Supplementary Document of SNPRuler

Xiang Wan[1], Can Yang[1],Qiang Yang[2], Hong Xue[3], Nelson L.S. Tang[4] and Weichuan Yu[1]
[1]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology.
[2] Department of Computer Science, The Hong Kong University of Science and Technology.
[3]Department of Biochemistry, The Hong Kong University of Science and Technology.
[4]Laboratory for Genetics of Disease Susceptibility, Li Ka Shing
Institute of Health Sciences, The Chinese University of Hong Kong.

1 Models used in simulation studies

1.1 Models with marginal effects

Table 1 lists the three 2-locus epistatic models used in Case 1 of Section 3.1. Table 2 shows the 3-locus epistatic model, which is used as Model 4 in Case 1 of Section 3.1.

Model 1	AA	Aa	aa
BB	1	$1+\theta$	$(1+\theta)^2$
Bb	$1+\theta$	$(1+\theta)^2$	$(1+\theta)^3$
bb	$(1+\theta)^2$	$(1+\theta)^3$	$(1+\theta)^4$
Model 2	**AA**	**Aa**	**aa**
BB	1	1	1
Bb	1	$(1+\theta)^2$	$(1+\theta)^3$
bb	1	$(1+\theta)^3$	$(1+\theta)^4$
Model 3	**AA**	**Aa**	**aa**
BB	1	1	1
Bb	1	$(1+\theta)$	$(1+\theta)$
bb	1	$(1+\theta)$	$(1+\theta)$

Table 1: Three models with marginal effects, which are proposed in [1] and also used in [4] and our paper.

	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
BB	1	1	1	1	1	$1+\theta$	1	$1+\theta$	1
Bb	1	1	$1+\theta$	1	$1+\theta$	1	$1+\theta$	1	1
bb	1	$1+\theta$	1	$1+\theta$	1	1	1	1	1

Table 2: The risk odds in a 3-locus epistatic model used in [4] and our paper.

1.2 Parameter setting for data generation in Case 1 of simulation experiments

The data generation follows the same procedure in [4]. The marginal effects of disease loci are controlled by the parameter λ (readers may check [4] for details). The Linkage Disequilibrium (LD) between SNPs is measured by r^2. The following settings are chosen for the practical concerns as discussed in [3]:

- Sample size: 2000 samples (1000 cases and 1000 controls) and 4000 samples (2000 cases and 2000 controls) are simulated under each parameter setting.
- MAF: MAF (minor allele frequency) is chosen from 0.1 to 0.5.
- Marginal effect size λ: $\lambda = 0.3$ for Model 1 and $\lambda = 0.2$ for Model 2, 3 and 4. The value of θ in each model is computed from the specified λ.
- Linkage Disequilibrium (LD) r^2: $r^2 = 1$ for directly genotyped disease loci and $r^2 = 0.7$ for genotyped LD markers.
- SNP size: 2000 SNP markers are simulated in each setting.

2 Models without marginal effects

The epistatic models without marginal effects used in simulation experiments are given in Table 3-8.

