

Published in final edited form as:

Gastroenterology. 2013 April ; 144(4): 799–807.e24. doi:10.1053/j.gastro.2012.12.020.

Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-wide Meta-analysis

Ulrike Peters^{1,2,*}, Shuo Jiao^{1,*}, Fredrick R. Schumacher^{3,*}, Carolyn M. Hutter^{1,2,*}, Aaron K. Aragaki¹, John A. Baron⁴, Sonja I. Berndt⁵, Stéphane Bézieau⁶, Hermann Brenner⁷, Katja Butterbach⁷, Bette J. Caan⁸, Peter T. Campbell⁹, Christopher S. Carlson^{1,2}, Graham Casey³, Andrew T. Chan^{10,11}, Jenny Chang-Claude¹², Stephen J. Chanock⁵, Lin S. Chen¹³, Gerhard A. Coetzee³, Simon G. Coetzee³, David V. Conti³, Keith R. Curtis¹, David Duggan¹⁴, Todd Edwards¹⁵, Charles S. Fuchs^{11,16}, Steven Gallinger¹⁷, Edward L. Giovannucci^{11,18}, Stephanie M. Gogarten¹⁹, Stephen B. Gruber³, Robert W. Haile³, Tabitha A. Harrison¹, Richard B. Hayes²⁰, Brian E. Henderson³, Michael Hoffmeister⁷, John L. Hopper²¹, Thomas J. Hudson^{22,23}, David J. Hunter¹⁸, Rebecca D. Jackson²⁴, Sun Ha Jee²⁵, Mark A. Jenkins²¹, Wei-Hua Jia²⁶, Laurence N. Kolonel²⁷, Charles Kooperberg¹, Sébastien Küry⁶, Andrea Z. Lacroix¹, Cathy C. Laurie¹⁹, Cecelia A. Laurie¹⁹, Loic Le Marchand²⁷, Mathieu Lemire²², David Levine¹⁹, Noralane M. Lindor²⁸, Yan Liu²⁹, Jing Ma¹¹, Karen W. Makar¹, Keitaro Matsuo³⁰, Polly A. Newcomb^{1,2}, John D. Potter^{1,31}, Ross L. Prentice¹, Conghui Qu¹, Thomas Rohan³², Stephanie A. Rosse^{1,2}, Robert E. Schoen³³, Daniela Seminara³⁴, Martha Shrubsole¹⁵, Xiao-Ou Shu¹⁵, Martha L. Slattery³⁵, Darin Taverna¹⁴, Stephen N. Thibodeau³⁶, Cornelia M. Ulrich^{1,2,37}, Emily White^{1,2}, Yongbing Xiang³⁸, Brent W. Zanke³⁹, Yi-Xin Zeng²⁶, Ben Zhang¹⁵, Wei Zheng⁴⁰, and Li Hsu¹ on behalf of the Colon Cancer Family Registry and the Genetics and Epidemiology of Colorectal Cancer Consortium

¹Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington
²School of Public Health, University of Washington, Seattle, Washington
³Keck School of Medicine, University of Southern California, Los Angeles, California
⁴Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, North Carolina
⁵Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland
⁶Service de Génétique Médicale, CHU Nantes, Nantes, France
⁷Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany
⁸Division of Research, Kaiser Permanente Medical Care Program, Oakland, California
⁹American Cancer Society, Atlanta, Georgia
¹⁰Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts
¹¹Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts
¹²Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany
¹³Department of Health Studies, University of Chicago, Chicago, Illinois
¹⁴Translational Genomics Research Institute, Phoenix,

© 2013 by the AGA Institute

Address requests for reprints to: Ulrike Peters, PhD, MPH, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M4-B402, PO Box 19024, Seattle, Washington 98109-1024. upeters@fhcrc.org; fax: (206) 667-7850; or Li Hsu, PhD, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M2-B500, Seattle, Washington 98109-1024. lih@fhcrc.org; fax: (206) 667-7004.

* Authors share co-first authorship.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <http://dx.doi:10.1053/j.gastro.2012.12.020>.

Conflicts of interest

The authors disclose no conflicts.

Arizona ¹⁵Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee ¹⁶Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts ¹⁷Department of Surgery, Toronto General Hospital, Toronto, Ontario, Canada ¹⁸School of Public Health, Harvard University, Boston, Massachusetts ¹⁹Department of Biostatistics, University of Washington, Seattle, Washington ²⁰Division of Epidemiology, New York University School of Medicine, New York, New York ²¹Melborne School of Population Health, University of Melbourne, Melbourne, Victoria, Australia ²²Ontario Institute for Cancer Research, Toronto, Ontario, Canada ²³Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Toronto, Ontario, Canada ²⁴Division of Endocrinology, Diabetes, and Metabolism, Ohio State University, Columbus, Ohio ²⁵Institute for Health Promotion, Yonsei University, Seoul, Korea ²⁶Cancer Center, Sun Yat-sen University, Guangzhou, China ²⁷Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii ²⁸Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona ²⁹Stephens and Associates, Carrollton, Texas ³⁰Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan ³¹Centre for Public Health Research, Massey University, Wellington, New Zealand ³²Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York ³³Department of Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania ³⁴Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland ³⁵Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah ³⁶Departments of Laboratory Medicine and Pathology and Laboratory Genetics, Mayo Clinic, Rochester, Minnesota ³⁷Division of Preventive Oncology, National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany ³⁸Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China ³⁹Division of Hematology, Faculty of Medicine, The University of Ottawa, Ottawa, Ontario, Canada ⁴⁰Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee

Abstract

BACKGROUND & AIMS—Heritable factors contribute to the development of colorectal cancer. Identifying the genetic loci associated with colorectal tumor formation could elucidate the mechanisms of pathogenesis.

METHODS—We conducted a genome-wide association study that included 14 studies, 12,696 cases of colorectal tumors (11,870 cancer, 826 adenoma), and 15,113 controls of European descent. The 10 most statistically significant, previously unreported findings were followed up in 6 studies; these included 3056 colorectal tumor cases (2098 cancer, 958 adenoma) and 6658 controls of European and Asian descent.

