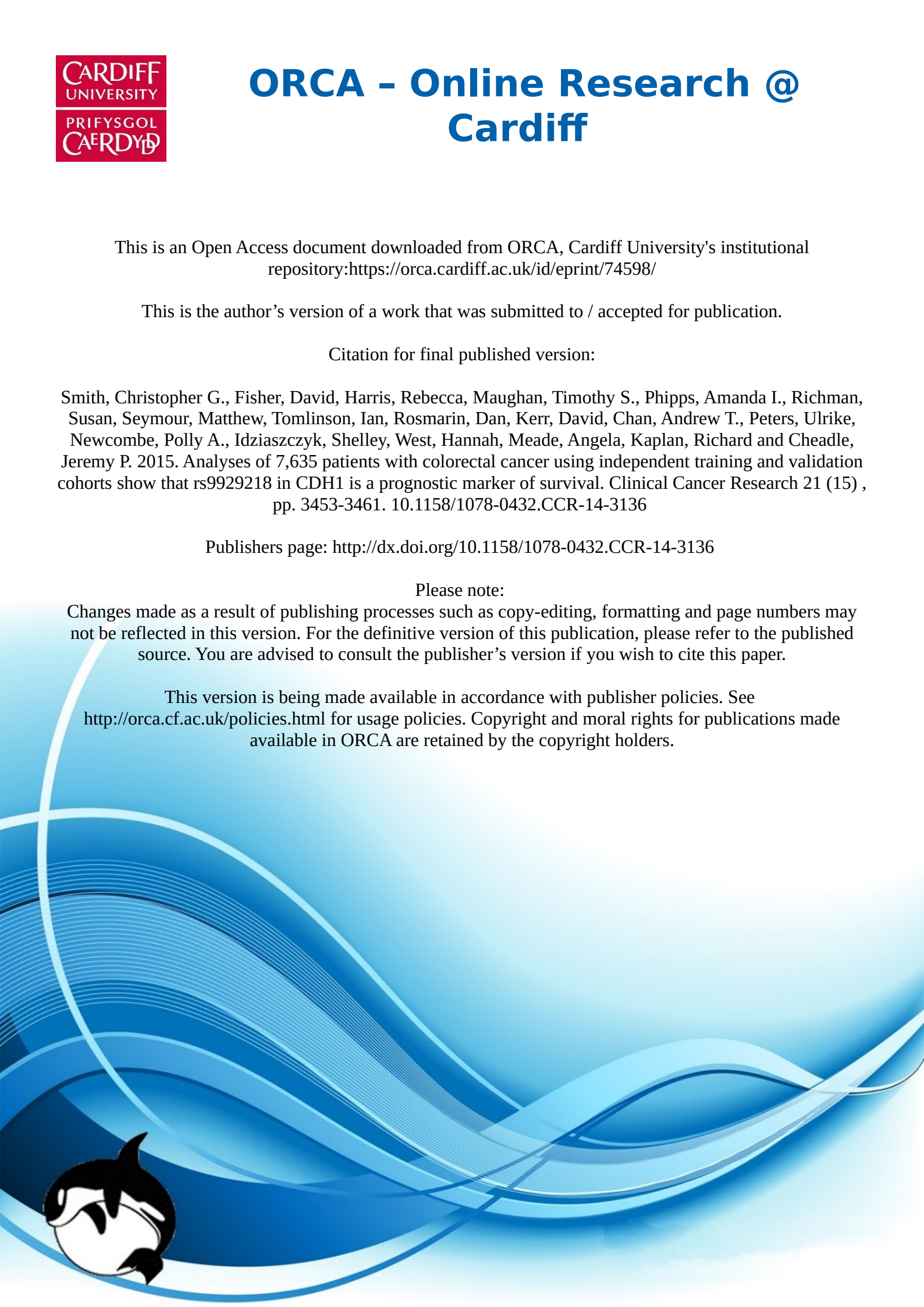


Analyses of 7,635 patients with colorectal cancer using independent training and validation cohorts show that rs9929218 in *CDH1* is a prognostic marker of survival

Christopher G. Smith[1*], David Fisher[2*], Rebecca Harris[1], Timothy S. Maughan[3], Amanda I. Phipps[4,5], Susan Richman[6], Matthew Seymour[6], Ian Tomlinson[7], Dan Rosmarin[7], David Kerr[8], Andrew T. Chan[9,10], Ulrike Peters[4,5], Polly A. Newcomb[4,5], Shelley Idziaszczyk[1], Hannah West[1], Angela Meade[2], Richard Kaplan[2] and Jeremy P. Cheadle[1]

*These authors contributed equally to this study

[1]Institute of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN; [2]MRC Clinical Trials Unit, Aviation House, 125 Kingsway, London, WC2B 6NH; [3]CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Roosevelt Drive, Oxford, OX3 7DQ; [4]Epidemiology Department, University of Washington, Seattle, WA; [5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; [6]Wellcome Trust Brenner Building, St James' University Hospital, University of Leeds, Leeds, LS9 7TF; [7]Molecular and Population Genetics Laboratory and NIHR Comprehensive Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN; [8]Nuffield Department of Clinical Laboratory Sciences, University of Oxford, Oxford, OX3 7DU; [9]Division of Gastroenterology, Massachusetts

General Hospital and Harvard Medical School, Boston, MA; [10]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA.