$h^2 = 0.4, MAF = 0.2$				$h^2 = 0.4, MAF = 0.4$			
Model epi1	AA	Aa	aa	Model epi6	AA	Aa	aa
BB	0.486	0.960	0.538	BB	0.077	0.656	0.880
Bb	0.947	0.004	0.811	Bb	0.892	0.235	0.312
bb	0.640	0.606	0.909	bb	0.174	0.842	0.106
$h^2 = 0.4, MAF = 0.2$				$h^2 = 0.4, MAF = 0.4$			
Model epi2	AA	Aa	aa	Model epi7	AA	Aa	aa
BB	0.469	0.956	0.697	BB	0.895	0.323	0.161
Bb	0.945	0.019	0.585	Bb	0.068	0.728	0.806
bb	0.786	0.407	0.013	bb	0.925	0.233	0.362
$h^2 = 0.4, MAF = 0.2$				$h^2 = 0.4, MAF = 0.4$			
Model epi3	AA	Aa	aa	Model epi8	AA	Aa	aa
BB	0.498	0.954	0.786	BB	0.805	0.251	0.085
Bb	0.978	0.038	0.428	Bb	0.002	0.668	0.638
bb	0.590	0.821	0.380	bb	0.830	0.079	0.542
$h^2 = 0.4, MAF = 0.2$				$h^2 = 0.4, MAF = 0.4$			
Model epi4	AA	Aa	aa	Model epi9	AA	Aa	aa
BB	0.505	0.988	0.624	BB	0.307	0.682	0.958
Bb	0.945	0.085	0.807	Bb	0.997	0.390	0.281
bb	0.969	0.116	0.159	bb	0.012	0.990	0.698
$h^2 = 0.4, MAF = 0.2$				$h^2 = 0.4, MAF = 0.4$			
Model epi5	AA	Aa	aa	Model epi10	AA	Aa	aa
BB	0.486	0.963	0.512	BB	0.083	0.891	0.037
Bb	0.941	0.006	0.899	Bb	0.619	0.271	0.691
bb	0.691	0.541	0.614	bb	0.853	0.079	0.742

Table 3: Models with $h^2 = 0.4, MAF = 0.2, 0.4$.

$h^2 = 0.3, MAF = 0.2$				$h^2 = 0.3, MAF = 0.4$			
Model epi11	AA	Aa	aa	Model epi16	AA	Aa	aa
BB	0.500	0.926	0.615	BB	0.891	0.362	0.480
Bb	0.895	0.131	0.647	Bb	0.213	0.829	0.601
bb	0.858	0.160	0.999	bb	0.925	0.267	0.685
$h^2 = 0.3, MAF = 0.2$				$h^2 = 0.3, MAF = 0.4$			
Model epi12	AA	Aa	aa	Model epi17	AA	Aa	aa
BB	0.413	0.851	0.535	BB	0.077	0.689	0.417
Bb	0.831	0.008	0.580	Bb	0.763	0.150	0.491
bb	0.692	0.268	0.736	bb	0.196	0.657	0.247
$h^2 = 0.3, MAF = 0.2$				$h^2 = 0.3, MAF = 0.4$			
Model epi13	AA	Aa	aa	Model epi18	AA	Aa	aa
BB	0.455	0.848	0.897	BB	0.132	0.793	0.274
Bb	0.890	0.088	0.016	Bb	0.799	0.213	0.514
bb	0.562	0.686	0.467	bb	0.255	0.528	0.793
$h^2 = 0.3, MAF = 0.2$				$h^2 = 0.3, MAF = 0.4$			
Model epi14	AA	Aa	aa	Model epi19	AA	Aa	aa
BB	0.609	0.980	0.980	BB	0.611	0.104	0.759
Bb	0.993	0.300	0.275	Bb	0.180	0.674	0.019
bb	0.876	0.483	0.683	bb	0.532	0.189	0.681
$h^2 = 0.3, MAF = 0.2$				$h^2 = 0.3, MAF = 0.4$			
Model epi15	AA	Aa	aa	Model epi20	AA	Aa	aa
BB	0.486	0.963	0.512	BB	0.091	0.827	0.863
Bb	0.941	0.006	0.899	Bb	0.869	0.393	0.415
bb	0.691	0.541	0.614	bb	0.738	0.508	0.363

Table 4: Models with $h^2 = 0.3, MAF = 0.2, 0.4$.

