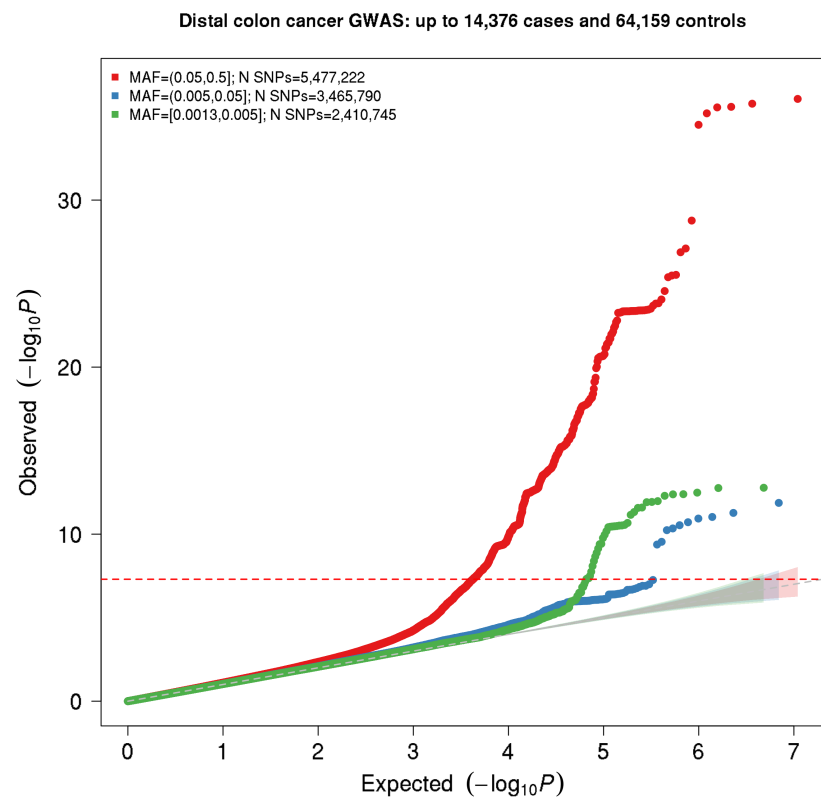
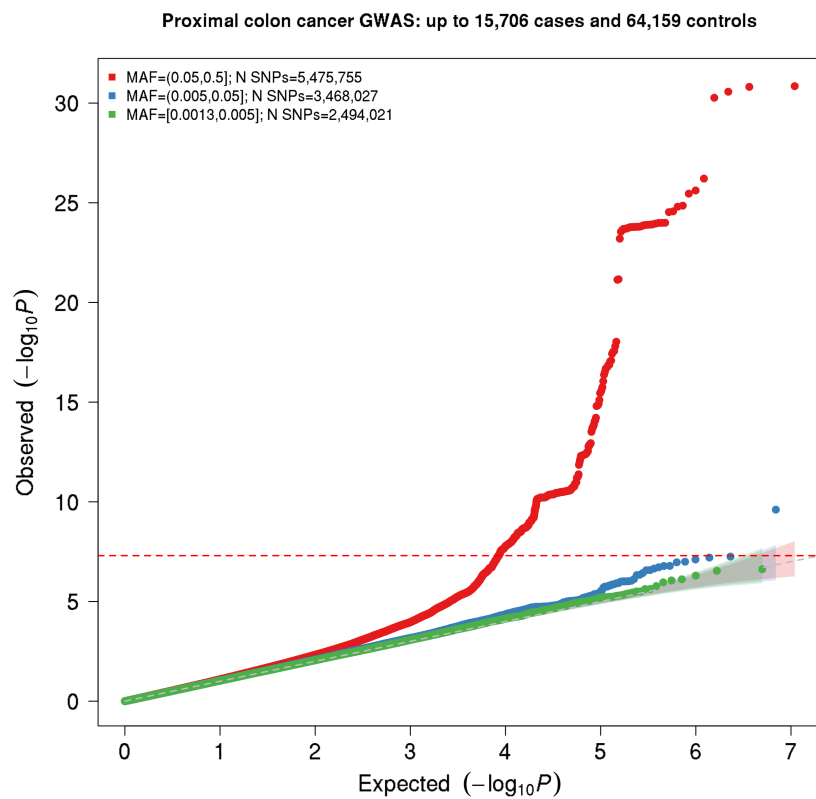
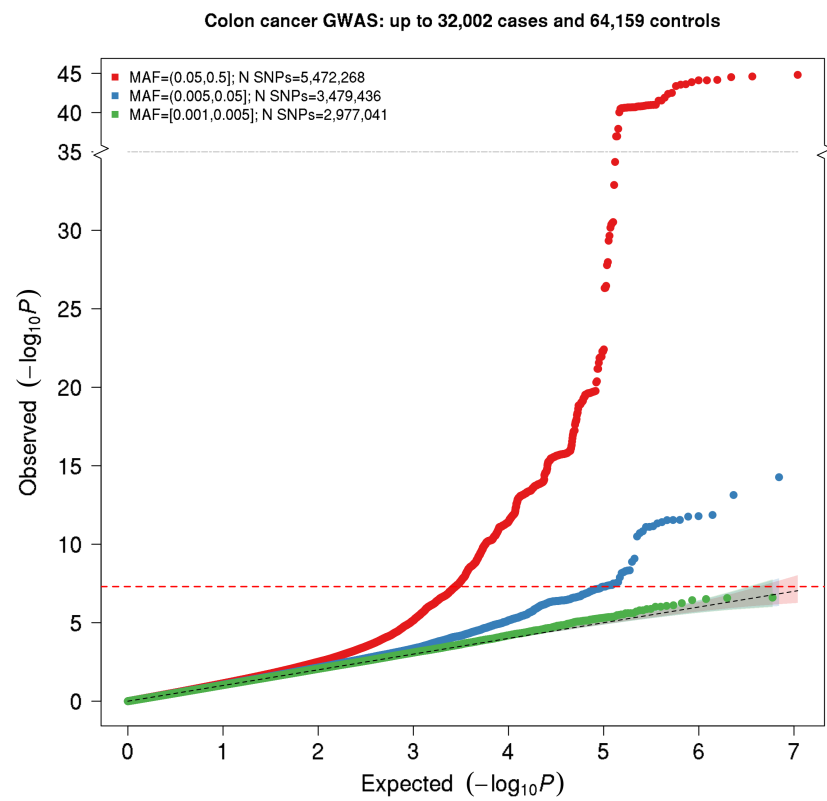
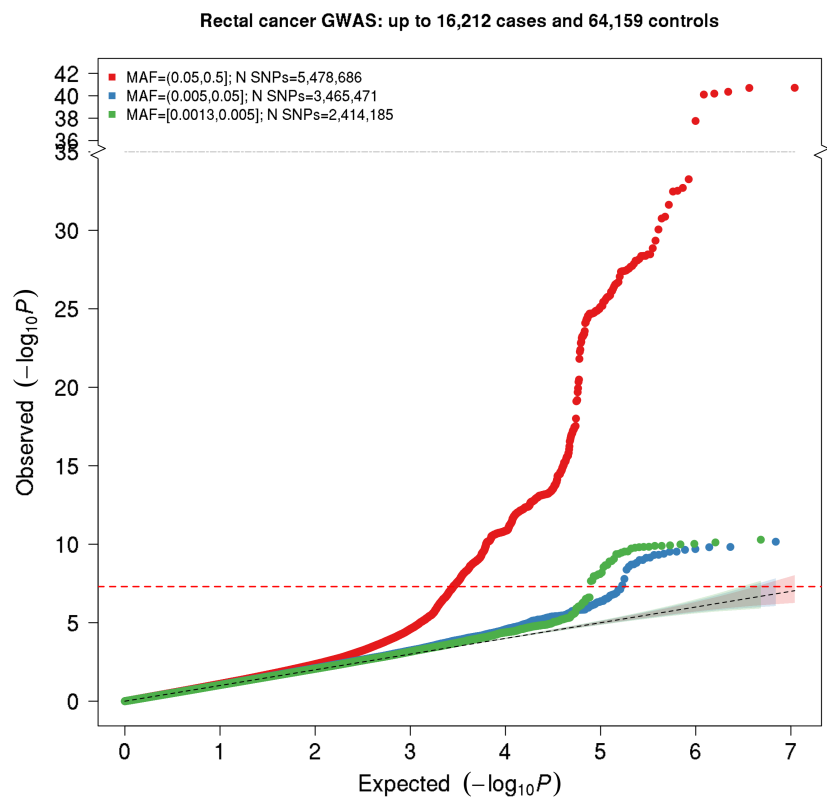


