ORIGINAL INVESTIGATION

Meta-analysis of new genome-wide association studies of colorectal cancer risk

Ulrike Peters · Carolyn M. Hutter · Li Hsu · Fredrick R. Schumacher · David V. Conti · Christopher S. Carlson · Christopher K. Edlund · Robert W. Haile · Steven Gallinger · Brent W. Zanke · Mathieu Lemire · Jagadish Rangrej · Raakhee Vijayaraghavan · Andrew T. Chan · Aditi Hazra · David J. Hunter · Jing Ma · Charles S. Fuchs · Edward L. Giovannucci · Peter Kraft · Yan Liu · Lin Chen · Shuo Jiao · Karen W. Makar · Darin Taverna · Stephen B. Gruber · Gad Rennert · Victor Moreno · Cornelia M. Ulrich · Michael O. Woods · Roger C. Green · Patrick S. Parfrey · Ross L. Prentice · Charles Kooperberg · Rebecca D. Jackson · Andrea Z. LaCroix · Bette J. Caan · Richard B. Hayes · Sonja I. Berndt · Stephen J. Chanock · Robert E. Schoen · Jenny Chang-Claude · Michael Hoffmeister · Hermann Brenner · Bernd Frank · Stéphane Bézieau · Sébastien Küry · Martha L. Slattery · John L. Hopper · Mark A. Jenkins · Loic Le Marchand · Noralane M. Lindor · Polly A. Newcomb · Daniela Seminara · Thomas J. Hudson · David J. Duggan · John D. Potter · Graham Casey

Abstract Colorectal cancer is the second leading cause of cancer death in developed countries. Genome-wide association studies (GWAS) have successfully identified novel susceptibility loci for colorectal cancer. To follow up on these findings, and try to identify novel colorectal cancer susceptibility loci, we present results for GWAS of colorectal cancer (2,906 cases, 3,416 controls) that have not previously published main associations. Specifically, we calculated odds ratios and 95% confidence intervals using log-additive models for each study. In order to improve our power to detect novel colorectal cancer susceptibility loci, we performed a meta-analysis combining the results across studies. We selected the most statistically significant single nucleotide polymorphisms (SNPs) for replication using ten independent studies (8,161 cases and 9,101 controls). We again used a meta-analysis to summarize results for the replication studies alone, and for a combined analysis of GWAS and replication studies. We measured ten SNPs previously identified in colorectal cancer susceptibility loci and found eight to be associated

On behalf of CCFR (Colon Cancer Family Registry) and GECCO (Genetics and Epidemiology of Colorectal Cancer Consortium).

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

U. Peters (✉) · C. M. Hutter · C. S. Carlson · S. Jiao · K. W. Makar · C. M. Ulrich · A. Z. LaCroix · P. A. Newcomb
Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, USA
e-mail: upeters@fhcrc.org

U. Peters · C. M. Ulrich · J. D. Potter
Department of Epidemiology, School of Public Health, University of Washington, Seattle, USA

L. Hsu
Biostatistics and Biomathematics Program, Fred Hutchinson Cancer Research Center, Seattle, USA

F. R. Schumacher · D. V. Conti · R. W. Haile · G. Casey (✉)
Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, USA
e-mail: gcasey@usc.edu

C. K. Edlund
Keck School of Medicine, University of Southern California, Los Angeles, USA

S. Gallinger
Department of Surgery, Toronto General Hospital, University Health Network, Toronto, Canada

B. W. Zanke
Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

M. Lemire · J. Rangrej · T. J. Hudson
Ontario Institute for Cancer Research, Toronto, Canada

R. Vijayaraghavan · D. Taverna · D. J. Duggan
Translational Genomics Research Institute, Phoenix, USA

with colorectal cancer (p value range 0.02 to 1.8×10^{-8}). When we excluded studies that have previously published on these SNPs, five SNPs remained significant at $p < 0.05$ in the combined analysis. No novel susceptibility loci were significant in the replication study after adjustment for multiple testing, and none reached genome-wide significance from a combined analysis of GWAS and replication. We observed marginally significant evidence for a second independent SNP in the *BMP2* region at chromosomal location 20p12 (rs4813802; replication p value 0.03; combined p value 7.3×10^{-5}). In a region on 5p33.15, which includes the coding regions of the *TERT-CLPTM1L* genes and has been identified in GWAS to be associated with susceptibility to at least seven other cancers, we observed a marginally significant association with rs2853668 (replication p value 0.03; combined p value 1.9×10^{-4}). Our study suggests a complex nature of the contribution of common genetic variants to risk for colorectal cancer.

Introduction

Colorectal cancer is the second leading cause of cancer death in developed countries, with the lifetime risk estimated to be 5–6% (Ries et al. 2007). Linkage studies have identified important rare germline mutations, such as those in the *APC* gene and DNA mismatch repair genes, leading to severe syndromes, e.g. familial adenomatous polyposis and Lynch syndrome (also called hereditary non-polyposis colorectal cancer) (de la Chapelle 2004). However, these high-penetrance mutations explain only a small fraction of the genetic risk. To date, genome-wide association studies (GWAS) have identified 14 low-penetrance genetic variants that, together, explain approximately 8% of the familial association of this disease (Broderick et al. 2007; Gruber et al. 2007; Houlston et al. 2008, 2010; Tenesa et al. 2008; Tomlinson et al. 2007, 2008; Zanke et al. 2007). Based on a recent method by Chatterjee and Park (Park et al. 2010) that estimates the amount of familial association explained by common genetic variants, we estimate that about 60–70 common variants [95% confidence interval (CI) 31–173] would explain approximately 17% (95% CI 11.6–35.8%) of the familial association in colorectal cancer. Accordingly, we hypothesize that additional common colorectal cancer susceptibility loci exist that yet have to be identified, and that these loci can be identified through a genome-wide analysis of single nucleotide polymorphism (SNP) data.

As has been demonstrated in studies of other common complex diseases, power to detect novel loci is enhanced by performing meta-analysis that combines GWAS results (Zeggini and Ioannidis 2009). Therefore, we conducted a combined analysis of two recently completed scans that

A. T. Chan
Division of Gastroenterology, Massachusetts General Hospital, Harvard Medical School, Boston, USA

A. T. Chan · A. Hazra · J. Ma · C. S. Fuchs · E. L. Giovannucci
Channing Laboratory, Brigham and Women's Hospital
and Harvard Medical School, Boston, USA

A. Hazra · D. J. Hunter · P. Kraft
Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, USA

C. S. Fuchs
Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, USA

E. L. Giovannucci
Departments of Epidemiology and Nutrition, Harvard School of Public Health, Boston, USA

Y. Liu
Dallas Research Center, Stephens & Associates, Dallas, USA

L. Chen
Department of Health Studies, University of Chicago, Chicago, USA

S. B. Gruber
Department of Internal Medicine, University of Michigan, Ann Arbor, USA

G. Rennert
Department of Community Medicine and Epidemiology, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel

V. Moreno
Biostatistics and Bioinformatics Unit, Catalan Institute of Oncology-IDIBELL, Barcelona, Spain

C. M. Ulrich
Division of Preventive Oncology, German Cancer Research Center, Heidelberg, Germany

M. O. Woods · R. C. Green
Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Canada

P. S. Parfrey
Discipline of Medicine, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Canada

R. L. Prentice · C. Kooperberg · J. D. Potter
Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, USA

R. D. Jackson
Division of Endocrinology, Diabetes and Metabolism, Ohio State University, Columbus, USA

have not previously published main associations, followed by a replication study of the most significant findings using ten independent studies (Table 1; Supplemental Note; Supplemental Table 1) to follow up on the currently established colorectal cancer susceptibility loci and to try to identify additional susceptibility loci. The GWAS meta-analysis included a total of 2,906 cases and 3,416 controls recruited as part of the Colon Cancer Family Registry (CCFR), the Diet, Activity and Lifestyle Study (DALS), the Prostate, Lung, Colorectal and Ovarian Screening Trial (PLCO), and the Women's Health Initiative (WHI). The replication included a total of 8,161 cases and 9,101 controls from the Nurses' Health Study (NHS), the Health Professionals Follow-up Study (HPFS), the Physicians' Health Study (PHS), the Assessment of Risk in Colorectal Tumors In Canada (ARCTIC), additional samples from DALS and CCFR, and case–control studies from Germany, France, Israel, and Newfoundland. Most of these studies are part of the Genetics and Epidemiology Colorectal Cancer Consortium (GECCO; details in Supplemental Note).

Results

From the fixed-effects meta-analysis of GWAS scans, the inflation factor λ was 1.008, indicating little evidence of residual population substructure, cryptic relatedness, or differential genotyping between cases and controls (Supplemental Figure 1). When analyzed separately, λ was similarly low for each scan (range 1.005 to 1.01).