RESULTS—Based on the combined analysis, we identified a locus that reached the conventional genome-wide significance level at less than 5.0×10^{-8} : an intergenic region on chromosome 2q32.3, close to *nucleic acid binding protein 1* (most significant single nucleotide polymorphism: rs11903757; odds ratio [OR], 1.15 per risk allele; $P = 3.7 \times 10^{-8}$). We also found evidence for 3 additional loci with P values less than 5.0×10^{-7} : a locus within the *laminin gamma 1* gene on chromosome 1q25.3 (rs10911251; OR, 1.10 per risk allele; $P = 9.5 \times 10^{-8}$), a locus within the *cyclin D2* gene on chromosome 12p13.32 (rs3217810 per risk allele; OR, 0.84; $P = 5.9 \times 10^{-8}$), and a locus in the *T-box 3* gene on chromosome 12q24.21 (rs59336; OR, 0.91 per risk allele; $P = 3.7 \times 10^{-7}$).

CONCLUSIONS—In a large genome-wide association study, we associated polymorphisms close to *nucleic acid binding protein 1* (which encodes a DNA-binding protein involved in DNA repair) with colorectal tumor risk. We also provided evidence for an association between colorectal tumor risk and polymorphisms in *laminin gamma 1* (this is the second gene in the

laminin family to be associated with colorectal cancers), *cyclin D2* (which encodes for cyclin D2), and *T-box 3* (which encodes a T-box transcription factor and is a target of Wnt signaling to β -catenin). The roles of these genes and their products in cancer pathogenesis warrant further investigation.

Keywords

Colon Cancer; Genetics; Risk Factors; SNP

Colorectal cancer has a sizable heritable component; a large twin study estimated that 35% of colorectal cancer risk may be explained by heritable factors.¹ Over the past several years, genome-wide association studies (GWAS), which focus on common single-nucleotide polymorphisms (SNPs), successfully have discovered low-penetrance loci for colorectal cancer.²⁻¹² These analyses have highlighted genes within the known *transforming growth factor- β* and *Wnt* signaling pathways (eg, *bone morphogenetic protein 2 & 4*, *SMAD7*), as well as regions and genes not previously strongly implicated in colorectal cancer (eg, *zinc finger protein 90*, *laminin alpha 5*, *disco-interacting protein 2*), thereby highlighting pathways previously not understood to be involved in colorectal carcinogenesis.²⁻¹²

To identify additional common genetic risk factors for colorectal tumors, we conducted a genome-wide scan across 14 independent studies including nearly 28,000 subjects and follow-up evaluation of nearly 10,000 independent subjects. We included both colorectal cancer cases and colorectal adenoma cases. Colorectal adenoma is a well-defined colorectal cancer precursor¹³ and the majority of colorectal cancers develop through the adenoma-cancer sequence.¹⁴ It has been estimated that the 10-year cumulative rate for advanced adenoma to transition to colorectal cancer is between 10% and 45%, depending on age and sex.^{13,15,16} Accordingly, the 2 phenotypes have overlapping etiology.¹⁷ Inclusion of adenoma cases can increase sample size, and hence statistical power, to identify genetic risk factors related to early events in the adenoma-carcinoma process, during which risk factor intervention strategies may offer the greatest potential benefit for cancer prevention.

Materials and Methods

Study Participants

Each study is described in detail in the Supplementary Materials and Methods section and the number of cases and controls as well as age and sex distributions are listed in Supplementary Table 1. In brief, colorectal cancer cases were defined as colorectal adenocarcinoma and confirmed by medical records, pathologic reports, or death certificate. Colorectal adenoma cases were confirmed by medical records, histopathology, or pathologic reports. Controls for adenoma cases had a negative colonoscopy (except for the Nurses' Health Study and the Health Professionals Follow-up Study controls matched to cases with distal adenoma, which either had a negative sigmoidoscopy or colonoscopy examination). All participants provided written informed consent and studies were approved by their respective institution's Institutional Review Boards.

Genotyping

GWAS in the Genetics and Epidemiology of Colorectal Cancer Consortium and the Colon Cancer Family Registry—We conducted a meta-analysis of GWAS from 13 studies within the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (10,729 cases and 13,328 controls) and additional GWAS within the Colon Cancer Family Registry (CCFR) (1967 cases and 1785 controls). Details on genotyping, quality assurance/quality control, and imputation can be found in the Supplementary

Materials and Methods section. Average sample and SNP call rates and concordance rates for blinded duplicates are listed in Supplementary Table 2. In brief, all analyses were restricted to European ancestry. Genotyped SNPs were excluded based on call rate (<98%), lack of Hardy–Weinberg equilibrium in controls (HWE, $P < 1 \times 10^{-4}$), and low minor allele frequency (MAF). Because imputation of genotypes is established as standard practice in the analysis of genotype array data, we imputed the autosomal SNPs of all studies to the Utah residents with Northern and Western European ancestry from the Centre d'étude du polymorphisme humain (CEPH) collection (CEU) population in HapMap II (available at: <http://hapmap.ncbi.nlm.nih.gov/>). Imputed SNPs were restricted based on MAF (>1%) and imputation accuracy ($R^2 > 0.3$). After imputation and quality control (QC), a total of 2,708,280 SNPs were used in the meta-analysis of GECCO studies and CCFR. In our detailed result table (Supplementary Table 3), we list for each SNP the number of studies with directly genotyped or imputed data and the mean imputation R^2 . These data show, as expected, that imputed SNPs tend to show very similar results as SNPs that were directly genotyped if the correlation is high between SNPs.

Follow-up studies—We selected the 10 most statistically significant regions (excluding known GWAS loci) based on the P value from the GECCO and CCFR meta-analysis for further follow-up evaluation in colorectal cancer studies in the Asian colorectal cancer consortium and a US-based colorectal adenoma study. Details on genotyping, quality assurance/quality control, and imputation can be found in the Supplementary Materials and Methods section. After quality control exclusions, 2098 colorectal cancer cases and 5749 controls, and 958 colorectal adenoma cases and 909 controls remained in the analysis.

Statistical Analysis

GWAS in GECCO and CCFR—Statistical analyses of the GECCO and CCFR samples were conducted centrally at the coordinating center on individual-level data to ensure a consistent analytic approach. For each study, we estimated the association between SNPs and risk for colorectal cancer by calculating β values, odds ratios (ORs), standard errors, 95% confidence intervals, and P values using logistic regression models with log-additive genetic effects. Each directly genotyped SNP was coded as 0, 1, or 2 copies of the risk allele. For imputed SNPs, we used the expected number of copies of the risk allele (the dosage), which has been shown to provide unbiased estimates in the association test for imputed SNPs.¹⁸ We adjusted for age, sex (when appropriate), center (when appropriate), smoking status (Physicians' Health Study only), batch effects (The french Association Study Evaluating RISK for sporadic colorectal cancer), and the first 3 principal components from EIGENSTRAT (available at: <http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm>) to account for population substructure. Because CCFR set 2 is a family-based study, we used a conditional logistic regression stratified by family identification while adjusting for age and sex. When analyzing genotyped SNPs on the X chromosome we need to account for different genotype variances between males and females. Therefore, we used the 1 degree of freedom modified Cochran–Armitage test¹⁹ to test for associations. This method has been shown to have robust and powerful performance across a wide range of scenarios.²⁰ We used logistic regression to model SNP \times SNP interaction effects for a log-additive model, in which the interaction term is the product of the 2 SNPs.

