American Journal of Human Genetics, Volume *93*

Supplemental Data

General Framework for Meta-analysis of Rare Variants in Sequencing Association Studies

Seunggeun Lee, Tanya M. Teslovich, Michael Boehnke, and Xihong Lin

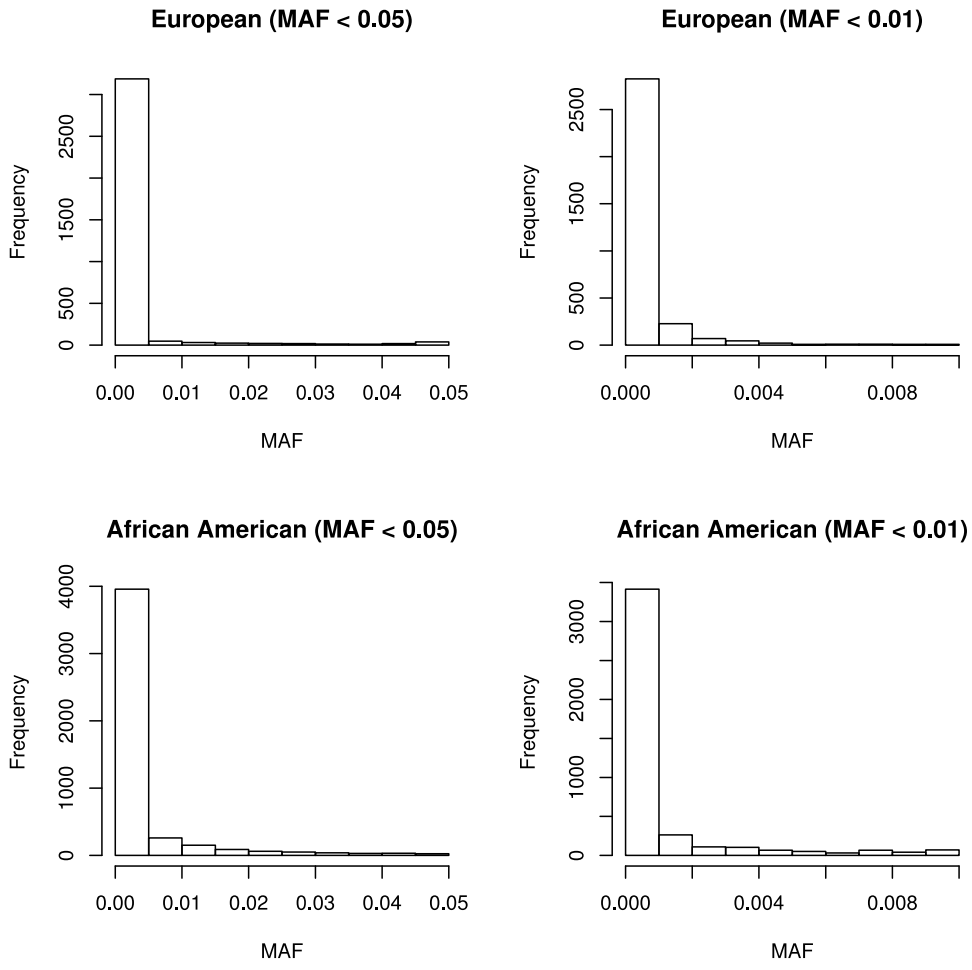

Figure S1: Distribution of allele frequencies for simulated data
MAFs for simulated Caucasian (top) and African American (bottom) data are based on the population allele frequencies under the coalescent model. The left panel considers MAF < 0.05, and the right panel zooms into a region MAF < 0.01.