Running Title:

rs9929218 is a biomarker of survival to colorectal cancer

Keywords:

rs9929218, *CDH1*, prognostic biomarker, colorectal cancer

Financial support:

COIN and COINB were funded by Cancer Research UK and the Medical Research Council, and the associated translational studies were supported by the Bobby Moore Fund from Cancer Research UK, Tenovus, the Kidani Trust, Cancer Research Wales and the National Institute for Social Care and Health Research Cancer Genetics Biomedical Research Unit (2010-2015). FOCUS2 (ISRCTN21221452) was funded jointly by the Medical Research Council and Cancer Research UK. PICCOLO (ISRCTN 93248876) was funded by Cancer Research UK, with support from Amgen. FOCUS3 was funded by the Medical Research Council Efficacy and Mechanism Evaluation programme. Core funding to the Wellcome Trust Centre for Human Genetics was provided by the Wellcome Trust (grant 090532/Z/09/Z). GECCO was supported by the National Cancer Institute (NCI), National Institutes of Health (NIH) and U.S. Department of Health and Human Services (DHHS) (U01 CA137088 and R01 CA059045). NIH also supported HPFS (P01 CA055075, UM1 CA167552, R01 CA137178 and P50 CA127003), NHS (R01 CA137178, P01 CA087969 and

P50 CA127003), PHS (R01 CA042182) and VITAL (K05 CA154337). PLCO was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, and supported by contracts from the Division of Cancer Prevention, NCI, NIH and DHHS. WHI was supported by the National Heart, Lung and Blood Institute, NIH and DHHS (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C).

Correspondence to:

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

COI statement:

The authors disclose no potential conflicts of interest

Word count (excluding references): 2601

Total number of figures and tables: 6

STATEMENT OF TRANSLATIONAL RELEVANCE

Numerous studies have attempted to identify common inherited variants that affect survival in patients with colorectal cancer (CRC). However, none of the proposed prognostic biomarkers have been confirmed, often because the original studies have used small numbers of patients and/or not used independent validation cohorts. We have overcome these limitations and sought robust prognostic biomarkers by analysing 20 genome-wide significant CRC-risk alleles in a large training phase cohort (n=2083 patients with CRC), with subsequent validation of positive associations in an independent study group (n=5552 patients with CRC). We found that rs9929218 (intron 2 of *CDH1*, encoding E-cadherin) was robustly associated with survival. Patients homozygous for the minor allele (AA genotype, ~8% of patients) had worse survival, which equated to a median decrease in life expectancy of 4.3 months, and was independent of known prognostic factors. Our findings clearly demonstrate that common germline variants influence life expectancy in patients with CRC.

ABSTRACT

Purpose

Genome wide association studies have identified numerous loci associated with colorectal cancer (CRC) risk. Several of these have also been associated with patient survival, although none have been validated. Here, we used large independent training and validation cohorts to identify robust prognostic biomarkers for CRC.

Experimental Design

In our training phase, we analysed 20 CRC-risk single nucleotide polymorphisms (SNPs) from 14 genome wide associated loci, for their effects on survival in 2083 patients with advanced CRC. A Cox survival model was used, stratified for treatment, adjusted for known prognostic factors and corrected for multiple testing. Three SNPs were subsequently analysed in an independent validation cohort of 5552 CRC patients. A validated SNP was analysed by disease stage and response to treatment.

Results

Three variants associated with survival in the training phase; however, only rs9929218 at 16q22 (intron 2 of *CDH1*, encoding E-cadherin) was significant in the validation phase. Patients homozygous for the minor allele (AA-genotype) had worse survival (training phase HR=1.43, 95%CI 1.20-1.71, P=5.8x10^{-5}; validation phase HR=1.18, 95%CI 1.01-1.37, P=3.2x10^{-2}; combined HR=1.28 95%CI 1.14-1.43, P=2.2x10^{-5}). This effect was independent of known prognostic factors, and was significant amongst

patients with stage 4 disease (P=2.7x10^{-5}). rs9929218 was also associated with poor response to chemotherapy (P=3.9x10^{-4}).

Conclusions

We demonstrate the potential of common inherited genetic variants to inform patient outcome and show that rs9929218 identifies ~8% of CRC patients with poor prognosis. rs9929218 may affect *CDH1* expression and E-cadherin plays a role in epithelial-mesenchymal transition providing a mechanism underlying its prognostic potential.

INTRODUCTION

Worldwide, over a million people are diagnosed with colorectal cancer (CRC) each year. Several factors influence survival after diagnosis, but the only routinely used prognostic marker is clinical stage which combines depth of tumour invasion, nodal status and distant metastasis (1). Other factors thought to influence prognosis include lifestyle (2,3), systemic inflammatory response to the tumour (4), the tumour immunologic microenvironment (5) and the tumour's somatic molecular profile (6-9).

The search for inherited factors that affect prognosis has primarily focussed on candidate genes that either function within the pharmacological pathways of the chemotherapeutic agents used in the treatment of CRC (10,11) or that influence tumour progression (12). Recently, high-throughput single nucleotide polymorphism (SNP) arrays have been used to search for CRC-susceptibility alleles by genome-wide association studies (GWAS) and, to-date, identified 27 genome-wide significant low penetrance loci mapping to 8q24 (13,14), 18q21 (15,16), 15q13 (17,18), 11q23 (16), 10p14 (19), 8q23 (19), 14q22 (20), 16q22 (20), 19q13 (20), 20p12 (20,21), 1q41 (22), 3q26 (22), 12q13 (22), 20q13 (22), 6p21 (23), 11q13 (23), Xp22 (23), 2q32 (24), 12p13 (21,25,26), 5q31 (21), 1q25.3 (24,25), 10q24 (25), 10q22 (26), 10q25 (26), 11q12 (26), 17p13 (26) and 19q13 (26). Studies have suggested that some of these risk alleles may also affect patient survival (27-32); however, none of these survival findings, nor any prognostic biomarkers identified through the candidate gene analyses, have been validated in independent studies (33-35).

Here, we sought robust biomarkers of patient survival by analysing 20 genome-wide significant CRC-susceptibility SNPs in a large training phase cohort, with subsequent validation of positive associations in an independent study group.

MATERIALS AND METHODS

Samples

Training phase

We prepared blood DNA samples from unrelated patients with advanced (Stage 4) CRC (aCRC) from the MRC clinical trial COIN (NCT00182715) (36). All patients had either previous or current histologically confirmed primary adenocarcinomas of the colon or rectum, together with clinical or radiological evidence of advanced and/or metastatic disease, or had histologically/cytologically confirmed metastatic adenocarcinomas, together with clinical and/or radiological evidence of a colorectal primary tumour. Patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (Arm A), continuous chemotherapy plus cetuximab (Arm B), or intermittent chemotherapy (Arm C). All patients gave informed consent for their samples to be used for bowel cancer research (approved by REC [04/MRE06/60]).