Quantile-quantile plots were assessed to determine whether the distribution of the P values in each study was consistent with the null distribution (except for the extreme tail). We also calculated the genomic inflation factor (λ) to measure the overdispersion of the test statistics from the association tests by dividing the median of the squared Z statistics by 0.455, the median of a chi-squared distribution with 1 degree of freedom. The inflation factor λ was

between 0.999 and 1.044 for individual studies based on all SNPs including both directly genotyped and imputed, indicating there is little evidence of residual population substructure, cryptic relatedness, or differential genotyping between cases and controls. This result was consistent with the visual inspection of the study-specific quantile-quantile plots.

We conducted inverse-variance weighted, fixed-effects meta-analysis to combine β estimates and standard errors across individual studies. In this approach, we weighed the β estimate of each study by its inverse variance and calculated a combined estimate by summing the weighted β estimates and dividing by the summed weights. For imputed SNPs, it has been shown that the inverse variance is approximately proportional to the imputation quality.¹⁸ Thus, the inverse variance weighting scheme automatically incorporates imputation quality in the meta-analysis for imputed SNPs. We calculated the heterogeneity P values based on Cochran's Q statistic²¹ and investigated sources for heterogeneity if the P value was less than .05 for the 10 most significant SNPs. For the most significant SNPs highlighted in this article, we also examined recessive and unrestricted genetic models and compared models by calculating the Akaike information criterion. We used PLINK (available at: pnu.mgh.harvard.edu/~purcell/plink/)²² and R (available at: <http://www.r-project.org/>)²³ to conduct the statistical analysis and summarized results graphically using LocusZOOM (available at: <http://csg.sph.umich.edu/locuszoom/>).²⁴

Follow-up studies—The 10 most significant SNPs from the GWAS meta-analysis described earlier were analyzed in the follow-up studies (P values from GWAS meta-analysis 2.5×10^{-7} to 6.5×10^{-6}). For the Asian colorectal cancer follow-up study, genotyped SNPs and dosage data of imputed SNPs were analyzed using the program mach2dat (available at: <http://www.sph.umich.edu/csg/abecasis/MACH/download/>).²⁵ The association between SNP and colorectal cancer risk was assessed using logistic regression with log-additive genetic effects after adjusting for age and sex. Meta-analyses were performed using the inverse-variance method based on a fixed-effects model, and calculations were implemented in the METAL package (available at: <http://www.sph.umich.edu/csg/abecasis/metal/>).²⁶ Because the MAF in the Asian follow-up population was very low for the locus on chromosome 14q23.1 (MAF, 0 – 0.006 in Han Chinese individuals from Beijing, China), we excluded this SNP from the follow-up evaluation in the Asian studies. Given potential differences in the linkage disequilibrium structure between European and Asian descent subjects, we also included all SNPs correlated with these 10 selected SNPs ($r^2 > 0.5$ in CEU).

For the adenoma follow-up study (all European descent), the association between each genetic marker and risk for colorectal adenoma was estimated by calculating ORs and 95% confidence intervals, using a log-additive genetic model. SNPTTESTv2.2.0 (available at: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) with the “-method score” option²⁷ was used for logistic regression with the frequentist test, and the model was adjusted for age and sex.

For a combined analysis of GWAS and follow-up results, we conducted inverse-variance weighted fixed-effects meta-analysis to combine ORs from log-additive models across individual studies and measured heterogeneity using Cochran's Q statistic, as discussed earlier.

Criterion for genome-wide significance—Based on an increasing number of articles^{28–33} providing a detailed discussion on the appropriate genome-wide significance threshold, which all arrive at similar values in the range of 5×10^{-7} to 5×10^{-8} for European populations, we decided to use a P value of 5×10^{-8} as the genome-wide significance threshold. In addition, we reported on SNPs with P values between less than $5 \times$

10^{-7} and greater than 5×10^{-8} as a potentially novel SNP that merited additional follow-up evaluation.

Heritability estimates—We estimated the additive heritability of colorectal cancer explained by all genotyped SNPs using the method by Yang et al³⁴ and implemented in the Genome-wide Complex Trait Analysis tool.³⁵ We set the prevalence of colorectal cancer to 0.004, based on Surveillance, Epidemiology and End Results incidence and National Center for Health Statistics mortality statistics.³⁶ We used all genotyped SNPs of Darmkrebs: Chancen der Verhütung durch Screening set II and Diet, Activity, and Lifestyle Study set I given the sizable sample set, different genotyping platforms, and inclusion of both sexes (Supplementary Table 1). We also estimated the heritability of previously and newly identified variants by using the method of So et al.³⁷ Furthermore, we used the method described by Park et al³⁸ to estimate the total number of loci expected to be identified for colorectal cancer based on the observed effect sizes and power for identifying the loci known to date (Table 1 and Supplementary Table 4).

Functional annotation of findings—We conducted a functional annotation for all tagging SNPs (and correlated SNPs) highlighted in this article. As detailed in the Supplementary Materials and Methods section, we queried multiple bioinformatic databases based on the University of California, Santa Cruz genome browser.

Results

Summary results of the GWAS meta-analysis of GECCO and CCFR are shown in the Manhattan plot (Supplementary Figure 1). Several of the previously identified GWAS SNPs were highly significantly associated with colorectal cancer, and overall we found a nominal significant association ($P < .05$) in the same direction for 16 of 18 previously identified GWAS loci (Supplementary Table 4). After excluding previously identified regions, we followed up the 10 most significant regions from the GWAS meta-analysis ($P = 2.5 \times 10^{-7}$ to 6.5×10^{-6} ; Supplementary Table 3). In 4 regions the follow-up studies showed evidence of replication with the association in the same directions as the GWAS and an overall improved significance level (Table 1). Of these 4 regions, 1 region reached the conventional genome-wide significance level at a P value less than 5.0×10^{-8} in the combined analysis (GWAS + follow-up evaluation). This region was on chromosome 2q32.3 (rs11903757: OR, 1.16 per risk allele; $P = 3.7 \times 10^{-8}$; Table 1 and Supplementary Figure 2). The SNP showed no evidence for heterogeneity ($P = .27$) across all studies. The SNP was correlated strongly ($r^2 > 0.9$) with several SNPs in the same region, which showed similar results (Supplementary Figure 3 and Supplementary Table 3).

