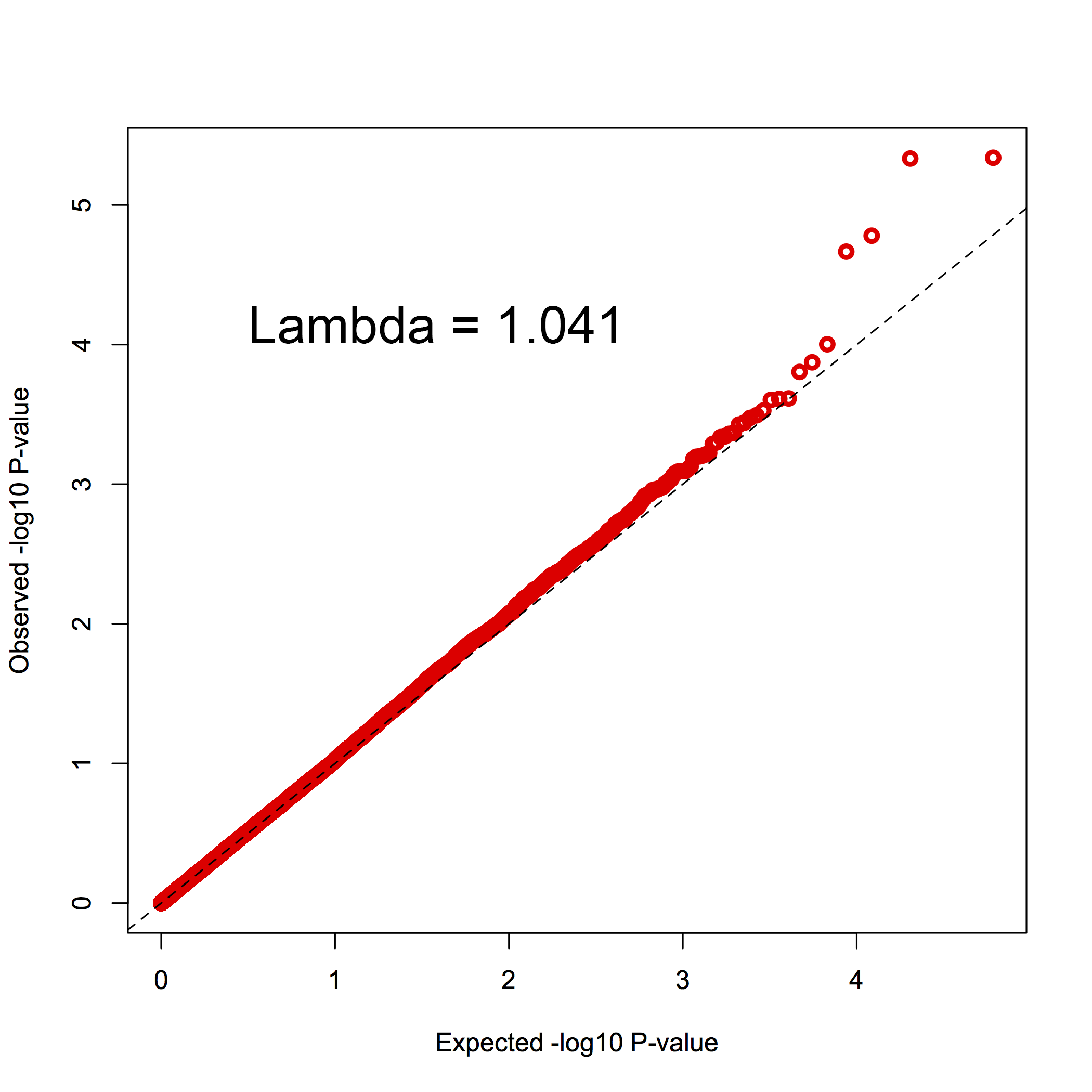
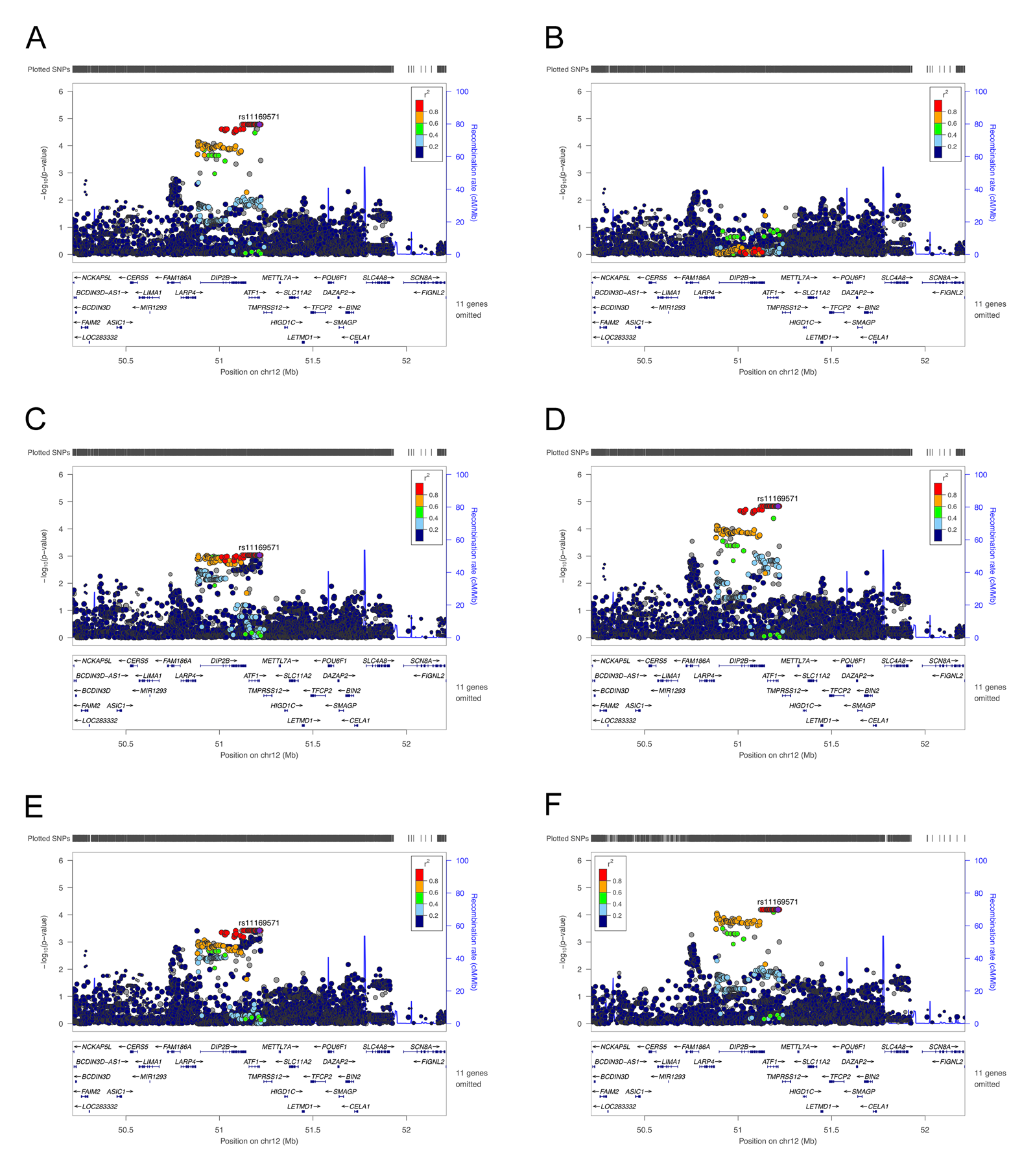