Validation phase

The validation phase consisted of samples from several different trials or prospective cohort studies. COINB is a MRC-funded phase II trial assessing

cetuximab efficacy in intermittent oxaliplatin-fluoropyrimidine chemotherapy of aCRC (NCT00640081) (37). FOCUS2 is a trial for patients with unpretreated aCRC judged unfit for full-dose combination chemotherapy (NCT00070213). FOCUS3 is a trial determining the feasibility of molecular selection of therapy using *KRAS*, *BRAF* and topoisomerase-1 in aCRC (NCT00975897). PICCOLO is a trial of the treatment for fluorouracil-resistant aCRC (NCT00389870) (patients from COIN or COINB that were subsequently recruited into PICCOLO were excluded). VICTOR is a trial of rofecoxib as post-adjuvant therapy for CRC (NCT00031863). Six prospective cohort studies from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (24,38) were also included: the Health Professionals Follow-up Study (HPFS), the Nurses' Health Study (NHS), the Physicians' Health Study (PHS), the VITamins And Lifestyle Study (VITAL), the Women's Health Initiative (WHI) and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) (see *Supplementary Information* for references). All of these studies used a prospective design, with follow-up for incident cancer diagnoses and survival outcomes. Cases of incident CRC arising in these studies were identified from self-report and confirmed by their medical records (HPFS, NHS, PHS, PLCO, WHI) and/or linkage to cancer registries (VITAL). Two subsets of cases were genotyped in the WHI: WHI1 included colon cancer patients diagnosed before September 2005 and WHI2 included unrelated CRC patients diagnosed before August 2009. Two subsets of cases were also genotyped in PLCO: PLCO1 included colon cancer patients and PLCO2 included unrelated CRC cases. All participants provided informed consent for genetic testing, and all studies were approved by their respective

Institutional Review Boards. Protocols for assessing survival in the GECCO studies have been described previously (see *Supplementary Information* for references).

Genotyping

Training phase

Genotyping of fifteen CRC risk alleles (rs6691170 and rs6687758 at 1q41, rs10936599 at 3q26, rs4444235 and rs1957636 at 14q22, rs9929218 at 16q22, rs10411210 at 19q13, rs961253 at 20p12, rs10795668 at 10p14, rs3802842 at 11q23, rs4925386 at 20q13, rs4939827 at 18q21, rs16892766 at 8q23, rs4779584 at 15q13 and rs6983267 at 8q24) was performed by Illumina's Fast-Track Genotyping Services (San Diego, CA) using their high throughput BeadArray™ technology. rs4925386 failed genotyping. For the remaining 14 SNPs, genotyping concordance rates for duplicate samples (n=110) was 100% (1540/1540 genotypes), GenTrain scores ranged from 0.6814 to 0.9500 and the overall genotype success rate was 99.44% (28868/29032 genotypes were called successfully). Genotyping of rs4925386 at 20q13, rs4813802 at 20p12 and, rs16969681 and rs11632715 at 15q13 was carried out by LGC genomics using their KASPar technology with a genotype success rate of 99.17% (8253/8322 genotypes called successfully) and concordance rate for duplicate samples (n=94) of 100% (376/376). Genotyping of rs11169552 and rs7136702 at 12q13 was carried out by Geneservice (Nottingham, UK) using TaqMan assays (Applied Biosystems) with a genotype success rate of 95.66% (3966/4146 genotypes called

successfully) and concordance rate for duplicate samples (n=94) of 100% (188/188).

Validation phase

rs16892766, rs9929218 and rs10795668 were genotyped in patients from COINB, FOCUS2, FOCUS3 and PICCOLO by LGC genomics (KASPar technology). In VICTOR, genotyping was carried out on Illumina HumanHap300 arrays and rs9929218 was directly genotyped, rs16892766 was imputed and rs706771 was genotyped as a proxy for rs10795668 (R^2=0.965, D'=1). All three SNPs were genotyped in cases from HPFS, NHS, and PHS using the TaqMan Open Array SNP genotyping platform. For the other GECCO studies, genotyping was performed on Illumina 300/240S (PLCO1), 550K (WHI1), 610K (WHI1, PLCO1), and HumanCytoSNP (VITAL, WHI2, PLCO2) arrays; rs9929218 was directly genotyped on these platforms in all studies, and, rs16892766 and rs10795668 were directly genotyped on the platform used in WHI1 and PLCO1, and imputed (using MACH and HapMap2 Release 24) in WHI2, VITAL, and PLCO2. Note – different genotyping platforms were often used because susceptibility SNPs were identified and assayed at different times by different investigators.

Statistical analyses

All SNPs were tested for their genotypes being consistent with the Hardy Weinberg Equilibrium (HWE) using a Pearson chi-square test. Linkage disequilibrium (LD) was examined using Haploview version 4.2. For survival analyses of the training phase, we used a Cox survival model with overall

survival (time from trial randomisation to death) as the primary measure. A co-dominant model was applied, analyses were stratified for treatment arm and type of fluoropyrimidine used, and *P*-values were corrected for multiple testing by Bonferroni correction. Significant SNPs were tested for independence to known prognostic factors using a closed-test procedure multiple fractional polynomial model with $P<0.05$ and the best-fitting genotype model (dominant or recessive) was identified. For survival analyses in the validation phase, time from randomisation to death (overall survival) was used for COINB, FOCUS2, FOCUS3, PICCOLO and VICTOR, and time from diagnosis to death for HPFS, NHS, PHS, VITAL, WHI and PLCO. A Cox survival model was fitted to the data from each trial or study separately, and an overall pooled result was calculated using a fixed-effects inverse-variance meta-analysis approach. Heterogeneity was assessed using the Q and I-squared statistics. If the pooled validation data generated a significant result, additional analyses were conducted: (i) a further meta-analysis including the training and validation data together, (ii) a sensitivity analysis replacing time from randomisation to death (considered left-truncated at randomisation to account for the fact that randomisation is conditional upon survival from diagnosis) with time from diagnosis to death - for those trials for which this information was available (COIN, COINB and FOCUS3; n=2446 patients genotyped with survival data), and, (iii) the effect on 12-week response to chemotherapy in COIN Arms A and C (those arms not confounded by treatment with cetuximab; n=1369 patients genotyped with this data). Response was defined as complete response or partial response at 12-weeks and non-response was defined as stable disease or progressive disease.