The other 3 regions had P values less than 5.0×10^{-7} (and $P > 5.0 \times 10^{-8}$) in the combined analysis (GWAS + follow-up evaluation). Reporting by chromosomal position, the first of these 3 regions was on chromosome 1q25.3. In this region, the association with rs10911251 had the lowest P value (OR, 1.09 per risk allele; $P = 9.5 \times 10^{-8}$; Table 1 and Supplementary Figure 2), showing no evidence of heterogeneity ($P = .69$) across studies. This was correlated strongly with a large number of SNPs in the same region showing similar allele frequencies, risk estimates, and P values spanning across the entire laminin gamma 1 (*LAMC1*) gene (Supplementary Figure 3 and Supplementary Table 3).

The second region with P values less than 5.0×10^{-7} and greater than 5.0×10^{-8} was on chromosome 12p13, within the *cyclin D2* (*CCND2*) gene. The most statistically significant SNP was rs3217810 (OR, 1.20 per risk allele; $P = 5.9 \times 10^{-8}$; Table 1 and Supplementary Figure 2). Furthermore, only 17.1 kb apart resides a second SNP, rs3217901, which was not strongly correlated with rs3217810 ($r^2 = 0.052 - 0.063$) and showed a slightly lower

significance level (OR, 1.10 per risk allele; $P = 4.9 \times 10^{-7}$). Although the risk allele frequency of rs3217810 in our European descent studies was on average 0.16, this SNP is very uncommon in Asian populations (0.03 in Japanese in Tokyo, Japan, and 0.01 in Han Chinese individuals from Beijing, China) and, hence, the follow-up evaluation of rs3217810 did not include the Asian cases and controls. Both SNPs were not heterogeneous across studies (P for heterogeneity = .51 and .91). When we included both SNPs simultaneously in the logistic regression analysis the significance of both SNPs was reduced (Supplementary Table 5).

The third region with P values less than 5.0×10^{-7} was in the *T-box 3 (TBX3)* gene on chromosome 12q24.21. The most statistically significant SNP in this region was rs59336 (OR, 1.09 per risk allele; $P = 3.7 \times 10^{-7}$; Table 1 and Supplementary Figure 2). Again, we observed no evidence for heterogeneity across studies ($P = .39$).

We investigated if the 4 regions listed earlier might be more significant (lower P value) under a different genetic model than the log-additive model. None of the variants was more significant when we modeled the unrestricted, dominant, or recessive mode of inheritance (Supplementary Table 6).

When we stratified results by colorectal adenoma and cancer we observed stronger associations for adenoma compared with cancer for rs11903757 at 2q32.3, similar associations for rs3217810 and rs3217901 at 12p13/*CCND2* and for rs59336 at 12q24.21/*TBX3*, and a weaker association for rs10911251 at 1q25.3/*LAMC1* (Supplementary Table 7). For previously identified loci, in particular, associations for rs16892766 at 8q23.3/*EIF3H* and rs4939827 at 18q21/*SMAD7* tended to be stronger for adenoma, whereas associations for other loci tended to be similar or weaker compared with cancer (Supplementary Table 4).

We observed no evidence for interaction between the SNPs in the newly identified regions or with SNPs in previously identified regions. The smallest P value for interaction was .017 for rs59336/*TBX3* and rs11632715/15q13 and was not significant after accounting for multiple comparisons.

As popularized by Yang et al,³⁴ we estimated that the additive heritability of colorectal cancer explained by all genotyped SNPs would be 14.2% (standard error, 8.2%). The newly identified loci (Table 1) and previously identified loci (Supplementary Table 4) explained about 11% of the additive heritability and cumulatively these newly and previously identified loci explain 1.6% of the variation of colorectal cancer. Based on the study by Park et al³⁸ we estimated that the total number of loci expected to be identified for colorectal cancer would be between 239 and 500 if the type I error rate was between 5×10^{-7} and 5×10^{-8} .

Discussion

In this large genome-wide scan meta-analysis and follow-up evaluation of a total of close to 38,000 subjects, we identified an intergenic region on chromosome 2q32.3 close to *nucleic acid binding protein 1 (NABPI)* that was associated with colorectal tumor risk with P values less than 5.0×10^{-8} , the conventional genome-wide significance level. Furthermore, we identified 3 regions with P values less than 5.0×10^{-7} : one on chromosome 1q31 in *LAMC1*, a second on chromosome 12p13 in *CCND2*, and a third on chromosome 12q24.21 in *TBX3*. All showed highly significant associations with P values less than 5×10^{-7} .

Our study provides strong support for an intergenic locus on chromosome 2q32.3. The most significant SNPs in this region are in closest proximity to the *NABPI* gene (44 kb centromeric) and the gene *serum deprivation response* (112 kb telomeric), which encodes for

the serum-deprivation response phosphatidylserine-binding protein. The SNPs are downstream of *NABP1*, which also is known as *human single-strand DNA binding protein 2* or *oligonucleotide/oligosaccharide binding fold-containing protein 2A*. This protein binds single-stranded DNA via the oligonucleotide/oligosaccharide binding fold domain.³⁹ Single-stranded DNA binding proteins are important for diverse DNA processes, such as DNA replication, recombination, transcription, and repair.^{40–42} Cells depleted of *NABP1* show hypersensitivity to DNA-damaging reagents; *NABP1* participates in repair of DNA double-strand breaks and *ataxia telangiectasia mutated*-dependent signaling pathways,⁴³ similar to the role of its homolog, *NABP2* (which is also known as human single-strand DNA binding protein 1).³⁹ Although our functional annotation did not provide further insights on the function of the SNPs, the biologic data described earlier support the importance of *NABP1* with respect to genomic stability, which could explain a link to the development of cancer.⁴⁴