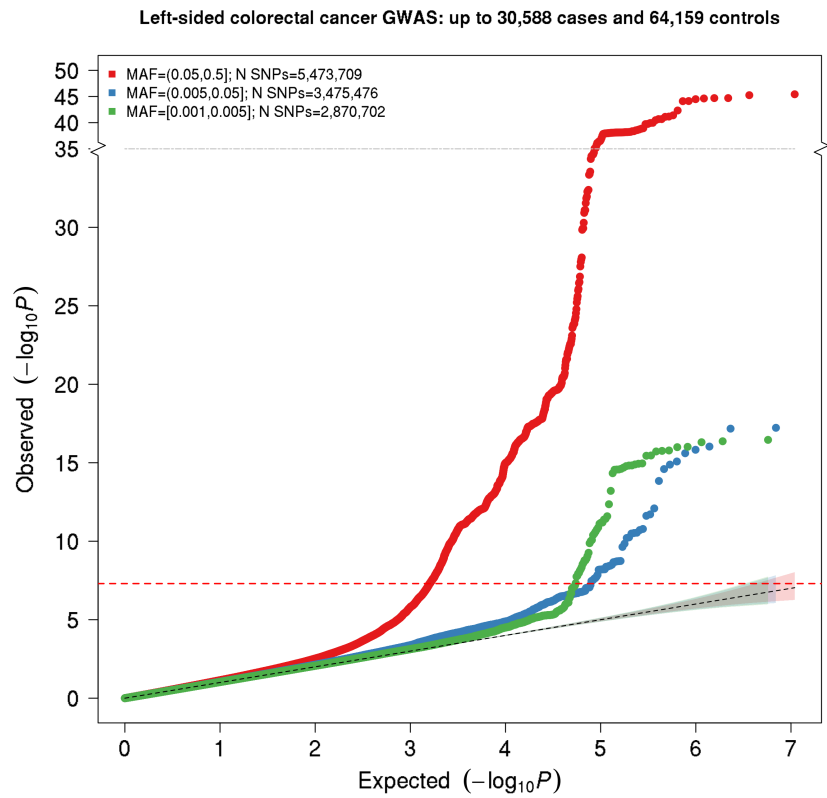
**Supplementary figure 1. Distributions of age of diagnosis by primary tumor anatomic subsite and sex.** Note that our data recapitulate the previously reported higher percentage of female proximal colon cancer cases, a male-to-female ratio that increases progressively from the proximal colon to the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



**Supplementary figure 2. Quantile-quantile (QQ) plots stratified by minor allele frequency (MAF) bins for the five GWAS meta-analyses of CRC case subgroups defined by primary tumor anatomic subsite.** GWAS studies were imputed to the Haplotype Reference Consortium (HRC) panel. The red dashed line indicates the genome-wide significance threshold ( $P=5 \times 10^{-8}$ ). The transparent regions around the equality line represent the analytically estimated 95% confidence bands for each MAF bin.

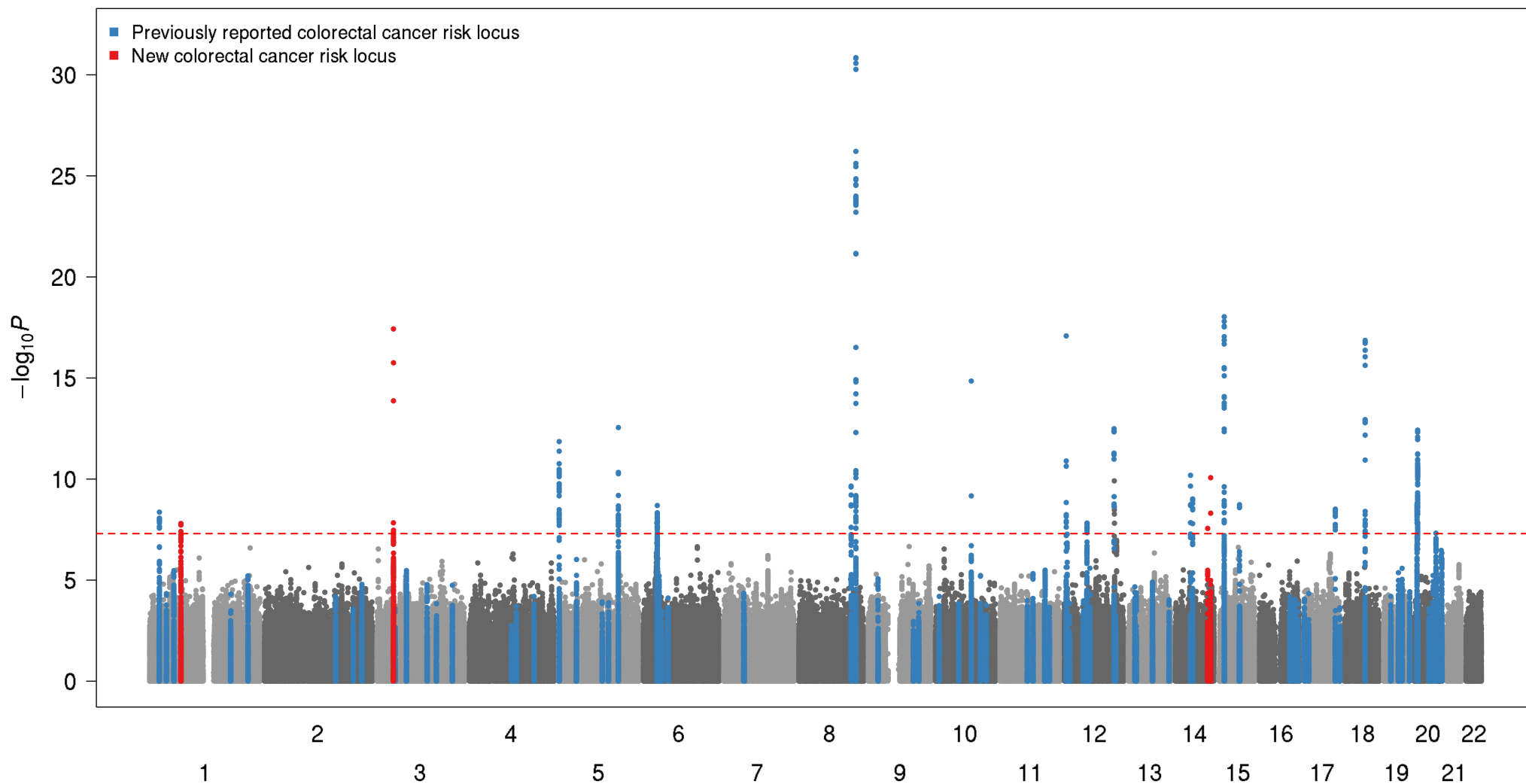


**Supplementary figure 2. (continued)**



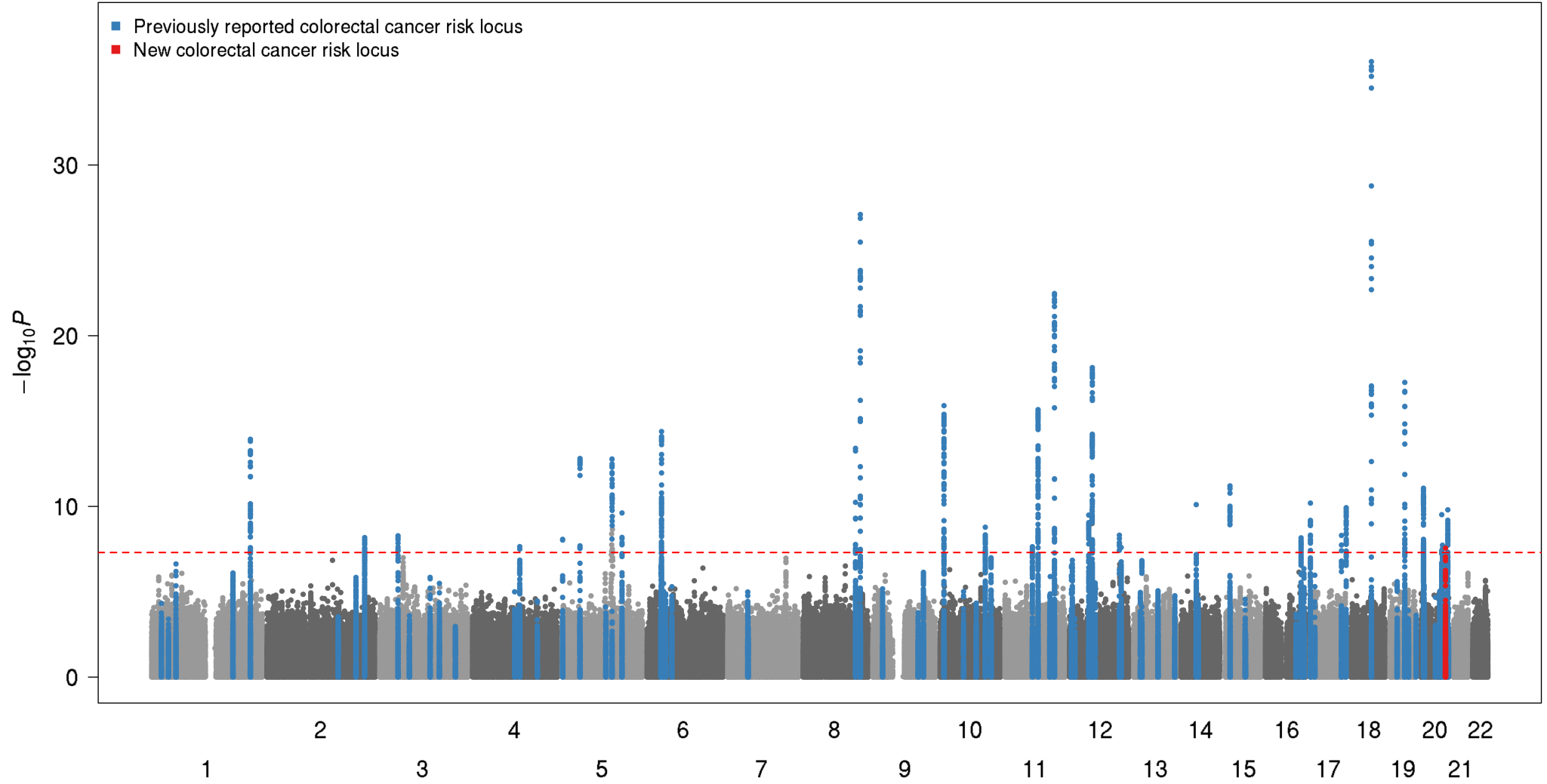
Supplementary figure 2. (continued)