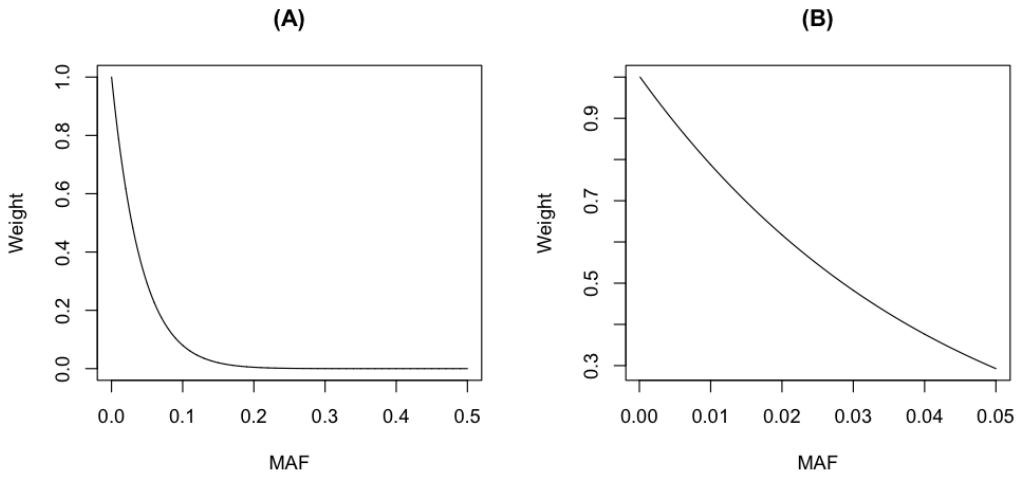

Figure S2: Default Beta(1,25) weight function
(A) and (B) represent the same functions, while (B) zooms into the region with MAF <0.05.