Initially, we attempted to validate the ten established susceptibility SNPs (p value 5×10^{-8}) that had been published at the time we selected SNPs for replication (Broderick et al. 2007; Gruber et al. 2007; Houlston et al. 2008; Tenesa et al. 2008; Tomlinson et al. 2007, 2008; Zanke et al. 2007). We found nominal evidence for association in the same direction with $p < 0.05$ from combined analyses of GWAS and replication for eight of these ten loci (rs4939827/SMAD7, rs4779584/GREM1, rs16892766/EIF3H, rs3802842/11q23, rs961253/BMP2, rs4444235/BMP4, rs9929218/CDH1, rs6983267/MYC; Table 2). When we excluded results from studies that had been previously published (Supplemental Table 2; Supplemental Figure 2), we found evidence for replication at $p < 0.05$ for five out of nine SNPs. This latter analysis did not include rs6983267, since that SNP has already been published on by a majority of the studies.

The most significant novel SNP in both the replication study, and in the combined analysis of the GWAS and replication, was rs7315438 located on chromosome 12q24 near MED13L [replication odds ratio (OR) = 0.92; replication p value 1.0×10^{-3}; combined OR = 0.92; combined fixed-effects p value (p_{fixed}) 5.6×10^{-6}; combined random-effects p value (p_{random}) 1.5×10^{-3}; Table 3; Supplemental Table 3]. For two other SNPs, their association with

B. J. Caan
Division of Research, Kaiser Permanente Medical Care Program, Oakland, USA

R. B. Hayes
Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, USA

S. I. Berndt · S. J. Chanock
Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, USA

R. E. Schoen
Department of Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, USA

J. Chang-Claude
Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

M. Hoffmeister · H. Brenner · B. Frank
Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany

S. Bézieau · S. Küry
Service de Génétique Médicale, Pôle de Biologie, Centre Hospitalier Universitaire (CHU) de Nantes, Nantes, France

M. L. Slattery
Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, USA

J. L. Hopper · M. A. Jenkins
Centre for Molecular, Environmental, Genetic, and Analytical Epidemiology, University of Melbourne, Melbourne, Australia

L. Le Marchand
Epidemiology Program, Cancer Research Center of Hawai'i, University of Hawai'i at Manoa, Honolulu, USA

N. M. Lindor
Department of Medical Genetics, Mayo Clinic, Rochester, USA

D. Seminara
Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, USA

T. J. Hudson
Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Toronto, Canada

Table 1 Studies participating in the genome-wide association study (GWAS) and replication meta-analyses

Study name	Abbreviation	Cases	Controls	Total
GWAS				
Colon Cancer Family Registry	CCFR	1,191	999	2,190
Women's Health Initiative	WHI	483	530	1,013
Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	PLCO	534	1,168	1,702
Diet, Activity and Lifestyle Survey	DALS	698	719	1,417
Replication				
Assessment of Risk in Colorectal Tumors in Canada	ARCTIC	769	665	1,434
Colon Cancer Family Registry Set II	CCFR-II	780	780	1,560
Darmkrebs: Chancen der Verhütung durch Screening	DACHS	1,731	1,742	3,473
Diet, Activity and Lifestyle Survey Set II	DALS-II	691	720	1,411
French Case–Control Study	FRENCH	954	1,060	2,014
Health Professionals Follow-up Study	HPFS	333	595	928
Molecular Epidemiology of Colorectal Cancer Study	MECC	1,686	1,779	3,465
Newfoundland Familial Colon Cancer Registry	NFCCR	409	321	730
Nurses' Health Study	NHS	432	946	1,378
Physician's Health Study	PHS	376	493	869

colorectal cancer was nominally significant within the replication study: rs4925386 located on chromosome 20q13 near *LAMA5* (replication p value 2.5×10^{-3}; p_{fixed} 2.1×10^{-4}; p_{random} 0.015) and rs16888522 located on 8q23.3 near *EIF3H* (replication p value 4.1×10^{-3}; p_{fixed} 1.7×10^{-5}; p_{random} 4.5×10^{-4}). We note that the rs4925386 SNP did not have strong evidence for association in the random-effects model. Since we selected rs4925386 for our replication study, it has been identified as being associated with colorectal cancer by a published GWAS (Houlston et al. 2010). The variant rs16888522 is in the region of rs16892766/*EIF3H*, previously identified to be associated with colorectal cancer by a GWAS (Tomlinson et al. 2008). The variant rs16888522/*EIF3H* was in weak linkage disequilibrium (LD) with rs16892766/*EIF3H* ($D' = 0.255$; $r^2 = 0.043$; Supplemental Figure 4). Conditional analysis, including both variants in the same model, resulted in less significant results for both variants (Supplemental Table 4) and showed weak correlation between the beta coefficients ($r = -0.269$), which suggests that these variants may not be independently associated SNPs.

We identified five other loci with $p < 0.05$ in our replication study and combined p value $<10^{-4}$ (Table 3). The associated SNPs were located near *BMP2*, *POLS*, *SLC8A1*, *TERT-CLPTM1L* and *TPK1*. One was in a region previously identified to be a colorectal cancer susceptibility locus by a GWAS (rs4813802/*BMP2*) (replication p value 0.03; p_{fixed} 7.3×10^{-5}; p_{random} 0.014; Table 3; Supplemental Figure 3; Supplemental Table 3). The higher p value for the random effects for this SNP reflects the fact that we observed evidence of heterogeneity among GWAS

for rs4813802 ($I^2 = 76.1\%$ and $p = 0.06$) (Ioannidis et al. 2007); however, this was less pronounced among the replication studies ($I^2 = 40.4\%$ and $p = 0.08$) suggesting that the result is consistent among studies after accounting for those that may be subject to the "winner's curse" (Garner 2007). rs4813802/*BMP2* was not in LD with the known colorectal cancer susceptibility SNP in the region rs961253/*BMP2* ($D' = 0.02$, $r^2 < 0.001$) (Houlston et al. 2008), and the joint conditional analysis demonstrates the independent association of both variants with colorectal cancer risk (correlation of beta coefficients = 0.018; Supplemental Table 4; Fig. 1).

For all SNPs in Table 3, we tested if the risk estimates of these variants may vary by mode of inheritance or sex. While for some variants (rs4813802/*BMP2*, rs275454/*POLS*, rs2373859/*SLC8A1*, and rs2853668/*CLPTM1L*), the recessive model tended to provide stronger risk estimates and slightly lower p values than the log-additive or dominant model, the AIC value was >2 in all cases, indicating no statistical evidence for improvement over the log-additive model (Supplemental Table 5). We also explored if results vary by sex and found that for rs16888522/*EIF3H* the statistical evidence for association was stronger in men (OR = 1.25; p value = 0.002) than in women (OR = 1.10; p value = 0.25), although the effect estimates were in the same direction and similar in magnitude for both men and women (Supplemental Table 6).

As a sensitivity analysis, we reran the combined fixed-effects meta-analysis leaving out one study at a time for all SNPs in Table 2. In no case did the point estimate change >3%. Further, all p_{fixed} remained $<5 \times 10^{-3}$ except for when we removed the French study from the analysis of

Table 2 Risk estimates for colorectal cancer for previously reported genome-wide association studies (GWAS) hits

SNP	Chr	Position (bp)	Gene/locus[a]	Ref. allele	MAF[b]	# Case[c]	# Control[c]	GWAS		Replication		GWAS + replication	
								OR (95% CI)	p^{d}	OR (95% CI)	p^{d}	OR (95% CI)	p^{d}
rs4939827	18	44707461	SMAD7	A	0.49	7,675	8,970	0.86 (0.79–0.93)	8.7×10^{-5}	0.90 (0.85–0.95)	2.1×10^{-4}	0.88 (0.85–0.93)	1.1×10^{-7}
rs4779584	15	30782048	SCG5, GREM1	G	0.19	7,669	8,795	1.15 (1.04–1.26)	5.2×10^{-3}	1.19 (1.11–1.28)	8.1×10^{-7}	1.18 (1.11–1.24)	1.8×10^{-8}
rs16892766	8	117699864	EIF3H	A	0.08	7,686	8,977	1.41 (1.23–1.62)	6.8×10^{-7}	1.15 (1.03–1.27)	9.3×10^{-3}	1.24 (1.14–1.34)	4.0×10^{-7}
rs3802842	11	110676919	LOC120376	A	0.28	7,677	8,782	1.13 (1.04–1.23)	4.1×10^{-3}	1.14 (1.07–1.22)	2.7×10^{-5}	1.14 (1.08–1.20)	3.8×10^{-7}
rs961253	20	6352281	LOC643503, BMP2	C	0.35	7,683	8,790	1.10 (1.01–1.19)	0.020	1.10 (1.04–1.17)	9.0×10^{-4}	1.10 (1.05–1.16)	5.2×10^{-5}
rs4444235	14	53480669	BMP4	A	0.47	7,678	8,792	1.07 (0.99–1.15)	0.091	1.06 (1–1.12)	0.055	1.06 (1.01–1.11)	0.011
rs9929218	16	67378447	CDH1	G	0.30	7,668	8,770	0.95 (0.87–1.03)	0.24	0.94 (0.88–1)	0.042	0.94 (0.90–0.99)	0.020
rs10795668	10	8741225	LOC338591	G	0.32	7,681	8,790	0.94 (0.87–1.02)	0.14	0.98 (0.92–1.04)	0.50	0.97 (0.92–1.01)	0.15
rs10411210	19	38224140	RHPN2	G	0.10	7,685	8,795	0.88 (0.77–1.01)	6.4×10^{-2}	0.98 (0.89–1.07)	0.62	0.95 (0.88–1.02)	0.14
rs6983267	8	128482487	POU5F1P1, MYC	C	0.49	4,166	4,990	0.88 (0.81–0.95)	7.5×10^{-4}	0.80 (0.69–0.93)	3.3×10^{-3}	0.86 (0.80–0.92)	1.4×10^{-5}