In addition to the genome-wide significant region we observed 3 regions that were slightly less significant with *P* values less than 5×10^{-7} but greater than 5×10^{-8} . As has been shown previously,⁴⁵ a large fraction of SNPs with borderline genome-wide-significant associations replicated when results from additional studies were added, suggesting that further follow-up evaluation of these regions is warranted. The first of these 3 regions was on chromosome 1q31 and included correlated SNPs showing associations that spanned across the *LAMC1* gene. Interestingly, previous genome-wide scans of colorectal cancer identified a different laminin gene on chromosome 20q13.33, *laminin alpha 5*, as associated with colorectal cancer,^{9,11} supporting the importance of this gene family for the development of colorectal cancer. Laminins are extracellular matrix glycoproteins that constitute a major component of the basement membrane in most tissues⁴⁶ and in the colon are part of the intestinal epithelial barrier. Laminins are involved in a wide variety of biological functions, such as regulation of cell adhesion, differentiation, migration, signaling, and metastasis.^{47–50} Loss of cell-surface laminin anchoring has been found in many cancer cells, particularly those with aggressive subtypes.⁵¹

LAMC1 is a large gene spanning 122 kb and containing 28 coding exons. rs10911251 is correlated strongly ($r^2 > 0.8$) with several other SNPs across the gene (Supplementary Figure 3 and Supplementary Table 3). Upon functional annotation, we identified a potential functional candidate (rs10911205) that is correlated strongly with the most significant tagSNP ($r^2 = 0.73$) and located 72 kb upstream within the first intron of *LAMC1*. As shown in the University of California, Santa Cruz Genome Browser view (Supplementary Figure 4), rs10911205 is located within a highly evolutionarily conserved region and, given its close proximity to the promoter, it is possible that this region influences gene transcription. In addition, the patterns of histone modifications and DNase signals indicating accessibility for transcription factors suggest that this variant may affect cell-type-specific enhancer activity. In summary, given the statistical evidence, support from functional annotation, and evidence from a previous GWAS that identified another laminin gene to be associated with colorectal cancer, we believe there is strong support for the importance of *LAMC1* in the development of colorectal cancer. It is of note that the biologic role of this gene family has not yet been studied substantially in relation to colorectal cancer, supporting the novelty of this finding.

A second region with *P* values less than 5×10^{-7} was on chromosome 12p13.32, with 2 independent SNPs both located in the intron of *CCND2*, which belongs to the highly conserved cyclin family, specifically encoding for the protein cyclin D2. Through regulation of CDK4 and CDK6, cyclin D2 affects the cell-cycle transition of the G1/S phase.^{52,53} Furthermore, cyclin D2 interacts with tumor-suppressor protein retinoblastoma. Recent studies have identified *CCND2* as a microRNA target gene in different colorectal cancer cell lines.^{54,55} Interestingly, genetic variants in *CCND1* also have been related to colorectal

cancer^{56,57} and a previous GWAS identified a SNP in *CCND1* to be associated with breast cancer.⁵⁸

The third region with *P* values less than 5×10^{-7} we identified was within the *TBX3* gene, which encodes the T-box transcription factor. *TBX3* is overexpressed in several cancers, including pancreas, liver, breast cancer, and melanoma,⁵⁹ playing multiple roles in normal development and cancer.⁶⁰ In liver cancer, *TBX3* was identified as a downstream target of the Wnt/ β -catenin pathway, mediating β -catenin activities on cell proliferation and survival.⁶¹ The Wnt/ β -catenin pathway plays a key role in colorectal cancer development.⁶² *TBX5*, another member of the T-Box gene family, has been suggested as an epigenetically inactivated tumor-suppressor gene in colon cancer⁶³ and provides an additional mechanism by which this gene family may influence colorectal cancer development.

Our study adds further support for all, except 3, previously identified GWAS loci for colorectal cancer. The 3 SNPs (on chromosomes 1q41, 3q26.2, and 6p21) that did not replicate are among the more recently identified GWAS loci^{9,12} and have smaller effect sizes (OR for risk allele, 1.1) compared with the earlier GWAS findings. As a result, larger sample sizes may be needed to fully replicate these SNPs. Furthermore, it is possible that the effect varies by environmental exposures, which may differ among the study populations. Overall, effect sizes from our study for previous GWAS loci tend to be weaker than in the initial reports, which may be explained by the fact that previous results were subject to the “winner’s curse.”⁶⁴

The large sample size of our GWAS and follow-up studies and availability of individual-level GWAS data are important advantages of our study. However, the study also had limitations. To increase the sample size, we included Asian descent subjects, who may have different linkage disequilibrium patterns, and the SNPs analyzed may be tagging different underlying causal variants. To address this potential limitation we included all SNPs correlated with the most significant SNPs, which likely will identify any variant that genuinely is associated with colorectal cancer risk across different ancestral groups, as shown for other GWAS loci.^{65–68} Given that genotyping platforms only capture a subset of the genome, we used imputation to HapMap II to obtain a better coverage of the common variation across the genome and to generate a common set of SNPs from the different platforms. Because imputed SNPs tend to result in less significant findings depending on their imputation accuracy,⁶⁹ we expect that our results provide relative conservative significance levels.⁷⁰ Similar to previous GWAS,^{2,4,6–10,12} we included colorectal adenoma as the major precursor of colorectal cancer to improve our statistical power and to identify genetic variants that act early in the adenoma-cancer sequence, where adenomas and cancer have a shared etiology. Although the inclusion of adenoma also may add heterogeneity because adenomas will not show an association for genetic variants that act later in the carcinogenic process (ie, on progression from adenoma to cancer) or for variants that act through adenoma-independent pathways, stratified analysis may provide insights into the mediating roles of genes within the normal to adenoma to cancer pathway. We show that for some of the newly and previously identified loci, associations are stronger for adenomas compared with cancer; however, we observed similar or weaker associations for other loci. These results may suggest that some genes are important in early stages of cancer development while others may be more important for the progression from adenoma to cancer. However, given the relatively small number of adenoma cases (only 6.5% of the GWAS and 31% of the follow-up cases were adenoma cases), it is important that our findings are replicated in studies with larger numbers of adenoma cases.

In summary, in this large study, we identified one novel susceptibility locus associated with the risk of colorectal tumor on chromosome 2q32.3 close to *NABPI*, and 3 potential loci

with borderline genome-wide significant results within *LAMC1*, *CCND2*, and *TBX3*. These findings are supported by biologic plausibility, functional annotation, and previous GWAS findings within the same gene family, emphasizing the potential relevance of these genes in the etiology of colorectal cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank the following:

Asian Consortium: The authors wish to thank the study participants and research staff for their contributions and commitment to this project, Regina Courtney for DNA preparation, and Jing He for data processing and analyses.

