Published in final edited form as:

Clin Gastroenterol Hepatol. 2020 November ; 18(12): 2717–2723.e3. doi:10.1016/j.cgh.2019.11.046.

Genetic Variant Associated with Survival of Patients with Stage II-III Colon Cancer

Kathryn L. Penney^{1,2}, Barbara L. Banbury³, Stephanie Bien³, Tabitha A. Harrison³, Xinwei Hua^{3,4}, Amanda I. Phipps^{3,4}, Wei Sun³, Mingyang Song^{2,5,6,7}, Amit D. Joshi⁵, Steven R. Alberts⁸, Carmen J. Allegra^{9,10}, James Atkins^{10,11}, Linda H. Colangelo^{10,12}, Thomas J. George^{9,10}, Richard M. Goldberg^{10,13}, Peter C. Lucas^{10,12}, Suresh G. Nair¹⁴, Qian Shi^{8,15}, Frank A. Sinicropo⁸, Norman Wolmark^{10,16}, Greg Yothers^{10,12}, Ulrike Peters^{3,4}, Polly A. Newcomb^{3,4}, Andrew T. Chan^{1,2,5,7}

¹Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA;

²Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA;

³Fred Hutchinson Cancer Research Center, Seattle, WA;

⁴University of Washington, Seattle, WA;

⁵Clinical and Translational Epidemiology Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA;

⁶Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA;

⁷Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, MA;

⁸Department of Health Science Research, Mayo Clinic, Rochester, MN;

⁹University of Florida, Gainesville, FL;

¹⁰NRG Oncology, Pittsburgh, PA;

Corresponding author: Andrew T. Chan, achan@mgh.harvard.edu, Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Bartlett Ext., 9th Floor, 55 Fruit St., Boston, MA 02114.

Authors' roles:

Study concept and design – KLP, BLB, SB, TAH, AIP, UP, PAN, ATC

Acquisition of data – SRA, CJA, JA, LHC, TJG, TMG, PCL, SGN, QS, FAS, NW, GY

Analysis and interpretation of data – KLP, BLB

Drafting of the manuscript – KLP, BLB, PAN, ATC

Critical review and revision of the manuscript – all authors

Statistical analysis – BLB

Study supervision – ATC

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ClinicalTrials.gov Identifiers: [NCT00096278](https://clinicaltrials.gov/ct2/show/study/NCT00096278) (NSABP C-08) and [NCT00079274](https://clinicaltrials.gov/ct2/show/study/NCT00079274) (NCCTG N0147)

Conflicts of Interest:

The following authors wish to disclose: UP received support for a research project from JUNO Therapeutics, outside of the scope of this work. PCL reports stock ownership outside of the scope of this work with Amgen and Bayer/Loxo. RMG received personal fees from Taiho, Merck KGA, Merck, and Novartis, all outside of the scope of this work.

¹¹Southeast Clinical Oncology Research (SCOR) Consortium NCORP, Winston-Salem, NC;

¹²University of Pittsburgh, Pittsburgh, PA;

¹³West Virginia University Cancer Institute, Morgantown, WV;

¹⁴Lehigh Valley Hospital-Cedar Crest, Allentown, PA;

¹⁵Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN;

¹⁶Allegheny Health Network Cancer Institute, Pittsburgh, PA

Abstract

Background & Aims: Many genetic variants have been associated with colorectal cancer risk, although few have been associated with survival times of patients. Identification of genetic variants associated with survival times might improve our understanding of disease progression and aid in outcome prediction. We performed a genome-wide association study to identify variants associated with colon cancer survival time.

Methods: We performed a post-hoc analysis of data from NCCTG N0147 (Alliance), a randomized phase 3 trial of patients with resected stage III colon cancer, and from NSABP C-08 (NRG), a phase 3 trial that compared therapy regimens for patients with resected stage II or III colon cancer. Genotype analyses were performed on DNA from blood samples from 4974 patients. We used Cox proportional hazards regression to evaluate the association of each single nucleotide polymorphism with times of overall survival and disease-free survival, adjusting for age at diagnosis, sex, treatment group, and principal components of genetic ancestry. We performed the analysis for studies N0147 and C-08 separately, and results were combined in a fixed-effects meta-analysis.

