Comprehensive analysis of colorectal cancer-risk loci and survival outcome; A prognostic role for *CDH1* variants.

Matthew G. Summers[1], Timothy S. Maughan[2], Richard Kaplan[3], Philip J. Law[4], Richard S. Houlston[4], Valentina Escott-Price[5] and Jeremy P. Cheadle[1]

[1]Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN; [2]CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Roosevelt Drive, Oxford OX3 7DQ; [3]MRC Clinical Trials Unit, Aviation House, 125 Kingsway, London, WC2B 6NH; [4]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP; [5]Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ.

Running title: Colorectal cancer-risk loci and survival outcome

Keywords: Colorectal cancer, *CDH1*, survival, risk-loci, prognostic biomarker

Financial support: This work was supported by Cancer Research Wales. The work of the Houlston laboratory was supported by Cancer Research UK (C1298/A8362). The COIN and COIN-B trials were funded by Cancer Research UK and an unrestricted educational grant from Merck-Serono.

Correspondence to: Professor Jeremy P. Cheadle, Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK. Tel: +442920742652, E-mail: cheadlejp@cardiff.ac.uk

Word count: 2,029

Total number of figures and tables: 4

ABSTRACT

Purpose

Genome wide association studies have identified common single nucleotide polymorphisms (SNPs) at 83 loci associated with colorectal cancer (CRC) risk in European populations. Since germline variation can also influence patient outcome, we studied the relationship between these SNPs and CRC survivorship.

Experimental Design

For the 83 risk loci, 10 lead SNPs were directly genotyped, 72 were imputed and 1 was not genotyped nor imputed, in 1,948 unrelated patients with advanced CRC from the clinical trials COIN and COIN-B (oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab). A Cox survival model was used for each variant, and variants classified by pathway, adjusting for known prognostic factors. We imposed a Bonferroni threshold of $P=6.6\text{x}10^{-4}$ for multiple testing. We carried out meta-analyses of published risk SNPs associated with survival.

Results

Univariate analysis identified six SNPs associated with overall survival (OS) ($P<0.05$); however, only rs9939049 in *CDH1* remained significant beyond the Bonferroni threshold (Hazard Ratio [HR] 1.44, 95% Confidence Intervals [CI] 1.21-1.71, $P=5.0\text{x}10^{-5}$). Fine mapping showed that rs12597188 was the most significant SNP at this locus and remained significant after adjustment for known prognostic factors beyond multiple testing thresholds (HR 1.23, 95% CI 1.13-1.34, $P=1.9\text{x}10^{-6}$). rs12597188 was also associated with poor response to therapy (OR 0.61, 95% CI 0.42–0.87, $P=6.6\text{x}10^{-3}$). No combinations of SNPs within pathways were more

significantly associated with survival compared to single variants alone and no other risk SNPs were associated with survival in meta-analyses.

Conclusions

The CRC susceptibility SNP rs9939049 in *CDH1* influences patient survival and warrants further evaluation as a prognostic biomarker.

INTRODUCTION

Each year, over a million people are diagnosed with colorectal cancer (CRC) worldwide. Clinical stage, which combines depth of tumour invasion, nodal status and distant metastasis [1], is the only routinely used marker of survival. Other factors thought to influence prognosis include lifestyle [2,3], systemic inflammatory response to the tumour [4], the tumour immunologic microenvironment [5] and the patient's germline and the tumour's somatic genetic profile [6-9].

Around 6% of CRC is associated with Mendelian susceptibility caused by the inheritance of rare high-impact germline mutations [10] including those responsible for familial adenomatous polyposis (FAP) (MIM 175100) [11], hereditary non-polyposis CRC (HNPCC; MIM 114500) [12] and MUTYH-associated polyposis (MAP; MIM 608456) [13]. Increasingly, it is being recognised that as well as influencing CRC risk, germline variation plays a role in patient outcome with HNPCC and MAP-associated CRC typically being associated with better prognosis than those with sporadic CRC [14-16].

Genome wide association studies (GWAS) have been successful in identifying single nucleotide polymorphisms (SNPs) robustly associated with an individual's risk of developing CRC. As well as influencing risk, studies have suggested that some of these alleles may affect patient survival [17-22]. However, most studies have not been performed in the context of a clinical trial but have been retrospective in design with the inherent biases from variation in patient management.

We have previously studied the relationship between SNP genotype and patient outcome for 14 of the GWAS risk-loci by analysing patient data from two clinical trials - COIN and COIN-B [23]. Since this study, an additional 69 loci have been identified which influence CRC risk in European populations [24]. To gain a comprehensive understanding of the role of genetic variation on patient outcome we assessed the prognostic effects of all known CRC-risk SNPs in 1,948 advanced disease patients by further utilising COIN and COIN-B trial data.