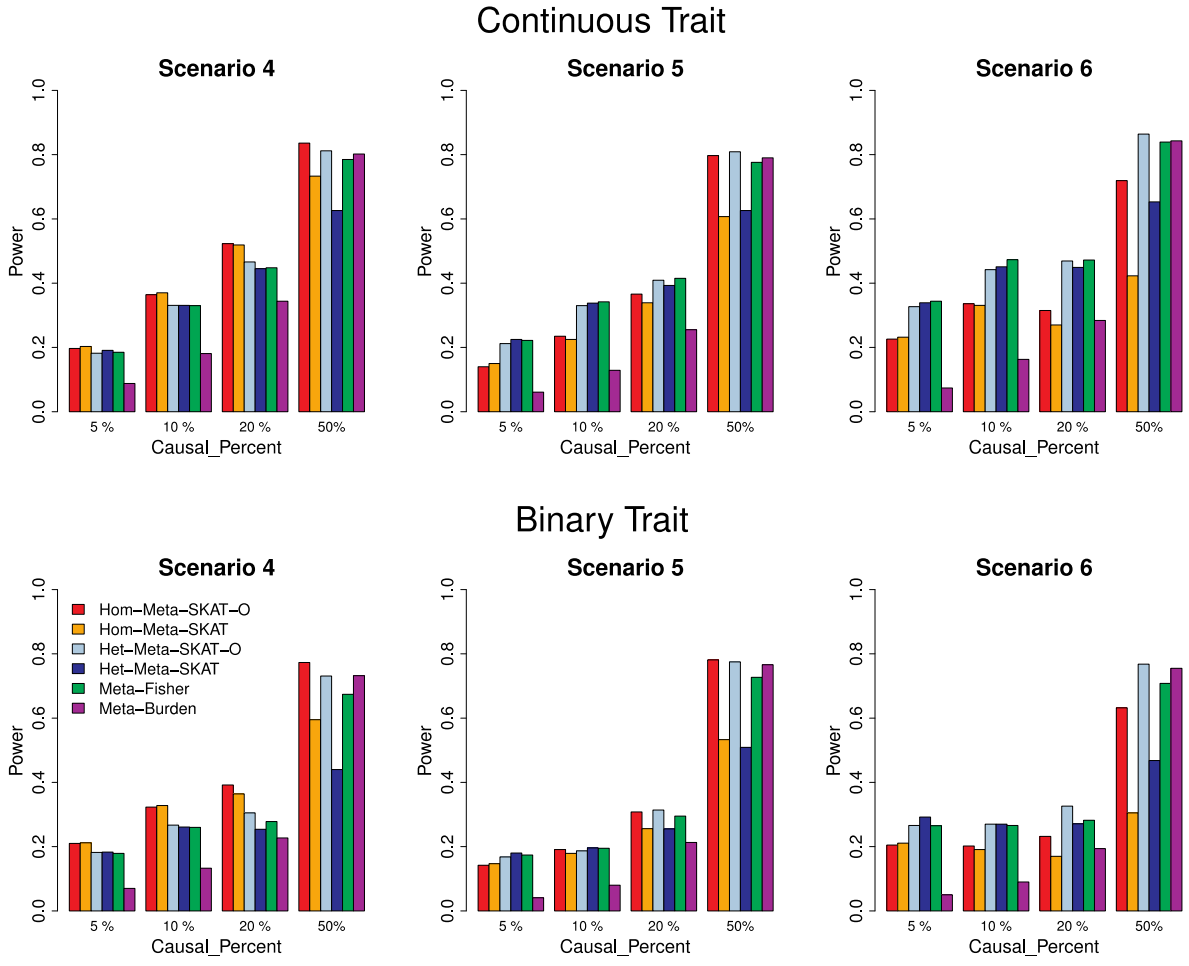

Figure S3: Power Comparisons of the Six Competing Methods with Equal Study Cohorts' Sizes when All Causal Variants Were Risk-Increasing
Empirical power at $\alpha = 2.5\times10^{-6}$ with equal study cohort sizes (Table 1) when all causal variants in a region were risk-increasing. All the other settings are the same as those in Figure 2 and 3 (main text).

.