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



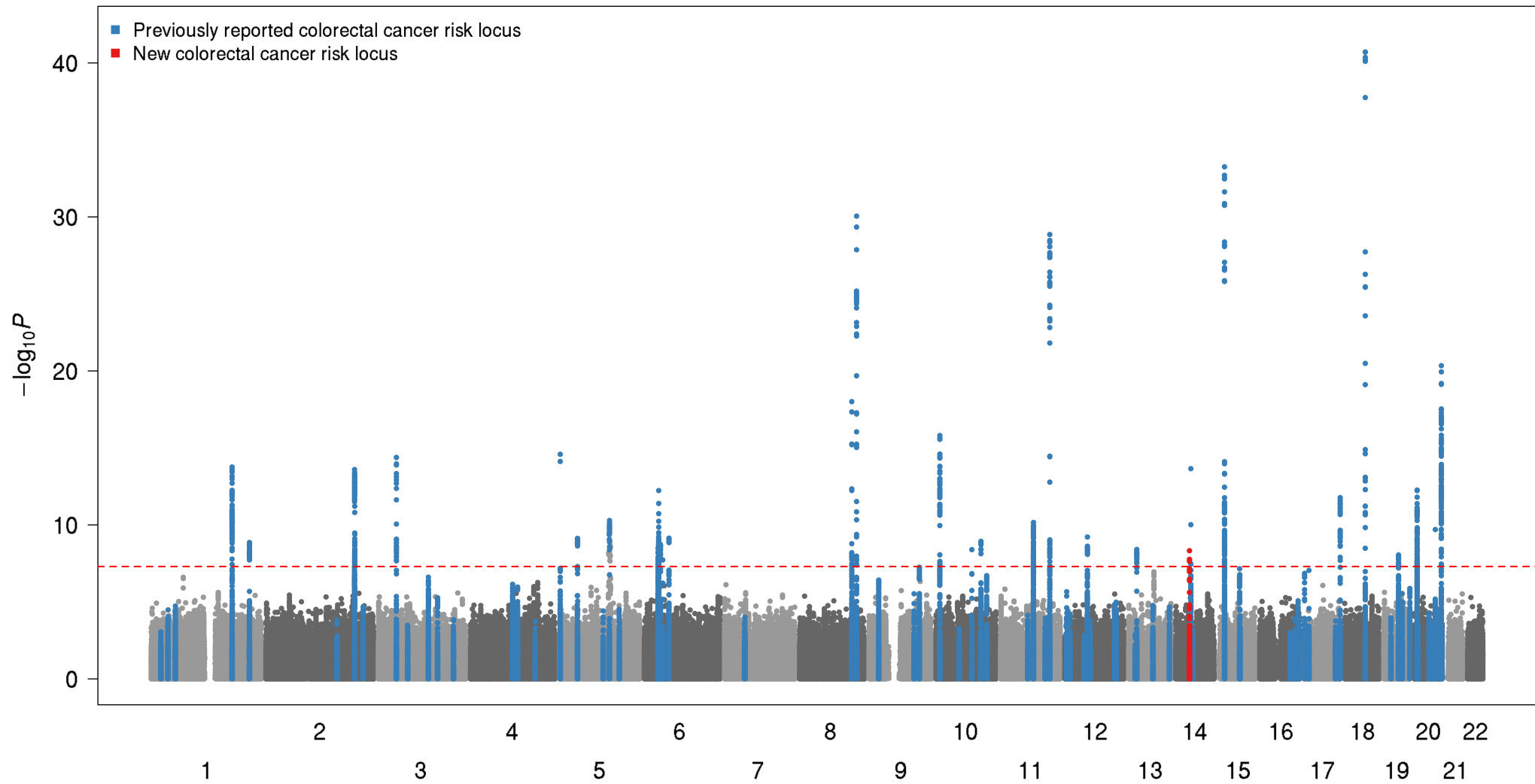
**Supplementary figure 3. Manhattan plots showing results of the five GWAS meta-analyses of CRC case subgroups defined by primary tumor anatomic subsite.** GWAS studies were imputed to the Haplotype Reference Consortium (HRC) panel. Association results for each variant ( $-\log_{10} P$  values) are plotted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

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



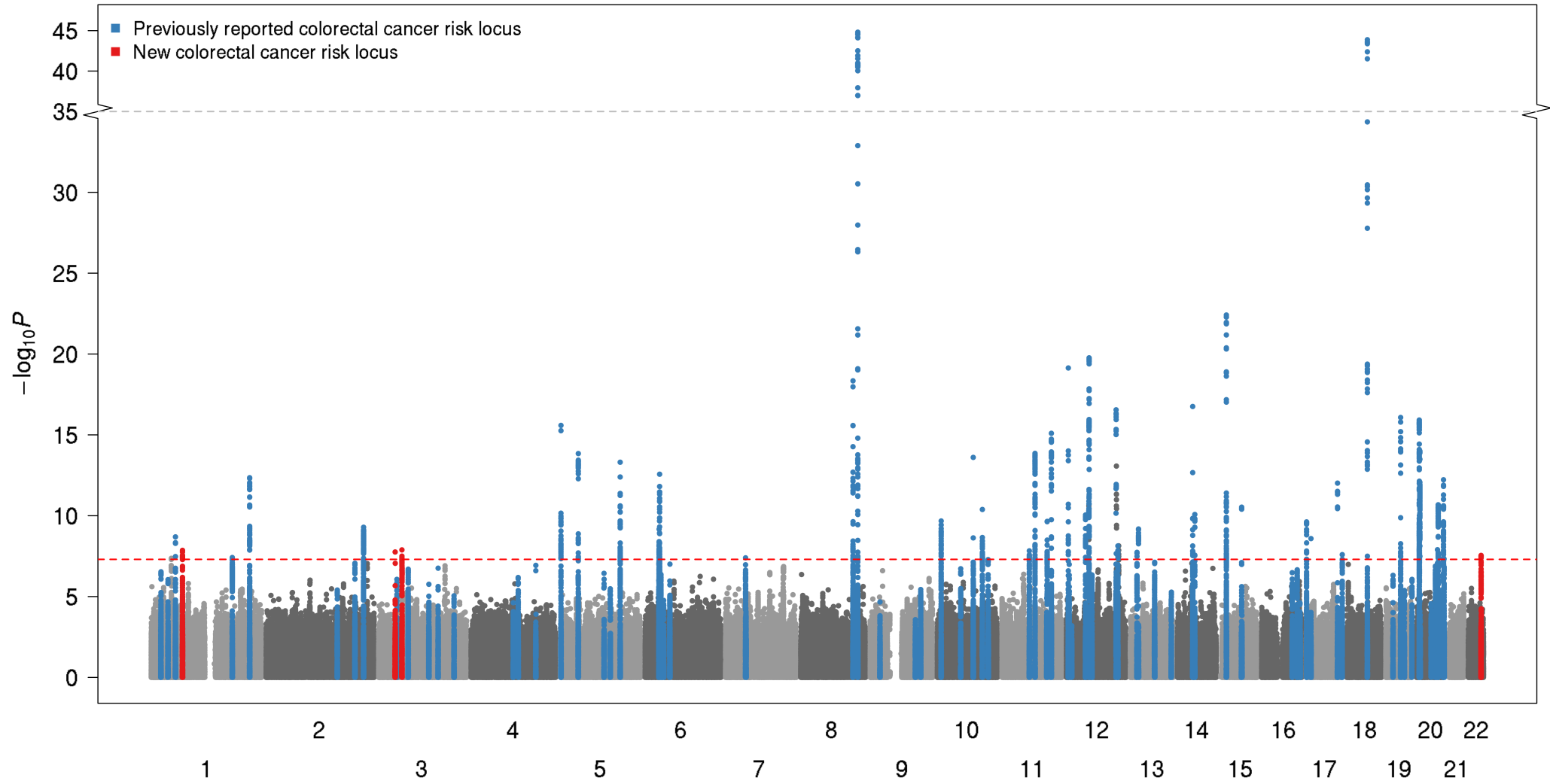
Supplementary figure 3. (continued)

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



Supplementary figure 3. (continued)