MATERIALS AND METHODS

Samples

We prepared blood DNA samples from unrelated patients with metastatic or locally advanced colorectal adenocarcinoma from the MRC clinical trials COIN (NCT00182715) [25] and COIN-B (NCT00640081) [26]. All patients gave fully informed consent for bowel cancer research (approved by REC [04/MRE06/60]). COIN patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy, continuous chemotherapy and cetuximab, or intermittent chemotherapy. COIN-B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab, or intermittent chemotherapy and continuous cetuximab.

Genotyping

As previously described [27], 2,244 cases from COIN and COIN-B were genotyped using Affymetrix Axiom Arrays according to the manufacturer's recommendations (Affymetrix, Santa Clara, CA 95051, USA). Individuals were excluded from analysis if they failed one or more of the following thresholds: overall successfully genotyped

SNPs <95% (n=122), discordant sex information (n=8), classed as out of bounds by Affymetrix (n=30), duplication or cryptic relatedness (n=4), and evidence of non-white European ancestry by PCA-based analysis (n=130). After quality control, we had whole genome SNP genotyping and derived imputation data on 1,950 patients, 2 of whom had no data on survival and were excluded (n=1,948). For the 83 CRC-risk loci, 10 lead SNPs were directly genotyped, 72 were imputed and one (rs2732875) was on the X-chromosome which was not genotyped nor imputed. Six SNPs (rs77776598, rs2735940, rs6933790, rs704017, rs6055286 and rs1741640) had info scores <0.7 and were excluded.

Statistical analysis

We used a Cox survival model with overall survival (OS; time from trial randomisation to death) as the primary measure. Univariate analyses were performed using *GenABEL* in R. Multivariate analyses were carried out using *survival* in R. The *coxph* function was used with prognostic covariates in COIN/COIN-B: sex (male vs. female: HR 0.87, 95% CI 0.78-0.97, P=9.7x10^{-4}), World Health Organization (WHO) performance status (HR 1.42, 95% CI 1.31-1.56, P<2.0x10^{-16}), resection status of the primary tumour (unresected/unresectable vs. local recurrence: HR 1.29, 95% CI 1.01-1.63, P=0.04), white blood cell (WBC) count (HR 1.03, 95% CI 1.03-1.04, P<2.0x10^{-16}), platelet count (HR 1.00, 95% CI 1.00-1.00, P<2.0x10^{-16}), number of metastatic sites (HR 1.21, 95% CI 1.14-1.28, P=2.5x10^{-10}), site of distant metastasis (yes vs. no: liver: HR 1.23, 95% CI 1.09-1.39, P=8.8x10^{-4}; peritoneum: HR 1.34, 95% CI 1.16-1.54, P=6.4x10^{-5}; nodal: HR 1.15, 95% CI 1.04-1.28, P=7.5x10^{-3}; other metastases: HR 1.31, 95% CI 1.15-1.51, P=7.9x10^{-5}), *KRAS* status (mutant vs. wild type: HR 1.46, 95% CI 1.29-1.66,

P=3.5x10^{-9}), *BRAF* status (mutant vs. wild type HR 2.29, 95% CI 1.79-2.93, *P*=4.8x10^{-11}) and *NRAS* status (mutant vs. wild type: HR 1.47, 95% CI 1.08-1.99, *P*=0.01), together with other factors in COIN/COIN-B (age at randomisation, cetuximab treatment, chemotherapy regimen, chemotherapy schedule, treatment arm and trial), none of which affected prognosis [25,28]. Response to treatment was defined as complete or partial response and non-response was defined as stable or progressive disease at 12 weeks, and analyses were performed with the *oddsratio* function from the *fmsb* package in R. We used Bonferroni correction to adjust for multiple testing with a significance threshold set at *P*=6.6x10^{-4} (0.05/76 SNPs after exclusion of SNPs with poor imputation).

Meta-analyses

We collected published data for 6 CRC-risk SNPs previously associated with survival, *albeit* at nominally significant levels (*P*<0.05) (rs4939827 [17,19], rs961253 [18,22], rs6983267 [18,21], rs10795668 [17,20], rs4444235 [22] and rs4925386 [17,22]). These SNPs had been analysed in different cohorts: Colorectal Neoplasia Repository and North Central Cancer Treatment Group (NCCTG) [29]; Study on Colorectal Cancer in Scotland (SOCCS) [30]; Seattle Colon Cancer Family Registry (CCFR) [31]; Health Professionals Follow-up Study (HPFS), Nurses' Health Study (NHS), Physicians' Health Study (PHS), VITamins and Lifestyle Study (VITAL), Women's Health Initiative (WH1 and WH2) [17]; Nurses' Health Study (NHS), Health Professionals Follow-up Study (HPFS) [19]; and National Study of Colorectal Cancer Genetics (NSCCG) [22]. Meta-analyses were performed in R using the *meta* package. The *metagen* function was used to perform all analyses under a fixed

effect model, or random effects model where there was significant heterogeneity. I^2 test and Cochran's Q tests were used for assessment of heterogeneity.