GWAS hits defined as variants with $p < 5 \times 10^{-8}$

[a] Nearest gene or locus for each SNP. For SNPs where two genes/loci are listed, the second gene represents a nearby gene thought to be a strong functional candidate

[b] *MAF* minor allele frequency based on controls of GWAS studies

[c] # of cases and controls differ for each SNP depending on genotyping platform used in replication and exclusions applied based on quality control measurements

[d] *p* value from fixed-effects meta-analysis

Table 3 Risk estimates for colorectal cancer for loci with $p < 5 \times 10^{-4}$ in combined genome-wide association studies (GWAS) and replication analysis

SNP	Chr	Position (bp)	Gene/locus[a]	Ref. allele	MAF[b]	# Case[c]	# Control[c]	GWAS (2,906 cases, 3,416 controls)		Replication (8,161 cases, 9,101 controls)		GWAS + replication	
								OR (95% CI)	p^{d}	OR (95% CI)	p^{d}	OR (95% CI)	p^{d}
SNPs in regions of prior GWAS hits for colorectal cancer													
rs16888522	8	117643094	EIF3H	G	0.06	9,308	10,708	1.32 (1.13–1.54)	5.0×10^{-4}	1.16 (1.05–1.28)	4.1×10^{-3}	1.20 (1.11–1.31)	1.7×10^{-5}
rs4813802	20	6647595	BMP2	A	0.35	11,498	13,075	1.18 (1.09–1.28)	5.5×10^{-5}	1.05 (1.01–1.10)	0.025	1.08 (1.04–1.13)	7.3×10^{-5}
SNPs in new regions													
rs7315438	12	114375786	MED13L	A	0.42	9,255	10,657	0.88 (0.81–0.95)	1.2×10^{-3}	0.92 (0.87–0.97)	0.001	0.90 (0.87–0.94)	5.6×10^{-6}
rs275454	5	6869013	LOC729434/POLS	G	0.37	11,348	12,925	1.16 (1.07–1.25)	2.9×10^{-4}	1.06 (1.01–1.11)	0.014	1.08 (1.04–1.13)	8.7×10^{-5}
rs2373859	2	40471324	SLC8A1	A	0.35	10,999	12,483	0.87 (0.8–0.95)	8.8×10^{-4}	0.95 (0.91–0.99)	0.010	0.93 (0.9–0.97)	1.4×10^{-4}
rs2853668	5	1353025	TERT-CLPTM1L	C	0.25	9,019	10,413	0.85 (0.78–0.93)	4.5×10^{-4}	0.94 (0.88–0.99)	0.032	0.91 (0.87–0.96)	1.9×10^{-4}
rs4925386	20	60354439	LAMA5	G	0.30	10,485	11,629	0.91 (0.84–0.99)	0.032	0.93 (0.88–0.97)	2.5×10^{-3}	0.92 (0.88–0.96)	2.1×10^{-4}
rs1525461	7	144359224	LOC643308/TPK1	A	0.19	10,978	12,483	1.20 (1.08–1.32)	3.6×10^{-4}	1.06 (1.0–1.12)	0.037	1.09 (1.04–1.14)	3.6×10^{-4}

[a] Nearest gene or locus for each SNP. For SNPs where two genes/loci are listed, the second gene represents a nearby gene thought to be a strong functional candidate

[b] MAF minor allele frequency based on controls of GWAS studies

[c] # of cases and controls differ for each single-nucleotide polymorphism (SNP) depending on genotyping platform used in replication and exclusions applied based on quality control measurements

[d] p value from fixed-effects meta-analysis

Fig. 1 Regional association results and LD structure for the associated region on chromosome 20p12.3/*BMP2* locus. The *top half* of the figure has physical position along the *x*-axis, and the $-\log_{10}$ of the meta-analysis *p* value on the *y*-axis. *Each dot on the plot* represents the result for one SNP. The *bottom half* of the figure shows pairwise linkage disequilibrium (LD) for the genotyped SNPs across the region. LD was measured as r^2 and calculated using the control individuals from the WHI, PLCO and DALS samples recruited from 53 centers across the USA. *Darker shading* indicates higher levels of LD. The *lines* between the *top and bottom half* of the figure connecting the same SNPs
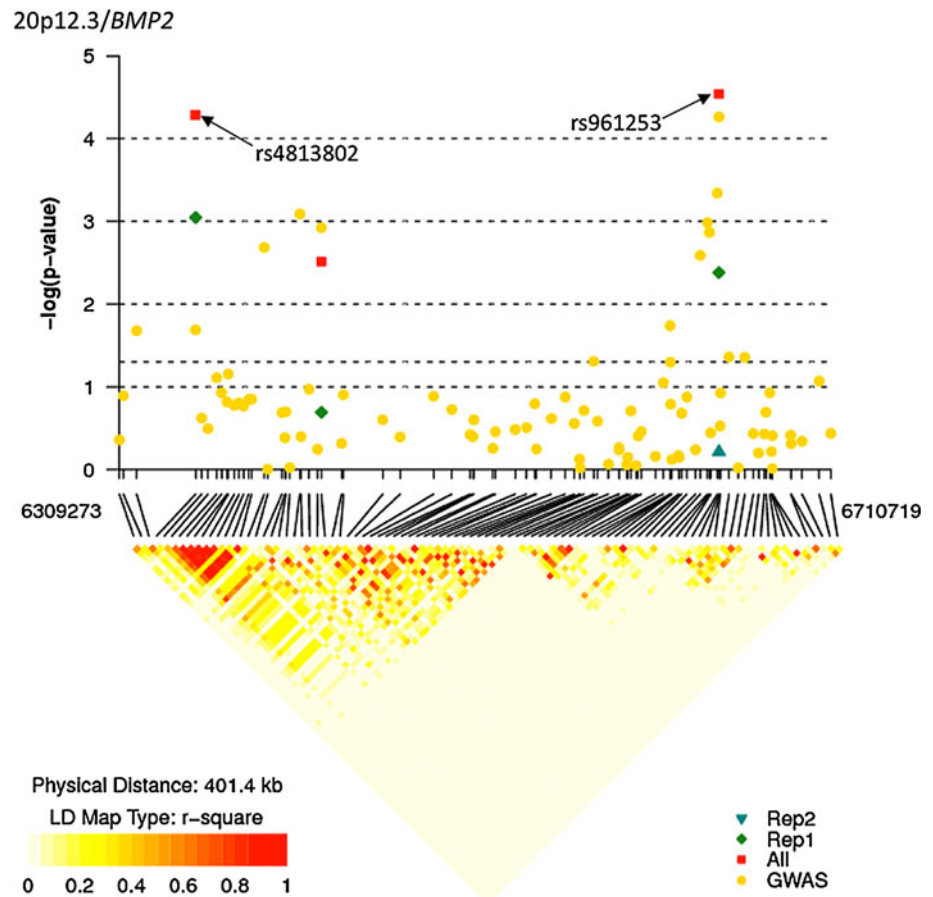
rs4925386. In that case the OR remained similar (OR = 0.94) but the *p* value was slightly attenuated $p_{\text{fixed}} = 8.2 \times 10^{-3}$.

Discussion

From the analysis of GWAS and replication, including a total of up to 11,067 cases and 12,517 controls, we found that SNPs in eight out of ten previously identified colorectal cancer susceptibility loci were associated with the disease in our replication study at $p < 0.05$. We found evidence that a second SNP (rs4813802) near the *BMP2* gene could be associated with colorectal cancer, independent of the association with the previously identified susceptibility SNP in that region (rs961253). Furthermore, our study reports for the first time a potential new association of a variant in the *TERT-CLPTM1L* region with colorectal cancer risk.

