

## **Supplementary Information for:**

# **Large-scale genetic study in East Asians identifies six novel loci associated with colorectal cancer risk**

Ben Zhang<sup>1</sup>, Wei-Hua Jia<sup>2</sup>, Koichi Matsuda<sup>3</sup>, Sun-Seog Kweon<sup>4,5</sup>, Keitaro Matsuo<sup>6</sup>, Yong-Bing Xiang<sup>7</sup>, Aesun Shin<sup>8,9</sup>, Sun Ha Jee<sup>10</sup>, Dong-Hyun Kim<sup>11</sup>, Qiuyin Cai<sup>1</sup>, Jirong Long<sup>1</sup>, Jiajun Shi<sup>1</sup>, Wanqing Wen<sup>1</sup>, Gong Yang<sup>1</sup>, Yanfeng Zhang<sup>1</sup>, Chun Li<sup>12</sup>, Bingshan Li<sup>13</sup>, Yan Guo<sup>14</sup>, Zefang Ren<sup>15</sup>, Bu-Tian Ji<sup>16</sup>, Zhi-Zhong Pan<sup>2</sup>, Atsushi Takahashi<sup>17</sup>, Min-Ho Shin<sup>4</sup>, Fumihiko Matsuda<sup>18</sup>, Yu-Tang Gao<sup>7</sup>, Jae Hwan Oh<sup>19</sup>, Soriul Kim<sup>10</sup>, Yoon-Ok Ahn<sup>9</sup>, Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)<sup>20</sup>, Andrew T. Chan<sup>21,22</sup>, Jenny Chang-Claude<sup>23</sup>, Martha L. Slattery<sup>24</sup>, Colorectal Transdisciplinary (CORECT) Study<sup>20</sup>, Stephen B. Gruber<sup>25</sup>, Fredrick R. Schumacher<sup>25</sup>, Stephanie L. Stenzel<sup>25</sup>, Colon Cancer Family Registry (CCFR)<sup>20</sup>, Graham Casey<sup>25</sup>, Hyeong-Rok Kim<sup>26</sup>, Jin-Young Jeong<sup>11</sup>, Ji Won Park<sup>19,27</sup>, Hong-Lan Li<sup>7</sup>, Satoyo Hosono<sup>6</sup>, Sang-Hee Cho<sup>28</sup>, Michiaki Kubo<sup>17</sup>, Xiao-Ou Shu<sup>1</sup>, Yi-Xin Zeng<sup>2</sup>, and Wei Zheng<sup>1</sup>

### **Corresponding author contact information:**

Wei Zheng, M.D., Ph.D.

Vanderbilt Epidemiology Center

Vanderbilt University School of Medicine

2525 West End Avenue, 8<sup>th</sup> Floor, Nashville, TN 37203-1738

Phone: (615) 936-0682; Fax: (615) 936-8241

E-mail: [wei.zheng@vanderbilt.edu](mailto:wei.zheng@vanderbilt.edu)

### **Affiliations**

<sup>1</sup>Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee, the United States.

<sup>2</sup>State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China.

<sup>3</sup>Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

<sup>4</sup>Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, South Korea.

<sup>5</sup>Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, South Korea.

<sup>6</sup>Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan.

<sup>7</sup>Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China.

<sup>8</sup>Molecular Epidemiology Branch, National Cancer Center, Goyang-si, South Korea.

<sup>9</sup>Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, South Korea.

<sup>10</sup>Institute for Health Promotion, Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, South Korea.

<sup>11</sup>Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, South Korea.

<sup>12</sup>Department of Biostatistics, Vanderbilt University School of Medicine, Tennessee, the United States.

<sup>13</sup>Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Tennessee, the United States.

<sup>14</sup>Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, Tennessee, the United States.

<sup>15</sup>School of Public Health, Sun Yat-sen University, Guangzhou, China.

<sup>16</sup>Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, Maryland, the United States.

<sup>17</sup>Center for Integrative Medical Sciences, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan.

<sup>18</sup>Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

<sup>19</sup>Center for Colorectal Cancer, National Cancer Center, Goyang-si, South Korea.

<sup>20</sup>A complete list of members is provided in the Acknowledgements.

<sup>21</sup>Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, the United States.

<sup>22</sup>Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, the United States.

<sup>23</sup>Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany.

<sup>24</sup>Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah, the United States.

<sup>25</sup>USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, the United States.

<sup>26</sup>Department of Surgery, Chonnam National University Medical School, Gwangju, South Korea.

<sup>27</sup>Department of Surgery, Seoul National University Hospital, Seoul, South Korea

<sup>28</sup>Department of Hemato-oncology, Chonnam National University Medical School, Gwangju, South Korea.

## Supplementary Information

<b>Content</b>	<b>Page</b>
Supplementary Note	4
Study participants	4
Laborotory procedures	7
Imputation	11
Evaluation of population structure	12
GECCO	13
CORECT	18
References	21
Supplementary Tables	24
Supplementary Table 1	26
Supplementary Table 2	27
Supplementary Table 3	28
Supplementary Table 4	29
Supplementary Table 5	33
Supplementary Table 6	34
Supplementary Table 7	35
Supplementary Table 8	36
Supplementary Table 9	37
Supplementary Table 10	38
Supplementary Table 11	39
Supplementary Table 12	40
Supplementary Table 13	41
Supplementary Table 14	42
Supplementary Table 15	48
Supplementary Table 16	49
Supplementary Table 17	50
Supplementary Table 18	51
Supplementary Table 19	52
Supplementary Table 20	53
Supplementary Figures	54
Supplementary Figure 1	56
Supplementary Figure 2	63
Supplementary Figure 3	72
Supplementary Figure 4	74
Supplementary Figure 5	81
Supplementary Figure 6	94

## SUPPLEMENTARY NOTE

### Study participants

This four-stage genome-wide association study (GWAS) consisted of 14 case-control studies conducted as part of the Asia Colorectal Cancer Consortium. Included in the final analyses were 14,963 colorectal cancer (CRC) cases and 31,945 controls recruited from eight centers located in China, Korea and Japan (**Supplementary Table 1**). All participants were of self-reported East Asian descent. The protocols of all participating studies were approved by the relevant review boards at the respective institutions and informed consents were obtained from all study participants. We summarize the participating studies in each center below.

**The Shanghai CRC Study:** This study included 2,368 cases and 8,219 controls from three case-control collections, Shanghai-1, Shanghai-2 and Shanghai-3. Details of Shanghai-1 and Shanghai-2 have been described elsewhere<sup>1</sup>. Briefly, 728 cases and 728 matched controls in these two studies were identified from among participants of the Shanghai Women's Health Study (SWHS)<sup>2</sup> and the Shanghai Men's Health Study (SMHS)<sup>3</sup>, two population-based cohort studies conducted in Shanghai, China, with a combined sample size of 136,631 participants. In addition to 474 cases and 497 controls scanned using Affymetrix arrays, Shanghai-1 also included 2,131 controls scanned using the same genotyping platform from an ongoing GWAS of breast cancer<sup>4</sup>. Shanghai-2 initially included 254 cases and 231 controls scanned using Illumina arrays. To increase statistical power, we expanded Shanghai-2 to include 423 additional controls genotyped using Illumina arrays from other studies<sup>5,6</sup>. These 4,010 participants (728 cases and 3,282 controls) were included in stage 1. In Shanghai-3, cases were newly recruited in Shanghai between January 2009 and February 2011. A total of 2,160 eligible patients with CRC were

identified through the population-based Shanghai Cancer Registry. Saliva samples were obtained using Oragene kits for 1,945 cases, 1,640 of whom are analyzed in this study. Controls in Shanghai-3 were cancer-free men and women identified from among SWHS and SMHS participants ( $n = 1,277$ ). We added 3,660 female controls to this study from an ongoing breast cancer project genotyped using the same platform. Samples in Shanghai-3 were included in stage 2.

**The Guangzhou CRC Study:** This study consisted of four case-control collections, Guangzhou-1, Guangzhou-2, Guangzhou-3 and Guangzhou-4. Cases were histopathologically confirmed patients with CRC recruited from Sun Yat-Sen University Cancer Center between January 2002 and January 2012 and controls were cancer-free men and women recruited from the physical examination centers of several large hospitals in Guangdong during the same time period. Details of the methods employed for case and control recruitment have been described in previous publications <sup>1, 7</sup>. The four Guangzhou studies contributed a total of 3,197 cases and 3,424 controls to this project. Participants in Guangzhou-1 ( $n = 1,603$ ) were included in stage 1, Guangzhou-2 ( $n = 809$ ) and Guangzhou-3 ( $n = 2,408$ ) in stage 2 and Guangzhou-4 ( $n = 1,791$ ) in stage 4.

**The Aichi CRC Study:** This study had two case-control collections (Aichi-1 and Aichi-2) with a total of 640 cases and 1,414 controls <sup>1</sup>. We included Aichi-1 in stage 1 ( $n = 1,346$ ) and Aichi-2 in stage 4 ( $n = 708$ ). In Aichi-1, cases were identified among 28,766 participants of the Hospital-based Epidemiologic Research Program at Aichi Cancer Center, Japan <sup>8</sup> and controls were cancer-free individuals recruited from this population <sup>9</sup>. Cases and frequency-matched controls in Aichi-2 were recently recruited from the same study site.

**The Korean Cancer Prevention Study-II CRC:** This study included 325 CRC cases and 976 cancer-free controls randomly selected from the Korean Cancer Prevention Study-II (KCPS-II), a cohort study with 266,258 individuals, aged 20-77 years, who visited 16 health promotion centers nationwide from April 2004 to December 2008 in South Korea <sup>10</sup>. Details of this study have been described elsewhere <sup>1, 10</sup>. KCPS-II samples were included in stage 1.

**The BioBank Japan CRC Study:** This study (BBJ) is an expansion of a previous Japanese GWAS of CRC, about which details of the study methodology have been previously described <sup>11</sup>. BBJ contributed 2,814 cases and 11,358 controls to this project and was included in stage 3. DNA samples of both CRC cases and cancer-free controls were obtained from the BioBank Japan of the Personalized Medicine Project (<http://biobankjp.org/>) supported by the Ministry of Education, Culture, Sports, Science and Technology, Japan. In addition to healthy individuals as controls, we also included participants with diabetes, myocardial infarction, brain infarction, arteriosclerosis obliterans, atrial fibrillation, cholangiocarcinoma, drug eruption, liver cirrhosis, amyotrophic lateral sclerosis and rheumatoid arthritis as controls to increase the statistical power.

**The Korean-National Cancer Center CRC Study:** This study (Korea-NCC) is a hospital-based case-control study of CRC conducted in South Korea. Cases were histologically confirmed patients with CRC who received surgery between 2000 and 2004 at the Korean National Cancer Center (NCC). A total of 1,392 patients participated in this study and provided a blood sample. Controls (n = 1,329), frequency-matched to cases by age and sex, were selected from participants of the Cancer Screening Cohort of the NCC recruited between August 2002 and December 2004. All cases and controls in this study were included in stage 4.

**The Seoul CRC Study:** This study (Korea-Seoul) is a multicenter case-control study conducted in South Korea <sup>12</sup>. Cases were CRC patients who were admitted to two university hospitals and one general cancer hospital in the Seoul Metropolitan Area between 1995 and 2004 <sup>12</sup>. Controls were selected from the same hospitals during the same time period from among a wide spectrum of inpatients with non-neoplastic conditions. A total of 849 cases and 673 controls providing a blood sample were included in stage 4 of this project.

**The Hwasun Cancer Epidemiology Study-Colon and Rectum Cancer (HCES-CRC):** The HCES is a hospital-based case-control study conducted in South Korea that includes multiple cancers <sup>13, 14</sup>. The HCES-CRC consisted of 3,378 cases and 4,552 controls who were included in stage 4 of this project. Cases were newly diagnosed patients with CRC at Chonnam National University Hwasun Hospital, a cancer-specific hospital in Jeollanam-do province, South Korea, between April 2004 and February 2013. Controls (n=4,552) were participants in the Korean Community Health Survey, an annual nationwide health interview survey, conducted from 2010 to 2012 in the Jindo and Bosung counties of Jeollanam-do province, Korea. Excluded from this study were patients with secondary or recurrent cancer. Demographic data of participants were collected through in-person interviews and review of medical records.

### **Laboratory procedures**

Methods of genotyping, genotype calling and quality control in stage 1 have been described elsewhere <sup>1, 4, 6, 7, 9, 10, 15</sup>. Briefly, genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix Inc., Santa Clara, CA, United States) for Shanghai-1 cases and controls; the Illumina HumanOmniExpress BeadChip (Illumina Inc., San Diego, CA, United States) for Shanghai-2 cases and controls, Guangzhou-1 cases and Aichi-1 cases; the Illumina

Infinium HumanHap550 BeadChip (Illumina Inc., San Diego, CA, United States) and the Illumina 660W-Quad BeadChip (Illumina Inc., San Diego, CA, United States) for Shanghai-2 controls; the Illumina Human610-Quad BeadChip (Illumina Inc., San Diego, CA, United States) for Guangzhou-1 controls; the Illumina Infinium HumanHap610 BeadChip (Illumina Inc., San Diego, CA, United States) for Aichi-1 controls and the Affymetrix Genome-Wide Human SNP Array 5.0 (Affymetrix Inc., Santa Clara, CA, United States) for KCPS-II cases and controls (**Supplementary Table 2**). Genotype calling was performed using the Birdseed algorithm v2 for Affymetrix arrays or GenomeStudio software for Illumina arrays based on the manufacturers' protocols. We only included shared SNPs in the analysis if samples in a study were genotyped on different platforms. A uniform quality control protocol was employed to exclude samples and SNPs from each of the stage 1 studies as described in the previous paper <sup>1</sup>, and samples or SNPs were excluded if they met any of the following criteria: (i) genotype call rate per sample <95%, (ii) genetically identical (i.e., PI\_HAT >0.9) or duplicate samples, (iii) sex determined using genotypes inconsistent with epidemiological or clinical data, (iv) first- or second-degree relatives (i.e., PI\_HAT >0.25), (v) ethnic outliers with population structure inconsistent with HapMap Asian samples (see **Evaluation of population structure**), (vi) genotype call rate per SNP <95%, (vii) minor allele frequency (MAF) <0.05, (viii) genotyping consistency rates <95% in quality control samples, (ix)  $P$  for Hardy-Weinberg equilibrium (HWE)  $<1\times10^{-5}$  in controls, or (x) SNPs not in autosomes. After these quality control filtering procedures, 502,145 SNPs for 3,102 individuals (474 cases and 2,628 controls) remained in the Shanghai-1 dataset; 245,961 SNPs for 908 individuals (254 cases and 654 controls) remained in the Shanghai-2 dataset; 250,612 SNPs for 1,613 participants (641 cases and 972 controls) remained in the Guangzhou-1 dataset; 232,426 SNPs for 1,346 participants (404 cases and 942 controls) remained in the Aichi-1

dataset; and 312,869 SNPs for 1,301 participants (325 cases and 976 controls) remained in KCPS-II dataset (**Supplementary Table 2**).

Genotyping in stage 2 was conducted for samples in Guangzhou-2 at the Genome Quebec Innovation Centre (Montreal, Canada) and for samples in Shanghai-3 and Guangzhou-3 at Genenergy (Shanghai, China) using Illumina Infinium assays as part of the customer add-on content for multiple projects to the Illumina HumanExome-12v1\_A Beadchip (Illumina Inc., San Diego, CA, United States). For clarity, we named this array The CRC GWAS replication component, and it included 7,113 SNPs. Details of SNP content and selection strategies for the Illumina HumanExome-12v1\_A Beadchip are described at [http://genome.sph.umich.edu/wiki/Exome\\_Chip\\_Design](http://genome.sph.umich.edu/wiki/Exome_Chip_Design). On each 96-well plate, two blind duplicate samples and two HapMap samples were included as positive quality controls. Genotype calling was conducted with the GenTrain clustering algorithm version 2.0 in GenomeStudio (version 2011.1). We calculated consistency rates for HapMap samples genotyped in this study and sequenced by the 1000 Genomes Project (<http://www.1000genomes.org/>). We also estimated pair-wise proportion of identity-by-descent (IBD) to identify potentially genetically identical, unexpected duplicate samples or close relatives. We further identified genetic outliers on the basis of principal components analysis (see **Evaluation of population structure**). Quality control filtering for samples and genetic markers was performed using PLINK version 1.0.7<sup>16</sup> (<http://pngu.mgh.harvard.edu/~purcell/plink/>) with a series of standard criteria. Specifically, samples were excluded if they met any of the following criteria: (i) genotype call rate <98%, (ii) genetically identical or duplicate samples, (iii) genetic sex inconsistent with epidemiological or clinical data, (iv) first- or second-degree relatives, (v) ethnic outliers, (vi) heterozygosity outliers or (vii) consistency rate <99% in duplicated samples.

Genetic markers were excluded if they met any of the following criteria: (i) MAF = 0, (ii) genotype call rate < 98%, (iii) consistency rate < 98% in QC samples, (iv)  $P$  for HWE <  $10^{-5}$  in controls, (v) duplicate SNPs, (vi) wrong SNPs or (vii) caution SNPs revealed by the Exome Chip Design group ([http://genome.sph.umich.edu/wiki/Exome\\_Chip\\_Design#Cautious\\_Sites](http://genome.sph.umich.edu/wiki/Exome_Chip_Design#Cautious_Sites)). After these quality control exclusions, we obtained a final dataset including 140,355 genetic markers genotyped on 3,519 cases and 6,275 controls. Of them, 6,899 SNPs were included in this project.

Details on genotyping and quality control procedures for stage 3 (BBJ) have been previously reported for the Japanese GWAS<sup>11</sup> and some procedures are updated herein. Briefly, both cases and controls in BBJ were genotyped using the Illumina HumanHap610-Quad BeadChip (Illumina Inc., San Diego, CA, United States). Samples were excluded if they met any of the following criteria: (i) genotype call rate per sample < 98%, (ii) first- or second-degree relatives, (iii) genetic sex inconsistent with epidemiological or clinical data or (iv) ethnic outliers with population structure inconsistent with HapMap Asian samples. SNPs were removed if they met any of the following criteria: (i) genotype call rate per SNP < 99%, (ii)  $P$  for HWE <  $10^{-6}$  in controls or (iii) MAF = 0. After sample and SNP exclusions, we obtained a final dataset including 2,814 cases and 11,358 controls with 460,463 SNPs.

Stage 4 genotyping for 29 SNPs was conducted with the iPLEX Sequenom MassARRAY platform (Sequenom, Inc., San Diego, CA, United States) at the Vanderbilt Molecular Epidemiology Laboratory (Nashville, TN, United States). Polymerase chain reaction (PCR) and extension primers were designed using MassARRAY Assay Design 4.0 software (Sequenom, Inc., San Diego, CA, United States). PCR and extension reactions were performed according to the manufacturer's instructions, and extension product sizes were determined by mass spectrometry using the Sequenom MassARRAY Analyzer 4. Four negative controls (water) and

eight positive quality controls (HapMap or duplicate samples) were included in each 384-well plate. We filtered out SNPs if they met any of the following criteria: (i) genotype call rate <95%, (ii) genotyping consistency rate <95% in positive control samples, (iii) unclear genotype call or (iv)  $P$  for HWE < $10^{-5}$  in controls. The number of SNPs in the final analysis was 28, 27, 28, 27 and 28 in Guangzhou-4, Aichi-2, Korea-NCC, Korea-Seoul and HCES-CRC, respectively. The mean genotype call rate of these SNPs was 98.5% with a median value of 99.1% across all five studies. The average concordance rate of data calculated from positive quality control samples was 99.9% with a median value of 100% for each of the five studies included in this stage.

## Imputation

For genome-wide imputation in stage 1, we used HapMap 2 CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan) panel ( $n = 2,416,663$ ) as references. Imputation was conducted separately for each of the five studies (Shanghai-1, Shanghai-2, Guangzhou-1, Aichi-1 and KCPS-II) using the MACH v1.0 program<sup>17</sup> (<http://www.sph.umich.edu/csg/abecasis/MACH/>). Dosage data were generated from imputation to represent the expected number of copies of the effect allele (a value between 0 and 2) for each SNP. For each study, we obtained approximately 2.4 million genotyped or imputed autosomal SNPs. Of them, we only included SNPs with MAF >0.05 and high imputation quality (RSQ >0.5) in the final analysis. Accordingly, the number of SNPs in each study was 1,917,071, 1,882,559, 1,875,120, 1,883,544, and 1,792,333, respectively. To evaluate the imputation quality, we masked a subset of SNPs with MAF >0.05 ( $n = 2,000$ ) during imputation and compared imputation data with those obtained from direct genotyping. The mean concordance rate was 98.8%. To further evaluate the imputation quality for the ten risk variants newly identified in our study, we directly genotyped them in approximately 2,800 samples included in stage 1. The

concordance between imputed and genotyped data was very high, with mean values ranging from 96.00% to 99.96% for the ten SNPs (**Supplementary Table 20**). Imputation of ungenotyped SNPs in samples included in BBJ was conducted with phased data of JPT/CHS/CHD subjects from the 1000 Genomes Project phase1 v3 (16-March-2012 release) as the reference using MACH v1.0<sup>17</sup> (<http://www.sph.umich.edu/csg/abecasis/MACH/>) and minimac<sup>18</sup> (<http://genome.sph.umich.edu/wiki/Minimac>). Before imputation was performed, we removed SNPs from the reference panel if they met any of the following criteria: (i) MAF <0.01, (ii) *P* for HWE <10<sup>-5</sup> and (iii) large allele frequency difference between BBJ and the reference panel (i.e., 0.16). After imputation, we obtained approximately 7.5 million genotyped or imputed autosomal SNPs. Of them, 6,089,958 are common SNPs (MAF >0.05) with high imputation quality (RSQ >0.5). Regional imputation of genotype data downloaded from the Cancer Genome Atlas<sup>19</sup> (<http://cancergenome.nih.gov/>) was also performed using MACH v1.0<sup>17</sup> (<http://www.sph.umich.edu/csg/abecasis/MACH/>) and minimac<sup>18</sup> (<http://genome.sph.umich.edu/wiki/Minimac>). A 1-mb region centered on the index SNP in each of the newly-identified loci was imputed using the GIANT ALL panel from the 1000 Genomes Project phase1 release v3 (16-March-2012 release) as the reference. We then called dosage data of the SNPs (RSQ >0.8) into genotypes after imputation.

### Evaluation of population structure

We evaluated population structure in studies included in stages 1 and 2 using principal components analysis with EIGENSTRAT software<sup>20</sup> (<http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm>). In stage 1, principal components analysis was carried out using a set of approximately 6,000 independent SNPs that met the following criteria: (i) a neighboring distance >200 kb, (ii) MAF >0.2, (iii)  $r^2$  <0.1 and (iv)

genotype call rate >99%, following the approach described in our previous GWAS<sup>1</sup>. Consistent with our previous reports<sup>1, 4, 6, 7, 9, 10, 15</sup>, no ethnic outliers (defined as 6 s.d. away from the means of principal components 1 and 2) were identified, and all 8,270 samples in stage 1 were clearly identified as being of East Asian ancestry (**Supplementary Fig. 1**). We estimated the genomic inflation factor  $\lambda$  using genome-wide SNPs with high imputation quality (RSQ >0.5). When the first ten principal components were included in the regression models, the  $\lambda$  was <1.04 in each of the five studies included in stage 1 and 1.0368 in the meta-analysis of all five studies (**Supplementary Fig. 2**). In stage 2, principal components analysis was performed on the basis of genotype data from approximately 2,500 ancestry information markers (AIMs) on the Asian Exome Array. Again, no ethnic outliers were identified. We estimated the inflation factor using 13,363 uncorrelated SNPs (including AIMs) with a MAF >0.05 on the Asian Exome Array but not included in the association analysis for CRC. The  $\lambda$  was <1.05 in each of the three studies included in stage 2 and 1.0525 in the meta-analysis of all studies (**Supplementary Fig. 3**). We also rescaled the inflation statistic to an equivalent value of a study with 1,000 cases and 1,000 controls ( $\lambda_{1000}$ ) using the formula:  $\lambda_{1000} = 1 + 500 \times (\lambda - 1) \times (1/N_{\text{cases}} + 1/N_{\text{controls}})$ <sup>21</sup>. Accordingly,  $\lambda_{1000}$  was 1.01 in both stage 1 and stage 2, indicating that little population substructure was present in our studies.

### The Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)

The GECCO GWAS meta-analysis included 9,421 cases and 11,934 controls of European ancestry in 17 studies from 11 study sites (**Supplementary Table 12**) conducted in the United States, Europe and Canada<sup>22, 23</sup>. Details about study participants, genotyping, quality control, imputation, statistical analysis and meta-analysis have been reported in elsewhere<sup>1, 22, 23</sup>. A summary description of each study is provided below.

**Ontario Familial Colorectal Cancer Registry (OFCCR):** In GECCO, a subset of the Assessment of Risk in Colorectal Tumors in Canada (ARCTIC) from the Ontario Registry for Studies of Familial Colorectal Cancer (OFCCR) was used. Both the case-control study<sup>24</sup> and the OFCCR<sup>25</sup> have been described in detail previously, as have GWAS results<sup>26</sup>.

**French Association Study Evaluating RISK for sporadic colorectal cancer (ASTERISK):**

Participants were recruited from the Pays de la Loire region in France between December 2002 and March 2006. Details of this study have been previously described<sup>27</sup>.

**Darmkrebs: Chancen der Verhütung durch Screening (DACHS):** This German study was initiated as a large population-based case-control study in 2003 in the Rhine-Neckar-Odenwald region to assess the potential of endoscopic screening for reduction of colorectal cancer risk and to investigate etiologic determinants of disease. Details of this study have been described previously<sup>28, 29</sup>. The Set 1 scan consisted of a subset of participants recruited through 2007, and samples were frequency-matched on age and gender. The Set 2 scan consisted of additional subjects that were recruited through 2010 as part of this ongoing study.

**Diet, Activity, and Lifestyle Study (DALS):** DALS is a population-based case-control study of colon cancer. Participants were recruited between 1991 and 1994 from three locations: the Kaiser Permanente Medical Care Program (KPMCP) of Northern California, an eight-county area in Utah, and the metropolitan Twin Cities area of Minnesota. Details of this study have been described previously<sup>30</sup>. The Set I scan consisted of a subset of the study designed above, from Utah, Minnesota, and KPMCP, and was restricted to subjects who self-reported as White non-Hispanic. The Set 2 scan consisted of subjects from Utah and Minnesota that were not genotyped

in Set 1. Set 2 was restricted to subjects who self-reported as White non-Hispanic and those that had appropriate consent to post data to dbGaP.

**Health Professionals Follow-up Study (HPFS):** The HPFS is a parallel prospective study to the Nurses' Health Study (NHS) that was conducted with the purpose of evaluating nutritional factors in relation to incidence of serious illnesses. The cohort includes 51,529 men in various health professions who responded to a mailed questionnaire in 1986. Details of the study have been described elsewhere<sup>31</sup>. Two case-control sets were constructed from which DNA was isolated from either buffy coat or buccal cells for genotyping: 1) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a blood sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the cases; 2) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a buccal sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the case. For both case-control sets, matching criteria included year of birth (within 1 year) and month/year of blood or buccal cell sampling (within six months). Cases were pair matched 1:1, 1:2, or 1:3 with a control participant(s).

**Nurses' Health Study (NHS):** The NHS cohort began in 1976 when 121,700 married female registered nurses ages 30 to 55 years returned the initial questionnaire that ascertained a variety of important health-related exposures. Details of this study have been described previously<sup>32</sup>, and study resources are available online (<http://www.channing.harvard.edu/nhs/>). We constructed two case-control sets from which DNA was isolated from either buffy coat or buccal cells for genotyping: 1) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a blood sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the case; 2) a case-control set with cases of

colorectal cancer matched to randomly selected controls who provided a buccal sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the cases. For both case-control sets, matching criteria included year of birth (within one year) and month/year of blood or buccal cell sampling (within six months). Cases were pair matched 1:1, 1:2, or 1:3 with a control participant(s).

**Physician's Health Study (PHS):** The PHS was established as a randomized, double-blind, placebo-controlled trial of aspirin and β-carotene among 22,071 healthy U.S. male physicians, between 40 and 84 years of age in 1982. Details have been described elsewhere<sup>33, 34</sup>. Among those who provided baseline blood samples, colorectal cases were ascertained through March 31, 2008, and controls were matched on age (within one year for younger participants, up to five years for older participants) and smoking status (never, past, current). Cases were “pair” matched 1:1, 1:2 or 1:3 with a control participant(s). Due to DNA availability, samples were genotyped in two batches on the same platform at the same genotyping center at different time points.

**Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO):** PLCO enrolled 154,934 participants (men and women between 55 and 74 years of age) at ten centers into a large, randomized, two-arm trial to determine the effectiveness of screening to reduce cancer mortality. Details of this study have been previously described<sup>35, 36</sup> and are available online (<http://dcp.cancer.gov/plco>). The Set 1 scan included a subset of 577 colon cancer cases self-reported as non-Hispanic White with available DNA samples, questionnaire data, and appropriate consent for ancillary epidemiologic studies. Controls come from the Cancer Genetic Markers of Susceptibility (CGEMS) prostate cancer scan<sup>37</sup> (all male) and the GWAS of Lung Cancer and Smoking<sup>38</sup> (enriched for smokers) along with an additional 92 non-Hispanic White female controls. The Set 2 scan included CRC cases from both arms of the trial, which were not

already included in Set 1. Controls were frequency matched 1:1 to cases without replacement, and cases were not eligible to be controls. Matching criteria were age at enrollment (two year blocks), enrollment date (two year blocks), sex, race/ethnicity, trial arm, and study year of diagnosis (i.e. controls must be cancer free into the case's year of diagnosis).

**Postmenopausal Hormones Supplementary Study to the Colon Cancer Family Registry**

**(PMH-CCFR)**: PMH is a population-based case-control study evaluating the effect of postmenopausal hormone use on CRC. Female participants from 13 counties in Washington State were enrolled from 1998-2002. Details of this study have been previously described<sup>39</sup>. Only participants that were not part of the CCFR Seattle site were included in the sample set.

**VITamins And Lifestyle (VITAL)**: The VITamins And Lifestyle (VITAL) cohort is comprised of 77,721 Washington State men and women ages 50 to 76 years, recruited from 2000 to 2002 to investigate the association of supplement use and lifestyle factors with cancer risk. Details of this study have been described elsewhere<sup>40</sup>. In GECCO, a nested case-control set was genotyped. Samples included CRC cases with DNA. Controls were matched on age at enrollment (within one year), enrollment date (within one year), sex, and race/ethnicity. One control was randomly selected per case among all controls that matched on the four factors above and where the control follow-up time was greater than follow-up time of the case until diagnosis.

**Women’s Health Initiative (WHI)**: WHI is a long-term health study of 161,808 post-menopausal women ages 50 to 79 years recruited from 40 clinical centers throughout the U.S. WHI is comprised of a Clinical Trial (CT) arm, an Observational Study (OS) arm, and several extension studies. Details of WHI have been previously described<sup>41, 42</sup> and are available online (<https://cleo.whi.org/SitePages/Home.aspx>). Set 1 cases were selected from a 2005 database and

were comprised of centrally adjudicated colon cancer cases from the OS who self-reported as White. Controls were first selected among controls previously genotyped as part of a Hip Fracture GWAS conducted within the WHI OS and matched to cases on age (within three years) enrollment date (within 365 days), hysterectomy status, and prevalent conditions at baseline. Set 2 scan cases were selected from the August 2009 database, and were comprised of centrally adjudicated colon and CRC cases from the OS and CT who were not genotyped in Set 1. Matching criteria included age (within years), race/ethnicity, WHI date (within three years), WHI Calcium and Vitamin D study date (within three years), and randomization arms (OS flag, hormone therapy assignments, dietary modification assignments, calcium/vitamin D assignments). In addition, they were matched on the four regions of randomization centers. Each case was matched with one control (1:1) that met the exact matching criteria. Control selection was done in a time-forward manner, selecting one control for each case first from the risk set at the time of the case's event.

