Gastroenterology 2021;160:1164–1178

BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

Identifying Novel Susceptibility Genes for Colorectal Cancer Risk From a Transcriptome-Wide Association Study of 125,478 Subjects

Xingyi Guo,[1,*] **Weiqiang Lin**,[2,*] Wanqing Wen,[1] Jeroen Huyghe,[3] Stephanie Bien,[3] Qiuyin Cai,[1] Tabitha Harrison,[3] Zhishan Chen,[1] Conghui Qu,[3] Jiandong Bao,[1] Jirong Long,[1] Yuan Yuan,[2] Fangqin Wang,[2] Mengqiu Bai,[2] Goncalo R. Abecasis,[4] Demetrius Albanes,[5] Sonja I. Berndt,[5] Stéphane Bézieau,[6] D. Timothy Bishop,[7] Hermann Brenner,[8,9] Stephan Buch,[10] Andrea Burnett-Hartman,[11] Peter T. Campbell,[12] Sergi Castellví-Bel,[13] Andrew T. Chan,[14,15] Jenny Chang-Claude,[16,17] Stephen J. Chanock,[5] Sang Hee Cho,[18] David V. Conti,[19] Albert de la Chapelle,[20] Edith J. M. Feskens,[21] Steven J. Gallinger,[22] Graham G. Giles,[23,24] Phyllis J. Goodman,[25] Andrea Gsur,[26] Mark Guinter,[12] Marc J. Gunter,[27] Jochen Hampe,[10] Heather Hampel,[28] Richard B. Hayes,[29] Michael Hoffmeister,[8] Ellen Kampman,[21] Hyun Min Kang,[4] Temitope O. Keku,[30] Hyeong Rok Kim,[31] Loic Le Marchand,[32] Soo Chin Lee,[33,34] Christopher I. Li,[3] Li Li,[35] Annika Lindblom,[36,37] Noralane Lindor,[38] Roger L. Milne,[23,24] Victor Moreno,[39,40] Neil Murphy,[12] Polly A. Newcomb,[3,41] Deborah A. Nickerson,[42] Kenneth Offit,[43,44] Rachel Pearlman,[28] Paul D. P. Pharoah,[45] Elizabeth A. Platz,[46] John D. Potter,[3] Gad Rennert,[47,48] Lori C. Sakoda,[3,49] Clemens Schafmayer,[50] Stephanie L. Schmit,[51] Robert E. Schoen,[52] Fredrick R. Schumacher,[53] Martha L. Slattery,[54] Yu-Ru Su,[3] Catherine M. Tangen,[25] Cornelia M. Ulrich,[55] Franzel J. B. van Duijnhoven,[21] Bethany Van Guelpen,[56] Kala Visvanathan,[46] Pavel Vodicka,[57,58] Ludmila Vodickova,[57,58] Veronika Vymetalkova,[57,58] Xiaoliang Wang,[3] Emily White,[3,59] Alicja Wolk,[60] Michael O. Woods,[61] Graham Casey,[62] Li Hsu,[3] Mark A. Jenkins,[24] Stephen B. Gruber,[63] Ulrike Peters,[3,59] and Wei Zheng[1]

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee; [2]The Kidney Disease Center, the First Affiliated Hospital, Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China; [3]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; [4]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan; [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; [6]Service de Génétique Médicale, Centre Hospitalier Universitaire, Nantes, France; [7]Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, United Kingdom; [8]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; [9]Division of Preventive Oncology, German Cancer Research Center and National Center for Tumor Diseases, Heidelberg, Germany; [10]Department of Medicine I, University Hospital Dresden, Technische Universität Dresden, Dresden, Germany; [11]Institute for Health Research, Kaiser Permanente Colorado, Denver, Colorado; [12]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, Georgia; [13]Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, University of Barcelona, Barcelona, Spain; [14]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; [15]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [16]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; [17]University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg, Hamburg, Germany; [18]Department of Hematology-Oncology, Chonnam National University Hospital, Hwasun, South Korea; [19]Department of Preventive Medicine and University of Southern California Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California; [20]Department of Cancer Biology and Genetics and the Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio; [21]Division of Human Nutrition and Health, Wageningen University and Research, Wageningen, The Netherlands; [22]Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; [23]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia; [24]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia; [25]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington; [26]Institute of Cancer Research, Department of Medicine I, Medical University Vienna, Vienna, Austria; [27]Nutrition and Metabolism Section, International Agency for Research on Cancer, World Health Organization, Lyon, France; [28]Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio; [29]Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, New York; [30]Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina; [31]Department of Surgery, Chonnam National University Hwasun Hospital and Medical School, Hwasun, Korea; [32]University of Hawaii Cancer Center,

Honolulu, Hawaii; [33]*National University Cancer Institute, Singapore;* [34]*Cancer Science Institute of Singapore, National University of Singapore, Singapore;* [35]*Department of Family Medicine, University of Virginia, Charlottesville, Virginia;* [36]*Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden;* [37]*Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden;* [38]*Health Sciences Research, Mayo Clinic, Scottsdale, Arizona;* [39]*Oncology Data Analytics Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain;* [40]*CIBER Epidemiología y Salud Pública, Madrid, Spain;* [41]*School of Public Health, University of Washington, Seattle, Washington;* [42]*Department of Genome Sciences, University of Washington, Seattle, Washington;* [43]*Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York;* [44]*Department of Medicine, Weill Cornell Medical College, New York, New York;* [45]*Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom;* [46]*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland;* [47]*Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel;* [48]*Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel;* [49]*Division of Research, Kaiser Permanente Northern California, Oakland, California;* [50]*Department of General Surgery, University Hospital Rostock, Rostock, Germany;* [51]*Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida;* [52]*Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania;* [53]*Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio;* [54]*Department of Internal Medicine, University of Utah, Salt Lake City, Utah;* [55]*Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah;* [56]*Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden;* [57]*Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic;* [58]*Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic;* [59]*Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington;* [60]*Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden;* [61]*Memorial University of Newfoundland, Discipline of Genetics, St John's, Newfoundland and Labrador, Canada;* [62]*Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia; and* [63]*Department of Preventive Medicine and University of Southern California Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California*

See Covering the Cover synopsis on page 976.

BACKGROUND AND AIMS: Susceptibility genes and the underlying mechanisms for the majority of risk loci identified by genome-wide association studies (GWAS) for colorectal cancer (CRC) risk remain largely unknown. We conducted a transcriptome-wide association study (TWAS) to identify putative susceptibility genes. **METHODS:** Gene-expression prediction models were built using transcriptome and genetic data from the 284 normal transverse colon tissues of European descendants from the Genotype-Tissue Expression (GTEx), and model performance was evaluated using data from The Cancer Genome Atlas (n = 355). We applied the gene-expression prediction models and GWAS data to evaluate associations of genetically predicted gene-expression with CRC risk in 58,131 CRC cases and 67,347 controls of European ancestry. Dual-luciferase reporter assays and knockdown experiments in CRC cells and tumor xenografts were conducted. **RESULTS:** We identified 25 genes associated with CRC risk at a Bonferroni-corrected threshold of $P < 9.1 \times 10^{-6}$, including genes in 4 novel loci, *PYGL* (14q22.1), *RPL28* (19q13.42), *CAPN12* (19q13.2), *MYH7B* (20q11.22), and *MAP1L3CA* (20q11.22). In 9 known GWAS-identified loci, we uncovered 9 genes that have not been reported previously, whereas 4 genes remained statistically significant after adjusting for the lead risk variant of the locus. Through colocalization analysis in GWAS loci, we additionally identified 12 putative susceptibility genes that were supported by TWAS analysis at $P < .01$. We showed that risk allele of the lead risk variant rs1741640 affected the promoter activity of *CABLES2*. Knockdown experiments confirmed that *CABLES2* plays a vital role in colorectal carcinogenesis. **CONCLUSIONS:** Our study reveals new putative susceptibility genes and provides new insight into the biological mechanisms underlying CRC development.

Keywords: Colorectal Cancer; Susceptibility Genes; TWAS; *CABLES2*.

Genetic factors play an important role in the etiology of both sporadic and familial colorectal cancer (CRC). Multiple CRC susceptibility genes, including *APC*, *MUTYH, MLH1, PMS2, MSH2, MSH6, PTEN, SMAD4, BMPR1A,* and *MYH,* have been identified as being responsible for family CRC syndromes, such as Lynch syndrome.[1–3] In addition to these known high-penetrance genes, approximately 150 common genetic variants have been found to be associated with CRC risk through genome-wide association studies (GWAS).[4–21] However, together, these variants explain only a small proportion of the familial relative risk of CRC.[18–21] Additional loci for genetic susceptibility for CRC remain to be identified. Furthermore, the target genes at most of the GWAS-identified risk loci remain largely unclear.

Most GWAS-identified risk variants are located in non-coding or intergenic regions without clear evidence of regulatory epigenetic signals. However, it has been shown that many functional variants in strong linkage disequilibrium (LD) with these variants play regulatory roles in gene expression.[22–26] Large genomics consortia, including the Genotype-Tissue Expression (GETx) and The Cancer Genome Atlas (TCGA)

* Authors share co-first authorship.

Abbreviations used in this paper: **CRC, colorectal cancer; eQTL, expression quantitative trait loci; GTEx, Genotype-Tissue Expression; GWAS, genome-wide association study; LD, linkage disequilibrium; PC, principal component; PEER, probabilistic estimation of expression residuals; SCNA, somatic copy number alteration; shRNA, short hairpin RNA; SMR, summary data–based Mendelian randomization; SNP, single-nucleotide polymorphism; TCGA, The Cancer Genome Atlas; TWAS, transcriptome-wide association study.**

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2020.08.062

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Large-scale transcriptome-wide association studies (TWAS) to identify genetic risk loci and susceptibility genes for colorectal cancer (CRC) are lacking.

NEW FINDINGS

Five genes in four novel loci were identified to be associated with CRC risk in individuals of European descent, along with an additional nine genes in known GWAS loci that have not been previously reported.

LIMITATIONS

Further functional assays are requested to establish the underlying regulatory mechanisms for the risk genetic variants and susceptibility genes identified in this study.

IMPACT

Identification of CRC risk loci and susceptibility genes can help identify individuals at high risk. These findings have significant implications for genetic screening.

projects, have generated massive quantities of high-dimensional genomics data, including data from matched genetic and transcriptome profiles from hundreds of normal and tumor tissues of the colon and rectum. These data are valuable for expression quantitative trait loci (eQTL) analyses, which evaluate the association of single-nucleotide polymorphism (SNP) genotypes with gene-expression levels. Using an eQTL approach, previous studies, together with our work, have revealed approximately 30 putative susceptibility genes linked to GWAS-identified SNPs (refer to index SNPs) or lead SNPs (refer to SNPs with the best association signals).[27–31] In addition to the eQTL approach, the advance of chromatin–chromatin interaction technology has produced large amounts of chromatin–chromatin interaction data in various normal and cancer cell lines, including CRC cell lines.[32–35]

Recently, transcriptome-wide association studies (TWAS)[36] have been initiated to systematically investigate the transcriptome for disease risk. First, prediction models are built to predict gene expression with cis-SNPs using a reference transcriptome. Then they are applied to GWAS data to evaluate the association of the predicted gene expression with disease risk. A recently performed TWAS, using approximately 27,000 cases and controls, reported several genes of which the predicted expressions were associated with CRC risk.[37,38] However, the sample size of that study was relatively small, and several reported associations were not statistically significant after Bonferroni correction. Herein, we report results from a large CRC TWAS, conducted among 125,478 subjects of European descent, to comprehensively search for susceptibility genes.

