



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Colorectal cancer risk variants rs10161980 and rs7495132 are associated with cancer survival outcome by a recessive mode of inheritance.**

**Citation for published version:**

He, Y, Timofeeva, M, Zhang, X, Xu, W, Din, FVN, Svinti MacLeod, V, Farrington, SM, Campbell, H, Dunlop, MG & Theodoratou, E 2021, 'Colorectal cancer risk variants rs10161980 and rs7495132 are associated with cancer survival outcome by a recessive mode of inheritance.', *International Journal of Cancer*.  
<https://doi.org/10.1002/ijc.33465>

**Digital Object Identifier (DOI):**

[10.1002/ijc.33465](https://doi.org/10.1002/ijc.33465)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

International Journal of Cancer

**Publisher Rights Statement:**

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi:  
[10.1002/ijc.33465](https://doi.org/10.1002/ijc.33465)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



**Colorectal cancer risk variants rs10161980 and rs7495132 are associated with cancer survival outcome by a recessive mode of inheritance.**

**Running title:** Genetic variants and colorectal cancer survival

**Authors:** Yazhou He<sup>1,2,3,4</sup> MD,PhD, Maria Timofeeva<sup>3,5</sup> PhD, Xiaomeng Zhang<sup>4</sup> MSc, Wei Xu<sup>4</sup> MSc, Xue Li<sup>4,6</sup> PhD, Farhat VN. Din<sup>2,3</sup> MD, Victoria Svinti<sup>3</sup> PhD, Susan M. Farrington<sup>2,3</sup> PhD, Harry Campbell<sup>4</sup> MD, Malcolm G. Dunlop<sup>2,3\*</sup> MD, Evropi Theodoratou<sup>2,4\*</sup> PhD

**Author affiliations:**

1. Department of Oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, P.R. China
2. Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics & Molecular Medicine, Western General Hospital, The University of Edinburgh, Edinburgh, UK
3. Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Medical Research Council Institute of Genetics & Molecular Medicine, Western General Hospital, The University of Edinburgh, Edinburgh, UK
4. Centre for Global Health, Usher Institute, The University of Edinburgh, Edinburgh, UK
5. Danish Institute for Advanced Study (DIAS), Department of Public Health, University of Southern Denmark, Odense, Denmark.
6. School of Public Health, Zhejiang University, Hangzhou, P.R. China

**\*Corresponding authors:**

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/ijc.33465](https://doi.org/10.1002/ijc.33465)

Malcolm G. Dunlop, Institute of Genetics and Molecular Medicine, The University of  
Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK, Tel +44 (0) 131 651 8631,

Email address: [malcolm.dunlop@ed.ac.uk](mailto:malcolm.dunlop@ed.ac.uk)

Evropi Theodoratou, Centre for Global Health Research, Usher Institute, University of  
Edinburgh, Teviot Place, Edinburgh, EH8 9AG, United Kingdom; Tel: (+44) 131-650-6194;

E-mail: [E.Theodoratou@ed.ac.uk](mailto:E.Theodoratou@ed.ac.uk)

Accepted Article

**Key words:** Colorectal cancer; Germline genetic variants; Survival outcomes; Recessive model

**Abbreviations:**

AJCC, American Joint Committee on Cancer; CI, confidence interval; CRC, colorectal cancer; GWAS, genome-wide association study; HR, hazard ratio; ; MAF, minor allele frequency; SOCCS, Study of Colorectal Cancer in Scotland.

**Article Category:**

Short reports; Cancer Epidemiology; Cancer Genetics and Epigenetics

**Manuscript word count:** 1,518

Microsoft Word for Windows, version 2016

**Novelty and Impact**

There is a dearth of well-powered studies examining the effect of common germline variants associated with colorectal cancer risk on subsequent survival outcomes. Studies using additive genetic models failed to reveal robust associations between CRC risk variants and survival outcomes. We identified two genetic variants (rs16181980 and rs7495132) with significant recessive effects on survival outcomes of colorectal cancer, which provides novel insight into genetic architecture of post-diagnosis survival of CRC patients.

