HHS Public Access

Author manuscript

Eur J Cancer. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:

Eur J Cancer. 2021 December ; 159: 247–258. doi:10.1016/j.ejca.2021.09.047.

A genome-wide search for determinants of survival in 1926 patients with advanced colorectal cancer with follow-up in over 22,000 patients

Christopher Wills[a], **Yazhou He**[b,c], **Matthew G. Summers**[a], **Yi Lin**[d,e], **Amanda I. Phipps**[d,e], **Katie Watts**[a], **Philip J. Law**[f], **Nada A. Al-Tassan**[g], **Timothy S. Maughan**[h], **Richard Kaplan**[i], **Richard S. Houlston**[f], **Ulrike Peters**[d,e], **Polly A. Newcomb**[d,e], **Andrew T. Chan**[j], **Daniel D. Buchanan**[k,l,m], **Steve Gallinger**[n,o,p], **Loic L. Marchand**[q], **Rish K. Pai**[r], **Qian Shi**[s], **Steven R. Alberts**[t], **Victoria Gray**[a], **Hannah D. West**[a], **Valentina Escott-Price**[u], **Malcolm G. Dunlop**[b], **Jeremy P. Cheadle**[a,*]

[a]Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK

[b]Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK

[c]Department of Oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, 610000, China

[d]Epidemiology Department, University of Washington, Seattle, WA, USA

[e]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[f]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP, UK

[g]Department of Genetics, King Faisal Specialist Hospital and Research Center, P.O. Box 3354, Riyadh 11211, Saudi Arabia

[h]CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Roosevelt Drive, Oxford OX3 7DQ, UK

[i]MRC Clinical Trials Unit, University College of London, 125 Kingsway, London, WC2B 6NH, UK

[*]*Corresponding author*. Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK. cheadlejp@cardiff.ac.uk (J.P. Cheadle).

Author contributions

JPC obtained funding for and directed this study. The study was designed by CW and JPC. TSM was CI of COIN and provided clinical advice and supported the translational research. RSK managed the COIN and COIN-B trials and facilitated access to the clinical data. NAA oversaw the genotyping of COIN and COIN-B. PJL and RSH oversaw the imputation and QC. YH and MGD provided data from SOCCS and YL, AIP, QS, SRA, UP, PAN, ATC, LLM, DDB, SG and RKP provided data from ISACC for replication analyses. CW undertook all of the GWAS statistical and meta-analyses with supervision from VEP and JPC, and with input from MGS. CW and JPC interpreted the data with input from KW, VG, HW and VEP. CW wrote the first draft of the paper with subsequent input from JPC, and all authors provided comments.

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2021.09.047.

[j]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[k]Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria, Australia

[l]Centre for Cancer Research, Victorian Comprehensive Cancer Centre, The University of Melbourne, Parkville, Victoria, Australia

[m]Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, Victoria, Australia

[n]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

[o]Ontario Institute for Cancer Research, Toronto, Ontario, Canada

[p]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

[q]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA

[r]Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, AZ, USA

[s]Department of Quantitative Science, Mayo Clinic, Rochester, MN, USA

[t]Division of Medical Oncology, Mayo Clinic, Rochester, MN, USA

[u]Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ, UK

Abstract

Background: While genome-wide association studies (GWAS) have identified germline variants influencing the risk of developing colorectal cancer (CRC), there has been limited examination of the possible role of inherited variation as a determinant of patient outcome.

Patients and methods: We performed a GWAS for overall survival (OS) in 1926 patients with advanced CRC from the COIN and COIN-B clinical trials. For single nucleotide polymorphisms (SNPs) showing an association with OS ($P < 1.0 \times 10^{-5}$), we conducted sensitivity analyses based on the time from diagnosis to death and sought independent replications in 5675 patients from the Study of Colorectal Cancer in Scotland (SOCCS) and 16,964 patients from the International Survival Analysis in Colorectal cancer Consortium (ISACC). We analysed the Human Protein Atlas to determine if *ERBB4* expression was associated with survival in 438 patients with colon adenocarcinomas.

Results: The most significant SNP associated with OS was rs79612564 in *ERBB4* (hazard ratio [HR] = 1.24, 95% confidence interval [CI] = 1.16–1.32, $P = 1.9 \times 10^{-7}$). SNPs at 17 loci had suggestive associations for OS and all had similar effects on the time from diagnosis to death. No lead SNPs were independently replicated in the meta-analysis of all patients from SOCCS and ISACC. However, rs79612564 was significant in stage-IV patients from SOCCS ($P = 2.1 \times 10^{-2}$) but not ISACC ($P = 0.89$) and SOCCS combined with COIN and COIN-B attained

genome-wide significance ($P = 1.7 \times 10^{-8}$). Patients with high *ERBB4* expression in their colon adenocarcinomas had worse survival (HR = 1.50, 95% CI = 1.1–1.9, $P = 4.6 \times 10^{-2}$).

Conclusions: Genetic and expression data support a potential role for rs79612564 in the receptor tyrosine kinase ERBB4 as a predictive biomarker of survival.

Keywords

Colorectal cancer; GWAS; Survival; Prognostic biomarkers

1. Introduction

Clinical stage, which combines the depth of tumour invasion, nodal status and distant metastasis [1], is currently the only routinely used marker of survival from colorectal cancer (CRC). Other factors thought to influence patient prognosis include lifestyle [2,3], systemic inflammatory response [4], immunologic microenvironment [5] and the patient's germline and the tumour's somatic profile [6,7]. The search for inherited prognostic factors has primarily focussed on candidate genes and single nucleotide polymorphisms (SNPs) that function in pharmacological pathways [8,9], influence tumour progression [10] or alter disease risk [11-16]. However, apart from rs9929218 in *CDH1*, most reported SNP associations have not been independently replicated [17].

Genome-wide association studies (GWAS) have been used successfully to identify 83 CRC-susceptibility alleles in the European population [18,19]. To date, the application of GWAS-based strategies for the identification of alleles influencing survival from CRC has been limited. SNPs near to *ELOVL5* and DCC have been associated with survival in a restricted discovery analysis but not replicated in the follow-up [20] and SNPs in *FHIT, EPHB1* and *MIR7515* have been associated with the time to metastasis but await independent replication [21]. Here, we report a GWAS of survival in 1926 patients with advanced CRC from two clinical trials with follow-up of promising SNP associations in over 22,000 CRC patients from clinical trials and population-based studies.