The french Association STudy Evaluating RISK for sporadic colorectal cancer: The authors are very grateful to Dr Bruno Buecher without whom this project would not have existed. The authors also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians, and students.

Darmkrebs: Chancen der Verhütung durch Screening: The authors thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benschaid, Muhabbet Celik, and Ursula Eilber for excellent technical assistance.

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping to bring together the data and people who made this project possible.

Health Professionals Follow-up Study, Nurses' Health Study, and Physicians' Health Study: The authors would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for Nurses' Health Study, Health Professionals Follow-up Study, and Physician's Health Study under the supervision of Dr Immaculata Devivo and Dr David Hunter, Qin (Carolyn) Guo, and Lixue Zhu who assisted in programming for Nurses' Health Study and Health Professionals Follow-up Study, and Haiyan Zhang who assisted in programming for the Physicians' Health Study. The authors would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-up Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, ME, MI, NC, ND, NE, NH, NJ, NY, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY.

Prostate, Lung, Colorectal Cancer, and Ovarian Cancer Screening Trial: The authors thank Drs Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff of the Prostate, Lung, Colorectal Cancer, and Ovarian Cancer Screening Trial, Mr Tom Riley and staff of Information Management Services, Inc, Ms Barbara O'Brien and staff of Westat, Inc, and Drs Bill Kopp, Wen Shao, and staff of SAIC-Frederick. Most importantly, the authors acknowledge the study participants for their contributions to making this study possible.

Postmenopausal Hormone study: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

Tennessee Colorectal Polyp Study: The authors thank the study participants and the research staff for their contributions and commitment to this project, and Regina Courtney for DNA preparation.

Women's Health Initiative: The authors thank the Women's Health Initiative investigators and staff for their dedication, and the study participants for making the program possible. A full listing of Women's Health Initiative investigators can be found at: <https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf>.

Funding

The Genetics and Epidemiology of Colorectal Cancer Consortium study was supported by the National Cancer Institute, National Institutes of Health, and the US Department of Health and Human Services (U01 CA137088; R01 CA059045). The Asian Consortium was supported by a Grant-in-aid for Cancer Research, the Grant for the Third Term Comprehensive Control Research for Cancer and Grants-in-Aid for Scientific Research from the

Japanese Ministry of Education, Culture, Sports, Science and Technology (17015018 and 221S0001). The French Association STudy Evaluating RISK for sporadic colorectal cancer was supported by a Hospital Clinical Research Program (PHRC) and by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte Contre le Cancer, the Association Anne de Bretagne Génétique, and the Ligue Régionale Contre le Cancer. The Assessment of Risk in Colorectal Tumours in Canada study was supported by the National Institutes of Health through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783; see the Colon Cancer Family Registry support section below); and by a GL2 grant from the Ontario Research Fund, the Canadian Institutes of Health Research, by a Cancer Risk Evaluation Program grant from the Canadian Cancer Society Research Institute, and by Senior Investigator Awards (T.J.H. and B.W.Z.) from the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Economic Development and Innovation. The Hawaii Colorectal Cancer Studies 2 and 3 studies were supported by the National Institutes of Health (R01 CA60987). The Colon Cancer Family Registry was supported by the National Institutes of Health (RFA CA-95-011) and through cooperative agreements with members of the Colon Cancer Family Registry and P.I.s. This genome-wide scan was supported by the National Cancer Institute, National Institutes of Health (U01 CA122839). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the cancer family registries, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the cancer family registries. The following colon cancer family registries centers contributed data to this article and were supported by National Institutes of Health: the Australasian Colorectal Cancer Family Registry (U01 CA097735), Seattle Colorectal Cancer Family Registry (U01 CA074794), and the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783). The Darmkrebs: Chancen der Verhütung durch Screening study was supported by the German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814). The Diet, Activity, and Lifestyle Study was supported by the National Institutes of Health (R01 CA48998 to M.L.S.); Guangzhou-1 was supported by the National Key Scientific and Technological Project (2011ZX09307-001-04) and the National Basic Research Program (2011CB504303) was supported by the People's Republic of China. The Health Professionals Follow-up Study was supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, and P50 CA 127003), the Nurses' Health Study was supported by the National Institutes of Health (R01 137178, P01 CA 087969, and P50 CA 127003), and the Physicians' Health Study was supported by the National Institutes of Health (CA42182). The Korean Cancer Prevention Study-II study was supported by the National R&D Program for cancer control (1220180), and the Seoul R&D Program (10526, Republic of Korea). The Multiethnic Cohort study was supported by the National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA63464). The Prostate, Lung, Colorectal Cancer, and Ovarian Cancer Screening Trial was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Department of Health and Human Services. Control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility prostate cancer scan, supported by the Intramural Research Program of the National Cancer Institute. The data sets used in this analysis were accessed with appropriate approval through the dbGaP online resource (http://www.cgems.cancer.gov/data_access.html) through dbGaP accession number 000207v.1p1. Control samples also were genotyped as part of the GWAS of Lung Cancer and Smoking (Yeager, M et al. Nat Genet 2008;124:161–170). Support for this work was provided through the National Institutes of Health, Genes, Environment and Health Initiative (Z01 CP 010200). The human subjects participating in the genome-wide association study were derived from the Prostate, Lung, Colon, and Ovarian Screening Trial and the study was supported by intramural resources of the National Cancer Institute. Assistance with genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, Geneva Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the National Institutes of Health, Genes, Environment and Health Initiative (U01 HG 004438). The data sets used for the analyses described in this article were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs000093 v2.p2. The Postmenopausal Hormone Study was supported by the National Institutes of Health (R01 CA076366 to P.A.N.). The Shanghai-1 and Shanghai-2 studies were supported by the National Institutes of Health (R37CA070867, R01CA082729, R01CA124558, R01CA148667, and R01CA122364), as well as an Ingram Professorship and Research Reward funds from the Vanderbilt University School of Medicine. The Tennessee Colorectal Polyp Study was supported by the National Institutes of Health (P50CA95103 and R01CA121060) and was conducted by the Survey and Biospecimen Shared Resource, which was supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA 68485). The VITamins And Lifestyle study was supported by the National Institutes of Health (K05 CA154337). The Women's Health Initiative program was funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services, through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