RESULTS

Training phase

We analysed blood DNA samples from 2083 unrelated patients with aCRC from the UK national trial COIN (36). In total, 34% of patients were female with a mean age at diagnosis of 62 years (range 18-84 years, Table 1). We assayed twenty independent, genome-wide significant, CRC-risk alleles (13,15-17,19,20,22) representing 14 loci; with a single SNP at nine loci, two SNPs at four loci and three SNPs at one locus (loci with ≥2 SNPs contain multiple independent risk alleles) (20,22). Fifteen SNPs were genotyped using the Illumina GoldenGate platform (one failed), four (including a repeat of the failed SNP) were successfully genotyped using KASPar technology and two were successfully genotyped using Taqman assays. All 20 SNPs, apart from rs7136702 (P=0.027), had genotype distributions consistent with the HWE with no imbalances between the treatment arms or according to the somatic mutation status of the CRCs (42.27%, 9.01% and 3.56% of CRCs were *KRAS*, *BRAF* and *NRAS* mutant, respectively) (39).

Fourteen SNPs did not influence survival under a co-dominant model (Table 2). Six SNPs were significant in the univariate analyses, of which three (rs16892766 at 8q23, rs9929218 at 16q22 and rs10795668 at 10p14) remained significant after correction for multiple testing (Table 2). We have previously shown that the WHO performance status, number of metastatic sites, white blood cell count, alkaline phosphatase levels and *KRAS* and *BRAF* mutation status are independent prognostic factors affecting survival in

patients from COIN (36). We therefore applied a multivariate model with these factors, together with the best genetic models that fitted the data, and showed that all three SNPs independently influenced survival (Supplementary Table S1).

Validation phase

We used samples from numerous independent trials and cohort studies to provide sufficient power to carry out our validation analyses. In total, we assayed rs16892766, rs9929218 and rs10795668 in 5552 patients with CRC (196 from COINB, 337 from FOCUS2, 172 from FOCUS3, 334 from PICCOLO, 918 from VICTOR, 259 from HPFS, 355 from NHS, 278 from PHS, 531 from PLCO1, 478 from PLCO2, 281 from VITAL, 450 from WHI1 and 963 from WHI2; Table 1). No significant heterogeneity was detected in any of the meta-analyses (I^2=0%). Only rs9929218 was found to be significantly associated with survival (P=2.5x10^{-2}, Supplementary Table S2).

Further analyses of rs9929218

Patients homozygous for the minor allele of rs9929218 (AA genotype), equating to ~8% of patients, showed significantly poorer survival as compared to patients with the AG or GG genotypes (training phase HR 1.47, 95% CI 1.24-1.75, P=1.4x10^{-5} unadjusted, HR=1.43, 95% CI 1.20-1.71, P=5.8x10^{-5} after adjustment for age, sex and time from diagnosis to randomisation; validation phase HR=1.19, 95% CI 1.02-1.38, P=2.5x10^{-2} unadjusted, HR=1.18, 95% CI 1.01-1.37, P=3.2x10^{-2} adjusted; combined HR=1.30 95% CI 1.16-1.46, P=6.1x10^{-6} unadjusted, HR=1.28 95% CI 1.14-1.43, P=2.2x10^{-5}

adjusted; Figure 1 and Table 3). This equated to a median decrease in life expectancy of 4.3 months (based on training phase data). Patients with a single variant allele (AG genotype) had similar survival outcomes to those with a wild type (GG) genotype (Supplementary Table S3).

We combined the training and validation phase data and analysed by disease stage. rs9929218 genotype did not deviate from the HWE according to stage (Supplementary Table S4). rs9929218 was not significantly associated with survival amongst patients with Stage 1-3 (pre-metastatic) disease (HR=1.19, 95% CI 0.93-1.52, P=0.18), with little statistical evidence of heterogeneity amongst the individual studies (P=0.39) (Figure 2). In contrast, rs9929218 was highly associated with survival in patients with Stage 4 (metastatic) CRC (HR=1.34, 95% CI 1.17-1.53, $P=2.7\times10^{-5}$), with no heterogeneity amongst the individual trials and cohorts (P=0.91) (Figure 2). There was, however, no significant difference between the associations of rs9929218 genotype and survival in patients with Stage 1-3 and Stage 4 disease ($P_{interaction}$= 0.48).

As a sensitivity analysis, we investigated whether overall survival accurately reflected survival from the time of diagnosis to death. For 2444 trial patients (from COIN, COINB and FOCUS3) we had relevant clinical information available and we found little difference in the effect of rs9929218 between the two survival measures (overall survival HR=1.50, 95% CI 1.27-1.76, $P=1.5\times10^{-6}$; survival time from diagnosis HR=1.46, 95% CI 1.24–1.73, $P=6.3\times10^{-6}$, Supplementary Figure).

We also investigated whether the type and duration of treatment influenced survival, by evaluating rs9929218 according to trial arm in COIN (the largest trial for which we had high quality clinical data). We did not find significant heterogeneity between the treatment arms (*P*=0.38) suggesting that treatment did not influence the association between rs9929218 genotype and survival (Supplementary Table S5).

We also sought whether rs9929218 was associated with response to treatment (likely to be correlated with survival). In COIN Arms A and C, treatment was identical for the first 12 weeks apart from the choice of fluoropyrimidine. At 12 weeks, patients from these arms that were homozygous for the rs9929218 minor allele had significantly worse response (36/112 responded, 32%), as compared to patients that were heterozygous or homozygous wild-type (626/1257 responded, 50%) (OR 0.47, 95% CI 0.31–0.72, $P=3.9\times10^{-4}$, adjusted for choice of fluoropyrimidine) (Table 4).