2. Materials and methods

2.1. Discovery GWAS

Unrelated patients with metastatic or locally advanced colorectal adenocarcinoma (n = 2671) were recruited for the MRC clinical trials COIN (NCT00182715) [22] and COIN-B (NCT00640081) [23]. COIN patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy, continuous chemotherapy with cetuximab or intermittent chemotherapy. COIN-B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab or intermittent chemotherapy and continuous cetuximab (Supplementary Figure S1). Patients from COIN and COIN-B were combined for survival analyses since there was no evidence of heterogeneity in overall survival (OS; the time from trial randomisation to death or end of the trial) between patients when analysed by trial ($P = 0.49$), trial arm ($P = 0.40$; Cochran Q test: p = 1.0, I^2 test: $P = 0.74$), type of chemotherapy received ($P = 0.60$) or cetuximab use ($P = 0.41$). Blood DNA samples were prepared from

2244 patients all of whom gave fully informed consent for bowel cancer research (approved by REC [04/MRE06/60]).

We genotyped DNA samples using Affymetrix Axiom Arrays according to the manufacturer's recommendations (Affymetrix, Santa Clara, CA 95051, USA) at the King Faisal Specialist Hospital and Research Center, Saudi Arabia (under IRB approval 2110033) [24]. After quality control (QC), 1950 patient samples remained for analyses, two of whom had no data on survival and were excluded (n = 1948, Supplementary Figure S1). Prediction of untyped SNPs was carried out using IMPUTE2 v2.3.0 [25] based on data from the 1000 Genomes Project as the reference [26,27]. In line with current GWAS guidelines [28,29], we excluded SNPs with minor allele frequencies (MAFs) < 5% or that had poor imputation scores (INFO score < 0.8, n = 29 million), missingness >0.02 (n = 3.5 million) or Hardy–Weinberg equilibrium (HWE) exact test [30] $P < 1.0 \times 10^{-6}$ (n = 47). After QC, 2.9 million SNPs remained. rs79612564 in *ERBB4* was independently genotyped by KASPar technology (LGC, Teddington, Middlesex, UK).

2.2. Statistical analysis

Somatic and clinicopathological factors available in COIN and COIN-B (trial, trial arm, cetuximab status, sex, age, *KRAS* status, *BRAF* status, *NRAS* status, MSI status, *PIK3CA* status, World Health Organisation [WHO] performance status, resection status of the primary tumour, site of the primary tumour, surface area, white blood cell [WBC] count, alkaline phosphatase level, platelet count, chemotherapy regimen, chemotherapy dose, radiotherapy, number of metastatic sites, liver metastases, lung metastases, nodal metastases, peritoneal metastases, other metastases, time to metastases, synchronous or metachronous metastases, creatinine clearance, glomerular filtration rate and carcinoembryonic antigen [CEA] level) were analysed for their effects on OS using either linear and logistic models (Supplementary Table S1). For those shown to be prognostic after Bonferroni correction ($P < 1.6 \times 10^{-3}$, n = 31 tests), we performed a GWAS for each factor to identify potential SNPs with pleiotropic effects on survival. Lead SNPs at credible independent loci (those with multiple SNPs in the linkage block and that reached the threshold for suggestive significance [$P < 1.0 \times 10^{-5}$]) were tested for their effects on OS.

We carried out a multivariate GWAS of OS under an additive model for patients in COIN and COIN-B using prognostic covariates that were available in the majority of patients (22 patients excluded leaving 1926 for analysis). The covariates included were WHO performance status, resection status of the primary tumour, WBC count, platelet count, alkaline phosphatase levels, number of metastatic sites, metastases in the liver, site of the primary tumour (encoded as seven binary variables), the surface area of the primary tumour, the time from diagnosis to metastases and metachronous versus synchronous metastases (Supplementary Table S1). For any SNPs that reached suggestive significance, we conducted a sensitivity analysis replacing OS (considered left-truncated at randomisation since randomisation is conditional upon survival from diagnosis) with time from diagnosis to death or end of the trial using Cox regressions. To test for differences in the association between the two measures of survival, for each SNP we calculated differences in beta-

coefficients and standard errors to produce a chi-squared distribution with 1 degree of freedom; from this, *P*-values were determined.

Gene and gene-set analyses were completed on the summary statistics from the association analysis to identify genes containing significant numbers of highly associated SNPs and significantly enriched gene-sets. The threshold for significance at the gene level was $P < 2.5 \times 10^{-6}$, with a Bonferroni correction for 20,000 independent tests [31]. Correction for multiple testing for gene-set analysis was completed by adjusting *P*-values for the false discovery rate to produce q-values [32,33], held to a significance threshold of $q < 0.05$.

Response at 12 weeks was assessed under a univariate dominant model and response was defined as a complete or partial response using RECIST 1.0 guidelines and no response was defined as stable or progressive disease.

2.3. Bioinformatic analyses

Discordant sex, individual/SNP missingness, heterozygosity, relatedness, principal component analysis, MAF and HWE quality control steps were performed using the *–sex-checks, –missing, –het, –genome, –pca* and *–hardy* commands in PLINK 1.9 (https://www.cog-genomics.org/plink2) [34] and clumping of GWAS summary statistics into independently associated loci was completed using the *–clump* command. INFO scores were obtained using SNPTEST v2.5.2 (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/). Linear and logistic SNP association tests were performed in PLINK v2.00a2 (https://www.cog-genomics.org/plink/2.0/). Regional association plots were created using LocusZoom (https://locuszoom.org). Multivariate OS analyses, genomic inflation factor calculation and Manhattan/quantile-quantile plots were performed using the *gwasurvivr* [35], *GenABEL* [36] and *qqman* R (https://www.r-project.org/) [37] packages, respectively.

Gene and gene-set analyses were performed using MAGMA [38] v1.07b (https://ctg.cncr.nl/software/magma). SNPs were annotated to genes (including those 35 kilobases before the gene transcription zone and ten kilobases after) using the *–annotate* command and the gene location file for hg19: 'NCBI37.3.loc'. SNP *P*-values were assessed with the linkage disequilibrium between them using the *multi = snp-wise* and *–gene-model* commands. This model takes advantage of the sum of the −log(*P*) for all SNPs, as well as the top SNP associations within each gene, to assess the association of their constituent genes. Genes were annotated to sets by gene-ontology terms [39] including experimental evidence, phylogenetically inferred annotation, computational analysis, author statement, curator statement and electronic annotation [40]. A competitive model (*–set-result* command) was used to assess each gene-set's association with OS. Expression quantitative trait loci analysis was completed by searching the Genotype-Tissue Expression project database (https://gtexportal.org/home/) [41] for significant associations between any relevant SNPs and gene expression.

The Human Protein Atlas [42] was used to find associations between *ERBB4* expression levels and survival in 438 patients with colon adenocarcinomas (+https://www.proteinatlas.org/ENSG00000178568-ERBB4/pathology/colorectal+cancer/COAD). RNA-seq data were reported as a median number of fragments

per kilobase of exon per million reads (FPKM) generated by The Cancer Genome Atlas. Samples were classified as high expression using a threshold of FPKM>0 as per The Human Protein Atlas recommendations [42].