Methods

Genome-Wide Association Study Data From Study Cohorts

The study utilized GWAS summary statistics data from 125,478 individuals of European descent (58,131 CRC cases and 67,347 controls) from 6 cohort studies. The cohort studies include Genetics and Epidemiology of Colorectal Cancer Consortium (n = 11,958 case, n = 14,740 controls), CORECT (n = 22,911 case, n = 14,311 controls), CROC (n = 12,007 cases, n = 12,000 controls), CONE (n = 4439 cases, n = 4115 controls), CORSA (n = 1460 cases, n = 774 controls), and UKBio (n = 5356 cases, n = 21,407 controls). All participants provided written informed consent, and each study was approved by the relevant research Ethics Committee or Institutional Review Boards. Details on sample selection and matching, sample numbers, and demographic characteristics of study participants have been described previously.[18]

Gene-Expression Model Building Using Genotype-Tissue Expression

The recent GTEx, release 8, includes RNA sequencing data of transverse colon samples obtained from 284 subjects and whole-genome sequencing data from these subjects. Gene-expression and whole-genome sequencing data from these samples were processed according to the GTEx protocol, as described in our previous work.[39–41] For gene-expression data, expression levels of each gene have been measured using reads per kilobase per million units from RNA-SeQC. We performed data quality control and normalization processing by filtering lowly expressed genes, log2 transforming, and Robust Multichip Average. We further performed rank-based inverse normal transformation for gene expression across all samples. We performed a probabilistic estimation of expression residuals (PEER) analysis to generate the top PEER factors (top 15 PEERs used in this study) to adjust batch and other potential confounding factors[42] for downstream prediction model building. Only common SNPs (minor allele frequency > 0.05) with "PASS" tags were included. SNPs with a call rate <98%, with a Hardy-Weinberg equilibrium $P < 10^{-6}$ (among subjects of European ancestry) or showing batch effects were excluded. Principal component (PC) analysis was conducted using EIGENSTRAT[43] to generate top PCs from the genotype data.

Genetic and transcriptome data from the 284 transverse colon samples from GTEx were used to build gene-expression prediction models for this study. We trained the gene-expression prediction model by flanking genetic variants (flanking ±1 Mb region) using an elastic-net approach, implemented in the MetaXcan tool (equation 1).

$$Y = G + PC + PEERs + \varepsilon \quad (1)$$

In equation 1, for each gene, Y is the expression level, G is the number of effect alleles (0–2) for each genetic variant, with an adjustment for top PCs, sex, age, potential batch effects, and other potential confounding factors (PEERs), and ε is the random error. We focused on the cis-regulation of a gene predicted by local genetic variants within 2 Mb flanking the gene region. The parameters of the model for each gene were assessed using 10-fold cross-validation, and the correlation (R^2) between predicted and observed gene-expression levels was used to evaluate the prediction performance.

Evaluating Prediction Performance of Gene-Expression Models Using The Cancer Genome Atlas

To further evaluate the prediction performance of these genes in GTEx externally, we downloaded TCGA Colorectal

Adenocarcinoma data, including data from gene expressions (RNA-seq V2, measured by median Z score), DNA methylations (Level 3, Infinium HumanMethylation450), and somatic copy number alteration (SCNAs) (data_linear_CNA.txt) from cBioPortal (https://www.cbioportal.org/). We also downloaded level 3 SNP data (COAD and READ) that were genotyped using the Affymetrix SNP 6.0 array from TCGA's data portal. A total of 355 tumor samples with gene expressions, DNA methylations, SCNA, and SNP data from TCGA were used to evaluate the prediction model. The processing protocol for genotype and gene-expression data was performed for the data from TCGA as described in our previous work.[25,31] To evaluate the prediction performance for each gene predicted by a set of SNPs in GTEx, we first constructed a residual model for gene expression as a dependent variable by adjusting DNA methylation, SCNA, and the first 5 PCs derived from the genotype data. We further used the derived residuals to construct a linear regression model with the genotype data from the set of SNPs. The prediction performance of the model for each gene was assessed by the adjusted R^2 for the linear model, including all genetic predictors determined in GTEx data. The prediction performance of gene-expression models has been summarized in Supplementary Table 1.

Association Analyses Between Predicted Gene Expression and Colorectal Cancer Risk

To identify susceptibility genes for CRC risk, we applied the weight matrix to the summary statistics data on SNPs from CRC GWAS data sets using the MetaXcan tool.[44] The MetaXcan method, which has been described elsewhere,[45,46] was used for the association analyses.

$$Z_g \approx \sum_{l \in \text{Model}_g} w_{lg} \frac{\widehat{\sigma}_l}{\widehat{\sigma}_g} \frac{\widehat{\beta}_l}{\text{se}(\widehat{\beta}_l)} \quad (2)$$

In equation 2, the Z score was used to estimate the association between predicted gene expression CRC risk. Here, w_{lg} is the weight of SNP l for predicting the expression of gene g. $\widehat{\beta}_l$and $\text{se}(\widehat{\beta}_l)$ are the association regression coefficients, and its standard error for SNP l in GWAS, and $\widehat{\sigma}_l$ and $\widehat{\sigma}_g$ are the estimated variances of SNP l and the predicted expression of gene g, respectively. For this study, we estimated the correlations between SNPs included in the prediction models using the phase 3, 1000 Genomes Project data of European populations. The summary statistics were analyzed based on a meta-analysis of GWAS data among 125,478 individuals of European descent from 6 cohort studies. An association result was included in the meta-analysis only if the estimated minor allele count was >50 and the imputation quality metric RsqR2 was >0.3. Notably, additional quality control steps for SNPs and samples have been performed in previous GWAS by the CRC consortia.[18]

For our TWAS analyses, a Bonferroni-corrected $P < 9.1 \times 10^{-6}$ (.05/5491) was used to identify a statistically significant association.

Conditional and Colocalization Analyses of Genome-Wide Association Study Association Signals

To determine whether the identified associations between genetically predicted gene expression and CRC risk were influenced by association signals identified in GWAS, we conducted conditional analyses by adjusting for index SNPs. We followed a previously reported approach[47] to estimate adjusted odds ratios and their standard errors for the association between selected SNPs and CRC risk, given the reported index SNP. Then we re-ran the MetaXcan analyses using the adjusted summary statistics.

We downloaded cis-eQTL results from normal transverse tissues in the GTEx database (version 8). Colocalization analysis of GWAS association signals and expression of TWAS-identified genes was conducted using summary data–based Mendelian randomization (SMR)[48] based on the eQTL results from the GTEx and the GWAS summary statistics data. The genotype data from the European populations of the 1000 Genomes Project Phase 3 was used for the LD estimation. The significant colocalized signals was determined based on the nominal threshold of $P_{\text{SMR}} < .05$. Additional data analyses and functional assays have been described in Supplementary Material and related data has been presented in Supplementary Table 1, 2, and 3.

Results

Gene Expression Predicted by Flanking Genetic Variants

We used transcriptome and genotype data from the GTEx to build gene-expression prediction models for normal transverse colon tissues from European descendants (n = 284) (see Methods). A total of 16,082 models were built for genes predicted by flanking genetic variants (flanking ±1 Mb region) using the elastic net approach.[36,49] Our results showed that the expression levels of 9503 genes could be predicted by local genetic variants at $R^2 > 0.01$ (10% correlation), with a median of 27 variants per gene (see Supplementary Table 4). To evaluate the prediction performances for each gene predicted by a set of SNPs in GTEx, we constructed a residual model for gene expression as a dependent variable by adjusting DNA methylation, SCNA, and the first 5 PCs using data from 355 primary CRC tissues from TCGA. We further used the derived residuals to construct a linear regression model with the genotype data from the same set of SNPs (see Methods). After removing poorly predicted genes, we focused on 5491 genes for a downstream association analysis between predicted gene expression and CRC risk. These genes had either a model prediction $R^2 > 0.01$ in the GTEx and replicated by TCGA, or $R^2 > 0.0625$ (25% correlation) in the GTEx regardless of TCGA; this second set included additional genes that could not be evaluated in TCGA due to a lack of data (see Supplementary Table 4).

Identified Colorectal Cancer Susceptibility Genes in Transcriptome-Wide Association Studies

We evaluated the associations of the predicted genes with CRC risk using the MetaXcan tool,[44] based on the weights of the SNPs in the expression-prediction models from GTEx, and their summary statistics from GWAS data (N = 125,487; see Methods). In total, we identified 25 genes with genetically predicted expressions associated with CRC risk at a Bonferroni-corrected threshold of $P < 9.1 \times 10^{-6}$

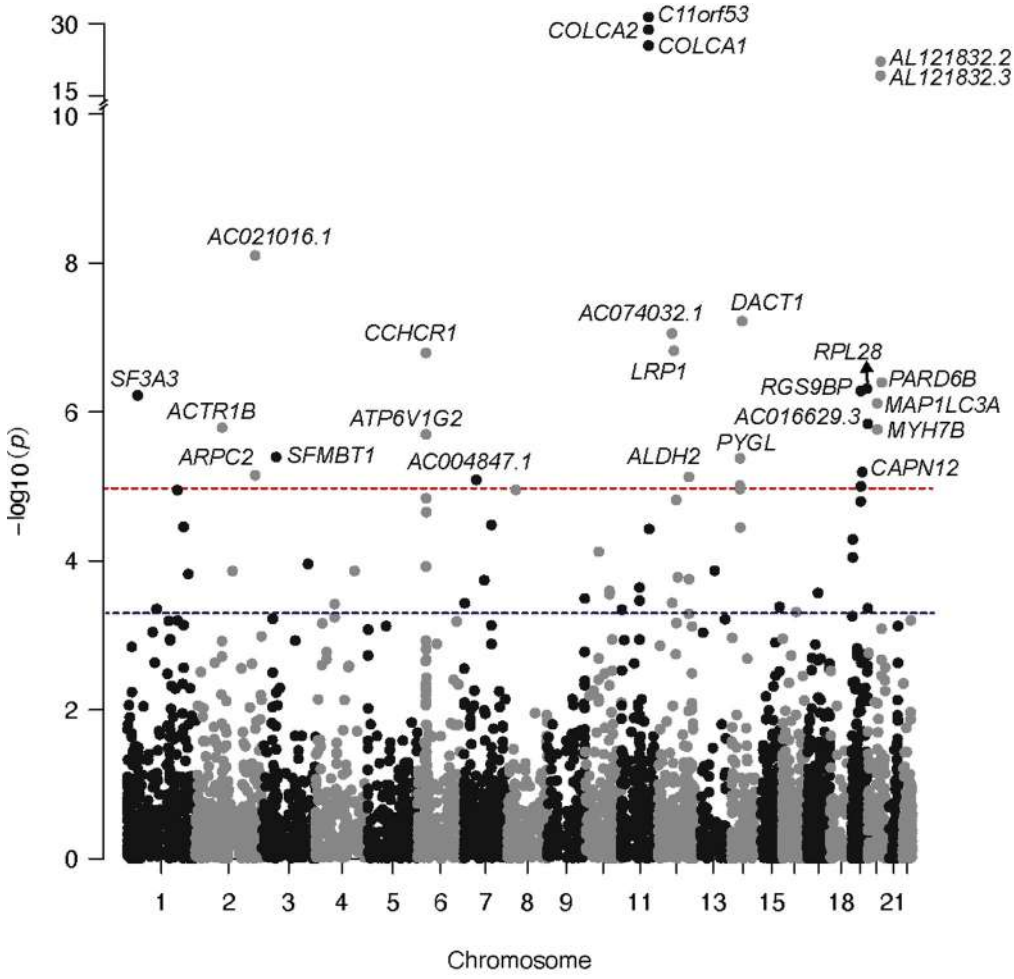


Figure 1. Manhattan plot of the association results from the TWAS of colorectal cancer (N = 125,478). The *blue* and *red lines* represent a false discovery rate (FDR)–corrected significance level of $P < 6.6 \times 10^{-4}$ and a Bonferroni-corrected threshold of $P < 9.1 \times 10^{-6}$, respectively.

