American Journal of Human Genetics, Volume *83*

Supplemental Data

A Combinatorial Approach to Detecting Gene-Gene and Gene-Environment Interactions in Family Studies

Xiang-Yang Lou, Guo-Bo Chen, Lei Yan, Jennie Z. Ma, Jamie E. Mangold, Jun Zhu, Robert C. Elston, and Ming D. Li

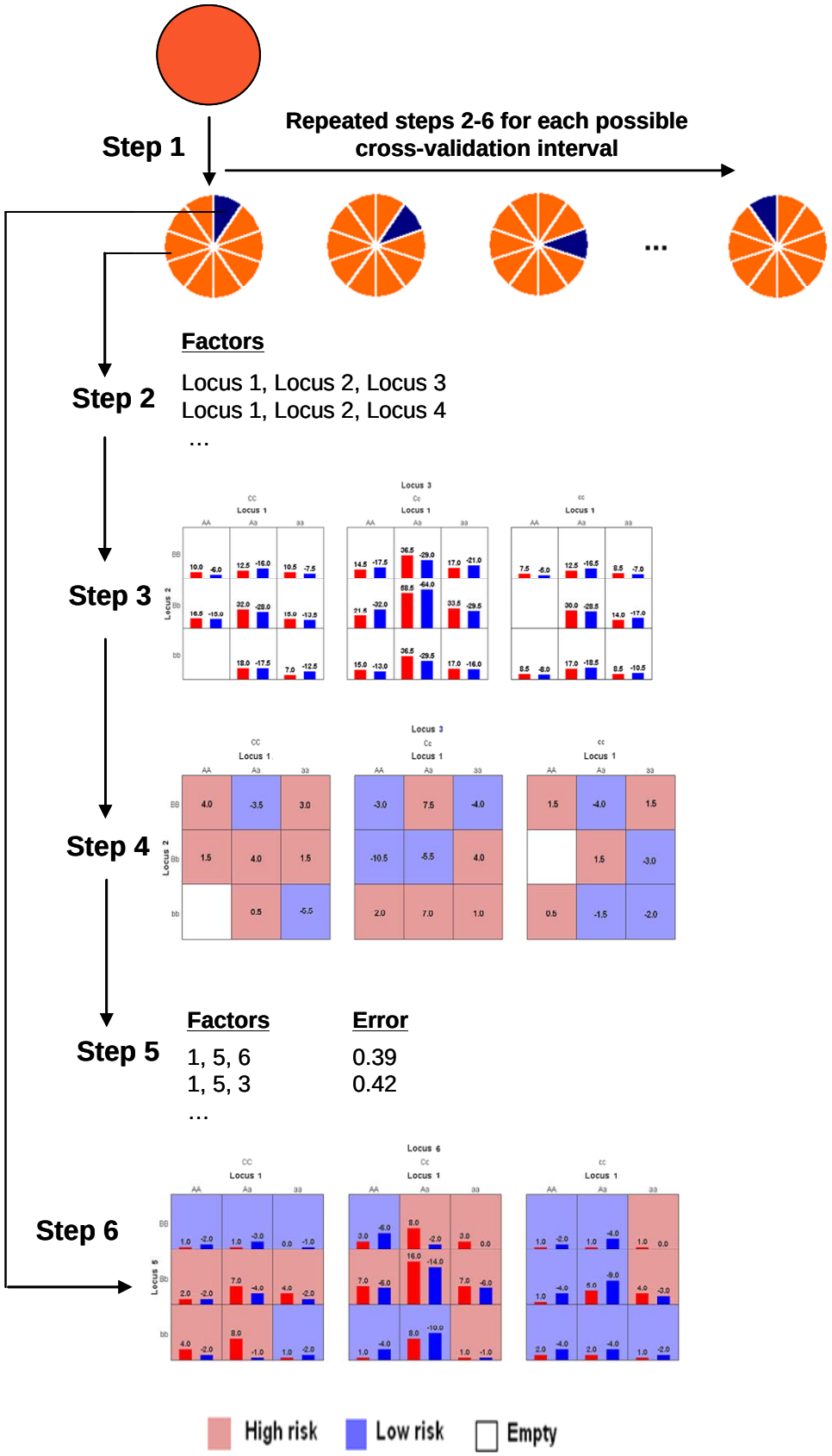

Figure S1. Summary of the steps involved in implementing the pedigree-based GMDR

(PGMDR) method [modified from the work of Ritchie et al. [1] and Lou et al. [2]]. For a detailed description of the Steps, see "A schematic illustration of the pedigree-based GMDR algorithm" in Appendix B. In Step 3, nontransmitted genotypes at loci of interest are constructed from family trio samples and the observed genotypes of affected children are considered as cases and the corresponding nontransmitted genotypes as internal controls. Bars represent hypothetical distributions of positive scores where each case contributes 0.5 (*left, dark red shading*) and negative scores where each control contributes -0.5 (*right, dark blue shading*), and numbers above bars are the sums. In Steps 4 and 6, numbers in the cells are the average statistic values. "High-risk" cells are indicated by *light red shading*, "low-risk" cell by *light blue shading*, and "empty" cells by *no shading*.

Table S1. Comparison of prediction accuracy, crossvalidation consistency and power between PGMDR with covariate adjustment and without covariate adjustment for a dichotomous trait

Model and No. of Loci	With Adjustment (Mean ± SEM)		Power [a]	Without Adjustment (Mean ± SEM)		Power [a]
	Prediction Accuracy	Crossvalidation Consistency		Prediction Accuracy	Crossvalidation Consistency	