2.4. Replication series

Independent replication of lead SNPs at 17 loci showing suggestive evidence of an association with OS in COIN and COIN-B was performed in two independent patient series:

(i) Study of Colorectal Cancer in Scotland—In total, 5675 patients (1358 CRC-specific deaths) of which 784 had stage-IV CRC (522 deaths) from the Study of Colorectal Cancer in Scotland (SOCCS; 1999–current [43,44]; ethics approval number MREC/01/0/5 obtained from the MultiCentre Research Ethics committee for Scotland) were included. Recruitment information, genotyping, QC and criteria for assigning the cause of death have been previously documented [45]. We considered CRC-specific survival, assigned as the time from diagnosis to death from CRC, and applied a Cox proportional hazards model and corrected for age, sex and AJCC stage.

(ii) International Survival Analysis in Colorectal cancer Consortium (ISACC).—In total, 16,964 patients (4010 deaths) of which 1847 had stage-IV CRC (1448 deaths) from ISACC which comprised 15 studies: the Cancer Prevention Study-II (CPS-II), the German Darmkrebs: Chancen der Verhutung durch Screening Study (DACHS), the Diet Activity and Lifestyle Study (DALS), the Early Detection Research Network (EDRN), the Swedish population of the European Prospective Investigation into Cancer (EPIC), the Health Professionals Follow-up Study (HPFS), the Melbourne Collaborative Cohort Study (MCCS), the Nurses' Health Study (NHS), the N9741 clinical trial, the Physician's Health Study (PHS), the Prostate, Lung, Colorectal, and Ovarian Study (PLCO), the UK Biobank (UKB), the VITamins And Lifestyle Study (VITAL), the Women's Health Initiative (WHI) and four Colon Cancer Family Registry (CCFR) sites: Seattle, Ontario, Australia, and the Mayo Clinic. References for each study are provided in the Supplementary Material. Study participants included individuals of European genetic ancestry diagnosed with CRC and with available genotyping and CRC-specific survival data. All participants provided informed consent for genetic testing, and all studies were approved by their respective Institutional Review Boards.

2.5. Meta-analyses of the follow-up cohorts

Meta-analyses were performed using the inverse variance-based method in the METAL software package [46]. $P < 0.05$ was considered significant for replication of the findings in the discovery cohort.

3. Results

We determined the influence of clinicopathological factors and somatic mutation status on OS in 1948 patients from COIN and COIN-B. We found that *KRAS* and *BRAF* mutation status, MSI status, platelet count, CEA levels, WHO performance status, resection status of the primary tumour, WBC count, alkaline phosphatase levels, number of metastatic sites,

metastases in the liver, lymph nodes and peritoneum, site and surface area of the primary tumour, time from diagnosis to metastases and metachronous versus synchronous metastases were all associated with OS after Bonferroni correction (Supplementary Table S1). We considered whether SNPs associated with these factors might influence OS and conducted independent GWASs for each factor (n = 16). One SNP was associated with WBC count (rs142358223 at 16p13.3, beta coefficient [beta] = 1.36, standard error [SE] = 0.25, $P = 3.5 \times 10^{-8}$) and two SNPs with CEA levels (rs17418475 at 1p21.2, beta = 932.53, SE = 163.05, $P = 1.3 \times 10^{-8}$ and rs72870425 at 2q24.2, beta = 1196.53, SE = 211.27, $P = 1.8 \times 10^{-8}$). We tested rs142358223, rs17418475, rs72870425 and 133 lead SNPs from other suggestive loci for their effects on OS; however, none were significant after adjustment for multiple testing ($P < 3.7 \times 10^{-4}$; Supplementary Table S2).

We carried out a multivariate GWAS for OS in 1926 patients from COIN and COIN-B using 11 prognostic covariates (Supplementary Figure S1, Fig. 1). No detectable genomic inflation was observed (1.08). We had >80% power to detect a HR of 1.3 for SNPs with MAFs ≥20%.

The most significant SNP associated with OS was rs79612564 in *ERBB4* (HR = 1.24, 95% CI = 1.16–1.32, $P = 1.9 \times 10^{-7}$). The median survival for patients in COIN and COIN-B carrying one minor allele was reduced by 46 days and for those homozygous for the minor allele by 81 days (Supplementary Figure S2, Supplementary Table S3). rs79612564 was not influenced by cetuximab treatment regardless of *KRAS* status (Supplementary Figure S3). The prognostic effect appeared to be independent of *KRAS* status and patients carrying at least one rs79612564 minor allele and *KRAS* mutant CRCs had the greatest effect on survival (HR = 1.51, CI = 1.29–1.77, $P = 3.7 \times 10^{-7}$) (Supplementary Figure S4). In terms of response to oxaliplatin and fluoropyrimidine-based chemotherapy, patients carrying one or more minor alleles showed less response (55.5% for heterozygotes and 55.9% for homozygotes) as compared to patients carrying both major alleles (60.2%), although this did not reach statistical significance ($P = 0.06$) (Supplementary Table S4). rs79612564 was not an eQTL.

rs79612564 had an INFO score of 0.99. We sought independent confirmation of the quality of genotyping and predictive score for this SNP by genotyping rs79612564 directly via KASPar technology. For those samples with both KASPar genotyping and an imputed genotype, we had >99% (1687/1703) genotype concordance (Supplementary Figure S5).

In total, we identified SNPs at 17 independent loci with suggestive associations with OS (Table 1, Fig. 1). We conducted a sensitivity analysis for lead SNPs at all 17 loci replacing OS with an alternative measure of survival – the time from diagnosis to death or end of the trial. There were no significant differences between the two measures of survival for any of the 17 SNPs ($P = 0.46$–0.95). rs6568761 at 6q21 (in a gene desert) passed the threshold for genome-wide significance ($P = 5.0 \times 10^{-8}$) with diagnosis to death (HR = 0.88, 95% CI = 0.78–0.98, $P = 4.5 \times 10^{-8}$).

We did not find any significantly associated genes (Supplementary Table S5) or gene-sets under competitive analyses (Supplementary Table S6) for OS after correction for multiple testing.

3.1. Replication analyses

We analysed lead SNPs at all 17 loci in 5675 patients with CRC from SOCCS and 16,964 patients with CRC from ISACC (Table 2, Fig. 2). Together, we had >98% power to replicate all 17 SNPs (Supplementary Table S7). After meta-analysis, no lead SNPs were independently replicated and only rs1352374 and rs2050337 reached nominal significance in SOCCS (Table 2).