### **The Colorectal Transdisciplinary (CORECT) Study**

#### **Study populations**

The CORECT study meta-analysis was conducted using germline DNA from 4 case-control studies (MECC<sup>43</sup>, CCFR<sup>44</sup>, Kentucky<sup>45</sup> and Newfoundland<sup>46</sup>) and 2 cohort studies (CPSII<sup>47</sup> and Melbourne<sup>48</sup>). Summaries of the 6 studies are provided in provided in **Supplementary**

#### **Table 12.**

#### **Genotyping and quality control**

Samples in the CORECT study were genotyped using the Affymetrix Axiom, Illumina Omni 2.5, Illumina 1M, Illumina 1M-Duo, or Illumina Omni1 platforms. Genotype data were cleaned

based on quality control (QC) metrics at the individual participant and SNP levels. Samples with <95% call rate, sex mismatches (between self-reported and genotypic predicted sex), low concordance with previous genotype data, duplicate samples, unanticipated genotype concordance, identity-by-descent (IBD) with another sample or ethnic outliers as identified by visual inspection of PCA cluster plots were removed. Prior to imputation, SNPs with <95% call rate, concordance <95% with the 1000 Genomes Project in samples genotyped for quality control or  $P$  for HWE  $<10^{-4}$  in controls were excluded. All SNPs overlapping the 1000 Genomes Project were matched to the forward strand.

### **Imputation**

To analyze genotype data generated from three different platforms that measure different genetic markers and to increase the coverage of variation that is measurable across the genome, imputation of genotypes was performed for both autosomal and X chromosome markers. IMPUTE2 was used to impute missing genotypes for study samples based on the cosmopolitan panel of reference haplotypes from Phase I of the 1000 Genomes Project (March 2012 release; n =1092). In order to enter subsequent statistical analysis steps, genetic markers resulting from the imputation had to pass stringent imputation quality and accuracy filters (info  $\geq 0.7$ , certainty  $\geq 0.9$ , concordance  $\geq 0.9$  between directly measured and imputed genotypes after masking input genotypes (for genotyped markers only). Version differences in Illumina 1M and Omni1 platforms led us to exclude SNPs where allele frequency differences in CFR cases were identified ( $P < 10^{-6}$ ). Further, we restricted our SNP list to those with study-specific MAF  $\geq 1\%$ .

### **GWAS meta-analysis**

Each contributing dataset was first analyzed in a study-specific fashion, allowing for adjustment for appropriate covariates, including age, sex, study center, genotyping batch and 2-4 principal

components. Then, study-specific results were analyzed using an inverse-variance-weighted, fixed-effects meta-analysis that assumed homogeneity of effects across all studies. To examine the association between each variant and CRC risk, we specified a log-additive genetic model, where each additional copy of the minor allele was assumed to confer the same magnitude of risk or protection. Each SNP was coded as a dosage for the expected number of effect alleles. We calculated beta coefficients and corresponding odds ratios (OR), standard errors, 95% confidence intervals and *P*-values, using unconditional logistic regression. These models were used to examine the ORs for CRC risk associated with each additional copy of the minor allele (or minor allele dosage) for a given SNP, after adjusting for all covariates in the model. Quantile-quantile (Q-Q) plots were generated for each study, as well as the overall meta-analysis, to examine the distribution of *P*-values compared with the distribution under null expectations. The genomic control lambda (GC  $\lambda$ ) associated with the observed *P*-value distribution was examined for each study and the summary meta-analysis, with little evidence of unadjusted population stratification. Statistical analysis and plotting were conducted using a combination of PLINK v1.07<sup>16</sup>, R v2.15.2 and METAL<sup>49</sup>.

## REFERENCES

1. Jia,W.H. *et al.* Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat. Genet.* **45**, 191-196 (2013).
2. Zheng,W. *et al.* The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am. J. Epidemiol.* **162**, 1123-1131 (2005).
3. Cai,H. *et al.* Dietary patterns and their correlates among middle-aged and elderly Chinese men: a report from the Shanghai Men's Health Study. *Br. J. Nutr.* **98**, 1006-1013 (2007).
4. Zheng,W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324-328 (2009).
5. Petersen,G.M. *et al.* A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **42**, 224-228 (2010).
6. Abnet,C.C. *et al.* A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat. Genet.* **42**, 764-767 (2010).
7. Bei,J.X. *et al.* A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat. Genet.* **42**, 599-603 (2010).
8. Matsuo,K. *et al.* Association between an 8q24 locus and the risk of colorectal cancer in Japanese. *BMC. Cancer* **9**, 379 (2009).
9. Nakata,I. *et al.* Association between the SERPING1 gene and age-related macular degeneration and polypoidal choroidal vasculopathy in Japanese. *PLoS. One.* **6**, e19108 (2011).
10. Jee,S.H. *et al.* Adiponectin concentrations: a genome-wide association study. *Am. J. Hum. Genet.* **87**, 545-552 (2010).
11. Cui,R. *et al.* Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut* **60**, 799-805 (2011).
12. Kim,J. *et al.* Dietary intake of folate and alcohol, MTHFR C677T polymorphism, and colorectal cancer risk in Korea. *Am. J. Clin. Nutr.* **95**, 405-412 (2012).
13. Song,H.R. *et al.* Sex-specific differences in the association between ABO genotype and gastric cancer risk in a Korean population. *Gastric. Cancer* **16**, 254-260 (2013).
14. Cui,L.H. *et al.* Methylenetetrahydrofolate reductase C677T polymorphism in patients with gastric and colorectal cancer in a Korean population. *BMC. Cancer* **10**, 236 (2010).
15. Amundadottir,L. *et al.* Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* **41**, 986-990 (2009).
16. Purcell,S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).
17. Li,Y., Willer,C.J., Ding,J., Scheet,P., & Abecasis,G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816-834 (2010).
18. Howie,B., Fuchsberger,C., Stephens,M., Marchini,J., & Abecasis,G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955-959 (2012).
19. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337 (2012).
20. Price,A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909 (2006).

21. Freedman,M.L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* **36**, 388-393 (2004).
22. Peters,U. *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.* **131**, 217-234 (2012).
23. Peters,U. *et al.* Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* **144**, 799-807 (2013).
24. Cotterchio,M., Manno,M., Klar,N., McLaughlin,J., & Gallinger,S. Colorectal screening is associated with reduced colorectal cancer risk: a case-control study within the population-based Ontario Familial Colorectal Cancer Registry. *Cancer Causes Control* **16**, 865-875 (2005).
25. Cotterchio,M. *et al.* Ontario familial colon cancer registry: methods and first-year response rates. *Chronic. Dis. Can.* **21**, 81-86 (2000).
26. Zanke,B.W. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39**, 989-994 (2007).
27. Kury,S. *et al.* The thorough screening of the MUTYH gene in a large French cohort of sporadic colorectal cancers. *Genet. Test.* **11**, 373-379 (2007).
28. Brenner,H., Chang-Claude,J., Seiler,C.M., Rickert,A., & Hoffmeister,M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. *Ann. Intern. Med.* **154**, 22-30 (2011).
29. Lilla,C. *et al.* Effect of NAT1 and NAT2 genetic polymorphisms on colorectal cancer risk associated with exposure to tobacco smoke and meat consumption. *Cancer Epidemiol. Biomarkers Prev.* **15**, 99-107 (2006).
30. Slattery,M.L. *et al.* Energy balance and colon cancer--beyond physical activity. *Cancer Res.* **57**, 75-80 (1997).
31. Rimm,E.B. *et al.* Validity of self-reported waist and hip circumferences in men and women. *Epidemiology* **1**, 466-473 (1990).
32. Belanger,C.F., Hennekens,C.H., Rosner,B., & Speizer,F.E. The nurses' health study. *Am J Nurs.* **78**, 1039-1040 (1978).
33. Hennekens,C.H. & Eberlein,K. A randomized trial of aspirin and beta-carotene among U.S. physicians. *Prev. Med.* **14**, 165-168 (1985).
34. Christen,W.G., Gaziano,J.M., & Hennekens,C.H. Design of Physicians' Health Study II--a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. *Ann. Epidemiol.* **10**, 125-134 (2000).
35. Gohagan,J.K., Prorok,P.C., Hayes,R.B., & Kramer,B.S. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control Clin. Trials* **21**, 251S-272S (2000).
36. Prorok,P.C. *et al.* Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin. Trials* **21**, 273S-309S (2000).
37. Yeager,M. *et al.* Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **41**, 1055-1057 (2009).
38. Landi,M.T. *et al.* 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 (2009).
39. Newcomb,P.A. *et al.* Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women. *Cancer Res.* **67**, 7534-7539 (2007).

40. White,E. *et al.* VITamins And Lifestyle cohort study: study design and characteristics of supplement users. *Am J Epidemiol.* **159**, 83-93 (2004).
41. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin. Trials* **19**, 61-109 (1998).
42. Hays,J. *et al.* The Women's Health Initiative recruitment methods and results. *Ann. Epidemiol.* **13**, S18-S77 (2003).
43. Gruber,S.B. *et al.* Genetic variation in 8q24 associated with risk of colorectal cancer. *Cancer Biol. Ther.* **6**, 1143-1147 (2007).
44. Newcomb,P.A. *et al.* Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol. Biomarkers Prev.* **16**, 2331-2343 (2007).
45. Li,L. *et al.* A common 8q24 variant and the risk of colon cancer: a population-based case-control study. *Cancer Epidemiol. Biomarkers Prev.* **17**, 339-342 (2008).
46. Sun,Z. *et al.* Calcium and vitamin D and risk of colorectal cancer: results from a large population-based case-control study in Newfoundland and Labrador and Ontario. *Can. J. Public Health* **102**, 382-389 (2011).
47. Calle,E.E. *et al.* The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* **94**, 2490-2501 (2002).
48. Bassett,J.K. *et al.* Body size, weight change, and risk of colon cancer. *Cancer Epidemiol. Biomarkers Prev.* **19**, 2978-2986 (2010).
49. Willer,C.J., Li,Y., & Abecasis,G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
50. Figueiredo,J.C. *et al.* 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 (2011).

## SUPPLEMENTARY TABLES

**Supplementary Table 1:** Selected characteristics of cases and controls in studies participating in the Asia Colorectal Cancer Consortium.

**Supplementary Table 2:** Sample size, SNPs and genotyping platforms in each of the five studies included stage 1.

**Supplementary Table 3:** Results of the 29 SNPs evaluated in stage 4.

**Supplementary Table 4:** Genotype counts of the ten newly identified genetic variants by study.

**Supplementary Table 5:** Genes in the newly identified loci.

**Supplementary Table 6:** Association of risk variants in the newly identified loci with colon and rectal cancer in East Asians.

**Supplementary Table 7:** Association of risk variants in the newly identified loci with CRC in Chinese, Korean and Japanese populations.

**Supplementary Table 8:** Association of risk variants in the newly identified loci with CRC by sex in East Asians.

**Supplementary Table 9:** Association of haplotypes in the 11q12.2 locus with CRC risk in East Asians.

**Supplementary Table 10:** Association of haplotypes in the 19q13.2 locus with CRC risk in East Asians.

**Supplementary Table 11:** Interactions between the six SNPs and SNPs in 31 known CRC susceptibility loci.

**Supplementary Table 12:** Characteristics of cases and controls in GECCO, CORECT and CCFR.

**Supplementary Table 13:** Risk allele frequencies in Europeans and East Asians from the HapMap Project.

**Supplementary Table 14:** Functional annotation of SNPs correlated with newly identified risk variants ( $r^2 > 0.5$ ) using data from ENCODE.

**Supplementary Table 15:** Analyses of expression quantitative trait locus (eQTL) in newly identified loci associated with CRC.

**Supplementary Table 16:** Expression levels of genes located in the newly identified loci in colon tumor tissue and normal colon tissue using data from TCGA.

**Supplementary Table 17:** Associations of 31 risk variants in previously reported susceptibility loci with CRC in East Asians.

**Supplementary Table 18:** Contribution of CRC risk variants identified to date to the familial relative risk in East Asians.

**Supplementary Table 19:** Contribution of CRC risk variants identified to date to the familial relative risk in European descendants.

**Supplementary Table 20:** Concordance between imputed and genotyped data.

**Supplementary Table 1: Selected characteristics of cases and controls in studies participating in the Asia Colorectal Cancer Consortium**

Study	Country	Population	Cases			Controls		
			N <sup>a</sup>	Age <sup>b</sup>	Female (%)	N <sup>a</sup>	Age <sup>b</sup>	Female (%)
<b>Stage 1</b>								
Shanghai-1	China	Chinese	474	60.02	73.84	2,628	51.99	94.82
Shanghai-2	China	Chinese	254	61.16	54.72	654	60.51	57.03
Guangzhou-1	China	Chinese	641	54.86	36.51	972	47.40	27.06
Aichi-1	Japan	Japanese	404	59.43	37.38	942	47.88	47.77
KCPS-II	Korea	Korean	325	51.38	27.08	976	41.27	43.34
<b>Subtotal</b>			<b>2,098</b>			<b>6,172</b>		
<b>Stage 2</b>								
Shanghai-3	China	Chinese	1,640	61.62	45.61	4,937	55.08	85.03
Guangzhou-2	China	Chinese	402	57.84	40.05	407	52.19	36.86
Guangzhou-3	China	Chinese	1,477	55.17	37.51	931	54.71	37.16
<b>Subtotal</b>			<b>3,519</b>			<b>6,275</b>		
<b>Stage 3</b>								
BBJ	Japan	Japanese	2,814	63.28	36.08	11,358	61.33	60.16
<b>Subtotal</b>			<b>2,814</b>			<b>11,358</b>		
<b>Stage 4</b>								
Guangzhou-4	China	Chinese	677	60.29	36.78	1,114	55.54	40.13
Aichi-2	Japan	Japanese	236	60.20	35.17	472	60.08	35.17
Korea-NCC	Korea	Korean	1,392	58.19	37.64	1,329	55.59	38.60
Korea-Seoul	Korea	Korean	849	59.05	40.99	673	57.19	47.85
HCES-CRC	Korea	Korean	3,378	62.64	35.49	4,552	58.06	87.52
<b>Subtotal</b>			<b>6,532</b>			<b>8,140</b>		
<b>Total (N=46,908)</b>			<b>14,963</b>			<b>31,945</b>		

<sup>a</sup> Final sample size used in statistical analysis.

<sup>b</sup> Mean age (years) of cases and controls.

**Supplementary Table 2: Sample size, SNPs and genotyping platforms in each of the five studies included stage 1**

Study	Genotyped			After quality-control			Genotyping Platform	
	Cases	Controls	SNPs	Cases	Controls	SNPs <sup>a</sup>	Cases	Controls
Shanghai-1	481	2,632	906,602	474	2,628	502,145	Affymetrix 6.0	Affymetrix 6.0
Shanghai-2	296	680	729,462/561,466/657,364	254	654	245,961	Illumina OmniExpress	Illumina OmniExpress/HumanHap550/660W-Quad
Guangzhou-1	694	972	729,462/620,901	641	972	250,612	Illumina OmniExpress	Illumina Human610-Quad
Aichi-1	497	942	729,462/592,044	404	942	232,426	Illumina OmniExpress	Illumina HumanHap610
KCPS-II	325	977	443,104	325	976	312,869	Affymetrix 5.0	Affymetrix 5.0

<sup>a</sup> Only shared SNPs were included in the analysis if samples in a study were genotyped using different platforms.

**Supplementary Table 3: Results of the 29 SNPs evaluated in stage 4**

SNP	Chr.	Position <sup>a</sup>	Alleles <sup>b</sup>	RAF <sup>c</sup>	Stage 1		Stage 2		Stage 3		Stage 4		Combined	
					OR (95% CI) <sup>d</sup>	P <sup>d</sup>	OR (95% CI) <sup>d</sup>	P <sup>d</sup>						
rs6691573	1	112414629	T/C	0.71	1.13 (1.03-1.23)	0.01	1.09 (1.01-1.18)	0.02	1.07 (1.00-1.14)	0.04	1.03 (0.98-1.09)	0.28	1.07 (1.03-1.10)	1.66×10 <sup>-4</sup>
rs10752881	1	181240114	A/G	0.39	1.07 (0.99-1.15)	0.09	1.04 (0.98-1.11)	0.20	1.13 (1.06-1.20)	1.08×10 <sup>-4</sup>	1.06 (1.01-1.11)	0.03	1.07 (1.04-1.10)	5.30×10 <sup>-6</sup>
rs2241975	2	113700974	T/C	0.10	1.20 (1.06-1.36)	0.006	1.19 (1.08-1.31)	6.27×10 <sup>-4</sup>	1.06 (0.95-1.19)	0.32	1.01 (0.94-1.09)	0.71	1.09 (1.04-1.14)	5.97×10 <sup>-4</sup>
rs12466239	2	223952329	C/T	0.93	1.21 (1.04-1.41)	0.02	1.22 (1.09-1.36)	3.44×10 <sup>-4</sup>	1.05 (0.93-1.19)	0.42	0.97 (0.88-1.07)	0.58	1.09 (1.03-1.15)	0.004
rs6469656	8	117716969	A/G	0.65	1.11 (1.02-1.21)	0.01	1.14 (1.07-1.23)	1.99×10 <sup>-4</sup>	1.08 (1.02-1.15)	0.01	1.07 (1.01-1.12)	0.01	1.09 (1.06-1.13)	5.38×10 <sup>-8</sup>
rs1333048	9	22115347	A/C	0.52	1.11 (1.02-1.20)	0.02	1.05 (0.99-1.12)	0.12	1.11 (1.05-1.18)	5.57×10 <sup>-4</sup>	1.01 (0.97-1.06)	0.63	1.06 (1.03-1.09)	3.00×10 <sup>-4</sup>
rs4948317	10	60241441	C/T	0.27	1.14 (1.05-1.24)	0.003	1.12 (1.04-1.21)	0.002	1.08 (1.01-1.15)	0.03	1.08 (1.02-1.14)	0.005	1.10 (1.06-1.13)	7.14×10 <sup>-8</sup>
rs704017	10	80489138	G/A	0.32	1.13 (1.03-1.23)	0.01	1.10 (1.02-1.18)	0.01	1.10 (1.03-1.17)	0.004	1.09 (1.03-1.14)	9.99×10 <sup>-4</sup>	1.10 (1.06-1.13)	2.07×10 <sup>-8</sup>
rs12412391	10	101278925	G/A	0.42	1.15 (1.06-1.24)	7.35×10 <sup>-4</sup>	1.07 (1.01-1.15)	0.03	1.08 (1.02-1.15)	0.008	1.05 (1.00-1.10)	0.03	1.08 (1.05-1.11)	7.41×10 <sup>-7</sup>
rs11196172	10	114716833	A/G	0.68	1.19 (1.03-1.34)	0.03	1.17 (1.09-1.25)	1.82×10 <sup>-5</sup>	1.08 (1.01-1.16)	0.03	1.14 (1.09-1.20)	5.18×10 <sup>-7</sup>	1.14 (1.10-1.18)	1.04×10 <sup>-12</sup>
rs174537	11	61309256	G/T	0.59	1.09 (1.01-1.18)	0.02	1.16 (1.08-1.24)	1.33×10 <sup>-5</sup>	1.12 (1.06-1.19)	1.61×10 <sup>-4</sup>	1.21 (1.15-1.27)	1.60×10 <sup>-13</sup>	1.16 (1.12-1.19)	9.22×10 <sup>-21</sup>
rs4246215	11	61320875	G/T	0.59	1.09 (1.01-1.18)	0.02	1.17 (1.10-1.26)	2.29×10 <sup>-6</sup>	1.12 (1.06-1.19)	1.83×10 <sup>-4</sup>	1.19 (1.13-1.25)	1.25×10 <sup>-11</sup>	1.15 (1.12-1.19)	7.65×10 <sup>-20</sup>
rs174550	11	61328054	T/C	0.59	1.11 (1.02-1.19)	0.01	1.17 (1.09-1.25)	5.71×10 <sup>-6</sup>	1.12 (1.06-1.19)	1.83×10 <sup>-4</sup>	1.19 (1.13-1.25)	2.70×10 <sup>-11</sup>	1.15 (1.12-1.19)	1.58×10 <sup>-19</sup>
rs1535	11	61354548	A/G	0.59	1.10 (1.02-1.19)	0.02	1.16 (1.09-1.24)	7.55×10 <sup>-6</sup>	1.13 (1.06-1.20)	1.24×10 <sup>-4</sup>	1.19 (1.13-1.25)	1.20×10 <sup>-11</sup>	1.15 (1.12-1.19)	8.21×10 <sup>-20</sup>
rs4980836	12	545702	A/T	0.39	1.10 (1.02-1.20)	0.01	1.05 (0.98-1.12)	0.18	1.11 (1.04-1.18)	0.002	1.00 (0.95-1.05)	0.97	1.05 (1.02-1.08)	0.002
rs12309274	12	846209	T/G	0.85	1.16 (1.03-1.31)	0.02	1.08 (0.96-1.20)	0.19	1.12 (1.04-1.20)	0.003	1.10 (1.02-1.18)	0.01	1.11 (1.06-1.16)	2.65×10 <sup>-6</sup>
rs10849432	12	6255988	T/C	0.82	1.21 (1.07-1.36)	0.002	1.12 (1.03-1.22)	0.007	1.09 (1.00-1.18)	0.06	1.15 (1.08-1.23)	6.95×10 <sup>-6</sup>	1.14 (1.09-1.18)	5.81×10 <sup>-10</sup>
rs725957	12	55618008	G/A	0.41	1.11 (1.03-1.20)	0.005	1.06 (0.99-1.13)	0.10	1.09 (1.03-1.16)	0.004	1.03 (0.98-1.08)	0.30	1.06 (1.03-1.09)	8.19×10 <sup>-5</sup>
rs4144229	13	72887159	T/A	0.74	1.17 (1.07-1.28)	6.02×10 <sup>-4</sup>	1.11 (1.04-1.19)	0.004	1.05 (0.97-1.13)	0.20	1.03 (0.98-1.09)	0.25	1.07 (1.04-1.11)	5.57×10 <sup>-5</sup>
rs12603526	17	747343	C/T	0.30	1.12 (1.02-1.22)	0.02	1.14 (1.06-1.23)	6.86×10 <sup>-4</sup>	1.06 (0.99-1.13)	0.08	1.11 (1.05-1.17)	3.80×10 <sup>-4</sup>	1.10 (1.06-1.14)	3.42×10 <sup>-8</sup>
rs6259	17	7477252	A/G	0.13	1.16 (1.04-1.29)	0.008	1.20 (1.09-1.31)	1.07×10 <sup>-4</sup>	1.04 (0.95-1.14)	0.36	1.03 (0.96-1.10)	0.36	1.09 (1.04-1.14)	1.06×10 <sup>-4</sup>
rs1641537	17	7486446	C/T	0.61	1.11 (1.02-1.20)	0.01	1.12 (1.04-1.20)	0.002	1.06 (1.00-1.13)	0.05	0.96 (0.88-1.04)	0.33	1.07 (1.03-1.11)	4.34×10 <sup>-4</sup>
rs17817050	17	10650478	A/G	0.91	1.32 (1.12-1.55)	7.50×10 <sup>-4</sup>	1.06 (0.95-1.19)	0.26	1.19 (1.06-1.34)	0.004	1.09 (1.01-1.18)	0.03	1.13 (1.07-1.19)	1.10×10 <sup>-5</sup>
rs2241716	19	46545926	T/C	0.34	1.16 (1.07-1.27)	8.80×10 <sup>-4</sup>	1.13 (1.05-1.21)	6.83×10 <sup>-4</sup>	1.12 (1.05-1.20)	0.001	1.04 (0.99-1.09)	0.15	1.09 (1.05-1.12)	1.86×10 <sup>-7</sup>
rs1800469	19	46552136	G/A	0.48	1.13 (1.04-1.22)	0.002	1.11 (1.04-1.19)	0.002	1.11 (1.04-1.17)	6.74×10 <sup>-4</sup>	1.06 (1.01-1.11)	0.03	1.09 (1.06-1.12)	1.17×10 <sup>-8</sup>
rs2241714	19	46561232	C/T	0.48	1.12 (1.04-1.20)	0.003	1.11 (1.04-1.19)	0.002	1.10 (1.04-1.17)	0.001	1.06 (1.01-1.11)	0.02	1.09 (1.06-1.12)	1.36×10 <sup>-8</sup>
rs6126948	20	36125647	A/G	0.45	1.10 (1.01-1.18)	0.02	1.11 (1.04-1.18)	0.003	1.08 (1.02-1.14)	0.01	0.97 (0.92-1.01)	0.14	1.04 (1.01-1.07)	0.01
rs909388	21	15273319	A/C	0.56	1.17 (1.08-1.26)	6.52×10 <sup>-5</sup>	1.08 (1.02-1.16)	0.02	1.08 (1.02-1.15)	0.01	1.00 (0.96-1.05)	0.93	1.06 (1.03-1.09)	6.67×10 <sup>-5</sup>
rs2249060	21	46597605	T/C	0.23	1.09 (1.00-1.20)	0.06	1.12 (1.04-1.22)	0.004	1.07 (1.00-1.15)	0.04	1.04 (0.98-1.10)	0.16	1.07 (1.04-1.11)	8.73×10 <sup>-5</sup>

Abbreviations: Chr., chromosome; RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

<sup>a</sup> The chromosome position (bp) is based on the National Center for Biotechnology Information (NCBI) database, build 36.

<sup>b</sup> Risk/reference alleles are based on forward allele coding in NCBI, build 36. OR was estimated for the risk allele (bold).

<sup>c</sup> RAF in controls.

<sup>d</sup> Summary OR (95% CI) and P value were obtained from fixed-effect meta-analysis in each stage.

**Supplementary Table 4: Genotype counts of the ten newly identified genetic variants by study**

SNP	Chr.	Position <sup>a</sup>	Alleles (1/2) <sup>b</sup>	Study	Cases			Controls		
					11	12	22	11	12	22
rs704017	10	80489138	G/A	Shanghai-1	57	207	210	245	1106	1277
rs704017	10	80489138	G/A	Shanghai-2	26	108	120	50	273	331
rs704017	10	80489138	G/A	Guangzhou-1	47	261	333	65	361	546
rs704017	10	80489138	G/A	Aichi-1	38	189	177	97	414	431
rs704017	10	80489138	G/A	KCPS-II	29	153	143	100	419	457
rs704017	10	80489138	G/A	Shanghai-3	163	676	801	413	2082	2442
rs704017	10	80489138	G/A	Guangzhou-2	37	159	206	20	172	215
rs704017	10	80489138	G/A	Guangzhou-3	111	662	704	72	357	502
rs704017	10	80489138	G/A	BBJ	349	1291	1105	1289	5062	4774
rs704017	10	80489138	G/A	Guangzhou-4	56	261	318	73	433	591
rs704017	10	80489138	G/A	Aichi-2	29	119	88	52	232	187
rs704017	10	80489138	G/A	Korea-NCC	151	606	628	138	594	590
rs704017	10	80489138	G/A	Korea-Seoul	106	364	356	86	283	279
rs704017	10	80489138	G/A	HCES-CRC	418	1564	1374	498	1973	2020
rs11196172	10	114716833	A/G	Shanghai-1	232	210	32	1208	1169	251
rs11196172	10	114716833	A/G	KCPS-II	145	161	19	452	442	82
rs11196172	10	114716833	A/G	Shanghai-3	833	680	127	2341	2154	441
rs11196172	10	114716833	A/G	Guangzhou-2	177	168	57	159	185	63
rs11196172	10	114716833	A/G	Guangzhou-3	615	689	173	332	446	153
rs11196172	10	114716833	A/G	BBJ	1388	1137	220	5444	4644	1037
rs11196172	10	114716833	A/G	Guangzhou-4	263	289	84	421	499	177
rs11196172	10	114716833	A/G	Aichi-2	114	102	18	217	198	54
rs11196172	10	114716833	A/G	Korea-NCC	778	505	104	687	525	110
rs11196172	10	114716833	A/G	Korea-Seoul	457	319	51	345	257	46
rs11196172	10	114716833	A/G	HCES-CRC	1884	1266	214	2296	1880	322
rs174537	11	61309256	G/T	Shanghai-1	153	245	76	953	1189	486
rs174537	11	61309256	G/T	Shanghai-2	76	145	33	213	304	137
rs174537	11	61309256	G/T	Guangzhou-1	85	306	250	95	429	448
rs174537	11	61309256	G/T	Aichi-1	161	188	55	345	469	128
rs174537	11	61309256	G/T	KCPS-II	145	155	25	444	431	101
rs174537	11	61309256	G/T	Shanghai-3	607	769	264	1686	2337	912

rs174537	11	61309256	G/T	Guangzhou-2	54	180	168	45	176	186
rs174537	11	61309256	G/T	Guangzhou-3	234	647	596	104	389	438
rs174537	11	61309256	G/T	BBJ	1136	1249	360	4191	5279	1655
rs174537	11	61309256	G/T	Guangzhou-4	113	252	249	108	452	519
rs174537	11	61309256	G/T	Aichi-2	101	106	28	168	237	63
rs174537	11	61309256	G/T	Korea-NCC	682	587	106	578	562	143
rs174537	11	61309256	G/T	Korea-Seoul	405	294	75	299	254	64
rs174537	11	61309256	G/T	HCES-CRC	1632	1411	297	2013	1869	571
rs4246215	11	61320875	G/T	Shanghai-1	153	245	76	961	1185	482
rs4246215	11	61320875	G/T	Shanghai-2	76	145	33	213	304	137
rs4246215	11	61320875	G/T	Guangzhou-1	87	305	249	96	429	447
rs4246215	11	61320875	G/T	Aichi-1	161	189	54	346	469	127
rs4246215	11	61320875	G/T	KCPS-II	145	155	25	444	431	101
rs4246215	11	61320875	G/T	Shanghai-3	622	736	253	1708	2327	884
rs4246215	11	61320875	G/T	Guangzhou-2	55	179	164	48	174	181
rs4246215	11	61320875	G/T	Guangzhou-3	239	646	583	106	387	430
rs4246215	11	61320875	G/T	BBJ	1135	1249	361	4191	5276	1658
rs4246215	11	61320875	G/T	Guangzhou-4	89	270	232	101	450	530
rs4246215	11	61320875	G/T	Aichi-2	100	106	28	166	240	64
rs4246215	11	61320875	G/T	Korea-NCC	680	594	107	580	598	132
rs4246215	11	61320875	G/T	Korea-Seoul	411	306	74	301	262	64
rs4246215	11	61320875	G/T	HCES-CRC	1623	1435	283	2001	1953	502
rs174550	11	61328054	T/C	Shanghai-1	161	239	74	975	1183	470
rs174550	11	61328054	T/C	Shanghai-2	76	145	33	213	306	135
rs174550	11	61328054	T/C	Guangzhou-1	86	306	249	96	428	448
rs174550	11	61328054	T/C	Aichi-1	161	189	54	346	469	127
rs174550	11	61328054	T/C	KCPS-II	147	154	24	446	430	100
rs174550	11	61328054	T/C	Shanghai-3	608	773	259	1688	2343	903
rs174550	11	61328054	T/C	Guangzhou-2	54	180	168	45	176	186
rs174550	11	61328054	T/C	Guangzhou-3	237	647	593	104	388	439
rs174550	11	61328054	T/C	BBJ	1134	1249	362	4186	5277	1662
rs174550	11	61328054	T/C	Guangzhou-4	101	287	273	107	458	539
rs174550	11	61328054	T/C	Aichi-2	100	108	26	167	239	64
rs174550	11	61328054	T/C	Korea-NCC	678	601	107	583	601	137
rs174550	11	61328054	T/C	Korea-Seoul	424	326	80	305	277	67

rs174550	11	61328054	T/C	HCES-CRC	1618	1449	283	2000	1986	501
rs1535	11	61354548	A/G	Shanghai-1	155	244	75	964	1184	480
rs1535	11	61354548	A/G	Shanghai-2	76	144	34	213	306	135
rs1535	11	61354548	A/G	Guangzhou-1	86	306	249	96	427	449
rs1535	11	61354548	A/G	Aichi-1	161	189	54	344	469	129
rs1535	11	61354548	A/G	KCPS-II	147	154	24	446	429	101
rs1535	11	61354548	A/G	Shanghai-3	608	773	259	1685	2350	900
rs1535	11	61354548	A/G	Guangzhou-2	53	181	168	45	174	188
rs1535	11	61354548	A/G	Guangzhou-3	235	647	595	104	389	438
rs1535	11	61354548	A/G	BBJ	1131	1249	365	4162	5285	1678
rs1535	11	61354548	A/G	Guangzhou-4	103	285	244	107	457	533
rs1535	11	61354548	A/G	Aichi-2	100	108	28	167	239	65
rs1535	11	61354548	A/G	Korea-NCC	680	598	110	583	603	136
rs1535	11	61354548	A/G	Korea-Seoul	422	327	76	303	278	68
rs1535	11	61354548	A/G	HCES-CRC	1616	1449	298	2000	1986	509
rs10849432	12	6255988	T/C	Shanghai-1	317	142	15	1691	830	107
rs10849432	12	6255988	T/C	Shanghai-2	169	74	11	372	254	28
rs10849432	12	6255988	T/C	Guangzhou-1	376	238	27	553	361	58
rs10849432	12	6255988	T/C	KCPS-II	239	81	5	682	269	25
rs10849432	12	6255988	T/C	Shanghai-3	1103	485	52	3141	1635	161
rs10849432	12	6255988	T/C	Guangzhou-2	245	143	14	246	141	20
rs10849432	12	6255988	T/C	Guangzhou-3	927	484	66	567	318	46
rs10849432	12	6255988	T/C	BBJ	1988	696	61	7873	2969	283
rs10849432	12	6255988	T/C	Guangzhou-4	418	193	25	635	406	56
rs10849432	12	6255988	T/C	Aichi-2	176	58	2	335	126	10
rs10849432	12	6255988	T/C	Korea-NCC	972	373	43	880	393	49
rs10849432	12	6255988	T/C	Korea-Seoul	567	234	25	441	189	20
rs10849432	12	6255988	T/C	HCES-CRC	2385	882	98	3035	1328	138
rs12603526	17	747343	C/T	Shanghai-1	38	201	235	161	1009	1458
rs12603526	17	747343	C/T	Shanghai-2	22	102	130	63	233	358
rs12603526	17	747343	C/T	Guangzhou-1	39	199	403	42	324	606
rs12603526	17	747343	C/T	Aichi-1	45	178	181	119	426	397
rs12603526	17	747343	C/T	KCPS-II	31	168	126	124	453	399
rs12603526	17	747343	C/T	Shanghai-3	113	644	883	332	1846	2759
rs12603526	17	747343	C/T	Guangzhou-2	16	145	241	15	104	288

rs12603526	17	747343	C/T	Guangzhou-3	81	496	900	39	283	609
rs12603526	17	747343	C/T	BBJ	336	1221	1188	1219	4961	4945
rs12603526	17	747343	C/T	Guangzhou-4	27	239	372	45	341	710
rs12603526	17	747343	C/T	Aichi-2	22	121	92	57	203	211
rs12603526	17	747343	C/T	Korea-Seoul	120	357	347	74	295	282
rs12603526	17	747343	C/T	HCES-CRC	431	1573	1361	527	2015	1956
rs1800469	19	46552136	G/A	Shanghai-1	137	232	105	582	1384	662
rs1800469	19	46552136	G/A	Shanghai-2	60	140	54	154	348	152
rs1800469	19	46552136	G/A	Guangzhou-1	147	273	221	167	474	331
rs1800469	19	46552136	G/A	Aichi-1	97	217	90	234	466	242
rs1800469	19	46552136	G/A	KCPS-II	93	157	75	258	486	232
rs1800469	19	46552136	G/A	Shanghai-3	439	818	383	1165	2520	1250
rs1800469	19	46552136	G/A	Guangzhou-2	71	194	136	77	190	139
rs1800469	19	46552136	G/A	Guangzhou-3	298	712	467	153	452	326
rs1800469	19	46552136	G/A	BBJ	727	1342	676	2631	5489	3005
rs1800469	19	46552136	G/A	Guangzhou-4	111	308	220	196	501	399
rs1800469	19	46552136	G/A	Aichi-2	52	125	59	119	228	123
rs1800469	19	46552136	G/A	Korea-NCC	394	692	301	346	683	293
rs1800469	19	46552136	G/A	HCES-CRC	985	1640	738	1221	2234	1047
rs2241714	19	46561232	C/T	Shanghai-1	134	235	105	579	1376	673
rs2241714	19	46561232	C/T	Shanghai-2	61	138	55	153	350	151
rs2241714	19	46561232	C/T	Guangzhou-1	146	275	220	166	474	332
rs2241714	19	46561232	C/T	Aichi-1	97	214	93	232	461	249
rs2241714	19	46561232	C/T	KCPS-II	92	155	78	253	486	237
rs2241714	19	46561232	C/T	Shanghai-3	437	818	385	1155	2526	1254
rs2241714	19	46561232	C/T	Guangzhou-2	71	191	139	74	192	139
rs2241714	19	46561232	C/T	Guangzhou-3	289	717	471	150	453	328
rs2241714	19	46561232	C/T	BBJ	712	1339	694	2591	5471	3063
rs2241714	19	46561232	C/T	Guangzhou-4	109	305	219	195	496	405
rs2241714	19	46561232	C/T	Aichi-2	51	126	59	119	227	125
rs2241714	19	46561232	C/T	Korea-NCC	392	699	298	341	682	298
rs2241714	19	46561232	C/T	HCES-CRC	978	1647	740	1214	2229	1051

Abbreviations: Chr., chromosome.