Results: A locus on chromosome 7p15.2 was significantly associated with overall survival time ($P = 5 \times 10^{-08}$). The most significant variant at this locus, rs76766811 ($P = 1.6 \times 10^{-08}$), is common among African Americans (minor allele frequency, approximately 18%) but rare in European Americans (minor allele frequency $< 0.1\%$). Within strata of self-reported ancestry, this variant was associated with times of overall survival and disease-free survival in only African Americans (hazard ratio for overall survival, 2.82; 95% CI, 1.88–4.23; $P = 5.0 \times 10^{-07}$ and hazard ratio for disease-free survival, 2.27; 95% CI, 1.62–3.18; $P = 1.8 \times 10^{-06}$).

Conclusions: In an analysis of data from 2 trials of patients with stage II or III colon cancer, we identified rs76766811 as a potential prognostic variant in African American patients. This finding should be confirmed in additional study populations.

Keywords

GWAS; DFS; NCCTG N0147; NSABP C-08

Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer, and the second leading cause of cancer death in the United States for men and women combined[1–3]. Five-year survival among colorectal cancer patients is 64%[2]. While pathological stage is the

most prognostic factor, patients diagnosed at the same stage show heterogeneous outcomes (National Cancer Institute SEER database. <http://seer.cancer.gov/>). To date, the etiology underlying CRC progression is poorly understood, though evidence suggests survival may have a genetic component. A study from Sweden of children and parents with the same cancer showed a significant increased risk of CRC death in the children whose parents had poor survival compared to those whose parents had better survival[4].

The importance of genetic susceptibility to CRC risk has been established by the identification of over 100 genetic variants associated with CRC risk in genome-wide association studies (GWAS)[5,6]; however, few genetic variants have been specifically associated with CRC survival outcomes[7,8]. The identification of genetic variants associated with survival may improve our understanding of the biology of disease progression, and potentially aid in the development of additional prognostic markers and facilitate tailored treatment. We therefore performed a GWAS in two large colon cancer clinical trials to identify genetic variants associated with time to disease recurrence and all-cause mortality.

Materials and Methods

Clinical Trial Description

North Central Cancer Treatment Group (NCCTG) N0147 was a multicenter randomized phase III clinical trial for adjuvant therapy in patients with resected stage III colon cancer.[9] NCCTG is now part of the Alliance for Clinical Trials in Oncology. It was originally designed to compare three different chemotherapy regimens, and later expanded to evaluate cetuximab in each original arm (leading to six total arms). Further modification added pre-screening for *KRAS* mutation status, with the resulting goal to assess the potential benefit of cetuximab added to the modified sixth version of the infusional fluorouracil, leucovorin, and oxaliplatin (mFOLFOX6) in patients with resected stage III wild-type *KRAS* colon cancer. The primary outcome of this trial was disease-free survival (DFS) in patients with wild-type *KRAS*, while secondary end points included overall survival (OS) and toxicity. For the purposes of the current GWAS, treatment groups were collapsed into FOLFOX or FOLFIRI, +/- cetuximab.

National Surgical Adjuvant Breast and Bowel Project (NSABP) C-08 was a phase III randomized clinical trial in patients with resected stages II and III colon cancer [10]. NSABP is now part of NRG Oncology. The primary aim of the trial was to compare the relative efficacy of mFOLFOX6 + bevacizumab with that of mFOLFOX6 alone in prolonging DFS; the secondary aim was to compare the relative efficacy of mFOLFOX6 + bevacizumab with that of mFOLFOX6 alone in prolonging OS.

For both trials, participants signed an IRB-approved informed consent document for use of samples, in accordance with federal and institutional guidelines.

