The American Journal of Human Genetics, Volume *94*

Supplemental Data

GeMes: Clusters of DNA Methylation under Genetic Control Can Inform Genetic and Epigenetic Analysis of Disease

Yun Liu, Xin Li, Martin J. Aryee, Tomas J. Ekström, Leonid Padyukov, Lars Klareskog, Amy Vandiver, Ann Zenobia Moore, Toshiko Tanaka, Luigi Ferrucci, M. Daniele Fallin, and Andrew P. Feinberg

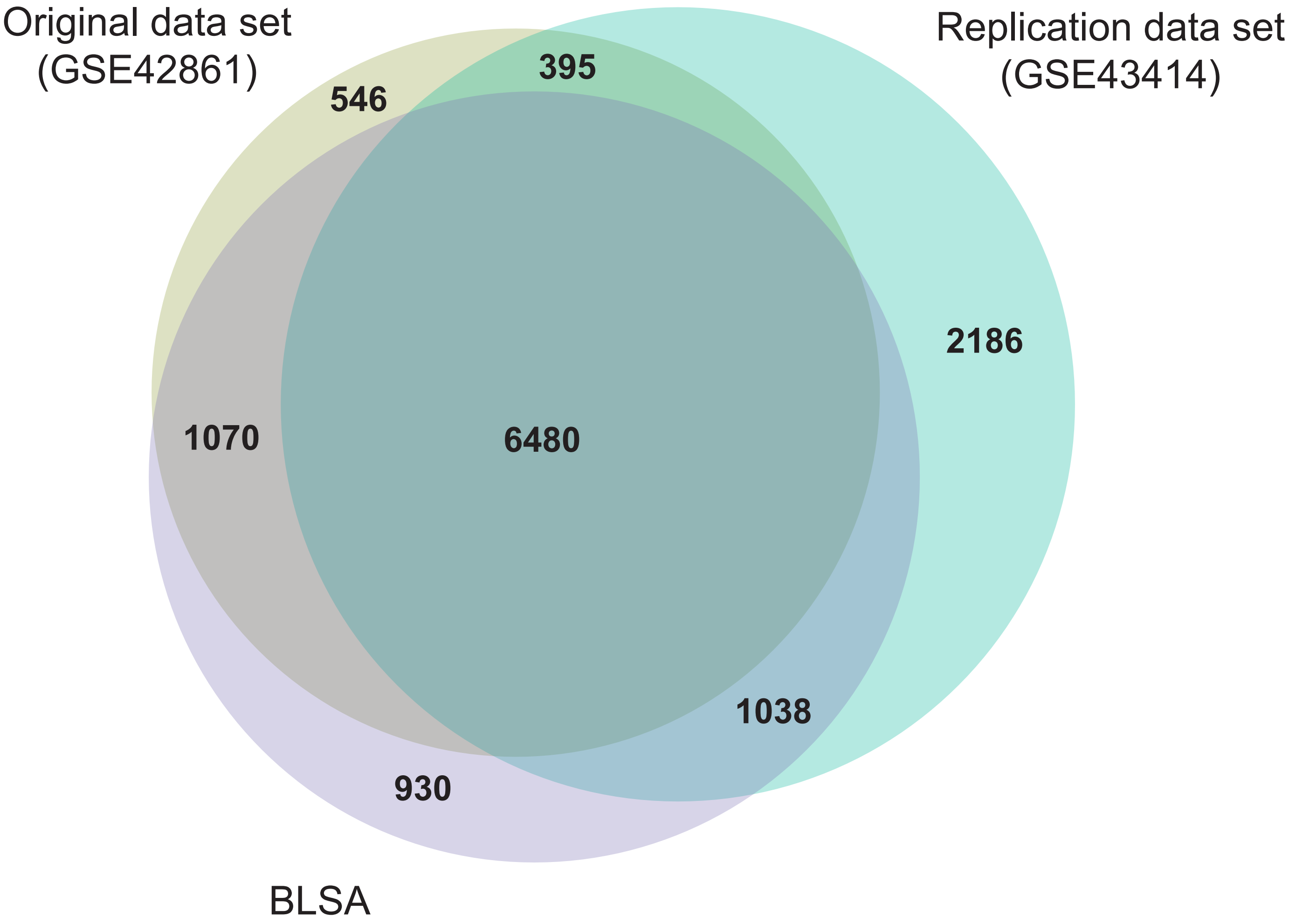


Figure S1. Venn diagram of contiguous methylation clusters identified from whole blood using three independent data sets.

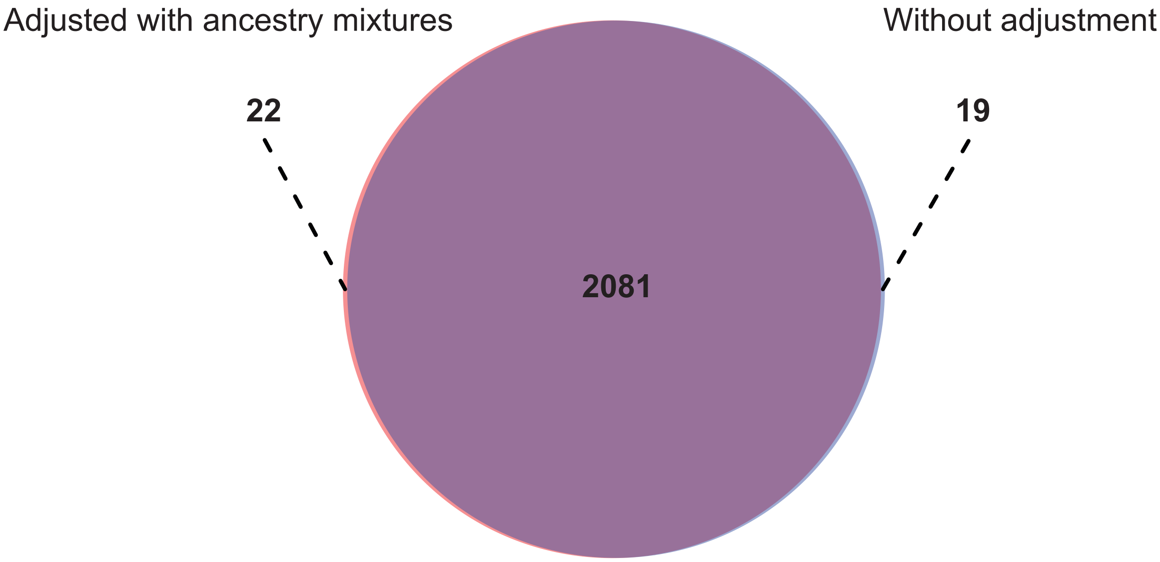


Figure S2. Venn diagram of contiguous methylation clusters identified from whole blood with or without adjusting for ancestry mixtures.