<sup>a</sup> The chromosome position (bp) is based on NCBI, build 36.

<sup>b</sup> Risk/reference alleles are based on forward allele coding in NCBI, build 36.

**Supplementary Table 5: Genes in the newly-identified loci**

Locus	SNP	Gene <sup>a</sup>	Position <sup>b</sup>	Functional annotation	Genes in a locus ( $r^2 > 0.5$ ) <sup>c</sup>
10q22.3	rs704017	<i>ZMIZ1</i> -AS1	80489138	Intron 3	<i>ZMIZ1</i> -AS1, <i>ZMIZ1</i>
10q25.2	rs11196172	<i>TCF7L2</i>	114716833	Intron 4	<i>TCF7L2</i>
11q12.2	rs174537	<i>MYRF</i>	61309256	Intron 24	<i>MYRF</i> , <i>TMEM258</i> , <i>MIR611</i> , <i>FEN1</i> , <i>FADS1</i> , <i>MIR1908</i> , <i>FADS2</i>
	rs4246215	<i>FEN1</i>	61320875	3'-UTR	<i>MYRF</i> , <i>TMEM258</i> , <i>MIR611</i> , <i>FEN1</i> , <i>FADS1</i> , <i>MIR1908</i> , <i>FADS2</i>
	rs174550	<i>FADS1</i>	61328054	Intron 7	<i>MYRF</i> , <i>TMEM258</i> , <i>MIR611</i> , <i>FEN1</i> , <i>FADS1</i> , <i>MIR1908</i> , <i>FADS2</i>
	rs1535	<i>FADS2</i>	61354548	Intron 1	<i>MYRF</i> , <i>TMEM258</i> , <i>MIR611</i> , <i>FEN1</i> , <i>FADS1</i> , <i>MIR1908</i> , <i>FADS2</i>
12p13.31	rs10849432	<i>CD9</i>	6255988	Intergenic	None
17p13.3	rs12603526	<i>NXN</i>	747343	Intron 1	<i>NXN</i>
19q13.2	rs1800469	<i>TGFB1</i>	46552136	Promoter	<i>TGFB1</i> , <i>B9D2</i> , <i>TMEM91</i> , <i>EXOSC5</i> , <i>BCKDHA</i> , <i>B3GNT8</i> , <i>ATP5SL</i>
	rs2241714	<i>B9D2</i>	46561232	Exon 1	<i>TGFB1</i> , <i>B9D2</i> , <i>TMEM91</i> , <i>EXOSC5</i> , <i>BCKDHA</i> , <i>B3GNT8</i> , <i>ATP5SL</i>

Abbreviations: Chr., chromosome.

<sup>a</sup> The closest gene(s).

<sup>b</sup> The chromosome position (bp) is based on NCBI, Build 36.

<sup>c</sup> Genes with SNPs in LD ( $r^2 > 0.5$ ) with risk variants in the newly-identified loci.

**Supplementary Table 6: Association of risk variants in the newly identified loci with colon and rectal cancer in East Asians**

Locus	SNP	Gene <sup>b</sup>	Position <sup>c</sup>	Alleles <sup>d</sup>	Colon cancer		Rectal cancer		<i>P</i> <sub>heterogeneity</sub> <sup>f</sup>
					OR (95% CI) <sup>e</sup>	<i>P</i> <sup>e</sup>	OR (95% CI) <sup>e</sup>	<i>P</i> <sup>e</sup>	
10q22.3	rs704017	<i>ZMIZ1 -AS1</i>	80489138	G/A	1.11 (1.06-1.17)	2.66×10 <sup>-5</sup>	1.11 (1.06-1.17)	7.65×10 <sup>-5</sup>	0.946
10q25.2	rs11196172	<i>TCF7L2</i>	114716833	A/G	1.15 (1.09-1.21)	1.03×10 <sup>-7</sup>	1.16 (1.10-1.23)	1.38×10 <sup>-7</sup>	0.747
11q12.2	rs174537	<i>MYRF</i>	61309256	G/T	1.18 (1.12-1.23)	9.76×10 <sup>-12</sup>	1.15 (1.09-1.21)	7.13×10 <sup>-8</sup>	0.551
	rs4246215	<i>FEN1</i>	61320875	G/T	1.18 (1.12-1.23)	2.24×10 <sup>-11</sup>	1.15 (1.09-1.21)	2.41×10 <sup>-7</sup>	0.497
	rs174550	<i>FADS1</i>	61328054	T/C	1.17 (1.12-1.23)	4.62×10 <sup>-11</sup>	1.14 (1.09-1.21)	3.45×10 <sup>-7</sup>	0.512
	rs1535	<i>FADS2</i>	61354548	A/G	1.17 (1.12-1.22)	5.88×10 <sup>-11</sup>	1.14 (1.08-1.20)	8.98×10 <sup>-7</sup>	0.465
12p13.31	rs10849432	<i>CD9</i>	6255988	T/C	1.10 (1.03-1.16)	0.004	1.12 (1.05-1.20)	8.36×10 <sup>-4</sup>	0.609
17p13.3	rs12603526	<i>NXN</i>	747343	C/T	1.11 (1.05-1.16)	6.44×10 <sup>-5</sup>	1.09 (1.03-1.15)	0.003	0.638
19q13.2	rs1800469	<i>TGFB1</i>	46552136	G/A	1.12 (1.07-1.17)	6.48×10 <sup>-7</sup>	1.08 (1.02-1.13)	0.004	0.226
	rs2241714	<i>B9D2</i>	46561232	C/T	1.12 (1.07-1.17)	8.10×10 <sup>-7</sup>	1.07 (1.02-1.13)	0.006	0.210

Abbreviations: RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

<sup>a</sup> Number of cases and controls included in the analysis.

<sup>b</sup> The closest gene(s).

<sup>c</sup> The chromosome position (bp) is based on NCBI, Build 36.

<sup>d</sup> Risk/reference alleles are based on forward allele coding in NCBI, Build 36. OR was estimated for the risk allele.

<sup>e</sup> Summary OR (95% CI) and *P* value were obtained from a fixed-effect meta-analysis of four studies (Shanghai-3, Guangzhou-3, BBJ and HCES-CRC).

<sup>f</sup> *P* for heterogeneity between colon and rectal cancer was calculated using a Cochran's *Q* test.

**Supplementary Table 7: Association of risk variants in the newly identified loci with CRC in Chinese, Korean and Japanese populations**

Locus	SNP	Gene <sup>b</sup>	Position <sup>c</sup>	Alleles <sup>d</sup>	Chinese		Korean		Japanese		<i>P</i> <sub>heterogeneity</sub> <sup>f</sup>
					(5,565/11,643) <sup>a</sup>	OR (95% CI) <sup>e</sup>	<i>P</i> <sup>e</sup>	(5,944/7,530) <sup>a</sup>	OR (95% CI) <sup>e</sup>	<i>P</i> <sup>e</sup>	
10q22.3	rs704017	<i>ZMIZ1-AS1</i>	80489138	G/A	1.12 (1.06-1.18)	9.94×10 <sup>-5</sup>	1.08 (1.02-1.14)	0.004	1.10 (1.03-1.16)	0.003	0.663
10q25.2	rs11196172	<i>TCF7L2</i>	114716833	A/G	1.16 (1.10-1.24)	5.15×10 <sup>-7</sup>	1.14 (1.08-1.21)	3.20×10 <sup>-6</sup>	1.09 (1.02-1.17)	0.01	0.330
11q12.2	rs174537	<i>MYRF</i>	61309256	G/T	1.17 (1.11-1.23)	1.64×10 <sup>-9</sup>	1.17 (1.11-1.23)	4.33×10 <sup>-9</sup>	1.13 (1.07-1.19)	2.63×10 <sup>-5</sup>	0.555
	rs4246215	<i>FEN1</i>	61320875	G/T	1.18 (1.12-1.24)	3.58×10 <sup>-10</sup>	1.15 (1.09-1.22)	1.56×10 <sup>-7</sup>	1.13 (1.07-1.19)	2.73×10 <sup>-5</sup>	0.521
	rs174550	<i>FADS1</i>	61328054	T/C	1.17 (1.11-1.23)	1.61×10 <sup>-9</sup>	1.16 (1.10-1.22)	1.04×10 <sup>-7</sup>	1.13 (1.07-1.19)	2.35×10 <sup>-5</sup>	0.647
	rs1535	<i>FADS2</i>	61354548	A/G	1.18 (1.12-1.24)	4.01×10 <sup>-10</sup>	1.15 (1.09-1.21)	2.55×10 <sup>-7</sup>	1.13 (1.07-1.20)	1.77×10 <sup>-5</sup>	0.576
12p13.31	rs10849432	<i>CD9</i>	6255988	T/C	1.16 (1.09-1.24)	3.53×10 <sup>-6</sup>	1.13 (1.06-1.21)	2.05×10 <sup>-4</sup>	1.10 (1.01-1.19)	0.02	0.533
17p13.3	rs12603526	<i>NXN</i>	747343	C/T	1.16 (1.10-1.23)	6.24×10 <sup>-7</sup>	1.09 (1.03-1.15)	0.005	1.05 (0.99-1.12)	0.08	0.069
19q13.2	rs1800469	<i>TGFB1</i>	46552136	G/A	1.11 (1.06-1.17)	3.35×10 <sup>-5</sup>	1.07 (1.01-1.12)	0.02	1.10 (1.04-1.16)	9.55×10 <sup>-4</sup>	0.507
	rs2241714	<i>B9D2</i>	46561232	C/T	1.11 (1.06-1.17)	3.44×10 <sup>-5</sup>	1.07 (1.01-1.13)	0.01	1.09 (1.03-1.15)	0.002	0.566

Abbreviations: RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

<sup>a</sup> Number of cases and controls included in the analysis.

<sup>b</sup> The closest gene(s).

<sup>c</sup> The chromosome position (bp) is based on NCBI, Build 36.

<sup>d</sup> Risk/reference alleles are based on forward allele coding in NCBI, Build 36. OR was estimated for the risk allele.

<sup>e</sup> Summary OR (95% CI) and *P* value were obtained from a fixed-effect meta-analysis.

<sup>f</sup> *P* for heterogeneity across different populations was calculated using a Cochran's *Q* test.

**Supplementary Table 8: Association of risk variants in the newly identified loci with CRC by sex in East Asians**

Locus	SNP	Gene <sup>b</sup>	Position <sup>c</sup>	Alleles <sup>d</sup>	Men		Women		
					OR (95% CI) <sup>e</sup>	P <sup>e</sup>	OR (95% CI) <sup>e</sup>	P <sup>e</sup>	P <sub>heterogeneity</sub> <sup>f</sup>
10q22.3	rs704017	ZMIZ1 -AS1	80489138	G/A	1.09 (1.05-1.14)	2.48×10 <sup>-5</sup>	1.10 (1.05-1.15)	1.38×10 <sup>-4</sup>	0.883
10q25.2	rs11196172	TCF7L2	114716833	A/G	1.11 (1.06-1.16)	9.97×10 <sup>-6</sup>	1.17 (1.11-1.23)	3.61×10 <sup>-9</sup>	0.113
11q12.2	rs174537	MYRF	61309256	G/T	1.16 (1.12-1.21)	9.35×10 <sup>-14</sup>	1.14 (1.09-1.20)	1.05×10 <sup>-8</sup>	0.559
	rs4246215	FEN1	61320875	G/T	1.16 (1.11-1.21)	2.02×10 <sup>-12</sup>	1.15 (1.09-1.20)	8.40×10 <sup>-9</sup>	0.748
	rs174550	FADS1	61328054	T/C	1.16 (1.11-1.21)	9.10×10 <sup>-13</sup>	1.14 (1.09-1.19)	2.88×10 <sup>-8</sup>	0.593
	rs1535	FADS2	61354548	A/G	1.16 (1.11-1.20)	1.39×10 <sup>-12</sup>	1.14 (1.09-1.20)	1.38×10 <sup>-8</sup>	0.696
12p13.31	rs10849432	CD9	6255988	T/C	1.12 (1.07-1.18)	1.13×10 <sup>-5</sup>	1.15 (1.08-1.22)	6.06×10 <sup>-6</sup>	0.600
17p13.3	rs12603526	NXN	747343	C/T	1.12 (1.08-1.17)	2.30×10 <sup>-7</sup>	1.07 (1.02-1.12)	0.009	0.146
19q13.2	rs1800469	TGFB1	46552136	G/A	1.10 (1.06-1.15)	1.42×10 <sup>-6</sup>	1.09 (1.04-1.14)	2.33×10 <sup>-4</sup>	0.699
	rs2241714	B9D2	46561232	C/T	1.10 (1.06-1.15)	1.30×10 <sup>-6</sup>	1.09 (1.04-1.14)	3.60×10 <sup>-4</sup>	0.630

Abbreviations: RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

<sup>a</sup> Number of cases and controls included in analysis.

<sup>b</sup> The closest gene(s).

<sup>c</sup> The chromosome position (bp) is based on NCBI, Build 36.

<sup>d</sup> Risk/reference alleles are based on forward allele coding in NCBI, Build 36. OR was estimated for the risk allele.

<sup>e</sup> Summary OR (95% CI) and P value were obtained from a fixed-effect meta-analysis.

<sup>f</sup> P for heterogeneity across different populations was calculated using a Cochran's Q test.

**Supplementary Table 9: Association of haplotypes in the 11q12.2 locus with CRC risk in East Asians**

Haplotype <sup>a</sup>	Cases frequency	Controls frequency	OR (95% CI) <sup>b</sup>	P <sup>b</sup>
T-T-C-G	0.38966	0.41799	1.00 (reference)	1.00
G-G-T-A	0.60270	0.57395	1.40 (1.29-1.51)	$3.69 \times 10^{-16}$
Other <sup>c</sup>	0.00763	0.00806	1.14 (0.74-1.75)	0.56

<sup>a</sup> Haplotypes were constructed based on rs174537, rs4246215, rs174550 and rs1535.

<sup>b</sup> Summary OR (95% CI) and P value were obtained from a fixed-effect meta-analysis of data from 10,051 CRC cases and 14,415 controls (stages 2 and 4).

<sup>c</sup> Ten other haplotypes, all with very low frequency.

**Supplementary Table 10: Association of haplotypes in the 19q13.2 locus with CRC risk in East Asians**

Haplotype <sup>a</sup>	Cases frequency	Controls frequency	OR (95% CI) <sup>b</sup>	P <sup>b</sup>
A-T	0.49550	0.50914	1.00 (reference)	1.00
G-C	0.49857	0.48517	1.16 (1.08-1.26)	1.18×10 <sup>-4</sup>
A-C	0.00209	0.00198	1.17 (0.50-2.75)	0.72
G-T	0.00384	0.00371	0.97 (0.51-1.81)	0.91

<sup>a</sup> Haplotypes were constructed based on rs1800469 and rs2241714.

<sup>b</sup> Summary OR (95% CI) and P value were obtained from a fixed-effect meta-analysis of data from 10,051 CRC cases and 14,415 controls (stages 2 and 4).

**Supplementary Table 11: Interactions between the six SNPs and SNPs in 31 known CRC susceptibility loci**

SNP	<i>P</i> for interaction					
	rs704017	rs11196172	rs174537	rs10849432	rs12603526	rs1800469
rs704017	NA	0.87	0.29	0.74	0.90	0.51
rs11196172		NA	0.22	0.55	0.28	0.80
rs174537			NA	0.001	0.96	0.55
rs10849432				NA	0.01	0.22
rs12603526					NA	0.63
rs1800469						NA
rs6691170	NA	NA	NA	NA	NA	NA
rs6687758	0.42	0.28	0.84	0.24	0.47	0.27
rs11903757	NA	NA	NA	NA	NA	NA
rs10936599	0.11	0.43	0.30	0.68	0.32	0.83
rs647161	0.77	0.68	0.38	0.70	0.42	0.56
rs1321311	0.78	0.05	0.92	0.15	0.32	0.80
rs7758229	0.30	0.74	0.28	0.35	0.97	0.61
rs16892766	NA	NA	NA	NA	NA	NA
rs10505477	0.91	0.41	0.67	0.82	0.63	0.13
rs6983267	0.73	0.13	0.83	0.69	0.13	0.004
rs7014346	0.74	0.39	0.26	0.32	0.41	0.69
rs10795668	NA	NA	NA	NA	NA	NA
rs3824999	NA	NA	NA	NA	NA	NA
rs3802842	0.82	0.17	0.98	0.59	0.92	0.14
rs10774214	0.94	0.99	0.73	0.10	0.70	0.27
rs7136702	0.17	0.46	0.37	0.93	0.49	0.04
rs11169552	0.17	0.58	0.84	0.78	0.70	0.81
rs4444235	0.22	0.67	0.71	0.37	0.48	0.07
rs1957636	NA	NA	NA	NA	NA	NA
rs16969681	NA	NA	NA	NA	NA	NA
rs4779584	0.65	0.20	0.51	0.57	0.80	0.59
rs11632715	NA	NA	NA	NA	NA	NA
rs9929218	0.02	0.02	0.76	0.24	0.23	0.46
rs7229639	0.34	0.08	0.96	0.64	0.11	0.61
rs4939827	0.47	0.61	0.79	0.43	0.20	0.30
rs10411210	0.55	0.61	0.83	0.82	0.94	0.98
rs961253	0.56	0.50	0.59	0.19	0.20	0.91
rs4813802	NA	NA	NA	NA	NA	NA
rs2423279	0.71	0.51	0.33	0.73	0.62	0.38
rs4925386	0.33	0.27	0.40	0.87	0.36	0.95
rs5934683	NA	NA	NA	NA	NA	NA

NA, not available.

**Supplementary Table 12: Characteristics of cases and controls in GECCO, CORECT and CCFR**

Study	Population	Cases			Controls		
		N <sup>a</sup>	Age <sup>b</sup>	Female (%)	N <sup>a</sup>	Age <sup>b</sup>	Female (%)
<b>GECCO</b>							
ASTERISK	Caucasian	892	68.73	38.12	947	61.91	44.67
DACHS Set 1	Caucasian	1,710	68.33	41.40	1,708	68.79	40.22
DACHS Set 2	Caucasian	666	68.75	39.04	498	69.73	35.14
DALS Set 1	Caucasian	706	63.78	43.34	710	63.70	43.52
DALS Set 2	Caucasian	410	63.69	46.83	464	64.36	47.84
HPFS Set 1	Caucasian	227	66.61	0.00	230	66.11	0.00
HPFS Set 2	Caucasian	176	63.34	0.00	172	64.04	0.00
NHS Set 1	Caucasian	391	59.97	100.00	774	59.98	100.00
NHS Set 2	Caucasian	158	58.46	100.00	181	59.46	100.00
OFCCR	Caucasian	650	61.45	58.92	522	62.69	43.49
PHS Set 1+2	Caucasian	375	58.78	0.00	389	58.07	0.00
PLCO Set 1	Caucasian	533	64.83	43.34	1,976	63.96	22.06
PLCO Set 2	Caucasian	486	63.71	42.80	415	63.58	42.17
PMH-CCFR	Caucasian	280	63.31	100.00	122	61.64	100.00
VITAL	Caucasian	285	66.38	47.37	288	66.59	47.92
WHI Set 1+Hip	Caucasian	470	67.26	100.00	1,528	69.47	100.00
WHI Set 2	Caucasian	1,006	65.90	100.00	1,010	65.64	100.00
<b>Subtotal</b>		<b>9,421</b>			<b>11,934</b>		
<b>CORECT/CCFR</b>							
Colon CFR	Caucasian	3,638	55.10	52.20	2,392	58.00	51.60
CPSII	Caucasian	548	75.40	49.60	538	75.20	48.50
Kentucky	Caucasian	1,038	64.00	50.70	1,134	65.70	50.70
MECC	Caucasian	1,605	71.50	48.80	1,318	72.90	48.50
Melbourne Cohort	Caucasian	539	69.90	48.60	469	70.00	48.40
Newfoundland	Caucasian	195	61.50	37.90	477	61.90	41.30
<b>Subtotal</b>		<b>7,563</b>			<b>6,328</b>		
<b>Total</b>		<b>16,984</b>			<b>18,262</b>		

<sup>a</sup> Final sample size used in statistical analysis.

<sup>b</sup> Mean age (years) of cases and controls.

**Supplementary Table 13: Risk allele frequencies in Europeans and East Asians from the HapMap Project**

Locus	SNP	Position (bp)	Risk Allele	Reference Allele	RAF based on data from HapMap		
					Build 36	CEU	CHB
10q22.3	rs704017	80489138	G	A	0.523	0.278	0.398
10q25.2	rs11196172	114716833	A	G	0.125	0.833	0.833
11q12.2	rs174537	61309256	G	T	0.655	0.624	0.681
	rs4246215	61320875	G	T	0.628	0.628	0.681
	rs174550	61328054	T	C	0.654	0.659	0.689
	rs1535	61354548	A	G	0.659	0.624	0.677
12p13.31	rs10849432	6255988	T	C	1.000	0.844	0.807
17p13.3	rs12603526	747343	C	T	0.013	0.234	0.335
19q13.2	rs1800469	46552136	G	A	0.712	0.526	0.544
	rs2241714	46561232	C	T	0.708	0.511	0.544

Abbreviations: RAF, risk allele frequency; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan.

**Supplementary Table 14: Functional annotation of SNPs correlated with newly identified risk variants ( $r^2 > 0.5$ ) using data from ENCODE**

SNP	Position <sup>a</sup>	Promoter histone marks <sup>b</sup>	Enhancer histone marks <sup>c</sup>	DNase <sup>d</sup>	Proteins bound <sup>e</sup>	Motifs changed <sup>f</sup>
<b>10q22.3</b>						
rs1249905	80481143		7 cell types	HSMMtube	BAF155,USF1	5 altered motifs
rs704017	80489138		HepG2, K562	7 cell types	23 bound proteins	
<b>10q25.2</b>						
rs11326257	114712127		K562	HRPEpiC		Spz1,TCF4
rs11196170	114712611		K562			5 altered motifs
rs6585195	114712944		K562			CEBPB,Mef2,THAP1
rs2296784	114713549		5 cell types		JUNB	
rs2296783	114713640		5 cell types	19 cell types	5 bound proteins	CHD2
rs2296782	114713847		5 cell types	14 cell types	7 bound proteins	12 altered motifs
rs11196171	114714463		NHLF, HMEC, HSMM			4 altered motifs
rs7068313	114715916		4 cell types			
rs10885399	114716167		4 cell types	9 cell types		AP-1,Pax-5
rs11196172	114716833		4 cell types			10 altered motifs
rs11196173	114717057		4 cell types	20 cell types	CFOS	
rs61875109	114719553			7 cell types	7 bound proteins	Pax-6
rs12573128	114720787					27 altered motifs
rs61875110	114721473					5 altered motifs
rs7897438	114722297					Zbtb3
rs7917983	114722872					Dmbx1,Ik-2
rs7901275	114722896					GCNF,HDAC2,Rad21
rs72826065	114725511					
rs7918749	114726460					Foxp1,Pou2f2,TATA
rs4297396	114734443		Huvec	BE2_C	ZNF263	7 altered motifs
<b>11q12.2</b>						
rs628993	61296267		H1, HepG2, Huvec	15 cell types		8 altered motifs
rs174528	61300075		H1, K562	5 cell types	CTCF	
rs174529	61300537		H1	Th1,pHTE	CTCF	Egr-1,Roaz,Zbtb3
rs174530	61303168			H1-hESC,LNCaP		SP1,ZBTB33,Zic
rs108499	61303813					E2F,ERalpha-a,Myb
rs509360	61305135		HMEC, NHEK	5 cell types	POL2,CMYC	GR,RFX5

**Supplementary Table 14 Continues**

**Supplementary Table 14 Continued**

<b>SNP</b>	<b>Position<sup>a</sup></b>	<b>Promoter histone marks<sup>b</sup></b>	<b>Enhancer histone marks<sup>c</sup></b>	<b>DNase<sup>d</sup></b>	<b>Proteins bound<sup>e</sup></b>	<b>Motifs changed<sup>f</sup></b>
rs174533	61305601			BE2_C,WERI-Rb-1		PU.1
rs174534	61306034			HRPEpiC	USF1	
rs174535	61307932			HSMM_emb	CTCF	Mrg
rs174536	61308503	HSMM		Caco-2		HEN1,ZBTB7A
rs174537	61309256			Osteobl		HNF4
rs102275	61314379					
rs102274	61314402					EWSR1-FLI1
rs174538	61316657	9 cell types		110 cell types	25 bound proteins	CTCF,Myf,Rad21
rs4246215	61320875		K562		CTCF	Nanog
rs174541	61322484		K562, NHLF	K562,WERI-Rb-1	POL2	5 altered motifs
rs174544	61324329		K562			Nanog,Nr2f2
rs174545	61325882		K562			4 altered motifs
rs174546	61326406		K562	K562,GM12892		BDP1,RXRA
rs174547	61327359		K562			AIRE,Egr-1
rs174548	61327924				POL2	AP-1
rs174549	61327958					Evi-1,ZBTB33
rs174550	61328054		Huvec	Th1	POL2,POL24H8	ERalpha-a
rs174551	61330260					8 altered motifs
rs174553	61331734					
rs72643557	61336003		K562			ERalpha-a,Smad4
rs174554	61336039		K562			4 altered motifs
rs174555	61336336		K562			Pou5f1
rs174556	61337211		K562		CEPB	Cpx, Rad21
rs174559	61338232		5 cell types		POL2,POL24H8	Evi-1,Osf2,SMC3
rs174560	61338340		5 cell types		POL2,POL24H8	Ascl2
rs174561	61339284	6 cell types	NHEK, HMEC, H1	94 cell types	16 bound proteins	7 altered motifs
rs174562	61341720	6 cell types	Huvec, HepG2, K562	4 cell types		Znf143
rs174564	61344881		7 cell types			4 altered motifs
rs28456	61346057		K562			ERalpha-a,TCF11::MafG,ZID
rs174565	61348212					9 altered motifs
rs57668028	61348574					10 altered motifs
rs174566	61348938		K562			

**Supplementary Table 14 Continues**

**Supplementary Table 14 Continued**

<b>SNP</b>	<b>Position<sup>a</sup></b>	<b>Promoter histone marks<sup>b</sup></b>	<b>Enhancer histone marks<sup>c</sup></b>	<b>DNase<sup>d</sup></b>	<b>Proteins bound<sup>e</sup></b>	<b>Motifs changed<sup>f</sup></b>
rs174567	61349581					HNF4,RXRA,VDR
rs174568	61350392		K562			
rs3834458	61351497	5 cell types	Huvec, HMEC	12 cell types	4 bound proteins	4 altered motifs
rs5792235	61352899	8 cell types		16 cell types	5 bound proteins	9 altered motifs
rs99780	61353209	8 cell types		15 cell types	POL2	4 altered motifs
rs174570	61353788	5 cell types	GM12878, Huvec, HSMM			GCNF
rs1535	61354548	HMEC, HepG2, NHLF	4 cell types	6 cell types	ERALPHA_A	AIRE,ERalpha-a,Nrf1
rs174574	61356918	GM12878	K562	4 cell types	POL2	4 altered motifs
rs174576	61360086		4 cell types	PANC-1		4 altered motifs
rs174577	61361390		K562			Zic
rs174578	61362075		K562	Medullo	STAT1,STAT2	4 altered motifs
rs174580	61363218		K562			4 altered motifs
rs174581	61363259		K562			VDR
rs174583	61366326		K562	pHTE	POL2	Nkx2
rs174584	61367326		K562			7 altered motifs
rs174592	61375184		K562	HA-h, HMVEC-dBl-Ad		6 altered motifs
rs174594	61376405			HConF		Nanog
rs72643559	61376850					Irf,TCF12
rs174598	61377770		K562			CTCF,NF-kappaB, Rad21
rs174599	61378132		K562	HSMMtube		RXRA
rs174600	61378803			HSMM	POL24H8	
rs174601	61379716					
rs97384	61380757				POL2	10 altered motifs
<b>12p13.31</b>						
rs12231078	6243834		4 cell types		PU1,KAP1	4 altered motifs
rs10849431	6246126		GM12878, HMEC	5 cell types		
rs12230499	6247228		GM12878, HMEC			ATF3,LXR,SREBP
rs4764547	6248473		H1	4 cell types		4 altered motifs
rs11064124	6251768		HMEC, NHEK	Urothelia,Caco-2	STAT3	13 altered motifs
rs10849432	6255988		H1, HepG2	LNCaP		STAT
rs7963944	6256186		HepG2			4 altered motifs
rs7964858	6256263		HepG2, HMEC, NHEK			Ets,NRSF,Znf143