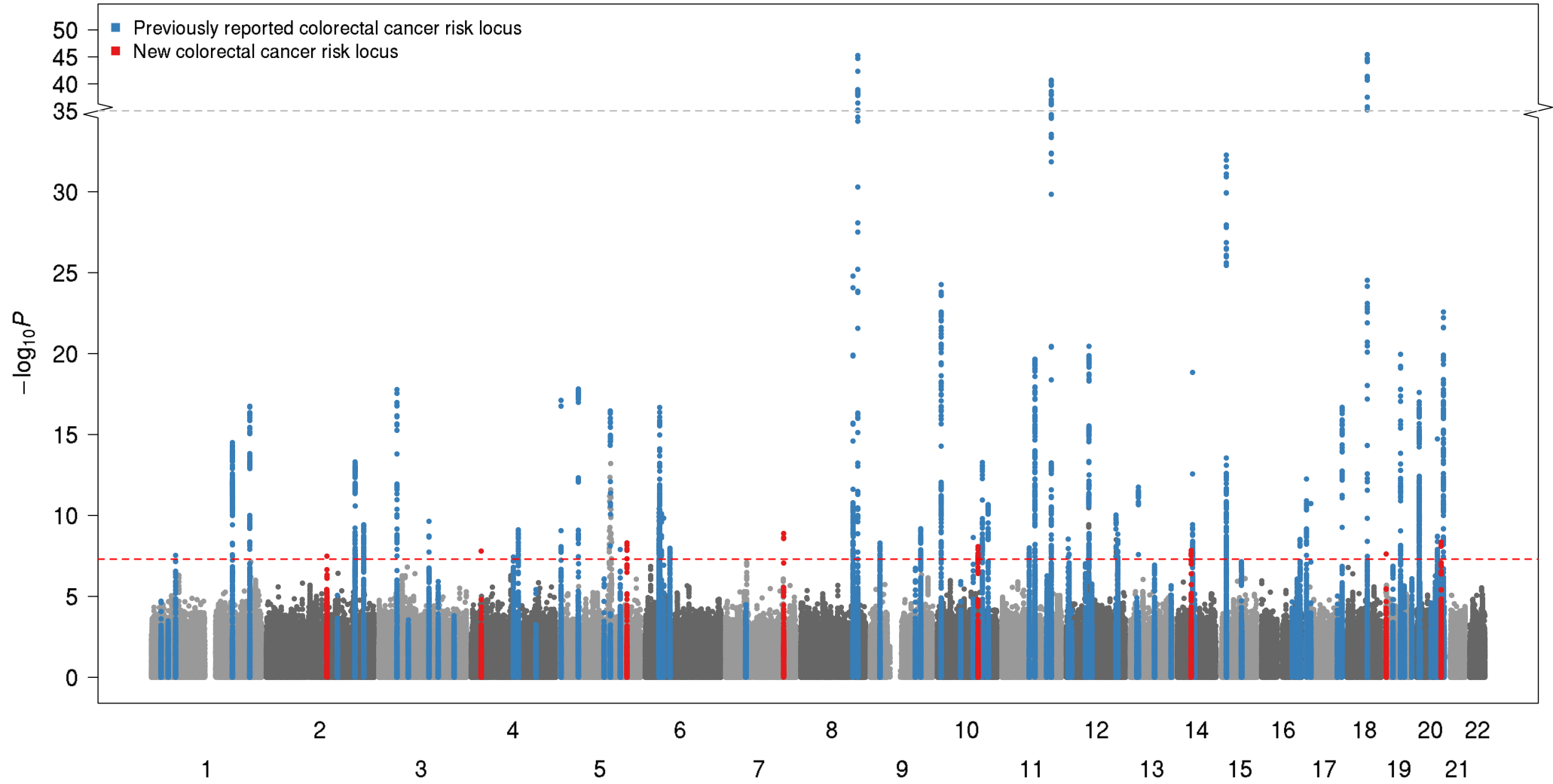
Colon cancer GWAS: up to 32,002 cases and 64,159 controls  
11,928,745 variants



Supplementary figure 3. (continued)

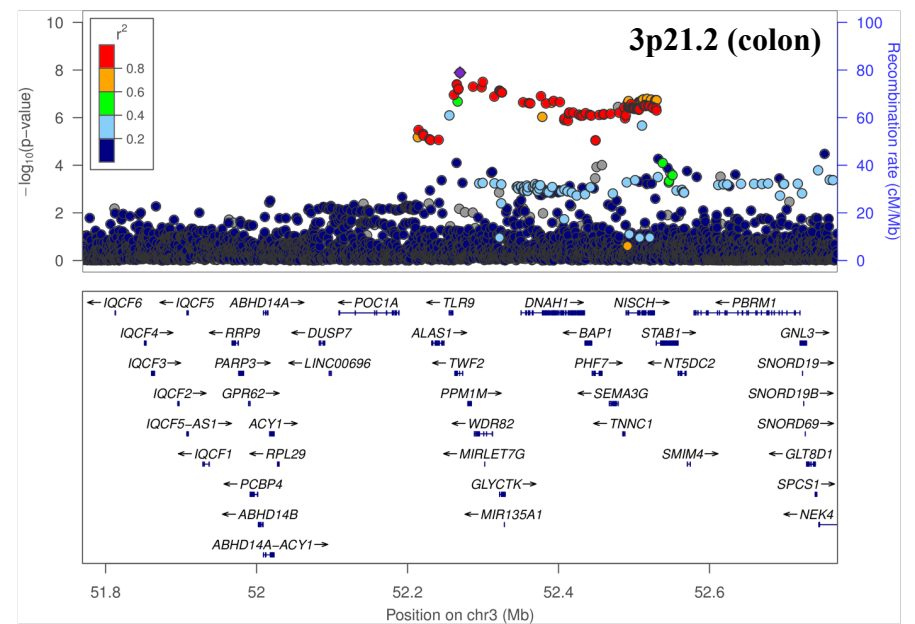
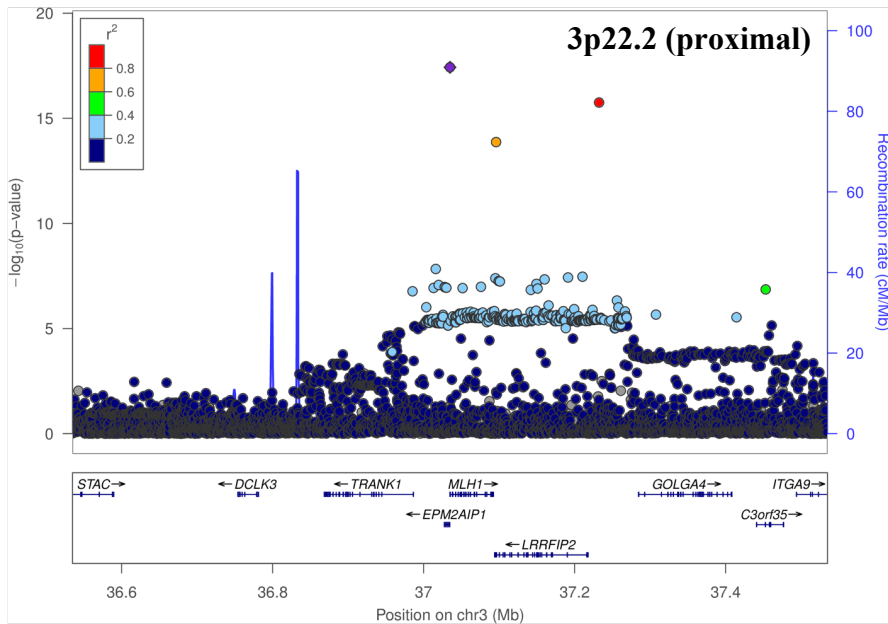
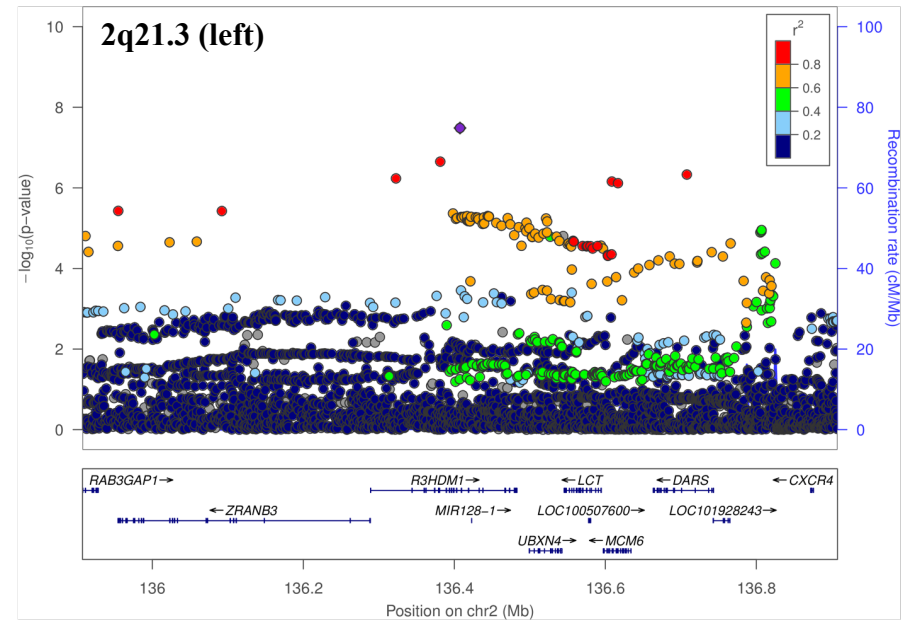
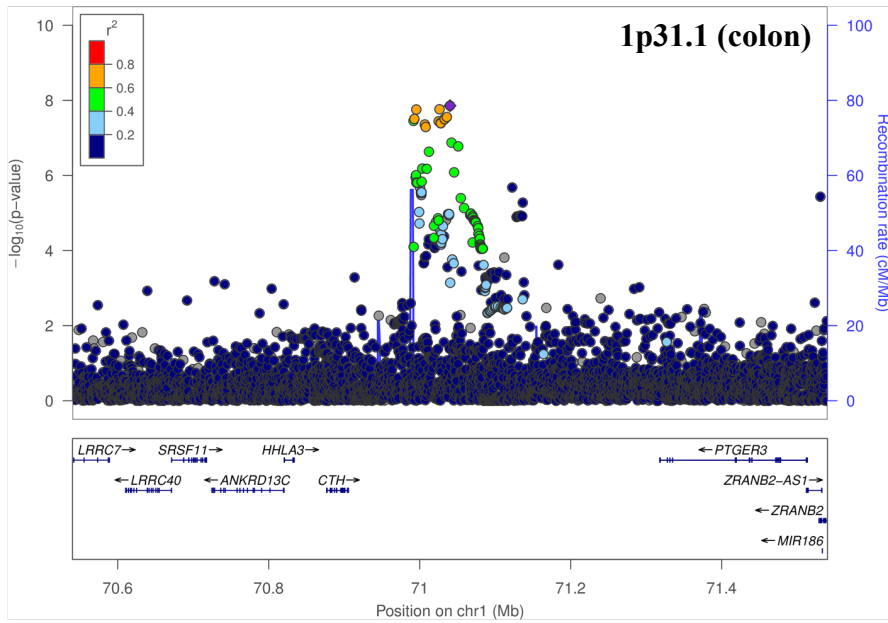