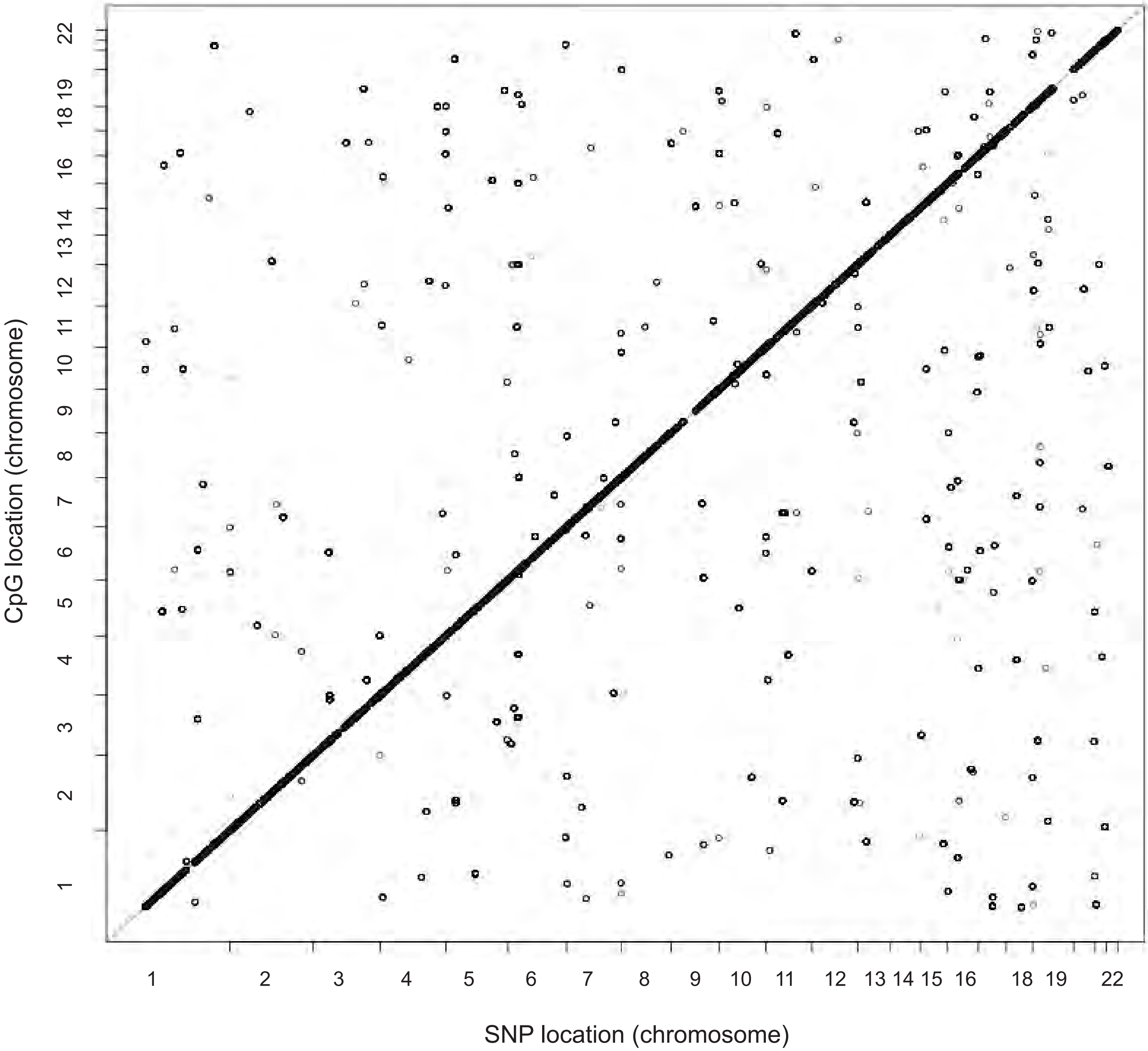


Figure S3. Distribution of significant vCpGs and SNPs with which they associate. Each point represents a significant association between a vCpG site and a SNP. The color of the dot indicates significance of the *P value*.

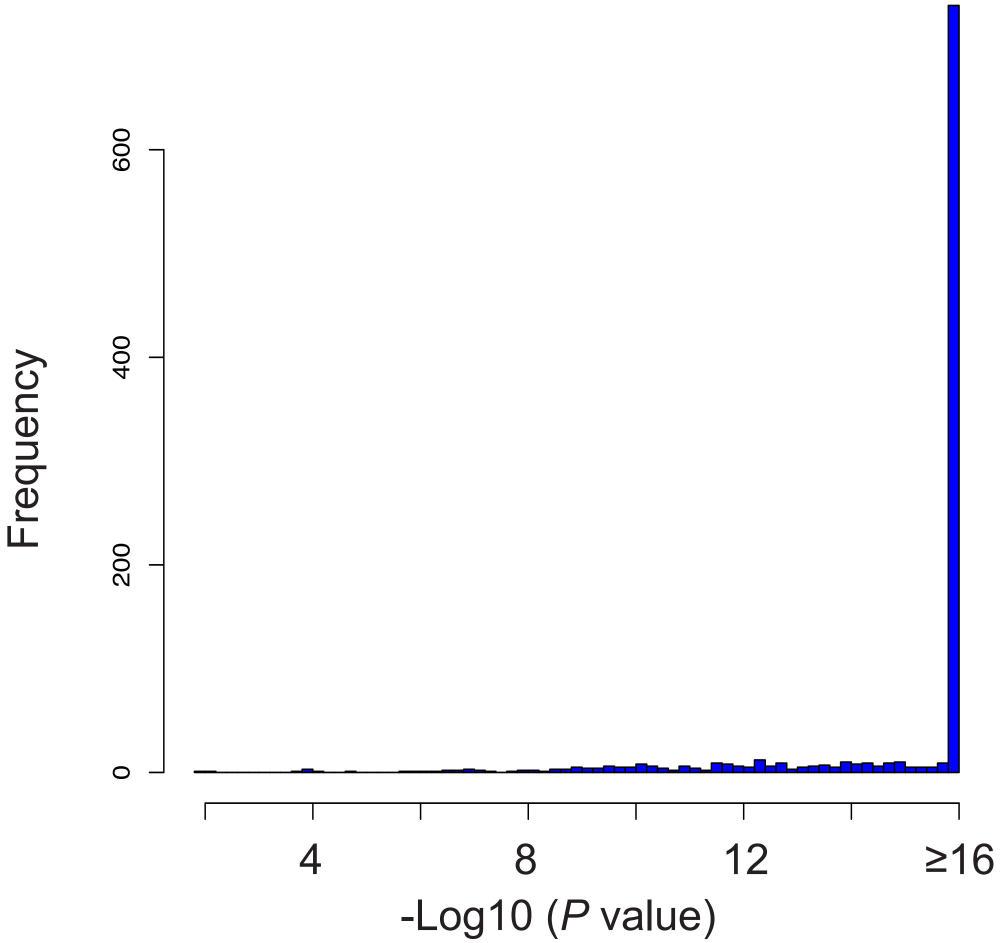


Figure S4. -Log10 (*P* value) distribution of genotype-methylation associations of 991 CpG-SNP pairs tested in the BLSA data set.