Our results provide further support for eight of ten previously identified GWAS hits. When excluding studies that have previously published results on these known loci, five loci showed evidence of replication in this independent subsample. The 8q24 SNP rs6983267 has already been heavily studied, including published reports for many of the studies included in this paper (Figueiredo et al. 2011;

Hutter et al. 2010), so we were not able to examine independent replication of this SNP in this study. Among the remaining four loci that did not show a significant association at $p < 0.05$, three showed a trend toward replication (with $p < 0.2$ and an OR in the same direction as the original GWAS report). However, one SNP, rs10795668, did not show any evidence for association with disease (OR = 1.00; 95% CI 0.93–1.08; $p = 0.96$; Supplemental Figure 2). Several papers have reviewed potential reasons for the lack of replication of GWAS findings (Chanock et al. 2007; Kraft et al. 2009). As in any observational study, it is possible this represents either a false positive in the initial report or a false negative in this replication; although that seems unlikely since both the discovery GWAS and this report are based on large, well-powered studies. We used the same genetic model and similar trait definitions as the discovery GWAS. Further, all studies were restricted to non-Hispanic Whites, limiting the possibility of differences in LD patterns. It is possible that there may be differences in the distribution of a key effect modifier between the studies used to identify rs10795668 and the studies presented in this paper. A full exploration of underlying gene–gene or gene–environment interactions is beyond the scope of the current paper, but we did explore if the effect of rs10795668 varied by sex. Although the results were not significant for either

sex, and the 95% CIs overlap, we do note an interesting pattern where the ORs are in opposite directions for women and men, with men showing a trend in the direction of the discovery GWAS. Specifically, we found $OR_{women} = 1.07$ (95% CI 0.92–1.26; $p_{fixed} = 0.38$) and $OR_{men} = 0.95$ (95% CI 0.84–1.07; $p_{fixed} = 0.40$).

The rs4925386/*LAMA5* SNP was also recently identified in another GWAS meta-analysis (Houlston et al. 2010). Although it was not a known colorectal cancer susceptibility locus at the time we selected SNPs for replication, this SNP met our criteria for selection, and showed evidence for association in our replication sample. The rs4925386 variant lies in the intron of the large laminin A5 protein encoding gene. As previously reported the variant is in LD ($r^2 > 0.5$) with four nonsynonymous SNPs in *LAMA5* (Houlston et al. 2010). However, the prediction of each of these amino acid changes is proposed to be benign. Overall, our finding provides additional independent support that this variant is associated with susceptibility to colorectal cancer.

None of the loci were significantly associated with colorectal cancer in our replication study after adjusting for multiple testing ($0.05/321 = 1.6 \times 10^{-4}$), and none of the loci reached "genome-wide significance" at the suggested p values of 1.6×10^{-7} after accounting for the two-stage design (for details, see "Materials and methods") (Dudbridge and Gusnanto 2008; Hoggart et al. 2008; International HapMap Consortium 2005; Pe'er et al. 2008; Risch and Merikangas 1996; Wellcome Trust Case Control Consortium 2007). However, for some of the variants with $p < 0.05$ in our replication and combined p value $<10^{-4}$, additional lines of evidence provide support for the hypothesis that we may have identified genomic regions harboring causal variants for colorectal cancer susceptibility. The variant rs4813802 is about 295.3 kb centromeric to the previously identified rs961253/*BMP2* GWAS hit (Houlston et al. 2008); both statistical models and LD data support the idea that these are independent signals. The closest gene is bone morphogenetic protein 2 (*BMP2*). The new variant of interest, rs4813802, is closer to *BMP2* (49.2 kb upstream) than the previously identified SNP rs961253 (344.5 kb upstream of *BMP2*). Interestingly, rs4813802 lies within an ENCODE Digital DNAseI Hypersensitivity Cluster; it is also within an ENCODE region showing H3K4Me1 enhancer associated histone marks (Rosenbloom et al. 2010), and the flanking 15 bp shows strong placental mammal conservation by PhastCons (Siepel et al. 2005). While not conclusive, all of these are consistent with the region flanking the SNP acting as a long-range enhancer element, plausibly for *BMP2*. The *BMP2* gene belongs to the transforming growth factor-β (TGFβ) superfamily, which plays an important role in cell proliferation, differentiation, and apoptosis (Massague 2000). SNPs in five out of the ten known colorectal cancer

SNPs have chromosomal locations in or near TGFβ superfamily genes (Tenesa and Dunlop 2009). Furthermore, loss in BMP signaling has been reported at the transition from advanced adenoma to early cancer stage, compatible with a role in tumor progression (Hardwick et al. 2008). Support for a role for BMP signaling in colorectal cancer comes from the identification of mutations in the bone morphogenetic protein receptor, type IA protein (*BMPR1A*) in juvenile polyposis (Howe et al. 2001). Individuals with familial juvenile polyposis have a 20% risk of colon cancer by age 35 and 68% by age 60 (Schreibman et al. 2005). Our finding supports the possibility of allelic heterogeneity at the *BMP2* locus, which is consistent with findings for the 8q24 cancer locus (Al Olama et al. 2009; Witte 2007) and recent findings for height showing evidence for allelic heterogeneity at as many as 19 loci (Lango et al. 2010). Similar to our finding, these 19 secondary signals in height were rather distant (on average 177 kb) from the initial index SNP that was found to be associated through GWAS (Lango et al. 2010). Accordingly, a comprehensive exploration of already discovered colorectal cancer loci may uncover additional independent variants. However, this example demonstrates that defining the boundaries of a susceptibility locus may be challenging, because the SNP we identified (rs4813802) would not have been included if we had defined the region around the initial index SNP (rs961253) by LD.

The 8q24 region has been shown to have multiple independent variants that are associated with cancers. Several of these variants are associated with more than one cancer, and some cancers are associated with multiple variants in this region (Al Olama et al. 2009; Witte 2007). Similarly, multiple variants associated with various cancer sites, including cancers of lung, pancreas, testes, and bladder, as well as glioma, basal cell carcinoma, and melanoma are found in the *TERT-CLPTM1L* region (Fig. 2) (Hsiung et al. 2010; Landi et al. 2009; McKay et al. 2008; Miki et al. 2010; Petersen et al. 2010; Rafnar et al. 2009; Shete et al. 2009; Stacey et al. 2009; Turnbull et al. 2010; Wang et al. 2008). Ours is the first report suggesting that a variant in the *TERT-CLPTM1L* region could be associated with colorectal cancer. The variant rs2853668 is 4.9 kb upstream of telomerase reverse transcriptase (*TERT*) and 18.0 kb downstream of cleft lip and palate transmembrane protein 1-like protein (*CLPTM1L*). Both genes have been implicated in cancer: *CLPTM1L* has been shown to be altered in cisplatin-resistant cell lines and potentially impacts apoptosis (Yamamoto et al. 2001); *TERT* encodes for the telomerase catalytic subunit that is important for the replication and stabilization of telomere ends, and subsequently impacts chromosome replication and suppression of cell senescence. Malfunction of telomerase can result in chromosomal abnormality and

Fig. 2 Genetic variants associated with different cancer sites in the *TERT-CLPTM1L* region, including the new finding for colorectal cancer for rs2853668. This figure shows the genomic region on chromosome 5.p.15.33 including the two genes *TERT* and *CLPTM1L*. The *top* of the figure shows the genes with each exon represented by a *short vertical line*. The *line* below the genes shows the location of the different SNPs that have been associated with various cancer sites relative to the position of the genes (for instance, the first two SNPs are located in the last intron of *TERT*). The *triangle* shows the pairwise linkage disequilibrium (LD) of the SNPs. LD was measured as r^2 and calculated using the control individuals from the WHI, PLCO and DALS samples recruited from 53 centers across the USA. *Darker shading* indicates higher levels of LD

rs2736100	glioma, lung, and testicular cancer
rs2853676	glioma
rs2736098	basal cell carcinoma, lung and bladder cancer
rs2853668	**colorectal cancer**
rs4635969	lung and testicular cancer
rs402710	lung cancer
rs401681C	basal cell carcinoma, lung, and bladder cancer
rs401681T	pancreas cancer and melanoma
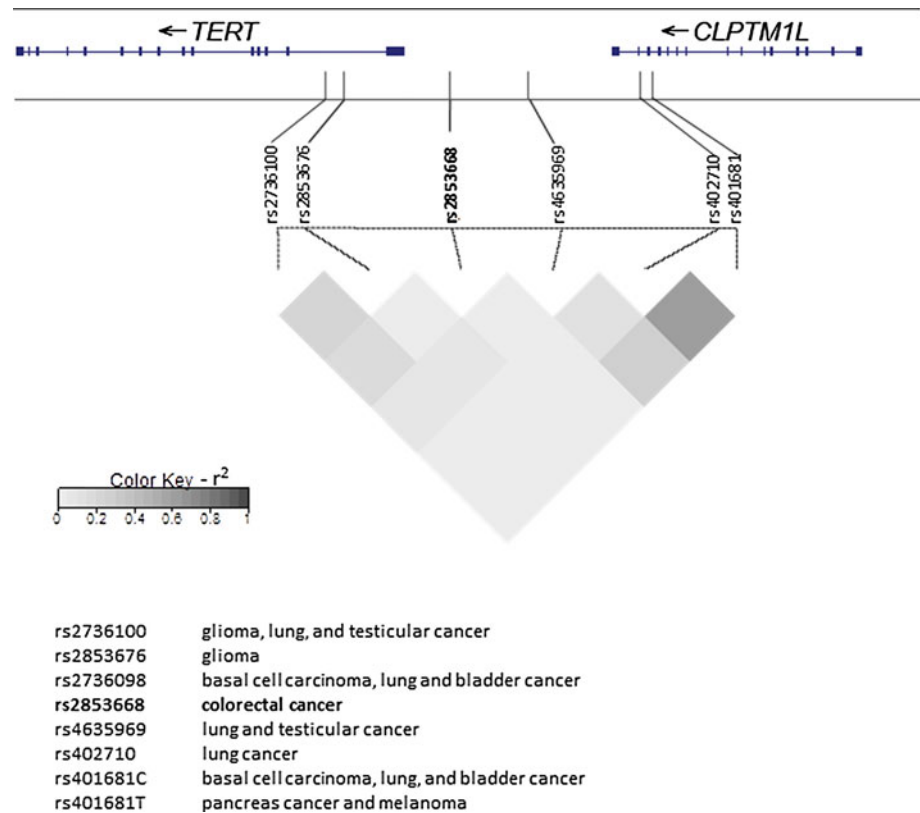
subsequent tumor formation (Rafnar et al. 2009). Our finding provides further evidence that *TERT-CLPTM1L* is a general cancer susceptibility locus that impacts critical function for cancer development, similar to the 8q24 region. The candidate gene in the 8q24 loci is *MYC*, and as noted by Johnatty et al. (2010), these two loci could act in concert. Specifically, the *TERT* promoter has several *MYC* (the nearest gene to the 8q24 locus) binding sites (Wu et al. 1999); however, we did not observe a statistically significant interaction between rs2853668/*TERT-CLPTM1L* and rs6983267/8q24, *MYC* (p for interaction term = 0.8).

