Cell Genomics, Volume *2*

Supplemental information

Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits

Kisung Nam, Jangho Kim, and Seunggeun Lee

Supplemental Figures

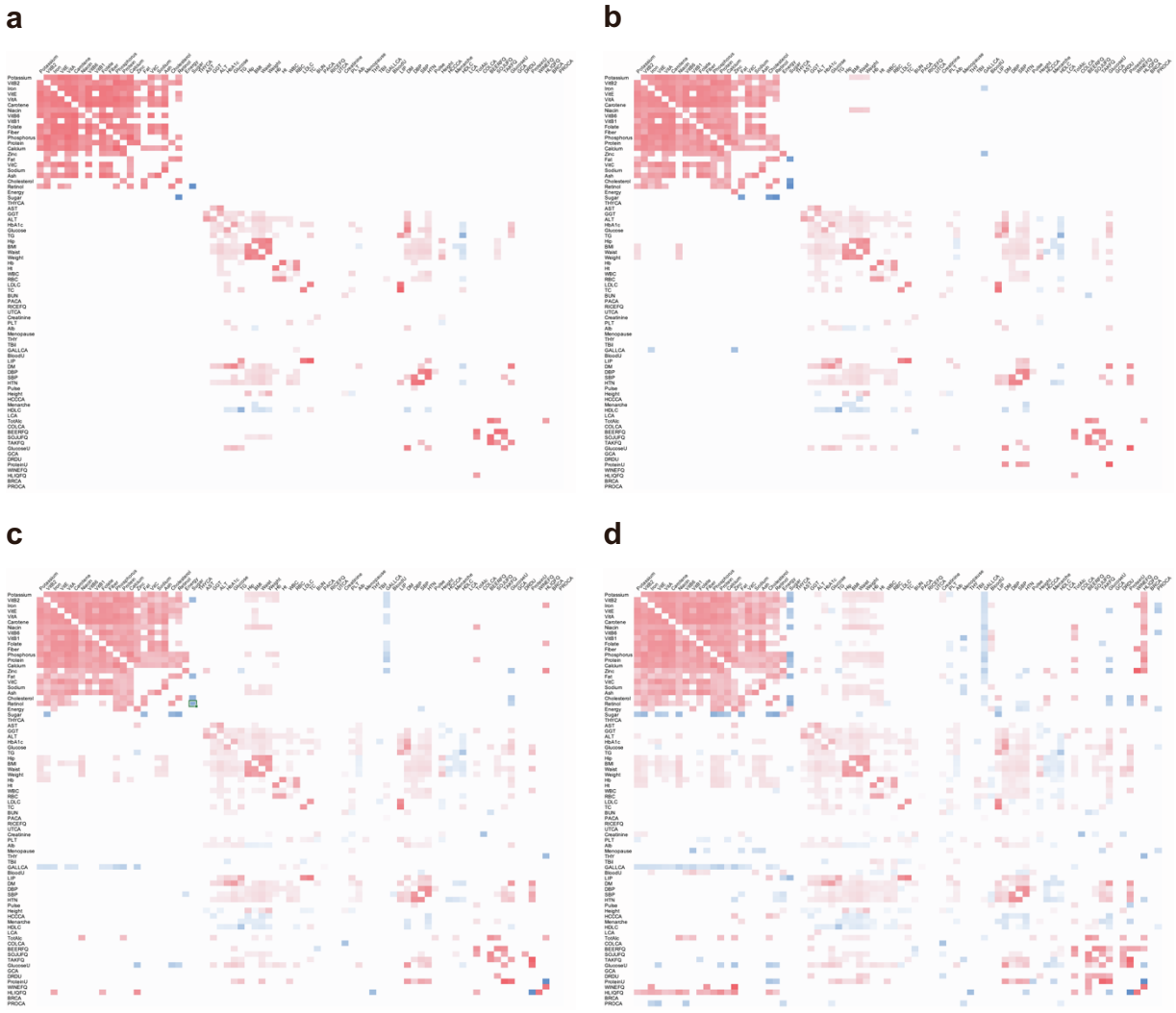
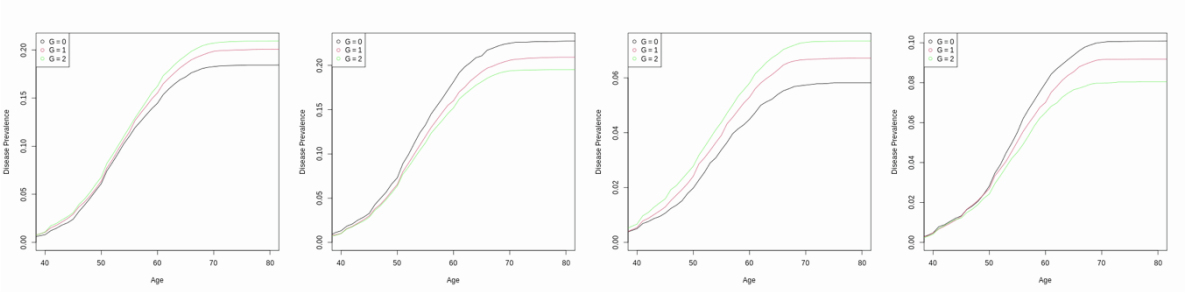