A

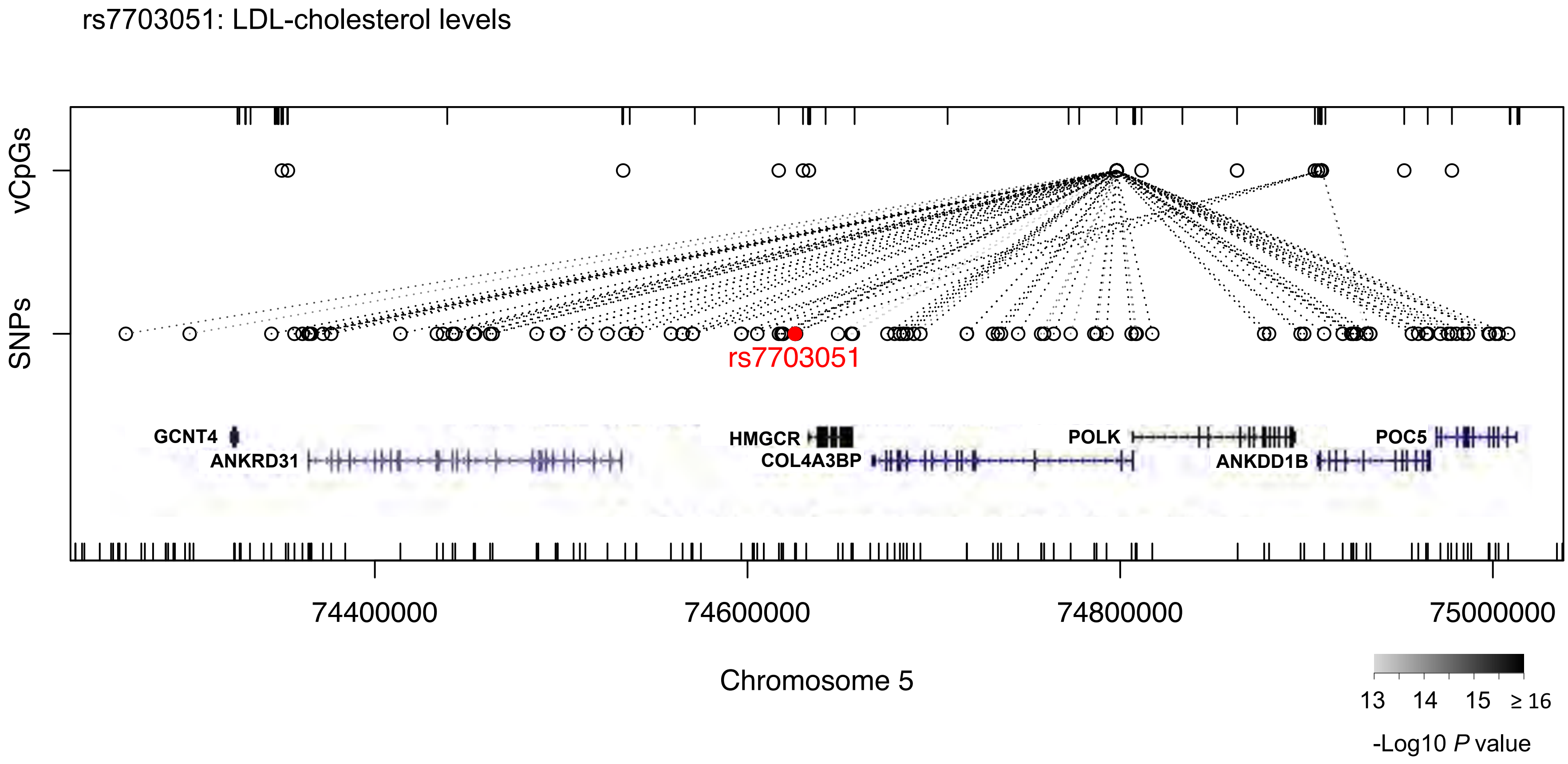


Figure S5. Some examples in which GWAS variants control DNA methylation at GeMes.

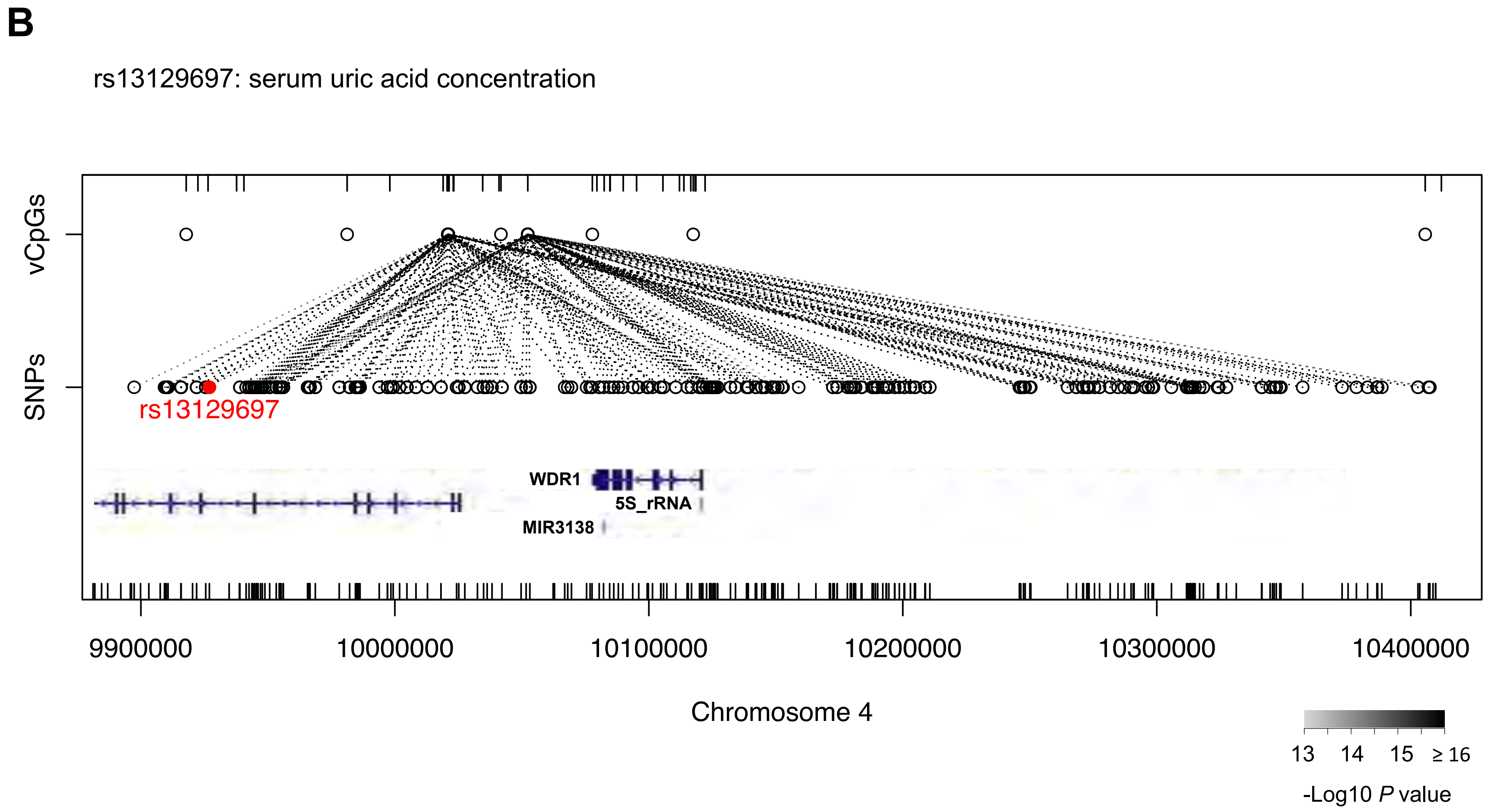


Figure S5. (continued).