We considered whether the lack of replication of the COIN and COIN-B data might be confounded by patients with different stages of disease in the follow-up cohorts. We therefore tested the 17 lead SNPs in a subset of 784 patients from SOCCS and 1847 patients from ISACC with stage-IV CRC (Table 3, Fig. 3). We had >80% power to replicate 16 of the SNPs (for rs3103204, we had 62% power) (Supplementary Table S7). rs79612564 was significant in stage-IV patients from SOCCS $P = 2.1 \times 10^{-2}$) but not in stage-IV patients from ISACC ($P = 0.89$, Table 3). When SOCCS was combined with COIN and COIN-B, rs79612564 reached genome-wide significance (HR = 1.22, 95% CI = 1.15–1.29, $P = 1.7 \times 10^{-8}$), but not when ISACC was also included (HR = 1.12, 95% CI = 1.06–1.17, $P = 3.4 \times 10^{-5}$).

rs6983214 was significant in the meta-analysis of stage-IV patients from SOCCS and ISACC ($P = 1.2 \times 10^{-3}$), however, the direction of effect was opposite to that found in COIN and COIN-B (Table 3). rs1352374 reached nominal significance in SOCCS ($P = 3.3 \times 10^{-2}$), but not in ISACC. rs2050337 reached nominal significance in the meta-analysis ($P = 1.1 \times 10^{-2}$, Table 3) with the same direction of effect in all cohorts tested (meta-analysis with COIN and COIN-B included HR = 1.13, 95% CI = 1.08–1.18, $P = 1.6 \times 10^{-6}$).

3.2. Relationship between ERBB4 expression and survival

We sought additional mechanistic data for a role for *ERBB4* in survival by studying 438 patients with colon adenocarcinomas from the Human Protein Atlas. Patients with high *ERBB4* expression in their tumours had worse survival (Cox-regression HR = 1.50, 95% CI = 1.10–1.90, $P = 4.6 \times 10^{-2}$, Supplementary Figure S6).

4. Discussion

Despite identifying 18 somatic and clinicopathological factors that significantly influenced survival in COIN and COIN-B, we found that SNPs associated with these factors did not themselves affect survival thereby excluding potential pleiotropic effects. To generate a comprehensive genome-wide analysis of survival, we included prognostic factors into our multivariate analyses and observed little genomic inflation supporting the validity of this approach.

The most significant SNP identified was rs79612564, which lies within intron 3 of *ERBB4*, a member of the epidermal growth factor receptor subfamily. We confirmed the quality of the genotyping and imputation for this SNP via an independent assay. Patients carrying the minor allele had an additive effect on survival with a median decrease in life expectancy of approximately 40 days per allele carried in the advanced disease setting. rs79612564 was also significant in stage-IV patients from SOCCS and, combined with COIN and COIN-B,

reached genome-wide significance. Our genetic data were supported by mechanistic data for this gene and we found that patients with high *ERBB4* expression in their colon adenocarcinomas had worse survival. Furthermore, it has previously been shown that *ERBB4* overexpression in experimental systems enhances the survival and growth of cells driven by *Ras* and/or *Wnt* signalling [47].

However, rs79612564 was not replicated in stage-IV patients from ISACC, nor in all patients from SOCCS and ISACC combined. This warrants further investigation although it is noteworthy that overexpression and heterodimerisation of ERBB4 and ERBB2 show a significant association with late-stage colorectal carcinomas [48]. Therefore, it is possible that the association for rs79612564 can only be seen in patients with later stages of disease and survival in these patients is confounded by numerous clinical and pathological prognostic covariates, which we accounted for in our GWAS but are, in general, not available in the population-based cohorts.

In terms of clinical application, it should be noted that the effect size for rs79612564 is modest and will need to be combined with other prognostic factors to have any role in patient management. For example, our data suggest that this SNP acts independently of *KRAS* mutational status, which itself is a prognostic factor. In isolation, rs79612564 has an OR of 1.24 but on a *KRAS* mutant background increases to 1.51. Although this effect size is still modest, it shows the potential for building germline, somatic and clinicopathological factors into a combined prognostic model.

Most of the other loci of interest failed to be replicated or their directions of effect were opposite to those found in our discovery cohort. However, rs2050337 at 10q25.1 reached significance in the stage-IV replication meta-analysis with a consistent direction of effect to COIN and COIN-B, and was also significant in all patients from SOCCS. It lies approximately 500 Kb upstream of *ADD3* which has been associated with tumour growth and cell migration in breast [49], glioblastoma multiforme [50] and lung cancer [51]. However, even combined with COIN and COIN-B, rs2050337 still did not achieve genome-wide significance in patients with stage-IV disease suggesting that its effects, if genuine, are modest.

Despite having 1926 patients with advanced CRC (with a 75% event rate) in our GWAS, we lacked sufficient power to detect common alleles with low effect sizes (HR < 1.3) at genome-wide significance levels. Even by considering loci at suggestive significance levels, as we have done, we only had 33% power to detect common alleles with HRs of 1.2. Future studies will therefore have to combine their datasets for meta-analyses to provide sufficient power to identify low impact alleles for survival. For example, to achieve 80% power to detect alleles with HRs of 1.2 and 1.1, 4907 and 18,022 patients with a 75% event rate would be required, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Hywel Williams for helpful advice, the patients and their families who participated and gave their consent for this research and the investigators throughout the UK who submitted samples for assessment. This research was conducted using the UK Biobank Resource. Acknowledgements for ISACC are provided in the Supplementary Material. None of the sponsors played a role in the study design; the collection, analysis and interpretation of data; the writing of the report and the decision to submit the paper for publication.

Financial support

This work was supported by Tenovus Cancer Care and Cancer Research Wales. NAA was funded and supported by KFSHRC. 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, the Medical Research Council and an unrestricted educational grant from Merck-Serono and were conducted with the support of the National Institute of Health Research Cancer Research Network. Sample and covariate data collection, generation of primary genotyping data and analysis related to SOCCS were supported by a Cancer Research UK Programme Grant (C348/A18927) and a Project Leader Grant to MGD (MRC Human Genetics Unit Centre Grant – U127527198). MGD is an MRC Investigator. The work was also supported by infrastructure and staffing of the Edinburgh Cancer Research UK Cancer Research Centre. ATC is a Stuart and Suzanne Steele MGH Research Scholar. For financial support for ISACC, please see Supplementary Materials.