Genotyping and Sample Quality Control

Study samples were genotyped at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. DNA sources were buffy coat (33.7% of samples) and blood (66.3%);

DNA extraction involved the Omega E.Z.N.A. Blood DNA kit, the Gentra Autopure LS, the Qiagen Midi, or the Autogen Flex Star. Genotyping controls were 133 HapMap samples.

Genotyping was performed using the Illumina HumanOmniExpress + Exome array (HumanOmniExpressExome-8v1-2, BPM annotation version A, genome build GRCh37/hg19) and GenomeStudio v2011.1 was used to call variants. The array consisted of 964,193 single nucleotide polymorphisms (SNPs). While the array contains both SNPs and non-SNP variants (i.e., indels), in this report, we use the term “SNP” more generically to refer to all genotyped variants.

Sample-level QC involved checking annotated vs genetic sex, chromosomal anomalies, >2% missing call rate (MCR), and relatedness among patients. Two samples were removed based on a discrepancy between annotated and genetic sex. After QC, the median call rate across samples was 99.96% and the error rate estimated from pairs of technical replicates was 3.15×10^{-6} . We included 4,974 study patients with complete genotype and clinical data in the final analysis. For genotype calls, each SNP was checked for technical failures, >2% MCR, discordance in technical duplicates, Mendelian errors, and sex differences in allele frequencies and heterozygosity. A total of 1.94% SNPs were excluded based on these quality control analyses; another 7.43% SNPs were removed from subsequent analysis because they were positional duplicates or monomorphic in our study population. After SNP QC, 873,829 SNPs passed this composite filter.

Principal Components Analysis and Imputation

To investigate population structure, we used principal components analysis (PCA) implemented in the R package SNPRelate[11], as described by Patterson et al.[11,12]. The first seven eigenvectors from the PCA described 95% of the genetic variation, and we used these as covariates in the analysis. All samples were phased prior to imputation using SHAPEIT2[13], and then missing genotypes were imputed to the 1000 Genomes Project phase 3 reference panel[14] using IMPUTE2 (version 2.3.2)[15]. After imputation, we converted genotype probabilities to allelic dosages using the formula $D_A = 2 \times P(AA) + 1 \times P(AB)$. We excluded all variants with less than 1% minor allele frequency and filtered out poorly imputed SNPs (info metric < 0.3) prior to analyses. A total of 10,564,363 SNPs were included in the final analysis.

Study Outcomes

We calculated overall survival (OS) as the time from randomization to death from any cause and disease-free survival (DFS) as the time from randomization to the first colon cancer recurrence or death due to any cause (whichever came first). Demographic characteristics and follow-up time after diagnosis were similar across the two trials (see Table 1).

Statistical Analysis

The data lock dates for clinical and outcome data were 8/7/2015 for N0147 and 9/30/2013 for C-08. We used Cox proportional hazards regression to evaluate hazard ratios (HRs) and 95% confidence intervals (CIs) for associations of each SNP with OS and DFS, adjusting for relevant baseline factors (age at randomization, sex, treatment arm, and principal

components). SNPs were modeled using the log-additive approach, relating genotype dosage (measure of genotype probability as described above) to each survival outcome. We tested for a non-zero slope of the scaled Schoenfeld residuals on ranked failure time to assess violations of proportional hazards assumptions prior to association analyses [16]. Two covariates significantly violated proportional hazards assumptions. The first was age at randomization ($r^2=5.34$; $p=0.02$); this violation was resolved after dichotomizing age (<57 , >57). The other was treatment arm, which we decided to stratify in the model. To account for population substructure, we adjusted for the first seven eigenvectors from the principal components analysis that described 95% of the genetic variation.