Bioinformatics

Linkage disequilibrium (LD) between SNPs was examined using the *--ld* command in PLINK. Forty-nine SNPs were located within, or close to, genes (https://www.ncbi.nlm.nih.gov/snp). Thirteen SNPs were associated with expression quantitative trait loci (eQTL) (https://gtexportal.org/home/). Data from GeneCards (https://www.genecards.org) was used to assess whether ≥2 genes or eQTLs had roles in signalling pathways: 8 genes/eQTLs functioned in GPCR, 6 in TGF-Beta, 5 in ERK, 3 in Wnt, 3 in BMP, 3 in Hedgehog, 3 in PI3K-Akt, 3 in E-cadherin and 2 in Notch signalling. Combinations of SNPs within the same pathway were analysed for survival outcome by the Log Likelihood Ratio test using the *coxph* and *anova* functions in R.

RESULTS

We analysed blood DNA samples and survival data from 1,948 unrelated patients with advanced CRC from the UK national trials COIN [25] and COIN-B [26] (Table 1). We found no evidence of heterogeneity in OS between patients when analysed by trial (COIN vs. COIN-B, *P*=0.33), trial arm (*P*=0.49), type of chemotherapy received (OxMdG/XELOX; *P*=0.46), or cetuximab use (*P*=0.24), so combined these groups for prognostic analyses. In total, 35% of patients were female with a mean age at randomisation of 63 years (range 18-87 years, Table 1). We had over 70% power under an additive model to detect a HR of 1.15 for survival for SNPs with minor allele frequencies (MAFs)>30% and a HR of 1.25 for SNPs with MAFs>10%.

For the 83 CRC-risk loci, 10 lead SNPs were directly genotyped, 72 were imputed and 1 was on the X-chromosome which was not genotyped nor imputed. Univariate analyses identified six SNPs (rs9831861 at 3p21.1, rs35470271 at 3p22.1, rs16892766 at 8q23.3, rs7894531 at 10p14, rs9939049 at 16q22.1 and rs1078643 at 17p12) that were nominally associated with OS (P<0.05) (Table 2). Only rs9939049 was significant beyond the Bonferroni corrected threshold (HR 1.44, 95% CI 1.21-1.71, P=5.0x10^{-5}). rs9939049 lies within *CDH1* in LD with rs9929218 (r^2=0.99, D'=0.99), which we have previously reported having a prognostic effect [23].

We consider whether other SNPs at 16q22.1 might be more significantly associated with survival and analysed all SNPs in LD with rs9939049 for which we had genetic data. rs12597188 (directly genotyped, r^2=0.75, D'=0.99) was the most significantly associated SNP (recessive model: HR 1.48, 95% CI 1.28-1.72, P=1.9x10^{-7}). We considered rs12597188 in multivariate analyses with known prognostic factors in COIN/COIN-B (sex, WHO performance status, resection status of the primary tumour, WBC and platelet count, number of metastatic sites, site of distant metastasis, and *KRAS*, *BRAF* and *NRAS* mutation status). rs12597188 remained significant beyond the Bonferroni corrected threshold (HR 1.23, 95% CI 1.13-1.34, P=1.9x10^{-6}). Patients that were homozygous for the minor allele had a median decrease in life expectancy of 5 months compared to patients that were homozygous or heterozygous for the wild type allele.

We sought whether rs12597188 was associated with response to oxaliplatin-fluoropyrimidine chemotherapy after 12 weeks of treatment (likely to be correlated

with survival, n=1162 patients). Patients that were homozygous for the minor allele had significantly worse response (54/142 responded, 38.0%), as compared to patients that were heterozygous or homozygous wild-type (512/1020 responded, 50.2%) (OR 0.61, 95% CI 0.42–0.87, $P=6.6x10^{-3}$). This association was not seen in patients who also received cetuximab (51.0% versus 49.1%, n=786) with significant heterogeneity between these groups ($I^2=75.2\%$, Cochran's Q test: $P=0.04$).