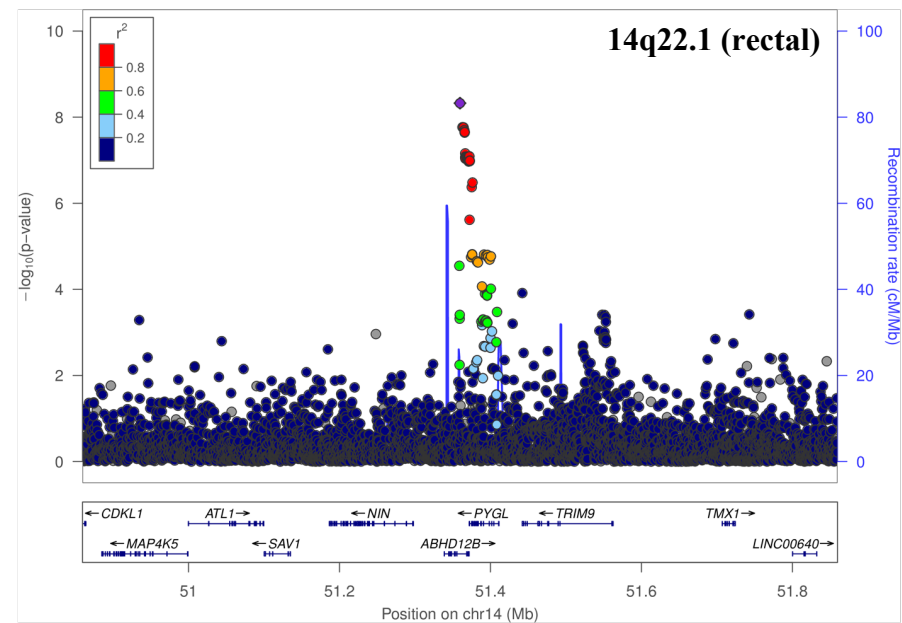
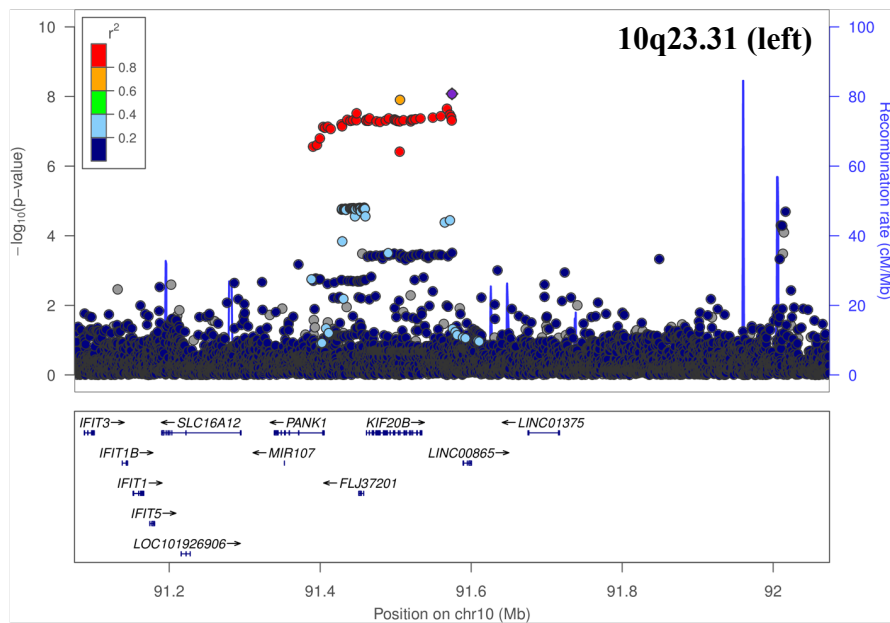
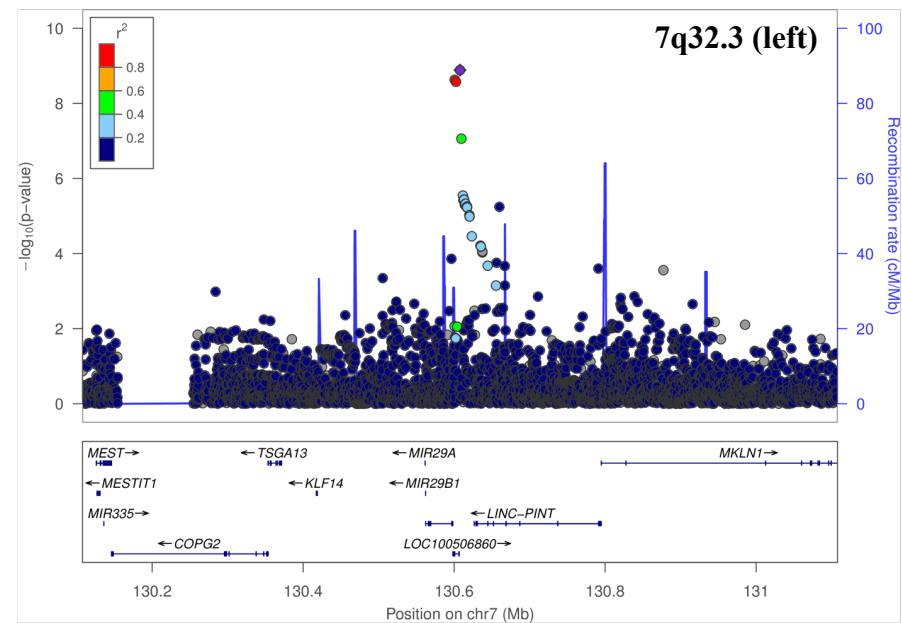
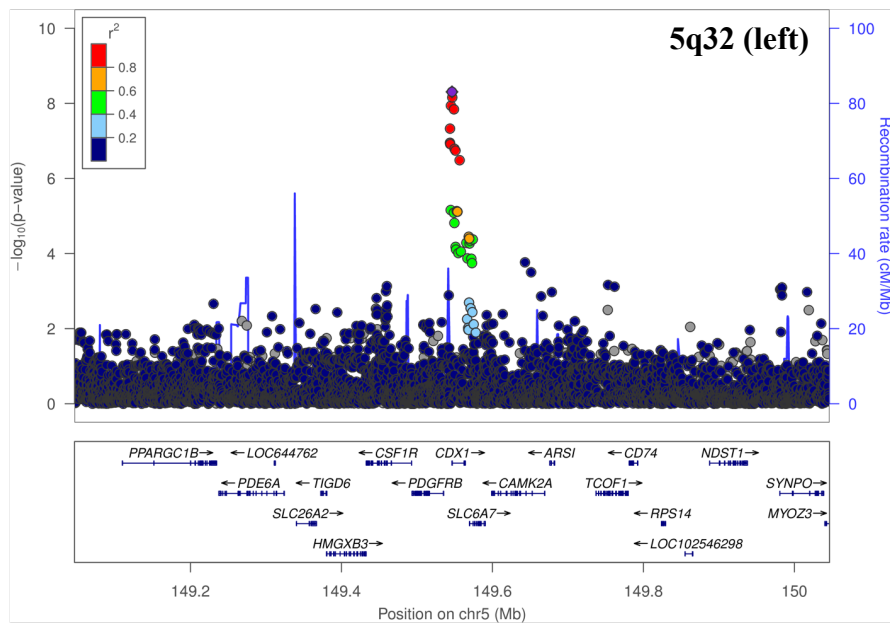
Left-sided colorectal cancer GWAS: up to 30,588 cases and 64,159 controls  
11,819,887 variants