The SNP rs7315438, which showed the most statistically significant association in both the replication study alone, as well as in the combined meta-analysis of GWAS and replication studies, is located on chromosome 12q24 about 76.9 kb upstream of the T-box 3 protein (*TXB3*). The SNP is also located 50.4 kb downstream of *MED13L*, which encodes for a subunit of the mediator complex, a large complex of proteins that functions as a transcriptional coactivator for most RNA polymerase II-transcribed genes. Since it has been implicated in transcription, this gene is a plausible candidate for further study. However, this SNP is in a large LD region containing numerous other potential candidate genes, including the kinase suppressor of RAS2 (*KSR2*).

Other SNPs identified as potentially associated with colorectal cancer in this study are rs27545 (*POLS*),

rs2373859 (*SLC8A1*) and rs1525461 (LOC643308/*TPK1*). The gene closest to rs27545 is *POLS* (59 kb downstream), a DNA polymerase that is likely involved in DNA repair and, hence, provides a potentially interesting candidate gene (Hubscher et al. 2002). Other genes close to rs27545 are *SRD5A1* (146 kb upstream), which converts testosterone into the more potent dihydrotestosterone, and the methyltransferase *NSUN2* (183 kb downstream), which methylates tRNA (Brzezicha et al. 2006). The SNP rs2373859 resides in the intronic region of *SLC8A1* also known as *NCX1*, which is a cell membrane protein that is involved in the rapid Ca(2+) transports (Annunziato et al. 2004). It is in a gene-rich region including other interesting candidates, such as *MAP4K3* (954 kb upstream), a member of the mitogen-activated protein kinases, which is involved in regulating both cell growth and death and has altered gene expression in many cancer types (Cuadrado and Nebreda 2010) and *SOS1*, which may act as a positive regulator of RAS (Freedman et al. 2006). The closest gene to rs1525461 is *TPK1* (195 kb upstream). *TPK1* is involved in the regulation of thiamine metabolism (Timm et al. 2001). *TPK1* flanks a gene-rich region, including several olfactory receptors but none of the genes has an obvious link to colorectal cancer development. However, the assignment of SNPs to candidate genes should be done with caution, as recently shown by additional fine mapping and in silico analysis of the previously identified colorectal

cancer loci 8q23.3 (*EIF3H*), 16q22.1 (*CDH1/CDH3*), which suggested functional variation in unexpected candidate target genes (Carvajal-Carmona et al. 2011).

Overall, our study suggests a complex nature of the contribution of common genetic variants to risk for colorectal cancer, and suggests the need for additional studies to identify variants with marginal effects, as well as studies to examine potential sources and role of heterogeneity, including gene–gene and gene–environment interactions. We note that this study focused on the log-additive model. Although we present results for other genetic models for our top findings, our results may have been biased for SNPs that do not follow this assumed log-additive model (Minelli et al. 2005). Further, this study was not set up to investigate less frequent (allele frequency 1–5%) and rare variants (allele frequency < 1%), which have the potential to contribute substantially to the genetic susceptibility of colorectal cancer (Bodmer and Bonilla 2008; Cirulli and Goldstein 2010; Manolio et al. 2009).

In summary, we replicated the majority of SNPs that have previously been found to be associated with CRC in GWAS studies. We also report suggestive evidence for an additional independent signal for colorectal cancer risk in the *BMP2* locus and a possible new association of colorectal cancer with a variant in the multi-cancer susceptibility locus around *TERT-CLPTM1L*. Future studies are needed to try to replicate these findings, and if successful, to identify the underlying variants directly responsible for the association, and to study the underlying molecular mechanisms.

Materials and methods

Study participants

The studies and their abbreviations are listed in Table 1, and each study is described in detail in the Supplemental Note. In brief, all cases were defined as colorectal adenocarcinoma (International Classification of Disease Code 153-154) and confirmed by medical records, pathologic reports, or death certificate. All cases and controls were self-reported as White, which was confirmed in GWAS samples based on genotype data. All participants gave written informed consent and studies were approved by the Institutional Review Board.

Study design

The GWAS meta-analysis results are based on two scans. One GWAS was conducted within the CCFR, including population-based cases and unrelated population-based controls from three sites: USA, Canada, and Australia

(Figueiredo et al. 2011). In total, 1,191 cases and 999 controls were successfully genotyped on the Illumina 1M/1M Duo platform and passed all quality-control (QC) steps. The second scan was conducted across three US studies: the WHI and PLCO cohorts and the DALS population-based case–control study. A total of 1,715 colon cancer cases and 2,417 controls were successfully genotyped on the Illumina HumanHap 550K, 610K or combined Illumina 300K and 240K platforms and passed all QC steps. After applying rigorous genotyping QC filters (see below), a total of 378,739 directly genotyped SNPs commonly shared among the scans were included in the GWAS meta-analysis. To further boost the power and inform the ranking of SNPs, we included summary statistics from a previously published colorectal cancer GWAS (Colorectal Tumour Gene Identification Consortium, CORGI) in the meta-analysis (The Institute of Cancer Research 2008; Tomlinson et al. 2008). However, to ensure independence of results from prior published scans, we did not include any CORGI results in any of the presented ORs or *p* values.

Fixed-effects *p* values from the GWAS meta-analysis were used to select SNPs for replication. We rank ordered the top SNPs. We used LD information in our controls to prune out "redundant" signals (defined as $r^2 > 0.5$ for SNPs ≥ 100 kb apart and $r^2 > 0.1$ for SNPs < 100 kb apart). For the top five SNPs, with $p < 10^{-5}$, we selected two other SNPs with $r^2 > 0.9$ to ensure against potential genotyping failure. We then went down the ranked list until we filled our SNP platform (total number of SNPs selected for this project = 343). SNPs were excluded based on *p* value for heterogeneity < 0.001 ($n = 1$) and poor clustering in visual inspection of cluster plots ($n = 3$). If SNPs had a low design score, we replaced them with an alternative SNP with $r^2 > 0.9$. The lowest ranked SNP had *p* value 1.2×10^{-3}. Our platform also included SNPs for the ten known colorectal cancer susceptibility loci published in previous GWAS at the time we designed the platform. These 343 SNPs were genotyped in samples from DACHS, DALS, French, HPFS, NHS and PHS studies ($N = 4,062$ cases and 4,718 controls) (Table 1; Supplemental Note) and 306 SNPs were successfully genotyped in all studies (see details below). After we selected SNPs for replication, the ARCTIC genome-wide scan became available (769 cases and 665 controls), and we used imputed data from that study for analysis of the 343 SNPs (12 SNPs were not included due to low imputation quality or low HWE *p* values). As of April 2010, we had genotyped and analyzed the GWAS data and replication data from ARCTIC, DACHS, DALS and the French case–control study. We selected 32 SNPs with $p < 0.1$ in this replication set and/or a $p_{\text{fixed}} < 10^{-4}$ in the combined replication and GWAS for further genotyping in 2,550 cases and 3,539 controls, including additional samples

from NHS, PHS, and HPFS, and samples from MECC and NFCCR. The top SNPs were also analyzed in a second set of data from the CCFR (780 cases and 780 controls). We present results for the total replication sample of 8,161 cases and 9,101 controls.

Genotyping

Genomic DNA was extracted from blood samples or, in the case of a subset of PLCO samples, from buccal cells using conventional methods.

GWAS for CCFR

Genotyping was completed on the Illumina Human1M and Human1M-Duo Bead Array in accordance with the manufacturer's protocol.