We tested whether combinations of variants classified by pathway influenced survival. Eight SNPs lie within or near to genes that function in the GPCR signalling pathway, six in the TGF-Beta signalling pathway, five in the ERK signalling pathway, three in each of the Wnt, BMP, Hedgehog, PI3K-Akt and E-cadherin signalling pathways and two in the Notch signalling pathway (Table 3). No combinations of SNPs within specific pathways were more significantly associated with survival beyond the single most significant SNP in that pathway alone.

Six CRC-risk SNPs (rs4939827, rs961253, rs6983267, rs10795668, rs4444235 and rs4925386) have previously been associated with survival [17-22], although none have been independently replicated [30,32,33]. We reviewed published survival data for these SNPs [17,19,22,29-31] and carried out meta-analysis with our data. No SNPs were associated with survival under fixed or random effects models (Figure).

DISCUSSION

Using an independent series of over 5,000 cases with CRC, we have previously validated rs9929218 in *CDH1* as a prognostic biomarker [23]. We have extended these analyses *herein* and shown that rs12597188 is the most significantly

associated SNP at this locus. Our data suggests that patients homozygous for the minor allele of rs12597188, equating to ~12% of patients, have worse survival, with a median decrease in life expectancy of 5 months (in the advanced disease setting). Another study has provided further support for *CDH1* variants having a genuine prognostic effect [20]. Our observations *herein* are limited to patients with stage 4 disease. It is noteworthy that we have previously shown rs9929218 in *CDH1* was not associated with survival amongst patients with Stage 1-3 (pre-metastatic) disease (HR=1.19, 95% CI 0.93-1.52, *P*=0.18) although there was no significant difference between the associations in patients with Stage 1-3 and Stage 4 disease ($P_{interaction}$= 0.48) [23]. Larger studies of pre-metastatic patients may help clarify the potential prognostic role of this biomarker in a population based setting. It is also important to note that the effect sizes for *CDH1* variants are modest and will need to be combined with other germline and somatic prognostic factors to have any role in patient management; we are currently modelling potential combined effects in the advanced disease setting.

rs12597188, rs9939049 and rs9929218 are in strong LD with rs16260 [34] in the *CDH1* promoter which down-regulates *CDH1* expression [35]. *CDH1* encodes E-cadherin. Patients homozygous for the minor alleles of these variants would be expected to have reduced E-cadherin expression. E-cadherin functions as a transmembrane glycoprotein involved in intercellular adhesion, cell polarity and tissue morphology and regeneration [36]. Critically, its loss represents a defining feature of the epithelial to mesenchymal transition during metastasis. *CDH1* variants are therefore plausible prognostic biomarkers which influence this process.

Beyond *CDH1* variants, the next CRC-risk loci most associated with survival was rs16892766 at 8q23.3. However, this variant was not significant after Bonferroni correction and was not significant in an independent cohort of >5000 patients with CRC [23]. Furthermore, our meta-analyses did not support a prognostic role for six other risk loci previously associated with survival. Given that our study was well powered to find variants with HRs>1.15, it is likely that no other low penetrance CRC-risk loci identified to-date have clinically-actionable effects on survival.

ACKNOWLEDGEMENTS

We thank the patients and their families who participated and gave their consent for this research, and the investigators and pathologists throughout the UK who submitted samples for assessment. COIN and COIN-B were coordinated by the Medical Research Council Clinical Trials Unit and conducted with the support of the National Institute of Health Research Cancer Research Network.

ROLE OF THE FUNDING SOURCE

The study sponsors had no involvement in the study design, collection, analysis and interpretation of the data, the writing of the report, nor the decision to submit this article for publication.

Conflict of interest statement: None declared.

REFERENCES

[1] Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 2009;9:489-99.

[2] Haydon AM, MacInnis RJ, English DR, Giles GG. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. Gut 2006;55:62-7.

[3] Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 2007;335:1134.

[4] Leitch EF, Chakrabarti M, Crozier JE, McKee RF, Anderson JH, Horgan PG, et al. Comparison of the prognostic value of selected markers of the systemic inflammatory response in patients with colorectal cancer. Br J Cancer 2007;97:1266-70.

[5] Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Largorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313:1960-4.

[6] Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 2005;23:609-18.

[7] Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 2008;57:941-50.

[8] Lochhead P, Kuchiba A, Imamura Y, Liao X, Yamauchi M, Nishihara R, et al. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. J Natl Cancer Inst 2013;105:1151-6.

[9] Eklöf V, Wikberg ML, Edin S, Dahlin AM, Jonsson BA, Öberg Ă, et al. The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. Br J Cancer 2013;108:2153-63.