C

rs780093: metabolic phenotypes

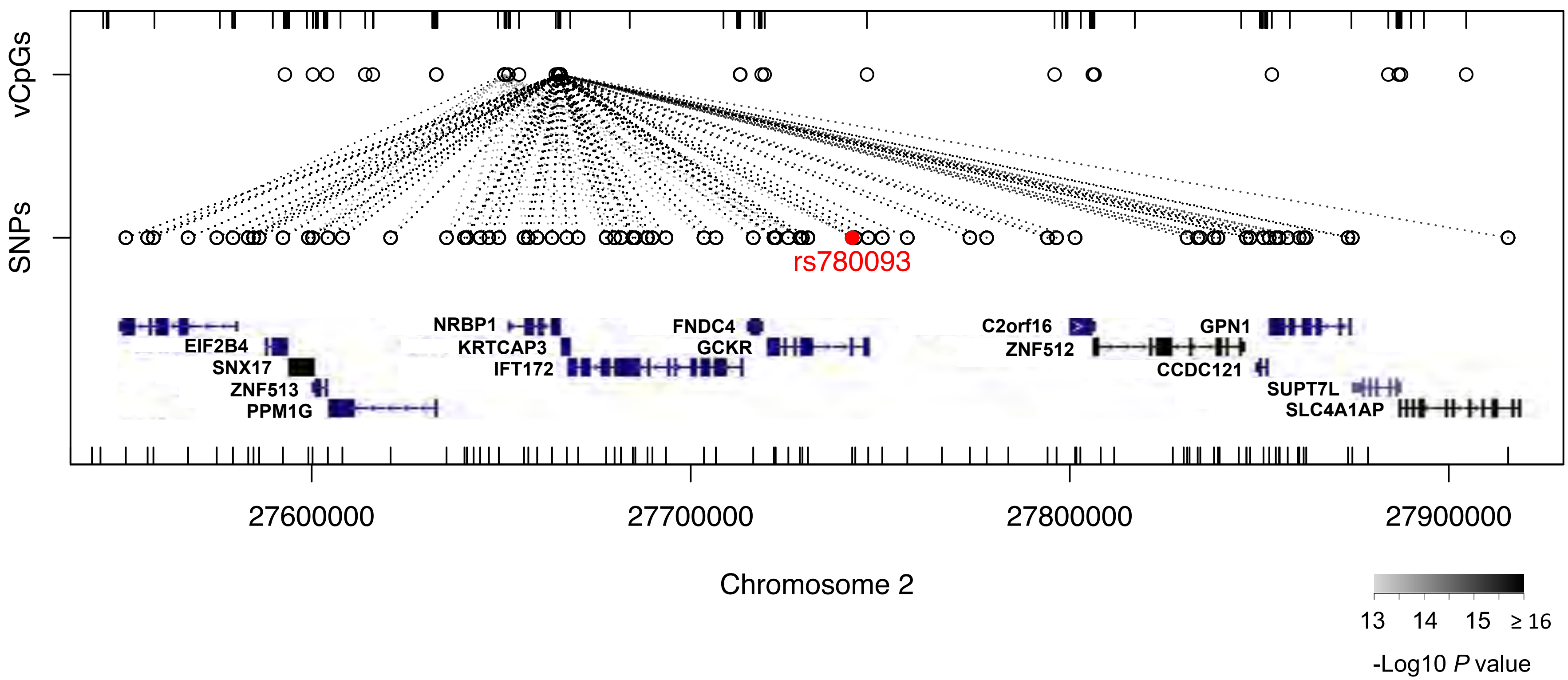


Figure S5. (continued).

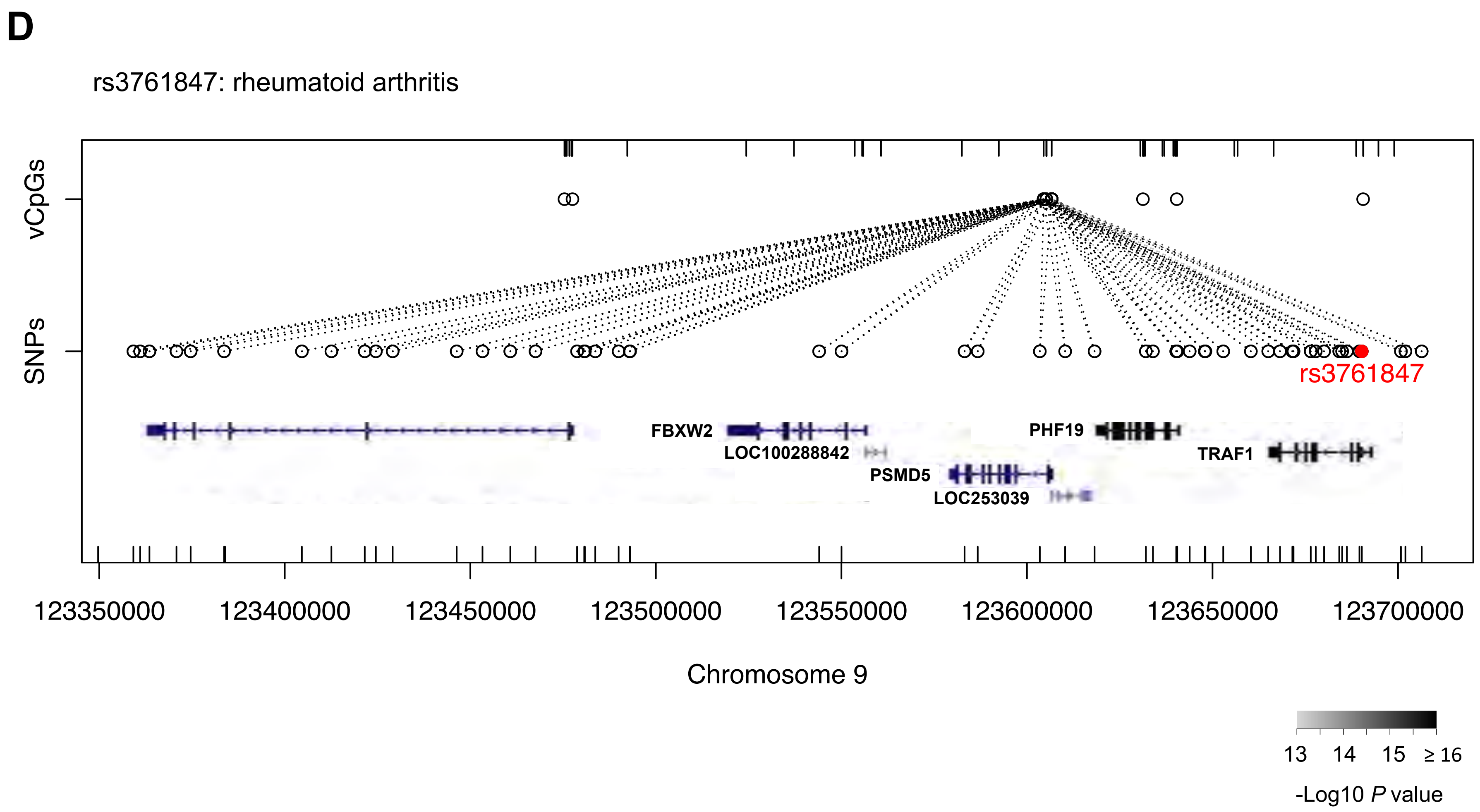


Figure S5. (continued).