$h^2 = 0.2, MAF = 0.2$				$h^2 = 0.2, MAF = 0.4$			
Model epi21	AA	Aa	aa	Model epi26	AA	Aa	aa
BB	0.428	0.757	0.812	BB	0.356	0.891	0.809
Bb	0.788	0.132	0.044	Bb	0.955	0.508	0.611
bb	0.559	0.548	0.373	bb	0.617	0.755	0.630
$h^2 = 0.2, MAF = 0.2$				$h^2 = 0.2, MAF = 0.4$			
Model epi22	AA	Aa	aa	Model epi27	AA	Aa	aa
BB	0.507	0.842	0.605	BB	0.086	0.536	0.641
Bb	0.845	0.162	0.629	Bb	0.677	0.275	0.096
bb	0.581	0.678	0.729	bb	0.219	0.413	0.712
$h^2 = 0.2, MAF = 0.2$				$h^2 = 0.2, MAF = 0.4$			
Model epi23	AA	Aa	aa	Model epi28	AA	Aa	aa
BB	0.577	0.247	0.428	BB	0.855	0.339	0.772
Bb	0.227	0.928	0.578	Bb	0.513	0.651	0.607
bb	0.586	0.262	0.158	bb	0.250	0.999	0.154
$h^2 = 0.2, MAF = 0.2$				$h^2 = 0.2, MAF = 0.4$			
Model epi24	AA	Aa	aa	Model epi29	AA	Aa	aa
BB	0.340	0.637	0.654	BB	0.506	0.838	0.024
Bb	0.689	0.017	0.041	Bb	0.603	0.454	0.957
bb	0.242	0.866	0.403	bb	0.729	0.427	0.753
$h^2 = 0.2, MAF = 0.2$				$h^2 = 0.2, MAF = 0.4$			
Model epi25	AA	Aa	aa	Model epi30	AA	Aa	aa
BB	0.387	0.726	0.734	BB	0.393	0.764	0.664
Bb	0.749	0.090	0.034	Bb	0.850	0.398	0.733
bb	0.551	0.401	0.724	bb	0.406	0.927	0.147

Table 5: Models with $h^2 = 0.2, MAF = 0.2, 0.4$.

$h^2 = 0.1, MAF = 0.2$				$h^2 = 0.1, MAF = 0.4$			
Model epi31	AA	Aa	aa	Model epi36	AA	Aa	aa
BB	0.463	0.703	0.431	BB	0.137	0.484	0.187
Bb	0.653	0.277	0.806	Bb	0.482	0.166	0.365
bb	0.830	0.008	0.129	bb	0.193	0.361	0.430
$h^2 = 0.1, MAF = 0.2$				$h^2 = 0.1, MAF = 0.4$			
Model epi32	AA	Aa	aa	Model epi37	AA	Aa	aa
BB	0.319	0.507	0.569	BB	0.469	0.198	0.754
Bb	0.553	0.105	0.045	Bb	0.337	0.502	0.141
bb	0.203	0.777	0.280	bb	0.339	0.453	0.285
$h^2 = 0.1, MAF = 0.2$				$h^2 = 0.1, MAF = 0.4$			
Model epi33	AA	Aa	aa	Model epi38	AA	Aa	aa
BB	0.627	0.393	0.335	BB	0.478	0.311	0.864
Bb	0.396	0.779	0.953	Bb	0.387	0.579	0.263
bb	0.174	0.842	0.106	bb	0.634	0.436	0.138
$h^2 = 0.1, MAF = 0.2$				$h^2 = 0.1, MAF = 0.4$			
Model epi34	AA	Aa	aa	Model epi39	AA	Aa	aa
BB	0.297	0.540	0.441	BB	0.068	0.299	0.017
Bb	0.541	0.072	0.278	Bb	0.289	0.044	0.285
bb	0.434	0.293	0.228	bb	0.048	0.262	0.174
$h^2 = 0.1, MAF = 0.2$				$h^2 = 0.1, MAF = 0.4$			
Model epi35	AA	Aa	aa	Model epi40	AA	Aa	aa
BB	0.332	0.562	0.573	BB	0.539	0.120	0.258
Bb	0.583	0.112	0.147	Bb	0.165	0.378	0.325
bb	0.399	0.496	0.033	bb	0.123	0.426	0.276

Table 6: Models with $h^2 = 0.1, MAF = 0.2, 0.4$.