Checkerboard [b]						
1	0.508 ± 0.022	7.285 ± 2.118		0.505 ± 0.014	7.355 ± 2.126	
2	***0.581 ± 0.013***	***9.990 ± 0.100***	***0.995***	***0.546 ± 0.011***	***9.835 ± 0.769***	***0.920***
3	0.565 ± 0.020	6.620 ± 1.999		0.535 ± 0.016	6.470 ± 2.198	
4	0.542 ± 0.024	4.965 ± 1.919		0.526 ± 0.017	5.375 ± 2.094	
5	0.515 ± 0.026	4.735 ± 2.130		0.512 ± 0.016	4.850 ± 2.051	
Checkerboard [c]						
1	0.501 ± 0.023	7.110 ± 2.133		0.503 ± 0.013	7.260 ± 2.023	
2	***0.576 ± 0.017***	***9.900 ± 0.657***	***0.980***	***0.542 ± 0.014***	***9.715 ± 0.865***	***0.830***
3	0.556 ± 0.021	6.520 ± 2.030		0.532 ± 0.016	6.475 ± 2.194	
4	0.541 ± 0.023	5.545 ± 2.207		0.524 ± 0.015	5.220 ± 2.164	
5	0.527 ± 0.023	4.985 ± 2.026		0.519 ± 0.015	4.965 ± 2.063	
Anti-diagonal [b]						
1	0.508 ± 0.022	7.350 ± 2.109		0.506 ± 0.015	7.435 ± 2.182	
2	***0.575 ± 0.017***	***9.875 ± 0.634***	***0.975***	***0.544 ± 0.012***	***9.670 ± 1.003***	***0.870***
3	0.560 ± 0.023	6.530 ± 2.025		0.535 ± 0.017	6.530 ± 2.198	
4	0.535 ± 0.028	5.100 ± 2.131		0.522 ± 0.016	5.020 ± 2.032	
5	0.507 ± 0.025	4.830 ± 2.055		0.509 ± 0.014	4.875 ± 1.980	
Anti-diagonal [c]						

1	0.514 ± 0.023	7.770 ± 2.024		0.510 ± 0.014	7.910 ± 2.096	
2	***0.578 ± 0.019***	***9.905 ± 0.507***	***0.965***	***0.541 ± 0.014***	***9.655 ± 1.054***	***0.880***
3	0.560 ± 0.021	6.590 ± 2.211		0.532 ± 0.014	6.335 ± 2.104	
4	0.542 ± 0.023	5.280 ± 2.003		0.524 ± 0.015	5.385 ± 2.029	
5	0.529 ± 0.023	4.875 ± 1.962		0.518 ± 0.015	4.710 ± 1.948	

[a] Power = the rate of true positives in all simulations at 5% level.
[b] The minor allele frequencies at both functional loci are 0.25.
[c] The minor allele frequencies at both functional loci are 0.50.