Abbreviations used in this paper

CFR	Colon Cancer Family Registry
CCND	cyclin D2
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
CEPH	Centre d'étude du polymorphisme humain
GECCO	Genetics and Epidemiology of Colorectal Cancer Consortium
GWAS	genome-wide association study
HWE	Hardy Weinberg Equilibrium
LAMC1	laminin gamma 1
MAF	minor allele frequency
NABP1	nucleic acid binding protein 1
ORs	odds ratios
QC	quality control
SNP	single nucleotide polymorphism
TBX3	T-box 3

References

1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343:78–85. [PubMed: 10891514]
2. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007; 39:984–988. [PubMed: 17618284]
3. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007; 39:989–994. [PubMed: 17618283]
4. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet.* 2007; 39:1315–1317. [PubMed: 17934461]
5. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008; 40:631–637. [PubMed: 18372901]
6. Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet.* 2008; 40:26–28. [PubMed: 18084292]
7. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet.* 2008; 40:623–630. [PubMed: 18372905]
8. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet.* 2008; 40:1426–1435. [PubMed: 19011631]
9. Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet.* 2010; 42:973–977. [PubMed: 20972440]

10. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.* 2011; 7:e1002105. [PubMed: 21655089]
11. Peters U, Hutter CM, Hsu L, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet.* 2012; 131:217–234. [PubMed: 21761138]
12. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet.* 2012; 44:770–776. [PubMed: 22634755]
13. Brenner H, Hoffmeister M, Stegmaier C, et al. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. *Gut.* 2007; 56:1585–1589. [PubMed: 17591622]
14. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell.* 1996; 87:159–170. [PubMed: 8861899]
15. Eide TJ. Risk of colorectal cancer in adenoma-bearing individuals within a defined population. *Int J Cancer.* 1986; 38:173–176. [PubMed: 3733258]
16. Stryker SJ, Wolff BG, Culp CE, et al. Natural history of untreated colonic polyps. *Gastroenterology.* 1987; 93:1009–1013. [PubMed: 3653628]
17. Potter JD. Colorectal cancer: molecules and populations. *J Natl Cancer Inst.* 1999; 91:916–932. [PubMed: 10359544]
18. Jiao S, Hsu L, Hutter CM, et al. The use of imputed values in the meta-analysis of genome-wide association studies. *Genet Epidemiol.* 2011; 35:597–605. [PubMed: 21769935]
19. Clayton D. Testing for association on the X chromosome. *Biostatistics.* 2008; 9:593–600. [PubMed: 18441336]
20. Hickey PF, Bahlo M. X chromosome association testing in genome wide association studies. *Genet Epidemiol.* 2011; 35:664–670. [PubMed: 21818774]
21. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS ONE.* 2007; 2:e841. [PubMed: 17786212]
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]
23. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
24. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26:2336–2337. [PubMed: 20634204]
25. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34:816–834. [PubMed: 21058334]
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26:2190–2191. [PubMed: 20616382]
27. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007; 39:906–913. [PubMed: 17572673]
28. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996; 273:1516–1517. [PubMed: 8801636]
29. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005; 437:1299–1320. [PubMed: 16255080]
30. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447:661–678. [PubMed: 17554300]
31. Hoggart CJ, Clark TG, De IM, et al. Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol.* 2008; 32:179–185. [PubMed: 18200594]
32. Pe'er I, Yelenski R, Altshuler D, et al. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol.* 2008; 32:381–385. [PubMed: 18348202]
33. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol.* 2008; 32:227–234. [PubMed: 18300295]

34. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010; 42:565–569. [PubMed: 20562875]
35. Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011; 88:76–82. [PubMed: 21167468]
36. Howlander, N.; Noone, AM.; Krapcho, M., et al. SEER cancer statistics review, 1975–2009. Bethesda, MD: National Cancer Institute; 2012.
37. So HC, Gui AH, Cherny SS, et al. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol.* 2011; 35:310–317. [PubMed: 21374718]
38. Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet.* 2010; 42:570–575. [PubMed: 20562874]
39. Richard DJ, Bolderson E, Cubeddu L, et al. Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature.* 2008; 453:677–681. [PubMed: 18449195]
40. Bochkarev A, Bochkareva E, Frappier L, et al. The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *EMBO J.* 1999; 18:4498–4504. [PubMed: 10449415]
41. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem.* 1997; 66:61–92. [PubMed: 9242902]
42. Yang H, Jeffrey PD, Miller J, et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science.* 2002; 297:1837–1848. [PubMed: 12228710]
43. Li Y, Bolderson E, Kumar R, et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem.* 2009; 284:23525–23531. [PubMed: 19605351]
44. Broderick S, Rehmet K, Concannon C, et al. Eukaryotic single-stranded DNA binding proteins: central factors in genome stability. *Subcell Biochem.* 2010; 50:143–163. [PubMed: 20012581]
45. Panagiotou OA, Ioannidis JP. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol.* 2012; 41:273–286. [PubMed: 22253303]
46. Kalluri R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat Rev Cancer.* 2003; 3:422–433. [PubMed: 12778132]
47. Turck N, Gross I, Gendry P, et al. Laminin isoforms: biological roles and effects on the intracellular distribution of nuclear proteins in intestinal epithelial cells. *Exp Cell Res.* 2005; 303:494–503. [PubMed: 15652360]
48. Pouliot N, Saunders NA, Kaur P. Laminin 10/11: an alternative adhesive ligand for epidermal keratinocytes with a functional role in promoting proliferation and migration. *Exp Dermatol.* 2002; 11:387–397. [PubMed: 12366691]
49. Gudjonsson T, Ronnov-Jessen L, Villadsen R, et al. Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. *J Cell Sci.* 2002; 115:39–50. [PubMed: 11801722]
50. Patarroyo M, Tryggvason K, Virtanen I. Laminin isoforms in tumor invasion, angiogenesis and metastasis. *Semin Cancer Biol.* 2002; 12:197–207. [PubMed: 12083850]
51. Akhavan A, Griffith OL, Soroceanu L, et al. Loss of cell-surface laminin anchoring promotes tumor growth and is associated with poor clinical outcomes. *Cancer Res.* 2012; 72:2578–2588. [PubMed: 22589276]
52. Lukas J, Muller H, Bartkova J, et al. DNA tumor virus oncoproteins and retinoblastoma gene mutations share the ability to relieve the cell's requirement for cyclin D1 function in G1. *J Cell Biol.* 1994; 125:625–638. [PubMed: 8175885]
53. Matsushime H, Quelle DE, Shurtleff SA, et al. D-type cyclin-dependent kinase activity in mammalian cells. *Mol Cell Biol.* 1994; 14:2066–2076. [PubMed: 8114738]
54. Ragusa M, Statello L, Maugeri M, et al. Specific alterations of the microRNA transcriptome and global network structure in colorectal cancer after treatment with MAPK/ERK inhibitors. *J Mol Med (Berl).* 2012; 90:1421–1438. [PubMed: 22660396]