(Figure 1 and Supplementary Table 3). Of those, 5 genes, *PYGL* (14q22.1), *RPL28* (19q13.42), *CAPN12* (19q13.2), *MYH7B* (20q11.22), and *MAP1L3CA* (20q11.22) are located in 4 independent loci at least 3 Mb away from any previously reported GWAS-identified locus for CRC risk, suggesting that they may be novel CRC susceptibility loci. Specifically, low predicted expression levels of all of these genes were associated with an increased risk of CRC (Table 1). Using a less stringent threshold at a $P < 6.6 \times 10^{-4}$, a false discovery rate–corrected significance level, we

Table 1. Association of Colorectal Cancer Risk With Predicted Expression for 5 Genes Located at Least 1 Mb Away From Any Genome-Wide Association Study Reported Colorectal Cancer Risk Variants

Locus	Gene	*Z* score	*P* value[a]	R^2 (GTEx)[b]	R^2 (TCGA)[b]	Lead SNP[c]	Distance, *Mb*[c]
14q22.1	*PYGL*	−4.60	4.20×10^{-6}	0.20	0.07	rs35107139	3.0
19q13.2	*CAPN12*	−4.51	6.24×10^{-6}	0.11	0.04	rs28840750	5.7
19q13.4	*RPL28*	−5.03	4.86×10^{-7}	0.23	0.04	rs73068325	3.2
20q11.22[d]	*MYH7B*	−4.78	1.72×10^{-6}	0.13	0.02	rs6031311	9.1
20q11.22[d]	*MAP1LC3A*	−4.94	7.66×10^{-7}	0.13	<0.01	rs6031311	9.5

[a]*P* value derived from association analyses in TWAS; statistically significant based on a Bonferroni-corrected threshold of $P < 9.1 \times 10^{-6}$ from 5491 tests (.05/5491);
[b]Prediction performance (R^2) was derived using data from GTEx, while adjusted R^2 was derived from TCGA data for validation of a gene expression prediction model.
[c]Distance between a gene with the closest lead SNP identified from previous CRC GWAS.
[d]The initial GWAS-identified locus (rs6058093) was excluded in the recent meta-analysis among 125,478 subjects of European descent.[18]

Table 2. Association of Colorectal Cancer Risk With Predicted Expression for 20 Genes Located in the Regions Within 500 kb From Genome-Wide Association Study Reported Colorectal Cancer Risk Variants

Locus	Gene[a]	*Z* score	*P* value[b]	R^2 (GTEx)[c]	R^2 (TCGA)[c]	Lead SNP[d]	Distance, *kb*[d]	*P* value after adjusting for lead SNP[e]
1p34.3	***SF3A3***	4.99	5.99×10^{-7}	0.13	0.01	rs4360494	144	.94
2q11.2	***ACTR1B***	−4.8	1.62×10^{-6}	0.28	0.02	rs11692435	0	.03
2q35	***AC021016.1***	5.77	7.90×10^{-9}	0.52	—[f]	rs35470271	27,542	.28
2q35	*ARPC2*	4.49	7.10×10^{-6}	0.02	0.01	rs3731861	72,177	.90
3p21.1	***SFMBT1***	4.61	4.03×10^{-6}	0.20	0.15	rs9831861	7519	—[g]
6p21.3	*ATP6V1G2*	−4.75	2.00×10^{-6}	0.06	0.08	rs2516420	62,619	7.6×10^{-3}
6p21.33	*CCHCR1*	5.24	1.61×10^{-7}	0.46	<0.01	rs2516420	323,605	8.78×10^{-7}
7p13	***AC004847.1***	−4.46	8.16×10^{-6}	0.15	—[f]	rs12672022	135,915	0.65
11q23.1	***C11orf53***	−11.66	1.94×10^{-31}	0.21	0.06	rs3087967	0	7.4×10^{-3}
11q23.1	***COLCA1***	−10.16	2.94×10^{-24}	0.24	0.01	rs3087967	4676	—[g]
11q23.1	***COLCA2***	−11.25	2.31×10^{-29}	0.36	0.06	rs3087967	12,444	—[g]
12q13.12	*AC074032.1*	−5.35	8.82×10^{-8}	0.19	—[f]	rs12372718	611,689	.86
12q13.3	*LRP1*	5.25	1.50×10^{-7}	0.16	0.01	rs4759277	0	4.47×10^{-4}
12q24.12	*ALDH2*	−4.48	7.47×10^{-6}	0.04	0.02	rs597808	231,333	.40
14q23.1	***DACT1***	−5.42	6.04×10^{-8}	0.06	<0.01	rs17094983	74,322	.09
19q13.11	*RGS9BP*	−5.02	5.22×10^{-7}	0.02	0.02	rs28840750	350,721	1.12×10^{-5}
19q13.43	*AC016629.3*	−4.82	1.45×10^{-6}	0.13	—[f]	rs73068325	7716	.13
20q13.13	*PARD6B*	5.07	4.02×10^{-7}	0.07	0.01	rs6063514	292,799	1.93×10^{-6}
20q13.33	***AL121832.2***	9.68	3.84×10^{-22}	0.23	—[f]	rs1741640	28,945	—[g]
20q13.33	***AL121832.3***	−8.59	8.99×10^{-18}	0.11	—[f]	rs1741640	44,878	—[g]

[a]The genes in bold type refer to those reported from the previous eQTL analysis in CRC and the colocalization analysis in this study.
[b]*P* value derived from association analyses in TWAS; statistically significant based on a Bonferroni-corrected threshold of $P < 9.1 \times 10^{-6}$ from 5491 tests (.05/5491).
[c]Prediction performance (R^2) was derived using data from GTEx, and adjusted R^2 was derived from TCGA data for validation of a gene-expression prediction model.
[d]Distance between a gene with the closest lead SNP identified from previous CRC GWAS.
[e]*P* value derived from association analyses in TWAS after adjusting for the association of the closest lead SNP.
[f]"—" indicates a model in GTEx that could not be validated in TCGA due to a lack of data.
[g]The lead SNP that has been included in the model or is in strong LD ($R^2 > 0.95$) with an SNP in the model that showed the best association signal with CRC risk.

identified an additional 48 genes with genetically predicted expressions associated with CRC risk (Figure 1 and Supplementary Table 5).

To determine whether the genes identified were implicated in the associations of established GWAS-identified risk variants, we investigated 20 (of the total of 25) genes located in 16 established independent CRC risk loci (Table 2). We performed conditional analyses for the associations between CRC risk and these genes, adjusting for the associations with the closest lead SNP for each locus (see Methods). We showed that an association with genes*CCHCR1 (*rs2516420, in a distance of 323kb), *LRP1 (*rs4759277, 0 kb*), RGS9BP (*rs28840750, 350kb*),* and *PARD6B* (rs6063514, 293kb) remained statistically significant at a $P < 2.5 \times 10^{-3}$ (with multiple comparisons corrections of .05/20) after adjusting for the closest risk SNP (Table 2). Of the remaining 16 genes, we observed that the expression prediction models for 5 of these, *SFMBT1* (rs9831861), *COLCA1* (rs3087967), *COLCA2* (rs3087967), *AL121832.2* (rs1741640), and *AL121832.3* (rs1741640), included the lead SNPs or SNPs in strong LD ($R^2 > 0.95$) with them. This supports the hypothesis that the observed association for these genes may be driven by these lead SNPs (Table 2). Further colocalization analysis between the expression of all TWAS-identified genes and previous GWAS association signals using the SMR approach showed that the effects of 8 lead SNPs on cancer risk might be mediated by expression of TWAS-identified genes, including *SF3A3,*

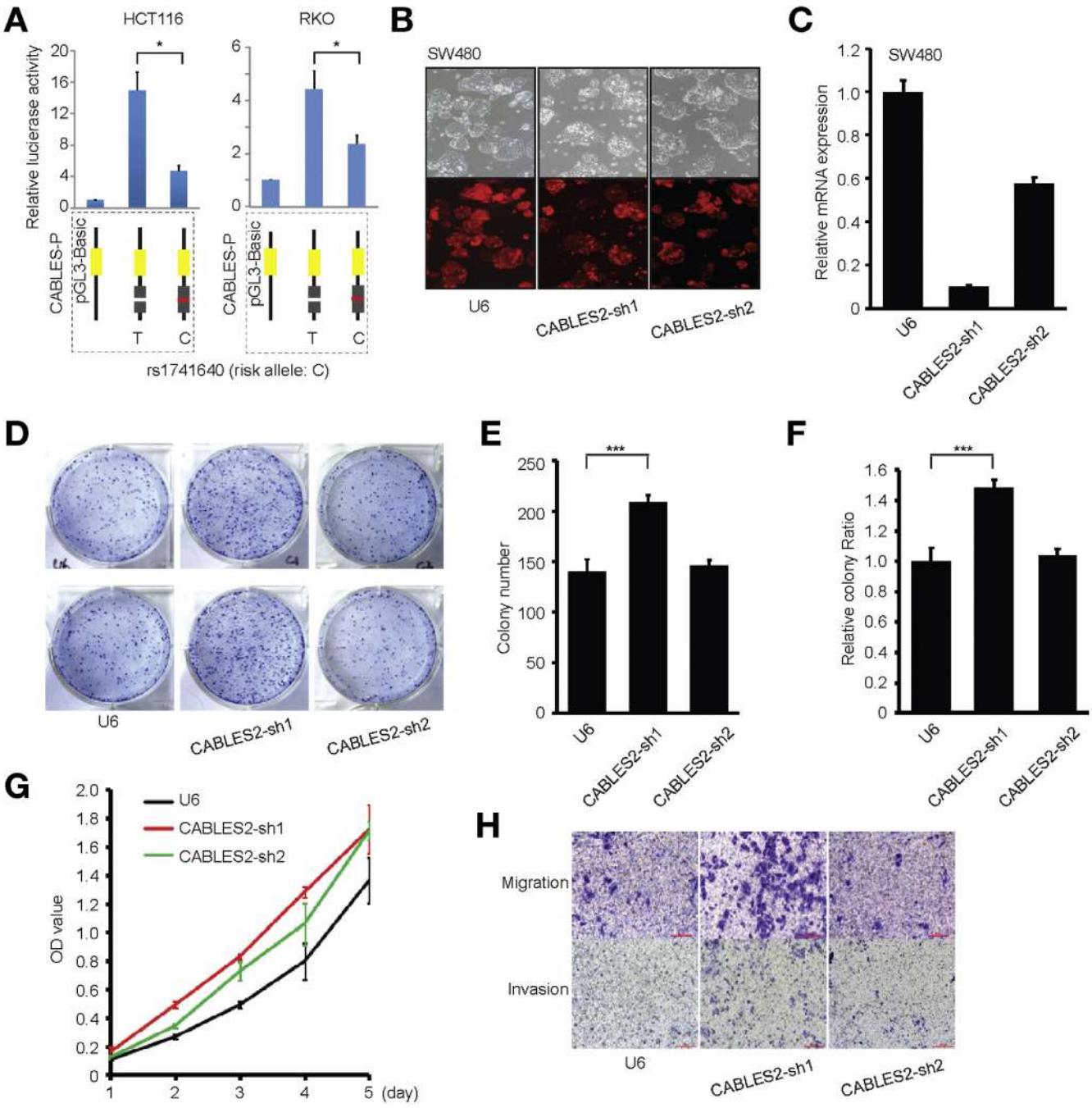


Figure 2. Dual-luciferase reporter assay and CRC cell proliferation, colony formation, migration, and invasion for the *CABLES2* gene. (*A*) *Boxplots* showing that alternative alleles (CC) of functional SNP rs1741640 can affect promoter activity of *CABLES2* compared to reference alleles (TT), based on the results using luciferase reporter assays in both the HCT116 and RKO cell lines. The *yellow* and *black boxes* represent the sequence of *CABLES2* promoter and the enhancer elements containing the candidate functional SNP individually. The *error bars* represent the SD of promoter activities. A paired *t* test was performed to derive *P* value. (*B*) The virus infection efficiency and (*C*) the knockdown efficiency were verified by red fluorescent protein fluorescence and quantitative polymerase chain reaction, respectively, in *CABLES2* shRNA-treated (sh1 and sh2) and vehicle control cells. After SW480 cells were transfected with *CABLES2* shRNA, the cell colonies were (*D*) imaged, (*E*) quantified, and (*F*) given a ratio using a colony formation assay for *the CABLES2 shRNA* and vehicle control cells. (*G*) Cell viabilities were measured using CCK8 assays at different time points (days 1, 2, 3, 4, and 5) and (*H*) migration and invasion were measured using Transwell assays for CABLES2 shRNA and vehicle cells. *P* values were determined by a *t* test from the comparison of knockdown and control cells. $^{*}P < .05$; $^{**}P < .01$; $^{***}P < .001$.

ACTR1B, AC021016.1, SFMTB1, AC004847.1, C11orf53, COLCA1, COLCA2, DACT1, AL121832.2, and *AL121832.3* (Supplementary Table 5). The remaining 9 genes have not been reported previously.

In addition, we systemically evaluated eQTL results in the established GWAS loci[18,23] using data from the GTEx and identified 42 eQTL genes for 19 lead variants at a Bonferroni-corrected threshold of $P < 2.4 \times 10^{-5}$ (Supplementary Table 6). Further colocalization analysis showed that the CRC susceptibility for 18 leading variants might be mediated by cis-effects on gene regulation, whereas 38 genes were involved (Supplementary Table 6). In addition to the genes identified in the above TWAS, we observed that an additional 12 genes, including *ASAH2B, ATF1, CABLES2, HCG20, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB5, PRRT1, TRPC6, ZNF132,* and *ZNF584* were supported by TWAS analysis at a less stringent threshold of $P <$ .01 (Supplementary Table 6).