DISCUSSION

The literature contains many reports of potential common inherited biomarkers of survival for CRC; however, most of these have been derived from poorly designed studies, with small numbers of samples and/or no validation of their results. As a consequence, very few of these prognostic biomarkers have been validated by independent groups. To address the critical shortcomings of previous studies, we have carried out an analysis using large independent training and validation phase cohorts as recommended by the REMARK guidelines (40) and produced robust evidence for the first common inherited

genetic variant affecting survival in patients with CRC. As such, this finding represents an important clinical milestone.

Our data suggest that patients homozygous for the minor allele of rs9929218, equating to ~8% of patients, have worse survival, with a median decrease in life expectancy of ~4 months (in the advanced disease setting). Another study recently reported that the major allele of rs9929218 was associated with improved prognosis (30), providing further support for this variant having a genuine prognostic effect. Although the effect size of rs9929218 identified *herein* is modest (HR=1.28, 95% CI 1.14-1.43), the identification of further prognostic alleles by well-powered GWAS-based approaches may help clinicians model the combined effects of common germline variants together with their somatic mutation profiles to help inform patient outcome. Our study therefore represents a critical first step in this endeavour.

We have shown a clear effect of rs9929218 on survival amongst patients with stage 4 disease. However, many of these patients would have received similar therapies raising the possibility that rs9929218 influences survival based upon an interaction with treatment, and we noted that patients carrying both minor alleles had poor response to chemotherapy. However, survival and response are likely to be related and we found similar prognostic effects across all arms of the COIN trial (including in those patients receiving intermittent therapy) and amongst many of the other trials and cohorts used in this study. These data suggest that the prognostic effect may therefore reflect an underlying influence on a biological process or pathway. rs9929218 lies

within intron 2 of *CDH1* encoding E-cadherin, in strong LD with rs16260 (41) in the *CDH1* promoter which down-regulates *CDH1* expression (42). Patients homozygous for the minor allele of rs9929218 would be expected to have reduced E-cadherin expression. E-cadherin functions as a transmembrane glycoprotein that is critical in the establishment and maintenance of intercellular adhesion, cell polarity and tissue morphology and regeneration (43) and its loss represents a defining feature of the epithelial to mesenchymal transition during metastasis. A clear mechanism therefore exists for the potential prognostic effect of rs9929218 by influencing this process.

ACKNOWLEDGEMENTS

We thank Ayman Madi, Richard Adams, Sarah Kenny and the COIN trial management group for their advice or support. We thank the patients and their families who participated in COIN and gave their consent for this research, and the investigators and pathologists throughout the UK who submitted samples for assessment. COIN, COINB, FOCUS3 and PICCOLO were conducted with the support of the National Institute of Health Research Cancer Research Network. We would also like to thank all staff at the GECCO Coordinating Center. For HPFS, NHS and PHS, we would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Immaculata Devivo and David Hunter, Qin Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for PHS. We would like to thank the participants and staff of the Nurses' Health Study and the

Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. For WHI, the authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. For PLCO, we would like to thank Christine Berg and Philip Prorok, Division of Cancer Prevention, NCI, the Screening Center investigators and staff of the PLCO Cancer Screening Trial, Tom Riley and staff, Information Management Services, Inc., Barbara O'Brien and staff, Westat, Inc., and Bill Kopp, Wen Shao, and staff, SAIC-Frederick. We also acknowledge the study participants for their contributions to making this study possible.

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, GG Giles. 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. Dotor E, Cuatrecases M, Martinez-Iniesta M, Navarro M, Vilardell F, Guinó E, et al. Tumor thymidylate synthase 1494del6 genotype as a prognostic factor in colorectal cancer patients receiving fluorouracil-based adjuvant treatment. J Clin Oncol 2006;24:1603-11.

11. Marcuello E, Altés A, Del Rio E, César A, Menoyo A, Baiget M. Single nucleotide polymorphism in the 5' tandem repeat sequences of thymidylate synthase gene predicts for response to fluorouracil-based chemotherapy in advanced colorectal cancer patients. Int J Cancer 2004;112:733-7.

12. Kim JG, Chae YS, Sohn SK, Cho YY, Moon JH, Park JY, et al. Vascular endothelial growth factor gene polymorphisms associated with prognosis for patients with colorectal cancer. Clin Cancer Res 2008;14:62-6.

13. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39:984-8.

14. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 2007;39:989-94.

15. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1315-7.

16. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Hag N, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 2008;40:631-7.

17. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 2008;40:26-8.

18. Tomlinson IP, Carvajal-Carmona L, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Multiple Common Susceptibility Variants near BMP Pathway Loci GREM1, BMP4, and BMP2 Explain Part of the Missing Heritability of Colorectal Cancer. PLoS Genet 2011;7:e1002105.

19. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 2008;40:623-30.

20. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40:1426-35.

21. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet 2013;45:191-6.

22. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet 2010;42:973-7.

23. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 2012;44:770-6.

24. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 2013;144:799-807.

25. Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Hum Mol Genet 2014;23:4729-37.

26. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet 2014;46:533-42.

27. 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.

28. 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.

29. 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet 2012;131:217-34.

39. 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.

40. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med 2012;9:e1001216.

41. 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.

42. 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.

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

TABLES

Table 1 – Clinical trial and population-based cohorts analysed in this study.