## Abstract

Previous studies using additive genetic models failed to identify robust evidence of associations between colorectal cancer (CRC) risk variants and survival outcomes. However, additive models can be prone to false negative detection if the underlying inheritance mode is recessive. Here, we tested all currently known CRC-risk variants (n=129) in a discovery analysis of 5,675 patients from a Scottish cohort. Significant associations were then validated in 2,474 CRC cases from UK Biobank. We found that the TT genotype of the intron variant rs7495132 in the *CRTC3* gene was associated with clinically relevant poorer CRC-specific survival in both the discovery (HR=1.97, 95% CI=1.41-2.74, p=6.1x10<sup>-5</sup>) and validation analysis (HR=1.69, 95% CI=1.03-2.79, p=0.038). In addition, the GG genotype of rs10161980 (intronic variant of *AL139383.1* lncRNA) was associated with worse overall survival in the discovery cohort (HR=1.24, 95% CI=1.10-1.39, p=3.4x10<sup>-4</sup>) and CRC-specific survival in the validation cohort (HR=1.26, 95% CI=1.01-1.56, p=0.040). Our findings show that common genetic risk factors can also influence CRC survival outcome.

## Introduction

Previous evidence have shown a familial contribution to survival outcomes of colorectal cancer (CRC), indicating that heritable germline genetic components can contribute to survival outcome<sup>1</sup>. Statistical association between germline variation and survival should be free from confounding and reverse causation, and so could have utility in *prediction* of cancer outcomes. Genetic variants involved in CRC pathogenesis may also influence subsequent survival outcomes in affected carriers. However, previous studies of low penetrance common genetic variance found inconsistent results using additive genetic models. Hence, no robust associations have been established between any common germline variation and CRC survival outcome<sup>2,3</sup>. Additive models assume a linear relationship between risk and each allele and is widely used in genome-wide association studies (GWAS)<sup>4</sup>. However, simulation studies show that the additive approach has low statistical power to identify effects of variants with true mode of inheritance being recessive<sup>4</sup>. We therefore investigated potential recessive genetic effects of variants known to be associated with CRC *risk*, in order to determine whether such variants also had recessive effects on survival outcomes.

## Materials and Methods

We tested associations between survival outcomes of CRC patients and a total of 129 known genetic variants (linkage disequilibrium  $r^2 < 0.2$ ) associated with CRC risk reported by two recent GWAS analyses<sup>4,5</sup> (details summarised in **supplementary Table 1**). Technical details of genotyping, imputation and quality control steps can be found in previous publications<sup>6</sup>.

We studied 5,675 CRC patients with GWAS data from the Study of Colorectal Cancer in Scotland (SOCCS) (discovery set). Age at diagnosis, sex and tumour stage were collected as covariates. Statistical power of discovery analysis was estimated for variants with varied

genotype frequencies and effect sizes using the method proposed by Freedman <sup>7</sup>(type I error rate=0.0005). The Cox hazards model was employed to estimate hazard ratios (HR) along with 95% confidence intervals (CIs) of homozygotes of minor alleles on overall and CRC-specific survival. We used the false positive rates (FDR) approach to correct for multiple testing ( $p < 0.05$  was the significance threshold after correction). We checked the proportional hazards assumption of the Cox model for each identified variant by testing the association between the Schoenfeld residuals and follow-up time using the SOCCS cohort.<sup>8</sup> If time-dependent trends of Schoenfeld residuals were detected ( $p < 0.05$ ), which indicated potential violation of the proportional hazards assumption, we fitted a parametric Weibull model to estimate the genetic effects on survival outcomes.<sup>9</sup> We also explored the associations between the 129 included variants and survival outcomes stratified by sex.

Significant associations identified from the SOCCS were then validated in an independent population-based prospective cohort composed of 2,474 incident CRC cases from the UK Biobank. Only age at diagnosis and sex were adjusted in the model due to CRC stage being unavailable in the UK Biobank. Given different covariates used, meta-analysis combining the two cohorts were not performed. Ethics approval of the study cohorts is presented elsewhere<sup>10</sup>. The diagram of patient selection for the two study cohorts is presented in **supplementary Figure S1**.

## **Results and discussion**

Patient characteristics for the study cohorts are presented in **Table 1**. Given 5,675 included cases, the discovery analysis yielded 88% (overall survival) and 66% (CRC-specific survival) power for 82 of the 129 (64%) included variants (homozygous genotype frequency  $> 0.05$ ) to

detect a HR of at least 1.50. Power estimates of variants with varied genotype frequencies and effect sizes are presented in **supplementary Table S2**.