[10] Aaltonen L, Johns L, Jarvinen H, Mecklin JP and Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res 2007;13,356-61.

[11] Fearnhead NS, Britton MP and Bodmer WF. The ABC of APC. Hum Mol Genet 2001;10,721-33.

[12] Peltomäki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 2001;10,735-40.

[13] Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. Nat. Genet. 2002;30,227-32.

[14] Watson P, Lin KM, Rodriguez-Bigas MA, Smyrk T, Lemon S, Shashidharan M, et al. Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members. Cancer 1998;83,259-66.

[15] Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354,2751-63.

[16] Nielsen M, van Steenbergen LN, Jones N, Vogt S, Vasen HF, Morreau H, et al. Survival of MUTYH-associated polyposis patients with colorectal cancer and matched control colorectal cancer patients. J Natl Cancer Inst. 2010;102:1724-30.

[17] Phipps AI, Newcomb PA, Garcia-Albeniz X, Hutter CM, White E, Fuchs CS, et al. Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. Gastroenterology 2012;143:51-4.

[18] Dai J, Gu J, Huang M, Eng C, Kopetz ES, Ellis LM, et al. GWAS-identified colorectal cancer susceptibility loci associated with clinical outcomes. Carcinogenesis 2012;33:1327-31.

[19] Garcia-Albeniz X, Nan H, Valeri L, Morikawa T, Kuchiba A, Phipps AI, et al. Phenotypic and tumor molecular characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. Carcinogenesis 2013;34:292-8.

[20] Abulí A, Lozano JJ, Rodríguez-Soler M, Jover R, Bessa X, Munoz J, et al. Genetic susceptibility variants associated with colorectal cancer prognosis. Carcinogenesis 2013;34:2286-91.

[21] Takatsuno Y, Mimori K, Yamamoto K, Sato T, Niida A, Inoue H, et al. The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. Ann Surg Oncol 2013;20:1395-402.

[22] Morris EJ, Penegar S, Whiffin N, Broderick P, Bishop DT, Northwood E, et al. A retrospective observational study of the relationship between single nucleotide polymorphisms associated with the risk of developing colorectal cancer and survival. PLoS One 2015;10:e0117816.

[23] Smith CG, Fisher D, Harris R, Maughan TS, Phipps AI, Richman SD, et al. Analyses of 7,635 patients with colorectal cancer using independent training and validation cohorts show that rs9929218 in CDH1 is a prognostic marker of survival. Clin Can Res 2015;21:3453-61.

[24] Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Comm 2019;10:2154.

[25] Maughan TS, Adams RA, Smith CG, Meade AM, Seymour MT, Wilson RH, et al. The addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. Lancet 2011;377:2103-14.

[26] Wasan H, Meade AM, Adams R, Wilson R, Pugh C, Fisher D, et al. Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. Lancet Oncol 2014;15:631-9.

[27] Al-Tassan NA, Whiffin N, Hosking FJ, Palles C, Farrington SM, Dobbins SE et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep. 2015;5:10442.

[28] Smith CG, Fisher D, Claes B, Maughan TS, Idziaszczyk S, Peuteman G, et al. Somatic profiling of the epidermal growth factor receptor pathway in tumors from patients with advanced colorectal cancer treated with chemotherapy ± cetuximab. Clin Cancer Res 2013;19:4104-13.

[29] Cicek MS, Slager SL, Achenbach SJ, French AJ, Blair HE, Fink SR, et al. Functional and clinical significance of variants localized to 8q24 in colon cancer. Cancer Epidemiol Biomarkers Prev 2009;18,2492-500.

[30] Tenesa A, Theodoratou E, Din FV, Farrington SM, Cetnarskyj R, Barneston RA, et al. Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. Clin Cancer Res 2010;16:3754-9.

[31] Passarelli MN, Coghill AE, Hutter CM, Zheng Y, Makar KW, Potter JD et al. Common colorectal cancer risk variants in SMAD7 are associated with survival among prediagnostic nonsteroidal anti-inflammatory drug users: a population-based study of postmenopausal women. Genes Chromosomes Cancer 2011;50,875-86.

[32] Hoskins JM, Ong PS, Keku TO, Galanko JA, Martin CF, Coleman CA, et al. Association of eleven common, low-penetrance colorectal cancer susceptibility genetic variants at six risk loci with clinical outcome. PLoS One 2012;7:e41954.

[33] Sanoff HK, Renfro LA, Poonnen P, Ambadwar P, Sargent DJ, Goldberg RM, et al. Germline variation in colorectal risk loci does not influence treatment effect or survival in metastatic colorectal cancer. PLoS One 2014;9:e94727.