Because of the different treatments across studies, we analyzed N0147 and C-08 separately, and meta-analyzed the results using a fixed-effects approach. We inspected quantile-quantile (QQ) plots of log-transformed p-values and assessed the calculated genomic control coefficients to assess possible inflation and bias [17]. We found no evidence of systemic inflation, as the genomic inflation factor estimates show little deviation from expectation ($\lambda = 0.989$; $OS_\lambda = 0.998$). We produced Manhattan plots and specified genome-wide statistical significance at $p < 5 \times 10^{-8}$; all p-values are 2-sided. We examined the association of significant SNPs with OS and DFS stratified by self-reported race (“White” and “Black or African American”). We examined the association of self-reported race with overall survival using a Cox proportional hazards model, with “White” as the referent group, adjusting for sex, study, and treatment group. Analyses were performed using R 3.2.2.

Bioinformatic Functional Follow-up

We performed *in silico* follow-up of the significant locus using the UCSC Genome Browser [18] and Haploreg Version 4.0 [19]. We first used rAggr (<http://raggr.usc.edu>) software to define the locus as the most significant variant, rs76766811, and all variants within 500 Kilobases having an LD $r^2 \geq 0.4$ in 1000 Genomes Phase 3 AFR super population. This list of variants was then aligned to a UCSC custom track hub [20] (<https://genome.ucsc.edu/>) populated with annotations related to chromatin states (enhancers/promoters and repressors) in normal colorectal crypts and primary CRC cell lines. Regulatory regions were previously defined using ChIP-seq H3K27ac and H3K4me1 signals for enhancers/promoters, and H3K27me3 for repressors [21,22]. Broad ChIP-seq signals were further refined to regions of open chromatin using annotations for DNaseI hypersensitivity sequencing (DHS) [21,22]. Transparent overlay tracks were designed to compare regulatory annotations in normal crypts (ChIP-seq $n=4$; DHS $n=3$) versus CRC cell lines (ChIP-seq $n=31$ H3K27ac, $n=28$ H3K4me1; DHS $n=3$). Additionally, we used the UCSC genome browser track “transcription factor (TF) ChIP-seq (161 factors)” to identify putative regulatory variants overlapping regions occupied by TF.

Results

Demographic and clinical characteristics of the patients in the two trials were very similar (Table 1; Supplementary Table 1). The fixed-effects meta-analysis results identified a locus at chromosome 7p15.2 as statistically significantly associated with OS. Two SNPs in complete linkage disequilibrium (LD) ($r^2=1.0$) were significant, with rs76766811 most

strongly associated with OS (HR=2.63, 95% CI: 1.88–3.67; $p=1.6\text{e-}08$; Table 2). The results for this SNP were not materially changed when the analysis was restricted to stage III patients or to specific treatment groups (Supplementary Table 2). The second SNP, rs75533594, was also strongly associated with OS (HR=2.63, 95% CI: 1.88–3.69; $p=1.9\text{e-}08$). Although no variants reached genome-wide significance for DFS, the association was similar for rs76766811 (HR=2.09, 95% CI: 1.57–2.78; $p=3.7\text{e-}07$). All OS and DFS results are presented in a Manhattan plot (Figure 1).

rs76766811, located on chromosome 7 (position 26660554 in NCBI dbGaNP Build 37), is rare in this primarily European American study population (<0.1% minor allele frequency [MAF]) but much more common in African ancestry populations (18% MAF). When we stratified the analysis by self-reported ancestry, this variant was not associated with OS or DFS in white participants, likely a reflection of low power to detect a signal with this very rare allele. The association with OS was evident in African Americans (HR=2.82, 95% CI: 1.88–4.23, $p=5.0\text{e-}07$, $\text{phet}=0.03$) (Table 2), and the association with DFS was similar (DFS HR=2.27, 95% CI: 1.62–3.18 $p=1.8\text{e-}06$), although both not statistically significant at a genome-wide level. Mean and median follow-up times for self-reported whites and African Americans were similar (Supplementary Table 1). There was no association between race and OS comparing African American to white patients ($p=0.86$). A forest plot of the OS results for self-reported whites and African American participants is presented in Supplementary Figure 1. We estimate rs76766811 genotypes based on the imputed dosage (info metric=0.86) and present a Kaplan-Meier curve of the association with OS among African American patients in Figure 2.