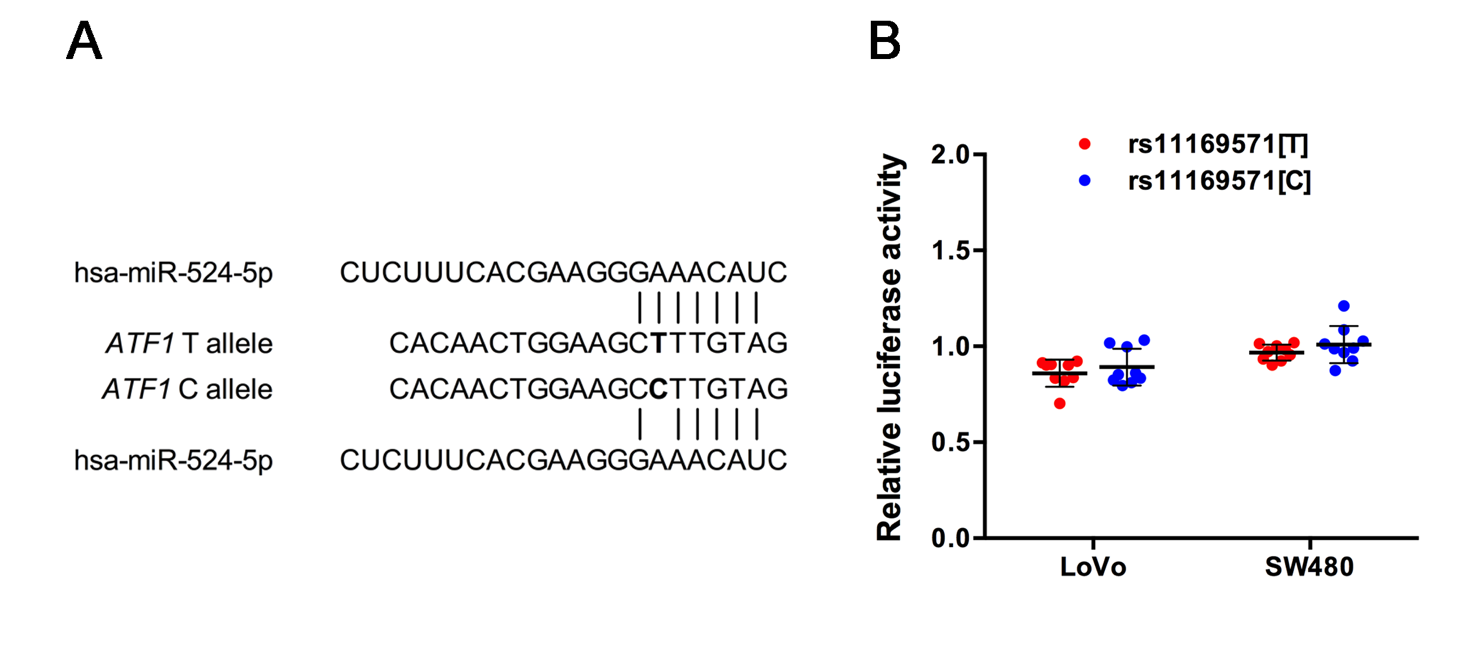
**Supplementary Figure 1. Plots for genetic matching of three principal components (PCs) derived from the PC analysis of 1,066 cases with CRC and 2,203 controls.** (**A**) Plots of the first two PCs from the CRC of all samples and 270 HapMap individuals were shown with different colors, (**B**) PC1 versus PC2 for cases and controls, (**C**) PC1 versus PC3 for cases and controls and (**D**) PC2 versus PC3 for cases and controls.



