The American Journal of Human Genetics, *91*

Supplemental Data

Prioritizing Genetic Variants for Causality on the Basis of Preferential Linkage Disequilibrium

Qianqian Zhu, Dongliang Ge, Erin L. Heinzen, Samuel P. Dickson, Thomas J. Urban, Mingfu Zhu, Jessica M. Maia, Min He, Qian Zhao, Kevin V. Shianna, and David B. Goldstein

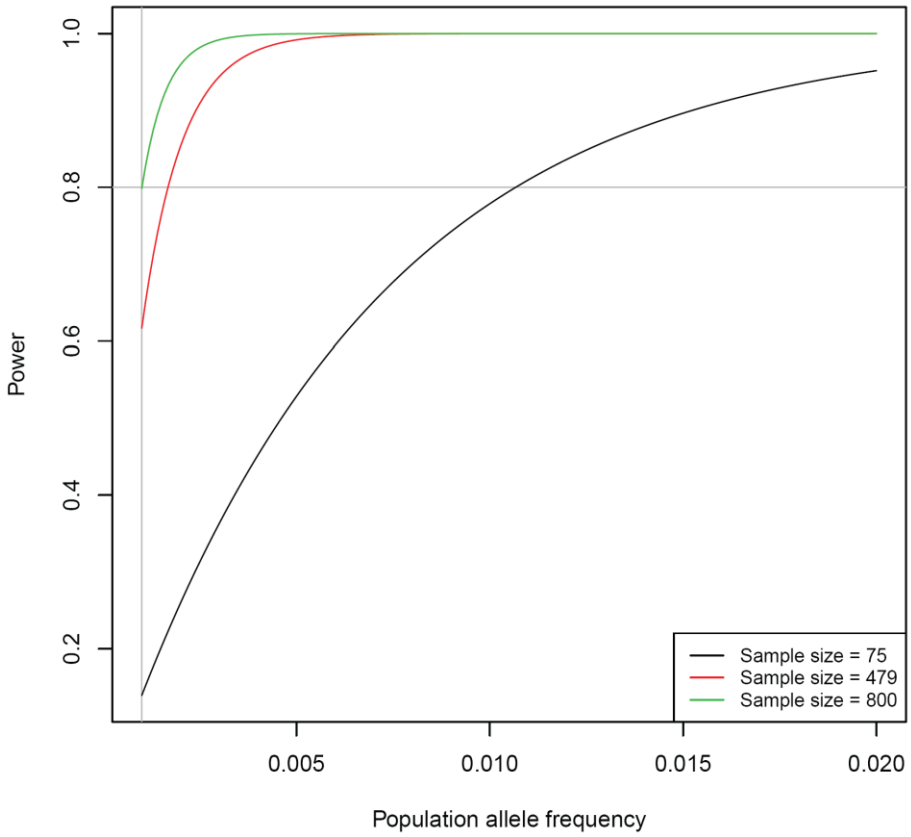

Figure S1. The Power of Detecting the Alleles of Variants in Our Samples

The grey vertical line corresponds to allele frequency 0.1%.

Table S1. The Ranking of the Causal Variants When Changing the MAF Requirements of the Candidate SNVs at Step 1*

Phenotype	Discovery SNP	Causal SNV	Rank of the Causal SNV among Final Candidates		
			MAF of Candidate SNVs ≤ min(15%, MAF of the Discovery SNP)	MAF of Candidate SNVs ≤ min(50%, MAF of the Discovery SNP)	MAF of Candidate SNVs ≤ 50%
Crohn disease	rs17221417	rs2066844	1	1	1
		rs2066845	2	2	2
Crohn disease	rs2076756	rs2066845	1	1	1
		rs2066844	2	2	2
Ribavirin induced haemolytic anaemia	rs6051702	rs1127354	1	2	2
		rs7270101	2	3	3
Therapeutic warfarin dose	rs4917639	rs1799853	2	2	2
Bladder cancer	rs11892031	rs17863783	23	23	30
Hearing Loss	rs870729	rs80338945	5	5	7
Hearing Loss	rs7329467	rs35887622	6	6	6

*: all other parameters were set as default.

Table S2. The Ranking of the Causal Variants When Changing the Weight for PhastCons Score at Step 4*

Phenotype	Discovery SNP	Causal SNV	Rank of the Causal SNV among Final Candidates					
			$w = 0$	$w = 0.1$	$w = 0.2$	$w = 0.3$	$w = 0.4$	$w = 0.5$
Crohn disease	rs17221417	rs2066844	34	1	1	1	1	2
		rs2066845	38	3	2	2	2	1
Crohn disease	rs2076756	rs2066845	27	2	2	1	1	1
		rs2066844	19	1	1	2	2	2
Ribavirin induced haemolytic anaemia	rs6051702	rs1127354	26	1	1	1	1	1
		rs7270101	19	2	2	2	3	3
Therapeutic warfarin dose	rs4917639	rs1799853	5	1	2	2	2	2
Bladder cancer	rs11892031	rs17863783	20	20	21	23	24	27
Hearing Loss	rs870729	rs80338945	150	99	84	5	4	2
Hearing Loss	rs7329467	rs35887622	12	2	5	6	7	7

*: all other parameters were set as default.