Bioinformatics follow-up of this locus showed that rs76766811 is in LD with SNPs rs75533594 ($r^2 = 1.0$), rs75256807 ($r^2 = 0.52$), and rs76007482 ($r^2 = 0.59$) that overlap with an enhancer that is active in CRC cell lines, but absent in normal colon crypts. This locus lies upstream of *Src Kinase Associated Phosphoprotein 2 (SKAP2)* (Supplementary Figure 2). rs75256807, which was filtered out of our main analysis due to its low MAF (0.008) but has a p-value of $5.9\text{e-}04$, is located in a transcription factor occupancy site for a number of genes including the transcriptional repressor early growth response protein 1 (*EGR1*), transcriptional repressor protein YY1 (*YY1*), and CCCTC-Binding Factor (*CTCF*). The variant is also predicted to decrease the binding efficiency for *CTCF*.

Discussion

This GWAS nested within two colon cancer clinical trials identified rs76766811 on chromosome 7p15 as statistically significantly associated with OS. As this SNP is more common among African ancestry populations, this association was driven by and only observed among the self-reported African American patients. While we found a significant finding for OS, we observed that this variant was strongly associated with DFS as well.

This SNP was not reported in the largest GWAS of CRC survival to date[8]; however, that study was restricted to participants of European ancestry. Other studies have identified variants associated with CRC survival. The GWAS from Phipps et al.[8] identified several SNPs at 6p12.1, including rs209489, as significantly associated with OS among individuals

with metastatic CRC. Smith et al.[7] tested the association of 20 known CRC risk SNPs and found that rs9929218 at 16q22 was associated with OS. Pander et al.[23] found rs885036 to be significantly associated with progression-free survival (PFS) in a GWAS of metastatic CRC patients. However, none of these variants were significant in our study. This is possibly due to the focus in prior studies on stage IV/metastatic CRC patients who were predominantly European American, while our study focused on earlier stage II-III colon cancer patients and included participants of multiple ancestries.

There is evidence that the SNP association we observed in this study has functional relevance. rs76766811 is located on chromosome 7, approximately 17kb upstream of *C7orf71* and approximately 46kb downstream from *SKAP2*. *SKAP2* is an adaptor protein that plays a role in activating the immune system. Based on the bioinformatic annotation, we hypothesize that variants in this locus disrupt the binding of transcriptional repressors, resulting in increased enhancer activity and expression of *SKAP2*. It has been shown that macrophages with high *SKAP2* expression promoted tumor invasion and metastasis[24]. This gene was also frequently amplified in pancreatic ductal adenocarcinoma and high tumor expression of *SKAP2* was associated with poor prognosis among non-small cell lung cancer patients[25,26].

This study has several strengths. By leveraging data within two well-conducted, Phase III randomized controlled trials, we were able to examine genetic associations within a population of patients in which clinical correlates were well-characterized, treatment was standardized, and follow-up for disease outcomes and serious adverse events was uniformly and meticulously conducted.

Our study also has limitations. Although we had a large enough sample size to identify the significant SNP in African American patients, we were constrained in our ability to detect other potentially significant SNPs in any ancestry group. Our sample size also did not allow for the examination of differential effects according to other characteristics, such as disease site or age. Additionally, although we did observe similar results in both trials, additional trial populations with similar data were not available for further validation. In the absence of a similar study population for formal replication, we cannot rule out that this is a chance finding.