Putative Susceptibility Genes Supported by Functional Genomic Data Analysis

To search for additional evidence for the 25 putative susceptibility genes identified from TWAS, we explored regulatory mechanisms for putative functional variants correlated with the SNP that has the strongest association with CRC risk in the prediction model. We performed functional annotations of the variants in strong LD ($R^2 >$ 0.8) with these SNPs (Supplementary Table 7; see Supplementary Material). We analyzed 349 putative functional variants (in strong LD with these SNPs), which showed evidence of the epigenetic signals from a data analysis of Roadmap (Supplementary Table 7; see Supplementary Material). Specifically, a total of 118 variants were mapped to promoter regions, whereas the remaining 231 variants were mapped to enhancer regions. To search for direct evidence that variants regulate the putative susceptibility genes identified from our TWAS analysis, we first examined whether the potential functional variants were positioned in proximal promoter regions, as such variants would most likely play a regulatory role in relation to the closest genes. The 13 genes, *C11orf53, COLCA1, COLCA2, AC074032, ALDH2, ARPC2, CAPN12, LRP1, PARD6B, PYGL, RPL28, SF3A3,* and *SFMBT1*, were likely regulated by nearby putative functional SNPs with promoter activities (Supplementary Table 8). We further examined whether the genes could be regulated by putative functional variants located in enhancer regions via long-distance promoter–enhancer interactions by analyzing chromatin-chromatin interaction data. An additional 6 genes, *AC004847.1, AC074032.1, MAP1LC3A, ACTR1B, AL121832.2,* and *DACT1*, showed evidence of potential regulation via distal promoter-enhancer interaction. Taken together, a total of 19 of the original 25 identified genes (76%) showed evidence of regulation by putative functional variants via proximal promoter or distal enhancer–promoter interactions. All of them were further supported with the evidence of their potential functional variants located in regions of histone modifications, DNase I hypersensitive or transcription factor binding sites in normal colorectal epithelium and CRC cell lines (Supplementary Table 8).

Functional Assays for Putative Tumor Suppressor CABLES2

Previous studies have indicated that the *CABLES1* gene might be involved in cell growth and differentiation,[50,51] and it appears to function as a tumor suppressor by inhibiting CRC formation and growth.[51–55] In line with these previous findings, our results showed that a lower predicted expression of *CABLES2* might mediate the effect of the lead SNP rs1741640 on an increased risk of CRC. A functional annotation of rs1741640 indicates that this SNP is located in a region with promoter activity. To confirm whether the rs1741640 can regulate *CABLES2* expression, we conducted luciferase reporter assays for rs1741640 and another SNP, rs477859, in strong LD with rs1741640 in both HCT116 and RKO cell lines (see Supplementary Material). Our results showed that the fragment containing the risk allele (C, rs1741640) significantly decreased the promoter activity of *CABLES2* compared with the reference allele T in both cell lines (Figure 2*A*). The observation was in line with the results from both colocalization and TWAS analyses (Supplementary Table 6). For rs477859, we did not observe that the fragment containing the alternative allele significantly affects the promoter activity compared with the reference allele (data not shown).

We further performed in vitro functional assays in CRC cell lines to investigate the cellular function of *CABLES2*. Quantitative polymerase chain reaction experiments were conducted to compare its relative expression level in 3 CRC cell lines (SW480, RKO, and HCT116). *CABLES2* showed higher expression levels in both the SW480 and HCT116 cell lines compared with the RKO cell line (Supplementary Figure 1). We designed knockdown experiments in the SW480 and HCT116 cell lines using short hairpin RNA (shRNA). We employed packaged lentivirus from lentiviral expression vectors containing the red fluorescent protein reporter gene with open reading frames and a fragment from *CABLES2* shRNA-1, shRNA-2, or a control vehicle shRNA. Using the red fluorescent protein fluorescence reporter system, we observed high transfection efficiencies for the control vehicles, shRNA-1 and shRNA-2, in the SW480 cells (Figure 2*B*). In SW480 cells, the shRNA-1 virus infection was able to reduce the endogenous messenger RNA level by approximately 90%, whereas shRNA-2 decreased the messenger RNA level by 40% in SW480 cells (Figure 2*C*). After performing the *CABLES2* shRNA-1 treatment, colony formation was significantly enhanced over that of the control vehicle; *CABLES2* shRNA-2 also led to a mild increase in colony formation (Figure 2*D–F*; $P <$ .001). Cell viabilities after the *CABLES2* shRNA-1 or shRNA-2 treatment were significantly increased over those of the control vehicle in the CCK8 assay ($P <$ .01 for both; Figure 2*G*). Further, we observed that shRNA-1 significantly increased cell migration and invasion compared with the control vehicle; shRNA-2 also led to an increase of cell invasion in

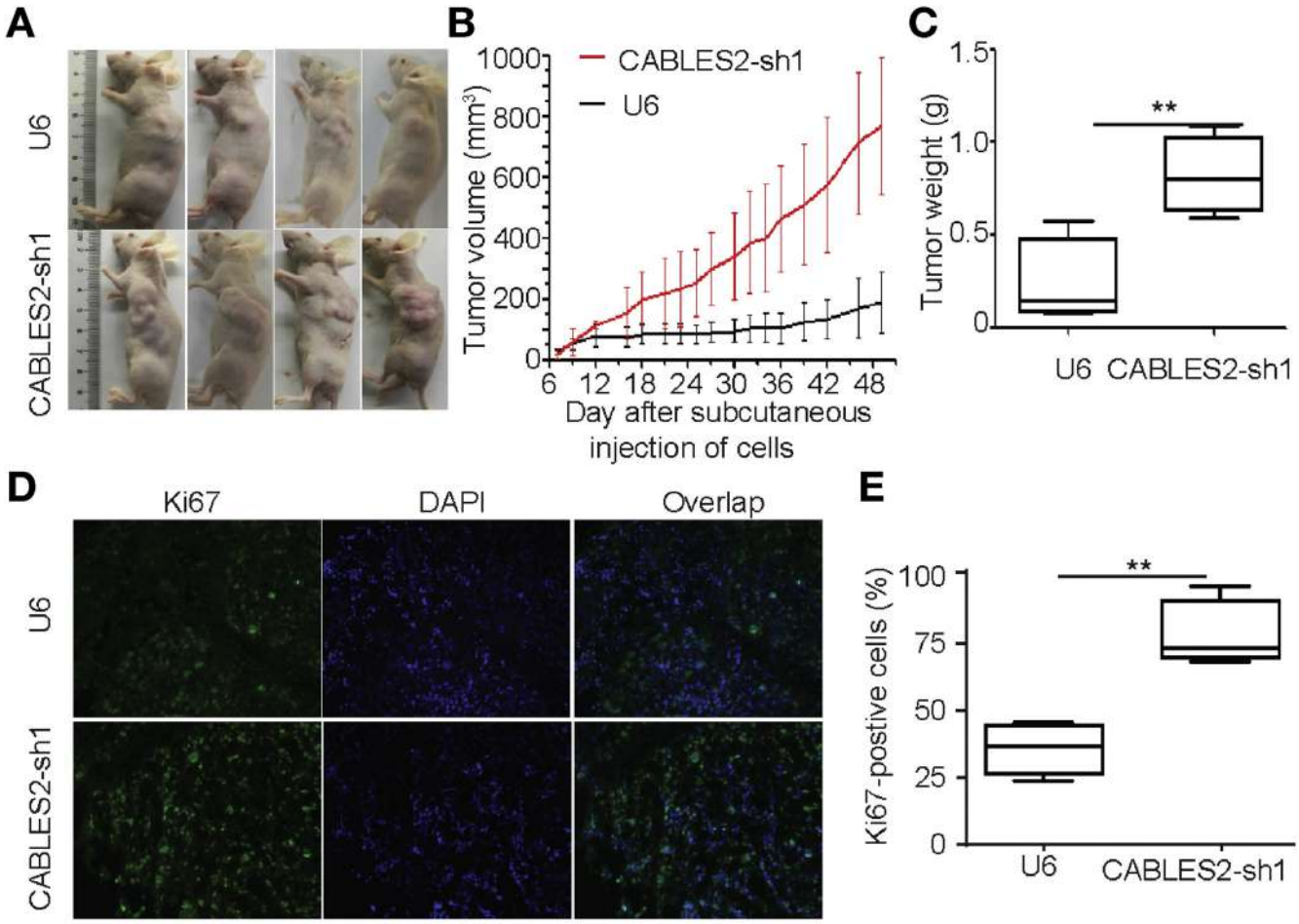


Figure 3. *CABLES2* inhibiting growth of colonic carcinoma in nude mice. (*A*) *Photograph* of tumor formation in nude mice after injection of cells at day 48. (*B*) Tumor volume was recorded by a measurement of tumor length, height, and width every 2-3 days during tumor proliferation, and calculated as length × width × height / 2. (*C*) Quantification of tumor weight at day 49 after the anesthetic execution of mice. (*D, E*) Ki67 staining of tumor tissue and quantification of Ki67-positive cells in knockdown and control cells.

the Transwell assay in the SW480 and HCT116 cell lines (Figure 2*H* and Supplementary Figure 2).

To explore potential genes and pathways regulated by *CABLES2*, we conducted RNA sequencing in the *CABLES2* shRNA-1 and control vehicle SW480 cells (see Supplementary Material). We identified 169 differentially expressed genes, which included 85 up-regulated and 84 down-regulated genes, with *CABLES2* shRNA-1, compared with control vehicle cells (Supplementary Figure 3*A*; see Supplementary Material). An enrichment analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG). Pathways revealed that these differentially expressed genes were significantly enriched in thePI3K/AKT cancer pathway ($P = 7.2 \times 10^{-3}$). A total of 10 genes were involved in that pathway, including 9 up-regulated genes: *TNC, ANTXR2, IL7R, CHN2, SGK1, TLR4, FN1, MET*, and *ARHGAP29.* The presence of only 1 down-regulated gene, *IL2RG,* suggests that the silenced *CABLES2 can* promote the PI3K-AKT signaling pathway (Supplementary Figure 3*B*). We further validated these expression changes for the top 7 genes using quantitative polymerase chain reaction in the cells (Supplementary Figure 3*C*).

To confirm that *CABLES2* is capable of inhibiting the growth of colonic carcinoma, we developed tumor xenografts by the subcutaneous injection of SW480 *CABLES2* knockdown and control shRNA cells in nude mice (see Supplementary Material). We showed that the injected *CABLES2* knockdown cells markedly increases tumor volume and weight compared with control cells (Figure 3*A–C*). To analyze cell proliferation in tumor tissue, Ki67 staining was conducted in tumor tissue from nude mice that had been injected with knockdown and control cells. We observed that *CABLES2* knockdown cells increase Ki67 foci compared with control cells (Figure 3*D*); 77% of Ki67-positive cells were observed in knockdown cells compared with 35% in control cells (Figure 3*E*).

Discussion

Utilizing data from a large-scale GWAS collaboration of GTEx and TCGA, we conducted a TWAS analysis to search for susceptibility genes for CRC risk. We disclosed 25 putative CRC predisposition genes, including 6 novel genes located in the regions far away from any previously reported susceptibility loci for CRC risk, and 9 novel genes as

targets for established CRC GWAS loci. We have also identified an additional 48 genes with genetically predicted expressions associated with CRC risk using a relax threshold at false discovery rate <0.05. These genes can be further verified by additional future large-scale genetic studies. These findings greatly increase the number of potential susceptibility genes identified for CRC risk. However, we also identified 38 putative target genes, including *CABLES2*, for the established GWAS loci through the colocalization analysis, whereas a majority of genes were supported by TWAS analysis. Using both in vitro and in vivo functional assays, we further showed that the putative tumor suppressor *CABLES2* plays a vital role in CRC tumorigenesis. Our findings provide novel insight into the genetic and biological basis for CRC development.