**Supplementary Figure 2. Quantile-quantile plot and genomic inflation factor lambda for associations.**The red circles represent the distribution of *P* values for the association of variants in 1,062 cases with colorectal cancer and 2,184 controls with adjustment of sex, age and the top 3 principal components.



**Supplementary Figure 3. The rs11169571 signal was the only one in the Chinses population.** The logistic regressions were performed without conditioning any variant (**A**) and conditioning on the rs11169571 (**B**), rs4768903 (**C**), rs7136702 (**D**), rs11169552 (**E**) or rs34245511 (**F**). The −log10 *P* values (y axis) of the SNPs are presented according to their chromosomal positions (x axis). The genetic recombination rates (cM/Mb) estimated using the 1000 Genomes June 2014 ASN samples are shown with a blue line; we annotated the genes within the interested region and these genes are shown as arrows. The top genotyped SNP is labeled by rs ID and the *r*2 values of the rest of the SNPs with the top genotyped SNP are indicated by different colors.

****

**Supplementary Figure 4. Luciferase reporter assays to measure T or C allele difference at rs11169571 with the presence of hsa-miR-524-5p in LoVo and SW480 cells.** (**A**) Predicted effect of allelic variation at rs11169571 on hsa-miR-524-5p recognition and the construct containing full-length 3’-UTR of *ATF1* gene with different alleles of rs11169571. SNP rs11169571 occurs in the 8-bp seed sequence of complementarity at hsa-miR-524-5p 5’ end. Base pairing is indicated by a solid (Watson–Crick) vertical line. (**B**) Cells were transiently cotransfected with constructs and 80 nmol/L hsa-miR-524-5p mimic or negative control (NC). Results are shown as relative luciferase activity versus NC. Data are from 3 independent transfection experiments with assays conducted in 3 replications.

