome**	**8**	**HMPS**			
D8S264	-1.56	-0.22	0.08	0.11	0.05	D8S264	-0.53	0.11	0.15	0.09	0.03
D8S1130	0.28	0.98	0.77	0.44	0.14	D8S1130	0.73	0.5	0.29	0.13	0.02
D8S1106	0.2	0.25	0.19	0.1	0.03	D8S1106	-0.24	-0.06	0	0.01	0.01
D8S1145	0.45	0.81	0.66	0.42	0.19	D8S1145	0.72	0.56	0.39	0.24	0.10
D8S136	-4.04	-1.16	-0.51	-0.17	-0.03	D8S136	-1.09	-0.48	-0.15	-0.01	0.03
D8S1477	-5.85	-1.28	-0.49	-0.15	-0.01	D8S1477	-1.1	-0.07	0.05	0.07	0.05
D8S1110	-3.51	-0.63	-0.08	0.11	0.12	D8S1110	-1.11	-0.3	-0.02	0.07	0.07
D8S1113	-2.21	-0.42	-0.03	0.08	0.06	D8S1113	-1.36	-0.45	-0.13	-0.01	0.01
D8S1136	-2.06	-0.64	-0.36	-0.17	-0.08	D8S1136	-1.25	-0.58	-0.24	-0.1	-0.04
GATA14E09	-3.73	-1.17	-0.54	-0.21	-0.06	GATA14E09	-0.73	0.28	0.28	0.18	0.08
D8S1119	-2.14	-0.82	-0.43	-0.14	-0.00	D8S1119	-1.32	-0.54	-0.21	-0.07	-0.01
GAAT1A4	-1.15	-0.44	-0.22	-0.09	-0.01	GAAT1A4	-1.31	-0.48	-0.23	-0.09	-0.02
D8S1132	-4.82	-1.79	-1.08	-0.62	-0.26	D8S1132	-2.88	-1.41	-0.75	-0.37	-0.14
D8S592	-2.29	-0.77	-0.31	-0.07	0.03	D8S592	-2.76	-1.24	-0.62	-0.27	-0.08
D8S1128	-1.2	-0.67	-0.25	-0.04	0.03	D8S1128	-1.22	-0.12	0.1	0.23	0.18
D8S256	-2.71	-1.06	-0.48	-0.2	-0.06	D8S256	-6.07	-2.16	-1.14	-0.54	-0.19
D8S373	-0.88	0.1	0.1	0.03	0.01	D8S373	-3.62	-1.15	-0.46	-0.14	-0.02

APPENDIX TWO

Chromosome	9	Q1				Chromosome	9	HMPS			
GATA62F03	-0.7	-0.49	-0.21	-0.08	-0.03	GATA62F03	-1.61	-0.56	-0.27	-0.12	-0.04
D9S925	-3.56	-0.9	-0.36	-0.07	0.05	D9S925	-1.5	-0.32	0	0.09	0.08
D9S1121	-2.16	-0.47	-0.19	-0.07	-0.02	D9S1121	0.04	0.02	0	-0.01	-0.01
D9S1118	-2.42	-0.78	-0.32	-0.09	0.00	D9S1118	0.33	0.2	0.1	0.04	0.01
D9S301	-0.44	-0.46	-0.32	-0.16	-0.05	D9S301	-1.32	-0.14	0.05	0.1	0.08
D9S1122	0.5	0.6	0.45	0.25	0.08	D9S1122	0.26	0.35	0.28	0.16	0.06
D9S922	-2	-0.99	-0.59	-0.32	-0.13	D9S922	-1.49	-0.8	-0.44	-0.21	-0.07
D9S910	-1.67	-0.66	-0.29	-0.1	-0.02	D9S910	-2.26	-0.57	-0.2	-0.06	-0.02
D9S938	-1.89	-0.53	-0.11	0.02	0.02	D9S938	-1.72	-0.53	-0.21	-0.08	-0.01
D9S930	-4.53	-1.36	-0.81	-0.56	-0.29	D9S930	-2.29	-1.06	-0.66	-0.37	-0.14
D9S934	-2.5	-0.75	-0.35	-0.17	-0.06	D9S934	-2.03	-0.84	-0.4	-0.17	-0.05
ATA59H06	-4.3	-0.94	-0.36	-0.12	-0.01	ATA59H06	-1.13	-0.03	0.07	0.06	0.03