[34] Pittman AM, Twiss P, Broderick P, Lubbe S, Chandler I, Penegar S, et al. The CDH1-160C>A polymorphism is a risk factor for colorectal cancer. Int J Cancer 2009;125:1622-5.

[35] Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, et al. A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. Cancer Res 2000;60:873-6.

[36] Takeichi M. Cadherin cell adhesion receptors as a morphogenetic regulator. Science 1991;251:1451-5.

TABLES

Table 1 – Clinicopathological data for patients in COIN and COIN-B

	COIN	COIN-B
No. cases with blood DNA	2078	196
No. with genotyping data after QC	1778	170
Total no. deaths (% of cases)	1557 (75)	99 (51)
Median follow-up (SD)	2.4 (2.2)	2.0 (4.4)
% Female	34	42
Age at randomisation, N (%)		
<65 years	1203 (58)	115 (59)
65–69	422 (20)	35 (18)
70–74	318 (15)	31 (16)
75–79	124 (6)	10 (5)
≥80 years	9 (<1)	5 (3)
Missing	2 (<1)	0 (0)
Mean (SD)	62.0 (9.6)	61.7 (10.4)
Stage (%)		
1	0 (0)	0 (0)
2-3	0 (0)	0 (0)
4	2078 (100)	196 (100)
Unknown	0 (0)	0 (0)
Tumour site, N (%)		
Colon[a]	1103 (53)	124 (63)
Rectum[b]	951 (46)	71 (36)
Unknown	24 (1)	1 (1)

[a]Colon defined as cecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon and sigmoid colon. [b]Rectum defined as rectosigmoid junction and rectum.