**Supplementary Table 14 Continues**

**Supplementary Table 14 Continued**

<b>SNP</b>	<b>Position<sup>a</sup></b>	<b>Promoter histone marks<sup>b</sup></b>	<b>Enhancer histone marks<sup>c</sup></b>	<b>DNase<sup>d</sup></b>	<b>Proteins bound<sup>e</sup></b>	<b>Motifs changed<sup>f</sup></b>
rs4764551	6256597		HepG2, HMEC, NHEK	35 cell types	CTCF,RAD21,SMC3	Ets,Pax-5,STAT
rs4764552	6256779		HMEC, NHEK, HepG2	12 cell types	CEBPB,STAT3	RFX5
rs10713506	6257246		6 cell types			9 altered motifs
rs6489708	6257314		6 cell types			Foxp1,Sox,TATA
<b>17p13.3</b>						
rs497425	733798		7 cell types	4 cell types	EGR1	
rs2467260	734903		NHEK	Osteobl		BCL,GLI,NF-kappaB
rs552258	735146		NHEK			NRSF
rs610378	735208					Maf,NRSF
rs2663319	737553		NHLF	5 cell types		
rs2457270	738088		NHLF			CTCF,Pax-3,SREBP
rs2467254	738859		GM12878			4 altered motifs
rs4471744	740457					8 altered motifs
rs657243	740755			Chorion,HVMF,NHDF-neo		11 altered motifs
rs670996	741583			Th1		
rs548176	743122			HMEC,pHTE		Ascl2
rs547165	743255	HMEC		HMEC		Myc,Nrf1,RREB-1
rs2467259	744667			Caco-2		HDAC2
rs2457268	744817	HMEC, HSMM				Foxo
rs7406806	745038					Ets,Gfi1,Znf143
rs62068450	745200					ZBTB33
rs62068451	745299					10 altered motifs
rs2940809	745515					Mef2,RXRA,SRF
rs2955623	745864					DMRT 1,DMRT 2
rs57694836	746856		Huvec, NHLF, HSMM	PanIsletD		25 altered motifs
rs537122	747318	NHLF, Huvec	HSMM	9 cell types		EWSR1-FLI1,Pax-4,Sox
rs12603526	747343	NHLF, Huvec	HSMM	6 cell types		4 altered motifs
rs7225048	747664		Huvec			7 altered motifs
rs9897377	747683		Huvec			HDAC2
<b>19q13.2</b>						
rs12461895	46540187				LRH1	
rs10416269	46540688					5 altered motifs

**Supplementary Table 14 Continues**

**Supplementary Table 14 Continued**

<b>SNP</b>	<b>Position<sup>a</sup></b>	<b>Promoter histone marks<sup>b</sup></b>	<b>Enhancer histone marks<sup>c</sup></b>	<b>DNase<sup>d</sup></b>	<b>Proteins bound<sup>e</sup></b>	<b>Motifs changed<sup>f</sup></b>
rs7408955	46541040					22 altered motifs
rs1549934	46541798					20 altered motifs
rs11466329	46541949					HDAC2
rs4803455	46543349					Hoxb7,ST AT
rs10406816	46543556					4 altered motifs
rs1989457	46543955					
rs2288874	46544615					Myf,ZEB1
rs2288873	46544819			4 cell types		TEF-1,Zbtb3
rs2241717	46545892					4 altered motifs
rs8108632	46546374					15 altered motifs
rs6508976	46546514					
rs7258445	46547355	GM12878, Huvec	K562			Pax-4
rs2241715	46548726	9 cell types		5 cell types	11 bound proteins	4 altered motifs
rs12462166	46549244	9 cell types		53 cell types	15 bound proteins	30 altered motifs
rs12983775	46549279	9 cell types		15 cell types	17 bound proteins	10 altered motifs
rs1800470	46550761	9 cell types		17 cell types	4 bound proteins	8 altered motifs
rs1800469	46552136	9 cell types		4 cell types	CTCF,POL2	HNF4,Nkx2
rs4803457	46553199		K562	Hepatocytes		6 altered motifs
rs2317130	46553514		K562	HAc		GATA,TAL1
rs11666933	46553514					Pax-4
rs1982072	46556349			Medullo		Foxc1,Hoxa13
rs11670143	46556811			Melano,HAEpiC		AP-1,NF-AT 1,Zbtb3
rs4803458	46557133					
rs11083616	46557483					GR,Pax-4,Pou2f2
rs11083617	46558113					
rs2241714	46561232	6 cell types	K562, GM12878, NHEK			ZBRK1
rs2241713	46561308	6 cell types	K562, GM12878, NHEK			
rs2241712	46561596	8 cell types	NHEK	115 cell types	11 bound proteins	6 altered motifs
rs1549933	46562982	GM12878, HepG2	6 cell types			AP-1
rs1025497	46563094	K562	HepG2, NHLF, GM12878			5 altered motifs
rs1963413	46563413		GM12878, HepG2, K562			Maf,NRSF,SETDB1
rs9797885	46564841		8 cell types			5 altered motifs

**Supplementary Table 14 Continues**

**Supplementary Table 14 Continued**

<b>SNP</b>	<b>Position<sup>a</sup></b>	<b>Promoter histone marks<sup>b</sup></b>	<b>Enhancer histone marks<sup>c</sup></b>	<b>DNase<sup>d</sup></b>	<b>Proteins bound<sup>e</sup></b>	<b>Motifs changed<sup>f</sup></b>
rs8103493	46566179		8 cell types			21 altered motifs
rs4803459	46568308					4 altered motifs
rs8108357	46570974			WERI-Rb-1		
rs34503210	46571623		5 cell types		ZNF263	Irf
rs7260340	46572366		5 cell types			CCNT2,p300
rs9710214	46573062		GM12878, K562			EBF
rs11670739	46573385		GM12878	HAEpiC,HVMF		Pax-4
rs57948035	46573891	5 cell types	GM12878, NHLF			12 altered motifs
rs1046909	46574552	8 cell types	K562	4 cell types	4 bound proteins	

Abbreviations: ENCODE, the Encyclopedia of DNA Elements; Chr., chromosome.

<sup>a</sup> The chromosome position (bp) is based on NCBI, Build 36.

<sup>b</sup> Evidence of local H3K4Me1 and H3K27Ac modification (cell lines/types: if >3, only the number is included).

<sup>c</sup> Evidence of local H3K4Me3 modification (cell lines/types: if >3, only the number is included).

<sup>d</sup> Evidence of chromatin hypersensitivity to DNase (cell lines/types: if >3, only the number is included).

<sup>e</sup> ChIP-seq experiments indicate alteration in binding of transcription factor ( if >3, only the number is included).

<sup>f</sup> Evidence of alteration in regulatory motif ( if >3, only the number is included).

**Supplementary Table 15: Analyses of expression quantitative trait locus (eQTL) in newly identified loci associated with CRC**

Locus	SNP	Gene <sup>a</sup>	Position <sup>b</sup>	Blood eQTL browser			eQTL Browser		
				Top eQTL ( $r^2$ ) <sup>c</sup>	Gene (Tissue)	P	Top eQTL ( $r^2$ )	Gene (Tissue)	P
11q12.2	rs174537	<i>MYRF</i>	61309256	rs174537 (1.00)	<i>TMEM258</i> (blood)	$1.59 \times 10^{-56}$	rs174548 (1.00)	<i>FADS1</i> (liver)	$1.52 \times 10^{-4}$
	rs4246215	<i>FEN1</i>	61320875	rs174537 (1.00)	<i>FADS1</i> (blood)	$1.80 \times 10^{-31}$	rs174548 (1.00)	<i>FADS2</i> (monocytes)	$3.24 \times 10^{-42}$
	rs174550	<i>FADS1</i>	61328054	rs174537 (1.00)	<i>FADS2</i> (blood)	$1.64 \times 10^{-14}$	rs174548 (1.00)	<i>FADS3</i> (liver)	$1.60 \times 10^{-9}$
	rs1535	<i>FADS2</i>	61354548						
12p13.31	rs10849432	<i>CD9</i>	6255988	rs4764552 (1.00)	<i>LTBR</i> (blood)	0.00315			
19q13.2	rs1800469	<i>TGFB1</i>	46552136	rs1800469 (1.00)	<i>B9D2</i> (blood)	$6.46 \times 10^{-12}$	rs2241714 (1.00)	<i>B3GNT8</i> (monocytes)	$4.39 \times 10^{-21}$
	rs2241714	<i>B9D2</i>	46561232	rs1800469 (1.00)	<i>BCKDHA</i> (blood)	$1.88 \times 10^{-8}$			
				rs1800469 (1.00)	<i>B3GNT8</i> (blood)	$1.97 \times 10^{-4}$			
				rs1800469 (1.00)	<i>HNRNPUL1</i> (blood)	$2.97 \times 10^{-4}$			
				rs1800469 (1.00)	<i>CYP2S1</i> (blood)	0.00229			
GTEx									
Locus	SNP	Gene <sup>a</sup>	Position <sup>b</sup>	GTEx			MuTHER		
				Top eQTL ( $r^2$ )	Gene (Tissue)	P	Top eQTL ( $r^2$ )	Gene (Tissue)	P
11q12.2	rs174537	<i>MYRF</i>	61309256	rs174537 (1.00)	<i>FADS2</i> (nerve)	$5.58 \times 10^{-6}$	rs174537 (1.00)	<i>TMEM258</i> (adipose)	0.0025
	rs4246215	<i>FEN1</i>	61320875	rs174537 (1.00)	<i>FADS2</i> (muscle)	$8.36 \times 10^{-6}$	rs174537 (1.00)	<i>TMEM258</i> (LCLs)	$1.04 \times 10^{-6}$
	rs174550	<i>FADS1</i>	61328054	rs174537 (1.00)	<i>FADS2</i> (blood)	$3.36 \times 10^{-9}$	rs174537 (1.00)	<i>TMEM258</i> (Skin)	$4.93 \times 10^{-4}$
	rs1535	<i>FADS2</i>	61354548	rs174537 (1.00)	<i>FADS2</i> (heart)	$7.09 \times 10^{-7}$	rs174537 (1.00)	<i>FADS1</i> (adipose)	$1.35 \times 10^{-5}$
19q13.2	rs1800469	<i>TGFB1</i>	46552136	rs1800469 (1.00)	<i>TGFB1</i> (adipose)	$9.00 \times 10^{-4}$			
	rs2241714	<i>B9D2</i>	46561232	rs2241714 (1.00)	<i>EGLN2</i> (muscle)	$6.80 \times 10^{-5}$			
				rs2241714 (1.00)	<i>B3GNT8</i> (blood)	$8.72 \times 10^{-5}$			
				rs2241714 (1.00)	<i>TGFB1</i> (adipose)	$6.00 \times 10^{-4}$			

Abbreviations: Chr., chromosome; LCLs, lymphoblastoid cell lines.

<sup>a</sup> The closest gene(s).

<sup>b</sup> The chromosome position (bp) is based on NCBI Build 36.

<sup>c</sup> LD ( $r^2$ ) between the eQTL and index SNP.

**Supplementary Table 16: Expression levels of genes located in the newly identified loci in colon tumor tissue and normal colon tissue using data from TCGA**

Locus	Gene	RefGene ID	Expression levels (RPKM)		
			Colon tumor tissue	Normal colon tissue	P
10q22.3	<i>ZMIZ1</i>	57178	8.656	14.281	$3.28 \times 10^{-6}$
10q25.2	<i>TCF7L2</i>	6934	9.982	14.397	$1.14 \times 10^{-4}$
11q12.2	<i>FEN1</i>	2237	18.761	9.253	$3.28 \times 10^{-6}$
	<i>FADS1</i>	3992	4.427	2.078	0.0067
	<i>FADS2</i>	9415	12.889	4.157	0.0098
12p13.31	<i>CD9</i>	928	373.429	430.685	0.025
	<i>PLEKHG6</i>	55200	14.976	22.202	$2.55 \times 10^{-4}$
	<i>TNFRSF1A</i>	7132	38.815	61.716	$4.89 \times 10^{-10}$
17p13.3	<i>NXN</i>	64359	4.318	12.234	$2.83 \times 10^{-5}$
19q13.2	<i>TGFB1</i>	7040	16.524	12.184	0.529
	<i>B9D2</i>	80776	5.129	5.072	0.535

Abbreviations: TCGA, The Cancer Genome Atlas; RPKM, per kilobase of exon per million mapped reads value.

**Supplementary Table 17: Associations of 31 risk variants in previously reported susceptibility loci with CRC in East Asians**

Locus	SNP <sup>a</sup>	Gene <sup>b</sup>	Annotation	Position <sup>c</sup>	Alleles <sup>d</sup>	East Asian combined			Published GWAS			P <sub>heterogeneity</sub> <sup>g</sup>
						RAF <sup>e</sup>	OR (95% CI)	P	RAF <sup>f</sup>	OR (95% CI) <sup>f</sup>	P <sup>f</sup>	
1q41	rs6691170	DUSP10	Intergenic	220112069	T/G	NA	NA	NA	0.36	1.06 (1.03-1.09)	9.55×10 <sup>-10</sup>	NA
1q41	rs6687758	DUSP10	Intergenic	220231571	<b>G/A</b>	0.24	1.12 (1.08-1.17)	8.99×10 <sup>-9</sup>	0.20	1.09 (1.06-1.12)	2.27×10 <sup>-9</sup>	0.232
2q32.3	rs11903757	NABP1	Intergenic	192295449	<b>C/T</b>	0.05	1.15 (1.03-1.28)	0.01	0.16	1.16 (1.10-1.22)	3.71×10 <sup>-8</sup>	0.892
3q26.2	rs10936599	MYNN	Exon 2	170974795	<b>C/T</b>	0.39	1.05 (1.01-1.08)	0.01	0.75	1.08 (1.05-1.10)	3.39×10 <sup>-8</sup>	0.221
5q31.1	rs647161	PITX1	Intergenic	134526991	<b>A/C</b>	0.31	1.15 (1.11-1.19)	1.87×10 <sup>-14</sup>	0.31	1.17 (1.11-1.22)	3.77×10 <sup>-10</sup>	0.512
6p21.31	rs1321311	CDKN1A	Intergenic	36730878	<b>A/C</b>	0.14	1.09 (1.03-1.15)	0.001	0.23	1.10 (1.07-1.13)	1.14×10 <sup>-10</sup>	0.769
6q25.3	rs7758229	SLC22A3	Intron 5	160760242	<b>T/G</b>	0.23	1.02 (0.98-1.06)	0.26	0.22	1.28 (1.18-1.39)	7.92×10 <sup>-9</sup>	1.43×10 <sup>-6</sup>
8q23.3	rs16892766	EIF3H	Intergenic	117699864	<b>C/A</b>	NA	NA	NA	0.07	1.25 (1.19-1.32)	3.30×10 <sup>-18</sup>	NA
8q24.21	rs10505477	Unknown	Intergenic	128476625	<b>A/G</b>	0.38	1.15 (1.11-1.20)	3.43×10 <sup>-13</sup>	0.51	1.17 (1.12-1.23)	3.16×10 <sup>-11</sup>	0.641
8q24.21	rs6983267	Unknown	Intergenic	128482487	<b>G/T</b>	0.38	1.14 (1.10-1.18)	4.85×10 <sup>-14</sup>	0.52	1.21 (1.15-1.27)	1.27×10 <sup>-14</sup>	0.057
8q24.21	rs7014346	Unknown	Intergenic	128493974	<b>A/G</b>	0.27	1.13 (1.08-1.17)	1.96×10 <sup>-8</sup>	0.37	1.19 (1.14-1.24)	8.60×10 <sup>-26</sup>	0.062
10p14	rs10795668	Unknown	Intergenic	8741225	<b>G/A</b>	0.60	1.15 (1.11-1.19)	4.91×10 <sup>-15</sup>	0.67	1.12 (1.09-1.16)	2.50×10 <sup>-13</sup>	0.297
11q13.4	rs3824999	POLD3	Intron 9	74023198	<b>G/T</b>	0.40	1.06 (1.02-1.11)	0.002	0.50	1.08 (1.05-1.10)	3.65×10 <sup>-10</sup>	0.543
11q23.1	rs3802842	Unknown	Intergenic	110676919	<b>C/A</b>	0.38	1.09 (1.05-1.12)	2.57×10 <sup>-7</sup>	0.29	1.11 (1.08-1.15)	5.82×10 <sup>-10</sup>	0.371
12p13.32	rs10774214	CCND2	Intergenic	4238613	<b>T/C</b>	0.37	1.14 (1.09-1.18)	1.40×10 <sup>-10</sup>	0.35	1.17 (1.11-1.23)	5.48×10 <sup>-10</sup>	0.385
12q13.13	rs7136702	LARP4	Intergenic	49166483	<b>T/C</b>	0.51	1.02 (0.98-1.06)	0.31	0.35	1.06 (1.04-1.08)	4.02×10 <sup>-8</sup>	0.052
12q13.13	rs11169552	ATF1	Intergenic	49441930	<b>C/T</b>	0.65	1.05 (1.01-1.09)	0.01	0.72	1.09 (1.06-1.12)	1.89×10 <sup>-10</sup>	0.112
14q22.2	rs4444235	BMP4	Intergenic	53480669	<b>C/T</b>	0.53	1.04 (1.01-1.08)	0.02	0.46	1.11 (1.08-1.15)	8.10×10 <sup>-10</sup>	0.007
14q22.2	rs1957636	BMP4	Intergenic	53629768	<b>T/C</b>	0.62	0.99 (0.95-1.04)	0.77	0.40	1.08 (1.06-1.11)	1.36×10 <sup>-9</sup>	0.001
15q13.3	rs16969681	SCG5	Intergenic	30780403	<b>T/C</b>	0.44	1.07 (1.03-1.12)	0.002	0.09	1.18 (1.11-1.25)	5.33×10 <sup>-8</sup>	0.013
15q13.3	rs4779584	SCG5	Intergenic	30782048	<b>T/C</b>	0.82	1.06 (1.01-1.11)	0.01	0.18	1.26 (1.19-1.34)	4.44×10 <sup>-14</sup>	5.48×10 <sup>-6</sup>
15q13.3	rs11632715	GREM1	Intergenic	30791539	<b>A/G</b>	0.81	0.95 (0.90-1.01)	0.11	0.47	1.12 (1.08-1.16)	2.30×10 <sup>-10</sup>	4.05×10 <sup>-6</sup>
16q22.1	rs9929218	CDH1	Intron 2	67378447	<b>G/A</b>	0.81	1.06 (1.00-1.11)	0.03	0.71	1.10 (1.07-1.13)	1.20×10 <sup>-8</sup>	0.193
18q21.1	rs7229639	SMAD7	Intron 3	44704974	<b>A/G</b>	0.16	1.20 (1.16-1.25)	3.05×10 <sup>-15</sup>	0.15	1.22 (1.15-1.29)	2.93×10 <sup>-11</sup>	0.720
18q21.1	rs4939827	SMAD7	Intron 3	44707461	<b>T/C</b>	0.24	1.12 (1.08-1.16)	1.53×10 <sup>-8</sup>	0.52	1.18 (1.12-1.23)	1.00×10 <sup>-12</sup>	0.109
19q13.11	rs10411210	RHPN2	Intron 2	38224140	<b>C/T</b>	0.82	1.12 (1.07-1.17)	3.14×10 <sup>-6</sup>	0.90	1.15 (1.10-1.20)	4.60×10 <sup>-9</sup>	0.389
20p12.3	rs961253	BMP2	Intergenic	6352281	<b>A/C</b>	0.09	1.10 (1.04-1.17)	7.74×10 <sup>-4</sup>	0.36	1.12 (1.08-1.16)	2.00×10 <sup>-10</sup>	0.662
20p12.3	rs4813802	BMP2	Intergenic	6647595	<b>G/T</b>	0.21	1.12 (1.06-1.17)	9.87×10 <sup>-6</sup>	0.36	1.09 (1.16-1.12)	7.52×10 <sup>-11</sup>	0.372
20p12.3	rs2423279	HAO1	Intergenic	7760350	<b>C/T</b>	0.31	1.13 (1.09-1.17)	3.04×10 <sup>-12</sup>	0.30	1.14 (1.08-1.19)	2.29×10 <sup>-7</sup>	0.859
20q13.33	rs4925386	LAMA5	Intron 10	60354439	<b>C/T</b>	0.77	1.05 (1.01-1.10)	0.01	0.68	1.08 (1.05-1.10)	1.89×10 <sup>-10</sup>	0.378
Xp22.2	rs5934683	SHROOM2	Intergenic	9711474	<b>T/C</b>	NA	NA	NA	0.33	1.07 (1.04-1.10)	7.30×10 <sup>-10</sup>	NA

Abbreviations: GWAS, genome-wide association study; RAF, risk allele frequency; OR, odds ratio; CI, confidence interval; NA, not available.

<sup>a</sup> SNP rs7758229 was initially associated with distal colon cancer in East Asian populations.

<sup>b</sup> The closest gene(s).

<sup>c</sup> The chromosome position (bp) is based on NCBI, Build 36.

<sup>d</sup> Risk/reference alleles (in published GWAS) are based on forward allele coding in NCBI, Build 36. OR was estimated for the risk allele (bold).

<sup>e</sup> RAF in controls.

<sup>f</sup> Results (RAF, ORs, and 95% CIs) from the original studies (ref. 7-20).

<sup>g</sup> P for heterogeneity between this study and published studies was calculated using a Cochran's Q test.

**Supplementary Table 18: Contribution of CRC risk variants identified to date to the familial relative risk in East Asians**

SNP	Alleles <sup>a</sup>	Per-allele OR <sup>b</sup>	p <sup>c</sup>	q <sup>c</sup>	Lambda <sup>d</sup>	FRR <sup>e</sup>
<b>Risk variants in new loci identified in East Asians</b>						
rs 704017	G/A	1.10	0.32	0.68	1.00204	0.00259
rs 11196172	A/G	1.14	0.68	0.32	1.00356	0.00450
rs 174537	G/T	1.16	0.59	0.41	1.00517	0.00654
rs 10849432	T/C	1.14	0.82	0.18	1.00233	0.00295
rs 12603526	C/T	1.10	0.30	0.70	1.00198	0.00251
rs 1800469	G/A	1.09	0.48	0.52	1.00186	0.00235
<b>Subtotal</b>						<b>0.02144</b>
<b>Risk variants in previously reported loci initially identified in East Asians</b>						
rs 647161	A/C	1.15	0.31	0.69	1.00439	0.00556
rs 10774214	T/C	1.14	0.37	0.63	1.00413	0.00523
rs 2423279	C/T	1.13	0.31	0.69	1.00334	0.00423
rs 7229639	A/G	1.20	0.15	0.85	1.00481	0.00608
<b>Subtotal</b>						<b>0.02110</b>
<b>Risk variants in previously reported loci initially identified in European descendants</b>						
rs 6687758	G/A	1.12	0.24	0.76	1.00248	0.00314
rs 11903757	C/T	1.15	0.05	0.95	1.00105	0.00133
rs 10936599	C/T	1.05	0.39	0.61	1.00057	0.00073
rs 1321311	A/C	1.09	0.14	0.86	1.00095	0.00121
rs 6983267	G/T	1.14	0.38	0.62	1.00416	0.00527
rs 10795668	G/A	1.15	0.60	0.40	1.00455	0.00575
rs 3824999	G/T	1.06	0.40	0.60	1.00082	0.00104
rs 3802842	C/A	1.09	0.38	0.62	1.00178	0.00226
rs 11169552	C/T	1.05	0.65	0.35	1.00053	0.00068
rs 4444235	C/T	1.04	0.53	0.47	1.00038	0.00048
rs 16969681	T/C	1.07	0.44	0.56	1.00114	0.00144
rs 4779584	T/C	1.06	0.82	0.18	1.00048	0.00061
rs 9929218	G/A	1.06	0.81	0.19	1.00050	0.00064
rs 4939827	T/C	1.12	0.24	0.76	1.00248	0.00314
rs 10411210	C/T	1.12	0.82	0.18	1.00176	0.00223
rs 961253	A/C	1.10	0.09	0.91	1.00080	0.00102
rs 4813802	G/T	1.12	0.21	0.79	1.00227	0.00288
rs 4925386	C/T	1.05	0.77	0.23	1.00041	0.00052
<b>Subtotal</b>						<b>0.03438</b>
<b>Total</b>						<b>0.07693</b>

<sup>a</sup> Risk/reference alleles.

<sup>b</sup> Odds ratios calculated based on risk allele.

<sup>c</sup> p is the frequency of risk allele and q=(1-p) is the frequency of the normal allele.

<sup>d</sup> lambda=(p\*OR<sup>2</sup>+q)/(p\*OR+q)<sup>2</sup>.

<sup>e</sup> Familial relative risk calculated based on λo=2.2 (ref. 79).

**Supplementary Table 19: Contribution of CRC risk variants identified to date to the familial relative risk in European descendants**

SNP	Alleles <sup>a</sup>	Per-allele OR <sup>b</sup>	p <sup>c</sup>	q <sup>c</sup>	Lambda <sup>d</sup>	FRR <sup>e</sup>
<b>Risk variants in new loci identified in East Asians</b>						
rs704017	G/A	1.06	0.57	0.43	1.00082	0.00105
rs174537	G/T	1.07	0.67	0.33	1.00099	0.00125
<b>Subtotal</b>						<b>0.00230</b>
<b>Risk variants in previously reported loci initially identified in East Asians</b>						
rs647161	A/C	1.07	0.38	0.62	1.00110	0.00139
rs10774214	T/C	1.04	0.67	0.33	1.00034	0.00043
rs2423279	C/T	1.07	0.25	0.75	1.00089	0.00113
<b>Subtotal</b>						<b>0.00294</b>
<b>Risk variants in previously reported loci initially identified in European descendants</b>						
rs6691170	T/G	1.06	0.36	0.64	1.00079	0.00101
rs6687758	G/A	1.09	0.20	0.80	1.00125	0.00159
rs11903757	C/T	1.16	0.16	0.84	1.00327	0.00414
rs10936599	C/T	1.08	0.75	0.25	1.00107	0.00135
rs1321311	A/C	1.10	0.23	0.77	1.00169	0.00214
rs16892766	C/A	1.25	0.07	0.93	1.00393	0.00498
rs6983267	G/T	1.21	0.52	0.48	1.00895	0.01130
rs10795668	G/A	1.12	0.67	0.33	1.00273	0.00346
rs3824999	G/T	1.08	0.50	0.50	1.00148	0.00188
rs3802842	C/A	1.11	0.29	0.71	1.00234	0.00296
rs7136702	T/C	1.06	0.35	0.65	1.00079	0.00100
rs11169552	C/T	1.09	0.72	0.28	1.00144	0.00183
rs4444235	C/T	1.11	0.46	0.54	1.00272	0.00345
rs1957636	T/C	1.08	0.40	0.60	1.00144	0.00183
rs16969681	T/C	1.18	0.09	0.91	1.00257	0.00326
rs4779584	T/C	1.26	0.18	0.82	1.00911	0.01150
rs11632715	A/G	1.12	0.47	0.53	1.00321	0.00407
rs9929218	G/A	1.10	0.71	0.29	1.00180	0.00228
rs4939827	T/C	1.18	0.52	0.48	1.00676	0.00855
rs10411210	C/T	1.15	0.90	0.10	1.00157	0.00199
rs961253	A/C	1.12	0.36	0.64	1.00305	0.00386
rs4813802	G/T	1.09	0.36	0.64	1.00175	0.00222
rs4925386	C/T	1.08	0.68	0.32	1.00125	0.00159
rs5934683	T/C	1.07	0.33	0.67	1.00104	0.00131
<b>Subtotal</b>						<b>0.08645</b>
<b>Total</b>						<b>0.08875</b>

<sup>a</sup> Risk/reference alleles.

<sup>b</sup> Odds ratios calculated based on risk allele.

<sup>c</sup> p is the frequency of risk allele and q=(1-p) is the frequency of the normal allele.

<sup>d</sup> lambda=(p\*OR<sup>2</sup>+q)/(p\*OR+q)<sup>2</sup>.

<sup>e</sup> Familial relative risk calculated based on λo=2.2 (ref. 79).

**Supplementary Table 20: Concordance between imputed and genotyped data**

SNP (alleles, genotypes) <sup>a</sup>	Concordance rate			
	AA	AB	BB	Overall
rs704017 ( <b>G/A, GG/AG/AA</b> )	0.8963	0.9587	0.9703	0.9600
rs11196172 ( <b>A/G, AA/AG/GG</b> )	0.9860	0.9964	1.0000	0.9939
rs174537 ( <b>G/T, GG/GT/TT</b> )	0.9954	0.9961	0.9853	0.9939
rs4246215 ( <b>G/T, GG/GT/TT</b> )	1.0000	0.9992	1.0000	0.9996
rs174550 ( <b>T/C, TT/CT/CC</b> )	1.0000	0.9992	1.0000	0.9996
rs1535 ( <b>A/G, AA/AG/GG</b> )	1.0000	0.9993	1.0000	0.9996
rs10849432 ( <b>T/C, TT/CT/CC</b> )	0.9801	0.9696	0.9531	0.9759
rs12603526 ( <b>C/T, CC/CT/TT</b> )	0.9947	0.9968	0.9992	0.9979
rs1800469 ( <b>G/A, GG/AG/AA</b> )	0.9937	0.9962	0.9970	0.9958
rs2241714 ( <b>C/T, CC/CT/TT</b> )	0.9970	0.9993	1.0000	0.9989

<sup>a</sup> Risk/reference alleles of SNPs as shown in Table 1.

## SUPPLEMENTARY FIGURES

**Supplementary Figure 1: Plots of the first two principal components in the five studies included in stage 1.** The six plots are (a) Shanghai-1, (b) Shanghai-2, (c) Guangzhou-1, (d) Aichi-1, (e) KCPS-II and (f) meta-analysis.

**Supplementary Figure 2: Q-Q plots of *P* values for the association of SNPs with CRC risk using data in stage 1.** The eight plots are (a) Shanghai-1 with additional controls ( $\lambda = 1.031142$ ), (b) Shanghai-1 without additional controls ( $\lambda = 1.02733$ ), (c) Shanghai-2 with additional controls ( $\lambda = 1.010774$ ), (d) Shanghai-2 without additional controls ( $\lambda = 1.033529$ ), (e) Guangzhou-1 ( $\lambda = 1.018318$ ), (f) Aichi-1 ( $\lambda = 1.036879$ ), (g) KCPS-II ( $\lambda = 1.0169$ ) and (h) meta-analysis ( $\lambda = 1.0368$ ;  $\lambda_{1000} = 1.0118$ ).

**Supplementary Figure 3: Q-Q plot of *P* values for the association of SNPs with CRC risk in meta-analysis of stage 2 data ( $\lambda = 1.0525$ ;  $\lambda_{1000} = 1.0116$ ).**

**Supplementary Figure 4: Regional association plots of the six newly identified loci.** The six plots represent (a) 10q22.3, (b) 10q25.2, (c) 11q12.2, (d) 12p13.31, (e) 17p13.3 and (f) 19q13.2. For each plot, the  $-\log_{10}$  (*P* values) (y axis) of SNPs are shown according to their chromosomal positions (x axis) in NCBI Build 36. Blue lines represent the estimated recombination rates from the HapMap Project (NCBI Build 36). Arrows indicate genomic locations of genes within the 1-mb regions centered on the index SNPs in the NCBI Build 36 human assembly. The color of SNPs represents their LD ( $r^2$ , HapMap Asian) with the index SNP at each locus. With the exception of the index SNPs, which are shown as purple diamonds for stage 1 and purple circles for the meta-analyses of all studies, data shown for all other SNPs are from stage 1 only.