$h^2 = 0.05, MAF = 0.2$				$h^2 = 0.05, MAF = 0.4$			
Model epi41	AA	Aa	aa	Model epi46	AA	Aa	aa
BB	0.492	0.664	0.481	BB	0.002	0.155	0.214
Bb	0.642	0.330	0.746	Bb	0.199	0.071	0.022
bb	0.656	0.396	0.000	bb	0.081	0.122	0.135
$h^2 = 0.05, MAF = 0.2$				$h^2 = 0.05, MAF = 0.4$			
Model epi42	AA	Aa	aa	Model epi47	AA	Aa	aa
BB	0.499	0.639	0.765	BB	0.188	0.020	0.171
Bb	0.666	0.389	0.083	Bb	0.032	0.174	0.059
bb	0.543	0.527	0.953	bb	0.134	0.087	0.092
$h^2 = 0.05, MAF = 0.2$				$h^2 = 0.05, MAF = 0.4$			
Model epi43	AA	Aa	aa	Model epi48	AA	Aa	aa
BB	0.212	0.350	0.116	BB	0.005	0.179	0.251
Bb	0.336	0.054	0.495	Bb	0.211	0.100	0.026
bb	0.227	0.273	0.495	bb	0.156	0.098	0.156
$h^2 = 0.05, MAF = 0.2$				$h^2 = 0.05, MAF = 0.4$			
Model epi44	AA	Aa	aa	Model epi49	AA	Aa	aa
BB	0.805	0.683	0.638	BB	0.174	0.321	0.154
Bb	0.657	0.936	0.989	Bb	0.223	0.254	0.245
bb	0.850	0.564	0.866	bb	0.448	0.025	0.424
$h^2 = 0.05, MAF = 0.2$				$h^2 = 0.05, MAF = 0.4$			
Model epi45	AA	Aa	aa	Model epi50	AA	Aa	aa
BB	0.638	0.488	0.383	BB	0.098	0.219	0.302
Bb	0.464	0.765	0.957	Bb	0.302	0.126	0.121
bb	0.580	0.562	0.719	bb	0.053	0.308	0.136

Table 7: Models with $h^2 = 0.05, MAF = 0.2, 0.4$.

$h^2 = 0.025, MAF = 0.2$				$h^2 = 0.025, MAF = 0.4$			
Model epi51	AA	Aa	aa	Model epi56	AA	Aa	aa
BB	0.495	0.415	0.657	BB	0.002	0.155	0.214
Bb	0.429	0.616	0.121	Bb	0.199	0.071	0.022
bb	0.552	0.331	0.419	bb	0.081	0.122	0.135
$h^2 = 0.025, MAF = 0.2$				$h^2 = 0.025, MAF = 0.4$			
Model epi52	AA	Aa	aa	Model epi57	AA	Aa	aa
BB	0.592	0.691	0.743	BB	0.188	0.020	0.171
Bb	0.712	0.493	0.419	Bb	0.032	0.174	0.059
bb	0.580	0.746	0.504	bb	0.134	0.087	0.092
$h^2 = 0.025, MAF = 0.2$				$h^2 = 0.025, MAF = 0.4$			
Model epi53	AA	Aa	aa	Model epi58	AA	Aa	aa
BB	0.108	0.194	0.186	BB	0.005	0.179	0.251
Bb	0.196	0.037	0.045	Bb	0.211	0.100	0.026
bb	0.172	0.073	0.130	bb	0.156	0.098	0.156
$h^2 = 0.025, MAF = 0.2$				$h^2 = 0.025, MAF = 0.4$			
Model epi54	AA	Aa	aa	Model epi59	AA	Aa	aa
BB	0.112	0.186	0.128	BB	0.174	0.321	0.154
Bb	0.193	0.024	0.138	Bb	0.223	0.254	0.245
bb	0.079	0.236	0.251	bb	0.448	0.025	0.424
$h^2 = 0.025, MAF = 0.2$				$h^2 = 0.025, MAF = 0.4$			
Model epi55	AA	Aa	aa	Model epi60	AA	Aa	aa
BB	0.272	0.192	0.185	BB	0.098	0.219	0.302
Bb	0.172	0.367	0.390	Bb	0.302	0.126	0.121
bb	0.345	0.069	0.005	bb	0.053	0.308	0.136

Table 8: Models with $h^2 = 0.025, MAF = 0.2, 0.4$.