In the discovery analysis, we found 12 variants associated with overall survival and nine variants associated with CRC-specific survival at nominal significance (uncorrected  $p < 0.05$ ) (**Table 2**). After FDR correction, we identified that the GG genotype of the variant rs10161980 was associated with clinically relevant inferior overall survival (HR=1.24, 95% CI=1.10-1.39, uncorrected  $p = 3.4 \times 10^{-4}$ , FDR corrected  $p = 0.022$ ). However, this effect was not significant on CRC-specific survival after FDR correction (HR=1.30, 95% CI=1.11-1.52, uncorrected  $p = 0.005$ , FDR corrected  $p = 0.196$ ). In addition to rs10161980, the TT genotype of the variant rs7495132 was significantly associated with worse CRC-specific survival (HR=1.97, 95% CI=1.41-2.74, uncorrected  $p = 6.1 \times 10^{-5}$ , FDR corrected  $p = 0.008$ ), whereas the association with overall survival did not survive the FDR correction (HR=1.40, 95% CI=1.01-1.93, uncorrected  $p = 0.045$ , FDR corrected  $p = 0.467$ ). No significant deviation from the proportional hazards assumption was observed for either variants ( $p > 0.05$ ). We then validated the two observed signals using the UK Biobank cohort (N=2,474). The GG genotype of rs10161980 was significantly associated with worse CRC-specific survival (HR=1.26, 95% CI=1.01-1.56,  $p = 0.040$ ), but not significantly associated with overall survival (HR=1.21, 95% CI=0.99-1.46,  $p = 0.057$ ). As for the other variant rs7495132, we replicated the observed association between the TT genotype and CRC-specific survival in the UK Biobank (HR=1.69, 95% CI=1.03-2.79,  $p = 0.038$ ), although no significant effect was detected of this variant on overall survival (HR=1.39, 95% CI=0.86-2.25,  $p = 0.179$ ). Kaplan-Meier survival estimates of these two variants in both discovery and validation analysis are plotted in **Figure 1** (CRC-specific survival) and **supplementary Figure S2** for overall survival.



With regard to stratified analysis by sex, the variant rs7495132 remained significantly associated with CRC-specific survival of male patients (TT genotype: HR=2.35, 95% CI=1.54-3.60, uncorrected  $p=8.0 \times 10^{-5}$ , FDR corrected  $p=0.010$ ) while rs10161980 was observed to be associated with overall survival of female patients from the SOCCS cohort (GG genotype: HR=1.38, 95% CI=1.15-1.65, uncorrected  $p=0.001$ , FDR corrected  $p=0.041$ ). In addition, we identified significant association between the TT genotype of the variant rs174537 and worse overall survival among female patients from the SOCCS. However, we failed to validate any of these associations by conducting sex-stratified analysis using the UK Biobank cohort (details in **supplementary Table S4**).

In this study, we identified two CRC-risk variants (rs7495132 and rs10161980) with significant effects on survival outcomes of CRC. Notably, both variants, although not reaching statistical significance, showed the same direction but smaller magnitude of effects in our previous analysis using additive model<sup>2</sup>, indicating the underlying recessive model of inheritance for these two variants.

The T allele of rs7495132 was identified as a CRC-risk increasing allele in the original GWAS meta-analysis<sup>6</sup>. Located in chromosome 15, rs7495132 is an intron variant of the *CRTC3* gene which encodes the protein CREB regulated transcription co-activator 3 (CRTC3) and is associated with Human T cell leukaemia virus infection. According to the results of Genotype-Tissue Expression (GTEx) project (<https://gtexportal.org/home/>), the T allele of rs7495132 is associated with lower expression of the *CRTC3* gene in multiple human tissues such as musculoskeletal tissue and subcutaneous adipose tissue, although no significant association is present in colonic tissue. Previous evidence showed that the CRTC3 protein can regulate energy balance and is associated with weight gain in a mouse model<sup>11</sup>.

To explore the potential role of adiposity in the observed genetic association, we compared the BMI level between CRC patients of different rs7495132 genotypes using the UK Biobank cohort. However, no significant difference was detected (mean BMI level of TT vs. TC+CC genotypes: 27.64 vs. 27.97,  $p=0.595$ ). We also added BMI as a covariate in the survival model, and observed that the BMI level was not associated with survival outcomes (CRC-specific survival: HR=1.00, 95%CI=0.98-1.02,  $p=0.938$ ) whilst the genetic association between rs7495132 and survival outcomes remained significant in the multivariable model. These results could not rule out adiposity as a potential mediator in the observed genetic association. Future research is needed to further reveal the possible biological function of rs7495132 in the adiposity-mediated pathway involved in CRC progression.