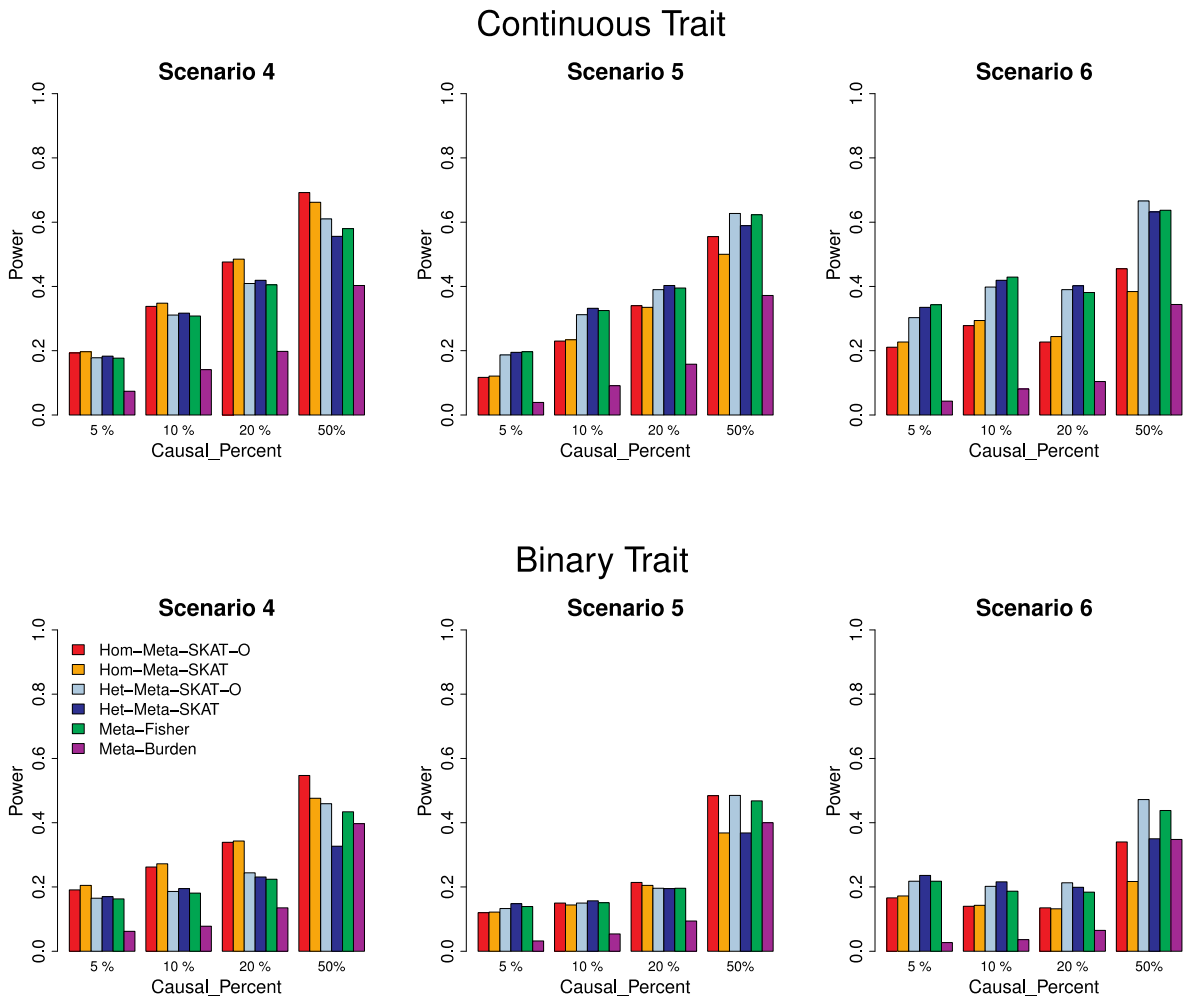

Figure S4: Power Comparisons of the Six Competing Methods with Equal Study Cohorts' Sizes when 20%/80% of Causal Variants Were Risk-Decreasing/Risk-Increasing

Empirical power at $\alpha = 2.5\times10^{-6}$ with equal study cohort sizes (Table 1) assuming 20% of the causal variants were risk-decreasing and 80% of the causal variants were risk-increasing. All the other settings are the same as those in Figure 2 and 3 (main text).

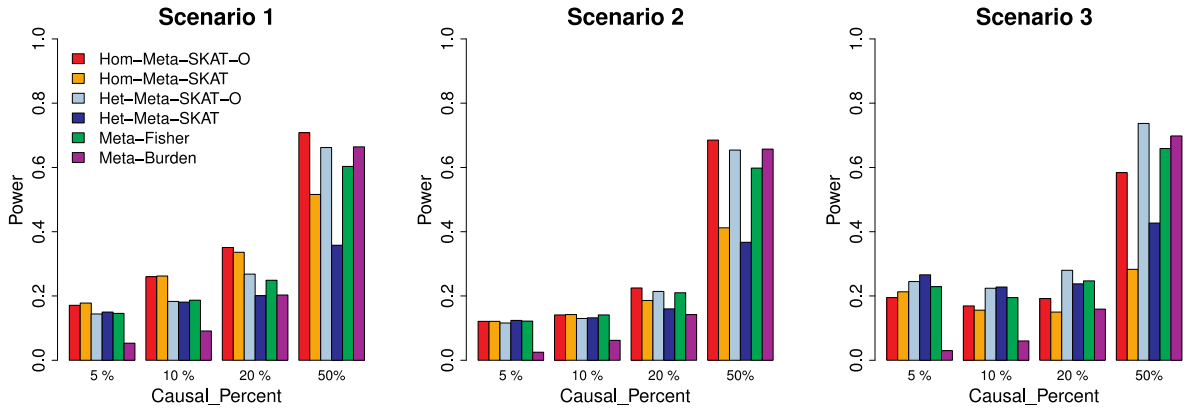

Figure S5: Power Comparisons of the Six Competing Methods with High Disease Prevalence when All Causal Variants Were Risk-Increasing

Empirical power from the binary trait simulations when the disease prevalence was 0.1, and all causal variants in a region were risk-increasing. The power was obtained at $\alpha = 2.5\times10^{-6}$ with different study cohort sizes (Table 1). All the other settings are the same as those in Figure 2 and 3 (main text).