3 Results on WTCCC data

Disease	SNP Groups	Location	Related Genes	Individual P-Values	Interaction P-value
Bipolar disease	(rs4844637,rs4844639)	Chr1	(PLXNA2,PLXNA2)	(0.001,0.774)	1.110×10^{-16}
	(rs10937420,rs1522930)	Chr3	(LEPREL1,LEPREL1)	(0.005,0.89)	4.441×10^{-15}
	(rs6438214,rs10511335)	Chr3	(LOC100132655,LOC100132655)	(0.004,0.987)	1.193×10^{-23}
	(rs1396113,rs9993195)	Chr4	(APBB2,APBB2)	(0.124,0.007)	2.509×10^{-24}
	(rs11733953,rs16893604)	Chr4	(LDB2,LDB2)	(0.007,0.311)	6.661×10^{-16}
	(rs1422672,rs10515786)	Chr5	(EBF1,EBF1)	(0.003,0.568)	7.447×10^{-27}
	(rs2992406,rs2935260)	Chr5	(CDC20B,CDC20B)	(0.009,0.689)	5.742×10^{-26}
	(rs1552835,rs17519558)	Chr5	(GRIA1,GRIA1)	(0.027,0.441)	1.665×10^{-15}
	(rs2768558,rs2768541)	Chr6	(ARMC2,ARMC2)	(0.767,0.004)	1.039×10^{-25}
	(rs9320174,rs13218960)	Chr6	(AIM1,AIM1)	(0.001,0.131)	4.441×10^{-16}
	(rs10266006,rs3793181)	Chr7	(NCAPG2,NCAPG2)	(0.039,0.129)	2.350×10^{-23}
	(rs1408526,rs7851968)	Chr9	(DFNB31,DFNB31)	(0.042,0.647)	1.110×10^{-16}
	(rs4242632,rs7857957)	Chr9	(NTRK2,NTRK2)	(0.112,0.498)	6.217×10^{-15}
	(rs11010228,rs4934551)	Chr10	(CCNY,CCNY)	(0.02,0.123)	1.332×10^{-15}
	(rs1530116,rs4372362)	Chr10	(BRWD2,BRWD2)	(0.49,0.004)	1.110×10^{-16}
	(rs6479769,rs7075575)	Chr10	(A1CF,A1CF)	(0.863,0.001)	6.661×10^{-16}
	(rs12244405,rs7075575)	Chr10	(A1CF,A1CF)	(0.878,0.001)	2.220×10^{-15}
	(rs12260411,rs11250624)	Chr10	(ADARB2,ADARB2)	(0.02,0.247)	1.221×10^{-15}
	(rs2422179,rs12790641)	Chr11	(LDLRAD3,LDLRAD3)	(0.004,0.15)	3.331×10^{-16}
	(rs12424737,rs11111782)	Chr12	(NT5DC3,NT5DC3)	(0.018,0.02)	1.110×10^{-16}
	(rs7963772,rs373529)	Chr12	(SCN8A,SCN8A)	(0.409,0.003)	7.105×10^{-15}
	(rs4882804,rs4597130)	Chr12	(TMEM132C,TMEM132C)	(0.816,0.001)	2.220×10^{-16}
	(rs730747,rs9533431)	Chr13	(ENOX1,ENOX1)	(0.249,0.061)	4.219×10^{-14}
	(rs10483596,rs7152370,rs4901281)	Chr14	(UNKNOWN,UNKNOWN,TXNDC16)	(0.156,0.006,0.011)	2.940×10^{-34}
	(rs16969478,rs16969475)	Chr15	(FSIP1,FSIP1)	(0.002,0.976)	2.098×10^{-14}