With respect to the other variant rs10161980, it is an intron variant of the *AL139383.1* LncRNA located in chromosome 13. Currently, little is known for potential biological implications of this gene, and there has been a paucity of evidence that rs10161980 can affect expression of any genes. Notably, the G allele of rs10161980 was identified as the CRC-risk decreasing allele in the original GWAS<sup>6</sup>. This indicates that the variant may be involved in different biological pathways of CRC carcinogenesis and tumour progression. It should be noted that a chance finding could not be confidently excluded despite the validation in our study. Therefore, future large population-based studies are also needed to confirm this association.

Given the limited sample size of the study cohorts and the number of variants tested, this study was underpowered to detect small recessive genetic effects (HR<1.20), especially for variants with low homozygous genotype frequencies (<0.30). Another limitation is that AJCC stage was not adjusted in replication analysis due to unavailable data. We re-estimate

the effects of these two variants in SOCCS adjusted for only age and sex, and the results remained statistically significant (**supplementary Table S3**).

This study identifies two CRC-risk variants (rs10161980 and rs7495132) that have recessive genetic effects on survival outcomes of CRC. Currently AJCC stage is used as a blunt indicator of the need for adjuvant chemotherapy. The genetic associations discovered in this study provide prospects for future investigations to incorporate rs10161980 and rs7495132 genotypes into individual prediction models for survival outcomes. In addition, our findings improve understanding of the genetic architecture for prognosis of CRC patients. Future research is warranted to illuminate biological mechanisms through which these two variants influence CRC progression. Our findings also suggest that recessive genetic effects should be considered in future genetic association studies on survival of CRC.

### **Authors' contributions**

MGD and ET conceived of the study and approved publication of this study. YH and MT performed all analyses. YH drafted the manuscript. MT supervised all statistical analysis. MT, VS, SMF collected and prepared the genotype data. FVD, SMF and MGD collected clinical data for the study. All authors reviewed and edited the manuscript.

### **Acknowledgements**

We acknowledge the excellent technical support from Stuart Reid and Marion Walker. We are grateful to Donna Markie and all those who continue to contribute to recruitment, data collection, and data curation for the Study of Colorectal Cancer in Scotland studies. We acknowledge that these studies would not be possible without the patients and surgeons who take part. We acknowledge the expert support on sample preparation from the Genetics Core of the Edinburgh Wellcome Trust Clinical Research Facility. This research has been conducted using the UK Biobank Resource under Application Number 7441.

### **Conflict of interest:**

All authors declare no conflicts of interest.

### **Data availability statement**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The data are not publicly available due to privacy or ethical restrictions.

### **Grant support**

This work was supported by CRUK programme grant C348/A18927 (MGD). It was also supported by funding for the infrastructure and staffing of the Edinburgh CRUK Cancer Research Centre. ET is supported by a CRUK Career Development Fellowship (C31250/A22804). The work received support from COST Action BM1206. This work was also funded by a grant to MGD as Project Leader with the MRC Human Genetics Unit Centre Grant (U127527202 and U127527198 from 1/4/18). YH was supported by the Henry Lester Trust and the Great Britain China Educational Trust. YH and XL was supported by China Scholarship Council. XZ was supported by the Darwin Trust.