**Supplementary Table 1.** **Summary of characteristics of study subjects participated in this study**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Discovery stage | |  | Replication stage I | |  | Replication stage II | |  | Combined | |
| Cases  (N = 1,062) | Controls  (N = 2,184) |  | Cases  (N = 2,478) | Controls  (N = 3,880) |  | Cases  (N = 3,761) | Controls  (N = 4,058) |  | Cases  (N = 7,301) | Controls  (N = 10,122) |
| Age (years), mean±S.D. | 60.52±12.83 | 63.37±9.72 |  | 58.31±12.05 | 63.66±10.24 |  | 57.72±11.85 | 58.10±12.22 |  | 58.33±12.10 | 61.37±11.30 |
| Gender, n (%) |  |  |  |  |  |  |  |  |  |  |  |
| Male | 711 (66.9) | 1,448 (66.3) |  | 1,537 (62.0) | 2,439 (62.9) |  | 2,269 (60.3) | 2,522 (62.1) |  | 4,517 (61.9) | 6,409 (63.3) |
| Female | 351(33.1) | 736 (33.7) |  | 941 (38.0) | 1,441 (37.1) |  | 1,492 (39.7) | 1,536 (37.9) |  | 2,784 (38.1) | 3,713 (36.7) |

**Supplementary Table 2. The results of variants that were found to be potentially associated with risk of CRC at the**

**discovery stage and were selected for further replication**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Chr | SNP | Gene | Location | Effect Allele | MAF | | OR (95% CI)† | *P*† | OR (95% CI)‡ | *P*‡ |
| Cases | Controls |
| 12 | rs671 | ALDH2 | Missense | A | 0.250 | 0.193 | 1.37 (1.21-1.55) | 5.14×10-7 | 1.34 (1.18-1.52) | 4.59×10-6 |
| 12 | rs3782886 | BRAP | Synonymous | G | 0.254 | 0.196 | 1.37 (1.21-1.55) | 5.25×10-7 | 1.34 (1.18-1.51) | 4.65×10-6 |
| 12 | rs11169571 | ATF1 | 3'-UTR | G | 0.349 | 0.297 | 1.27 (1.13-1.42) | 3.91×10-5 | 1.28 (1.15-1.44) | 1.66×10-5 |
| 12 | rs1129406 | ATF1 | Synonymous | A | 0.349 | 0.297 | 1.27 (1.13-1.42) | 3.91×10-5 | 1.28 (1.15-1.44) | 1.66×10-5 |
| 10 | rs138649767 | TCF7L2 | Missense | A | 0.022 | 0.009 | 2.35 (1.53-3.60) | 9.95×10-5 | 2.44 (1.59-3.76) | 4.90×10-5 |
| 6 | rs2395269 | TAP1 | Intron | C | 0.274 | 0.214 | 1.26 (1.12-1.43) | 1.63×10-4 | 1.27 (1.12-1.43) | 1.34×10-4 |
| 17 | rs2304977 | KIAA0753 | Missense | A | 0.356 | 0.309 | 1.24 (1.11-1.38) | 1.66×10-4 | 1.24 (1.11-1.38) | 1.57×10-4 |
| 3 | rs145045076 | CNTN6 | Missense | G | 0.036 | 0.019 | 1.84 (1.33-2.54) | 2.03×10-4 | 1.83 (1.32-2.52) | 2.43×10-4 |
| 6 | rs406113 | GPX6 | Missense | A | 0.439 | 0.488 | 0.82 (0.74-0.91) | 1.68×10-4 | 0.82 (0.74-0.91) | 2.45×10-4 |
| 2 | rs1377656 | LOC105376755 | Intron | A | 0.416 | 0.366 | 1.22 (1.10-1.36) | 2.13×10-4 | 1.22 (1.10-1.35) | 2.49×10-4 |
| 6 | rs2270190 | C6orf15 | Upstream | G | 0.109 | 0.083 | 1.36 (1.14-1.62) | 5.00×10-4 | 1.38 (1.16-1.65) | 2.97×10-4 |
| 8 | rs13257322 | LOC102724623 | Intron | A | 0.452 | 0.500 | 0.82 (0.74-0.91) | 2.76×10-4 | 0.82 (0.74-0.92) | 3.21×10-4 |
| 1 | rs2840532 | RER1 | Intron | A | 0.414 | 0.462 | 0.83 (0.75-0.92) | 5.42×10-4 | 0.82 (0.74-0.92) | 3.35×10-4 |
| 10 | rs2489392 | DOCK1 | Intron | G | 0.217 | 0.256 | 0.79 (0.70-0.89) | 1.96×10-4 | 0.80 (0.70-0.90) | 3.62×10-4 |
| 15 | rs74505897 | ACAN | Missense | A | 0.046 | 0.070 | 0.65 (0.51-0.82) | 3.43×10-4 | 0.65 (0.51-0.82) | 3.74×10-4 |
| 1 | rs61744267 | TPR | Missense | A | 0.094 | 0.067 | 1.41 (1.17-1.70) | 3.15×10-4 | 1.40 (1.16-1.69) | 4.28×10-4 |
| 4 | rs3733197 | BANK1 | Missense | A | 0.179 | 0.218 | 0.78 (0.69-0.89) | 2.84×10-4 | 0.79 (0.69-0.90) | 4.36×10-4 |
| 2 | rs75011196 | XIRP2 | Missense | G | 0.036 | 0.020 | 1.76 (1.29-2.41) | 3.95×10-4 | 1.75 (1.28-2.40) | 4.55×10-4 |
| 3 | rs2271077 | GALNTL2 | Missense | A | 0.219 | 0.182 | 1.26 (1.11-1.44) | 4.37×10-4 | 1.26 (1.11-1.44) | 4.60×10-4 |