	(rs8037901,rs967514)	Chr15	(TCF12,TCF12)	(0.134,0.002)	3.963×10^{-14}
	(rs2561037,rs2561036)	Chr16	(SNX29,SNX29)	(0.01,0.003)	1.454×10^{-38}
	(rs8056538,rs7186084)	Chr16	(CDH1,CDH1)	(0.381,0.011)	3.919×10^{-14}
	(rs10408464,rs3746210)	Chr19	(ZNF211,ZNF211)	(0.663,0.005)	1.406×10^{-13}
	(rs3746210,rs9749085)	Chr19	(ZNF211,LOC100130710)	(0.005,0.842)	9.420×10^{-13}
	(rs2079033,rs3746210)	Chr19	(ZNF134,ZNF211)	(0.005,0.005)	5.750×10^{-23}
	(rs17310782,rs6060163)	Chr20	(MYH7B,TRPC4AP)	(0.002,0.249)	1.800×10^{-23}
	(rs2295701,rs6060163)	Chr20	(TRPC4AP,TRPC4AP)	(0.009,0.249)	2.220×10^{-16}
	(rs6088660,rs6060163)	Chr20	(GSS,TRPC4AP)	(0.006,0.249)	3.109×10^{-15}
	(rs1206808,rs1206809)	Chr20	(EYA2,EYA2)	(0.048,0.343)	1.443×10^{-14}
	(rs1885119,rs6060163)	Chr20	(TRPC4AP,TRPC4AP)	(0.04,0.249)	8.515×10^{-14}
	(rs220503,rs6027163)	Chr20	(C20ORF95,C20ORF95)	(0.036,0.344)	1.427×10^{-23}
	(rs10854233,rs1487328)	Chr20	(WFDC11,WFDC9)	(0.036,0.572)	9.992×10^{-14}
Crohn's disease	(rs2104962,rs1931363,rs5744303)	Chr1	(UNKNOWN,CLCA2,CLCA1)	(0.022,0.528,0.499)	4.154×10^{-34}
	(rs2104962,rs272504,rs5744303)	Chr1	(UNKNOWN,ODF2L,CLCA1)	(0.022,0.441,0.499)	6.899×10^{-33}
	(rs2104962,rs6692784,rs5744303)	Chr1	(UNKNOWN,UNKNOWN,CLCA1)	(0.022,0.023,0.499)	5.895×10^{-35}
	(rs1531939,rs9826424)	Chr3	(GRM7,GRM7)	(0.002,0.301)	2.665×10^{-15}
	(rs584852,rs7897278)	Chr10	(SORCS1,SORCS1)	(0.003,0.178)	1.872×10^{-23}
	(rs7154773,rs10130695)	Chr14	(PPM1A,PPM1A)	(0.004,0.476)	4.435×10^{-43}
	(rs17569609,rs4998386)	Chr16	(GRIN2A,GRIN2A)	(0.335,0.018)	3.542×10^{-14}
	(rs783239,rs6501727)	Chr17	(C17ORF77,LOC100133702)	(0.007,0.055)	4.862×10^{-23}
	(rs11867262,rs8065239)	Chr17	(AATF,AATF)	(0.002,0.918)	8.089×10^{-26}
	(rs9896922,rs228879)	Chr17	(PRKCA,PRKCA)	(0.096,0.027)	1.890×10^{-13}
	(rs1006746,rs6016951)	Chr20	(PTPRT,PTPRT)	(0.652,0.004)	4.319×10^{-14}
	(rs2837821,rs9647188)	Chr21	(LOC100134046,LOC100134046)	(0.896,0.003)	3.133×10^{-13}
	(rs8138080,rs739296)	Chr22	(WBP2NL,SEPT3)	(0.172,0.134)	9.326×10^{-15}