## References

1. Lindstrom LS, Hall P, Hartman M, Wiklund F, Gronberg H, Czene K. Familial concordance in cancer survival: a Swedish population-based study. *Lancet Oncol* 2007;**8**: 1001-6.
2. He Y, Theodoratou E, Li X, Din FVN, Vaughan-Shaw P, Svinti V, Farrington SM, Campbell H, Dunlop MG, Timofeeva M. 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**: 2427-32.
3. Phipps AI, Newcomb PA, Garcia-Albeniz X, Hutter CM, White E, Fuchs CS, Hazra A, Ogino S, Nan H, Ma J, Campbell PT, Figueiredo JC, et al. Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. *Gastroenterology* 2012;**143**: 51-4 e4.
4. Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet Epidemiol* 2007;**31**: 358-62.
5. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, Conti DV, Qu C, Jeon J, Edlund CK, Greenside P, Wainberg M, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;**51**: 76-87.
6. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, Farrington S, Svinti V, Palles C, Orlando G, Sud A, Holroyd A, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* 2019;**10**: 2154.
7. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in medicine* 1982;**1**: 121-9.
8. Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* 1980;**67**: 145-53.
9. Carroll KJ. On the use and utility of the Weibull model in the analysis of survival data. *Control Clin Trials* 2003;**24**: 682-701.
10. He Y, Timofeeva M, Farrington SM, Vaughan-Shaw P, Svinti V, Walker M, Zgaga L, Meng X, Li X, Spiliopoulou A, Jiang X, Hypponen E, et al. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. *BMC medicine* 2018;**16**: 142.
11. Song Y, Altarejos J, Goodarzi MO, Inoue H, Guo X, Berdeaux R, Kim JH, Goode J, Igata M, Paz JC, Hogan MF, Singh PK, et al. CRT3 links catecholamine signalling to energy balance. *Nature* 2010;**468**: 933-9.

## Figure legend

Figure 1 Kaplan-Meier survival estimates of CRC-specific survival stratified by rs10161980 and rs7495132 (A: rs10161980 in SOCCS, B: rs10161980 in UK Biobank; C: rs7495132 in SOCCS; D: rs7495132 in UK Biobank)

Table 1 Summarised characteristics of the discovery and validation cohorts

Variables*	SOCCS (n=5,675)	UK Biobank (n=2,474)
Age at diagnosis (years)	64.5(54.6-71.6)	64.9(61.2-69.6)
Sex		
Male	3,235(57%)	1,035(42%)
Female	2,440(43%)	1,439(58%)
AJCC stage		
I	1,005(17.7%)	NA
II	1,891(33.3%)	
III	1,995(35.2%)	
IV	784(13.8%)	

\*Continuous variables are presented with median and interquartile range. AJCC, the American Joint Committee on Cancer; NA, not available.



Table 2 Summary of associations ( $p < 0.05$ ) between CRC-risk variants and survival outcomes of CRC patients in the SOCCS study under a recessive model

	Variant	MA	RGF	HR(95%CI)	P(uncorrected)	Pfdr
OS						
	<b>rs10161980</b>	<b>G</b>	<b>0.12</b>	<b>1.24(1.10-1.39)</b>	<b>3.40E-04</b>	<b>0.022</b>
	rs174537	T	0.09	1.23(1.07-1.41)	0.003	0.091
	rs6066825	G	0.24	1.22(1.07-1.40)	0.003	0.091
	rs73975588	C	0.01	0.55(0.35-0.86)	0.009	0.235
	rs3087967	C	0.07	0.89(0.82-0.98)	0.014	0.311
	rs3217810	T	0.003	1.42(1.06-1.92)	0.021	0.380
	rs78341008	C	0.001	1.79(1.08-2.98)	0.024	0.395
	rs35509282	A	0.05	0.61(0.40-0.96)	0.030	0.437
	rs10951878	T	0.20	0.89(0.80-0.99)	0.035	0.454
	rs35360328	A	0.01	1.35(1.01-1.80)	0.041	0.467
	rs7495132	T	0.02	1.40(1.01-1.93)	0.045	0.467
	rs16892766	C	0.01	0.55(0.30-1.00)	0.049	0.467
CSS						
	<b>rs7495132</b>	<b>T</b>	<b>0.02</b>	<b>1.97(1.41-2.74)</b>	<b>6.10E-05</b>	<b>0.008</b>
	rs6066825	G	0.24	1.30(1.11-1.52)	0.001	0.063
	rs10161980	G	0.12	1.22(1.06-1.40)	0.005	0.196
	rs9537521	A	0.03	1.22(1.06-1.41)	0.006	0.201
	rs4811050	A	0.04	1.38(1.08-1.78)	0.012	0.301
	rs35509282	A	0.05	0.52(0.30-0.90)	0.019	0.414
	rs10951878	T	0.20	0.87(0.76-0.99)	0.030	0.492
	rs3217810	T	0.003	1.47(1.04-2.09)	0.030	0.492
	rs13020391	T	0.09	1.17(1.01-1.35)	0.037	0.536

MA, minor allele; RGF, rare genotype frequency; HR, hazard ratio; CI, confidence interval; Pfdr, p-values adjusted using the false positive rate approach. OS, overall survival, CSS, CRC-specific survival.