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(7):489–99. 10.1038/nrc2645. [PubMed: 19536109]

[2]. Haydon AMM, Macinnis RJ, English DR, Giles GG. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. Gut 2006;55(1):62–7. 10.1136/gut.2005.068189. [PubMed: 15972299]

[3]. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 2007;335(7630):1134. 10.n36/bmj.39367.495995.AE. [PubMed: 17986716]

[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(9):1266–70. 10.1038/sj.bjc.6604027. [PubMed: 17923866]

[5]. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313(5795):1960–4. 10.n26/science.1129139. [PubMed: 17008531]

[6]. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 2005;23(3):609–18. 10.1200/jco.2005.01.086. [PubMed: 15659508]

[7]. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 2008;57(7):941–50. 10.1136/gut.2007.135004. [PubMed: 18364437]

[8]. Dotor E, Cuatrecases M, Martinez-Iniesta M, Navarro M, Vilardell F, Guino 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(10):1603–11. 10.1200/jco.2005.03.5253. [PubMed: 16575011]

[9]. Marcuello E, Altes A, del Rio E, Cesar 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(5):733–7. 10.1002/ijc.20487. [PubMed: 15386371]

[10]. 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(1):62–6. 10.1158/1078-0432.Ccr-07-1537. [PubMed: 18172253]

[11]. 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(1):51–4. 10.1053/j.gastro.2012.04.052. [PubMed: 22580541]

[12]. 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(7):1327–31. 10.1093/carcin/bgs147. [PubMed: 22505654]

[13]. 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(2):292–8. 10.1093/carcin/bgs335. [PubMed: 23104301]

[14]. Abuli A, Lozano JJ, Rodriguez-Soler M, Jover R, Bessa X, Munoz J, et al. Genetic susceptibility variants associated with colorectal cancer prognosis. Carcinogenesis 2013;34(10):2286–91. 10.1093/carcin/bgt179. [PubMed: 23712746]

[15]. 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(4):1395–402. 10.1245/s10434-012-2657-z. [PubMed: 22976378]

[16]. 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(2):117816. 10.1371/journal.pone.0117816.

[17]. Smith CG, Fisher D, Harris R, Maughan TS, Phipps AI, Richman S, 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 Cancer Res 2015;21(15):3453–61. 10.1158/1078-0432.Ccr-14-3136. [PubMed: 25873087]

[18]. Houlston RS. COGENT (COlorectal cancer GENeTics) revisited. Mutagenesis 2012;27(2):143–51. 10.1093/mutage/ger059. [PubMed: 22294761]

[19]. 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 Commun 2019;10(1):2154. 10.1038/s41467-019-09775-w. [PubMed: 31089142]

[20]. Phipps AI, Passarelli MN, Chan AT, Harrison TA, Jeon J, Hutter CM, et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. Carcinogenesis 2016;37(1):87–95. 10.1093/carcin/bgv161. [PubMed: 26586795]

[21]. Penney ME, Parfrey PS, Savas S, Yilmaz YE. A genome-wide association study identifies single nucleotide polymorphisms associated with time-to-metastasis in colorectal cancer. BMC Cancer 2019;19(1):133. 10.1186/s12885-019-5346-5. [PubMed: 30738427]

[22]. Maughan TS, Adams RA, Smith CG, Meade AM, Seymour MT, Wilson RH, et al. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. Lancet 2011;377(9783):2103–14. 10.1016/s0140-6736(11)60613-2. [PubMed: 21641636]

[23]. 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(6):631–9. 10.1016/s1470-2045(14)70106-8. [PubMed: 24703531]

[24]. 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. 10.1038/srep10442. [PubMed: 25990418]

[25]. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5(6):1000529. 10.1371/journal.pgen.1000529.

[26]. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. A global reference for human genetic variation. Nature 2015;526(7571):68. 10.1038/nature15393. [PubMed: 26432245]

[27]. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3: Genes|Genomes|Genetics 2011;1(6):457. 10.1534/g3.111.001198. [PubMed: 22384356]

[28]. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int J Methods Psychiatr Res 2018;27(2):1608. 10.1002/mpr.1608.

[29]. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, et al. Quality control procedures for genome-wide association studies. Curr Protoc Human Genet 2011. 10.1002/0471142905.hg0119s68. Chapter 1:Unit1.19-Unit1.

[30]. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of hardy-weinberg Equilibrium. Am J Hum Genet 2005;76(5):887–93. 10.1086/429864. [PubMed: 15789306]

[31]. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. Nat Genet 2012;44(6):623–30. 10.1038/ng.2303. [PubMed: 22641211]

[32]. Storey JD. A direct approach to false discovery rates. J Roy Stat Soc B Stat Methodol 2002;64:479–98. 10.1111/1467-9868.00346.

[33]. Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control. 2019.

[34]. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559–75. 10.1086/519795. [PubMed: 17701901]

[35]. Rizvi AA, Karaesmen E, Morgan M, Preus L, Wang J, Sovic M, et al. gwasurvivr: an R package for genome-wide survival analysis. Bioinformatics 2019;35(11):1968–70. 10.1093/bioinformatics/bty920. [PubMed: 30395168]

[36]. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23(10):1294–6. 10.1093/bioinformatics/btm108. [PubMed: 17384015]

[37]. R_Core_Team. R: A language and environment for statistical computing. 2018.

[38]. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11(4):19. 10.1371/journal.pcbi.1004219.

[39]. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000;25(1):25–9. 10.1038/75556. [PubMed: 10802651]

[40]. Balakrishnan R, Harris MA, Huntley R, Van Auken K, Cherry JM. A guide to best practices for Gene Ontology (GO) manual annotation. Database – J Biol Database Curation 2013;2013:bat054–bat. 10.1093/database/bat054.

[41]. Carithers LJ, Moore HM. The genotype-tissue expression (GTEx) project. Biopreserv Biobanking 2015;13(5):307–8. 10.1089/bio.2015.29031.hmm.

[42]. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science 2015;347(6220):10. 10.1126/science.1260419. [PubMed: 25554765]

[43]. Theodoratou E, Kyle J, Cetnarskyj R, Farrington SM, Tenesa A, Barnetson R, et al. Dietary flavonoids and the risk of colorectal cancer. Cancer Epidemiol Biomark Prev 2007;16(4):684–93. 10.1158/1055-9965.epi-06-0785.

[44]. He Y, Theodoratou E, Li X, Din FVN, Vaughan-Shaw P, Svinti V, et al. Effects of common genetic variants associated with colorectal cancer risk on survival outcomes after diagnosis: a large population-based cohort study. Int J Cancer 2019;145(9):2427–32. 10.1002/ijc.32550.117 [PubMed: 31271446]

[45]. Theodoratou E, Farrington SM, Timofeeva M, Din FVN, Svinti V, Tenesa A, et al. Genome-wide scan of the effect of common nsSNPs on colorectal cancer survival outcome. Br J Cancer 2018;119(8):988–93. 10.1038/s41416-018-0117-7. [PubMed: 30135471]

[46]. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. Bioinformatics 2010;26(17):2190–1. 10.1093/bioinformatics/btq340. [PubMed: 20616382]

[47]. Williams CS, Bernard JK, Demory Beckler M, Almohazey D, Washington MK, Smith JJ, et al. ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. Carcinogenesis 2015;36(7):710–8. 10.1093/carcin/bgv049. [PubMed: 25916654]