	Training Phase	Validation Phase												
	COIN	**COINB**	**FOCUS2**	**FOCUS3**	**PICCOLO**	**VICTOR**	**HPFS**	**NHS**	**PHS**	**PLCO1**	**PLCO2**	**VITAL**	**WHI1**	**WHI2**
No. cases with rs9929218 genotype	2078[a]	196	337	172	334	918	259	355	278	531	478	281	450	963
GG	1061	106	170	83	170	485	128	186	134	273	261	141	217	471
GA	853	73	143	75	137	361	109	132	123	217	173	112	190	399
AA	164	17	24	14	27	72	22	37	21	41	44	28	43	93
Total no. deaths (% of cases)	1557 (75)	99 (51)	301 (89)	78 (45)	312 (93)	108 (12)	124 (48)	145 (41)	128 (46)	180 (34)	103 (22)	94 (33)	165 (37)	310 (32)
GG	783 (74)	58 (55)	153 (90)	32 (39)	159 (94)	56 (12)	65 (51)	71 (38)	67 (50)	84 (31)	62 (24)	42 (30)	77 (35)	146 (31)
GA	634 (74)	30 (41)	124 (87)	38 (51)	128 (93)	41 (11)	47 (43)	64 (48)	50 (41)	79 (36)	34 (20)	37 (33)	69 (36)	133 (33)
AA	140 (85)	11 (65)	24 (100)	8 (57)	25 (93)	11 (15)	12 (55)	10 (27)	11 (52)	17 (41)	7 (16)	15 (54)	19 (44)	31 (33)
Median follow-up (SD)	2.4 (2.2)	2.0 (4.4)	3.7 (n/a)[b]	1.0 (0.8)	3.0 (3.1)	5.3 (1.4)	5.0 (3.8)	5.4 (4.9)	9.3 (7.4)	6.7 (3.4)	3.4 (3.6)	3.6 (2.3)	5.2 (3.5)	2.9 (3.4)
% Female	34	42	37	37	34	65	0	100	0	43	43	47	100	100
Age at diagnosis, N (%)														
<65 years	1203 (58)	115 (59)	39 (12)	110 (64)			55 (21)	115 (32)	91 (33)	125 (24)	98 (21)	51 (18)	87 (19)	149 (16)
65–69	422 (20)	35 (18)	54 (16)	32 (19)			32 (13)	75 (25)	42 (15)	145 (27)	115 (24)	59 (20)	87 (19)	205 (21)
70–74	318 (15)	31 (16)	104 (31)	17 (10)			55 (21)	78 (22)	37 (13)	161 (30)	131 (27)	90 (31)	133 (30)	248 (26)
75–79	124 (6)	10 (5)	94 (28)	13 (8)	Not collected	Not collected	53 (21)	60 (17)	43 (16)	88 (17)	88 (18)	67 (23)	96 (21)	199 (21)
≥80 years	9 (<1)	5 (3)	46 (14)	0 (0)			62 (24)	27 (8)	65 (23)	12 (2)	46 (10)	14 (5)	47 (10)	162 (17)
Missing	2 (<1)	0 (0)	0 (0)	0 (0)			2 (1)	0 (0)	0 (0)	0 (0)	0 (0)	8 (3)	0 (0)	0 (0)
Mean (SD)	62.0 (9.6)	61.7 (10.4)	72.7 (7.1)	60.9 (10.0)			72.3 (8.7)	68.5 (7.7)	71.3 (9.8)	69 (5.9)	70 (6.6)	70.4 (6.5)	70.9 (7.1)	72.1 (7.2)
Stage (%)														
1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5 (1)	72 (28)	78 (22)	57 (21)	193 (36)	166 (35)	105 (37)	126 (28)	293 (30)
2-3	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	913 (99)	89 (34)	183 (52)	108 (39)	282 (53)	246 (52)	126 (45)	252 (56)	493 (51)
4	2078 (100)	196 (100)	337 (100)	172 (100)	334 (100)	0 (0)	33 (13)	54 (15)	24 (9)	51 (10)	65 (14)	46 (16)	66 (15)	123 (13)
Unknown	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	65 (25)	40 (11)	89 (32)	5 (1)	1 (<1)	4 (1)	6 (1)	54 (6)
Tumour site, N (%)														
Colon[c]	1103 (53)	124 (63)	240 (71)	83 (48)	225 (64)	574 (63)	173 (67)	273 (77)	195 (70)	514 (97)	314 (66)	211 (75)	436 (97)	678 (70)
Rectum[d]	951 (46)	71 (36)	94 (28)	86 (50)	121 (34)	344 (37)	54 (21)	73 (21)	55 (20)	5 (1)	159 (33)	64 (23)	11 (2)	232 (24)
Unknown	24 (1)	1 (1)	3 (1)	3 (2)	7 (2)	0 (0)	32 (12)	9 (3)	28 (10)	12 (2)	5 (1)	6 (2)	3 (1)	53 (6)

Data provided for those samples with an rs9929218 genotype. [a]Of the 2083 COIN patients, 5 failed genotyping for rs9929218. [b]Follow-up never dropped below 50%, so figure represents the median time from patient entry to the cut-off date for analysis. [c]Colon defined as cecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon and sigmoid colon. [d]Rectum defined as rectosigmoid junction and rectum.