Our identification of these 25 putative susceptibility genes is supported by several lines of additional evidence. For example, for 76% of them, we observed evidence of cis-regulation by putative functional risk SNPs via proximal promoter or distal enhancer–promoter interactions. These data suggest a possible link of a gene with a potential regulatory SNP. Functional assays, such as luciferase reporter assays in CRC cell lines, are needed to establish the underlying regulatory mechanisms, as demonstrated by *CABLES2* (rs1741640) in our study. Furthermore, 11 of these genes have been implicated as putative target genes for lead SNPs based on previous eQTL and our current colocalization analyses[27–31] (Table 2). A particular lead SNP might be a surrogate for multiple variants for cancer risk in the locus, while some target genes for potential causal variants, which are in weak LD with index SNPs, might not be detected by an eQTL analysis. In addition, we conducted a differential gene expression analysis for the genes identified in TWAS among normal colon mucosa, adenoma, and adenocarcinoma using publicly available gene expression data, including 135 normal colon mucosas, 363 colon adenomas, and 2760 colon adenocarcinomas (Supplementary Tables 5 and 6). Of the 25 genes identified in TWAS, we observed 7 genes, including *DACT1*, *CCHCR1*, *PARD6B*, *MYH7B*, *SFMBT1*, *PYGL*, and *ALDH2*, which showed significant differential expression between carcinoma and adenoma tissues at $P <$ .05. For 3 genes, *MYH7B*, *PYGL*, and *ALDH2,* we observed that low predicted expression levels were associated with an increased risk of CRC and showed lower expression in carcinoma than in adenoma. For the remaining 4 genes where high predicted expression levels were associated with an increased risk of CRC, they showed higher expression in carcinoma than in adenoma, except for the gene *DCAT1* (Supplementary Table 5). The results provide additional evidence to support a possible role of these genes in tumorigenesis. However, we did not observe the same pattern when we compared adenoma with mucosa. This might be due to their potential oncogenic role in the initiation of carcinogenesis in adenoma stage. It should be noted that more than half of our identified genes, including, *RPL28*,[56] *CCHCR1*,[57,58] *PARD6B*,[59] *DACT1*,[60,61] *LRP1*,[62–64] *SF3A3*,[65] *MAP1LC3A*,[66] *ACTR1B*,[67] *ATP6V1G2*,[68] *PYGL*,[69,70] *ARPC2*,[71,72] *ALDH2*,[73,74] *KCNQ1*,[75] and *SFMBT1*,[76] have been implicated in cancer-driver events or cancer-related functions in in vitro or in vivo functional experimental studies in CRC or other cancer types.

In our gene-expression prediction model building from GTEx, although we used a relatively low cutoff of prediction performance at $R^2 > 0.01$ (10% correlation) for downstream TWAS analysis, all of the reported genes can still be well predicted with a minimal prediction performance at $R^2 > 0.04$. In addition, we used tumor tissues from TCGA to evaluate the prediction performance of gene-expression built in GTEx. Although we performed additional analyses to adjust for possible confounders of gene expression in tumor tissues of TCGA, large independent normal tissues samples are desirable for the evaluation of the prediction performance of gene-expression. For our prediction model, the reference panels were built using only the data from European populations. Both GWAS and transcriptome data from GTEx are currently limited for other racial groups, preventing us from evaluating these susceptibility genes in non-European populations. Additional efforts are still needed to generate transcriptome data for TWAS in those populations.

Although many of the genes we identified have been implicated previously in CRC susceptibility genes, we provided additional evidence to support their role. Further, we have identified 15 genes not reported previously for CRC susceptibility, including 5 genes located in the regions not yet identified by GWAS. We performed functional genomic experiments for 1 particular gene, *CABLES2*, to demonstrate that some genes identified in TWAS are functionally important in colorectal carcinogenesis. Additional in vitro and in vivo experiments will be needed to firmly establish a causal association between the other reported genes and CRC risk.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://doi.org/10.1053/j.gastro.2020.08.062

References

1. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 2013;45:136–144.
2. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–337.
3. Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol 2011;6:479–507.
4. Zeng C, Matsuda K, Jia WH, et al. Identification of susceptibility loci and genes for colorectal cancer risk. Gastroenterology 2016;150:1633–1645.
5. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet 2015;47:373–380.

6. Al-Tassan NA, Whiffin N, Hosking FJ, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep 2015;5:10442.
7. Wang H, Burnett T, Kono S, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. Nat Commun 2014; 5:4613.
8. Schmit SL, Schumacher FR, Edlund CK, et al. A novel colorectal cancer risk locus at 4q32.2 identified from an international genome-wide association study. Carcinogenesis 2014;35:2512–2519.
9. Zhang B, Jia WH, Matsuda K, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet 2014;46:533–542.
10. Figueiredo JC, Hsu L, Hutter CM, et al. Genome-wide diet-gene interaction analyses for risk of colorectal cancer. PLoS Genet 2014;10:e1004228.
11. Whiffin N, Hosking FJ, Farrington SM, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Hum Mol Genet 2014;23:4729–4737.
12. Zhang B, Jia WH, Matsuo K, et al. Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. Int J Cancer 2014;135:948–955.
13. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. Gastroenterology 2013; 144:799–807.e24.
14. 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.
15. 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.
16. 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.
17. Cui R, Okada Y, Jang SG, et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut 2011;60:799–805.
18. 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.
19. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. J Natl Cancer Inst 2019;111:146–157.
20. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun 2019; 10:2154.
21. Lu Y, Kweon SS, Tanikawa C, et al. Large-scale genome-wide association study of east asians identifies loci associated with risk for colorectal cancer. Gastroenterology 2019;156:1455–1466.
22. Gusev A, Lee SH, Trynka G, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet 2014;95:535–552.
23. Guo X, Cai Q, Bao P, et al. Long-term soy consumption and tumor tissue MicroRNA and gene expression in triple-negative breast cancer. Cancer 2016;122:2544–2551.
24. Zeng C, Guo X, Long J, et al. Identification of independent association signals and putative functional variants for breast cancer risk through fine-scale mapping of the 12p11 locus. Breast Cancer Res 2016;18:64.
25. Guo X, Lin W, Bao J, et al. A comprehensive cis-eQTL analysis revealed target genes in breast cancer susceptibility loci identified in genome-wide association studies. Am J Hum Genet 2018;102:890–903.
26. Chen Z, Wen W, Beeghly-Fadiel A, et al. Identifying putative susceptibility genes and evaluating their associations with somatic mutations in human cancers. Am J Hum Genet 2019;105:477–492.
27. Biancolella M, Fortini BK, Tring S, et al. Identification and characterization of functional risk variants for colorectal cancer mapping to chromosome 11q23.1. Hum Mol Genet 2014;23:2198–2209.
28. Closa A, Cordero D, Sanz-Pamplona R, et al. Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. Carcinogenesis 2014; 35:2039–2046.
29. Peltekova VD, Lemire M, Qazi AM, et al. Identification of genes expressed by immune cells of the colon that are regulated by colorectal cancer-associated variants. Int J Cancer 2014;134:2330–2341.
30. Hofer P, Hagmann M, Brezina S, et al. Bayesian and frequentist analysis of an Austrian genome-wide association study of colorectal cancer and advanced adenomas. Oncotarget 2017;8:98623–98634.
31. Chen Z, Wen W, Beeghly-Fadiel A, et al. Identifying putative susceptibility genes and evaluating their associations with somatic mutations in human cancers. Am J Hum Genet 2019;105:477–492.
32. Rao SS, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–1680.
33. Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 2013;503:290–294.
34. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 2015;47:598–606.
35. Jager R, Migliorini G, Henrion M, et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nat Commun 2015;6:6178.
36. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091–1098.
37. Bien SA, Su YR, Conti DV, et al. Genetic variant predictors of gene expression provide new insight into risk of colorectal cancer. Hum Genet 2019;138:307–326.
38. Su YR, Di C, Bien S, et al. A mixed-effects model for powerful association tests in integrative functional genomics. Am J Hum Genet 2018;102:904–919.

39. Wu L, Shu X, Bao J, et al. Analysis of over 140,000 european descendants identifies genetically predicted blood protein biomarkers associated with prostate cancer risk. Cancer Res 2019;79:4592–4598.
40. Guo X, Long J, Chen Z, et al. Discovery of rare coding variants in OGDHL and BRCA2 in relation to breast cancer risk in Chinese women. Int J Cancer 2020; 146:2175–2181.
41. Guo X, Shi J, Cai Q, et al. Use of deep whole-genome sequencing data to identify structure risk variants in breast cancer susceptibility genes. Hum Mol Genet 2018;27:853–859.
42. Stegle O, Parts L, Piipari M, et al. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 2012;7:500–507.
43. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38:904–909.
44. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 2018;9:1825.
45. Lu Y, Beeghly-Fadiel A, Wu L, et al. A transcriptome-wide association study among 97,898 women to identify candidate susceptibility genes for epithelial ovarian cancer risk. Cancer Res 2018;78:5419–5430.
46. Wu L, Shi W, Long J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet 2018; 50:968–978.
47. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 2012;44:369–375, S1–S3.
48. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48:481–487.
49. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245–252.
50. Zukerberg LR, DeBernardo RL, Kirley SD, et al. Loss of cables, a cyclin-dependent kinase regulatory protein, is associated with the development of endometrial hyperplasia and endometrial cancer. Cancer Res 2004; 64:202–208.
51. Wu CL, Kirley SD, Xiao H, et al. Cables enhances cdk2 tyrosine 15 phosphorylation by Wee1, inhibits cell growth, and is lost in many human colon and squamous cancers. Cancer Res 2001;61:7325–7332.
52. Bonifant CL, Waldman T. 'Cables' suspends cancer in mice. Cancer Biol Ther 2005;4:864–865.
53. Kirley SD, D'Apuzzo M, Lauwers GY, et al. The Cables gene on chromosome 18Q regulates colon cancer progression in vivo. Cancer Biol Ther 2005;4:861–863.
54. Park DY, Sakamoto H, Kirley SD, et al. The Cables gene on chromosome 18q is silenced by promoter hypermethylation and allelic loss in human colorectal cancer. Am J Pathol 2007;171:1509–1519.
55. Arnason T, Pino MS, Yilmaz O, et al. Cables1 is a tumor suppressor gene that regulates intestinal tumor progression in Apc(Min) mice. Cancer Biol Ther 2013; 14:672–678.
56. Labriet A, Levesque E, Cecchin E, et al. Germline variability and tumor expression level of ribosomal protein gene RPL28 are associated with survival of metastatic colorectal cancer patients. Sci Rep 2019;9:13008.
57. Suomela S, Elomaa O, Skoog T, et al. CCHCR1 is upregulated in skin cancer and associated with EGFR expression. PLoS One 2009;4:e6030.
58. Chang J, Zhong R, Tian J, et al. Exome-wide analyses identify low-frequency variant in CYP26B1 and additional coding variants associated with esophageal squamous cell carcinoma. Nat Genet 2018;50:338–343.
59. Cunliffe HE, Jiang Y, Fornace KM, et al. PAR6B is required for tight junction formation and activated PKCzeta localization in breast cancer. Am J Cancer Res 2012;2:478–491.
60. Shi X, Huo J, Gao X, et al. A newly identified lncRNA H1FX-AS1 targets DACT1 to inhibit cervical cancer via sponging miR-324-3p. Cancer Cell Int 2020;20:358.
61. Zhu K, Jiang B, Yang Y, et al. DACT1 overexpression inhibits proliferation, enhances apoptosis, and increases daunorubicin chemosensitivity in KG-1alpha cells. Tumour Biol 2017;39:1010428317711089.
62. Song H, Li Y, Lee J, et al. Low-density lipoprotein receptor-related protein 1 promotes cancer cell migration and invasion by inducing the expression of matrix metalloproteinases 2 and 9. Cancer Res 2009;69:879–886.
63. Boyé K, Pujol N, D Alves I, et al. The role of CXCR3/LRP1 cross-talk in the invasion of primary brain tumors. Nat Commun 2017;8:1571.
64. Tian Y, Wang C, Chen S, et al. Extracellular Hsp90alpha and clusterin synergistically promote breast cancer epithelial-to-mesenchymal transition and metastasis via LRP1. J Cell Sci 2019;132.
65. Siebring-van Olst E, Blijlevens M, de Menezes RX, et al. A genome-wide siRNA screen for regulators of tumor suppressor p53 activity in human non-small cell lung cancer cells identifies components of the RNA splicing machinery as targets for anticancer treatment. Mol Oncol 2017;11:534–551.
66. Hamurcu Z, Delibasi N, Gecene S, et al. Targeting LC3 and Beclin-1 autophagy genes suppresses proliferation, survival, migration and invasion by inhibition of cyclin-D1 and uPAR/Integrin beta1/Src signaling in triple negative breast cancer cells. J Cancer Res Clin Oncol 2018; 144:415–430.
67. Di Simone N, Hall HA, Welt C, et al. Activin regulates betaA-subunit and activin receptor messenger ribonucleic acid and cellular proliferation in activin-responsive testicular tumor cells. Endocrinology 1998;139:1147–1155.
68. Pacifici R, Civitelli R, Rifas L, et al. Does interleukin-1 affect intracellular calcium in osteoblast-like cells (UMR-106)? J Bone Miner Res 1988;3:107–111.
69. Favaro E, Bensaad K, Chong MG, et al. Glucose utilization via glycogen phosphorylase sustains proliferation

and prevents premature senescence in cancer cells. Cell Metab 2012;16:751–764.
70. Terashima M, Fujita Y, Togashi Y, et al. KIAA1199 interacts with glycogen phosphorylase kinase beta-subunit (PHKB) to promote glycogen breakdown and cancer cell survival. Oncotarget 2014;5:7040–7050.
71. Cheng Z, Wei W, Wu Z, et al. ARPC2 promotes breast cancer proliferation and metastasis. Oncol Rep 2019; 41:3189–3200.
72. Yoon YJ, Han YM, Choi J, et al. Benproperine, an ARPC2 inhibitor, suppresses cancer cell migration and tumor metastasis. Biochem Pharmacol 2019;163:46–59.
73. Li K, Guo W, Li Z, et al. ALDH2 repression promotes lung tumor progression via accumulated acetaldehyde and DNA damage. Neoplasia 2019;21:602–614.
74. Seo W, Gao Y, He Y, et al. ALDH2 deficiency promotes alcohol-associated liver cancer by activating oncogenic pathways via oxidized DNA-enriched extracellular vesicles. J Hepatol 2019;71:1000–1011.
75. Rapetti-Mauss R, Bustos V, Thomas W, et al. Bidirectional KCNQ1:beta-catenin interaction drives colorectal cancer cell differentiation. Proc Natl Acad Sci U S A 2017;114:4159–4164.
76. Hulur I, Gamazon ER, Skol AD, et al. Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. BMC Genomics 2015;16:138.