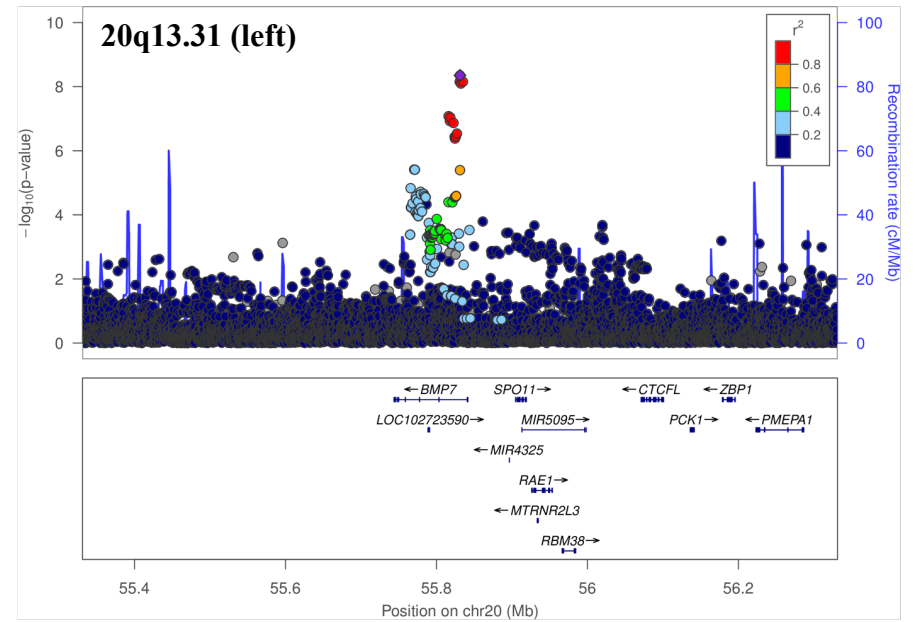
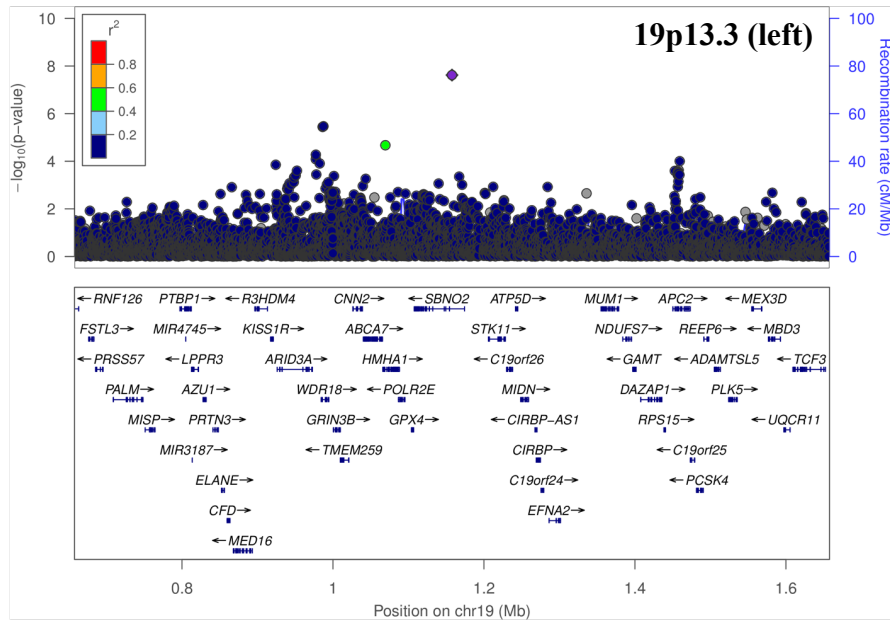
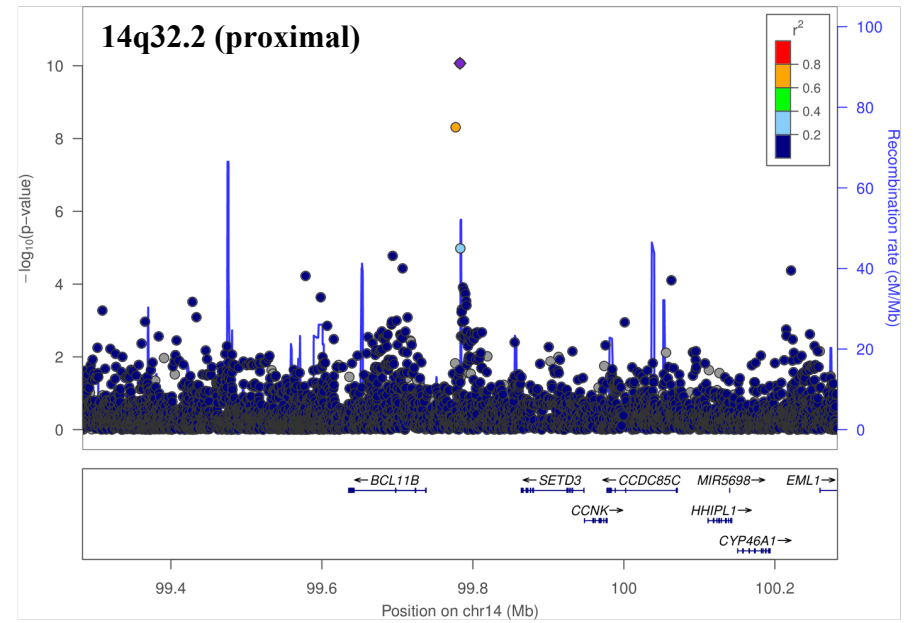
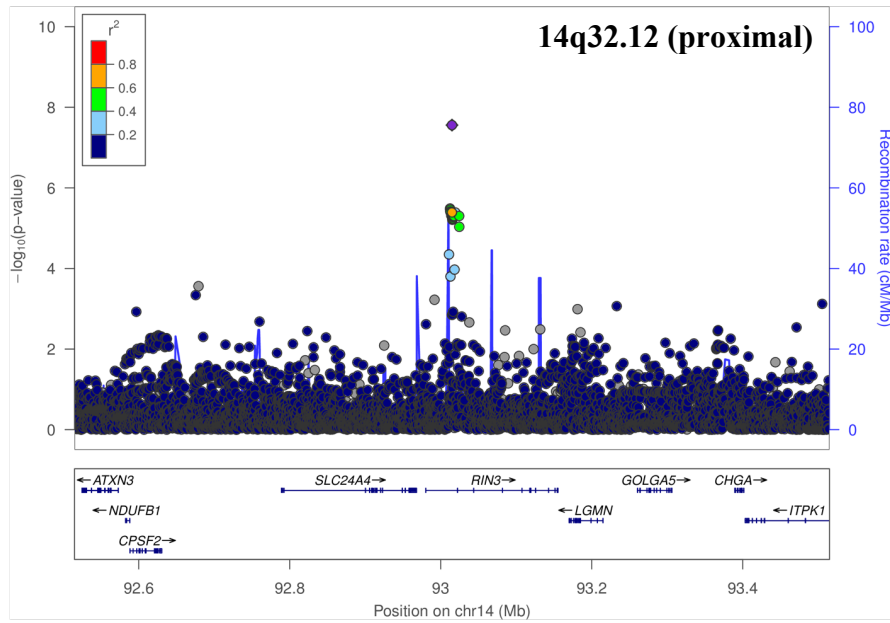


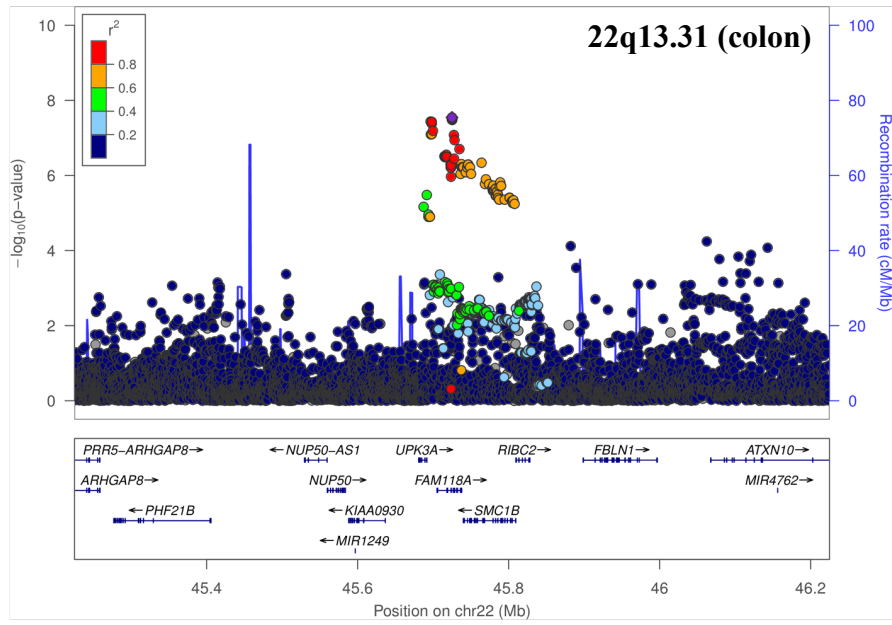
Supplementary figure 3. (continued)