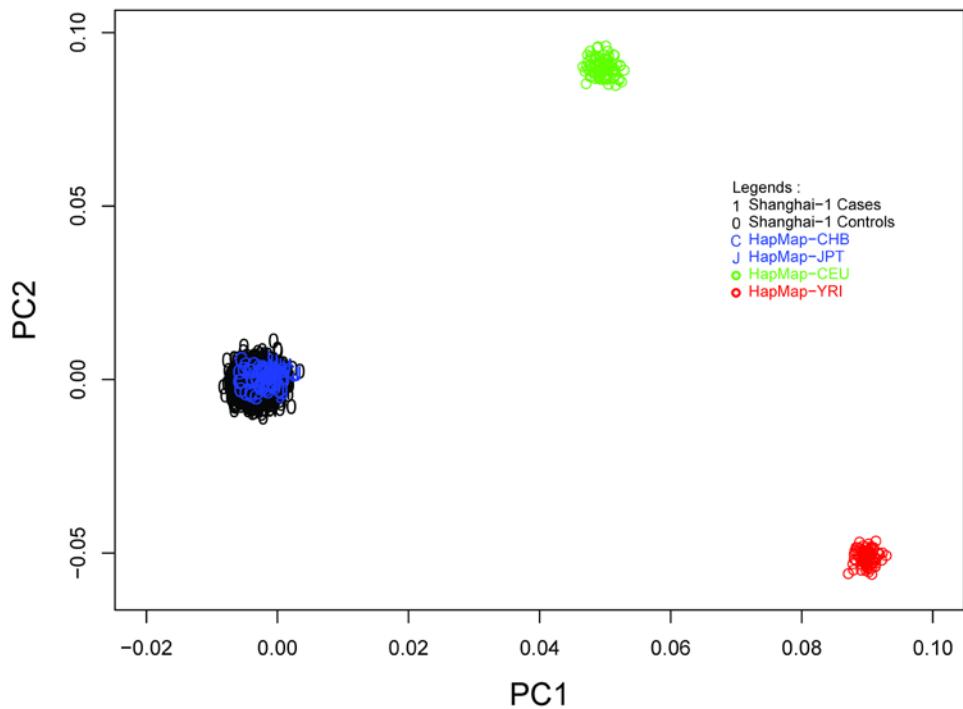
**Supplementary Figure 5: LD patterns for six newly identified loci in Asian and European populations using data from the 1000 Genomes Project.** The 12 plots represent (a) rs704017 (CHBJPT), (b) rs704017 (CEU), (c) rs11196172 (CHBJPT), (d) rs11196172 (CEU), (e) rs174537 (CHBJPT), (f) rs174537 (CEU), (g) rs10849432 (CHBJPT), (h) rs10849432 (CEU), (i) rs12603526 (CHBJPT), (j) rs12603526 (CEU), (k) rs1800469 (CHBJPT), and (l) rs1800469 (CEU).

**Supplementary Figure 6: Evidence from ENCODE data for regulatory function of SNPs in the six newly identified loci using the UCSC Genome Browser (see PDF file).** The six plots represent (a) 10q22.3, (b) 10q25.2, (c) 11q12.2, (d) 12p13.31, (e) 17p13.3 and (f) 19q13.2, within a 500-kb window centered on rs704017, rs11196172, rs174537, rs10849432, rs12603526 and rs1800469, respectively. Tracks (from top to bottom) in each of the plots are Genome Base Position, Chromosome Bands, UCSC Genes, Simple Nucleotide Polymorphisms (dbSNP build 130), LD structure (HapMap 2, CHB+JPT), ENCODE Enhancer- and Promoter-Associated Histone Mark (H3K4Me1) on 8 Cell Lines, ENCODE Promoter-Associated Histone Mark (H3K4Me3) on 9 Cell Lines, ENCODE Digital DNaseI Hypersensitivity Clusters, ENCODE Transcription Levels Assayed by RNA-seq on 6 Cell Lines, ENCODE Enhancer- and Promoter-

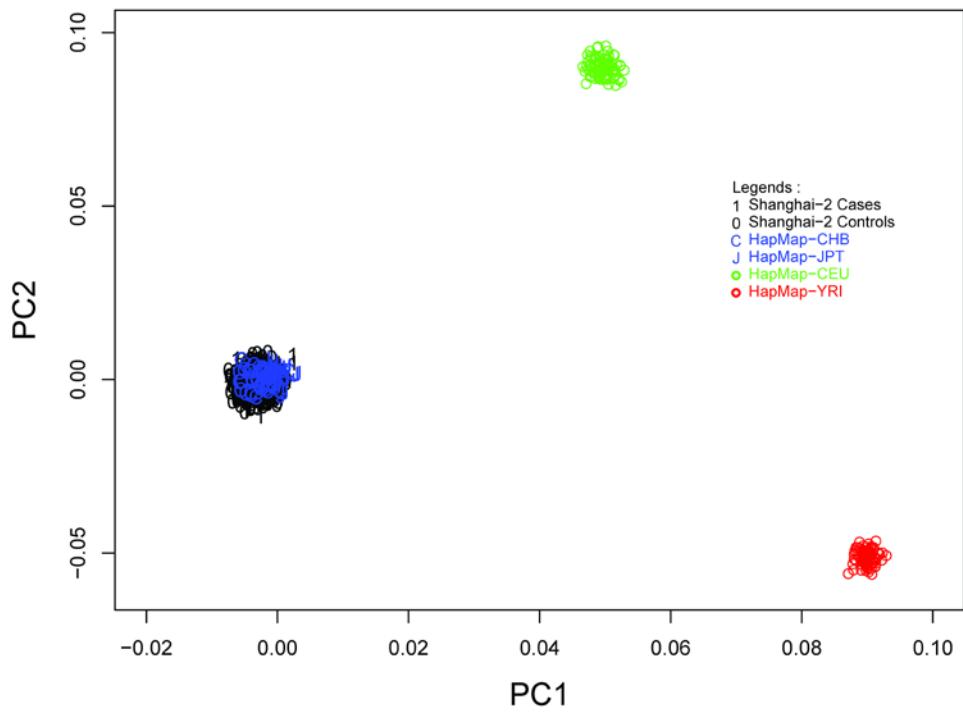
Associated Histone Mark (H3K27Ac) on 8 Cell Lines, ENCODE Broad Chromatin State Segmentation by HMM on 9 Cell Lines (bright red, active promoter; light red, weak promoter; purple, inactive/poised promoter; orange, strong enhancer; yellow, weak/poised enhancer; blue, insulator; dark green, transcriptional transition/elongation; light green, weak transcribed; gray, polycomb-repressed; light gray, heterochromatin/low signal/repetitive/copy number variation).

**Supplementary Figure 1: Plots of the first two principal components in the five studies included in stage 1.** The six plots are (a) Shanghai-1, (b) Shanghai-2, (c) Guangzhou-1, (d) Aichi-1, (e) KCPS-II and (f) meta-analysis.

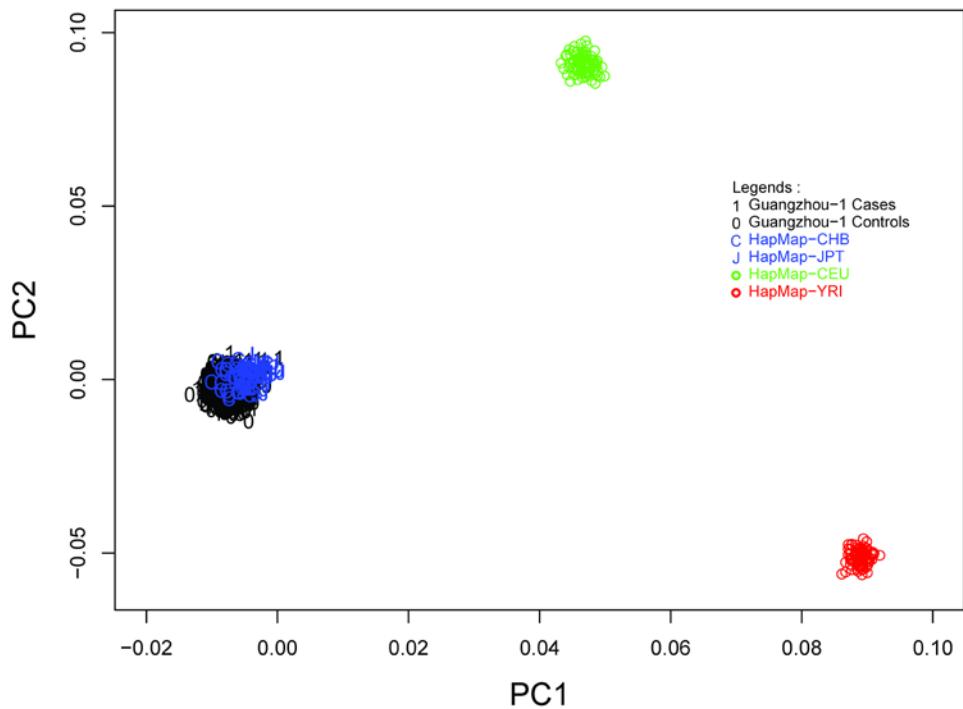
**(a)**



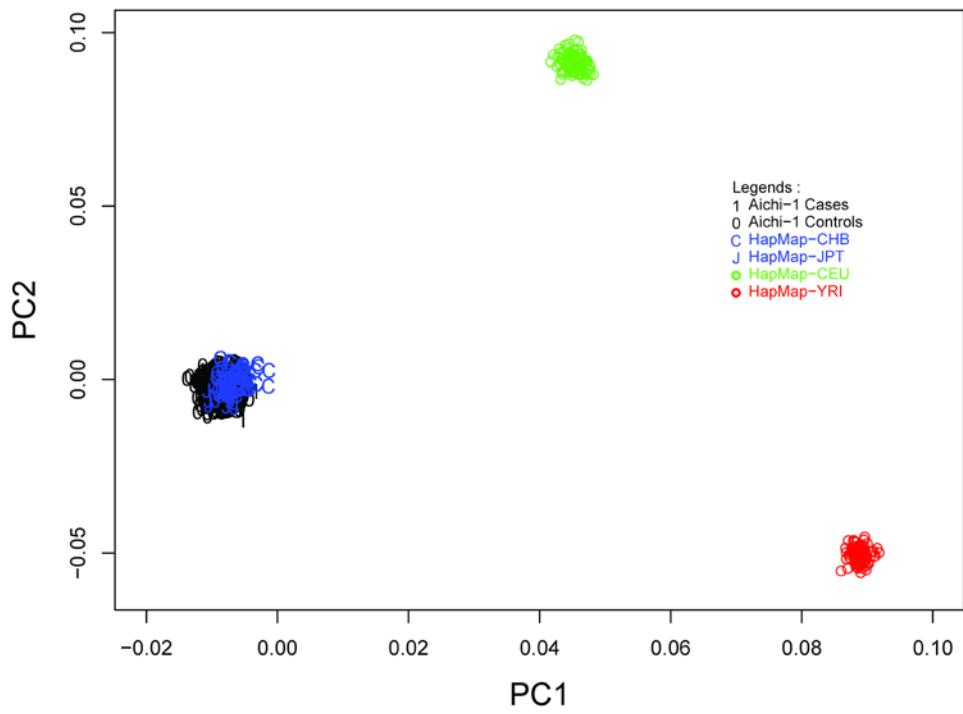
**(b)**



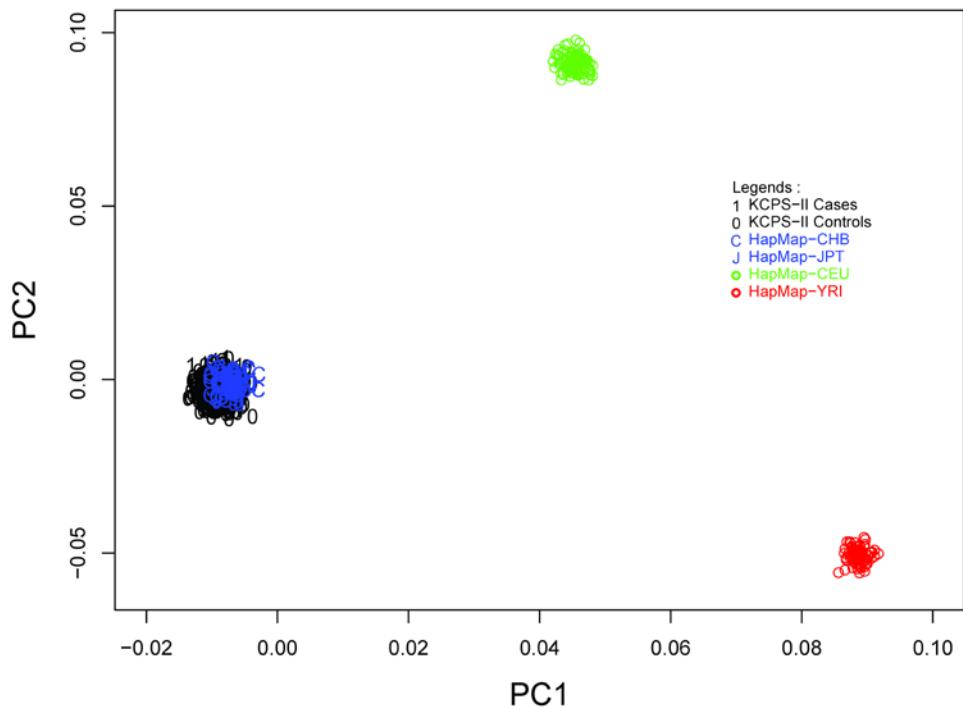
**(c)**



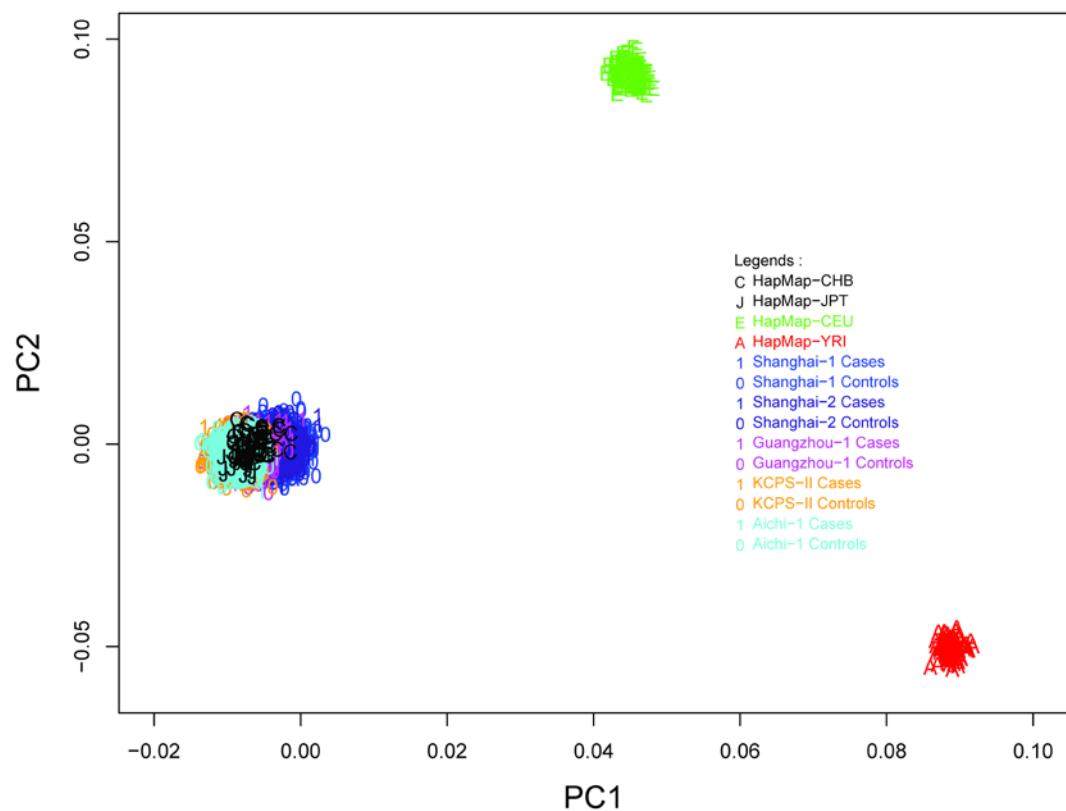
**(d)**



(e)

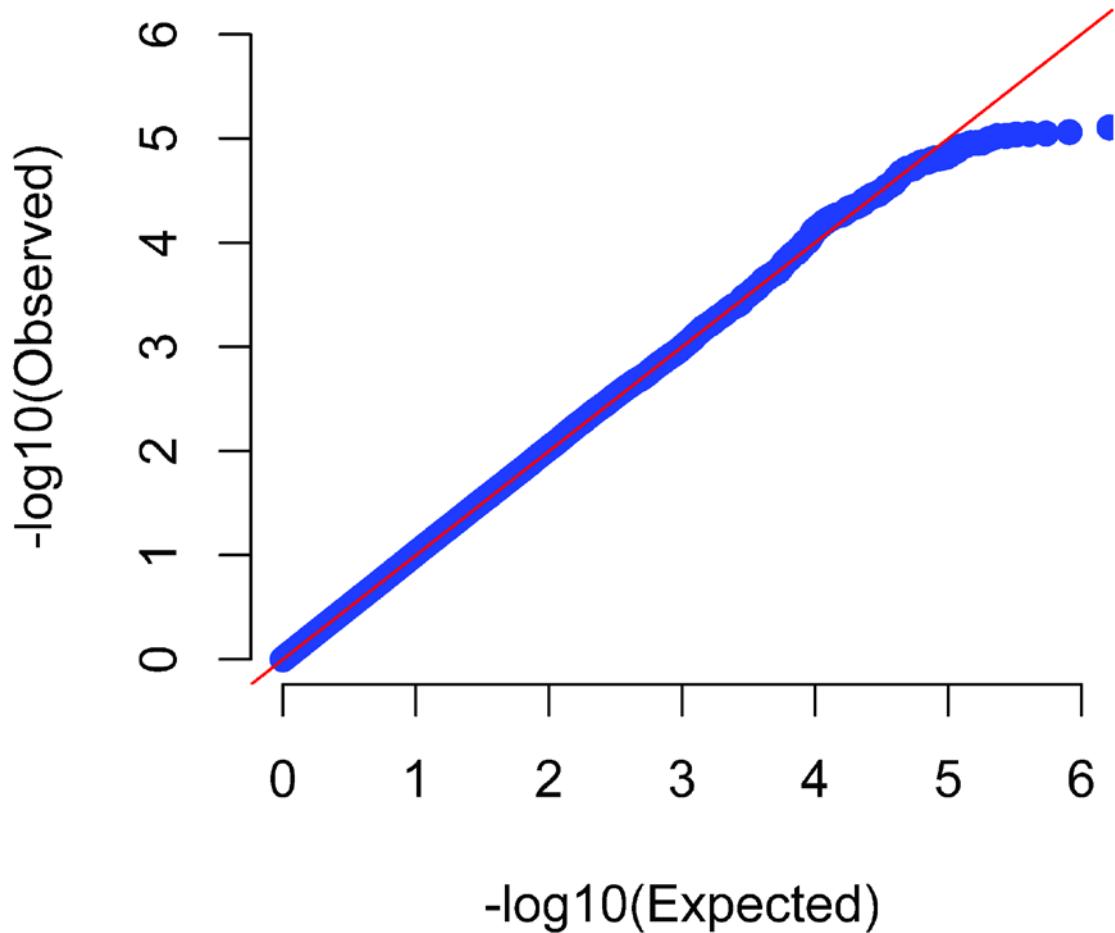


(f)

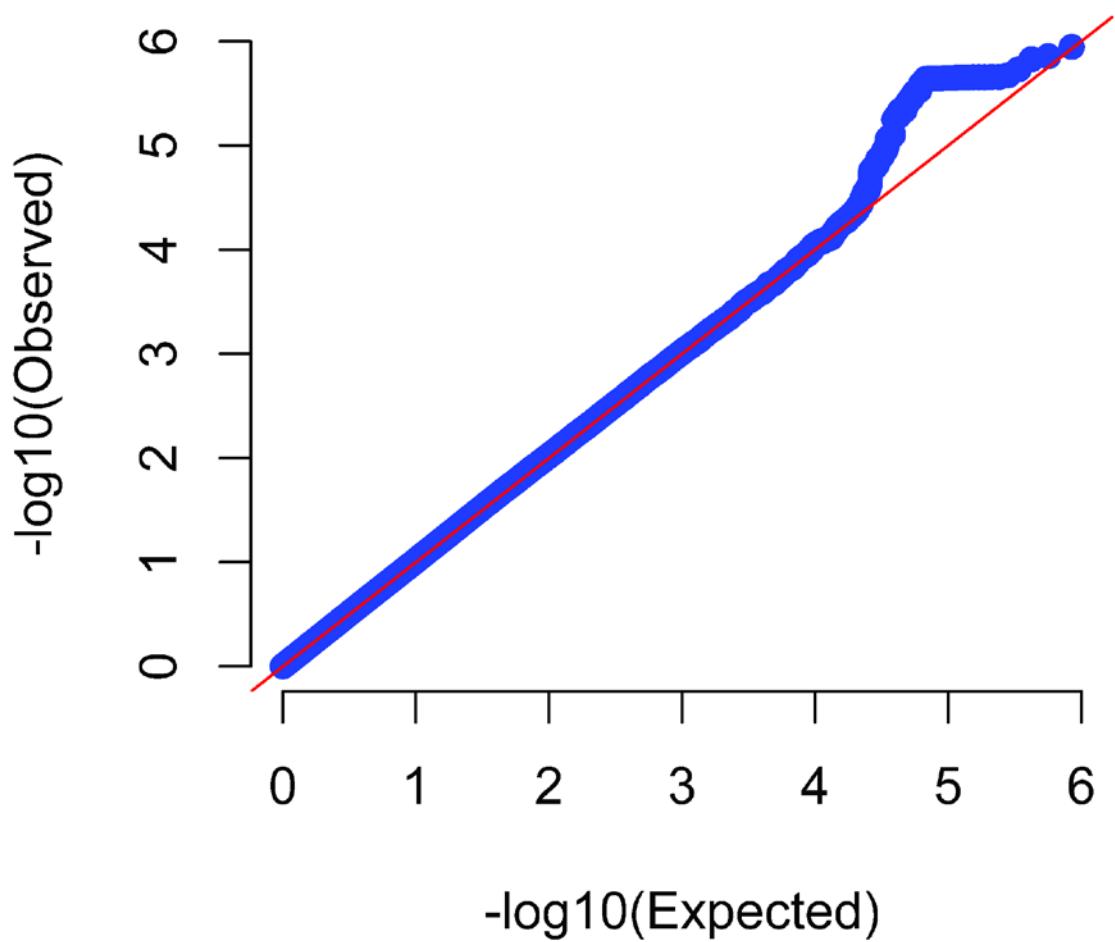


**Supplementary Figure 2: Q-Q plots of  $P$  values for the association of SNPs with CRC risk using data in stage 1.** The eight plots are (a) Shanghai-1 with additional controls ( $\lambda = 1.031142$ ), (b) Shanghai-1 without additional controls ( $\lambda = 1.02733$ ), (c) Shanghai-2 with additional controls ( $\lambda = 1.010774$ ), (d) Shanghai-2 without additional controls ( $\lambda = 1.033529$ ), (e) Guangzhou-1 ( $\lambda = 1.018318$ ), (f) Aichi-1 ( $\lambda = 1.036879$ ), (g) KCPS-II ( $\lambda = 1.0169$ ) and (h) meta-analysis ( $\lambda = 1.0368$ ;  $\lambda_{1000} = 1.0118$ ).

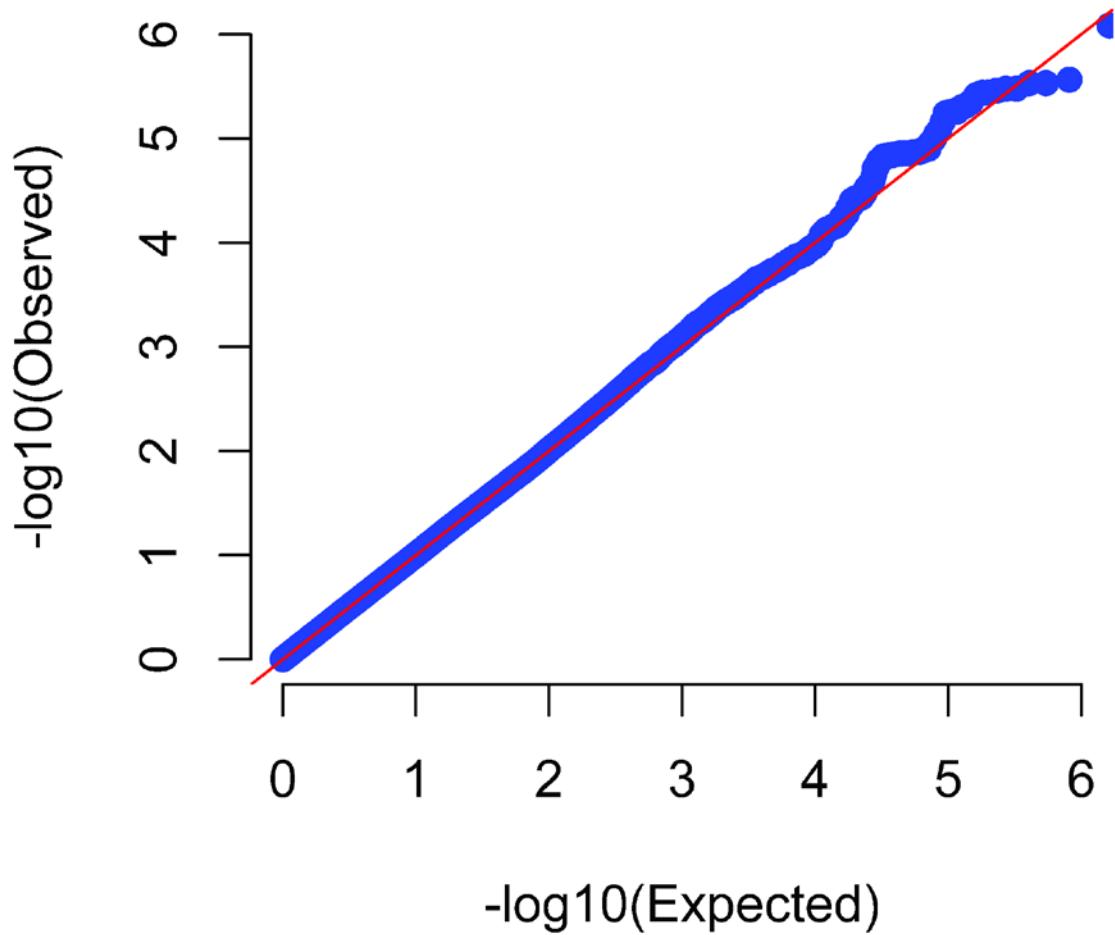
(a)



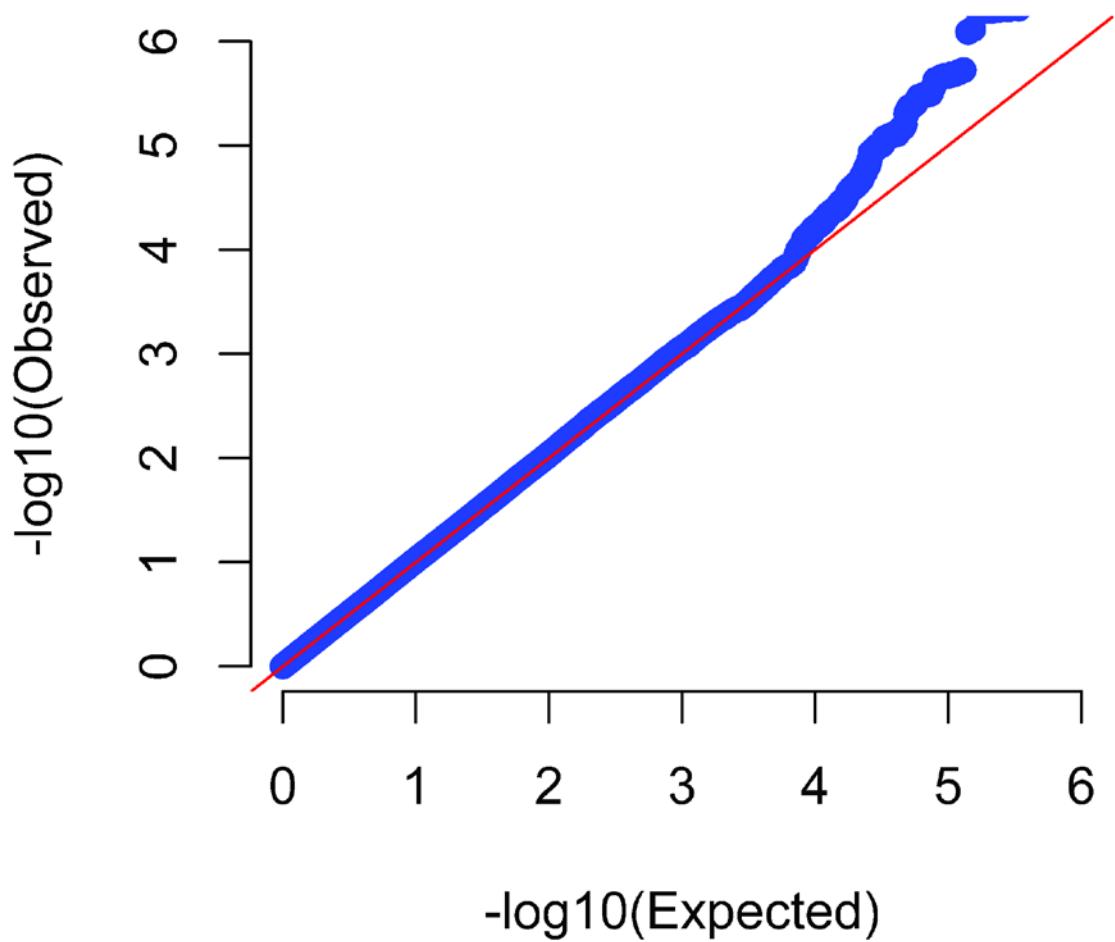
**(b)**



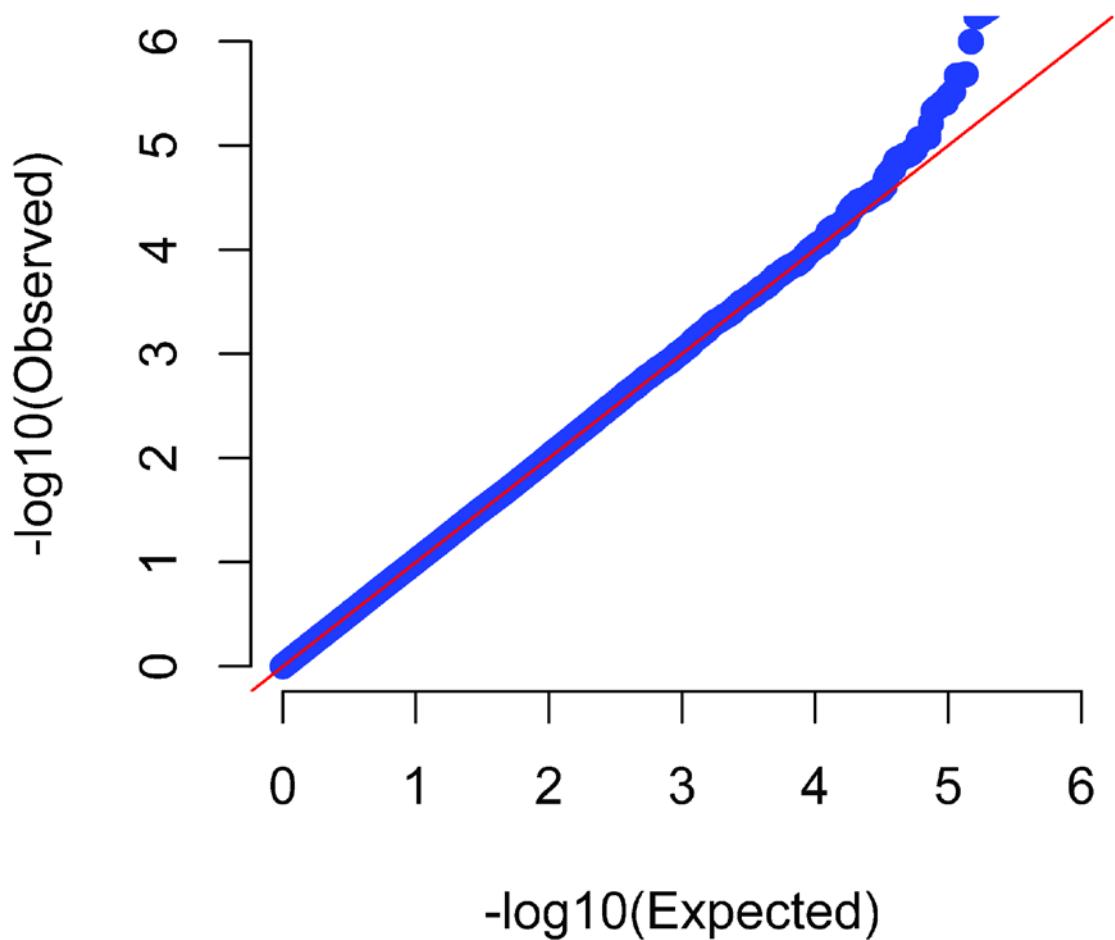
**(c)**



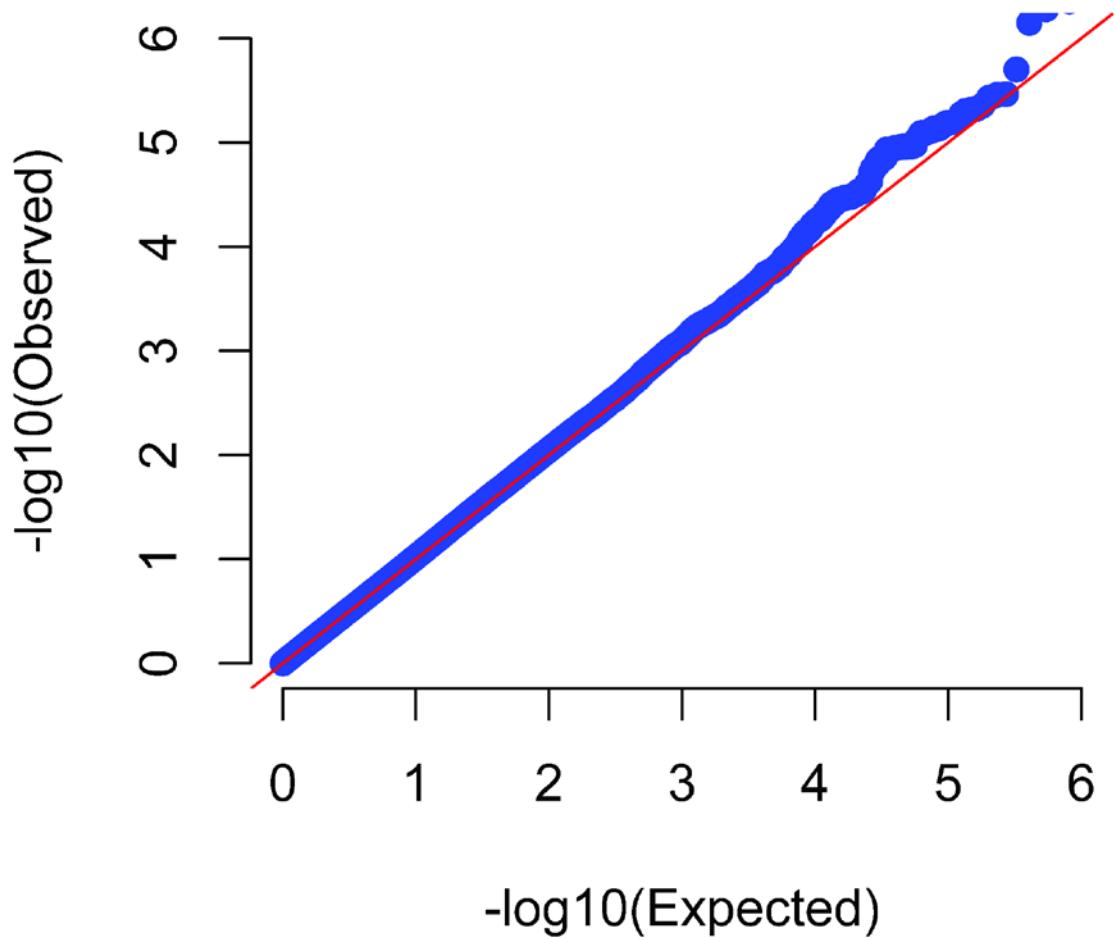
(d)