Note: Chr, chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval. †*P* values are two sided and were calculated by an additive model in logistic regression analysis adjusted for sex and age. ‡*P* values are two sided and were calculated by an additive model in logistic regression analysis adjusted for sex, age and the first three principle components.

**Supplementary Table 3. The results of variants that were found to be potentially associated with risk of CRC at the replication stage**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SNP | Replication stage I | | | |  | Replication stage II | | | |  | Combined replication samples | | | |
| EAF | | OR (95% CI) | *P* |  | EAF | | OR (95% CI) | *P* |  | EAF | | OR (95% CI) | *P* |
| Cases | Controls |  | Cases | Controls |  | Cases | Controls |
| rs671 | 0.196 | 0.199 | 0.98 (0.89-1.08) | 0.6816 |  |  |  |  |  |  |  |  |  |  |
| rs3782886 | 0.207 | 0.202 | 1.03 (0.94-1.13) | 0.5266 |  |  |  |  |  |  |  |  |  |  |
| **rs11169571** | **0.333** | **0.303** | **1.14 (1.05-1.24)** | **0.0011** |  | **0.337** | **0.302** | **1.18 (1.10-1.27)** | **2.07×10-6** |  | **0.335** | **0.303** | **1.17 (1.11-1.23)** | **4.06×10-9** |
| **rs138649767** | **0.015** | **0.008** | **2.12 (1.49-3.02)** | **3.16×10-5** |  | **0.014** | **0.007** | **2.09 (1.50-2.92)** | **1.37×10-5** |  | **0.014** | **0.007** | **2.07 (1.63-2.63)** | **2.51×10-9** |
| rs2395269 | 0.216 | 0.216 | 1.01 (0.92-1.10) | 0.8853 |  |  |  |  |  |  |  |  |  |  |
| rs2304977 | 0.313 | 0.324 | 0.94 (0.87-1.02) | 0.1503 |  |  |  |  |  |  |  |  |  |  |
| rs145045076 | 0.020 | 0.024 | 0.82 (0.63-1.05) | 0.1148 |  |  |  |  |  |  |  |  |  |  |
| rs406113 | 0.455 | 0.455 | 1.00 (0.93-1.07) | 0.9961 |  |  |  |  |  |  |  |  |  |  |
| rs1377656 | 0.372 | 0.363 | 1.04 (0.97-1.13) | 0.2694 |  |  |  |  |  |  |  |  |  |  |
| rs2270190 | 0.084 | 0.089 | 0.93 (0.82-1.06) | 0.3051 |  |  |  |  |  |  |  |  |  |  |
| rs13257322 | 0.492 | 0.475 | 1.05 (0.98-1.13) | 0.1556 |  |  |  |  |  |  |  |  |  |  |
| rs2840532 | 0.446 | 0.457 | 0.96 (0.89-1.03) | 0.2492 |  |  |  |  |  |  |  |  |  |  |
| rs2489392 | 0.243 | 0.236 | 1.04 (0.96-1.13) | 0.3526 |  |  |  |  |  |  |  |  |  |  |
| rs74505897 | 0.067 | 0.070 | 0.95 (0.82-1.10) | 0.5094 |  |  |  |  |  |  |  |  |  |  |
| rs61744267 | 0.065 | 0.072 | 0.90 (0.78-1.04) | 0.1555 |  |  |  |  |  |  |  |  |  |  |
| rs3733197 | 0.208 | 0.217 | 0.94 (0.86-1.03) | 0.1927 |  |  |  |  |  |  |  |  |  |  |
| rs75011196 | 0.022 | 0.023 | 0.97 (0.76-1.25) | 0.8348 |  |  |  |  |  |  |  |  |  |  |
| rs2271077 | 0.181 | 0.190 | 0.92 (0.84-1.01) | 0.0974 |  |  |  |  |  |  |  |  |  |  |

Note: Chr, chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval. *P* values are two sided and were calculated by an additive model in logistic regression analysis adjusted for sex and age. Two SNPs with *P* < 0.0028 (0.05/18) in replication stage I were selected for further genotyping in replication stage II. Variants passed the significant threshold were in bold.