E

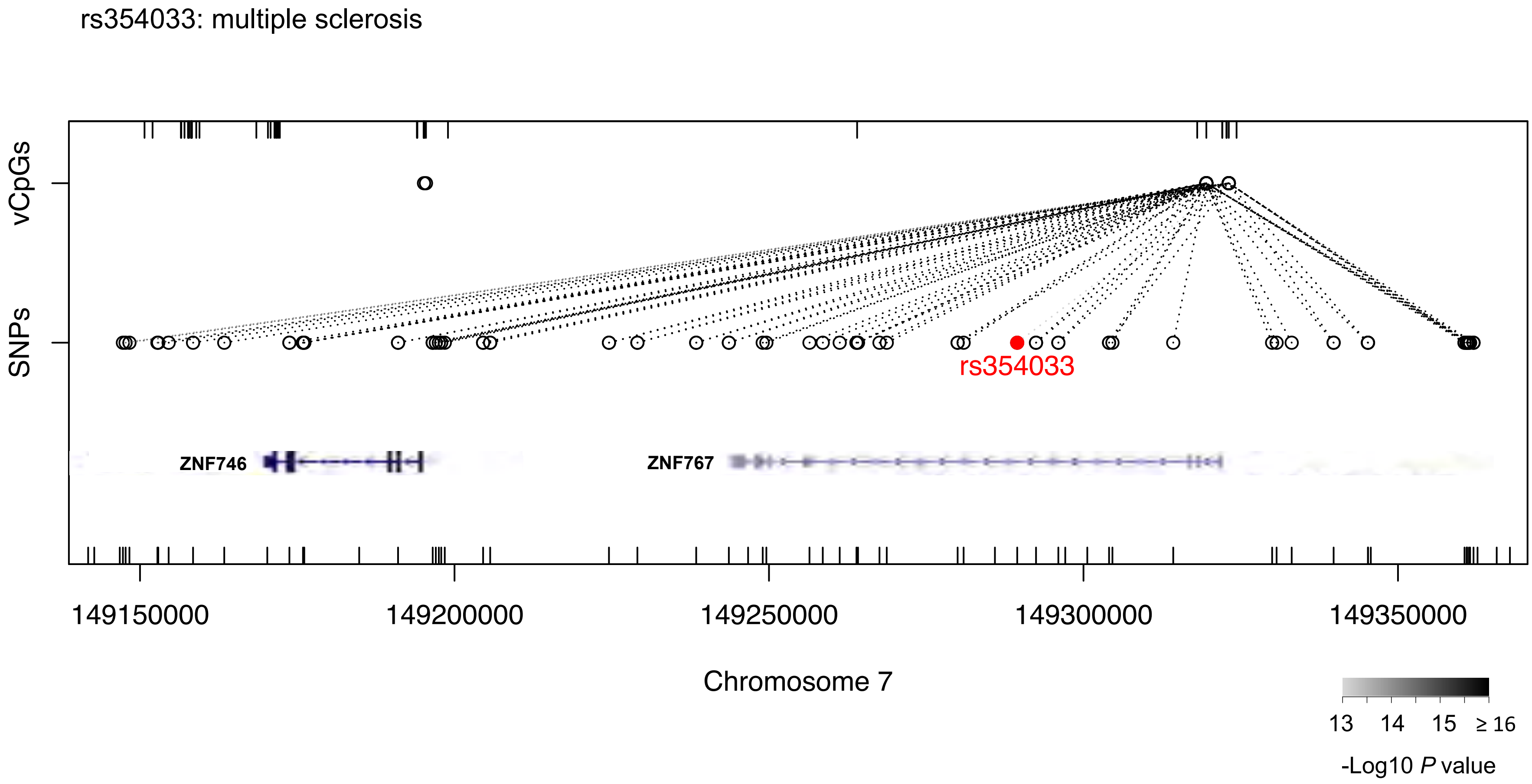


Figure S5. (continued).

F

rs401681: melanoma, lung cancer, pancreatic cancer, bladder cancer

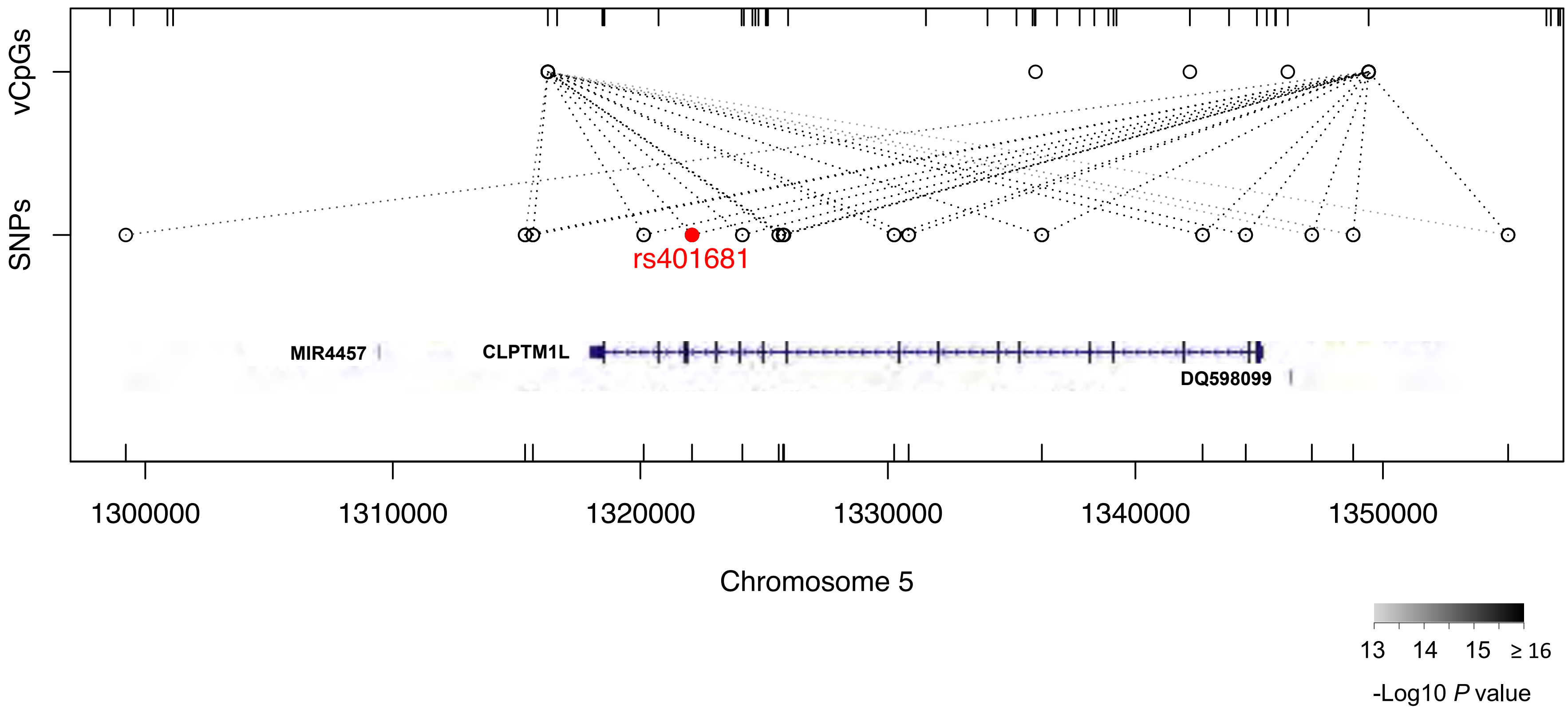


Figure S5. (continued).