Received November 14, 2019. Accepted August 28, 2020.

Correspondence
Address correspondence to: Xingyi Guo, PhD, Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 2525 West End Avenue, Suite 330, Nashville, Tennessee 37203. e-mail: xingyi.guo@vumc.org.

Acknowledgments
The authors thank the CLUE cohort GTEx, TCGA, ENCODE, and Roadmap for providing valuable data resources for the research. The authors thank Drs Lang Wu, Yingchang Lu, and Chenjie Zeng for valuable discussion and Marshal Younger for assistance with editing and manuscript preparation.

CRediT Authorship Contributions
Xingyi Guo, PhD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Methodology: Lead; Resources: Lead; Supervision: Lead; Writing – original draft: Lead)
Weiqiang Lin, PhD (Investigation: Equal; Writing – original draft: Equal),
Wanqing Wen, PhD (Formal analysis: Supporting; Writing – review & editing: Supporting),
Jeroen Huyghe, PhD (Formal analysis: Equal; Writing – review & editing: Supporting) Stephanie Bien, PhD (Formal analysis: Supporting; Writing – review & editing: Supporting)
Qiuyin Cai, PhD (Data curation: Equal). Tabitha Harrison, PhD (Data curation: Supporting; Writing – review & editing: Supporting)
Zhishan Chen, PhD (Formal analysis: Supporting; Writing – review & editing: Supporting)
Conghui Qu, PhD (Data curation: Supporting; Writing – review & editing: Supporting) Jiandong Bao, PhD (Formal analysis: Equal; Writing – review & editing: Supporting) Jirong Long, PhD (Writing – review & editing: Supporting),
Yuan Yuan, PhD (Data curation: Supporting; Writing – review & editing: Supporting) Fanqing Wang, PhD (Writing – review & editing: Supporting),
Mengqiu Bai, PhD (Writing – review & editing: Supporting),
Goncalo R Abecasis, PhD (Writing – review & editing: Supporting),
Demetrius Albanes, PhD (Writing – review & editing: Supporting),
Sonja I Berndt, PhD (Writing – review & editing: Supporting),
Stephane Bezieau, PhD (Writing – review & editing: Supporting),
D Timothy Bishop, PhD (Writing – review & editing: Supporting),
Hermann Brenner, PhD (Writing – review & editing: Supporting)
Stephan Buch, PhD (Writing – review & editing: Supporting)
Andrea Burnett-Hartman, PhD (Writing – review & editing: Supporting)
Peter T Campbell, PhD (Writing – review & editing: Supporting)
Sergi Castellvi-Bel, PhD (Writing – review & editing: Supporting)
Andrew T Chan, PhD (Writing – review & editing: Supporting)
Jenny Chang-Claude, PhD (Writing – review & editing: Supporting)
Stephen J Chanock, PhD (Writing – review & editing: Supporting)
Sang Hee Cho, PhD (Writing – review & editing: Supporting)
David V Conti, PhD (Writing – review & editing: Supporting)
Albert de la Chapelle, PhD (Writing – review & editing: Supporting)
Edith JM Feskens, PhD (Writing – review & editing: Supporting)
Steven J Gallinger, PhD (Writing – review & editing: Supporting)
Graham G Giles, PhD (Writing – review & editing: Supporting)
Phyllis J Goodman, PhD (Writing – review & editing: Supporting)
Andrea Gsur, PhD (Writing – review & editing: Supporting)
Mark Guinter, PhD (Writing – review & editing: Supporting)
Marc J Gunter, PhD (Writing – review & editing: Supporting)
Jochen Hampe, PhD (Writing – review & editing: Supporting)
Heather Hampel, PhD (Writing – review & editing: Supporting)
Richard B Hayes, PhD (Writing – review & editing: Supporting)
Michael Hoffmeister, PhD (Writing – review & editing: Supporting)
Ellen Kampman, PhD (Writing – review & editing: Supporting)
Hyun Min Kang, PhD (Writing – review & editing: Supporting)
Temitope O Keku, PhD (Writing – review & editing: Supporting)
Hyeong Rok Kim, PhD (Writing – review & editing: Supporting)
Loic Le Marchand, PhD (Writing – review & editing: Supporting)
Soo Chin Lee, PhD (Writing – review & editing: Supporting)
Christopher I Li, PhD (Writing – review & editing: Supporting)
Li Li, PhD (Writing – review & editing: Supporting)
Annika Lindblom, PhD (Writing – review & editing: Supporting)
Noralane Lindor, PhD (Writing – review & editing: Supporting)
Roger L Milne, PhD (Writing – review & editing: Equal)
Victor Moreno, PhD (Writing – review & editing: Supporting)
Neil Murphy, PhD (Writing – review & editing: Supporting)
Polly A Newcomb, PhD (Writing – review & editing: Supporting)
Deborah A Nickerson, PhD (Writing – review & editing: Supporting)
Kenneth Offit, PhD (Writing – review & editing: Supporting)
Rachel Pearlman, PhD (Writing – review & editing: Supporting)
Paul DP Pharoah, PhD (Writing – review & editing: Supporting)
Elizabeth A Platz, PhD (Writing – review & editing: Supporting)
John D Potter, PhD (Writing – review & editing: Equal)
Gad Rennert, PhD (Writing – review & editing: Supporting)
Lori C Sakoda, PhD (Writing – review & editing: Supporting)
Clemens Schafmayer, PhD (Writing – review & editing: Supporting)
Stephanie L Schmit, PhD (Writing – review & editing: Supporting)
Robert E Schoen, PhD (Writing – review & editing: Supporting)
Fredrick R Schumacher, PhD (Writing – review & editing: Supporting)
Martha L Slattery, PhD (Writing – review & editing: Supporting)
Yu-Ru Su, PhD (Writing – review & editing: Supporting)
Catherine M Tangen, PhD (Writing – review & editing: Supporting)
Cornelia M Ulrich, PhD (Writing – review & editing: Supporting)
Franzel JB van Duijnhoven, PhD (Writing – review & editing: Supporting)
Bethany Van Guelpen, PhD (Writing – review & editing: Supporting)
Kala Visvanathan, PhD (Writing – review & editing: Supporting)
Pavel Vodicka, PhD (Writing – review & editing: Supporting)
Ludmila Vodickova, PhD (Writing – review & editing: Supporting)
Veronika Vymetalkova, PhD (Writing– review & editing: Supporting)
Xiaolong Wang, PhD (Writing – review & editing: Supporting)
Emily White, PhD (Writing – review & editing: Supporting)
Alicja Wolk, PhD (Writing – review & editing: Supporting)
Michael O Woods, PhD (Writing – review & editing: Supporting)
Graham Casey, PhD (Writing – review & editing: Supporting)
Li Hsu, PhD (Writing – review & editing: Supporting)
Mark A Jenkins, PhD (Writing – review & editing: Supporting)
Stephen B Gruber, PhD (Writing – review & editing: Supporting)
Ulrike Peters, PhD (Writing – review & editing: Equal)
Wei Zheng, PhD, MD (Conceptualization: Lead; Methodology: Lead; Writing – review & editing: Equal)

Conflicts of interest
The authors disclose no conflicts.

Funding
The data analyses were conducted using the Advanced Computing Center for Research and Education at Vanderbilt University. This research is supported primarily by the grant from US National Institutes of Health (NIH) grant R37 CA227130 to Xingyi Guo and R01 CA188214 to Wei Zheng. Genetics and Epidemiology of Colorectal Cancer Consortium: National Cancer Institute (NCI), NIH, US Department of Health and Human Services (U01 CA164930, U01 CA137088, R01 CA059045, R21 CA191312, R01 CA201407). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) (X01-HG008596 and X01-HG007585). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN268201200008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. PCS is supported by the NIH (R01

CA160356, R01 CA193677, R01 CA204279 and R01 CA143237). ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes and supported 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 ATBC Study is supported by the Intramural Research Program of the US NCI, NIH, and by US Public Health Service contract HHSN261201500005C from the NCI, Department of Health and Human Services. CLUE II: This research was funded by the American Institute for Cancer Research and the NCI (P30 CA006973 to W.G. Nelson). COLO2&3: NIH (R01 CA60987). ColoCare: This work was supported by the NIH (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueireido/Colditz), 2P30CA015704-40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative. The Colon Cancer Family Registry (CFR) Illumina GWAS was supported by funding from the NCI, NIH (grant numbers U01 CA122839, R01 CA143247). The Colon CFR/CORECT Affymetrix Axiom GWAS and OncoArray GWAS were supported by funding from NCI, NIH (grant number U19 CA148107 to S Gruber). The Colon CFR participant recruitment and collection of data and biospecimens used in this study were supported by the NCI, NIH (grant number UM1 CA167551) and through cooperative agreements with the following Colon CFR centers: Australasian Colorectal Cancer Family Registry (NCI/NIH grant numbers U01 CA074778 and U01/U24 CA097735), USC Consortium Colorectal Cancer Family Registry (NCI/NIH grant numbers U01/U24 CA074799), Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (NCI/NIH grant number U01/U24 CA074800), Ontario Familial Colorectal Cancer Registry (NCI/NIH grant number U01/U24 CA074783), Seattle Colorectal Cancer Family Registry (NCI/NIH grant number U01/U24 CA074794), and University of Hawaii Colorectal Cancer Family Registry (NCI/NIH grant number U01/U24 CA074806), Additional support for case ascertainment was provided by the Surveillance, Epidemiology and End Results Program of the NCI to Fred Hutchinson Cancer Research Center (Control Nos. N01-CN-67009 and N01-PC-35142, and Contract No. HHSN2612013000121), the Hawai'i Department of Health (Control Nos. N01-PC-67001 and N01-PC-35137, and Contract No. HHSN26120100037C, and the California Department of Public Health (contracts HHSN261201000035C awarded to the University of Southern California, and the following state cancer registries: AZ, CO, MN, NC, NH, and by the Victoria Cancer Registry and Ontario Cancer Registry. COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Funds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The NQplus study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project PREVention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project, which received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and Nutrition (cardiovascular health theme), a public–private partnership on pre-competitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life. Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the NCI, NIH, US Department of Health and Human Services (grant numbers U19 CA148107, R01 CA81488, P30 CA014089, R01 CA197350, P01 CA196569, R01 CA201407, and R01 CA143237) and National Institutes of Environmental Health Sciences, NIH (grant number T32 ES013678). CORSA: This study was funded by FFG BRIDGE (grant 829675 to Andrea Gsur), the "Herzfelder'sche Familienstiftung" (grant to Andrea Gsur) and was supported by COST Action BM1206. Cancer Prevention Study-II (CPS-II): The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. This study was conducted with Institutional Review Board approval. Colorectal Cancer Genetics & Genomics, Spanish study, was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants of the Catalan Government (grant 2017SGR723), and Junta de Castilla y León (grant LE22A10-2). Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. Czech Republic CCS: This work was supported by the Czech Science Foundation (grants 17-16857S and 18-09709S), the Grant Agency of the Ministry of Health of the Czech Republic (grants 15-27580A and 17-30920A), the Charles University Research Centre program (UNCE/MED/006), the Charles University Research Fund (Progres Q39 and Q28), and MEYS CR, financed from EFRR (CZ.02.1.01/0.0/0.0/16_019/0000787). DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, BR 1704/6-6, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases, Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B). DALS: NIH (R01 CA48998 to M. L. Slattery). EDRN: This work is funded and supported by the NCI, EDRN grant (U01 CA 84968-06). EPIC: The coordination of EPIC is financially supported by the School of Public Health, Imperial College London and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (France); German Cancer Aid, German Cancer Research Center, Federal Ministry of Education and Research, Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports, Netherlands Cancer Registry, LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund, Statistics Netherlands (The Netherlands); ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); Health Research Fund, PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom).

Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization. EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, P111/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and COST Action BM1206. CIBERehd is funded by the Instituto de Salud Carlos III. ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid. Harvard cohorts (HPFS, NHS, PHS): HPFS is supported by the NIH (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003), NHS by the NIH (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003) and PHS by the NIH (R01 CA042182). Hawaii Adenoma Study: NCI grants R01 CA72520. Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer: grants from Chonnam National University Hwasun Hospital (HCRI15011-1), and the NIH (R01 CA188214). Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726. The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167). Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian National Health and Medical Research Council grants 209057, 396414, and 1074383, and by infrastructure provided by Cancer Council Victoria. Cases and their vital statuses were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. MEC: NIH (R37 CA54281, P01 CA033619, R01 CA063464, and U01 CA164973). MECC: This work was supported by the NIH, US Department of Health and Human Services (R01 CA81488 to SBG, R01 CA197350 to SBG, U19 CA148107 to SBG, N01 CN043302 assigned to SBG, 5P01 CA196569 to WG, P30 CA014089 to SBG, R01 CA144040 to SDM, and P50 CA150964 to SDM), as well as funding from the Ravitz Foundation, the Irving Weinstein Foundation, the Anton B. Burg Foundation, the Jane and Kris Popovich Chair in Cancer Research, and a generous gift from Daniel and Maryann Fong. Memorial Sloan Kettering Cancer Center: The work at Memorial Sloan Kettering Cancer Center in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. It is also supported by the NCI-designated Comprehensive Cancer Center (grant number P30 CA008748). NCCCS I & II: We acknowledge funding support for this project from the NIH, R01 CA66635 and P30 DK034987. NFCCR:

This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the NIH, US Department of Health and Human Serivces (U01 CA74783); and NCI of Canada grants (18223 and 18226). Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute. Northern Sweden Health and Disease Study (NSHDS) investigators thank the Biobank Research Unit at Umeå University, the Västerbotten Intervention Programme, the Northern Sweden MONICA study and Region Västerbotten for providing data and samples and acknowledge the contribution from Biobank Sweden, supported by the Swedish Research Council (VR 2017-00650). This research was also supported by funding to BVG from the Swedish Cancer Society (CAN 2017/581); the Swedish Research Council (VR 2017-01737); Region Västerbotten (VLL-841671, VLL-833291); the Lion's Cancer Research Foundation (several grants), the Faculty of Medicine and Insamlingsstiftelsen at Umeå University; and the Margareta Dannborg Memorial Fund. Ontario Registry for Studies of Familial Colorectal Cancer (OFCCR): NIH, through funding allocated to the OFCCR (U01 CA074783); see Colon Cancer Family Registry section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation. *Ohio State University Wexner Medical Center: Ohio Colorectal* Cancer Prevention Initiative funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA16058 and CA67941). PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, NCI, NIH, Department of Health and Human Services. Funding was provided by NIH, Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. PMH: NIH (R01 CA076366 to P.A. Newcomb). SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge. The Selenium and Vitamin E Cancer Prevention Trial (SELECT) was supported by the NCI of the NIH under award numbers UM1CA182883 and U10CA37429. SMS: This work was supported by the NCI (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the NIH (grant KL2 TR000421 to A.N.B.-H.) The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project). Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by grants from the Swedish Cancer Foundation, the Swedish Research Council for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER) and Karolinska Institute's Distinguished Professor Award to Alicja Wolk. VITAL: NIH (K05 CA154337). Women's Health Initiative (WHI): The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

Supplementary Material

Chromatin–Chromatin Interaction Data Analysis

We analyzed chromatin–chromatin interactions between the regions for potential functional SNPs in strong LD ($R^2 >$ 0.8, European populations) with the SNPs that had the strongest association with CRC risk in the prediction models and promoter regions of the identified putative susceptibility genes using various functional genomics data. For each SNP, we investigated whether it was mapped to functional regions (ie, promoter or enhancer) using chromatin states annotation, from the database HaploReg, version 4 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php).[e1] We also downloaded a total of 384 chromatin immunoprecipitation sequencing and DNase I hypersensitive sites sequencing peak files for histone modifications, DNase hypersensitivity, and transcription factors in normal colorectal epithelium and CRC cell lines from the Cistrome database (http://cistrome.org/ Supplementary Table 2). We only analyzed the peak data in high quality control with a fraction of the reads in peaks and uniquely mapped ration based on the evaluation from Cistrome. For each SNP, we examined whether it was mapped to a region of histone modifications, DNase I hypersensitive, or transcription factors binding sites using in-house PERL script.

Experimentally derived chromatin interactions generated by Hi-C, chromatin interaction analysis with paired-end-tag sequencing, and integrated method for predicting enhancer targets were collected from 4DGenome (https://4dgenome.research.chop.edu/).[e2] Additional chromatin interactions data from CRC relevant cells were also analyzed.[e3,e4] We examined the position of a potential functional SNP and ±2 kb flanking regions of the gene transcription start site of the identified putative susceptibility genes, to determine their chromatin–chromatin interactions.

Plasmid Construction and Dual-Luciferase Reporter Assay

Construction for a *CABLES2* promoter-driven luciferase reporter plasmid was generated by polymerase chain reaction (PCR) using custom-designed primers (Supplementary Table 3). The PCR product was double-digested by the enzymes KpnI and XhoI, and then inserted into the pGL3-Basic vector, named pGL3-CABLES2. In order to investigate whether the candidate functional SNP rs1741640 can affect *CABLES2* promoter activity, the 1.5-kb fragment containing the SNP (flanking 750 bp) was synthesized and introduced into the pGL3-CABLES2 by SalI to construct pGL3-CABLES2-rs1741640 by homologous recombination reaction. We also conducted the same experiment for another SNP rs477859 in strong LD with rs1741640 (LD $R^2 = 0.75$ in European populations). All constructed target fragments were confirmed by sequencing. The minor allele of the individual SNP of the 2 constructions was introduced into the plasmid using site-directed mutagenesis. The primers are listed in Supplementary Table 3. We sequenced all constructed fragments to confirm variant incorporation. The Dual-Luciferase Reporter Assay Kit (Promega E1910) was conducted following the manufacturer's instructions (as detailed in our previous work[25]). Briefly, luciferase reporter plasmid and pGL-TK transfection control plasmid were cotransfected into the HCT116 and RKO cells individually. The cells were collected and lysed, 24 hours post-transfection, in 1× Passive Lysis Buffer at room temperature for 25 minutes. Twenty microliters of the cell lysate and 100 μL of LAR II were mixed into each well of a 96-well plate. Firefly luciferase activity and Renilla luciferase activity were further measured. We normalized firefly luciferase activity to Renilla luciferase activity to correct the potential effects of transfection efficiency. All reporter assays were performed in triplicate and repeated in 3 independent experiments.

Lentiviral Vector Construction, Viral Production, and Viral Infection

An shRNA fragment targeting the gene was generated using a pair of annealed primers and subcloned into the pLent-U6-RFP-Puro vector (Vigene Biosciences) via BamHI and MluI sites. The resulting plasmid was confirmed by sequencing. Viral production and infection have been described in detail previously.[e5] In most cases, cells were subjected to various analyses 72 hours post viral infection.

Real-Time Quantitative Polymerase Chain Reaction for Gene Expression Assays

Total messenger RNA (mRNA) was extracted using an RNA-plus reagent (TaKaRa, catalog no. 9109), and the complementary DNA was synthesized using a PrimeScript RT reagent kit with gDNA eraser (TaKaRa, catalog no. RR047A), according to the manufacturer's instructions. Real-time quantitative PCR was performed in a MJ Chromo 4 (Bio-Rad DNA engine) by using a reaction mixture of SsoFastTM EvaGreen SuperMix (Bio-Rad, catalog no. 172-5202). The primers used in this study are listed in Supplementary Table 4. All of the PCR amplifications were performed in triplicate and repeated in 3 independent experiments. The relative quantities of selected mRNAs were normalized to that of glyceraldehyde 3-phosphate dehydrogenase.

Cell Proliferation Assay

We measured the proliferation of shRNA knockdown cells and control cells in 96-well plates. We used a CCK8 assay to determine proliferation according to the manufacturer's instructions. Briefly, cells were plated into 96-well plates in triplicate and cultured on different days, then the medium was changed to a solution of 100 μL fresh medium and 10 μL CCK-8 for a 2-hour incubation. Finally, the OD values were measured by Infinite M1000 pro (Tecan) at 450 nm.

Colony Formation Assay

The indicated cells were trypsinized into a single-cell suspension and seeded into 6-well plates at densities of

200, 400, 800, or 1000 cells per well, according to cell types. After 10–15 days, these colonies were washed and fixed with 4% paraformaldehyde for 30 minutes at room temperature, and further stained with Coomassie Blue (Beyotime Biotechnology, catalog no. P0017B). Clones containing at least 50 cells were considered 1 formation. All experiments were performed 3 times.

Migration and Invasion Assay

Cell migration and invasion ability were examined using an insert Transwell containing an 8-μm pores membrane (Corning, catalog no. 3422). For the cell invasion assay, the Transwell membranes were precoated with Matrigel (Corning, catalog no. 356234), and cells of different groups were trypsinized. Two hundred microliters of the cell suspension in a serum-free medium containing 5×10^4 cells were plated into each well of the inserts. Six hundred microliters of the media containing 20% fetal bovine serum were then added to the lower chamber. Cells were incubated at 37°C for 48–72 hours, depending on the cell lines. The cells that invaded the bottom of the membrane were fixed with 4% formaldehyde for 30 minutes at room temperature and stained with crystal violet. The results were imaged using a microscope. For the cell-invasion assay, we repeated all of the procedures we employed for the cell migration assay, except that the upper chamber of the cell migration assay was performed without the Matrigel precoating.

RNA Sequencing and Data Analysis

We conducted RNA-sequencing (read depth approximately 26M) in the *CABLES2* shRNA-1 and control vehicle cells with 2 technical replicates. In brief, the protocol for RNA sequencing included mRNA isolation and fragmentation, complementary DNA synthesis, Adapter ligation, and PCR. The libraries of mRNA were first amplified within the flow cell on the cBot instrument for cluster generation and the clustered flow cell was loaded onto the BGISEQ-500 Sequencer for single end sequencing, with length 50 bp. We used bioinformatics analyses to process the raw RNA-seq reads, following our previously developed pipeline.[e6] All sequencing reads (FASTQ files) were mapped to the human genome (hg38) using the Bowtie2 tool.[e7] We used the DESeq tool[e8] to perform a differential gene expression analysis comparing the *CABLES2* shRNA-1 and control vehicle cells. Significantly differentially expressed genes were identified using a cutoff at an adjusted $P < .01$ and a fold-change >2.