Figure S1: Heatmap for pairwise genetic correlations, Related to STAR Methods. To reduce false positives, genetic correlation was treated as zero when the corresponding p-value is greater than (a) 0.0001, (b) 0.001, (c) 0.01, (d) 0.05, respectively.

a



b

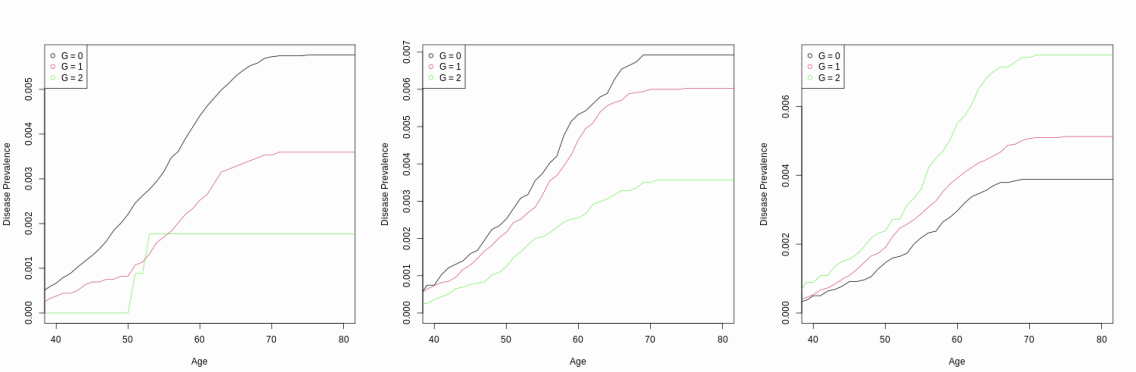


Figure S2: Incidence plot of diseases by genotype of significant variants, Related to STAR Methods. (a) Incidence plot of three disease prevalence by genotype of significant variant. First two plots show the disease prevalence of hypertension (rs71037444 and rs113628671, respectively), and next two plots are of diabetes (rs201174461) and hyperlipidemia (chr16:72,003,267), respectively. (b) Incidence plot of gastric cancer by genotype of most significant loci. Each plot demonstrates the disease prevalence according to the genotype of rs760077, rs2978977, rs866605438, respectively.

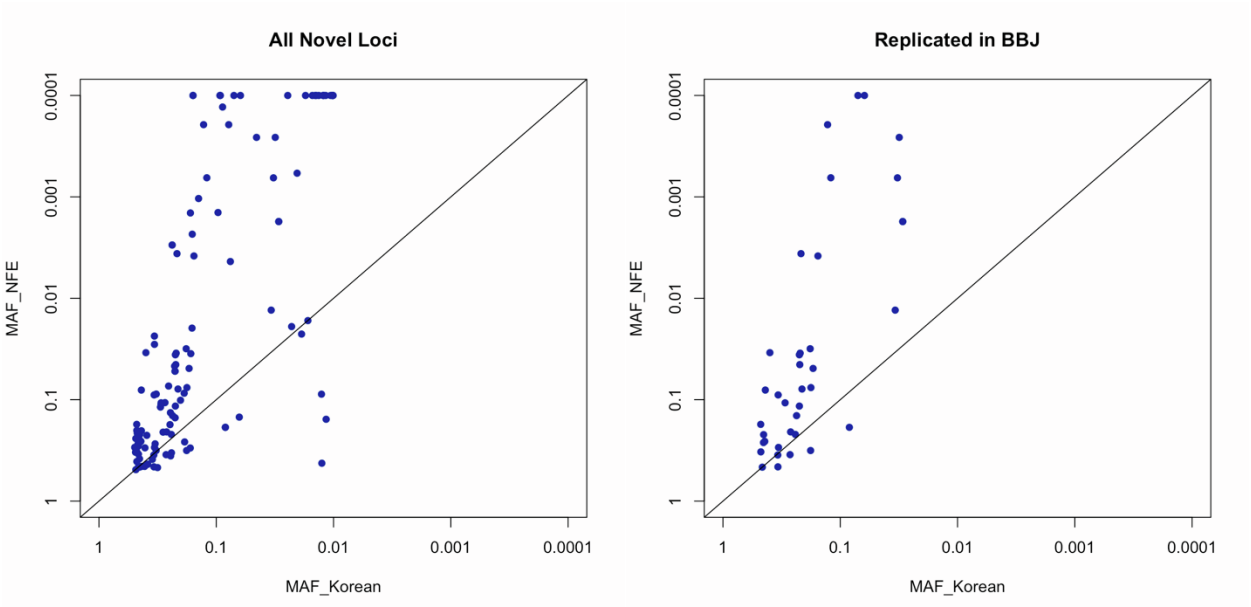
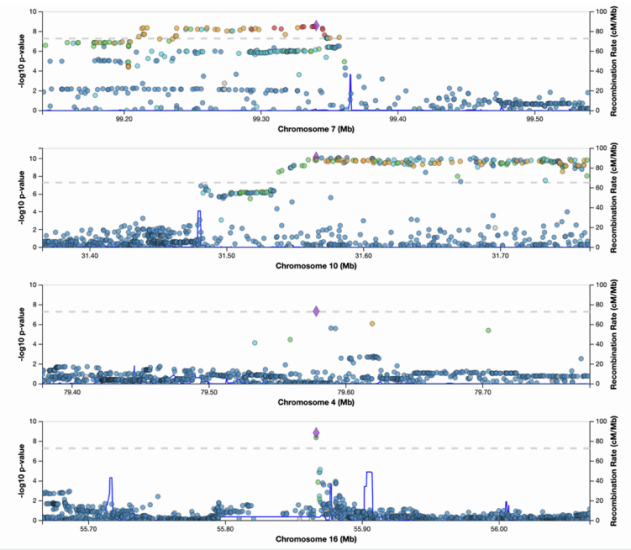