G

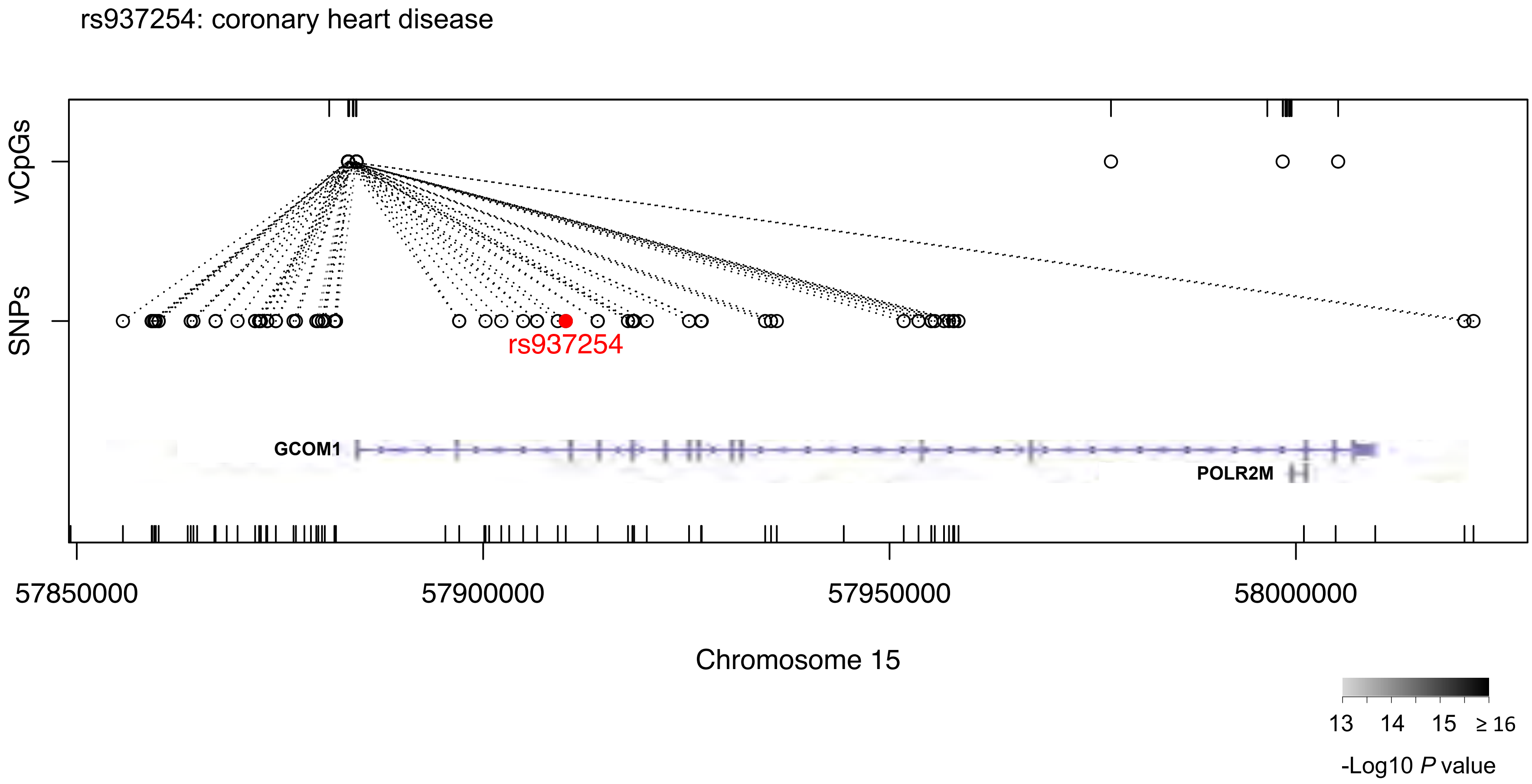


Figure S5. (continued).

H

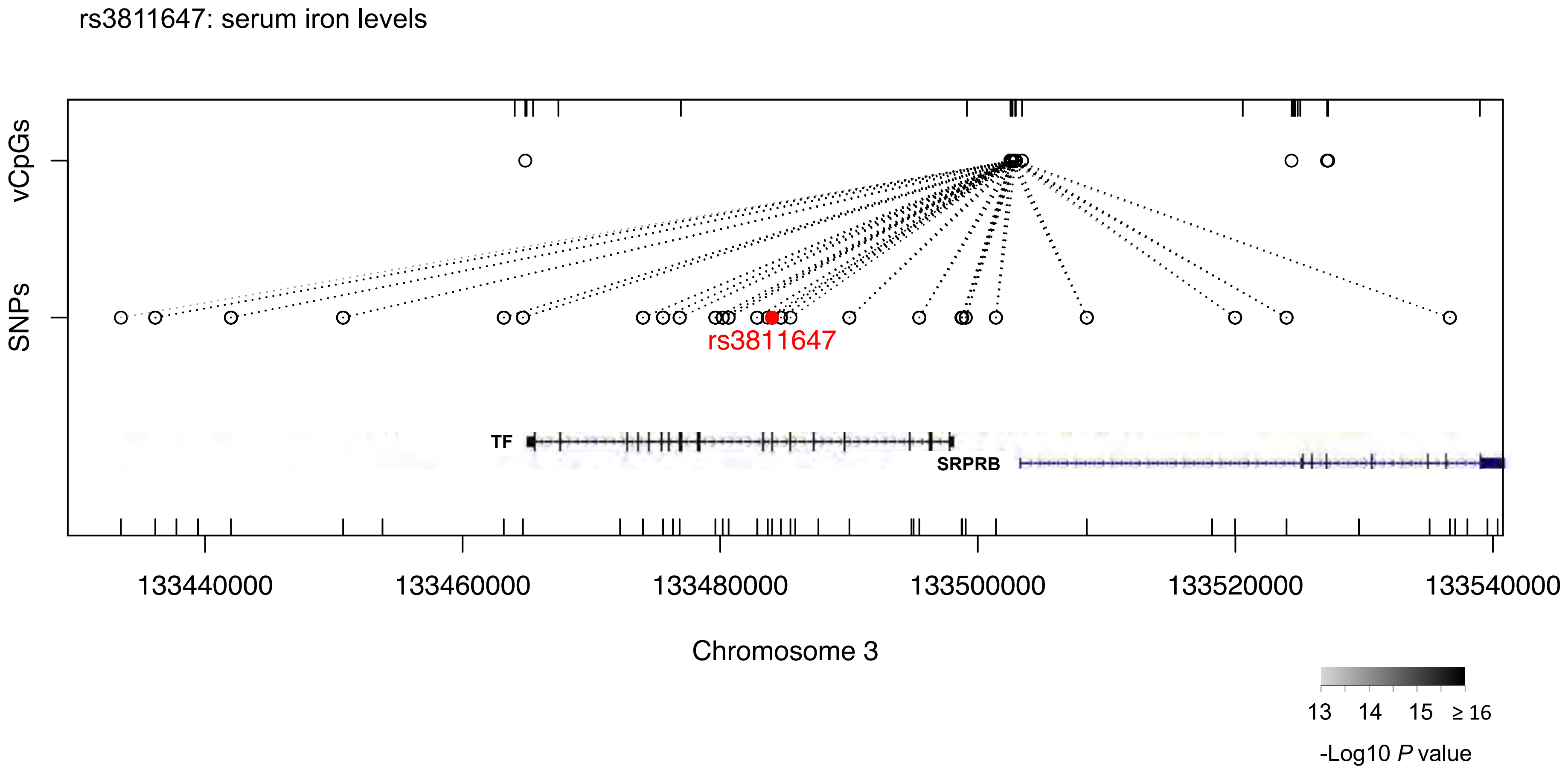


Figure S5. (continued).