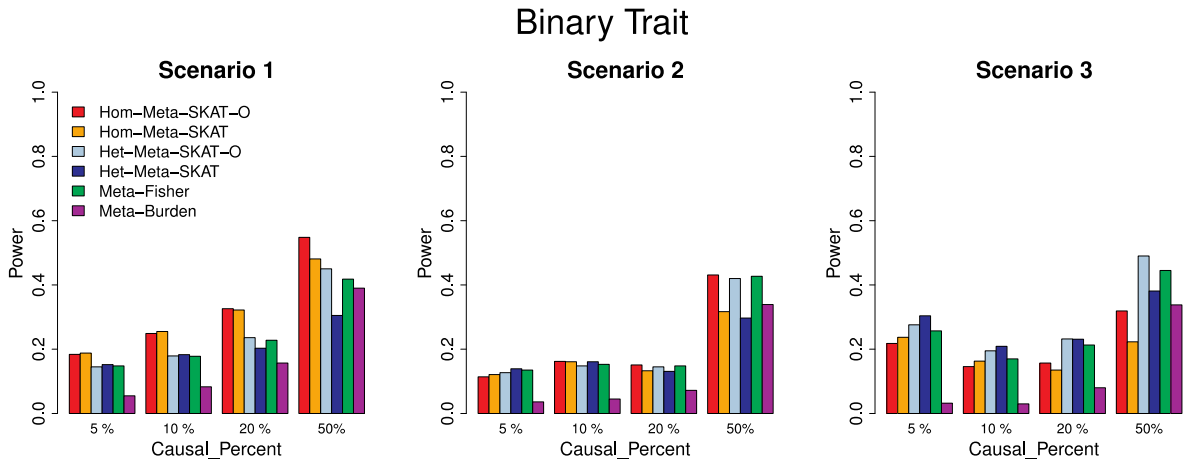

Figure S6: Power Comparisons of the Six Competing Methods with High Disease Prevalence when 20%/80% of Causal Variants Were Risk-Decreasing/Risk-Increasing

Empirical power from the binary trait simulations when the disease prevalence was 0.1. 20% of the causal variants were risk-decreasing, and 80% of the causal variants were risk-increasing. The power was obtained at $\alpha = 2.5\times10^{-6}$ with different study cohort sizes (Table 1). All the other settings are the same as those in Figure 2 and 3 (main text).