Table S3. The Ranking of the Causal Variants When Using PhyloP Score at Step 4*

Phenotype	Discovery SNP	Causal SNV	Rank of the Causal SNV among Final Candidates					
			$w = 0$	$w = 0.1$	$w = 0.2$	$w = 0.3$	$w = 0.4$	$w = 0.5$
Crohn disease	rs17221417	rs2066844	34	10	9	8	6	6
		rs2066845	38	12	11	10	8	5
Crohn disease	rs2076756	rs2066845	27	8	5	1	1	1
		rs2066844	19	2	2	3	3	3
Ribavirin induced haemolytic anaemia	rs6051702	rs1127354	26	7	5	5	4	3
		rs7270101	19	5	4	4	3	4
Therapeutic warfarin dose	rs4917639	rs1799853	5	4	5	5	5	6
Bladder cancer	rs11892031	rs17863783	20	1	1	1	1	1
Hearing Loss	rs870729	rs80338945	150	128	113	53	46	43
Hearing Loss	rs7329467	rs35887622	12	1	1	1	1	1

*: the ranking score for the i th candidate SNV

$S_i = w \times \frac{1 + \mathrm{sgn}(PhyloP_i) \times \left(1 - 10^{-|PhyloP_i|}\right)}{2} + (1 - w) \times \left(1 - \frac{P_{LD,i}}{0.05}\right)$, where sgn is a sign function, $PhyloP_i$ is the PhyloP score for primates at the corresponding genomic position, w is the weight for the PhyloP score, and $P_{LD,i}$ is the corresponding Preferential LD value. All other parameters were set as default.

Table S4. Performance of the Preferential LD Approach on GWAS with Known Causal Variants Using Genotypes from the 1000 Genomes Project

Phenotype	Discovery SNP	# of Candidate SNVs				Rank of the Causal SNV in Step 4
		Step 1	Step 2	Step 3	Step 4	
Crohn disease	rs17221417	539	536	536	121	1 (rs2066844)
						2 (rs2066845)
Crohn disease	rs2076756	367	366	366	127	2 (rs2066845)
						3 (rs2066844)
Ribavirin induced haemolytic anaemia	rs6051702	809	800	579	88	1 (rs1127354)
						2 (rs7270101)
Therapeutic warfarin dose	rs4917639	1678	1674	1440	178	2 (rs1799853)
Bladder cancer	rs11892031	489	489	434	191	40 (rs17863783)
Hearing Loss	rs7329467	440	440	280	223	60 (rs35887622)

Table S5. The Ranking of the Common Causal Variants Using Different Statistics[a]

Phenotype	Discovery SNP	Causal SNV	MAF	Rank of Causal SNV in Step 4[b]	Rank of the Causal SNV in Step 1[c]				
					D'	r^2	P_{LD}	PhastCons	Sorting Score
Coronary artery disease	rs1333049[1]	rs10757278[2]	0.496	7	1927	3	8	184	7
		rs10811656[2]	0.420	8	1934	6	8	226	8
Prostate Cancer	rs1447295[3]	rs11986220[4]	0.088	9	604	25	29	209	15

[a]: the preferential LD approach was applied without any MAF requirement of the candidate SNVs while all other parameters were unchanged.

[b]: the rank of causal SNV among final candidates obtained at step 4 of the preferential LD approach.

[c]: the rank of the causal SNV among candidate obtained at step 1 of the preferential LD approach using the corresponding statistic; the rank of the causal variant is the number of candidate SNVs with an equal or better value of the corresponding statistic than the causal variant.

References:

1. Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.-E., et al. (2007). Genomewide Association Analysis of Coronary Artery Disease. New England Journal of Medicine 357, 443-453.
2. Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.-D., Topol, E.J., Rosenfeld, M.G., et al. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature 470, 264-268.
3. Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39, 645-649.
4. Jia, L., Landan, G., Pomerantz, M., Jaschek, R., Herman, P., Reich, D., Yan, C., Khalid, O., Kantoff, P., Oh, W., et al. (2009). Functional Enhancers at the Gene-Poor 8q24 Cancer-Linked Locus. PLoS Genet 5, e1000597.