Sample exclusions The following sample exclusion criteria were applied: call rate $< 95\%$ ($n = 75$), any stripe (physical/analytical location on BeadChip) call rate $< 80\%$ ($n = 9$), discordance with prior genotyping ($n = 3$), non-White ($n = 29$), samples that showed admixture identified using the program STRUCTURE ($n = 33$) (Falush et al. 2003; Pritchard et al. 2000), high identity by descent using PLINK ($n = 2$), and mismatch between called and phenotypic sex ($n = 4$). The final analysis was based on 1,191 cases and 999 controls.

SNP exclusions SNPs were excluded if they did not overlap between the Illumina Human1M and Human1M-Duo ($n = 190,301$), were annotated as "Intensity Only" ($n = 8,263$), had call rates $< 90\%$ on either the Illumina Human1M or Human1M-Duo ($n = 9,229$), or by study center or case–control status ($n = 12,695$). When further restricting analysis to SNPs with Hardy–Weinberg Equilibrium (HWE) $p > 0.0001$, MAF > 0.05, and SNP call rate > 0.98, a total of 739,733 SNPs remained in the analysis.

Average sample call rate was equal to 98.6% with $>94\%$ of samples having a call rate $> 98\%$. Intra- and interplate replicate concordance rates were equal to 99.97 and 98.7%, respectively.

GWAS for DALS, PLCO and WHI

Genotyping was completed using Illumina HumanHap300 and HumanHap240S (PLCO), 550K (WHI, DALS) and 610K (DALS, PLCO) BeadChip Array System on the Infinium platform in accordance with the manufacturer's protocol or as previously described for HumanHap300 and HumanHap240S (Yeager et al. 2007).

Sample exclusions Samples were excluded if the average call rate was $<97\%$ (DALS: $n = 110$, PLCO: $n = 63$, WHI: $n = 66$) or there was a mismatch between called and phenotypic sex (DALS: $n = 6$, PLCO: $n = 1$). To search for unexpected duplicates and closely related individuals we calculated identity-by-state values. We excluded unexpected duplicates (DALS, $n = 2$). Additionally, we excluded samples based on low concordance with prior genotyping (DALS: $n = 10$, WHI: $n = 1$) as well as samples that did not cluster with the CEU samples in principal component analysis including the three HapMap populations as a reference (DALS: $n = 20$, PLCO: $n = 2$, WHI: $n = 6$). The final analysis was based on 698 cases and 719 controls in DALS, 534 cases and 1,168 controls in PLCO, and 483 cases and 530 controls in WHI.

SNP exclusions Because we combined data from different platforms, we took precautions to exclude SNPs that do not perform consistently across platforms. This included SNPs reported by Illumina as not performing consistently across platforms ($n = 78$), SNPs found to have more than one discordant call across the 550K and 610K platforms in HapMap Data or our interplatform duplicates ($n = 185$); and SNPs with different MAF calls on the two platforms in our control populations ($n = 9$). We further filtered SNPs within each study (DALS, PLCO, WHI) based on MAF $< 0.05\%$ or HWE in controls < 0.0001. We applied a call rate per chip type per study of $>98\%$. A total of 392,361 SNPs passed all QC checks for all three studies.

The average sample call rate was $\geq98.8\%$ in any of the three studies, and the concordance rate of blinded duplicates ($n = 98$ pairs) was $>97\%$.

When we combined data across all scans, a total of 378,739 autosomal SNPs were successfully genotyped across all studies and used in our final GWAS meta-analysis of 2,906 cases and 3,416 controls.

Replication

Genotyping of 343 SNPs in DACHS, DALS, French, and the first sub-sets of HPFS, NHS and PHS were carried out using BeadXpress technology according to the manufacturer's protocol. Problematic genotype clusters were visually inspected by lot number and the calling algorithm was adjusted, if indicated. 35 SNPs were excluded from the analysis due to poor cluster quality and 2 SNPs were excluded for being out of HWE ($p < 0.0001$) in controls of at least one study. The 306 SNPs in the replication had call rates $> 92\%$ across studies (average call rate per SNP per study 97.8%). MECC and NFCCR samples were genotyped using Matrix-assisted Laser Desorption/Ionization Time-of-Flight on the Sequenom® MassARRAY 7K platform (Sequenom, Inc., San Diego, CA). A total of 23 and 30 SNPs were successfully genotyped in MECC and NFCCR, respectively. Additional samples from NHS, HPFS and PHS were genotyped on 29 SNPs using the TaqMan® OpenArray® Genotyping Instrument Platform

Assays (Applied Biosystems, Carlsbad, CA). Overall, 32 SNPs had call rates > 98% across studies (average call rate per SNP per study 99.5%; Supplemental Table 7), indicating excellent quality.

Two GWAS data sets (ARCTIC and CCFR II) became available after the GWAS meta-analysis and were used only for replication as described above. ARCTIC has been previously published (Zanke et al. 2007). Because ARCTIC was genotyped on the Affymetrix platform with limited overlap of SNPs with the Illumina platforms, we made use of imputed data for this study. Imputation was done with BEAGLE, using the phased HapMap release 22 as the reference sample (http://ftp.hapmap.org/phasing/2007-08_rel22/) (Browning and Browning 2009). SNPs were removed if they were out of HWE ($p < 0.0001$) in the controls ($n = 1$) or had an imputation $r^2 < 0.3$ ($n = 11$). For CCFR phase II, samples were genotyped using the Illumina 1M Omni. Inclusion/exclusion criteria for cases in phase II were consistent to those described for phase I.

Statistical analysis

Study-specific analysis of GWAS data

To estimate the association between each genetic marker and risk for colorectal cancer we calculated ORs and 95% CIs using log-additive genetic models relating the genotype dose (0, 1 or 2 copies of the minor allele) to risk of colorectal cancer. We adjusted for age, sex (when appropriate), center and the first three principal components from EIGENSTRAT to account for population substructure. The CCFR calculated these estimates with Cochran–Mantel–Haenzsel analysis with strata defined by age, sex, and center.

Quantile–quantile (Q–Q) plots were assessed to determine whether the distribution of the p values in each study population was consistent with the null distribution (except for the extreme tail; Supplemental Figure 1). To quantify the data in the QQ plots, we calculated the inflation factor (λ) to measure the over-dispersion of the test statistics from association tests by dividing the mean of the test statistics by the mean of the expected values from a Chi-square distribution with 1 degree of freedom.

Combined analysis of GWAS

We conducted inverse-variance weighted fixed-effects meta-analysis to combine OR estimates from log-additive models or multiplicative methods across individual studies as described above. In this approach, we weighted the beta estimates of each study by their inverse variance and calculated a combined estimate by summing the weighted betas and dividing by the summed weights. We chose to

focus on fixed effects because we only had a small number of studies. When the number of studies is small, the between study variance may be poorly estimated, resulting in deflated test statistics for association. As such, fixed-effects analysis is better powered for discovery of novel variants (Kraft et al. 2009). We calculated I^2, which is a measure of the percentage of total variation across studies due to heterogeneity beyond chance, and obtained the heterogeneity p values based on Cochran's Q statistic (Ioannidis et al. 2007).

Study-specific analysis of replication data

To estimate the association between each genetic marker and risk for colorectal cancer, we calculated ORs and 95% CIs using a log-additive genetic model relating the genotype dose (0, 1 or 2 copies of the minor allele) to risk of colorectal cancer and adjusting for age, sex, and study center (as appropriate) in logistic regression analysis.

Combined analysis of replication data

We conducted inverse-variance weighted fixed-effects meta-analysis to combine OR estimates from log-additive models across individual studies and measured heterogeneity using I^2 and Cochran's Q statistic, as discussed above.

Analysis of combined GWAS and replication data

We again combined across studies using inverse-variance weighted fixed-effects meta-analysis. For novel SNPs with $p < 5 \times 10^{-4}$ based on combined analysis of GWAS and replication, we also report random effects that incorporate potential heterogeneity into the effect estimate. For these SNPs, we also examined dominant, recessive and unrestricted genetic models and compared models by calculating the Akaike information criterion (AIC). We performed stratified analyses and evaluated whether the effects differed by sex. For novel SNPs in regions identified by previous GWAS, we also performed a conditional analysis including both the newly and previously identified SNPs in the region in one model to examine whether the effect of the newly identified SNP can be explained by the existing one. To quantify the independence of the novel SNPs from prior GWAS hits in the same region, we calculated the variance–covariance matrix and reported the correlation between the two betas. Finally, we performed a sensitivity analysis where we removed the studies one at a time and examined the results from the fixed-effect meta-analysis. We report any situations where removing one study resulted in a >5% change in the OR point estimate and/or reduced the p value of the combined fixed-effects

meta-analysis to be $<5 \times 10^{-3}$, since that would indicate the results might be being driven by only one study.