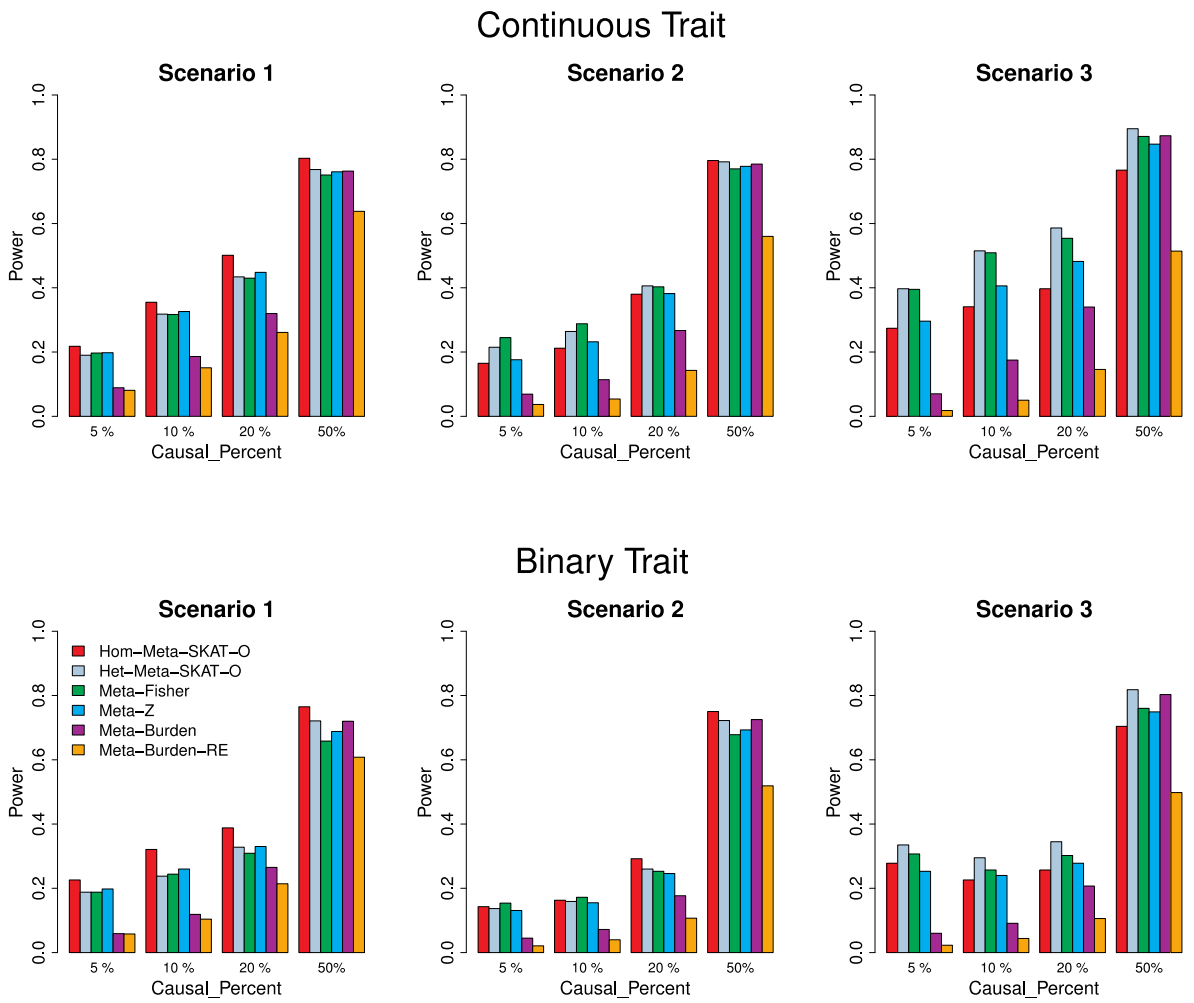

Figure S7: Power Comparisons of Meta-Z, Meta-Burden-RE and the Other Competing Methods with Equal Study Cohorts' Sizes when All Causal Variants Were Risk-Increasing

Empirical power at $\alpha = 2.5\times10^{-6}$ with equal study cohort sizes (Table 1) when all causal variants in a region were risk-increasing. All the other settings are the same as those in Figure 2 and 3 (main text). The power estimates of all methods except Meta-Z and Meta-Burden-RE are the exactly same as the estimates in Figure 2.