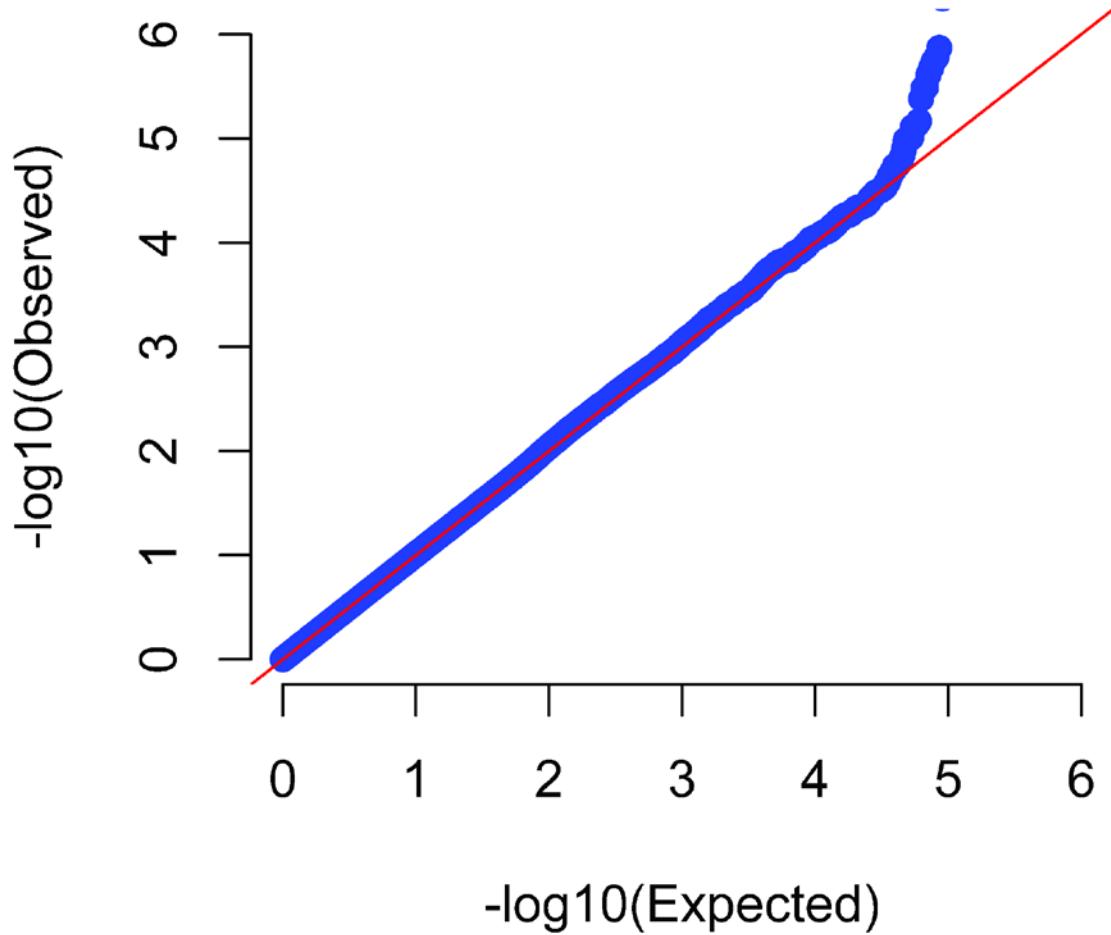
(e)



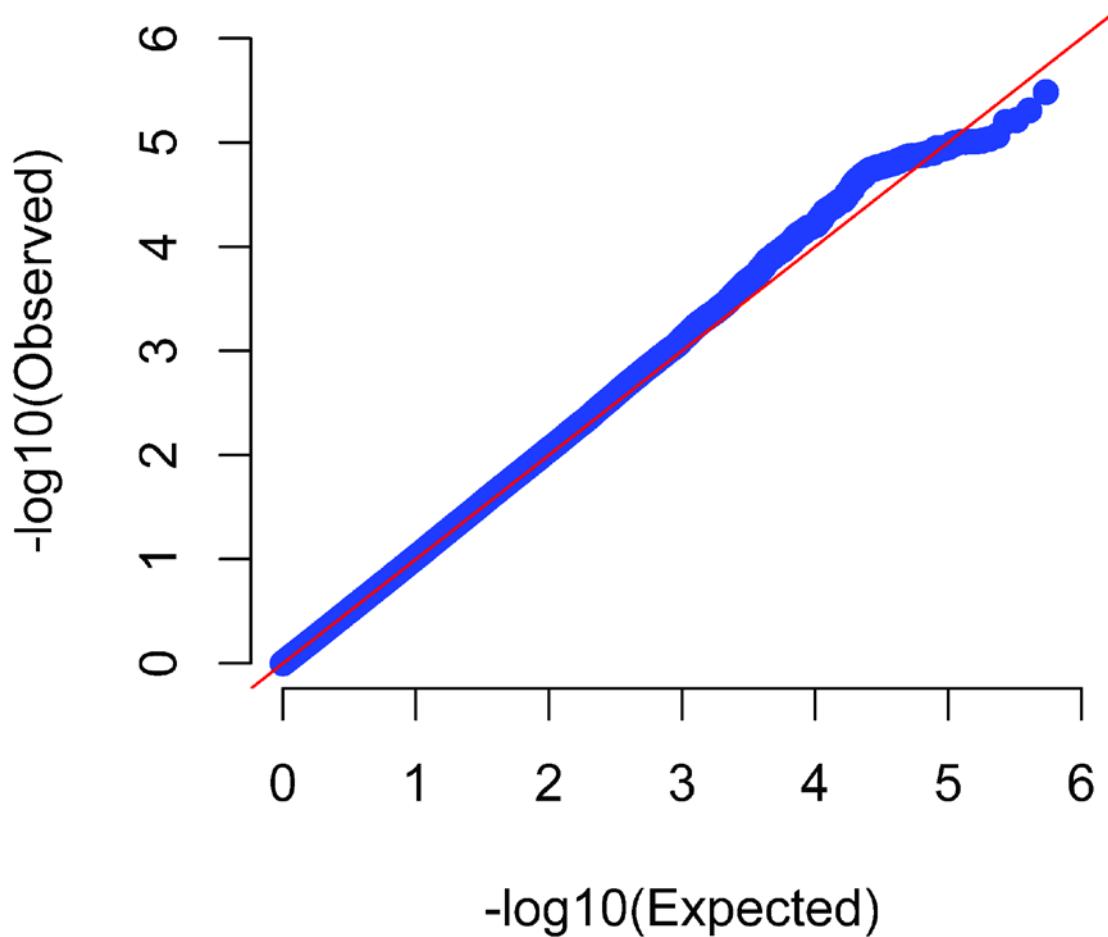
(f)



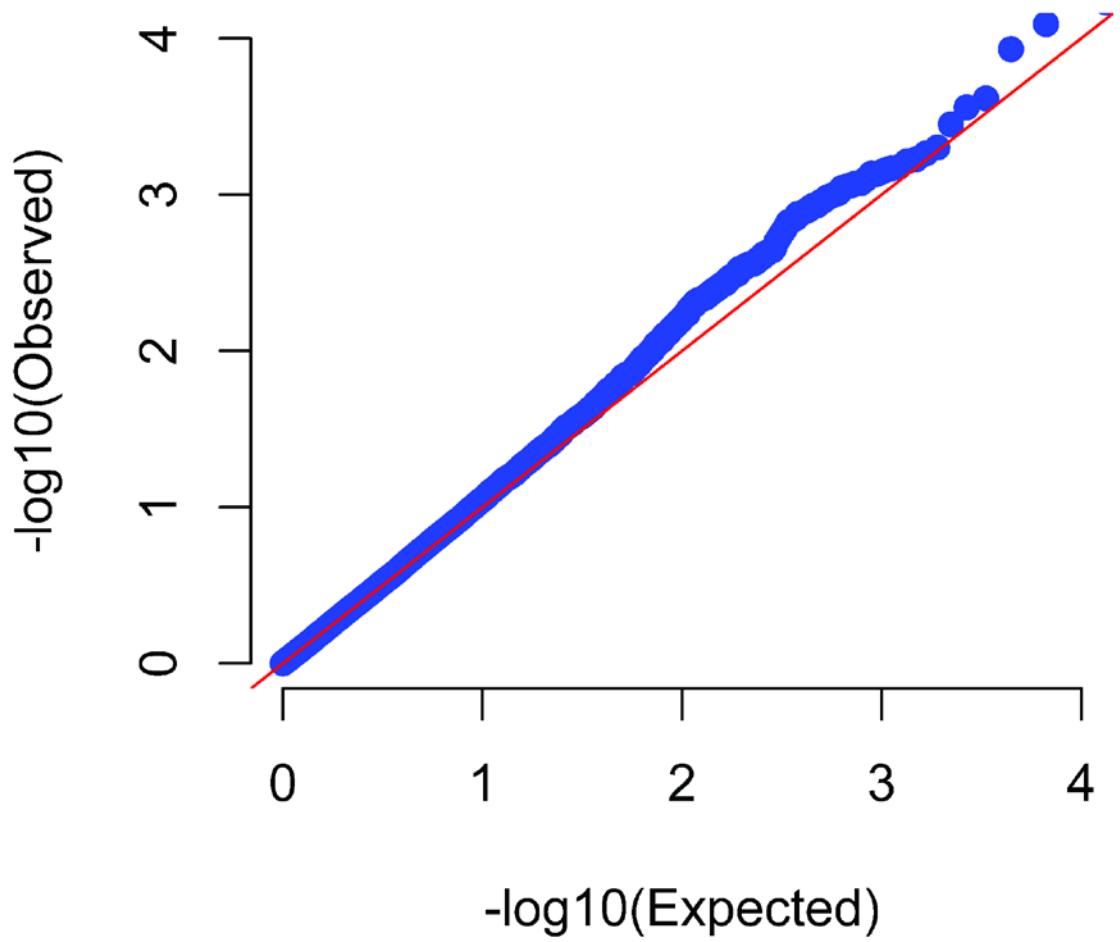
(g)



(h)

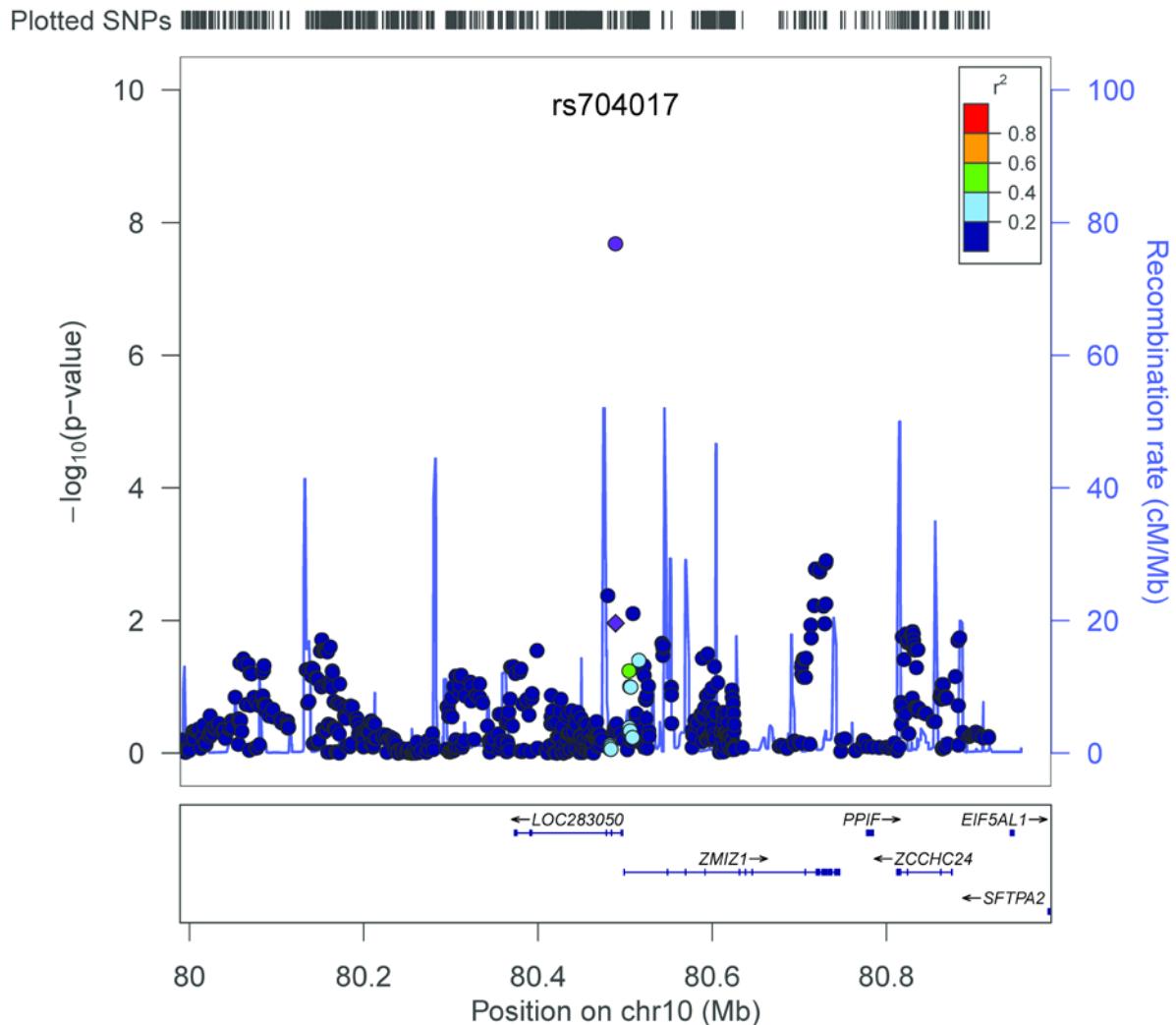


**Supplementary Figure 3: Q-Q plot of  $P$  values for the association of SNPs with CRC risk in meta-analysis of stage 2 data ( $\lambda = 1.0525$ ;  $\lambda_{1000} = 1.0116$ ).**

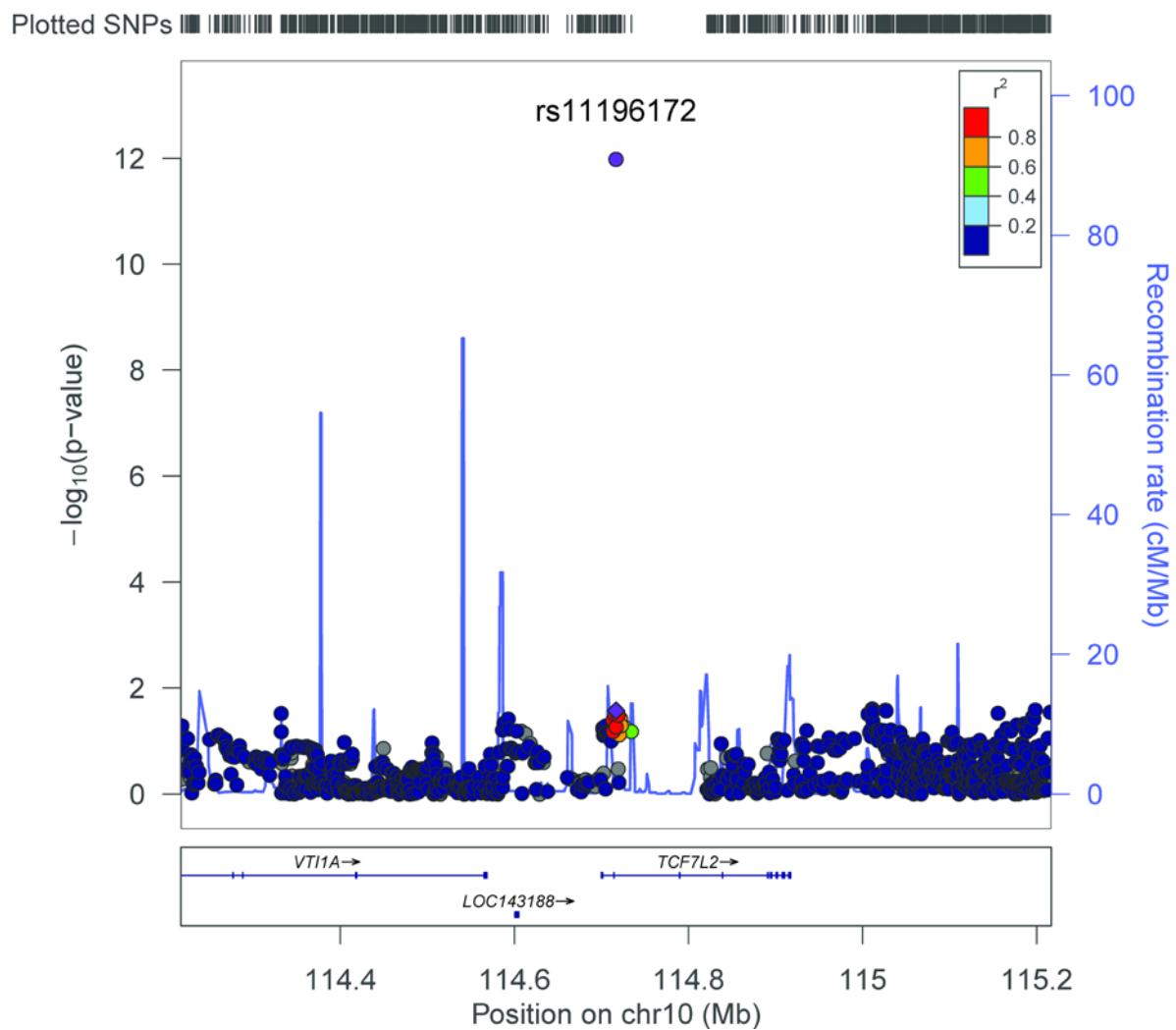


**Supplementary Figure 4: Regional association plots of the six newly identified loci.** The six plots represent (a) 10q22.3, (b) 10q25.2, (c) 11q12.2, (d) 12p13.31, (e) 17p13.3 and (f) 19q13.2. For each plot, the  $-\log_{10}(P$  values) (y axis) of SNPs are shown according to their chromosomal positions (x axis) in NCBI Build 36. Blue lines represent the estimated recombination rates from the HapMap Project (NCBI Build 36). Arrows indicate genomic locations of genes within the 1-mb regions centered on the index SNPs in the NCBI Build 36 human assembly. The color of SNPs represents their LD ( $r^2$ , HapMap Asian) with the index SNP at each locus. With the exception of the index SNPs, which are shown as purple diamonds for stage 1 and purple circles for the meta-analyses of all studies, data shown for all other SNPs are from stage 1 only.