Criterion for genome-wide significance

Based on an increasing number of papers (Dudbridge and Gusnanto 2008; Hoggart et al. 2008; International HapMap Consortium 2005; Pe'er et al. 2008; Risch and Merikangas 1996; Wellcome Trust Case Control Consortium 2007) providing a detailed discussion on the appropriate genome-wide significance threshold, which all arrive at similar values in the range of 5×10^{-7} to 5×10^{-8} for White populations, we decided to use a p value of 5×10^{-8} as genome-wide significance threshold. To account for the two-stage approach (GWAS and replication), we calculated that an overall p value of 5×10^{-8} is equal to a combined two-stage p value of 1.6×10^{-7} given our sample sizes in the GWAS and replication and a threshold for selecting SNPs from the GWAS of 1.2×10^{-3} as used here.

We used PLINK (Purcell et al. 2007; Purcell 2011) and R (R Development Core Team 2011) to conduct the statistical analysis and summarized results graphically using STATA (StataCorp 2009), snp.plotter (Luna and Nicodemus 2007), and LocusZOOM (Pruim et al. 2010).

References

Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S, Saunders E, Hopper JL, Southey MC, Muir KR, English DR, Dearnaley DP, rdern-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Sawyer E, Lophatananon A, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper C, Donovan JL, Hamdy FC, Neal DE, Eeles RA, Easton DF (2009) Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet 41:1058–1060

Annunziato L, Pignataro G, Di Renzo GF (2004) Pharmacology of brain Na+/Ca2+ exchanger: from molecular biology to therapeutic perspectives. Pharmacol Rev 56:633–654

Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40:695–701

Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I, Houlston RS (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 39:1315–1317

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223

Brzezicha B, Schmidt M, Makalowska I, Jarmolowski A, Pienkowska J, Szweykowska-Kulinska Z (2006) Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). Nucleic Acids Res 34:6034–6043

Carvajal-Carmona LG, Cazier JB, Jones AM, Howarth K, Broderick P, Pittman A, Dobbins S, Tenesa A, Farrington S, Prendergast J, Theodoratou E, Barnetson R, Conti D, Newcomb P, Hopper JL, Jenkins MA, Gallinger S, Duggan DJ, Campbell H, Kerr D, Casey G, Houlston R, Dunlop M, Tomlinson I (2011) Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. Hum Mol Genet 20:2879–2888

Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA,

Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS (2007) Replicating genotype–phenotype associations. Nature 447:655–660

Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11:415–425

Cuadrado A, Nebreda AR (2010) Mechanisms and functions of p38 MAPK signalling. Biochem J 429:403–417

de la Chapelle A (2004) Genetic predisposition to colorectal cancer. Nat Rev Cancer 4:769–780

Dudbridge F, Gusnanto A (2008) Estimation of significance thresholds for genomewide association scans. Genet Epidemiol 32:227–234

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Figueiredo JC, Lewinger JP, Song C, Campbell PT, Conti DV, Edlund CK, Duggan DJ, Rangrej J, Lemire M, Hudson T, Zanke B, Cotterchio M, Gallinger S, Jenkins M, Hopper J, Haile R, Newcomb P, Potter J, Baron JA, Marchand LL, Casey G (2011) Genotype–environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol Biomarkers Prev 20:758–766

Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, Ingles SA, Altshuler D, Henderson BE, Reich D (2006) Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci USA 103:14068–14073

Garner C (2007) Upward bias in odds ratio estimates from genome-wide association studies. Genet Epidemiol 31:288–295

Gruber SB, Moreno V, Rozek LS, Rennert HS, Lejbkowicz F, Bonner JD, Greenson JK, Giordano TJ, Fearon ER, Rennert G (2007) Genetic variation in 8q24 associated with risk of colorectal cancer. Cancer Biol Ther 6:1143–1147

Hardwick JC, Kodach LL, Offerhaus GJ, Van Den Brink GR (2008) Bone morphogenetic protein signalling in colorectal cancer. Nat Rev Cancer 8:806–812

Hoggart CJ, Clark TG, De Iorio IM, Whittaker JC, Balding DJ (2008) Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol 32:179–185

Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS, Kerr D, Cazier JB, Niittymaki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IPM, Farrington SM, Tenesa A, Prendergast JGD, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PDP, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 40:1426–1435

Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, Spain SL, Broderick P, Domingo E, Farrington S, Prendergast JG, Pittman AM, Theodoratou E, Smith CG, Olver B, Walther A, Barnetson RA, Churchman M, Jaeger EE, Penegar S, Barclay E, Martin L, Gorman M, Mager R, Johnstone E, Midgley R, Niittymaki I, Tuupanen S, Colley J, Idziaszczyk S, Thomas HJ, Lucassen AM, Evans DG, Maher ER, Maughan T, Dimas A, Dermitzakis E, Cazier JB, Aaltonen LA, Pharoah P, Kerr DJ, Carvajal-Carmona LG, Campbell H, Dunlop MG, Tomlinson IP (2010) Meta-analysis of three genome-wide association studies

identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet 42:973–977

Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, Petersen GM, Velculescu VE, Traverso G, Vogelstein B (2001) Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet 28:184–187

Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, Chatterjee N, Brennan P, Wu C, Zheng W, Chang GC, Wu T, Park JY, Hsiao CF, Kim YH, Shen H, Seow A, Yeager M, Tsai YH, Kim YT, Chow WH, Guo H, Wang WC, Sung SW, Hu Z, Chen KY, Kim JH, Chen Y, Huang L, Lee KM, Lo YL, Gao YT, Kim JH, Liu L, Huang MS, Jung TH, Jin G, Caporaso N, Yu D, Kim CH, Su WC, Shu XO, Xu P, Kim IS, Chen YM, Ma H, Shen M, Cha SI, Tan W, Chang CH, Sung JS, Zhang M, Yang TY, Park KH, Yuenger J, Wang CL, Ryu JS, Xiang Y, Deng Q, Hutchinson A, Kim JS, Cai Q, Landi MT, Yu CJ, Park JY, Tucker M, Hung JY, Lin CC, Perng RP, Boffetta P, Chen CY, Chen KC, Yang SY, Hu CY, Chang CK, Fraumeni JF, Jr., Chanock S, Yang PC, Rothman N, Lin D (2010) The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. PLoS Genet 6

Hubscher U, Maga G, Spadari S (2002) Eukaryotic DNA polymerases. Annu Rev Biochem 71:133–163

Hutter CM, Slattery ML, Duggan DJ, Muehling J, Curtin K, Hsu L, Beresford SA, Rajkovic A, Sarto GE, Marshall JR, Hammad N, Wallace R, Makar KW, Prentice RL, Caan BJ, Potter JD, Peters U (2010) Characterization of the association between 8q24 and colon cancer: gene–environment exploration and meta-analysis. BMC Cancer 10:670

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320

Ioannidis JP, Patsopoulos NA, Evangelou E (2007) Heterogeneity in meta-analyses of genome-wide association investigations. PLoS ONE 2:e841

Johnatty SE, Beesley J, Chen X, Macgregor S, Duffy DL, Spurdle AB, de Fazio A, Gava N, Webb PM, Rossing MA, Doherty JA, Goodman MT, Lurie G, Thompson PJ, Wilkens LR, Ness RB, Moysich KB, Chang-Claude J, Wang-Gohrke S, Cramer DW, Terry KL, Hankinson SE, Tworoger SS, Garcia-Closas M, Yang H, Lissowska J, Chanock SJ, Pharoah PD, Song H, Whitemore AS, Pearce CL, Stram DO, Wu AH, Pike MC, Gayther SA, Ramus SJ, Menon U, Gentry-Maharaj A, Anton-Culver H, Ziogas A, Hogdall E, Kjaer SK, Hogdall C, Berchuck A, Schildkraut JM, Iversen ES, Moorman PG, Phelan CM, Sellers TA, Cunningham JM, Vierkant RA, Rider DN, Goode EL, Haviv I, Chenevix-Trench G (2010) Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility "hot-spot". PLoS Genet 6:e1001016

Kraft P, Zeggini E, Ioannidis JP (2009) Replication in genome-wide association studies. Stat Sci 24:561–573

Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Wang Y, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metsapalu A, Vooder T, Nelis M, Valk K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeboller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF Jr, Savage SA, Bertazzi PA, Tucker MA, Chanock S, Caporaso NE (2009) A genome-wide association study of lung cancer identifies a region of

chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet 85:679–691

Lango AH, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon SA, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin LK, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Carola ZM, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Hua ZJ, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani JM, Kaplan LM, Kettunen J, Konig IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh NJ, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haiqing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpelainen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Pare G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietilainen KH, Pouta A, Ridderstrale M, Rotter JI, Sambrook JG, Sanders AR, Oliver SC, Sinisalo J, Smit JH, Stringham HM, Bragi WG, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S, van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kahonen M, Kaprio J, Kathiresan S, Kiemeney L, Kocher T, Launer LJ, Lehtimaki T, Melander O, Mosley TH, Jr., Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tonjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Gronberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Mark LG, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Adrienne CL, Erdmann J, Fox CS, Gudnason V, Gyllensten U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838

Luna A, Nicodemus KK (2007) snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics 23:774–776

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461:747–753

Massague J (2000) How cells read TGF-beta signals. Nat Rev Mol Cell Biol 1:169–178

McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njolstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubinski J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, Gonzalez CA, Ramon QJ, Martinez C, Navarro C, Ardanaz E, Larranaga N, Kham KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjonneland A, Kumle M, Riboli E, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P (2008) Lung cancer susceptibility locus at 5p15.33. Nat Genet 40:1404–1406

Miki D, Kubo M, Takahashi A, Yoon KA, Kim J, Lee GK, Zo JI, Lee JS, Hosono N, Morizono T, Tsunoda T, Kamatani N, Chayama K, Takahashi T, Inazawa J, Nakamura Y, Daigo Y (2010) Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. Nat Genet 42:893–896

Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J (2005) The choice of a genetic model in the meta-analysis of molecular association studies. Int J Epidemiol 34:1319–1328

National Cancer Institute (2009) Cancer Genetic Markers of Susceptibility (CGEMS) data website. http://cgems.cancer.gov/data_access.html

Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42:570–575

Pe'er I, Yelensk R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32:381–385

Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, Arslan AA, Bueno-de-Mesquita HB, Gallinger S, Gross M, Helzlsouer K, Holly EA, Jacobs EJ, Klein AP, Lacroix A, Li D, Mandelson MT, Olson SH, Risch HA, Zheng W, Albanes D, Bamlet WR, Berg CD, Boutron-Ruault MC, Buring JE, Bracci PM, Canzian F, Clipp S, Cotterchio M, de Andrade M, Duell EJ, Gaziano JM, Giovannucci EL, Goggins M, Hallmans G, Hankinson SE, Hassan M, Howard B, Hunter DJ, Hutchinson A, Jenab M, Kaaks R, Kooperberg C, Krogh V, Kurtz RC, Lynch SM, McWilliams RR, Mendelsohn JB, Michaud DS, Parikh H, Patel AV, Peeters PH, Rajkovic A, Riboli E, Rodriguez L, Seminara D, Shu XO, Thomas G, Tjonneland A, Tobias GS, Trichopoulos D, Van den Eeden SK, Virtamo J, Wactawski-Wende J, Wang Z, Wolpin BM, Yu H, Yu K, Zeleniuch-Jacquotte A, Fraumeni JF, Jr, Hoover RN, Hartge P, Chanock SJ (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet 42:224–228

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26:2336–2337

Purcell (2011) PLINK. http://pngu.mgh.harvard.edu/purcell/plink/

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, Jakobsdottir M, Helgadottir H, Thorlacius S, Aben KK, Blondal T, Thorgeirsson TE, Thorleifsson G, Kristjansson K, Thorisdottir K, Ragnarsson R, Sigurgeirsson B, Skuladottir H, Gudbjartsson T, Isaksson HJ, Einarsson GV, Benediktsdottir KR, Agnarsson BA, Olafsson K, Salvarsdottir A, Bjarnason H, Asgeirsdottir M, Kristinsson KT, Matthiasdottir S, Sveinsdottir SG, Polidoro S, Hoiom V, Botella-Estrada R, Hemminki K, Rudnai P, Bishop DT, Campagna M, Kellen E, Zeegers MP, de Verdier P, Ferrer A, Isla D, Vidal MJ, Andres R, Saez B, Juberias P, Banzo J, Navarrete S, Tres A, Kan D, Lindblom A, Gurzau E, Koppova K, de Vegt F, Schalken JA, van der Heijden HF, Smit HJ, Termeer RA, Oosterwijk E, van Hooij O, Nagore E, Porru S, Steineck G, Hansson J, Buntinx F, Catalona WJ, Matullo G, Vineis P, Kiltie AE, Mayordomo JI, Kumar R, Kiemeney LA, Frigge ML, Jonsson T, Saemundsson H, Barkardottir RB, Jonsson E, Jonsson S, Olafsson JH, Gulcher JR, Masson G, Gudbjartsson DF, Kong A, Thorsteinsdottir U, Stefansson K (2009) Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet 41:221–227

Ries LAG, Melbert D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Horner MJ, Howlader N, Eisner MP, Reichman M, Edwards BK (2007) SEER cancer statistics review, 1975–2004. National Cancer Institute, Bethesda, MD

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, Raney BJ, Wang T, Hinrichs AS, Zweig AS, Fujita PA, Learned K, Rhead B, Smith KE, Kuhn RM, Karolchik D, Haussler D, Kent WJ (2010) ENCODE whole-genome data in the UCSC Genome Browser. Nucleic Acids Res 38:D620–D625

Schreibman IR, Baker M, Amos C, McGarrity TJ (2005) The hamartomatous polyposis syndromes: a clinical and molecular review. Am J Gastroenterol 100:476–490

Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, Simon M, Marie Y, Boisselier B, Delattre JY, Hoang-Xuan K, El HS, Idbaih A, Zelenika D, Andersson U, Henriksson R, Bergenheim AT, Feychting M, Lonn S, Ahlbom A, Schramm J, Linnebank M, Hemminki K, Kumar R, Hepworth SJ, Price A, Armstrong G, Liu Y, Gu X, Yu R, Lau C, Schoemaker M, Muir K, Swerdlow A, Lathrop M, Bondy M, Houlston RS (2009) Genome-wide association study identifies five susceptibility loci for glioma. Nat Genet 41:899–904

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034–1050

Stacey SN, Sulem P, Masson G, Gudjonsson SA, Thorleifsson G, Jakobsdottir M, Sigurdsson A, Gudbjartsson DF, Sigurgeirsson B, Benediktsdottir KR, Thorisdottir K, Ragnarsson R, Scherer D, Hemminki K, Rudnai P, Gurzau E, Koppova K, Botella-Estrada R, Soriano V, Juberias P, Saez B, Gilaberte Y, Fuentelsaz V, Corredera C, Grasa M, Hoiom V, Lindblom A, Bonenkamp JJ, van Rossum MM, Aben KK, de Vries E, Santinami M, Di Mauro MG, Maurichi A, Wendt J, Hochleitner P, Pehamberger H, Gudmundsson J, Magnusdottir DN, Gretarsdottir S, Holm H, Steinthorsdottir V, Frigge ML, Blondal T, Saemundsdottir J, Bjarnason H, Kristjansson K, Bjornsdottir G, Okamoto I, Rivoltini L, Rodolfo M, Kiemeney LA, Hansson J, Nagore E, Mayordomo JI, Kumar R, Karagas MR, Nelson HH, Gulcher JR, Rafnar T, Thorsteinsdottir U, Olafsson JH, Kong A, Stefansson K (2009) New common variants affecting susceptibility to basal cell carcinoma. Nat Genet 41:909–914

StataCorp (2009) Stata statistical software: release 11. StataCorp LP, College Station, TX

Tenesa A, Dunlop MG (2009) New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet 10:353–358

Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IP, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CM, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 40:631–637

The Institute of Cancer Research (2008) Software and databases, molecular and population genetics team, section of cancer genetics phase 1 GWAS data

Timm DE, Liu J, Baker LJ, Harris RA (2001) Crystal structure of thiamin pyrophosphokinase. J Mol Biol 310:195–204

Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB, Houlston R (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39:984–988

Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijayakrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, Gray R, Maher ER, Lucassen A, Kerr D, Evans DG, Schafmayer C, Buch S, Volzke H, Hampe J, Schreiber S, John U, Koessler T, Pharoah P, van Wezel T, Morreau H, Wijnen JT, Hopper JL, Southey MC, Giles GG, Severi G, Castellvi-Bel S, Ruiz-Ponte C, Carracedo A, Castells A, Forsti A, Hemminki K, Vodicka P, Naccarati A, Lipton L, Ho JW, Cheng KK, Sham PC, Luk J, Agundez JA, Ladero JM, de la Hoya M, Caldes T, Niittymaki I, Tuupanen S, Karhu A, Aaltonen L, Cazier JB, Campbell H, Dunlop MG, Houlston RS (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 40:623–630

Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, Ricketts M, Linger R, Nsengimana J, Deloukas P, Huddart RA, Bishop DT, Easton DF, Stratton MR, Rahman N (2010) Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet 42:604–607

Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet 40:1407–1409

Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

Witte JS (2007) Multiple prostate cancer risk variants on 8q24. Nat Genet 39:579–580

Wu KJ, Grandori C, Amacker M, Simon-Vermot N, Polack A, Lingner J, Dalla-Favera R (1999) Direct activation of TERT transcription by c-MYC. Nat Genet 21:220–224

Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y (2001) A novel gene, CRR9, which was up-regulated in CDDP-resistant

ovarian tumor cell line, was associated with apoptosis. Biochem Biophys Res Commun 280:1148–1154

Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hunter DJ, Chanock SJ, Thomas G (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39:645–649

Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S, O'shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, Blanche H, Sahbatou M, Tubacher E, Bonaiti-Pellie C, Buecher B, Riboli E, Kury S, Chanock SJ, Potter J, Thomas G, Gallinger S, Hudson TJ, Dunlop MG (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 39:989–994

Zeggini E, Ioannidis JP (2009) Meta-analysis in genome-wide association studies. Pharmacogenomics 10:191–201