[48]. Lee JC, Wang ST, Chow NH, Yang HB. Investigation of the prognostic value of coexpressed erbB family members for the survival of colorectal cancer patients after curative surgery. Eur J Cancer 2002;38(8):1065–71. 10.1016/s0959-8049(02)00004-7. [PubMed: 12008194]

[49]. Yang LF, Yang F, Zhang FL, Xie YF, Hu ZX, Huang SL, et al. Discrete functional and mechanistic roles of chromodomain Y-like 2 (CDYL2) transcript variants in breast cancer growth and metastasis. Theranostics 2020;10(12):5242–58. 10.7150/thno.43744. [PubMed: 32373210]

[50]. Kiang KMY, Zhang PD, Li N, Zhu ZY, Jin L, Leung GKK. Loss of cytoskeleton protein ADD3 promotes tumor growth and angiogenesis in glioblastoma multiforme. Cancer Lett 2020;474:118–26. 10.1016/j.canlet.2020.01.007. [PubMed: 31958485]

[51]. Lechuga S, Amin PH, Wolen AR, Ivanov AI. Adducins inhibit lung cancer cell migration through mechanisms involving regulation of cell-matrix adhesion and cadherin-11 expression. Biochim Biophys Acta Mol Cell Res 2019;1866(3):395–408. 10.1016/j.bbamcr.2018.10.001. [PubMed: 30290240]

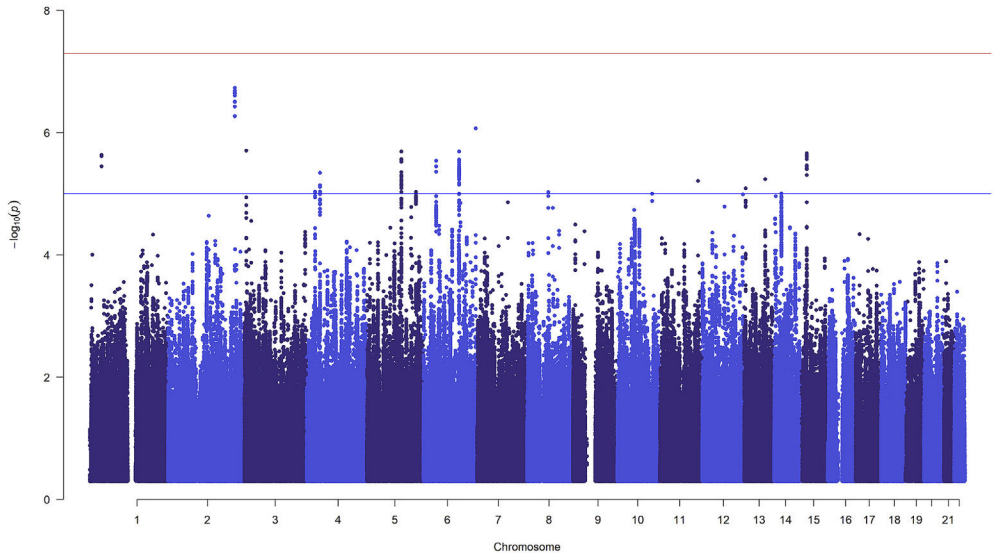

Fig. 1. Manhattan plot of SNP associations with overall survival (OS) (n = 1926 patients with advanced CRC from COIN and COIN-B).

SNPs are ordered by chromosome position and plotted against the −log10(*P*) for their association with OS. The red line represents the threshold for genome-wide significance ($P = 5.0 \times 10^{-8}$) and the blue line is the threshold for suggestive significance ($P = 1.0 \times 10^{-5}$). Covariates included: World Health Organisation performance status, resection status of the primary tumour, white blood cell count, platelet count, alkaline phosphatase levels, number of metastatic sites, metastases within or outside of the liver, site of the primary tumour, surface area of the primary tumour, time from diagnosis to metastases and metachronous versus synchronous metastases.

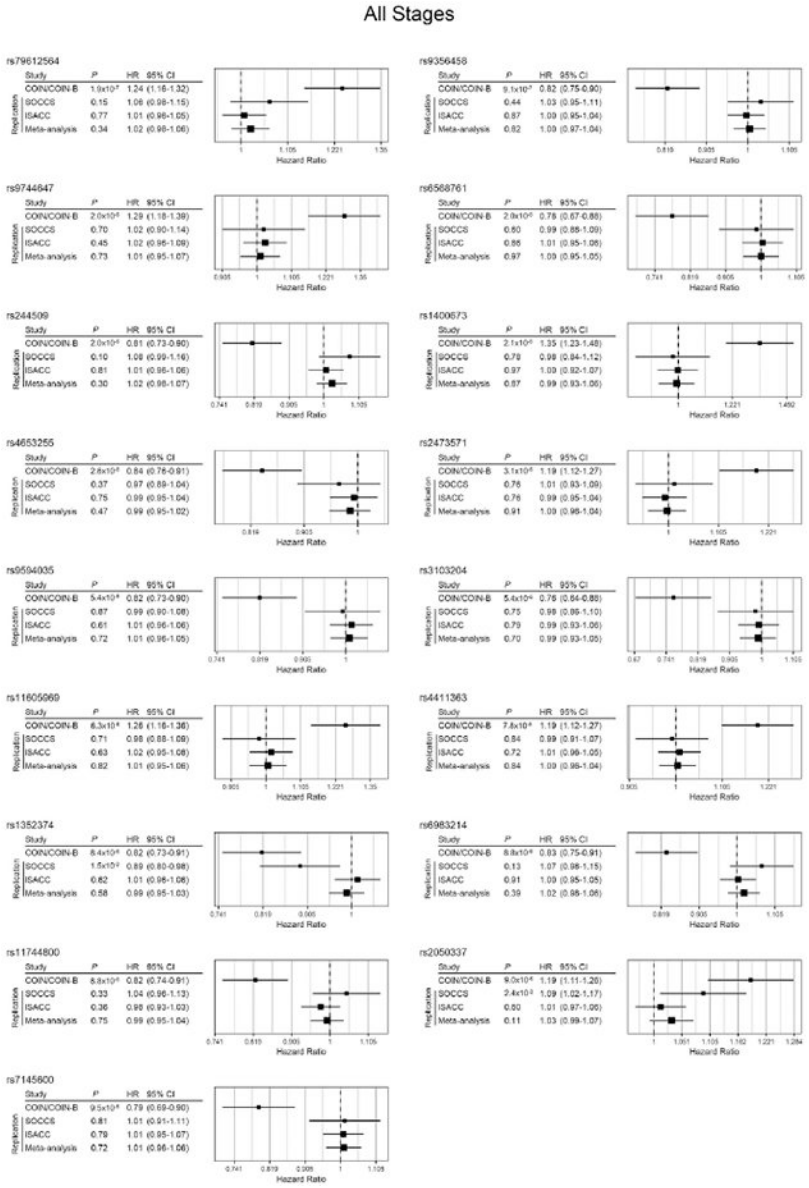

Fig. 2. Forest plots for lead SNPs at 17 loci identified in COIN and COIN-B and the independent replication cohorts (all stages).