I

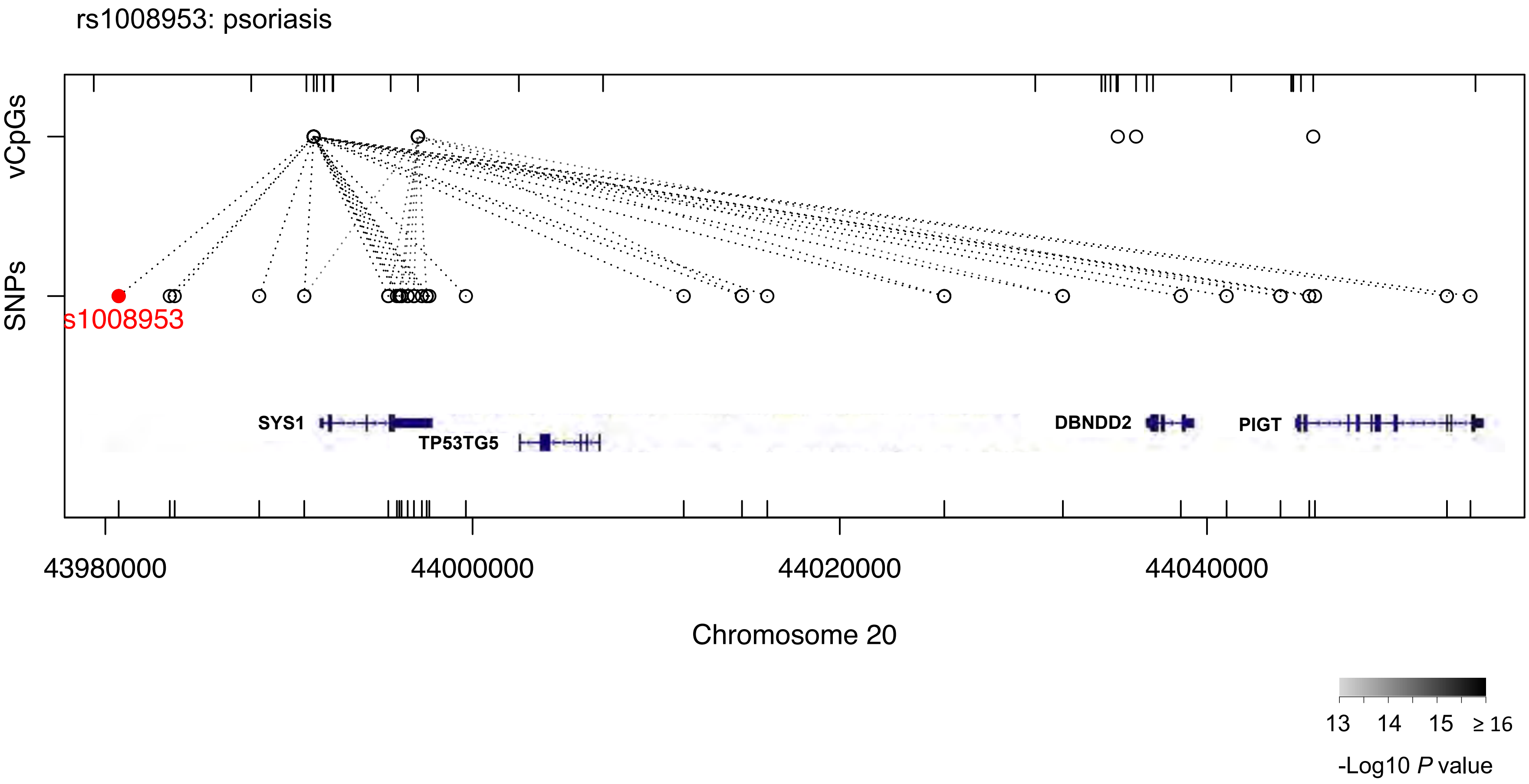


Figure S5. (continued).

J

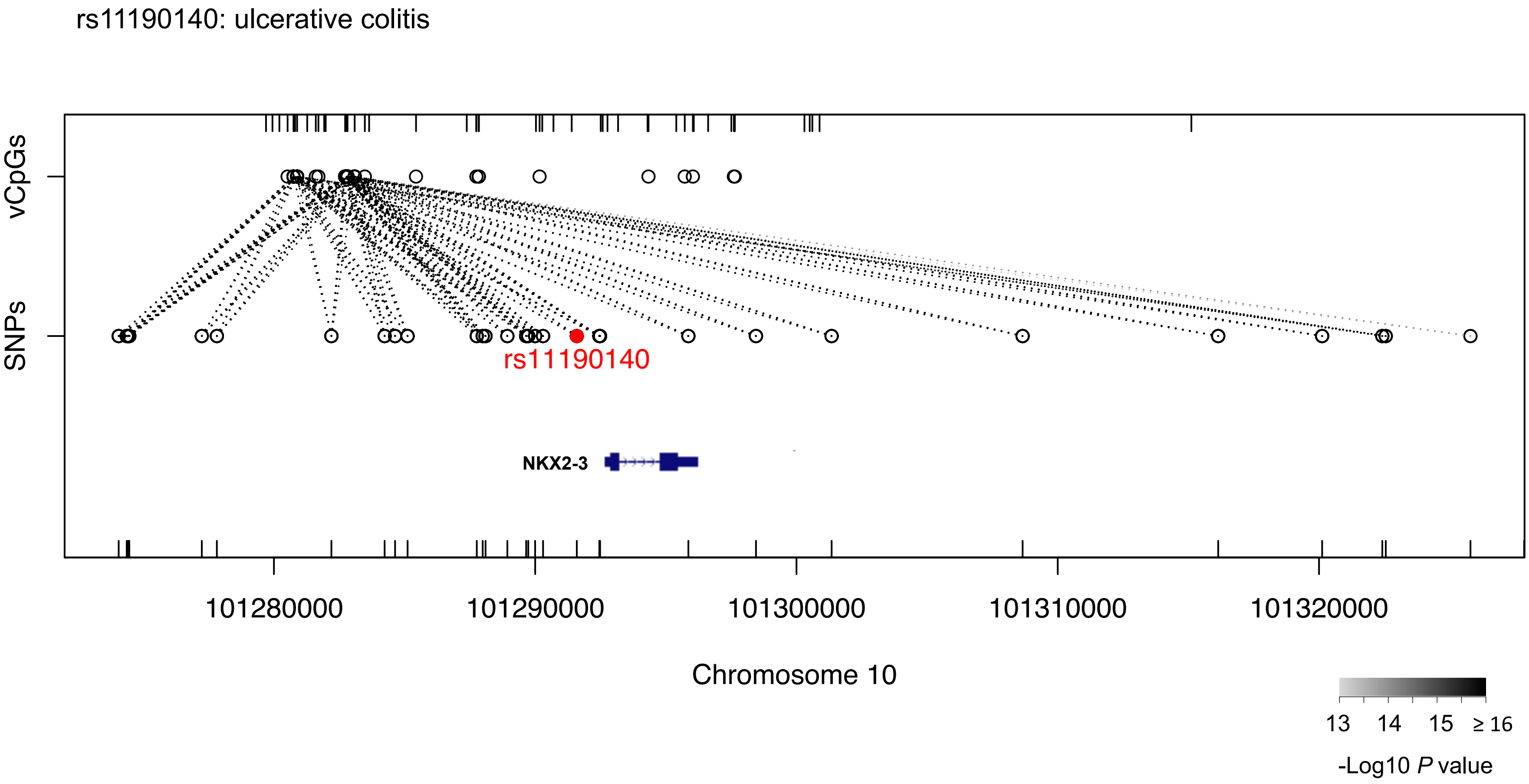
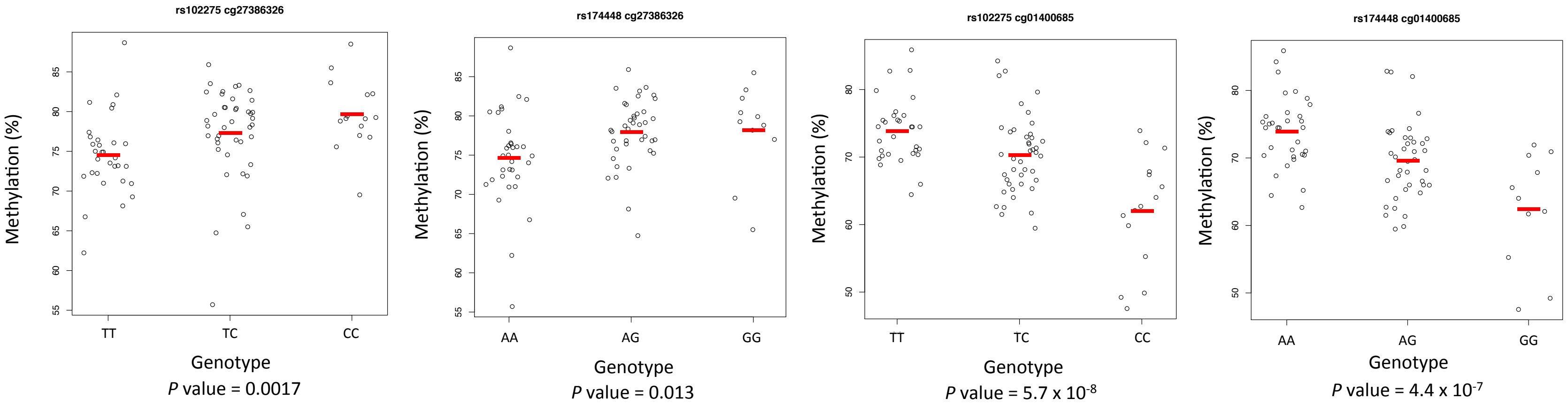