Table 2: Univariate analysis of CRC-risk SNPs and overall survival

			Additive Model			Recessive Model		
Locus	**SNP**	**Directly genotyped or imputed info score**	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
1p32.3	rs12143541	0.98	1.01	0.92-1.12	0.80	0.88	0.64-1.21	0.42
1p34.3	rs61776719	0.81	1.01	0.92-1.11	0.77	0.98	0.83-1.15	0.78
1p36.12	rs72647484	0.88	1.03	0.89-1.19	0.67	1.12	0.56-2.24	0.76
1q25.3	rs4546885	0.92	0.98	0.91-1.07	0.69	0.90	0.77-1.06	0.20
1q41	rs6658977	0.99	0.93	0.87-1.01	0.11	0.87	0.75-1.01	$7.5x10^{-2}$
2q11.2	rs11692435	0.85	1.01	0.85-1.19	0.92	2.48	1.23-4.97	0.01
2q33.1	rs11893063	0.92	1.05	0.96-1.14	0.28	1.10	0.96-1.26	0.19
2q33.1	rs7593422	0.98	1.01	0.94-1.09	0.77	1.00	0.88-1.14	0.99
2q35	rs13020391	0.95	0.96	0.89-1.04	0.36	0.89	0.76-1.04	0.13
3p21.1	rs9831861	1.00	0.94	0.87-1.01	0.11	0.86	0.75-1.00	$4.6x10^{-2}$
3p22.1	rs35470271	0.94	1.01	0.91-1.12	0.81	0.64	0.42-0.98	$4.1x10^{-2}$
3q13.2	rs12635946	0.97	0.98	0.91-1.06	0.60	0.97	0.83-1.13	0.71
3q26.2	rs35446936	0.85	0.98	0.88-1.08	0.65	1.03	0.74-1.42	0.87
4q24	rs17035289	DG	0.94	0.85-1.05	0.28	1.06	0.73-1.56	0.74
4q31.21	rs75686861	0.97	1.05	0.93-1.19	0.40	0.92	0.70-1.19	0.52
5p13.1	rs1445011	0.99	1.02	0.94-1.11	0.59	1.02	0.85-1.21	0.85
5q31.1	rs639933	0.80	1.01	0.91-1.12	0.87	1.01	0.81-1.26	0.95
6p12.1	rs62404966	0.97	1.05	0.96-1.14	0.30	1.21	0.98-1.48	$7.2x10^{-2}$
6p21.2	rs1321310	0.98	0.94	0.86-1.03	0.19	0.91	0.71-1.16	0.44
6p21.31	rs16878812	0.99	1.04	0.93-1.17	0.49	1.30	0.86-1.96	0.22
6p21.32	rs9271770	0.98	1.04	0.95-1.15	0.39	0.96	0.71-1.35	0.88
6p21.33	rs3131043	DG	1.03	0.95-1.10	0.50	1.01	0.88-1.15	0.91
6p24.1	rs2070699	DG	1.01	0.93-1.08	0.86	1.02	0.90-1.16	0.77
6q21	rs6928864	0.97	0.90	0.79-1.03	0.13	0.72	0.37-1.38	0.32
7p12.3	rs10951878	0.99	1.04	0.96-1.11	0.34	1.02	0.91-1.15	0.71
7p12.3	rs3801081	1.00	1.05	0.97-1.14	0.19	1.02	0.86-1.22	0.78
8q23.3	rs16892766	DG	1.23	1.08-1.39	$1.3x10^{-3}$	1.83	0.95-3.53	$7.1x10^{-2}$
8q24.21	rs6983267	DG	1.06	0.99-1.15	0.11	1.10	0.97-1.26	0.13
9p21.3	rs1412834	1.00	1.01	0.94-1.08	0.83	0.99	0.87-1.12	0.84
10p14	rs7894531	1.00	0.88	0.81-0.96	$2.6x10^{-3}$	0.77	0.63-0.93	$8.4 x10^{-3}$
10q24.2	rs2193352	1.00	1.01	0.93-1.10	0.81	0.97	0.76-1.24	0.82
10q25.2	rs12255141	0.95	0.94	0.83-1.07	0.35	1.29	0.71-2.34	0.39
11p15.4	rs4450168	0.76	1.09	0.94-1.26	0.24	1.21	0.63-2.34	0.56
11q13.4	rs57796856	0.99	1.02	0.94-1.09	0.65	1.00	0.88-1.13	1.00
11q13.4	rs4944940	0.86	1.01	0.81-1.26	0.93	0.66	0.16-2.64	0.56
11q23.1	rs3087967	DG	1.05	0.98-1.14	0.18	1.12	0.94-1.32	0.21
12p13.31	rs10849438	0.89	1.00	0.88-1.14	0.97	1.37	0.79-2.36	0.26
12p13.32	rs12818766	0.96	0.96	0.87-1.06	0.43	1.08	0.79-1.46	0.64
12p13.32	rs3217810	0.73	0.98	0.84-1.14	0.75	0.73	0.35-1.54	0.41
12q13.13	rs11169572	0.99	1.05	0.97-1.13	0.24	1.12	0.98-1.28	0.10
12q13.3	rs7398375	0.73	0.99	0.87-1.11	0.91	0.83	0.61-1.14	0.26
12q24.12	rs597808	0.99	0.96	0.89-1.04	0.29	0.97	0.85-1.11	0.64
12q24.21	rs7315438	0.97	1.04	0.96-1.13	0.30	1.09	0.95-1.26	0.21
13q13.2	rs9537521	0.86	0.99	0.90-1.08	0.77	0.91	0.75-1.10	0.33
13q13.3	rs12427600	0.97	1.02	0.94-1.11	0.68	0.92	0.74-1.15	0.46
13q22.1	rs45597035	0.94	1.00	0.92-1.09	1.00	0.93	0.78-1.12	0.45
13q22.3	rs1330889	0.96	1.02	0.91-1.14	0.78	0.89	0.58-1.36	0.59
13q34	rs7993934	0.93	1.00	0.92-1.09	0.97	0.96	0.80-1.14	0.62
14q22.2	rs35107139	0.86	1.00	0.91-1.09	0.93	0.96	0.89-1.05	0.42
14q22.2	rs1570405	0.97	1.01	0.93-1.09	0.78	0.95	0.87-1.04	0.23
15q13.3	rs16969681	0.99	1.03	0.91-1.15	0.65	1.04	0.66-1.65	0.85
15q13.3	rs73376930	0.95	1.04	0.95-1.13	0.40	1.02	0.81-1.28	0.89
15q13.3	rs16959063	0.97	0.93	0.80-1.25	0.61	0.93	0.69-1.25	0.61
15q13.3	rs17816465	0.96	1.07	0.97-1.17	0.16	1.07	0.83-1.39	0.58
15q22.31	rs4776316	0.74	1.00	0.88-1.13	0.94	1.03	0.72-1.47	0.88
15q23	rs10152518	0.90	0.97	0.87-1.07	0.52	1.21	0.89-1.63	0.23
15q26.1	rs7495132	0.97	1.00	0.89-1.11	0.95	0.91	0.63-1.33	0.63
16q22.1	**rs9939049**	**1.00**	**1.12**	**1.03-1.21**	**$8.1x10^{-3}$**	**1.44**	**1.21-1.71**	**$5.0x10^{-5}$**
16q23.2	rs61336918	0.99	0.96	0.88-1.04	0.31	0.87	0.72-1.05	0.14