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

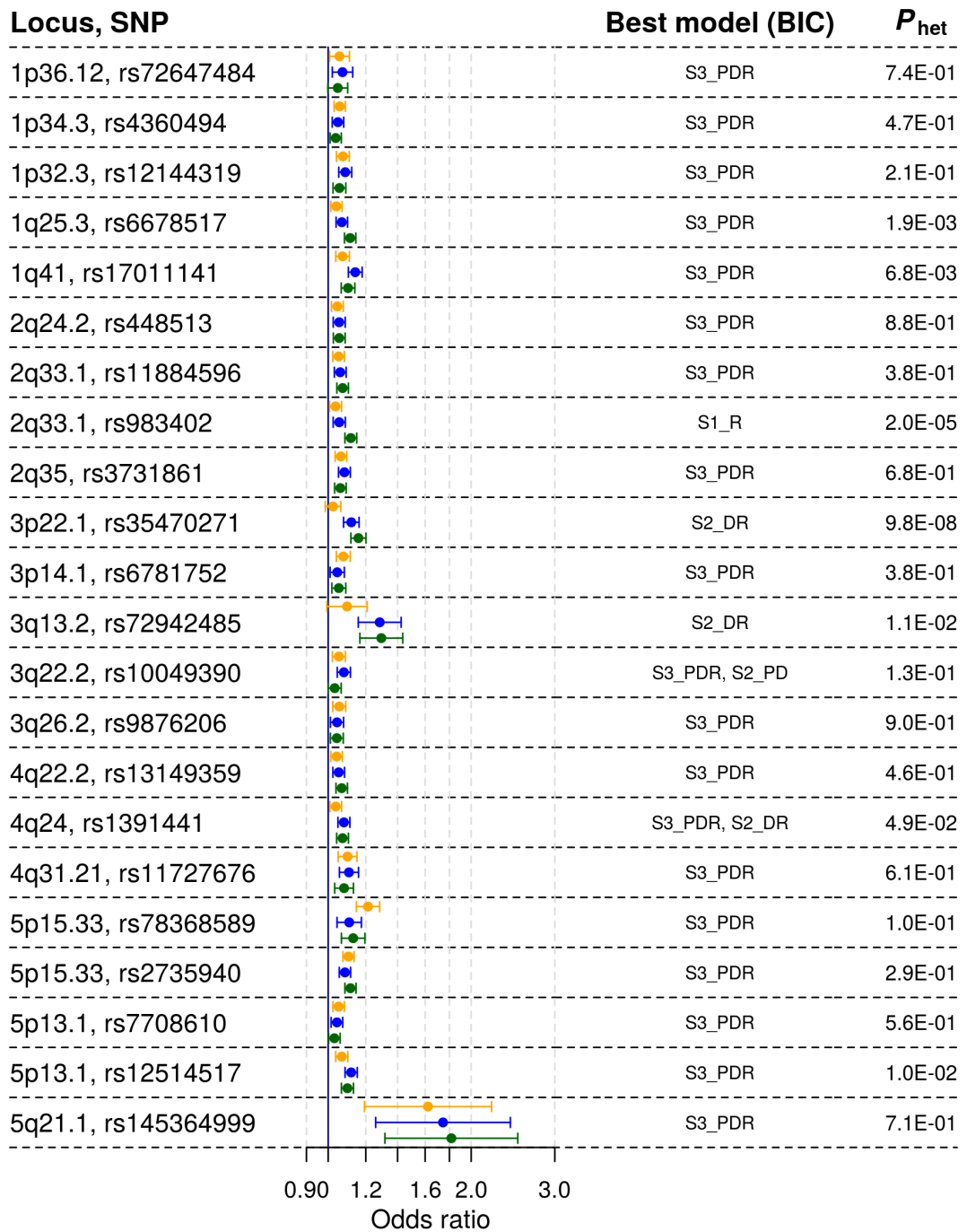






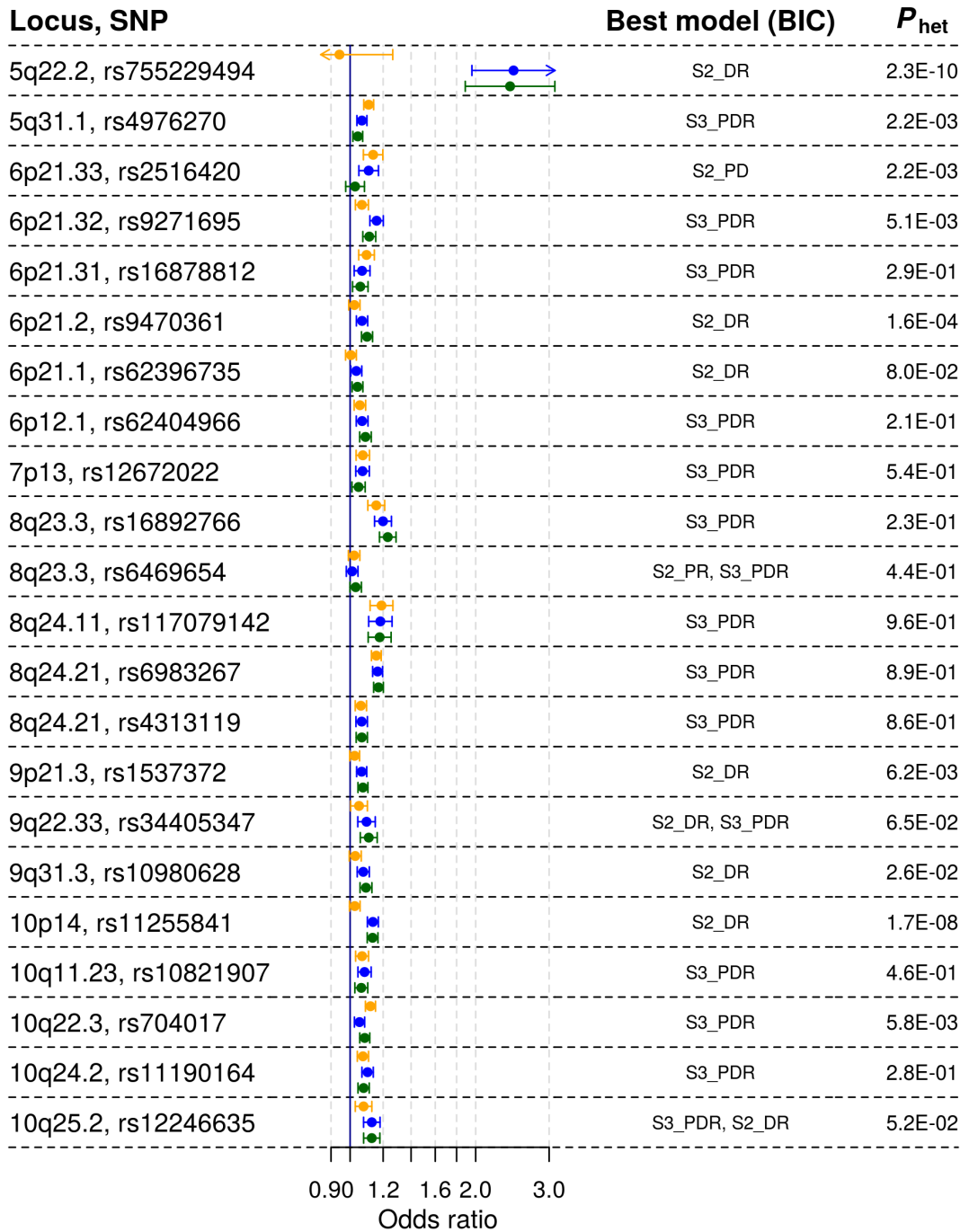


● proximal colon cancer ● distal colon cancer ● rectal cancer



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

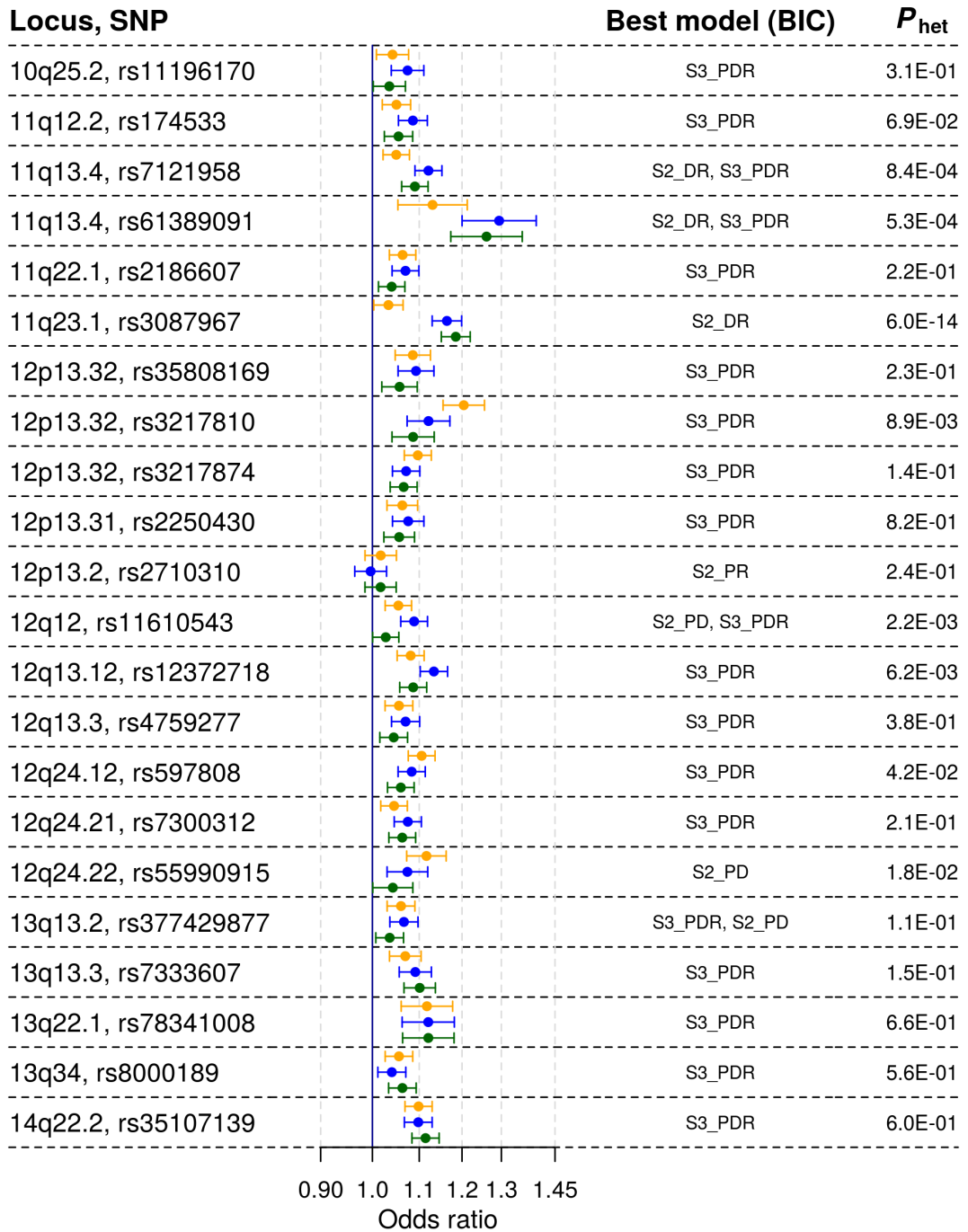
● proximal colon cancer ● distal colon cancer ● rectal cancer



Supplementary figure 5 (continued).