Figure S5. (continued).

A

FADS1/2



B

GSDMA/B

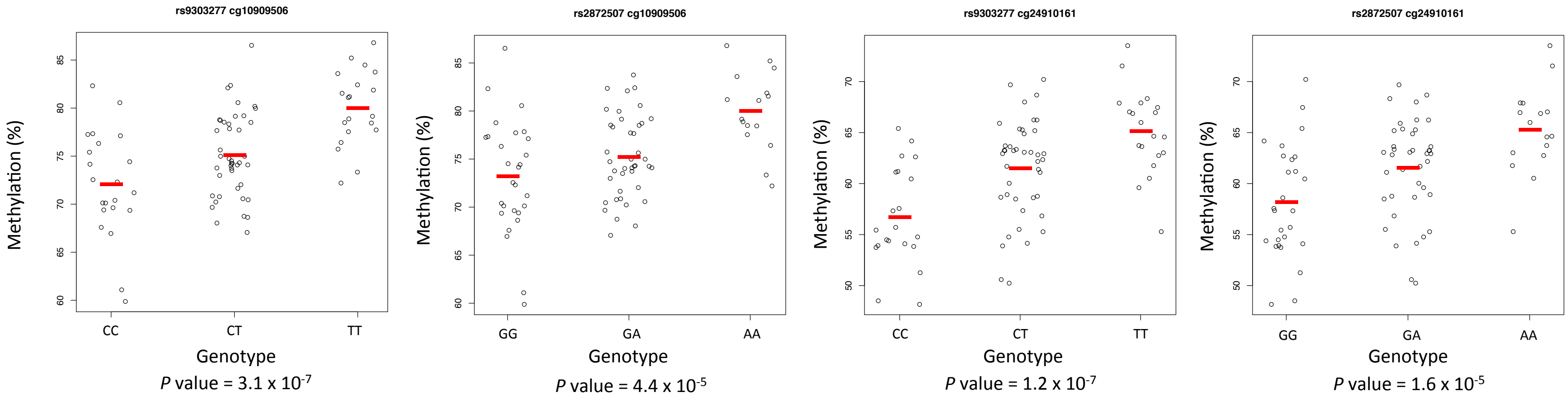


Figure S6. Replication of genotype and methylation associations at two GeMes shown in Figures 3B and 3C using bisulfite pyrosequencing from whole blood of an independent sample set of 90 healthy individuals.

Table S3. Primer sequences used for bisulfite pyrosequencing and location of CpGs interrogated. Chromosomal coordinates are based on the UCSC Genome Browser Human Feb. 2009 Assembly (NCBI37/hg19).

CpG	Chromosome	Position	Gene	Primer	Sequence (5' → 3')
cg27386326	11	61587980	*FADS1/2*	Forward	TTTAGGATTTAAATAGTTTATATGATATGA
				Reverse	ATTTTAATAAAAACAAAATTTCACC
				Nested forward	TATTTATGTTTTTTTGTTTTTTATG
				Nested reverse	/5Biosg/TAAAACTCCTAACCTCAAATAATCC
				Sequencing(F)	GGTTAGTGTTATGTATTTTGTT
cg01400685	11	61598025	*FADS1/2*	Forward	TTTGGTAATTGTTTTGATTTTTTTG
				Reverse	AATAAACCCAATAATAACATCTTTC
				Nested forward	TGTTTTGGATGTTTTTAGTGTTTATTT
				Nested reverse	/5Biosg/CAATTCTAAATCTATTTACAACCCC
				Sequencing(F)	TTTTTTTGTTTTATAAAAATGTAAT
cg10909506	17	38081995	*GSDMA/B*	Forward	GGGAATGTAGAAAAGAGTTTTAGATG
				Reverse	CTCAATTATCCAACCTAATAAACTA
				Nested forward	GGAGATTGAGTTTTTAGGGGTTAGT
				Nested reverse	/5Biosg/AAATCCCAAAAACATCTCTTCAA
				Sequencing(F)	GATTTTAGAGAAAATGTGATTTAT
cg24910161	17	38119198	*GSDMA/B*	Forward	GGTATTTGTTTGGTAGTTTTTTATTTAT
				Reverse	CTTAATCCACCCATCTTCCTATATC
				Nested forward	/5Biosg/TTTTAGTTTTTAGGAGGTTAGGGTATA
				Nested reverse	CATCCCTTACCAAAACAAACTACTT
				Sequencing(R)	CTACTTTATACAAAAACAAAAAC

/5Biosg/ = 5’ biotin added

F = forward   R = reverse

Table S4. Enrichment of DHSs within contiguous methylation clusters

	Contiguous methylation clusters (n >=3)	Contiguous methylation clusters (n >=5)	450K Array
DHSs	2,046	901	53,289
All	8,491	3,243	417,457
Enrichment	1.89	2.17	

Table S5. Smoking DMPs are enriched in contiguous methylation clusters

	Smoking DMPs	450K Array
Clusters from GSE42861	16	8,491
Clusters from GSE43414	17	10,099
Clusters from BLSA	14	9,518
Total on the 450K Array	26	417,457