Figure S3: Comparison of minor allele frequencies (MAF) in Korean and non-Finnish European (NFE) for the top variants of 122 novel associations, Related to STAR Methods. MAF estimates in gnomAD were used for NFE. When the MAF is less than 0.0001 or unavailable for NFE, we regarded them as 0.0001 in the plot. 97 variants (79.5%) had lower MAF in NFE than in Korean among all novel loci. After filtering variants replicated in BBJ (p-value less than 0.05 in BBJ), 32 variants (76.3%) had lower MAF in NFE.

a



b

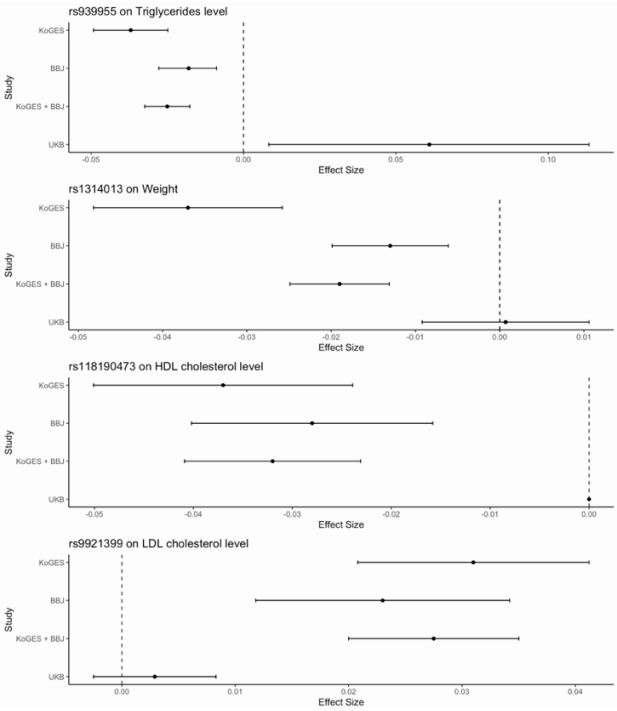


Figure S4: Locus zoom plots and forest plots for selected novel loci, Related to STAR Methods. (a) Locus zoom plots for selected novel loci. (b) Forest plots for effect sizes of top variant of selected novel loci. Each plot represents (1) rs939955 on triglycerides level, (2) rs1314013 on weight, (3) rs118190473 on HDL cholesterol level and (4) rs9921399 on LDL cholesterol level, respectively.

Supplemental Tables

Phenotype	Number of significant loci		Minimum p-value	
	SAIGE	SPACox	SAIGE	SPACox
Hypertension	29	40	3.60E-37	7.20E-40
Diabetes	11	14	9.70E-43	3.80E-46
Hyperlipidemia	11	12	1.00E-24	2.40E-25

Table S4: The number of significant loci and minimum p-value of plain GWAS (SAIGE) and survival analysis (SPACox) for 3 disease endpoints in KoGES data, Related to STAR Methods.

Phenotype	CHR	SNP	Average of adjusted phenotype		
			G = 0	G = 1	G = 2
Gastric cancer	1	rs760077	0.0383 (51,228)	0.0297 (14,552)	0.0299 (1,025)
Gastric cancer	8	rs2978977	0.0421 (9,930)	0.0377 (31,591)	0.0323 (25,284)
Gastric cancer	5	rs866605438	0.0325 (20,174)	0.0362 (33,029)	0.0422 (13,602)
Gallbladder cancer	4	rs75867949	0.008 (376)	0.0028 (9,616)	0.0014 (56,813)

Table S5: The average value of the TAPE-adjusted phenotypes by the genotype of the top SNP of newly found loci by TAPE, Related to STAR Methods. The numbers in parentheses are the sample size of the corresponding group.