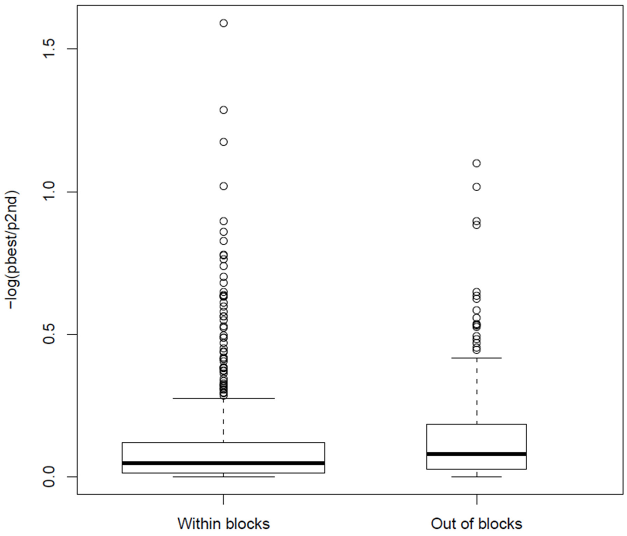
SUPPLEMENTARY INFORMATION

Performance Comparison of Two Gene Set Analysis Methods for Genome-wide Association Study Results: GSA-SNP vs i-GSEA4GWAS

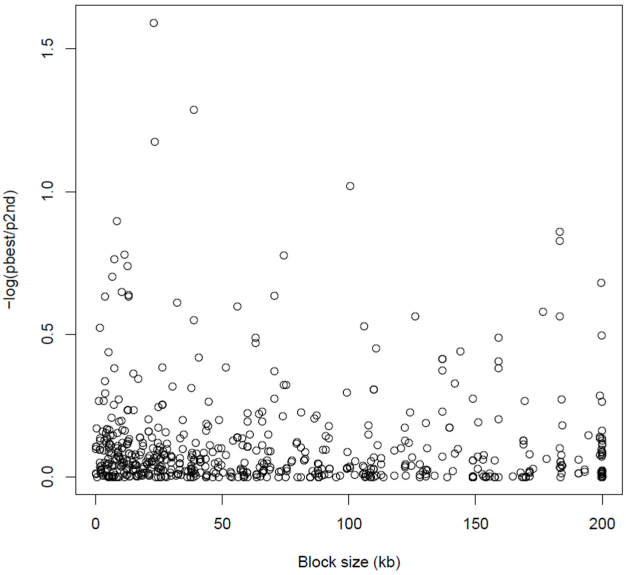
Ji-sun Kwon[1], Jihye Kim[1], Dougu Nam[2], Sangsoo Kim[1]*

[1]Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Korea,
[2]School of Nano-Bioscience and Chemical Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Korea

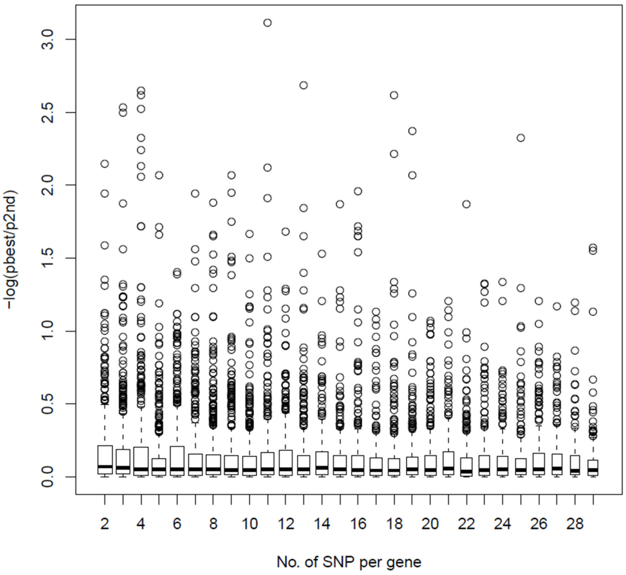
http://dx.doi.org/10.5808/GI.2012.10.2.123



Supplementary Fig. 1. Distributions of the p-value difference. The ratios of the best and the second-best p-values were -log-transformed, and their distribution was represented by a box plot. The left plot includes 586 genes on chromosome 5 for which the best single nucleotide polymorphism (SNP) was located within a haplotype block, while the right plot represents 146 genes on chromosome 5 for which the best SNP could not be located within a haplotype block. The difference between these two distributions was significant (p= 0.01329). Haplotype blocks were calculated with PLINK.



Supplementary Fig. 2. Relationship between haplotype block size and the p-value difference. The ratios of the best and the second-best p-values were -log-transformed and plotted against the size of the haplotype block for which the best single nucleotide polymorphism was located; 586 genes on chromosome 5 were plotted. Haplotype blocks were calculated with PLINK.



Supplementary Fig. 3. Distribution of the p-value difference by the number of single nucleotide polymorphisms (SNPs) per gene. The ratios of the best and the second-best p-values were -log-transformed, and their distribution was represented by a box plot. For a gene, the number of SNPs located inside or within 20 kb of the gene boundary was counted.