a

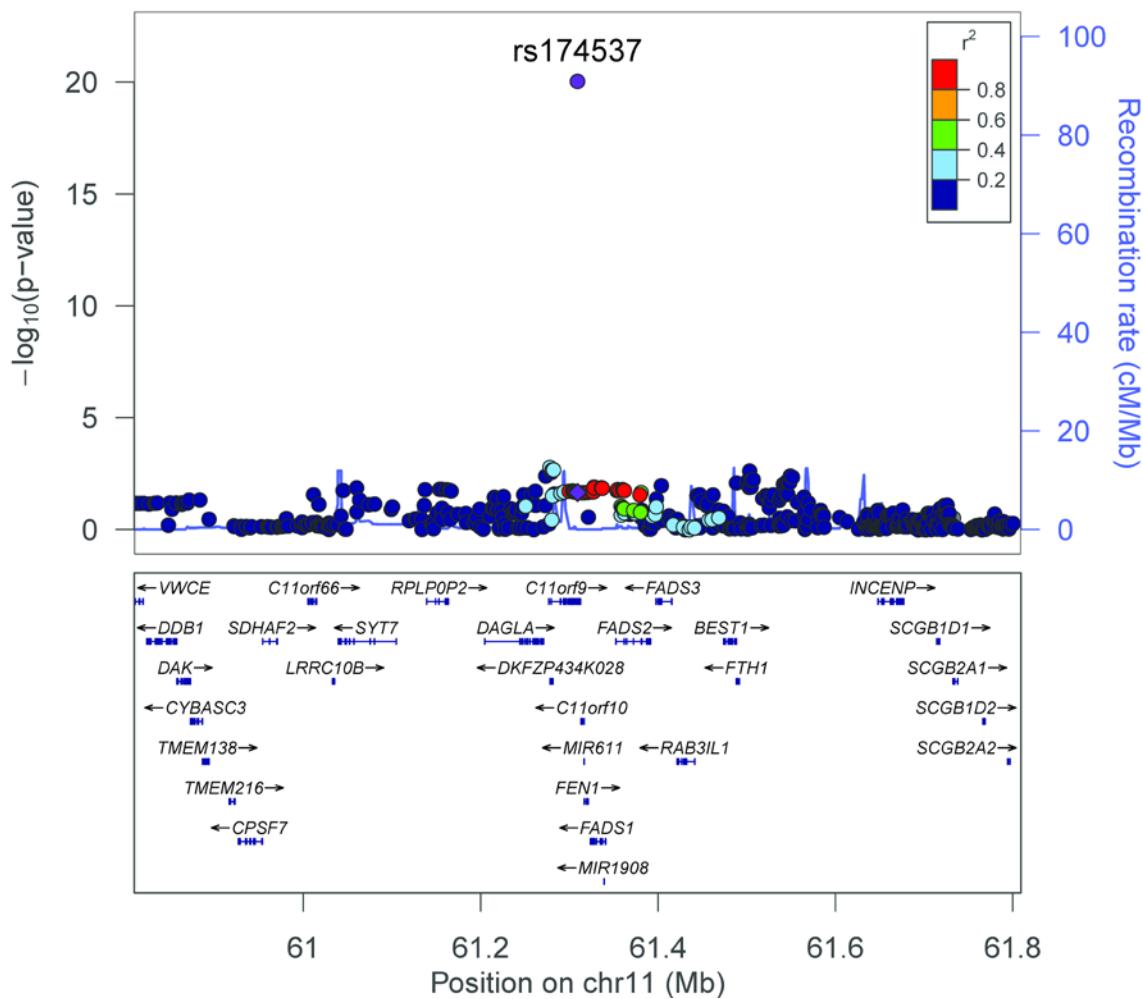


b



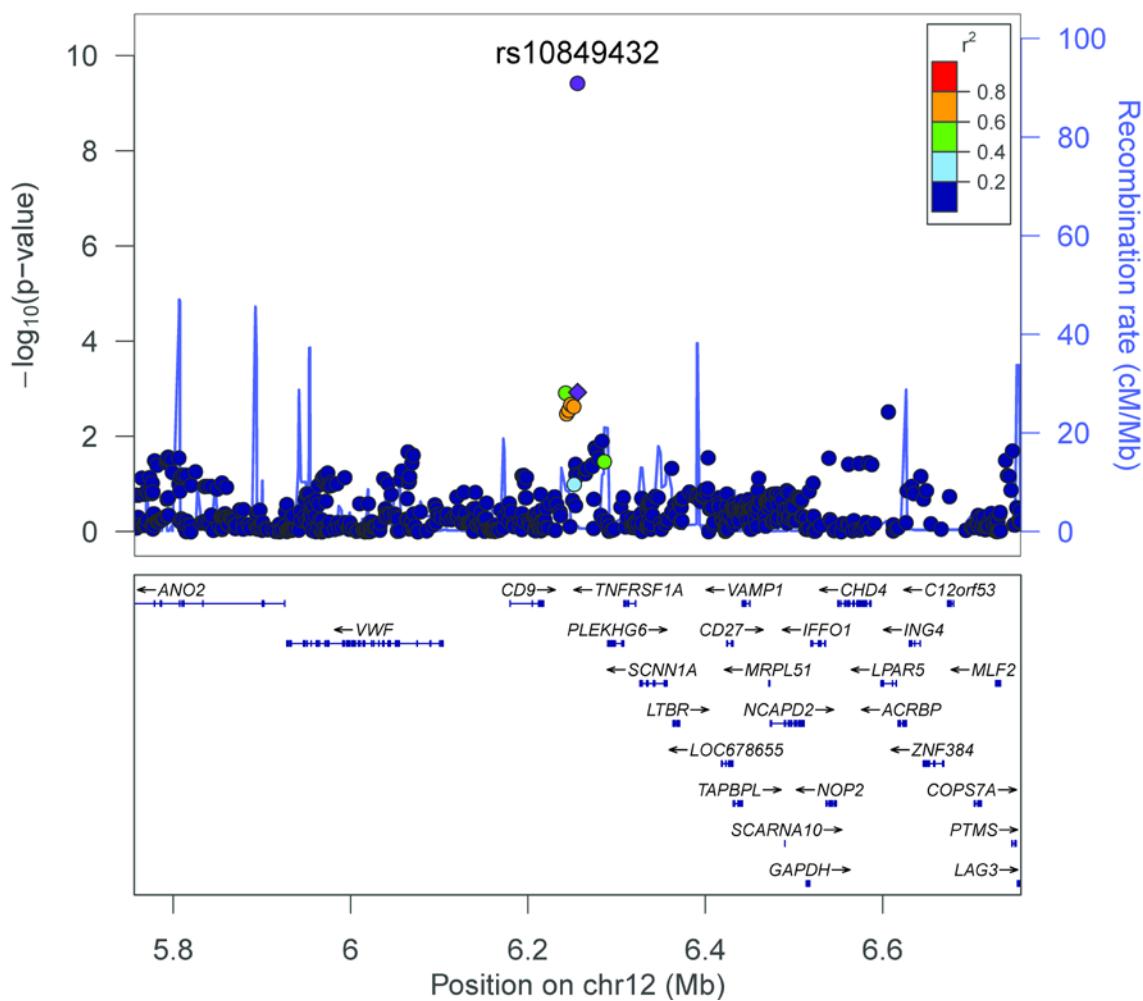
C

Plotted SNPs

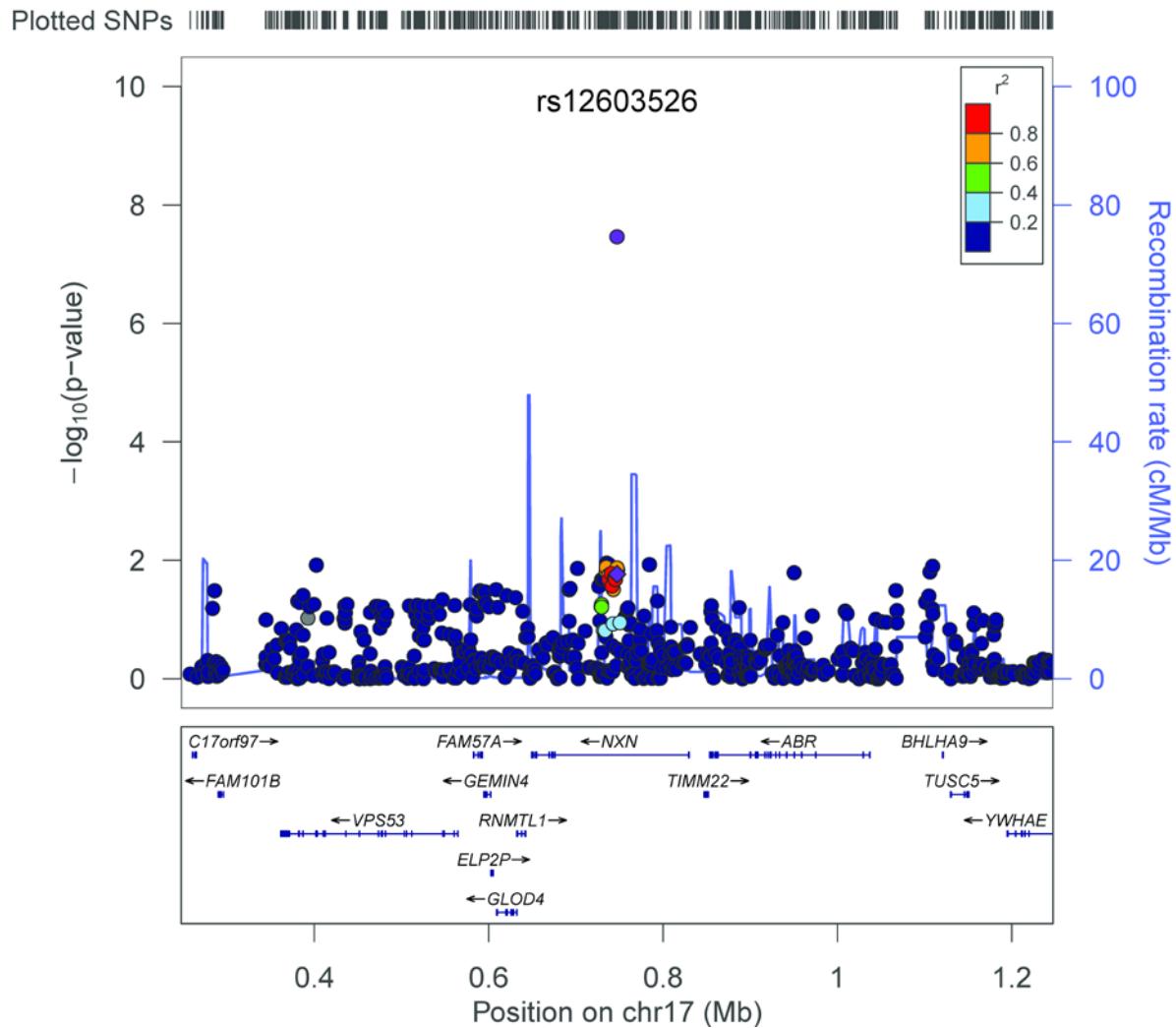


d

Plotted SNPs

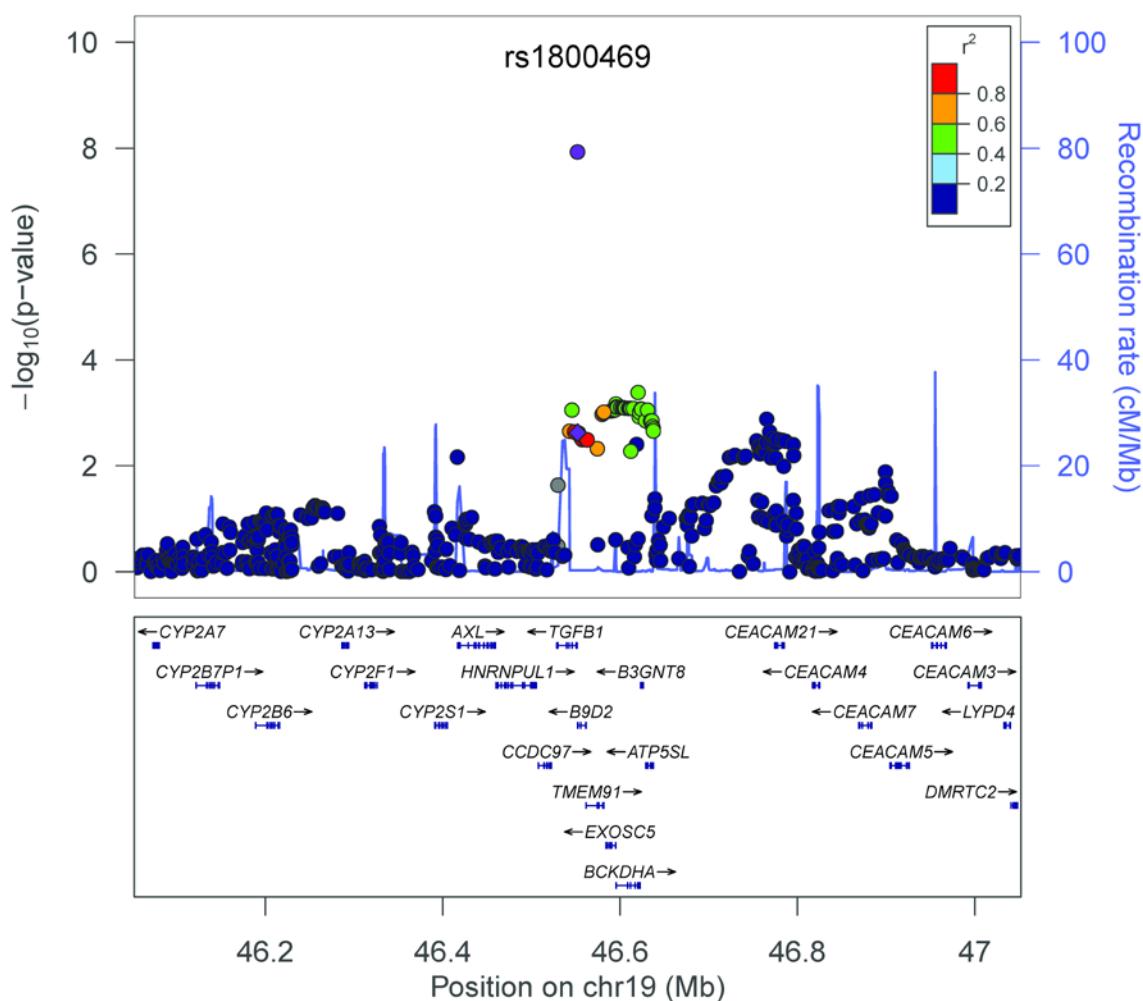


e



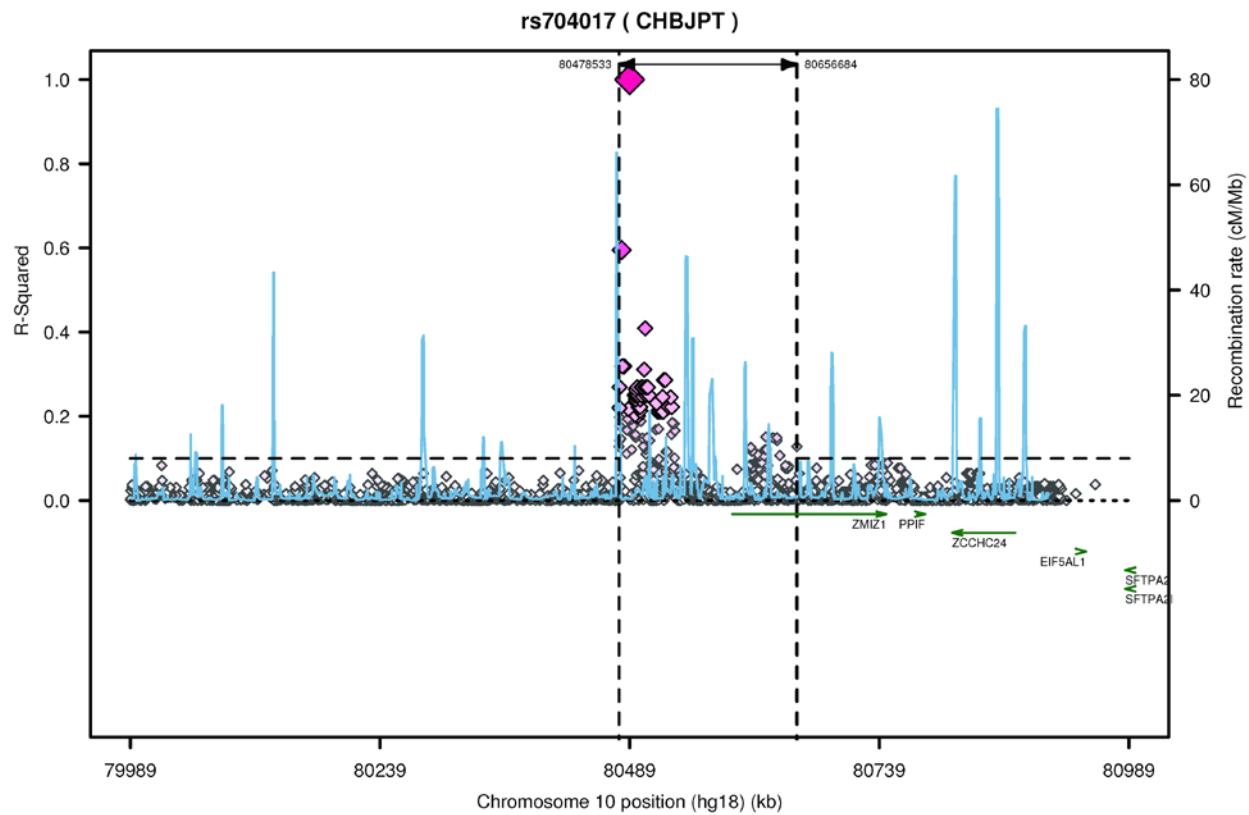
f

Plotted SNPs

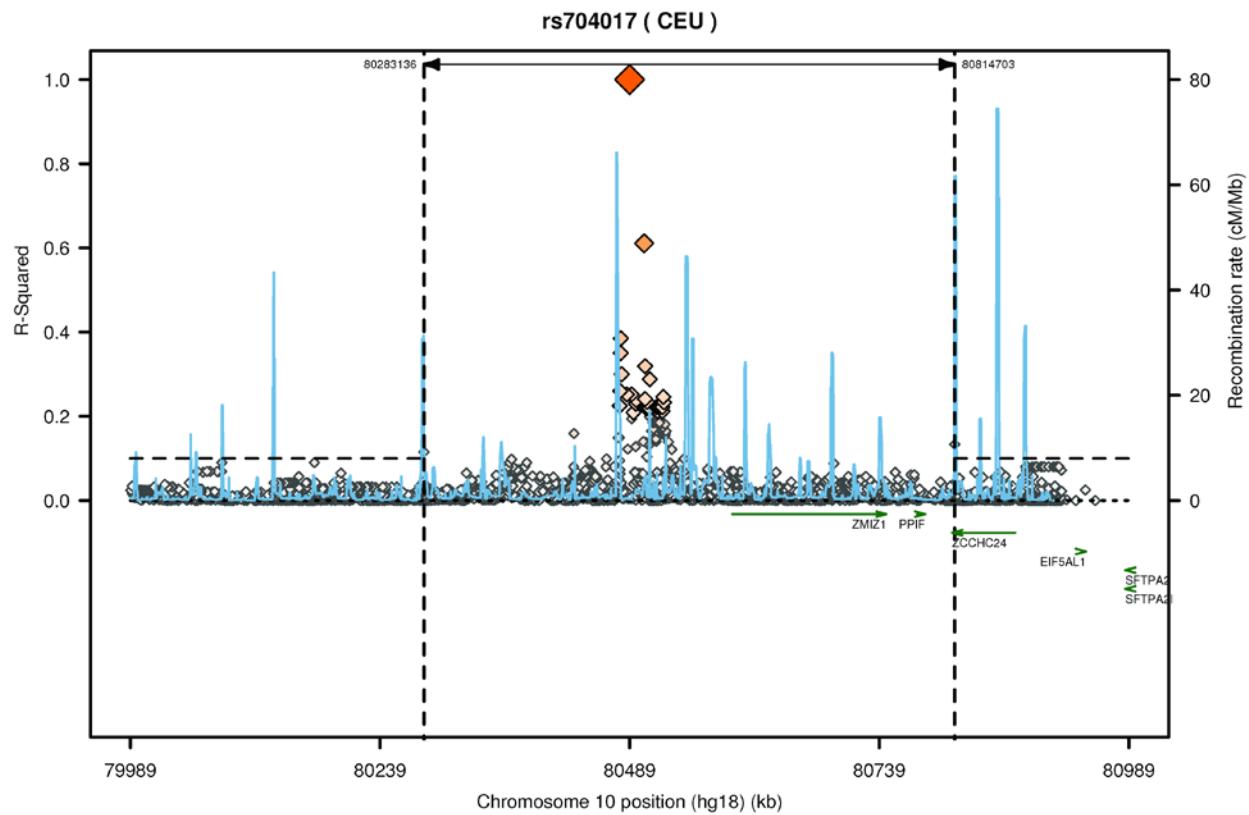


**Supplementary Figure 5: LD patterns for six newly identified loci in Asian and European populations using data from the 1000 Genomes Project.** The 12 plots represent (a) rs704017 (CHBJPT), (b) rs704017 (CEU), (c) rs11196172 (CHBJPT), (d) rs11196172 (CEU), (e) rs174537 (CHBJPT), (f) rs174537 (CEU), (g) rs10849432 (CHBJPT), (h) rs10849432 (CEU), (i) rs12603526 (CHBJPT), (j) rs12603526 (CEU), (k) rs1800469 (CHBJPT), and (l) rs1800469 (CEU).

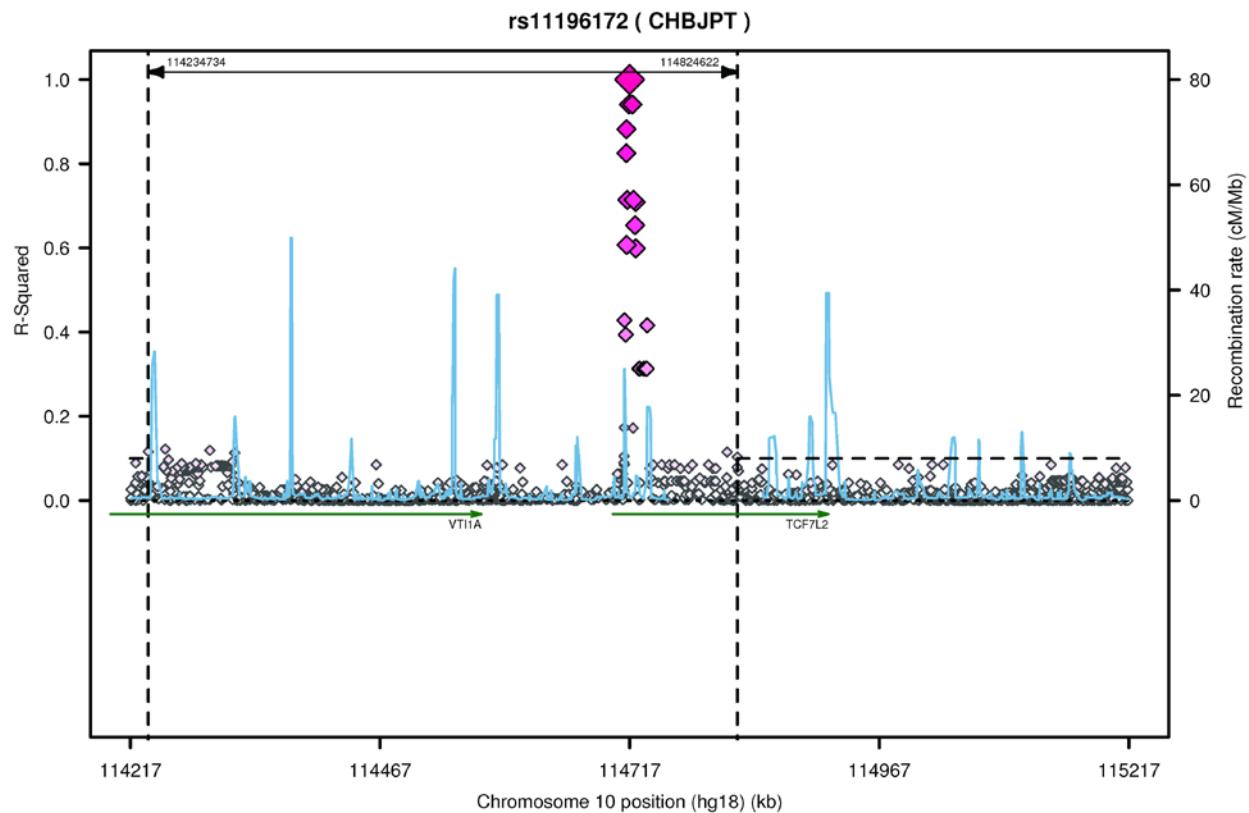
(a)



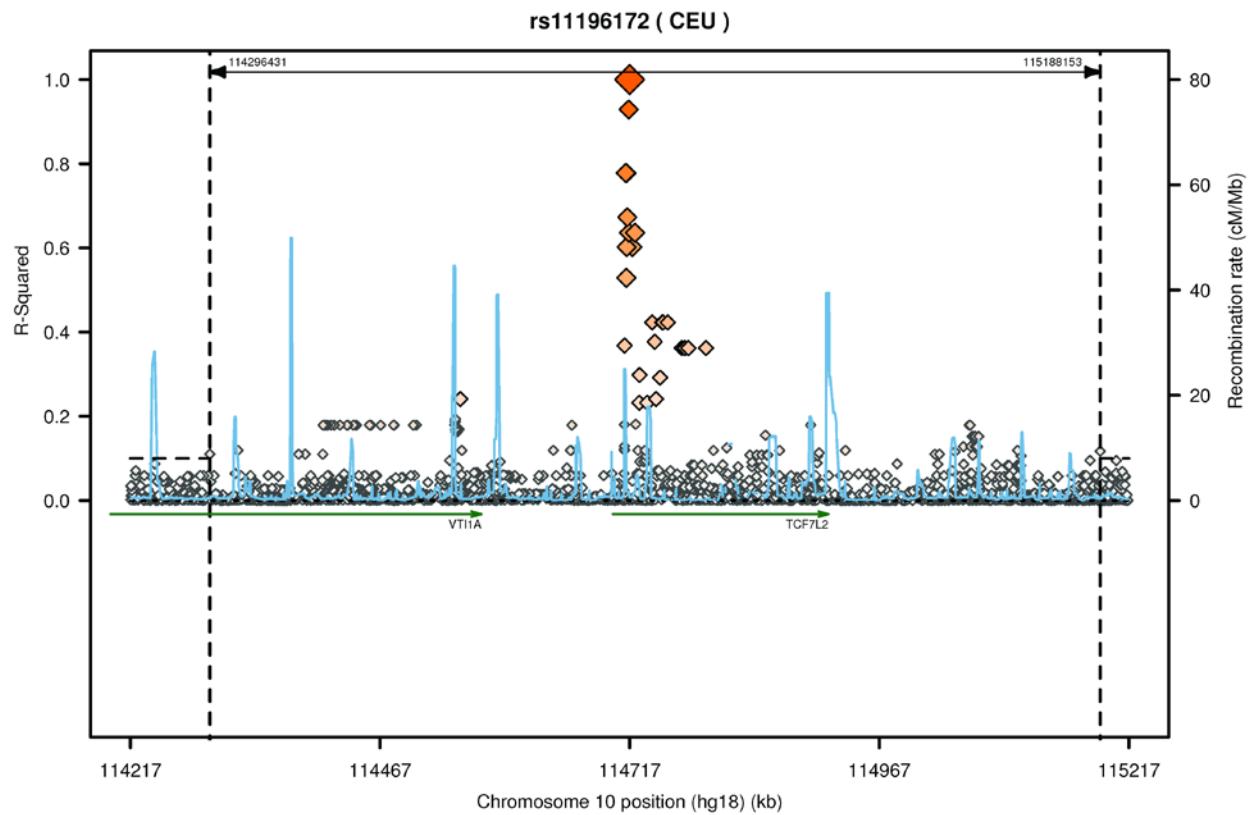
(b)



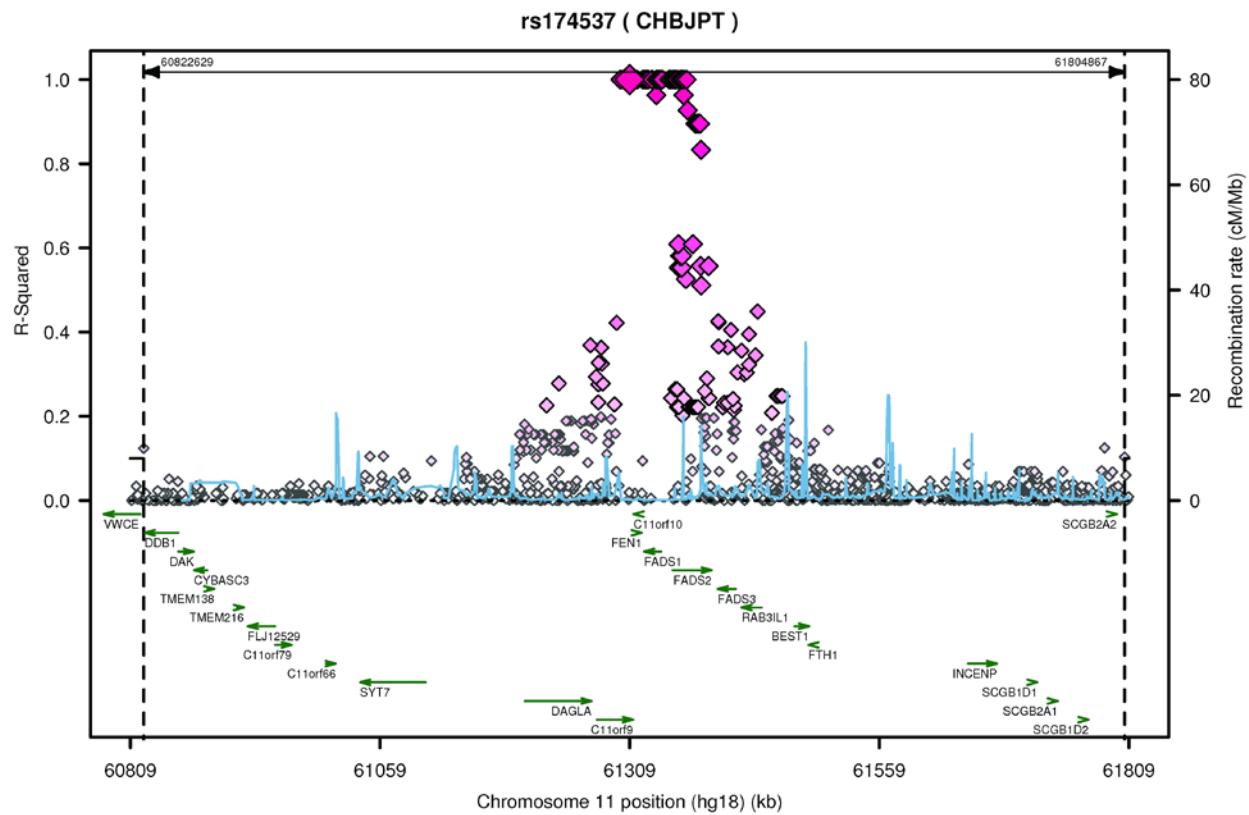
(c)



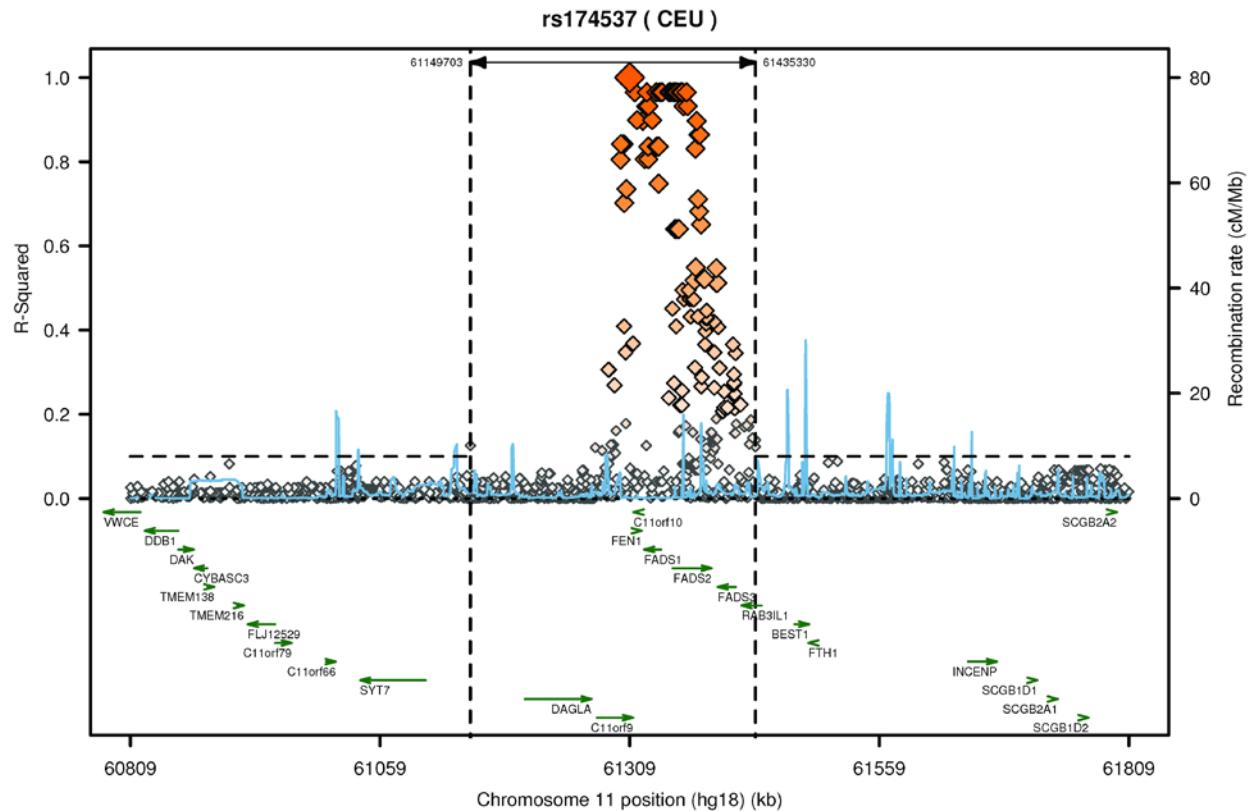
(d)



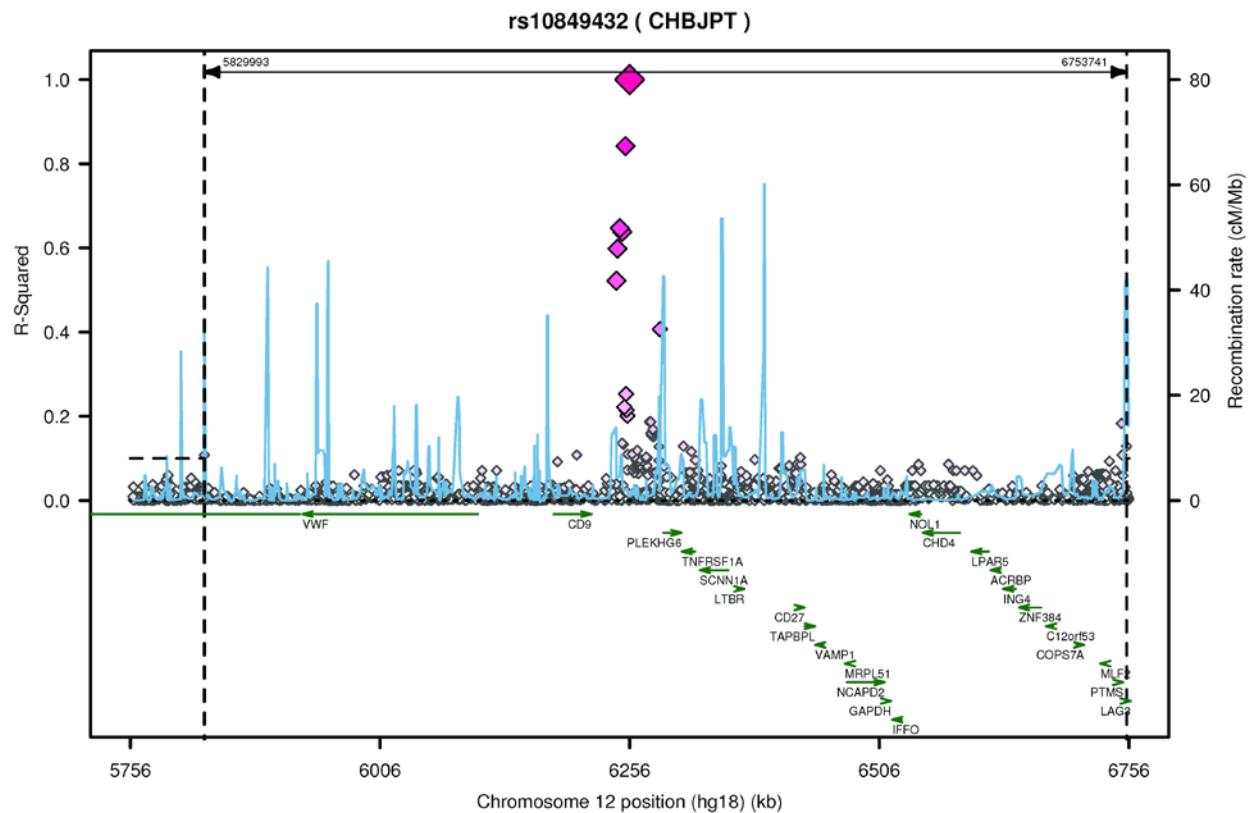
(e)



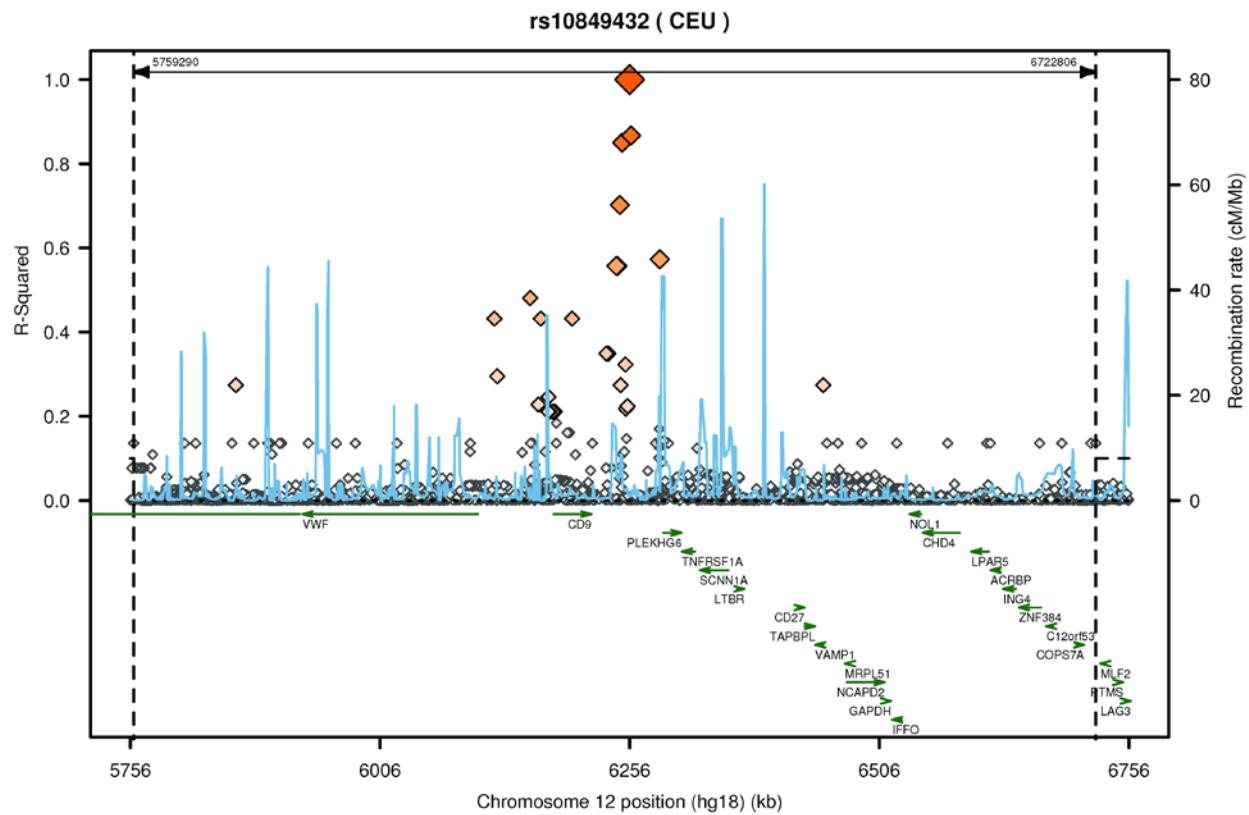
(f)