Chromosome	10	Q1				Chromosome	10	HMPS			
D10S189	-0.63	-0.13	-0.06	-0.03	-0.00	D10S189	-1.43	-0.44	-0.2	-0.09	-0.02
D10S1412	-3.58	-0.95	-0.49	-0.28	-0.12	D10S1412	-1.23	-0.16	-0.1	-0.06	-0.02
D10S1423	-0.23	-0.11	0.05	0.11	0.09	D10S1423	-1.18	-0.49	-0.24	-0.11	-0.03
D10S1426	-2.35	-0.29	-0.08	-0.05	-0.04	D10S1426	0.62	0.34	0.11	-0.01	-0.03
D10S1208	-1.49	-0.27	-0.09	-0.02	-0.00	D10S1208	0.16	0.1	0.06	0.03	0.01
D10S1221	-4.28	-1.49	-0.74	-0.29	-0.07	D10S1221	-1.4	-0.92	-0.41	-0.15	-0.03
D10S1225	-0.15	0.7	0.52	0.29	0.12	D10S1225	0.81	0.57	0.36	0.19	0.07
GATA121A08	0.48	0.37	0.28	0.2	0.11	GATA121A08	0.21	0.16	0.12	0.08	0.04
D10S2327	-5.87	-1.79	-0.92	-0.4	-0.11	D10S2327	-2.75	-0.94	-0.41	-0.16	-0.05
d10s2470	-2.42	-0.33	-0.1	-0.05	-0.04	d10s2470	-1.31	-0.37	-0.21	-0.13	-0.06
D10S1239	-3.96	-0.88	-0.31	-0.09	-0.00	D10S1239	-1.12	-0.14	0	0.03	0.02
D10S1230	-1.98	-0.25	0.08	0.09	0.00	D10S1230	-2.42	-0.39	-0.08	0	-0.00
D10S1248	-2.42	-0.33	-0.1	-0.05	-0.04	D10S1248	-0.2	0.32	0.25	0.12	0.03
D10S212	-3.96	-0.88	-0.31	-0.09	-0.00	D10S212	-0.04	0.85	0.74	0.48	0.18

Chromosome	11	Q1				Chromosome	11	HMPS			
D11S1984	-3.07	-0.39	0.13	0.22	0.13	D11S1984	-3.17	-1.31	-0.68	-0.35	-0.14
D11S1999	1.38	1.05	0.74	0.46	0.21	D11S1999	0.18	0.16	0.11	0.07	0.03
D11S1981	0.23	0.2	0.27	0.21	0.09	D11S1981	-1.35	-0.12	0.04	0.08	0.05
ATA34E08	1.38	1.9	1.44	0.84	0.26	ATA34E08	-0.77	0.42	0.34	0.17	0.04
D11S1392	-1.45	0.64	0.64	0.38	0.07	D11S1392	-0.97	0.08	0.13	0.08	0.02
D11S2371	-0.37	0.28	0.3	0.23	0.12	D11S2371	0.43	0.33	0.25	0.17	0.08
D11S1998	-1.45	-0.56	-0.23	-0.09	-0.03	D11S1998	-0.55	-0.24	-0.06	0	0.01
D11S4464	-1.71	-0.2	0.06	0.11	0.05	D11S4464	-1.78	-0.78	-0.33	-0.12	-0.03
d11s2359	-0.6	0.1	0.14	0.09	0.03	d11s2359	0.9	0.66	0.43	0.22	0.06

Chromosome	12	Q1				Chromosome	12	HMPS			
D12S372	-1.12	0.01	0.11	0.1	0.05	D12S372	0.29	0.15	0.07	0.02	-0.00
D12S1042	-2.32	-0.85	-0.36	-0.13	-0.03	D12S1042	0.18	0.21	0.16	0.09	0.03