Sample size, number of events, *P*-value, hazard ratio and 95% confidence intervals are listed.

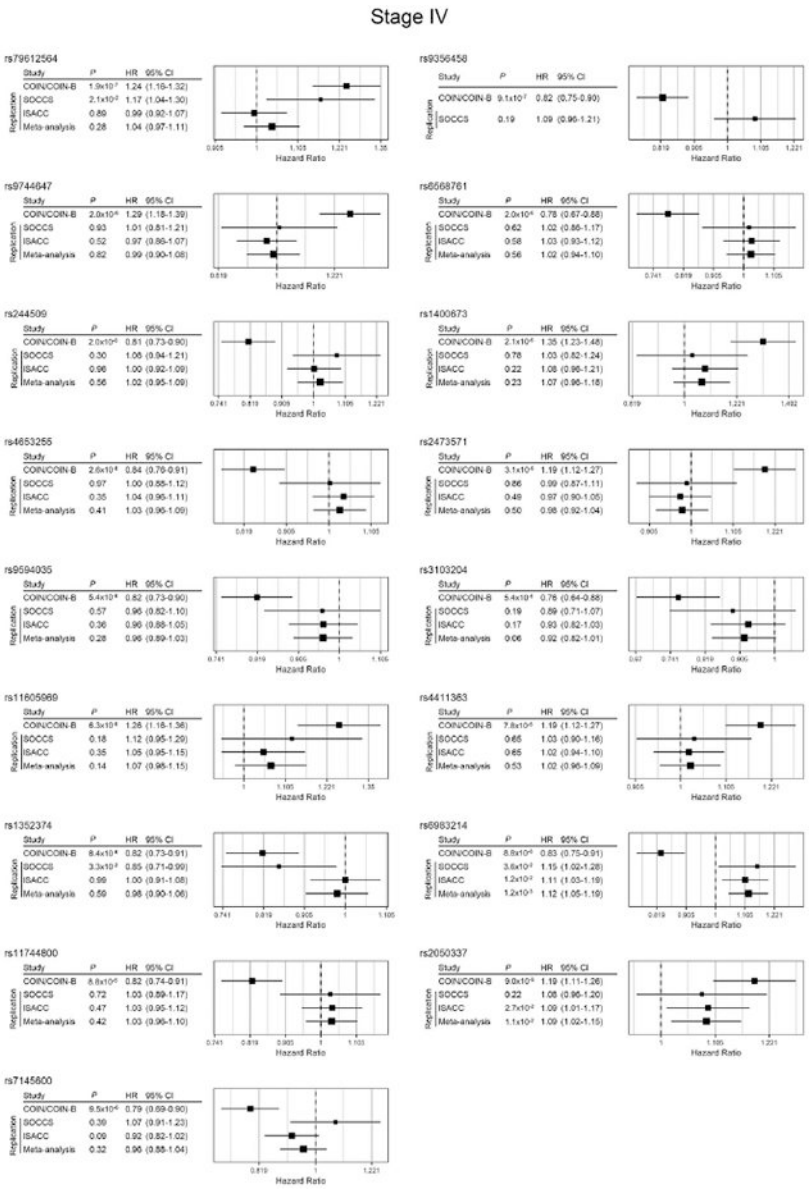

Fig. 3. Forest plots for lead SNPs at 17 loci identified in COIN and COIN-B and the independent replication cohorts (stage-IV disease).

Sample size, number of events, *P*-value, hazard ratio and 95% confidence intervals are listed.

Table 1

Lead SNPs from independent loci that reached suggestive significance in a multivariate analysis of overall survival in COIN and COIN-B.

Cytogenic band, minor allele, *P*-value, hazard ratio and 95% confidence intervals are shown for overall survival (the time from trial recruitment to death or end of the study) and time from diagnosis to death or end of the trial. Only rs6568761 reached the threshold for genome-wide significance ($P < 5.0 \times 10^{-8}$, in bold). Genes overlapping with the SNPs attributed to each locus are listed.

SNP	Locus	Minor allele	Genes	Overall survival			Diagnosis to death		
				HR	95% CI	*P*	HR	95% CI	*P*
rs79612564	2q34	C	*ERBB4*	1.24	1.16–1.32	1.9×10^{-7}	1.08	1.00–1.16	4.7×10^{-5}
rs9356458	6q27	A		0.82	0.75–0.90	9.1×10^{-7}	0.92	0.85–1.00	1.1×10^{-5}
rs9744647	15q14	T	*C145orf51*	1.29	1.18–1.39	2.0×10^{-6}	1.11	1.03–1.20	4.3×10^{-6}
rs6568761	6q21	G		0.78	0.67–0.88	2.0×10^{-6}	0.88	0.78–0.98	**4.5×10^{-8}**
rs244509	5q22.1	C	*CAMK4*	0.81	0.73–0.90	2.0×10^{-6}	0.91	0.83–0.99	1.0×10^{-6}
rs1400673	3p25.1	G		1.35	1.23–1.48	2.1×10^{-6}	1.13	1.01–1.25	1.4×10^{-5}
rs4653255	1p34.3	A		0.84	0.76–0.91	2.6×10^{-6}	0.94	0.86–1.02	1.1×10^{-4}
rs2473571	6p21.1	G	*LRFN2*	1.19	1.12–1.27	3.1×10^{-6}	1.06	0.98–1.14	3.5×10^{-4}
rs9594035	13q31.1	T		0.82	0.73–0.90	5.4×10^{-6}	0.92	0.84–1.00	5.4×10^{-6}
rs3103204	4p13	T	*ATP8A1, SHISA3*	0.76	0.64–0.88	5.4×10^{-6}	0.89	0.78–1.01	2.5×10^{-5}
rs11605969	11q24.1	T	*SORL1*	1.26	1.16–1.36	6.3×10^{-6}	1.08	0.98–1.18	3.3×10^{-4}
rs4411363	13q12.12	G	*TNFRSF19*	1.19	1.12–1.27	7.8×10^{-6}	1.06	0.98–1.14	1.1×10^{-3}
rs1352374	4p15.2	C		0.82	0.73–0.91	00×10^{-6}	0.92	0.80–1.03	1.8×10^{-5}
rs6983214	8q13.1	T	*C8orf44, C8orf44-SGK3, VCPIP1*	0.83	0.75–0.91	8.8×10^{-6}	0.92	0.84–1.00	4.9×10^{-6}
rs11744800	5q33.3	C	*ADAM19*	0.82	0.74–0.91	8.8×10^{-6}	0.93	0.85–1.01	3.5×10^{-4}
rs2050337	10q25.1	G		1.19	1.11–1.26	9.0×10^{-6}	1.07	0.99–1.15	6.5×10^{-5}
rs7145600	14q21.1	T		0.79	0.69–0.90	9.5×10^{-6}	0.91	0.81–1.01	5.2×10^{-5}

Table 2

Independent replication of lead SNPs in SOCCS and ISACC.