Table 2 - Univariate analyses of overall survival in our training phase cohort

SNP	locus	n genotyped	AA	AB	BB	n deaths	[a]HR (95% CI) AB vs AA	[a]HR (95% CI) BB vs AA	X^2	*P*-value	Corrected *P*
rs4939827	18q21	2068	637	1028	403	1552	1.00 (0.89-1.12)	1.02 (0.88-1.17)	0.06	0.97	-
rs16892766	8q23	2079	1688	378	13	1557	1.28 (1.13-1.45)	1.26 (0.67-2.35)	15.14	$5.2x10^{-4}$	$1.0x10^{-2}$
rs4779584	15q13	2070	1245	710	115	1554	0.97 (0.87-1.08)	0.96 (0.77-1.19)	0.36	0.84	-
rs6983267	8q24	2065	674	979	412	1549	1.01 (0.90-1.14)	1.15 (1.00-1.32)	4.41	0.11	-
rs11169552	12q13	2002	1086	785	131	1506	0.91 (0.82-1.01)	0.92 (0.75-1.14)	3.28	0.19	-
rs7136702	12q13	1964	807	868	289	1474	1.00 (0.89-1.11)	1.15 (0.98-1.34)	3.63	0.16	-
rs6691170	1q41	2070	760	1019	291	1554	1.01 (0.90-1.12)	0.89 (0.76-1.04)	2.56	0.28	-
rs6687758	1q41	2066	1302	666	98	1551	0.92 (0.83-1.03)	0.97 (0.78-1.22)	2.08	0.35	-
rs10936599	3q26.2	2070	1218	739	113	1554	0.99 (0.89-1.10)	1.09 (0.87-1.36)	0.61	0.74	-
rs4925386	20q13	2061	973	886	202	1544	0.92 (0.83-1.02)	0.88 (0.74-1.05)	3.48	0.18	-
rs4444235	14q22	2066	571	1008	487	1552	1.00 (0.89-1.12)	0.92 (0.80-1.05)	1.93	0.38	-
rs9929218	16q22	2078	1061	853	164	1557	1.01 (0.91-1.12)	1.47 (1.23-1.76)	18.79	$8.3x10^{-5}$	$1.7x10^{-3}$
rs10411210	19q13	2070	1686	360	24	1554	1.24 (1.09-1.41)	0.94 (0.58-1.52)	10.81	$4.5x10^{-3}$	0.09
rs961253	20p12	2069	808	972	289	1553	1.04 (0.93-1.16)	1.00 (0.85-1.16)	0.65	0.72	-
rs10795668	10p14	1993	940	868	185	1491	0.95 (0.86-1.06)	0.70 (0.58-0.85)	12.42	$2.0x10^{-3}$	$4.0x10^{-2}$
rs3802842	11q23	2070	993	870	207	1554	0.98 (0.88-1.09)	1.13 (0.96-1.34)	2.61	0.27	-
rs1957636	14q22	2069	656	1029	384	1554	0.99 (0.88-1.10)	0.95 (0.82-1.09)	0.59	0.74	-
rs4813802	20p12	2051	795	958	298	1543	0.86 (0.77-0.96)	1.01 (0.87-1.18)	9.26	$9.8x10^{-3}$	0.196
rs16969681	15q13	2060	1637	394	29	1544	1.04 (0.92-1.18)	1.35 (0.92-2.00)	2.61	0.27	-
rs11632715	15q13	2063	535	1034	494	1548	0.86 (0.76-0.97)	0.97 (0.85-1.12)	7.47	$2.4x10^{-2}$	0.48

Analyses used a Cox proportional-hazard model (co-dominant analyses) with the outcome of overall survival, adjusted for treatment arm and chemotherapy regimen (*P*-value) and corrected for multiple testing (corrected *P*-value). [a]The co-dominant model tests for the joint effect of AB vs AA and BB vs AA. n values give the numbers of patients with their respective genotypes and for whom survival data was available. Note – rs4939827, rs961253, rs6983267 and rs4444235 have all been previously associated with survival (27-29,31,32), but none were validated in our study.

Author Manuscript Published OnlineFirst on April 14, 2015; DOI: 10.1158/1078-0432.CCR-14-3136
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Table 3 - Univariate analysis of rs9929218 on survival according to training phase, validation phase and combined

Analysis phase	Alleles	n genotyped	n deaths	HR (95% CI)	*P*-value
Training phase	GG/GA AA	1913 163	1416 139	1.43 (1.20-1.71)	5.8×10^{-5}
Validation phase	GG/GA AA	5069 483	1946 201	1.18 (1.01-1.37)	3.2×10^{-2}
Combined	GG/GA AA	6982 646	3362 340	1.28 (1.14-1.43)	2.2×10^{-5}

Data are shown for recessive analyses with *P*-values adjusted for age, sex and time of diagnosis. HRs for the validation phase and the combined analysis are pooled effects using fixed-effects inverse-variance meta-analysis.

Table 4 - Prognostic effect of rs9929218 on response to chemotherapy

Outcome	GG/AG n (%)	AA n (%)	*P*-value
Response	626 (49.8)	36 (32.1)	χ^2=12.8, 1 d.f. $P=3.9 \times 10^{-4}$
No response	631 (50.2)	76 (67.9)	

Patients were from Arms A and C of COIN in which treatment was identical for the first 12 weeks apart from the choice of fluoropyrimidine. *P*-value is adjusted for choice of fluoropyrimidine.

Legends to Figures

Figure 1 – Forest plot of rs9929218 analysed for survival in the training phase, validation phase and all data combined (adjusted for age, sex and time of diagnosis).

Figure 2 - Forest plot of rs9929218 analysed for survival and stratified by disease stage (adjusted for age, sex and time of diagnosis).