GATA91H06	-3.08	-1.17	-0.52	-0.22	-0.08	GATA91H06	-2.88	-1.23	-0.6	-0.27	-0.09
D12S398	-2.63	-1.3	-0.54	-0.15	0.01	D12S398	-2.94	-1.39	-0.72	-0.33	-0.11
D12S1294	0.01	0.26	0.24	0.14	0.04	D12S1294	-0.27	0.04	0.08	0.05	0.01
D12S375	-1.63	-0.59	-0.28	-0.13	-0.04	D12S375	-1.28	-0.88	-0.39	-0.14	-0.02
D12S1052	-2.7	-1.2	-0.58	-0.25	-0.08	D12S1052	-0.98	-0.29	-0.17	-0.11	-0.05
D12S1064	-3.31	-1	-0.55	-0.32	-0.14	D12S1064	-2.52	-0.88	-0.5	-0.28	-0.12
D12S1300	-2.47	-0.92	-0.45	-0.15	-0.00	D12S1300	-1.23	-0.71	-0.23	-0.01	0.05
D12S2070	0.71	0.49	0.31	0.17	0.06	D12S2070	-0.02	-0.04	-0.02	0.01	0.01
D12S395	-0.33	0.07	0.18	0.16	0.09	D12S395	-1	-0.05	0.03	0.02	0.00
D12S2078	-2.48	-0.81	-0.3	-0.06	0.02	D12S2078	-1.22	-0.31	-0.03	0.05	0.04
D12S1045	0	-0.07	0.03	0.11	0.10	D12S1045	-1.36	-0.63	-0.27	-0.1	-0.02
Chromosome	**13**	**Q1**				**Chromosome**	**13**	**HMPS**			
D13S787	-0.06	0.58	0.45	0.27	0.10	D13S787	-0.29	-0.19	-0.13	-0.08	-0.04
D13S1493	-2.4	-0.95	-0.45	-0.2	-0.07	D13S1493	-0.37	0.12	0.14	0.09	0.03
D13S894	0.32	0.21	0.13	0.07	0.03	D13S894	0.14	0.1	0.07	0.04	0.02
D13S788	-3.66	-1.07	-0.57	-0.29	-0.11	D13S788	-0.88	-0.36	-0.13	-0.07	-0.04
D13S800	0.05	0	0.05	0.05	0.02	D13S800	-0.49	0.08	0.09	0.03	-0.00
D13S317	-1.79	-0.59	-0.23	-0.1	-0.04	D13S317	-0.36	-0.15	0.07	0.1	0.05
D13S793	0.24	0.16	0.1	0.06	0.03	D13S793	1.08	0.8	0.55	0.33	0.15
D13S285	-2.68	-0.62	-0.24	-0.1	-0.05	D13S285	-3.18	-0.06	0.15	0.03	-0.11
Chromosome	**14**	**Q1**				**Chromosome**	**14**	**HMPS**			
D14S742	-2.1	-0.76	-0.35	-0.12	-0.01	D14S742	-1.41	-0.69	-0.29	-0.1	-0.01
D14S1280	0.84	1.07	0.98	0.66	0.26	D14S1280	-1.03	0.06	0.16	0.11	0.03
D14S599	-2.92	-1.14	-0.58	-0.28	-0.10	D14S599	-1.18	-0.3	-0.14	-0.07	-0.03
D14S306	-0.11	0.66	0.56	0.34	0.10	D14S306	1.32	1	0.67	0.36	0.11
D14S587	-3.83	-1.38	-0.71	-0.34	-0.13	D14S587	-1.39	-0.42	-0.19	-0.1	-0.05
D14S592	-3.38	-0.64	-0.09	0.12	0.12	D14S592	-1.06	-0.41	-0.07	0.05	0.07
D14S588	-1.71	-0.66	-0.23	-0.06	-0.01	D14S588	-0.96	-0.32	-0.07	0.01	0.02
D14S53	-2.49	-0.2	0.14	0.15	0.03	D14S53	-1.51	-0.35	-0.06	0.02	0.01
D14S606	-1.33	0.09	0.22	0.16	0.04	D14S606	0	0.2	0.2	0.13	0.05
GATA193A07	-2.94	-1.37	-0.7	-0.32	-0.10	GATA193A07	-1.39	-0.53	-0.22	-0.07	-0.01
GATA168F06	-2.16	-0.53	-0.13	0	0.01	GATA168F06	-2.47	-0.44	-0.12	-0.02	-0.01
GATA136B01	-3.77	-0.97	-0.34	-0.06	0.02	GATA136B01	-1.5	-0.46	-0.15	-0.04	0
Chromosome	**15**	**Q1**				**Chromosome**	**15**	**HMPS**			
D15S822	-0.08	0.38	0.47	0.31	0.08	D15S822	-1.19	0.05	0.11	0.06	0.01
D15S165	1	1.51	1.32	0.93	0.45	D15S165	-0.7	0.55	0.52	0.35	0.15
ACTC	3.32	2.66	1.96	1.22	0.47	ACTC	1.76	1.4	1.02	0.61	0.22
D15S659	1.34	1.08	0.82	0.55	0.28	D15S659	0.01	-0.01	-0.01	-0.01	-0.00
D15S643	-3.27	-0.7	-0.22	0	0.05	D15S643	-1.08	-0.15	-0.09	-0.06	-0.03
GATA151F03	-2.23	-0.72	-0.33	-0.14	-0.05	GATA151F03	-3.4	-1.16	-0.53	-0.22	-0.07
D15S818	-3.96	-1.23	-0.61	-0.3	-0.12	D15S818	-3.06	-0.52	-0.16	-0.05	-0.02
D15S655	-0.67	-0.14	-0.03	0	0.01	D15S655	-0.05	-0.13	-0.12	-0.07	-0.02
D15S652	-1.65	-0.47	-0.25	-0.14	-0.06	D15S652	-0.09	-0.04	-0.01	0	0.01

D15S816	-1.37	-0.17	-0.01	0.03	0.01	D15S816	0.19	0.18	0.12	0.07	0.02
D15S657	-2.12	-0.54	-0.12	0.03	0.05	D15S657	-1.15	-0.03	0.06	0.05	0.02
D15S642	0.79	1.6	1.36	0.93	0.41	D15S642	1.29	1.05	0.75	0.44	0.16
Chromosome	**16**	**Q1**				**Chromosome**	**16**	**HMPS**			
ATA41E04	-2.02	-0.12	0.08	0.12	0.09	ATA41E04	0.35	0.25	0.18	0.11	0.05
D16S748	-2.3	-0.87	-0.47	-0.19	-0.04	D16S748	-1.35	-0.51	-0.2	-0.07	-0.01
D16S764	-4.19	-1.54	-0.76	-0.3	-0.06	D16S764	-1.4	-0.49	-0.16	-0.04	-0.00
D16S403	-1.91	-1	-0.32	-0.02	0.05	D16S403	-0.23	0.77	0.65	0.42	0.17
ATA55A11	-0.85	0.25	0.4	0.34	0.16	ATA55A11	-0.42	0.56	0.48	0.3	0.12
D16S3253	-2.88	-1.09	-0.54	-0.22	-0.05	D16S3253	-0.94	-0.09	-0.02	-0.02	-0.01
GATA67G11	-1.56	-0.09	0.11	0.11	0.05	GATA67G11	-0.43	0	0.07	0.05	0.02
D16S2624	-4.85	-1.73	-0.9	-0.44	-0.16	D16S2624	-2.05	-0.99	-0.5	-0.23	-0.07
D16S516	0.02	0.33	0.31	0.16	0.04	D16S516	-1.16	-0.53	-0.28	-0.14	-0.05
D16S621	-2.12	-0.93	-0.53	-0.29	-0.12	D16S621	-1.08	-0.76	-0.44	-0.21	-0.06
Chromosome	**17**	**Q1**				**Chromosome**	**17**	**HMPS**			
D17S1308	0.37	0.66	0.63	0.44	0.22	D17S1308	1.28	1	0.71	0.41	0.15
D17S1298	-1.76	-0.48	-0.27	-0.08	0.01	D17S1298	-1.56	-0.51	-0.13	0.01	0.04
D17S974	-1.56	-0.37	-0.19	-0.09	-0.03	D17S974	-0.12	-0.14	-0.1	-0.04	-0.01
D17S1303	-4.2	-1	-0.4	-0.13	-0.02	D17S1303	0.51	0.36	0.24	0.15	0.07
GATA185H04	-3.31	-0.76	-0.3	-0.11	-0.03	GATA185H04	-1.6	-0.32	-0.15	-0.08	-0.04
D17S1294	-0.08	0.65	0.51	0.29	0.07	D17S1294	-0.7	0.43	0.43	0.28	0.10
D17S1293	-0.59	0	0.14	0.12	0.04	D17S1293	0.37	0.28	0.17	0.08	0.02
D17S1299	-1.92	-0.6	-0.28	-0.13	-0.05	D17S1299	-1.46	-0.35	-0.09	-0.01	0.01
ATC6A06	-1.49	-0.52	-0.26	-0.12	-0.05	ATC6A06	-1.59	-0.38	-0.21	-0.12	-0.05
D17S1290	-3.66	-1.29	-0.81	-0.44	-0.16	D17S1290	-3.66	-0.96	-0.34	-0.09	0.01
ATA43A10	-3.32	-0.75	-0.35	-0.22	-0.14	ATA43A10	-2.87	-0.63	-0.31	-0.16	-0.05
D17S784	-2.22	-0.76	-0.38	-0.15	-0.03	D17S784	-1.71	-0.79	-0.32	-0.1	-0.00
D17S928	-1.64	-0.26	0.07	0.17	0.14	D17S928	-2.11	-1.01	-0.48	-0.2	-0.06
Chromosome	**18**	**Q1**				**Chromosome**	**18**	**HMPS**			
GATA178F11	-4.29	-1.28	-0.58	-0.24	-0.08	GATA178F11	-1.54	-0.67	-0.29	-0.11	-0.03
D18S481	-1.67	-0.16	0.21	0.24	0.14	D18S481	-0.86	-0.24	0.03	0.1	0.08
D18S976	-4.29	-1.22	-0.55	-0.23	-0.07	D18S976	-2.79	-0.73	-0.34	-0.16	-0.06
D18S843	-2.32	-0.72	-0.31	-0.07	0.02	D18S843	-1.18	-0.31	-0.03	0.04	0.03
D18S542	-1.35	-0.43	-0.12	-0.01	0.01	D18S542	-2.48	-0.31	-0.09	-0.05	-0.04
D18S877	-0.54	0.19	0.19	0.11	0.02	D18S877	-0.74	0.23	0.23	0.13	0.03
D18S535	-1.13	-0.32	-0.08	-0.03	-0.02	D18S535	-0.14	-0.24	-0.06	-0.01	-0.02
D18S851	-1.18	-0.71	-0.34	-0.15	-0.05	D18S851	-2.34	-0.79	-0.42	-0.22	-0.10
D18S858	-1.54	-0.72	-0.45	-0.24	-0.09	D18S858	-2.44	-1	-0.44	-0.13	-0.01
ATA7D07	-0.29	-0.23	-0.13	-0.06	-0.01	ATA7D07	-1.16	-0.33	-0.14	-0.05	-0.01
GATA7E12	-0.92	0.07	0.12	0.08	0.03	GATA7E12	-2.23	-0.26	0	0.05	0.03
ATA82B02	-0.8	0	0.02	-0.02	-0.04	ATA82B02	-1.78	-0.01	0.26	0.2	0.06
GATA177C03	-1.01	0.1	0.12	0.06	0.01	GATA177C03	-0.47	-0.51	-0.25	-0.08	-0.02
D18S844	-0.56	-0.2	-0.11	-0.08	-0.05	D18S844	-1.43	-0.54	-0.18	-0.06	-0.01