55. Zhang P, Ma Y, Wang F, et al. Comprehensive gene and microRNA expression profiling reveals the crucial role of hsa-let-7i and its target genes in colorectal cancer metastasis. *Mol Biol Rep.* 2012; 39:1471–1478. [PubMed: 21625861]
56. Yang Y, Wang F, Shi C, et al. Cyclin D1 G870A polymorphism contributes to colorectal cancer susceptibility: evidence from a systematic review of 22 case-control studies. *PLoS One.* 2012; 7:e36813. [PubMed: 22606291]
57. Yang J, Zhang G, Chen J. CCND1 G870A polymorphism is associated with increased risk of colorectal cancer, especially for sporadic colorectal cancer and in Caucasians: a meta-analysis. *Clin Res Hepatol Gastroenterol.* 2012; 36:169–177. [PubMed: 22322158]
58. Turnbull C, Rapley EA, Seal S, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat Genet.* 2010; 42:604–607. [PubMed: 20543847]
59. Smith J, Mowla S, Prince S. Basal transcription of the human TBX3 gene, a key developmental regulator which is overexpressed in several cancers, requires functional NF-Y and Sp1 sites. *Gene.* 2011; 486:41–46. [PubMed: 21784138]
60. Washkowitz AJ, Gavrilov S, Begum S, et al. Diverse functional networks of Tbx3 in development and disease. *Wiley Interdiscip Rev Syst Biol Med.* 2012; 4:273–283. [PubMed: 22334480]
61. Renard CA, Labalette C, Armengol C, et al. Tbx3 is a downstream target of the Wnt/beta-catenin pathway and a critical mediator of beta-catenin survival functions in liver cancer. *Cancer Res.* 2007; 67:901–910. [PubMed: 17283120]
62. Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science.* 1997; 275:1787–1790. [PubMed: 9065402]
63. Yu J, Ma X, Cheung KF, et al. Epigenetic inactivation of T-box transcription factor 5, a novel tumor suppressor gene, is associated with colon cancer. *Oncogene.* 2010; 29:6464–6474. [PubMed: 20802524]
64. Garner C. Upward bias in odds ratio estimates from genome-wide association studies. *Genet Epidemiol.* 2007; 31:288–295. [PubMed: 17266119]
65. Setiawan VW, Haessler J, Schumacher F, et al. HNF1B and endometrial cancer risk: results from the PAGE study. *PLoS One.* 2012; 7:e30390. [PubMed: 22299039]
66. Lindstrom S, Schumacher FR, Campa D, et al. Replication of five prostate cancer loci identified in an Asian population—results from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol Biomarkers Prev.* 2012; 21:212–216. [PubMed: 22056501]
67. Dumitrescu L, Carty CL, Taylor K, et al. Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. *PLoS Genet.* 2011; 7:e1002138. [PubMed: 21738485]
68. Buyske S, Wu Y, Carty CL, et al. Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS One.* 2012; 7:e35651. [PubMed: 22539988]
69. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet.* 2001; 69:1–14. [PubMed: 11410837]
70. Newton-Cheh CN, Eijgelsheim M, Rice KM, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet.* 2009; 41:399–406. [PubMed: 19305408]

Table 1

Risk Estimates for Newly Identified SNPs Associated With Colorectal Cancer at a P Value Less Than 5×10^{-7}

SNP	Chromosome (gene) ^a	Risk allele	Stage ^b	RAF (range) ^c	OR (95% CI)	P value	P heterogeneity
SNP with $P < 5 \times 10^{-8}$							
rs11903757	2q32.3 (<i>NABP1</i>)	C	GWAS	0.16 (0.11–0.23)	1.15 (1.09–1.22)	1.38E-06	
			Asian	0.05 (0.04–0.08)	1.16 (0.95–1.41)	1.34E-01	
			Adenoma	0.15	1.23 (1.03–1.47)	2.30E-02	
			Overall		1.16 (1.10–1.22)	3.71E-08	.27
SNPs with $P < 5 \times 10^{-7}$ and $P > 5 \times 10^{-8}$							
rs10911251	1q25.3 (<i>LAMC1</i>)	A	GWAS	0.57 (0.49–0.63)	1.10 (1.06–1.14)	1.34E-06	
			Asian	0.54 (0.50–0.55)	1.09 (1.01–1.17)	3.20E-02	
			Adenoma	0.58	1.06 (0.93–1.21)	3.66E-01	
			Overall		1.09 (1.06–1.13)	9.45E-08	.69
rs3217810	12p13.32 (<i>CCND2</i>)	T	GWAS	0.16 (0.18–0.10)	1.19 (1.11–1.28)	3.40E-07	
			Asian	NA	NA	NA	
			Adenoma	0.15	1.31 (1.00–1.71)	5.07E-02	
			Overall		1.20 (1.12–1.28)	5.86E-08	.91
rs3217901	12p13.32 (<i>CCND2</i>)	G	GWAS	0.41 (0.43–0.39)	1.10 (1.06–1.15)	1.71E-06	
			Asian	0.56 (0.54–0.58)	1.08 (0.98–1.18)	1.04E-01	
			Adenoma	0.42	1.08 (0.91–1.27)	3.72E-01	
			Overall		1.10 (1.06–1.14)	3.31E-07	.51
rs59336	12q24.21 (<i>TBX3</i>)	T	GWAS	0.48 (0.51–0.42)	1.10 (1.06–1.14)	7.64E-07	
			Asian	0.60 (0.56–0.62)	1.05 (0.94–1.18)	3.61E-01	
			Adenoma	0.48	1.13 (0.90–1.40)	2.89E-01	
			Overall		1.09 (1.06–1.13)	3.67E-07	.39

CI, confidence interval; RAF, risk allele frequency.

NOTE. Bolded entries signify combined results.

^aChromosome position and build from Genome Browser based on build 37.^bGWAS (n = 12,696 cases and 15,113 controls); Asian (2098 cases and 5749 controls); adenoma (958 cases and 909 controls); overall (15,752 cases and 21,771 controls except for rs3217810, which has 13,654 cases and 16,022 controls).

^cMean and (range) computed from respective studies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript