

Supplementary figure 1. Distributions of age of diagnosis by primary tumor anatomic subsite and sex. Note that our data recapitulate the previously reported higher percentage of female proximal colon cancer cases, a male-to-female ratio that increases progressively from the proximal colon to the rectum, and differences in age of onset by primary tumor site, with an earlier age of onset for rectal cancer.



Supplementary figure 2. Quantile-quantile (QQ) plots stratified by minor allele frequency (MAF) bins for the five GWAS meta-analyses of CRC case subgroups defined by primary tumor anatomic subsite. GWAS studies were imputed to the Haplotype Reference Consortium (HRC) panel. The red dashed line indicates the genome-wide significance threshold ($P=5\times10^{-8}$). The transparent regions around the equality line represent the analytically estimated 95% confidence bands for each MAF bin.



Rectal cancer GWAS: up to 16,212 cases and 64,159 controls

Colon cancer GWAS: up to 32,002 cases and 64,159 controls



Supplementary figure 2. (continued)



Left-sided colorectal cancer GWAS: up to 30,588 cases and 64,159 controls

Supplementary figure 2. (continued)

Proximal colon cancer GWAS: up to 15,706 cases and 64,159 controls 11,437,803 variants



Supplementary figure 3. Manhattan plots showing results of the five GWAS meta-analyses of CRC case subgroups defined by primary tumor anatomic subsite. GWAS studies were imputed to the Haplotype Reference Consortium (HRC) panel. Association results for each variant ($-\log 10 P$ values) are plotted against genomic position (NCBI Build 37). The red dashed line indicates the genome-wide significance threshold (P=5×10-8). New loci are shown in red in the Manhattan plot. Loci previously associated with overall colorectal cancer risk at genome-wide significance are denoted in blue. (figure continued on following pages)

Distal colon cancer GWAS: up to 14,376 cases and 64,159 controls 11,353,757 variants



Supplementary figure 3. (continued)

Rectal cancer GWAS: up to 16,212 cases and 64,159 controls 11,358,342 variants



Supplementary figure 3. (continued)

Colon cancer GWAS: up to 32,002 cases and 64,159 controls 11,928,745 variants



Supplementary figure 3. (continued)

Left-sided colorectal cancer GWAS: up to 30,588 cases and 64,159 controls 11,819,887 variants



Supplementary figure 3. (continued)

Supplementary figure 4. Regional association plots for the new CRC risk loci reaching genome-wide significance (*P*-value $< 5 \times 10^{-8}$) in the GWAS meta-analyses for CRC case subgroups defined by primary tumor anatomical subsites. Case subgroups were defined as follows: proximal colon cancer (*n*=15,706), distal colon cancer (*n*=14,376), rectal cancer (*n*=16,212), colon cancer (*n*=32,002), and distal/left-sided CRC (*n*=30,588). Analyses were based on 64,159 shared controls. LocusZoom plots show the – $\log_{10}(P$ -value) for the association with risk for the CRC case subgroup as a function of genomic position (NCBI Build 37) for each variant within a 1-Mb window centered at the lead variant of the locus. Lead variants are indicated by the purple diamond symbol. The color labeling of other variants indicates LD with the lead variant estimated from our previously published whole-genome sequence (WGS) data on 2,159 European ancestry study participants (Huyghe *et al.*). Gray dots indicate that the variant was not found in our WGS panel and that LD could not be calculated. Recombination rates are based on Phase 2 HapMap and gene models are RefSeq genes taken from the UCSC Genome Browser.





























Locus, SNP		Best model (BIC)	\boldsymbol{P}_{het}
1p36.12, rs72647484		S3_PDR	7.4E-01
1p34.3, rs4360494		S3_PDR	4.7E-01
1p32.3, rs12144319		S3_PDR	2.1E-01
1q25.3, rs6678517		S3_PDR	1.9E-03
1q41, rs17011141		S3_PDR	6.8E-03
2q24.2, rs448513		S3_PDR	8.8E-01
2q33.1, rs11884596		S3_PDR	3.8E-01
2q33.1, rs983402		S1_R	2.0E-05
2q35, rs3731861		S3_PDR	6.8E-01
3p22.1, rs35470271		S2_DR	9.8E-08
3p14.1, rs6781752		S3_PDR	3.8E-01
3q13.2, rs72942485		S2_DR	1.1E-02
3q22.2, rs10049390		S3_PDR, S2_PD	1.3E-01
3q26.2, rs9876206		S3_PDR	9.0E-01
4q22.2, rs13149359		S3_PDR	4.6E-01
4q24, rs1391441		S3_PDR, S2_DR	4.9E-02
4q31.21, rs11727676		S3_PDR	6.1E-01
5p15.33, rs78368589		S3_PDR	1.0E-01
5p15.33, rs2735940		S3_PDR	2.9E-01
5p13.1, rs7708610		S3_PDR	5.6E-01
5p13.1, rs12514517		S3_PDR	1.0E-02
5q21.1, rs145364999		S3_PDR	7.1E-01
	0.90 1.2 1 Odd	.6 2.0 3.0 s ratio	

Supplementary figure 5. Forest plots and multinomial modeling results for previously reported CRC risk variants. Best model is the best-fitting multinomial logistic regression model according to the Bayesian Information Criterion (BIC). Please refer to supplementary table 2 for model definitions. P_{het} is the *P*-value from a heterogeneity test, testing the null hypothesis that odds ratios are fixed across CRC subtypes defined by primary tumor site.



Locus, SNP			Best model (BIC)	$m{P}_{het}$
5q22.2, rs755229494			S2_DR	2.3E-10
5q31.1, rs4976270			S3_PDR	2.2E-03
6p21.33, rs2516420			S2_PD	2.2E-03
6p21.32, rs9271695			S3_PDR	5.1E-03
6p21.31, rs16878812			S3_PDR	2.9E-01
6p21.2, rs9470361			S2_DR	1.6E-04
6p21.1, rs62396735			S2_DR	8.0E-02
6p12.1, rs62404966			S3_PDR	2.1E-01
7p13, rs12672022			S3_PDR	5.4E-01
8q23.3, rs16892766			S3_PDR	2.3E-01
8q23.3, rs6469654			S2_PR, S3_PDR	4.4E-01
8q24.11, rs117079142			S3_PDR	9.6E-01
8q24.21, rs6983267			S3_PDR	8.9E-01
8q24.21, rs4313119			S3_PDR	8.6E-01
9p21.3, rs1537372			S2_DR	6.2E-03
9q22.33, rs34405347			S2_DR, S3_PDR	6.5E-02
9q31.3, rs10980628			S2_DR	2.6E-02
10p14, rs11255841			S2_DR	1.7E-08
10q11.23, rs10821907			S3_PDR	4.6E-01
10q22.3, rs704017			S3_PDR	5.8E-03
10q24.2, rs11190164			S3_PDR	2.8E-01
10q25.2, rs12246635			S3_PDR, S2_DR	5.2E-02
C).90 1.2 1 Odd:	.6 2.0 3.0 s ratio		



Locus, SNP	Best model (BIC)	$m{P}_{het}$
10q25.2, rs11196170	S3_PDR	3.1E-01
11q12.2, rs174533	S3_PDR	6.9E-02
11q13.4, rs7121958	S2_DR, S3_PDR	8.4E-04
11q13.4, rs61389091	S2_DR, S3_PDR	5.3E-04
11q22.1, rs2186607	S3_PDR	2.2E-01
11q23.1, rs3087967	S2_DR	6.0E-14
12p13.32, rs35808169	S3_PDR	2.3E-01
12p13.32, rs3217810	S3_PDR	8.9E-03
12p13.32, rs3217874	S3_PDR	1.4E-01
12p13.31, rs2250430	S3_PDR	8.2E-01
12p13.2, rs2710310	S2_PR	2.4E-01
12q12, rs11610543	S2_PD, S3_PDR	2.2E-03
12q13.12, rs12372718	S3_PDR	6.2E-03
12q13.3, rs4759277	S3_PDR	3.8E-01
12q24.12, rs597808	S3_PDR	4.2E-02
12q24.21, rs7300312	S3_PDR	2.1E-01
12q24.22, rs55990915	S2_PD	1.8E-02
13q13.2, rs377429877	S3_PDR, S2_PD	1.1E-01
13q13.3, rs7333607	S3_PDR	1.5E-01
13q22.1, rs78341008	 S3_PDR	6.6E-01
13q34, rs8000189	S3_PDR	5.6E-01
14q22.2, rs35107139	S3_PDR	6.0E-01

0.90 1.0 1.1 1.2 1.3 1.45 Odds ratio

Locus, SNP		Best model (BIC)	$oldsymbol{P}_{het}$
14q22.2, rs4901473		S3_PDR	9.8E-02
14q23.1, rs17094983		S3_PDR, S2_PD	3.4E-02
15q13.3, rs12708491		S3_PDR	7.0E-02
15q13.3, rs2293581		S3_PDR	6.1E-05
15q13.3, rs17816465		S2_PR	1.9E-02
15q22.33, rs56324967		S3_PDR	1.7E-01
16q22.1, rs9924886		S3_PDR	7.3E-01
16q23.2, rs9930005		S1_D, S3_PDR	5.6E-03
16q24.1, rs12149163	⊢ ⊢ ⊢	S2_DR	2.0E-03
16q24.1, rs62042090		S2_DR	6.1E-03
17p13.3, rs4968127		S3_PDR	9.9E-02
17p12, rs1078643		S3_PDR	6.2E-01
17q24.3, rs983318		S2_PD	1.1E-04
17q25.3, rs75954926		S2_DR	3.6E-06
18q21.1, rs11874392		S3_PDR_HET, S3_PDR	8.1E-06
19p13.11, rs34797592		S3_PDR	3.3E-01
19q13.11, rs28840750		S3_PDR	2.7E-03
19q13.2, rs1963413		S3_PDR	6.7E-01
19q13.43, rs73068325		S3_PDR	2.4E-01
20p12.3, rs189583		S3_PDR	6.2E-01
20p12.3, rs994308		S3_PDR	3.0E-01
20p12.3, rs4813802		S3_PDR	2.3E-01
_			

0.90 1.0 1.1 1.2 1.3 1.45 Odds ratio