APPENDIX TWO

Chromosome	19	Q1			
D19S591	-2.06	-0.84	-0.37	-0.14	-0.04
D19S586	-3.03	-1.35	-0.72	-0.31	-0.07
D19S714	-4.81	-1.58	-0.77	-0.31	-0.07
D19S433	-2.15	-0.57	-0.15	0.01	0.04
D19S178	-0.36	-0.43	-0.34	-0.19	-0.07
D19S246	-3.56	-1.3	-0.58	-0.17	0.01
D19S254	-2.34	-0.5	-0.13	0.01	0.03

Chromosome	19	HMPS			
D19S591	-3.38	-1.04	-0.49	-0.22	-0.07
D19S586	-1.36	-0.54	-0.11	0.05	0.07
D19S714	-1.56	-0.51	-0.13	0.01	0.05
D19S433	0.63	0.46	0.31	0.19	0.09
D19S178	-1.84	-0.63	-0.35	-0.19	-0.08
D19S246	-2.41	-1.4	-0.68	-0.29	-0.09
D19S254	-1.42	-0.21	-0.03	0.02	0.02

Chromosome	20	Q1			
D20S103	0.07	0.08	0.08	0.06	0.02
D20S482	-0.52	-0.3	-0.12	-0.03	0.00
D20S851	-1.63	-0.28	-0.06	-0.01	-0.01
D20S604	-0.73	-0.06	0.04	0.05	0.02
D20S470	-1.95	-1.21	-0.75	-0.38	-0.13
D20S481	-3.61	-1.59	-0.83	-0.36	-0.10
D20S480	-2.01	-0.93	-0.47	-0.21	-0.07
D20S171	-1.69	0.22	0.54	0.44	0.19

Chromosome	20	HMPS			
D20S103	-0.43	-0.22	-0.1	-0.03	-0.00
D20S482	-1.25	-0.53	-0.24	-0.1	-0.03
D20S851	0.03	0.32	0.27	0.15	0.04
D20S604	0.71	0.49	0.3	0.14	0.04
D20S470	0.57	0.38	0.17	0.02	-0.03
D20S481	-1.56	-0.7	-0.4	-0.22	-0.09
D20S480	-1.38	-0.54	-0.28	-0.14	-0.05
D20S171	-1.02	0.03	0.2	0.17	0.07

Chromosome	21	Q1			
D21S1432	-0.25	0.15	0.17	0.11	0.04
D21S1437	-4.27	-1.58	-0.87	-0.43	-0.16
GATA129D11	-4.27	-1.47	-0.7	-0.28	-0.07
D21S1440	-3.78	-1.38	-0.61	-0.24	-0.08
GATA188F04	-1.85	-0.53	-0.12	0.04	0.06

Chromosome	21	HMPS			
D21S1432	-0.2	-0.08	-0.03	-0.02	-0.01
D21S1437	-1.38	-0.6	-0.31	-0.17	-0.06
GATA129D11	-1.27	-0.7	-0.29	-0.09	-0.01
D21S1440	-0.4	0.45	0.42	0.27	0.09
GATA188F04	-0.71	-0.58	-0.29	-0.1	-0.01

Chromosome	22	Q1			
D22S420	-0.49	0.47	0.62	0.5	0.25
D22S345	-1.12	0.34	0.56	0.45	0.19
D22S689	-0.84	-0.08	0.22	0.28	0.19
D22S685	-0.37	0.53	0.58	0.45	0.25
D22S683	-3.78	-0.87	-0.43	-0.21	-0.07

Chromosome	22	HMPS			
D22S420	-0.52	0.15	0.23	0.19	0.10
D22S345	0.46	0.69	0.57	0.35	0.11
D22S689	-1.86	-1.26	-0.67	-0.31	-0.10
D22S685	-1.7	-0.72	-0.41	-0.22	-0.10
D22S683	-1.27	-0.48	-0.23	-0.1	-0.03