Figure 1

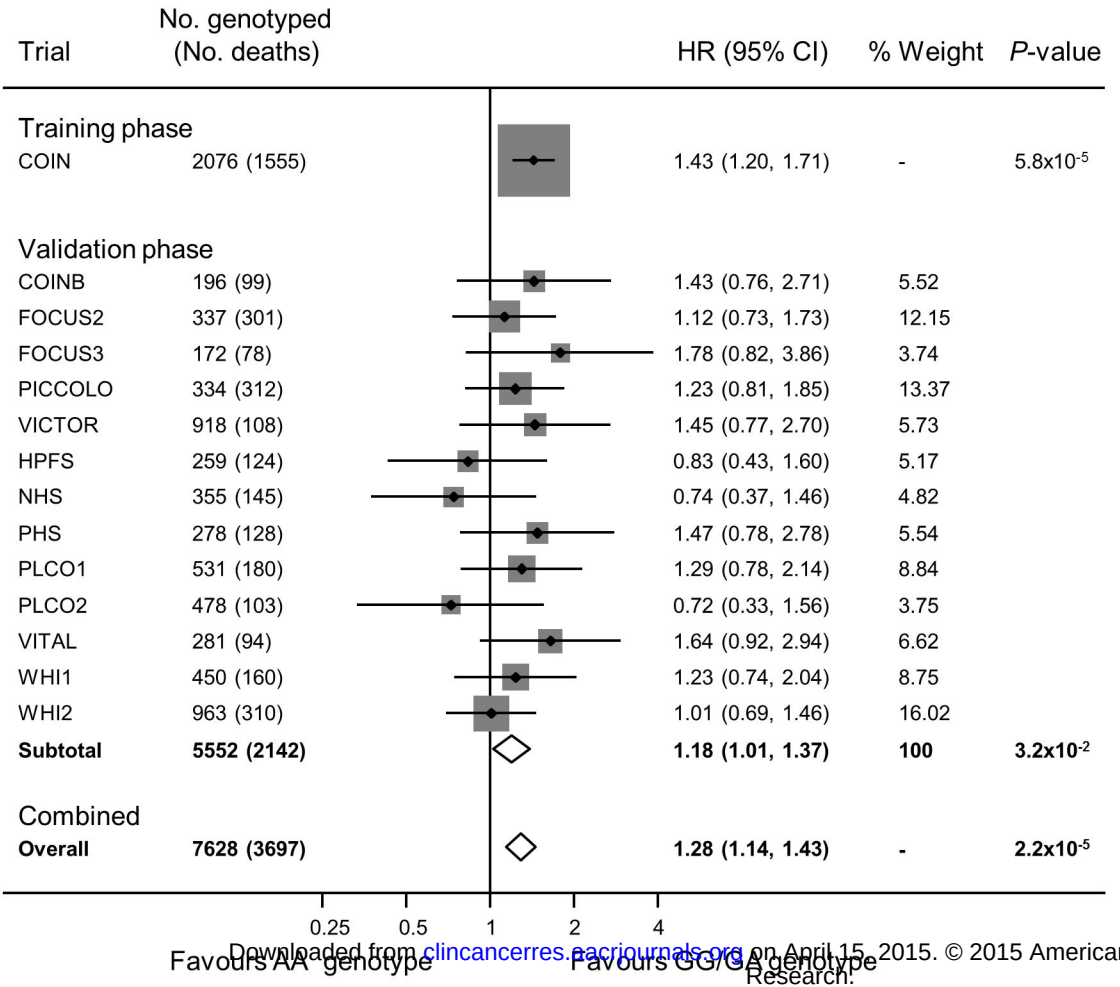
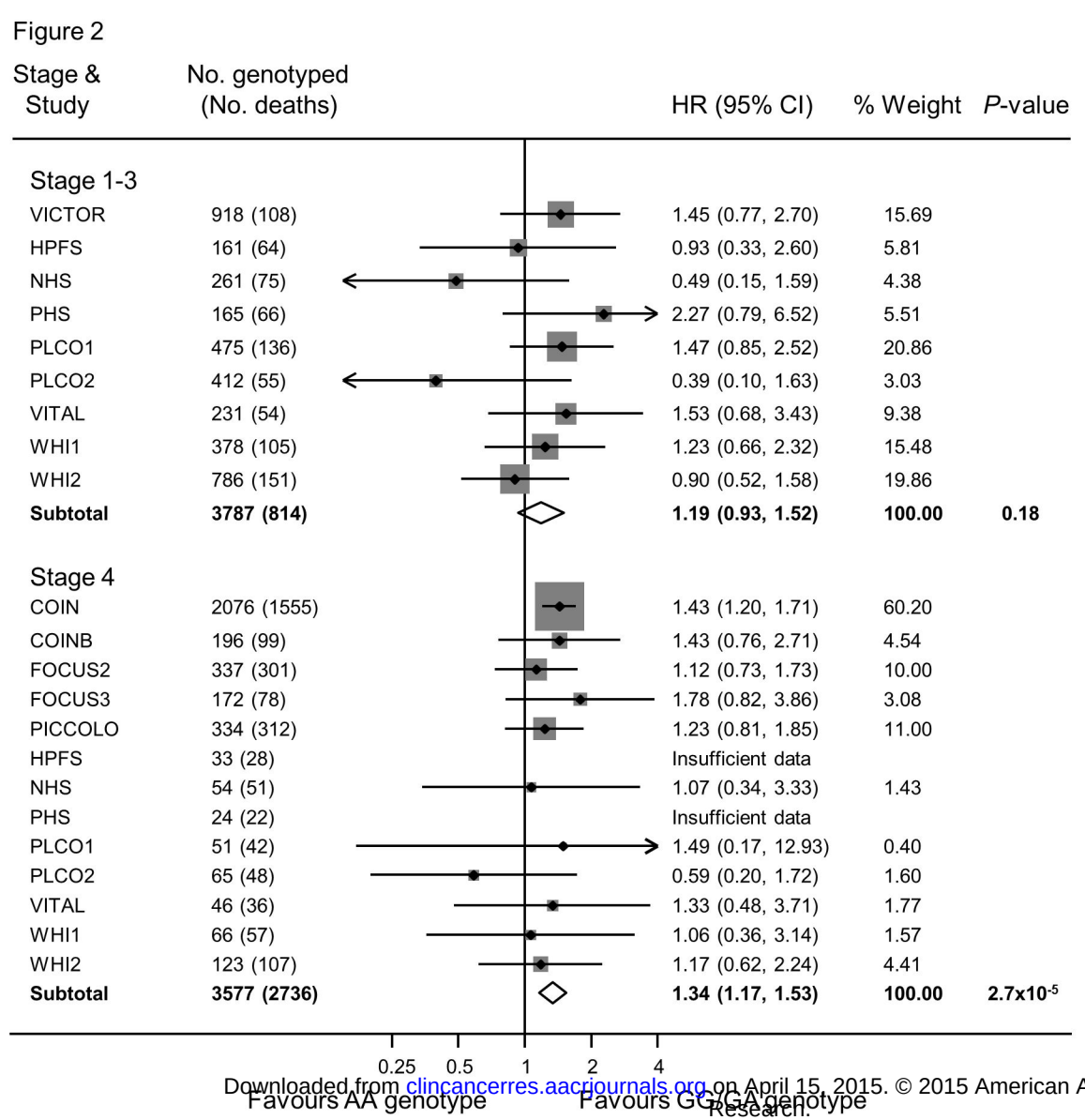


Figure 2



AACR American Association for Cancer Research

Clinical Cancer Research

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

Christopher G Smith, David Fisher, Rebecca Harris, et al.

Clin Cancer Res Published OnlineFirst April 14, 2015.