Table S2. Comparison of crossvalidation consistency, prediction accuracy and power between PGMDR with covariate adjustment and without covariate adjustment for a quantitative trait

Model and No. of Loci	With Adjustment (Mean ± SEM)		Power [a]	Without Adjustment (Mean ± SEM)		Power [a]
	Prediction Accuracy	Crossvalidation Consistency		Prediction Accuracy	Crossvalidation Consistency	
Checkerboard [b]						
1	0.514 ± 0.018	7.870 ± 2.028		0.511 ± 0.016	7.860 ± 2.079	
2	***0.588 ± 0.011***	***10.000 ± 0.000***	***1.000***	***0.569 ± 0.010***	***9.995 ± 0.071***	***0.995***
3	0.580 ± 0.015	6.770 ± 2.116		0.559 ± 0.014	6.510 ± 2.072	
4	0.565 ± 0.017	5.750 ± 2.034		0.546 ± 0.017	5.440 ± 2.029	
5	0.545 ± 0.024	5.250 ± 2.126		0.526 ± 0.019	4.770 ± 1.984	
Checkerboard [c]						
1	0.499 ± 0.018	6.700 ± 2.185		0.501 ± 0.014	7.065 ± 2.110	
2	***0.585 ± 0.012***	***10.000 ± 0.000***	***1.000***	***0.566 ± 0.011***	***9.995 ± 0.071***	***1.000***
3	0.572 ± 0.016	6.730 ± 2.131		0.553 ± 0.014	6.690 ± 2.028	
4	0.555 ± 0.018	5.520 ± 2.059		0.539 ± 0.015	5.455 ± 2.052	
5	0.539 ± 0.017	4.975 ± 2.137		0.529 ± 0.016	5.020 ± 2.088	
Anti-diagonal [b]						
1	0.536 ± 0.014	8.940 ± 1.513		0.527 ± 0.013	8.760 ± 1.651	
2	***0.573 ± 0.012***	***9.975 ± 0.354***	***0.995***	***0.558 ± 0.011***	***9.955 ± 0.322***	***0.995***
3	0.570 ± 0.015	6.885 ± 1.980		0.554 ± 0.013	6.910 ± 2.099	
4	0.557 ± 0.019	5.680 ± 2.102		0.539 ± 0.018	5.135 ± 1.869	
5	0.526 ± 0.024	4.860 ± 1.977		0.513 ± 0.020	4.615 ± 1.922	
Anti-diagonal [c]						
1	0.514 ±	7.770 ± 2.024		0.498 ±	6.750 ± 2.135	

2	0.023 ***0.578 ± 0.019***	***9.905 ± 0.507***	***0.975***	0.017 ***0.558 ± 0.014***	***9.960 ± 0.242***	***0.770***
3	0.560 ± 0.021	6.590 ± 2.211		0.545 ± 0.016	6.590 ± 2.113	
4	0.542 ± 0.023	5.280 ± 2.003		0.533 ± 0.018	5.450 ± 2.064	
5	0.529 ± 0.023	4.875 ± 1.962		0.522 ± 0.019	4.890 ± 2.044	

[a] Power = the rate of true positives in all simulations at 5% level.
[b] The minor allele frequencies at both functional loci are 0.25.
[c] The minor allele frequencies at both functional loci are 0.50.

Table S3. Information on *TAS2R38* and *TAS2R16* genes*

Gene	Chromosome	Location	Gene Size (bp)	No. of Exon	mRNA size (bp)	Protein size (aa)
TAS2R38	7	141,318,900~ 141,320,042	1,143	1	1,143	333
TAS2R16	7	122,421,995~ 122,422,990	996	1	996	291

* All information was obtained from NCBI Entrez Gene database: http://www.ncbi.nlm.nih.gov/entrez/

Table S4. Positions, nucleotide variation, and allele frequencies for *TAS2R16* and *TAS2R38*

Gene	dbSNP ID	Chromosome Position [a]	Alleles [b]	MAF [c]
TAS2R16	rs1204014	122,422,079	**A**/G	0.31
	rs846664	122,422,409	**G**/T	0.30
	rs2233989	122,422,465	**C**/T	0.13
TAS2R38	rs10246939	141,319,073	**C**/T	0.49
	rs1726866	141,319,174	C/**T**	0.34
	rs713598	141,319,814	C/**G**	0.47

[a] From NCBI dbSNP database build 128
[b] The nucleotide of each SNP shown in bold font represents the minor allele in our AA samples
[c] MAF = the minor allele frequency in our AA sample

Supplemental References

1. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF and Moore JH (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 69: 138-47
2. Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC and Li MD (2007) A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet* 80: 1125-37