16q24.1	rs2696839	0.93	1.00	0.92-1.08	0.98	1.00	0.87-1.14	0.97
16q24.1	rs899244	0.96	1.02	0.93-1.12	0.68	0.92	0.70-1.20	0.53
17p12	rs1078643	0.85	0.94	0.84-1.05	0.30	1.45	1.03-2.03	3.2×10^{-2}
17p13.3	rs73975588	0.97	1.05	0.93-1.19	0.43	0.96	0.54-1.69	0.88
18q21.1	rs7226855	DG	0.99	0.92-1.07	0.83	0.95	0.83-1.10	0.50
19p13.11	rs285245	0.98	0.95	0.79-1.14	0.57	NA	NA	NA
19q13.11	rs73039434	0.76	1.09	0.84-1.43	0.51	NA	NA	NA
19q13.2	rs9797885	0.91	1.00	0.92-1.09	0.94	1.05	0.95-1.16	0.35
19q13.33	rs12979278	0.92	0.98	0.91-1.06	0.65	1.02	0.95-1.10	0.51
20p12.3	rs961253	DG	0.99	0.92-1.07	0.84	0.97	0.84-1.13	0.71
20p12.3	rs6085661	0.99	1.00	0.93-1.08	0.90	1.12	0.97-1.28	0.11
20q13.12	rs2179593	0.98	0.96	0.88-1.05	0.37	0.89	0.72-1.11	0.31
20q13.13	rs6066825	DG	0.97	0.90-1.05	0.46	1.01	0.86-1.18	0.93
20q13.13	rs4811050	DG	1.01	0.92-1.11	0.86	1.00	0.77-1.31	0.98
20q13.13	rs1810502	0.83	1.07	0.98-1.17	0.15	1.07	0.90-1.26	0.44
20q13.13	rs6091213	0.91	1.03	0.94-1.12	0.54	1.01	0.82-1.25	0.92
20q13.33	rs3787089	0.78	1.03	0.93-1.14	0.54	1.03	0.82-1.31	0.79

HR: Hazard ratio, CI: Confidence interval, *P*: *P*-value, SNP: Single nucleotide polymorphism, NA: Not applicable. DG: Directly genotyped. Only rs9939049 at 16q22.1 (bold) was significant beyond the Bonferroni corrected threshold of $P=6.6 \times 10^{-4}$.

Table 3: Variants classified by signalling pathway

Signalling pathway	SNPs
GPCR	rs2070699, rs4776316, rs3801081, rs9537521, rs35107139, rs6066825, rs73039434, rs62404966
TGF-Beta	rs62404966, rs12427600, rs4776316, rs35107139, rs7226855, rs73376930
ERK	rs4546885, rs62404966, rs7993934, rs35107139, rs9939049
Wnt	rs75686861, rs12427600, rs73376930
BMP	rs12427600, rs4776316, rs73376930
Hedgehog	rs75686861, rs12427600, rs73376930
PI3K-Akt	rs4546885, rs16878812, rs597808
E-cadherin	rs17816465, rs16959063, rs9939049
Notch	rs12427600, rs73376930

LEGEND TO FIGURE

Meta-analysis of six CRC risk-SNPs previously associated with survival. Forest plots shown using a fixed effect model. rs10795668 and rs4939827 showed evidence of between-study heterogeneity (I^2=72%, Cochran's Q *P*=0.01 and I^2=69%, Cochran's Q *P*<0.01, respectively); however, neither were associated with survival when also considered under a random effects model (*P*=0.58 and *P*=0.22, respectively). Note - Results from Tenesa et al., 2010 [30] and Morris et al., 2015 [22] were not adjusted for prognostic factors; Cicek et al., 2009 [29] adjusted for tumour characteristics at diagnosis, mismatch repair status, tumour site and stage; Passarelli et al., 2011 [31] adjusted for age at diagnosis and race; Phipps et al., 2012 [17] adjusted for age, and sex (VITamins and Lifestyle Study); Garcia-Albeniz et al., 2013 [19] adjusted for age, race, sex, tumour stage, grade of differentiation, aspirin use, smoking status, alcohol consumption, consumption of meat, and, calcium and folate intake. The survival measure used by Phipps et al., 2012 [17], Garcia-Albeniz et al., 2013 [19], Tenesa et al., 2010 [30] and Passarelli et al., 2011 [31] was diagnosis to death with any cause mortality, Cicek et al., 2009 [29] used overall survival or when censored at eight years, and Morris et al., 2015 [22] used the date of recruitment to date of death or when censored at five years. Where appropriate, the inverse HR of those reported is shown to ensure the allele analysed for each study is consistent. HR: Hazard ratio. CI: Confidence Interval.