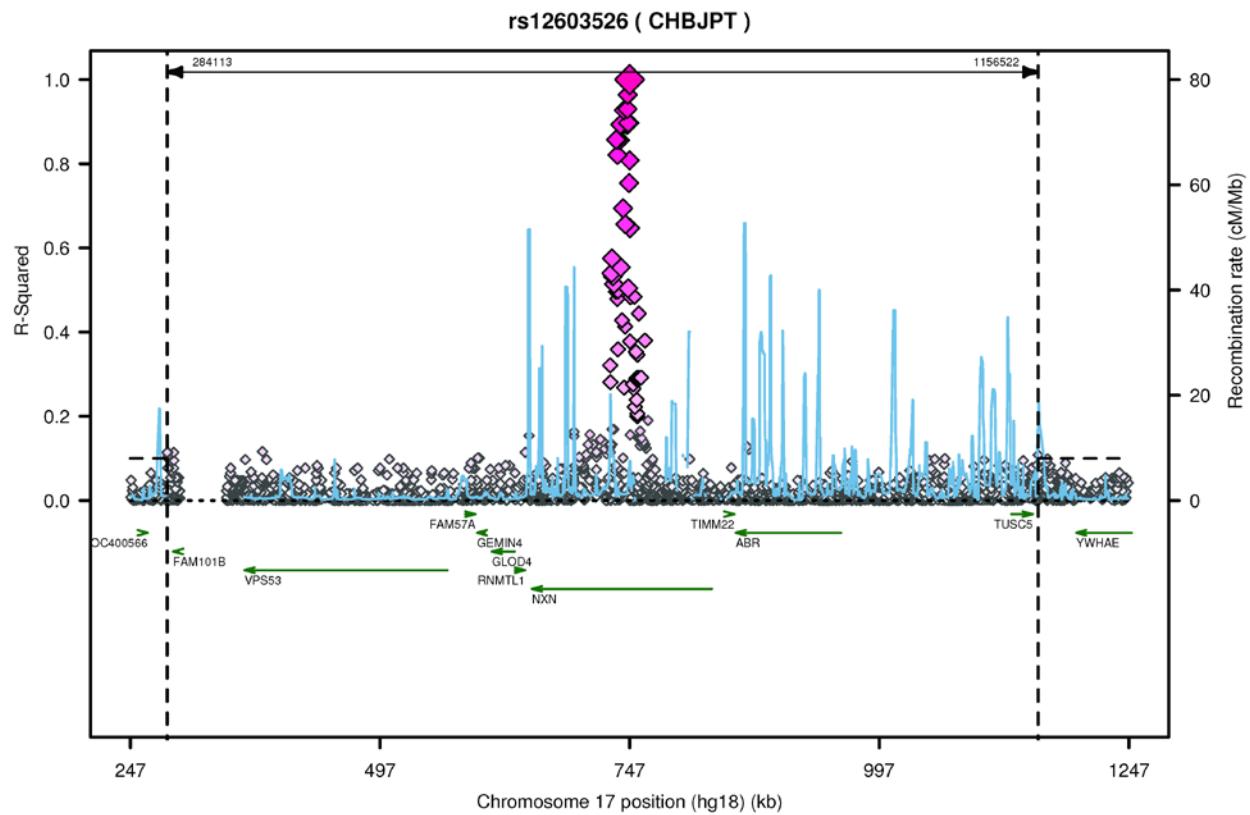
(g)



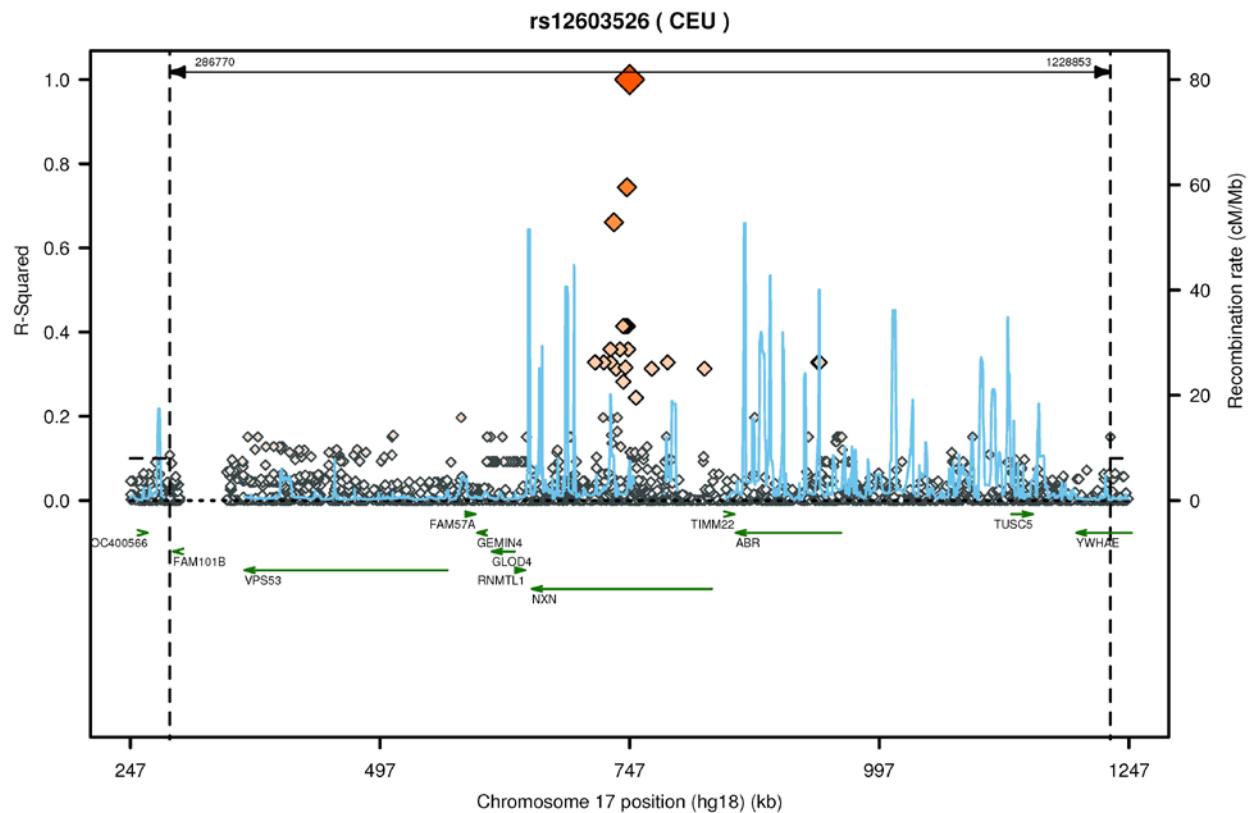
(h)



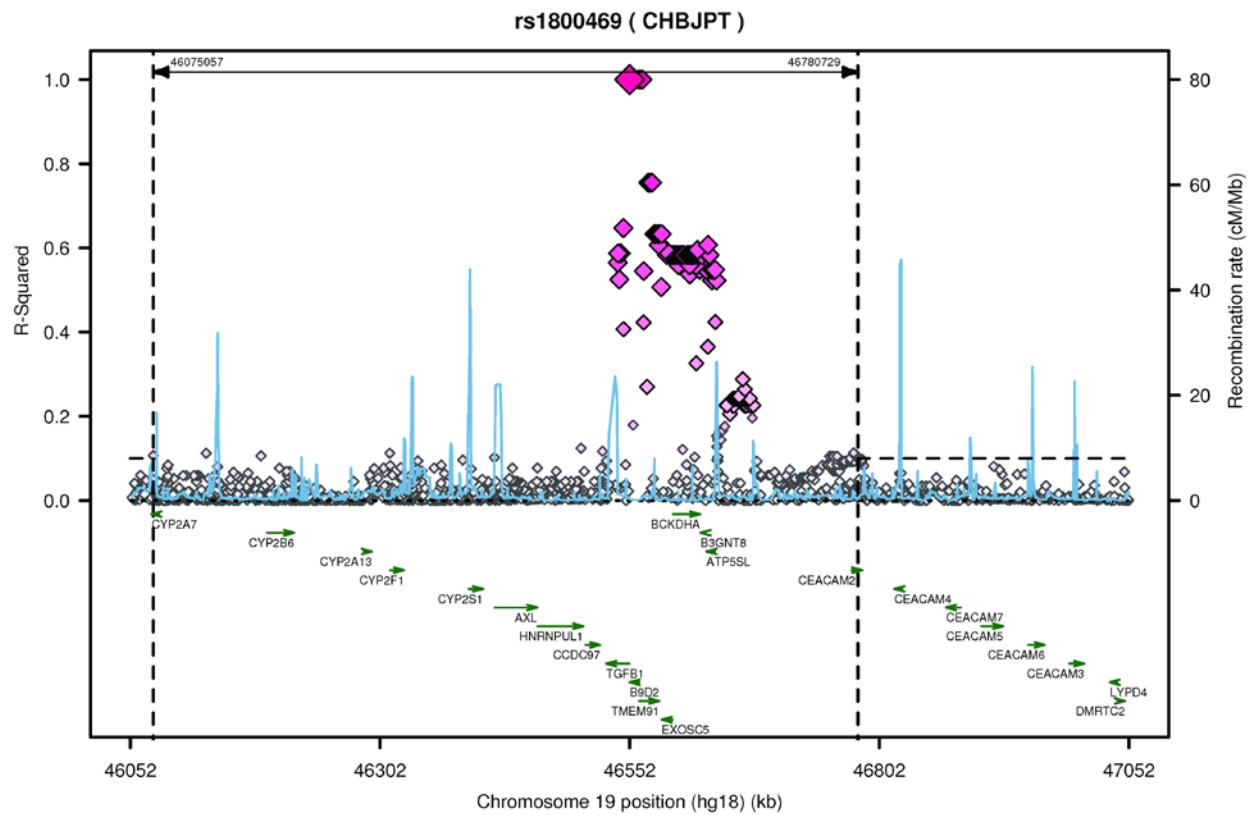
(i)



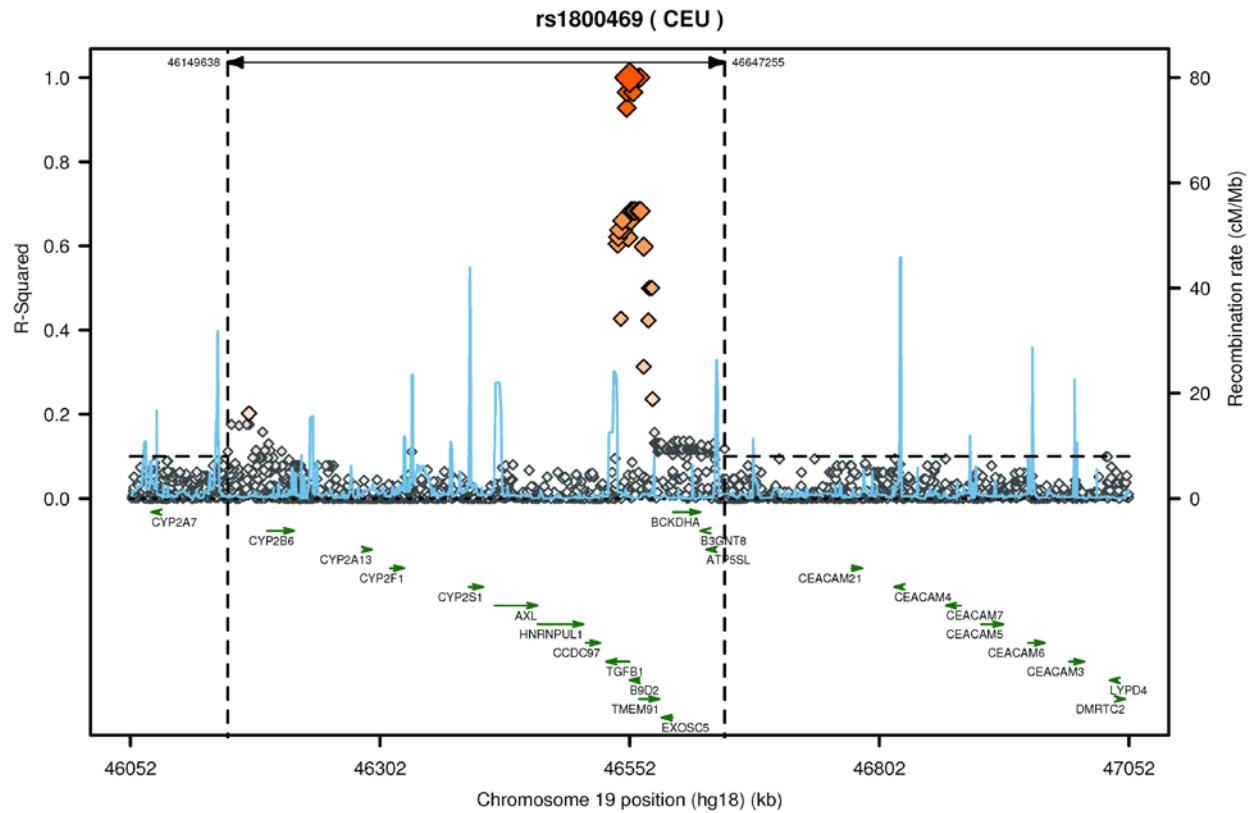
(j)



(k)

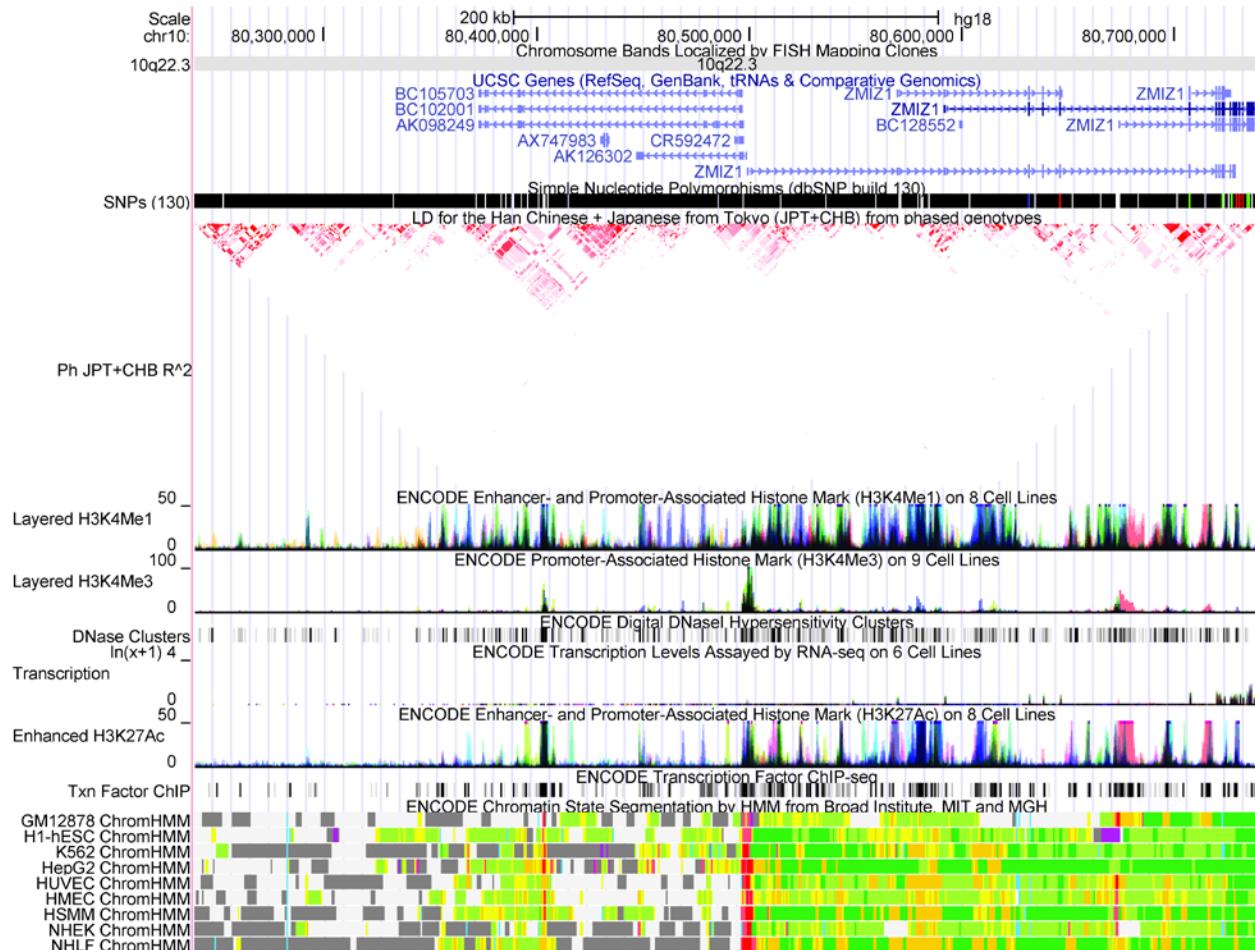


(I)

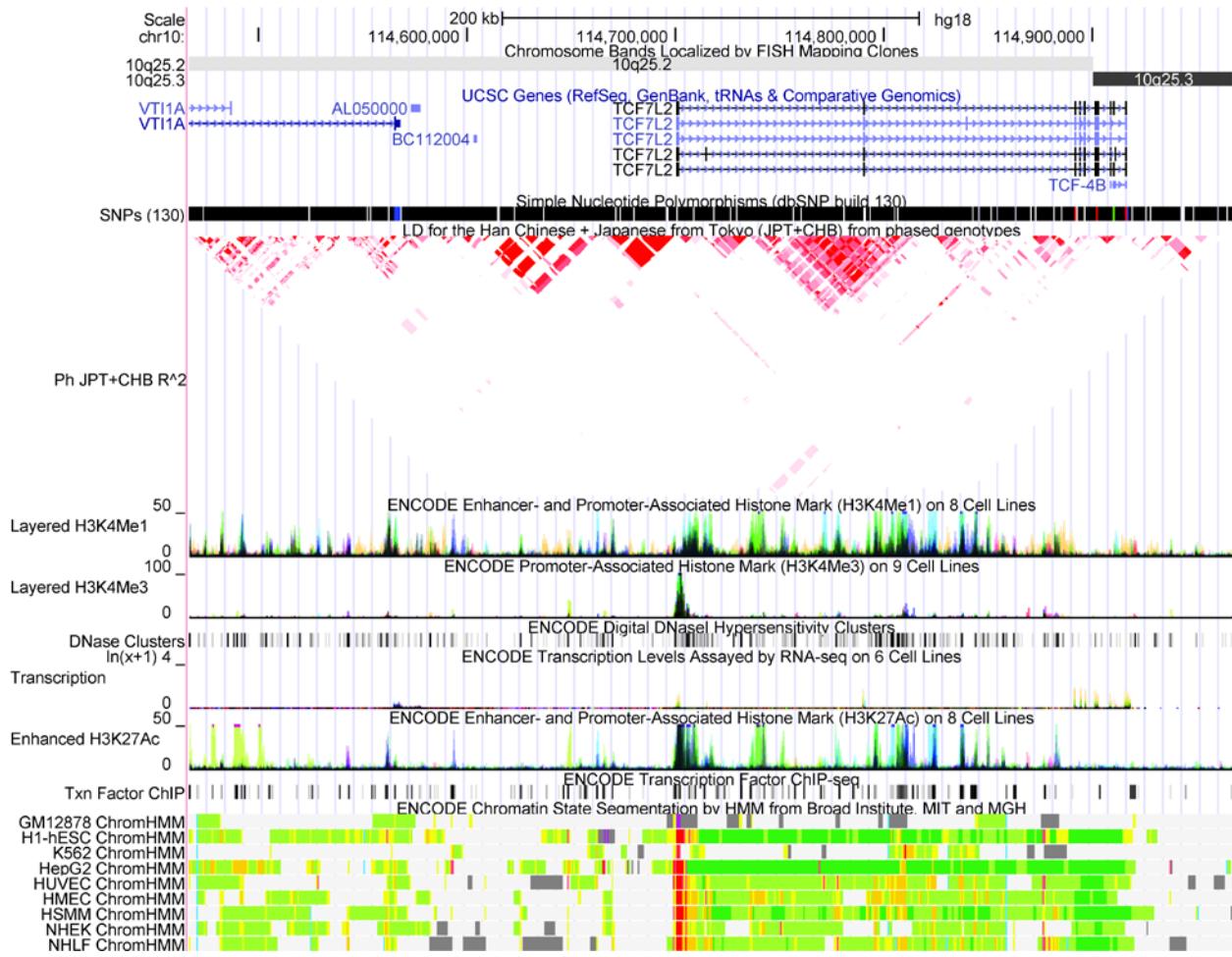


**Supplementary Figure 6: Evidence from ENCODE data for regulatory function of SNPs in the six newly identified loci using the UCSC Genome Browser (see PDF file).** The six plots represent (a) 10q22.3, (b) 10q25.2, (c) 11q12.2, (d) 12p13.31, (e) 17p13.3 and (f) 19q13.2, within a 500-kb window centered on rs704017, rs11196172, rs174537, rs10849432, rs12603526 and rs1800469, respectively. Tracks (from top to bottom) in each of the plots are Genome Base Position, Chromosome Bands, UCSC Genes, Simple Nucleotide Polymorphisms (dbSNP build 130), LD structure (HapMap 2, CHB+JPT), ENCODE Enhancer- and Promoter-Associated Histone Mark (H3K4Me1) on 8 Cell Lines, ENCODE Promoter-Associated Histone Mark (H3K4Me3) on 9 Cell Lines, ENCODE Digital DNaseI Hypersensitivity Clusters, ENCODE Transcription Levels Assayed by RNA-seq on 6 Cell Lines, ENCODE Enhancer- and Promoter-Associated Histone Mark (H3K27Ac) on 8 Cell Lines, ENCODE Broad Chromatin State Segmentation by HMM on 9 Cell Lines (bright red, active promoter; light red, weak promoter; purple, inactive/poised promoter; orange, strong enhancer; yellow, weak/poised enhancer; blue, insulator; dark green, transcriptional transition/elongation; light green, weak transcribed; gray, polycomb-repressed; light gray, heterochromatin/low signal/repetitive/copy number variation).

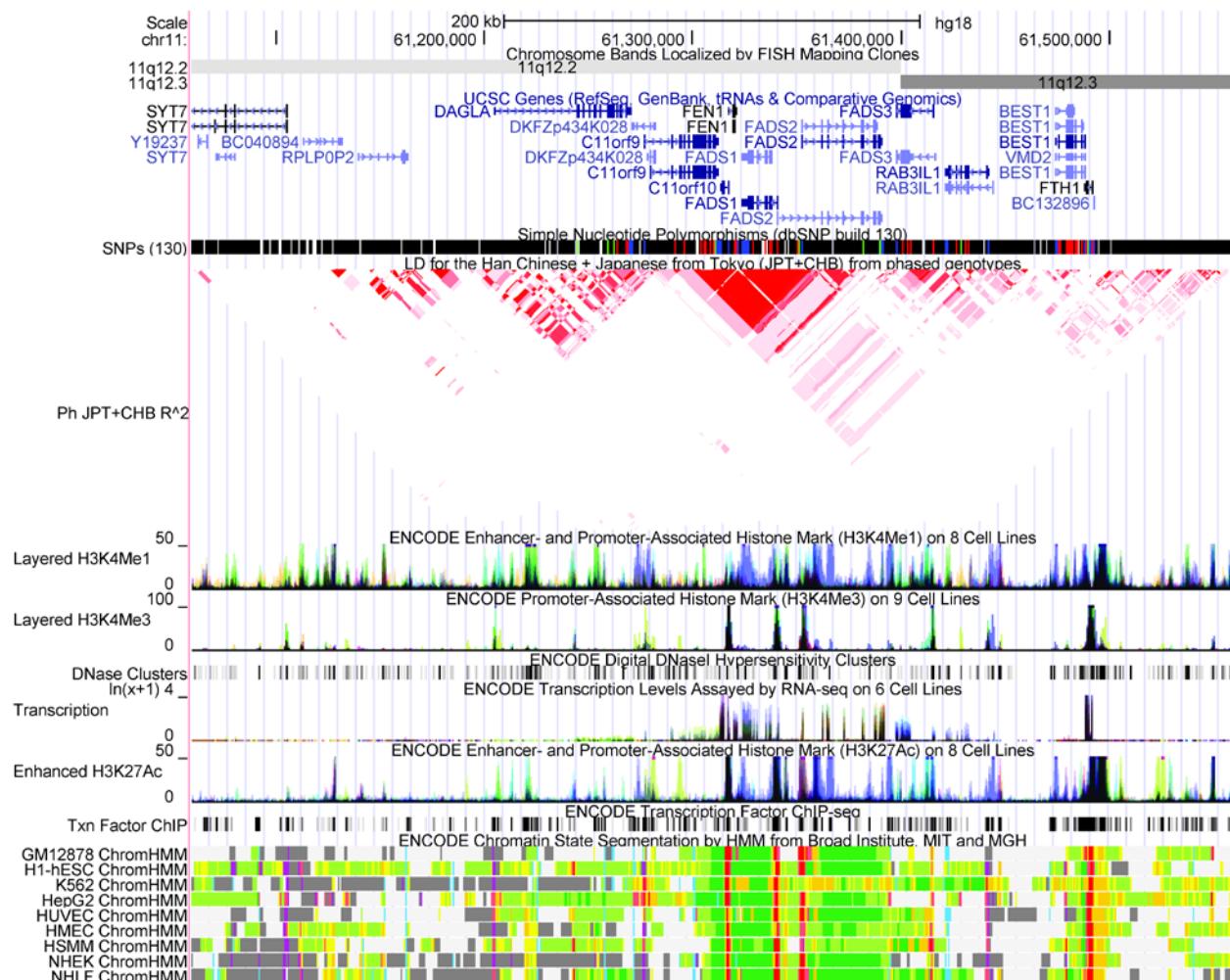
(a)



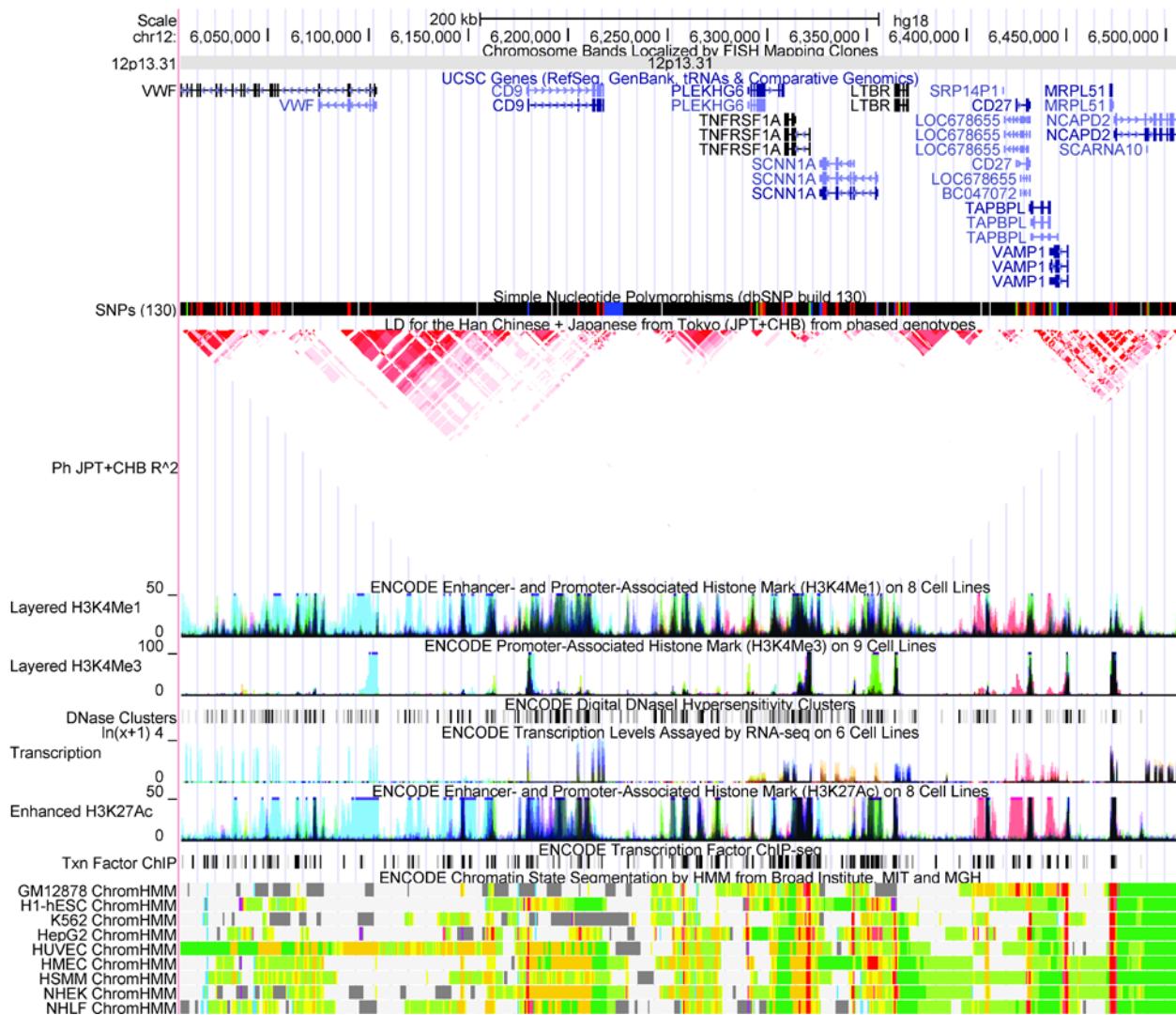
(b)



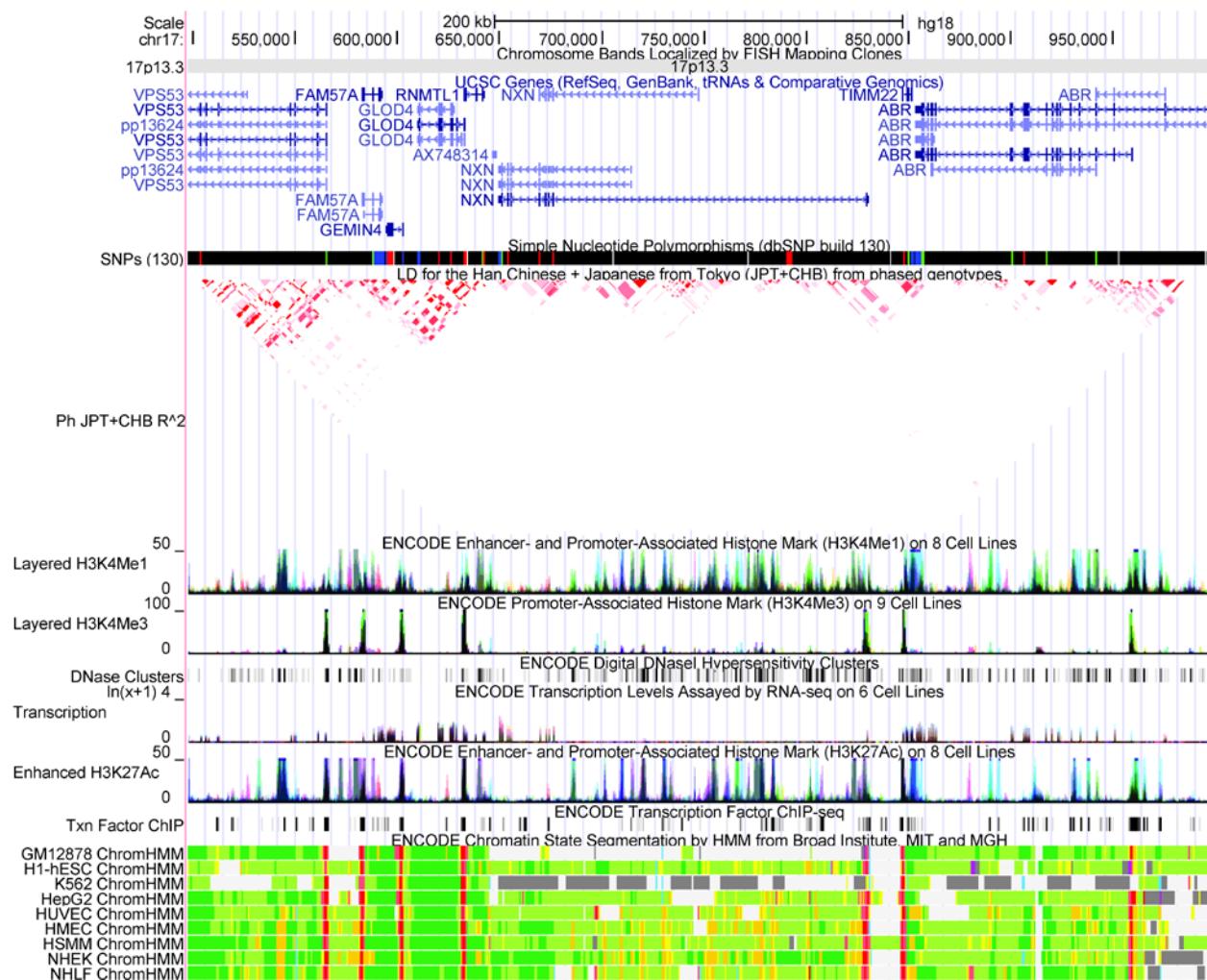
(c)



(d)



(e)



(f)