Hazard Ratio, 95% confidence intervals and *P*-value are listed for overall survival (the time from trial recruitment to death or end of the study) in COIN and COIN-B, and CRC-specific survival (the time from diagnosis to death due to CRC) in SOCCS and ISACC. Nominally significant *P*-values are highlighted in bold.

SNP			Independent replication						
	COIN and COIN-B 1926 patients (1435 deaths)		SOCCS 5675 patients (1358 deaths)			ISACC 16,964 patients (4010 deaths)			Meta
	HR	95% CI	HR	95% CI	*P*	HR	95% CI	*P*	*P*
rs79612564	1.24	1.16–1.32	1.06	0.98–1.15	0.15	1.01	0.96–1.05	0.77	0.34
rs9356458	0.82	0.75–0.90	1.03	0.95–1.11	0.44	1.00	0.95–1.04	0.87	0.82
rs9744647	1.29	1.18–1.39	1.02	0.90–1.14	0.70	1.02	0.96–1.09	0.45	0.73
rs6568761	0.78	0.67–0.88	0.99	0.88–1.09	0.60	1.01	0.95–1.06	0.86	0.97
rs244509	0.81	0.73–0.90	1.08	0.99–1.16	0.10	1.01	0.96–1.06	0.81	0.30
rs1400673	1.35	1.23–1.48	0.98	0.84–1.12	0.78	1.00	0.92–1.07	0.97	0.87
rs4653255	0.84	0.76–0.91	0.97	0.89–1.04	0.37	0.99	0.95–1.04	0.75	0.47
rs2473571	1.19	1.12–1.27	1.01	0.93–1.09	0.76	0.99	0.95–1.04	0.76	0.91
rs9594035	0.82	0.73–0.90	0.99	0.90–1.08	0.87	1.01	0.96–1.06	0.61	0.72
rs3103204	0.76	0.64–0.88	0.98	0.86–1.10	0.75	0.99	0.93–1.06	0.79	0.70
rs11605969	1.26	1.16–1.36	0.98	0.88–1.09	0.71	1.02	0.95–1.08	0.63	0.82
rs4411363	1.19	1.12–1.27	0.99	0.91–1.07	0.84	1.01	0.96–1.05	0.72	0.84
rs1352374	0.82	0.73–0.91	0.89	0.80–0.98	**1.5×10^{-2}**	1.01	0.96–1.06	0.62	0.58
rs6983214	0.83	0.75–0.91	1.07	0.98–1.15	0.13	1.00	0.95–1.05	0.91	0.39
rs11744800	0.82	0.74–0.91	1.04	0.96–1.13	0.33	0.98	0.93–1.03	0.36	0.75
rs2050337	1.19	1.11 −1.26	1.09	1.02–1.17	**2.4×10^{-2}**	1.01	0.97–1.06	0.60	0.11
rs7145600	0.79	0.69–0.90	1.01	0.91–1.11	0.81	1.01	0.95–1.07	0.79	0.72

Table 3

Independent replication of lead SNPs in patients from SOCCS and ISACC with stage-IV CRC.

Hazard ratio, 95% confidence intervals and *P*-value are listed for overall survival (the time from trial recruitment to death or end of the study) in COIN and COIN-B, and CRC-specific survival (the time from diagnosis to death due to CRC) in SOCCS and ISACC. Nominally significant *P*-values are highlighted in bold. *Opposite direction of effect to COIN and COIN-B, so not validated. Data for rs9356458, nor any proxies, were available for stage-IV patients from ISACC.

SNP			Independent replication						
	COIN and COIN-B 1926 patients (1435 deaths)		SOCCS stage-IV 784 patients (522 deaths)			ISACC stage-IV 1847 patients (1448 deaths)			Meta
	HR	95% CI	HR	95% CI	*P*	HR	95% CI	*P*	*P*
rs79612564	1.24	1.16–1.32	1.17	1.04–1.30	**2.1×10^{-2}**	0.99	0.92–1.07	0.89	0.28
rs9356458	0.82	0.75–0.90	1.03	0.96–1.21	0.19	–	–	–	–
rs9744647	1.29	1.18–1.39	1.01	0.81–1.21	0.93	0.97	0.86–1.07	0.52	0.82
rs6568761	0.78	0.67–0.88	1.02	0.86–1.17	0.62	1.03	0.93–1.12	0.58	0.56
rs244509	0.81	0.73–0.90	1.08	0.94–1.21	0.30	1.00	0.92–1.09	0.96	0.56
rs1400673	1.35	1.23–1.48	1.03	0.82–1.24	0.78	1.08	0.96–1.21	0.22	0.23
rs4653255	0.84	0.76–0.91	1.00	0.88–1.12	0.97	1.04	0.96–1.11	0.35	0.41
rs2473571	1.19	1.12–1.27	0.99	0.87–1.11	0.86	0.97	0.90–1.05	0.49	0.50
rs9594035	0.82	0.73–0.90	0.96	0.82–1.10	0.57	0.96	0.88–1.05	0.36	0.28
rs3103204	0.76	0.64–0.88	0.89	0.71–1.07	0.19	0.93	0.82–1.03	0.17	0.06
rs11605969	1.26	1.16–1.36	1.12	0.95–1.29	0.18	1.05	0.95–1.15	0.35	0.14
rs4411363	1.19	1.12–1.27	1.03	0.90–1.16	0.65	1.02	0.94–1.10	0.65	0.53
rs1352374	0.82	0.73–0.91	0.85	0.71–0.99	**3.3×10^{-2}**	1.00	0.91–1.08	0.99	0.59
rs6983214	0.83	0.75–0.91	1.15	1.02–1.28	**3.6×10^{-2}**	1.11	1.03–1.19	**1.2×10^{-2}**	**1.2×10^{-3}***
rs11744800	0.82	0.74–0.91	1.03	0.89–1.17	0.72	1.03	0.95–1.12	0.47	0.42
rs2050337	1.19	1.11 −1.26	1.08	0.96–1.20	0.22	1.09	1.01–1.17	**2.7×10^{-2}**	**1.1×10^{-2}**
rs7145600	0.79	0.69–0.90	1.07	0.91–1.23	0.39	0.92	0.82–1.02	0.09	0.32