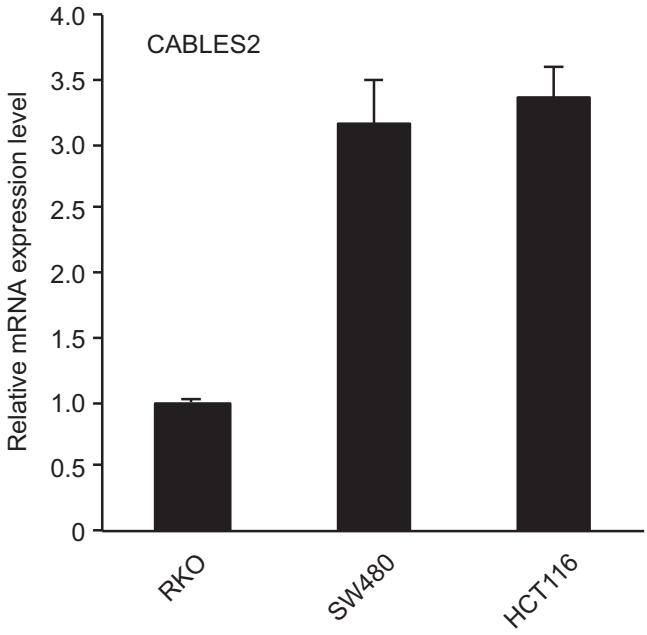
In Vivo Tumor Formation and Ki67 Staining

BALB/c athymic nude mice, approximately 6 weeks of age, obtained from the Shanghai SLAC Laboratory Animal Co, Ltd, were used for tumor xenografts. These mice were maintained under specific pathogen-free conditions at 25°C under a 12-hour light/12-hour dark cycle, with free access to food and water. We divided nude mice into 2 groups for subcutaneous injection of SW480 *CABLES2* knockdown and control shRNA cells. Each mouse was anesthetized, and 1×10^7 cells in 100 μL phosphate-buffered saline were injected subcutaneously. Tumor volume was measured every 2–3 days, and was calculated as length $\times$ width $\times$ height / 2. The mice were anesthetized and killed 49 days post-injection, and tumors from the 2 groups were excised and weighed. Tissues were fixed in 10% formalin, and tissue sections were stained with Ki67 antibody and 4′,6-diamidino-2-phenylindole. Slides were examined under a fluorescence microscope (DM4000; Leica). All protocols that involved animals were approved by the Research Animal Care Committee of Zhejiang University, China.

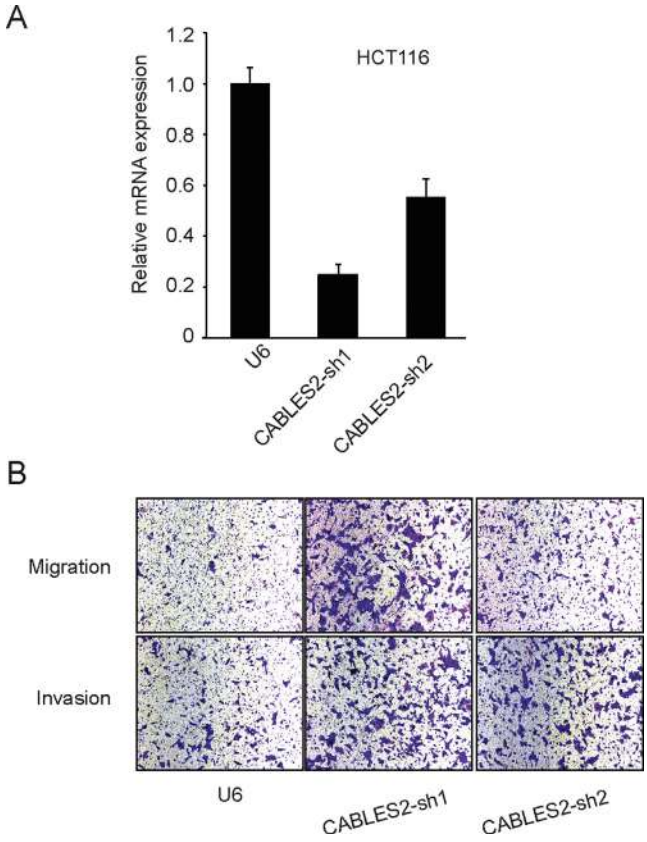
Supplementary Table 1. Supplementary Table 2. Supplementary Table 5. Supplementary Table 6. Supplementary Table 7. Supplementary Table 8.

Supplementary References

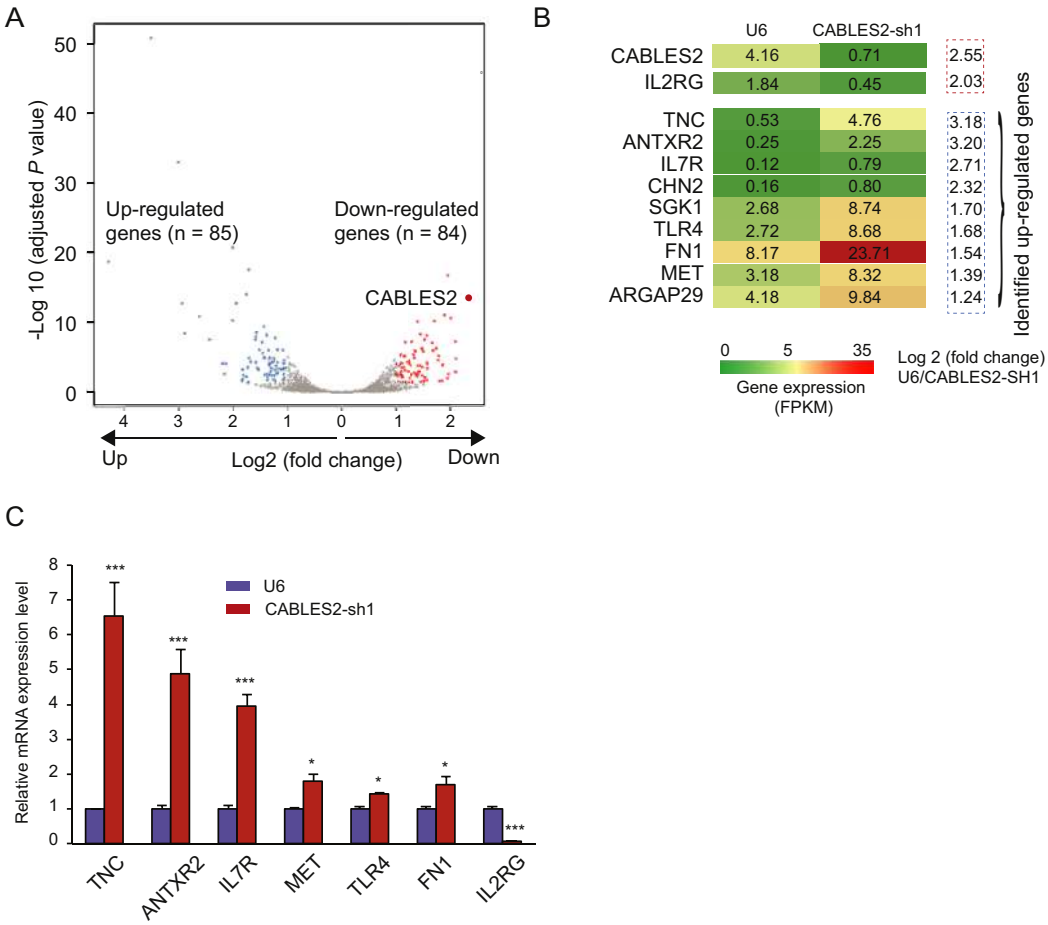
1. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016;44:D877–D881.
2. Teng L, He B, Wang J, et al. 4DGenome: a comprehensive database of chromatin interactions. Bioinformatics 2015;31:2560–2564.
3. Jung I, Schmitt A, Diao Y, et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat Genet 2019;51:1442–1449.
4. Orlando G, Law PJ, Cornish AJ, et al. Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. Nat Genet 2018;50:1375–1380.
5. Lin Y, Yang Y, Li W, et al. Reciprocal regulation of Akt and Oct4 promotes the self-renewal and survival of embryonal carcinoma cells. Mol Cell 2012;48:627–640.
6. Guo X, Lin M, Rockowitz S, et al. Characterization of human pseudogene-derived non-coding RNAs for functional potential. PLoS One 2014;9:e93972.
7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.
8. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.



Supplementary Figure 1. The relative messenger RNA expression levels of the *CABLES2* gene, measured by quantitative polymerase chain reaction in RKO, SW480, and HCT116.



Supplementary Figure 2. The effects on CRC migration and invasion for the *CABLES2* gene using in vitro functional assays in cell HCT116. (*A*) The knockdown efficiency was verified by qPCR in *CABLES2* shRNA treated (sh1 and sh2), and vehicle control cells. (*B*) Migration and invasion were measured using Transwell assays for *CABLES2* shRNA and vehicle cells.



Supplementary Figure 3. The *CABLES2* gene can promote PI3K/AKT signaling pathways and fatty acid metabolism. (*A*) Differentially expressed genes were identified by RNA sequencing in the *CABLES2* shRNA-1 and vehicle control cells. *Blue* and *red* refer to significantly up-regulated and down-regulated genes, respectively, by the silenced *CABLES2*. The top 2 genes with significant expression changes, *CABLES2* and *ALDH3A1*, are *highlighted*. (*B*) Differentially expressed genes in the PI3K/AKT pathways and fatty acid metabolism are presented. Gene-expression levels measured by FPKM are indicated by *heat-map*. (*C*) A total of 7 differentially expressed genes in the PI3K/AKT pathway are validated by quantitative polymerase chain reaction between *CABLES2* shRNA-1 and vehicle cells ("U6"). *P* values were determined by *t* test from the comparison of knockdown and control cells. **P* < .05; ***P* < .01; ****P* < .001.

Supplementary Table 2. Primer Pairs Used for the Construction of the Vectors

Primer[a]	Sequence
*CABLES2*_P_KpnI	GGGGTACCATCACACGAAACAAAAGCCGTG
*CABLES2*_P_XhoI	CCGCTCGAGCCTCAGACTGCGCCCGCC
rs1741640_H_F	AATCGATAAGGATCCGTCGACGGTACCGAGCTCTTACGCGTG
rs1741640_H_R	CTCTCAAGGGCATCGGTCGACTATCGATAGAGAAATGTTCTGGCAC
rs477859_H_F	AATCGATAAGGATCCGTCGACTCAGGGCTGGCAGCCAGG
rs477859_H_R	CTCTCAAGGGCATCGGTCGACTCGGAAGCAGCAGCCAGG
rs1741640_M_F	GGGCCAGCACCCTGACCGCACTGCTC
rs1741640_M _R	GAGCAGTGCGGTCAGGGTGCTGGCCC
rs477859_M_F	GCCGTGCAGATGTCGCAGTACTGGCCCTG
rs477859_M_R	CAGGGCCAGTACTGCGACATCTGCACGGC

[a]"P" represents the primers for the construction of the *CABLES2* promoter vector. "H" represents the primers for sub-cloning the fragment containing the candidate functional SNP by homologues recombination."M" represents the associated primers for mutagenesis.

Supplementary Table 3. shRNA primers for CABLES and Real-Time Quantitative Polymerase Chain Reaction Primers Used for the Genes *CABLES2*, *GAPDH, TNC, ANTXR2, IL7R, MET, TLR4, FN1*, and *IL2RG*

Primer	Sequence
CABLES2_H1_F1	GATCCGAGAACCAAGTGTTACCTCATCTCGAGATGAGGTAACACTTGGTTCTCTTTTTA
CABLES2_H1_R1	CGCGTAAAAAGAGAACCAAGTGTTACCTCATCTCGAGATGAGGTAACACTTGGTTCTCG
CABLES2_H2_F2	GATCCCCCTTCAGTTATGTCCCATTTCTCGAGAAATGGGACATAACTGAAGGGTTTTTA
CABLES2_H2_R2	CGCGTAAAAACCCTTCAGTTATGTCCCATTTCTCGAGAAATGGGACATAACTGAAGGGG
GAPDH_F	GGGGAGCCAAAAGGGTCATCATCT
GAPDH_R	GAGGGGCCATCCACAGTCTTCT
TNC_F1	TCTGATGGGGGTGGATGGAT
TNC_R1	TCTCCCACGCTGAACTTGTC
ANTXR2_F	GGGGATCGGTTTGATGTGG
ANTXR2_R	GTGGGTTTGGGTCGAGGTG
IL7R_F1	CTCCAACCGGCAGCAATGTAT
IL7R_R1	AGATGACCAACAGAGCGACAG
MET_F1	TCCTCTGGGAGCTGATGACA
MET_R1	CTGGGCAGTATTCGGGTTGT
TLR4_F1	AGTTGATCTACCAAGCCTTGAGT
TLR4_R1	GCTGGTTGTCCCAAAATCACTTT
FN1_F1	TGGATGTTCCTTCCACAGTTCA
FN1_R1	TCCCTGGGGATGTGACCAAT
IL2RG_F1	GTGCAGCCACTATCTATTCTCTG
IL2RG_R1	GTGAAGTGTTAGGTTCTCTGGAG