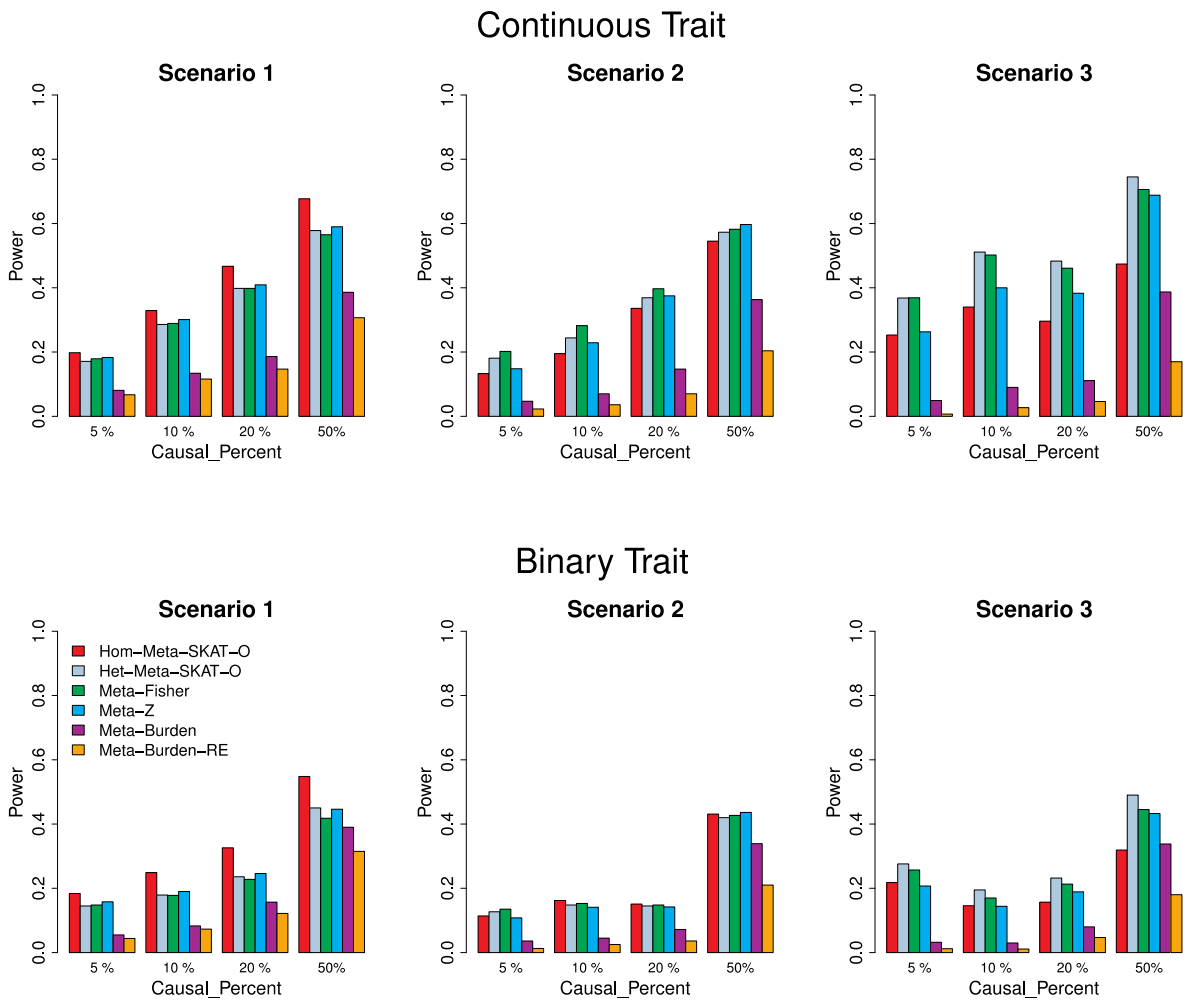

Figure S8: Power Comparisons of Meta-Z, Meta-Burden-RE and the Other Competing Methods with Equal Study Cohorts' Sizes when 20%/80% of Causal Variants Were Risk-Decreasing/Risk-Increasing
Empirical power at $\alpha = 2.5\times10^{-6}$ with equal study cohort sizes (Table 1) assuming 20% of the causal variants were risk-decreasing and 80% of the causal variants were risk-increasing. All the other settings are the same as those in Figure 2 and 3 (main text). The power estimates of all methods except Meta-Z and Meta-Burden-RE are the exactly same as the estimates in Figure 3.

Table S1: Observed number of variants and sum of risk & protective allele frequencies in the power simulations.

Each entry represents an average number of observed variants and allele frequencies over 1000 simulated datasets.

Scenario	% of Causal Variants	% of Protective Variants	Number of Observed Variants	Sum of Risk allele Freq.	Sum of Protective allele Freq.
Continuous Traits					
Scenario 1	5 %	0 %	42.49	0.0039	0
		20 %	42.2	0.0025	0.0008
	10 %	0 %	42.75	0.0079	0
		20 %	42.35	0.006	0.0014
	20 %	0 %	42	0.0177	0
		20 %	42.27	0.0137	0.0035
	50 %	0 %	42.34	0.0436	0
		20 %	42.58	0.0352	0.0089
Scenario 2	5 %	0 %	42.4	0.0036	0
		20 %	42.26	0.0027	0.0005
	10 %	0 %	42.53	0.0064	0
		20 %	42.51	0.0051	0.0016
	20 %	0 %	42.52	0.0162	0
		20 %	42.6	0.0131	0.0037
	50 %	0 %	42.26	0.0427	0
		20 %	42.61	0.0343	0.008
Scenario 3	5 %	0 %	44.58	0.0052	0
		20 %	44.53	0.0041	0.001
	10 %	0 %	44.41	0.0094	0
		20 %	44.29	0.0076	0.0018
	20 %	0 %	44.48	0.0186	0
		20 %	43.97	0.0143	0.0036
	50 %	0 %	44.46	0.0444	0
Binary Traits					
Scenario 1	5 %	0 %	43.06	0.0082	0
		20 %	42.83	0.0069	0.0005
	10 %	0 %	43.12	0.0151	0
		20 %	43.09	0.0118	0.0009
	20 %	0 %	43.63	0.0284	0
		20 %	43.65	0.0223	0.0024
	50 %	0 %	45.85	0.0696	0
		20 %	44.38	0.053	0.0071
Scenario 2	5 %	0 %	42.56	0.0076	0
		20 %	42.8	0.0062	0.0004
	10 %	0 %	43.24	0.0132	0
		20 %	42.63	0.0109	0.0009
	20 %	0 %	43.75	0.0262	0
		20 %	43.39	0.0198	0.0025
	50 %	0 %	45.63	0.0673	0
		20 %	44.66	0.0509	0.0072
Scenario 3	5 %	0 %	45.06	0.0127	0
		20 %	45.13	0.0101	0.0006
	10 %	0 %	45.74	0.0177	0
		20 %	45.06	0.0139	0.0012
	20 %	0 %	46.26	0.0297	0
		20 %	45.61	0.0232	0.0027
	50 %	0 %	47.9	0.0642	0
		20 %	46.84	0.0509	0.0072

Table S2. Annotation information and single variant meta-analysis results of the seven SNVs in *LPL*

				HDL		LDL		TG	
Chr	Position	Function	MAF	Fisher	IVW	Fisher	IVW	Fisher	IVW
8	19840949	5UTR	7.3×10^{-3}	5.6×10^{-1}	5.8×10^{-1}	5.1×10^{-2}	7.5×10^{-1}	3.2×10^{-1}	5.3×10^{-1}
8	19840951	5UTR	6.5×10^{-3}	4.1×10^{-2}	1.3×10^{-2}	9.9×10^{-1}	5.6×10^{-1}	1.3×10^{-4}	6.3×10^{-3}
8	19849988	Coding (non-synonymous)	6.1×10^{-3}	5.0×10^{-2}	2.1×10^{-2}	9.8×10^{-1}	8.8×10^{-1}	2.6×10^{-4}	5.2×10^{-3}
8	19853715	Coding (synonymous)	2.6×10^{-2}	7.2×10^{-1}	2.0×10^{-1}	5.6×10^{-1}	7.9×10^{-1}	5.7×10^{-1}	1.7×10^{-1}
8	19867130	3UTR	3.9×10^{-4}	3.2×10^{-1}	1.7×10^{-1}	4.1×10^{-1}	1.6×10^{-1}	1.5×10^{-1}	7.6×10^{-1}
8	19867472	3UTR	1.9×10^{-2}	5.5×10^{-3}	4.7×10^{-2}	1.4×10^{-1}	3.3×10^{-1}	3.2×10^{-3}	1.6×10^{-2}
8	19868517	3UTR	1.8×10^{-4}	3.9×10^{-1}	4.1×10^{-1}	6.1×10^{-2}	1.3×10^{-1}	7.6×10^{-1}	6.8×10^{-1}
Min-P				5.5×10^{-3}	1.3×10^{-2}	5.1×10^{-2}	1.3×10^{-1}	1.3×10^{-4}	5.2×10^{-3}
Min-P*				3.8×10^{-2}	8.8×10^{-2}	3.6×10^{-1}	> 0.5	9.0×10^{-3}	3.6×10^{-2}

SNV annotations were conducted using hg18 (NCBI build 36) assembly. “Fisher” represents p-values from the Fisher's inverse chi-square method and “IVW” represents p-values from the inverse variance weighting method. “Min-P” represents the minimum p-value over the seven SNV p-values, and “Min-P*” represents the multiple test adjusted minimum p-value