African Americans have a higher risk of CRC recurrence and disease-specific mortality[27]. As GWAS have predominantly focused on Europeans and East Asians, it is critical to expand GWAS efforts to include multiple ancestry populations to better understand the biology of disease and address disease prediction and prognosis in diverse groups to ensure that all populations benefit from the evolving field of genomic medicine. Future colon cancer GWAS should focus on identifying variants associated with disease-specific survival in more diverse populations. Such efforts may lead to the identification of germline genetic variants that can improve prognostication and treatment plans for the larger population. The association of rs76766811 with OS should be confirmed in additional African American study populations. If validated, a precision medicine strategy using this variant could risk-stratify colon cancer survivors to identify those who should receive more intense surveillance colonoscopy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers R01CA176272, U10CA180821, U10CA180882, and U24CA196171 (to the Alliance for Clinical Trials in Oncology), U10CA180833, U10CA180867, UG1CA189858. NRG Oncology/NSABP C-08 acknowledges support from NIH U10CA180868, U10CA180822, UG1CA189867, and U24CA196067. (NRG Oncology Lab disclaimer: Under a grant from the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analysis, interpretations, or conclusions). Also supported in part by funds from Genentech and Sanofi. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. American Cancer Society. Cancer Facts & Figures 2018. 2018<https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html> (accessed 13 Sep 2018).
2. American Cancer Society. Colorectal Cancer Facts & Figures | Facts About Colon Cancer. 2017<https://www.cancer.org/research/cancer-facts-statistics/colorectal-cancer-facts-figures.html> (accessed 14 Sep 2018).
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;68:7–30. doi:10.3322/caac.21442 [PubMed: 29313949]
4. Lindström LS, Hall P, Hartman M, et al. Familial concordance in cancer survival: a Swedish population-based study. *Lancet Oncol* 2007;8:1001–6. doi:10.1016/S1470-2045(07)70282-6 [PubMed: 17921068]
5. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. *J Natl Cancer Inst* 2019;111:146–57. doi:10.1093/jnci/djy099 [PubMed: 29917119]
6. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;51:76–87. doi:10.1038/s41588-018-0286-6 [PubMed: 30510241]
7. Smith CG, Fisher D, Harris R, 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;21:3453–61. doi:10.1158/1078-0432.CCR-14-3136 [PubMed: 25873087]
8. Phipps AI, Passarelli MN, Chan AT, et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis* 2016;37:87–95. doi:10.1093/carcin/bgv161 [PubMed: 26586795]
9. Alberts SR, Sinicrope FA, Grothey A. N0147: a randomized phase III trial of oxaliplatin plus 5-fluorouracil/leucovorin with or without cetuximab after curative resection of stage III colon cancer. *Clin Colorectal Cancer* 2005;5:211–3. [PubMed: 16197625]
10. Allegra CJ, Yothers G, O'Connell MJ, et al. Phase III trial assessing bevacizumab in stages II and III carcinoma of the colon: results of NSABP protocol C-08. *J Clin Oncol* 2011;29:11–6. doi:10.1200/JCO.2010.30.0855 [PubMed: 20940184]
11. Zheng X, Levine D, Shen J, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012;28:3326–8. doi:10.1093/bioinformatics/bts606 [PubMed: 23060615]
12. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:e190. doi:10.1371/journal.pgen.0020190 [PubMed: 17194218]
13. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;10:5–6. doi:10.1038/nmeth.2307 [PubMed: 23269371]

14. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015;526:68–74. doi:10.1038/nature15393 [PubMed: 26432245]
15. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* 2011;1:457–70. doi:10.1534/g3.111.001198 [PubMed: 22384356]
16. Schoenfeld D Partial residuals for the proportional hazards regression model. *Biometrika* 1982;69:239–41. doi:10.1093/biomet/69.1.239
17. Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* 2001;60:155–66. doi:10.1006/tpbi.2001.1542 [PubMed: 11855950]
18. Casper J, Zweig AS, Villarreal C, et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res* 2018;46:D762–9. doi:10.1093/nar/gkx1020 [PubMed: 29106570]
19. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–4. doi:10.1093/nar/gkr917 [PubMed: 22064851]
20. Raney BJ, Dreszer TR, Barber GP, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 2014;30:1003–5. doi:10.1093/bioinformatics/btt637 [PubMed: 24227676]
21. Cohen AJ, Saiakhova A, Corradin O, et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat Commun* 2017;8:14400. doi:10.1038/ncomms14400 [PubMed: 28169291]
22. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. *Genome Med* 2014;6:85. doi:10.1186/s13073-014-0085-3 [PubMed: 25473424]
23. Pander J, van Huis-Tanja L, Böhringer S, et al. Genome Wide Association Study for Predictors of Progression Free Survival in Patients on Capecitabine, Oxaliplatin, Bevacizumab and Cetuximab in First-Line Therapy of Metastatic Colorectal Cancer. *PLoS ONE* 2015;10:e0131091. doi:10.1371/journal.pone.0131091 [PubMed: 26222057]
24. Tanaka M, Shimamura S, Kuriyama S, et al. SKAP2 Promotes Podosome Formation to Facilitate Tumor-Associated Macrophage Infiltration and Metastatic Progression. *Cancer Res* 2016;76:358–69. doi:10.1158/0008-5472.CAN-15-1879 [PubMed: 26577701]
25. Harada T, Chelala C, Bhakta V, et al. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. *Oncogene* 2008;27:1951–60. doi:10.1038/sj.onc.1210832 [PubMed: 17952125]
26. Kuranami S, Yokobori T, Mogi A, et al. Src kinase-associated phosphoprotein2 expression is associated with poor prognosis in non-small cell lung cancer. *Anticancer Res* 2015;35:2411–5. [PubMed: 25862907]
27. Simon MS, Thomson CA, Pettijohn E, et al. Racial differences in colorectal cancer incidence and mortality in the Women's Health Initiative. *Cancer Epidemiol Biomarkers Prev* 2011;20:1368–78. doi:10.1158/1055-9965.EPI-11-0027 [PubMed: 21602308]

Figure 1. Manhattan plot of SNP associations with OS (top) and DFS (bottom) among colon cancer patients.

Figure 2.
Kaplan-Meier survival curve by rs76766811 genotype in self-reported African American patients

Table 1.

Characteristics of colon cancer patients in study populations included in association analyses.

	NSABP C-08 (N=1,982)	NCCTG N0147 (N=2,992)
Deaths, N (%)	412 (21)	646 (22)
Median time to death, months (SD)	74 (23)	56 (23)
DFS Events, N (%)	530 (27)	950 (32)
Median time to event, months (SD)	73 (28)	50 (26)
Female, N (%)	992 (50)	1424 (48)
Age at diagnosis, N (%)		
<50y	574 (29)	767 (26)
50–60y	644 (32)	964 (32)
60–70y	510 (26)	916 (31)
70–80y	238 (12)	328 (11)
>80y	16 (1)	17 (1)
Self-reported race, N (%)		
White	1749 (88)	2570 (86)
African American	145 (7)	213 (7)
Asian	53 (3)	137 (5)
Other or not reported	35 (2)	72 (2)
Stage at diagnosis, N (%)		
Stage II	484 (24)	0 (0)
Stage III	1498 (76)	2992 (100)
Primary tumor location, N (%)		
Proximal	999 (50)	1510 (50)
Distal	927 (47)	1435 (48)
Other colon	56 (3)	47 (2)

Table 2.

Association of rs76766811 (ref=C; risk=T) with overall survival (OS) and disease-free survival (DFS) among all patients and among African American patients.

Study population	risk allele freq.	Outcome	Events / Total n	HR (CI)	p-value
All patients	1.6%	OS	1,058 / 4,974	2.63 (1.88–3.67)	1.60E-08
All patients	1.6%	DFS	1,480 / 4,974	2.09 (1.57–2.78)	3.66E-07
Self-reported African American patients	18%	OS	75 / 358	2.82 (1.88–4.23)	5.00E-07
Self-reported African American patients	18%	DFS	110 / 358	2.27 (1.62–3.18)	1.83E-06

rsID based on NCBI dbGaNP Build 37. Risk allele frequency is from this study population. HR: hazard ratio, CI: confidence interval. P-values are based on fixed-effects, inverse variance-weighted meta-analysis adjusted for sex, dichotomized age, treatment, and the first seven principal components for ancestry.