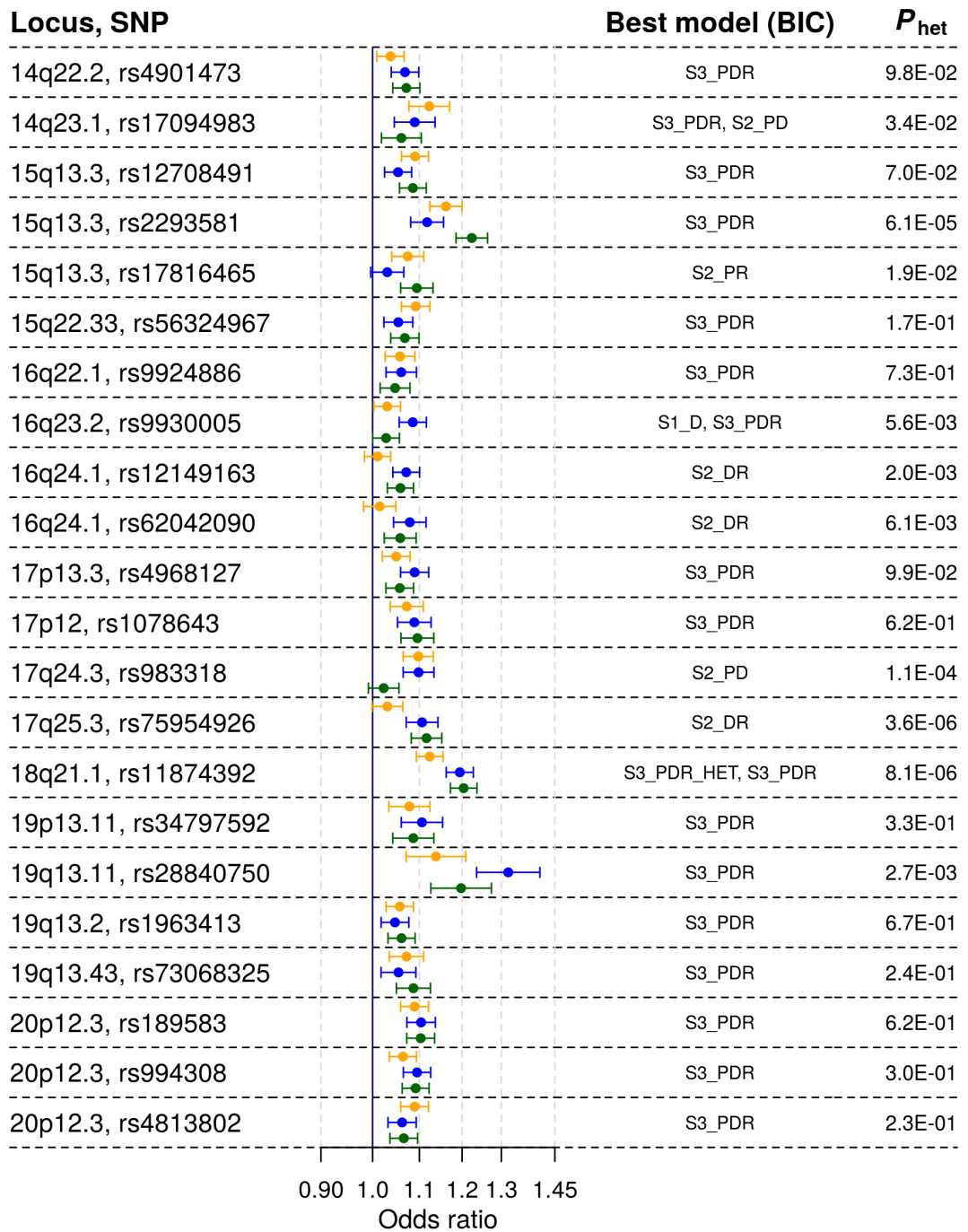


● proximal colon cancer ● distal colon cancer ● rectal cancer



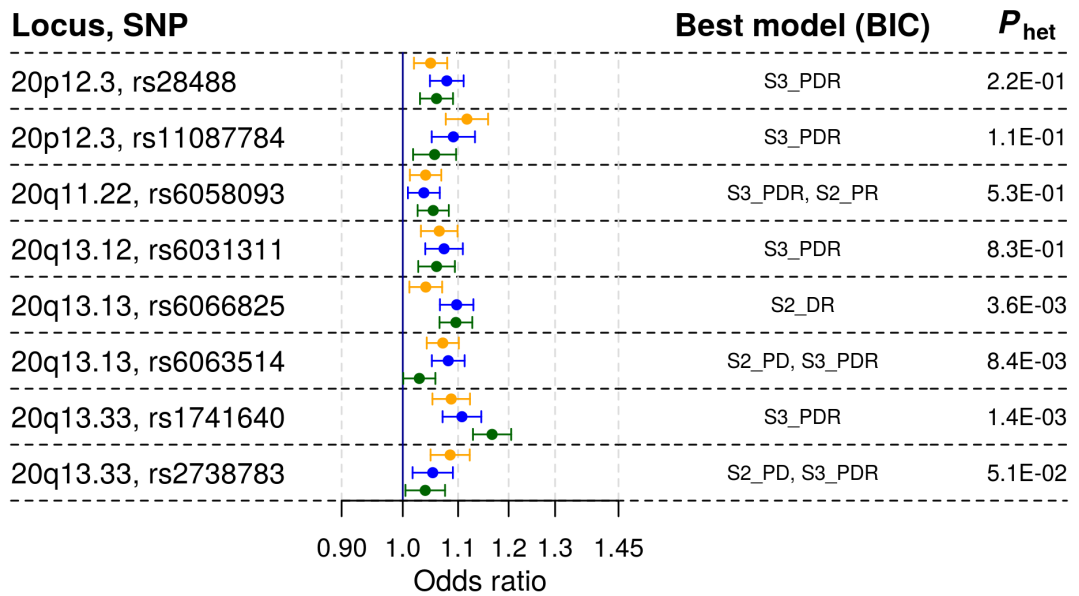
Supplementary figure 5 (continued).

● proximal colon cancer ● distal colon cancer ● rectal cancer



Supplementary figure 5 (continued).

● proximal colon cancer ● distal colon cancer ● rectal cancer



Supplementary figure 5 (continued).