Disease	SNP Groups	Location	Related Genes	Individual P-Values	Interaction P-value
Rheumatoid arthritis	(rs2809345,rs2808250)	Chr1	(DDX59,DDX59)	(0.012,0.299)	4.772×10^{-26}
	(rs2273936,rs748405)	Chr1	(C1ORF142,C1ORF142)	(0.003,0.593)	2.331×10^{-15}
	(rs4844637,rs4844639)	Chr1	(PLXNA2,PLXNA2)	(0.002,0.678)	4.219×10^{-15}
	(rs6545864,rs7579073)	Chr2	(LOC729723,LOC729723)	(0.011,0.007)	5.605×10^{-29}
	(rs6545864,rs893589)	Chr2	(LOC729723,EFR3B)	(0.011,0.104)	1.396×10^{-26}
	(rs6545864,rs955855)	Chr2	(LOC729723,RBJ)	(0.011,0.015)	4.135×10^{-26}
	(rs1605705,rs2136152)	Chr3	(GRM7,GRM7)	(0.014,0.038)	2.630×10^{-30}
	(rs9870678,rs10866009)	Chr3	(DNAH12L,DNAH12L)	(0.072,0.138)	2.663×10^{-24}
	(rs2610201,rs2610204)	Chr4	(GALNT17,GALNT17)	(0.002,0.747)	5.773×10^{-15}
	(rs1422672,rs10515786)	Chr5	(EBF1,EBF1)	(0.012,0.386)	1.110×10^{-16}
	(rs2563335,rs2245643)	Chr5	(WDR55,IK)	(0.76,0.004)	1.221×10^{-15}
	(rs801168,rs2245643)	Chr5	(ZMAT2,IK)	(0.785,0.004)	3.442×10^{-15}
	(rs801399,rs2245643)	Chr5	(IK,IK)	(0.803,0.004)	7.438×10^{-15}
	(rs801167,rs2245643)	Chr5	(ZMAT2,IK)	(0.932,0.004)	1.554×10^{-15}
	(rs2803101,rs2846530)	Chr6	(PARK2,PARK2)	(0.1,0.002)	2.220×10^{-16}
	(rs9394893,rs13205512)	Chr6	(TRERF1,TRERF1)	(0.454,0.589)	4.275×10^{-23}
	(rs9293855,rs1891698)	Chr6	(RIMS1,RIMS1)	(0.738,0.074)	6.560×10^{-24}
	(rs10216301,rs860010)	Chr7	(CALN1,CALN1)	(0.26,0.008)	2.887×10^{-15}
	(rs4732651,rs2237812)	Chr8	(EXTL3,EXTL3)	(0.12,0.319)	1.728×10^{-25}
	(rs7085631,rs10883586)	Chr10	(TD1,TD1)	(0.005,0.002)	1.669×10^{-33}
	(rs2961593,rs2924264)	Chr10	(AKR1CL2,AKR1CL2)	(0.228,0.005)	2.531×10^{-26}
	(rs12424737,rs11111782)	Chr12	(NT5DC3,NT5DC3)	(0.022,0.038)	1.887×10^{-15}
	(rs6490403,rs7997274)	Chr13	(KIAA0774,KIAA0774)	(0.002,0.387)	9.481×10^{-25}
	(rs9603602,rs9315724)	Chr13	(COG6,COG6)	(0.071,0.428)	2.998×10^{-15}
	(rs7154773,rs10130695)	Chr14	(PPM1A,PPM1A)	(0.112,0.274)	1.485×10^{-42}
	(rs1420247,rs4238755)	Chr16	(LOC100132440,LOC100132440)	(0.001,0.929)	1.110×10^{-16}
	(rs8058964,rs12598771)	Chr16	(CDH13,CDH13)	(0.002,0.889)	3.061×10^{-26}
	(rs2561037,rs2561036)	Chr16	(SNX29,SNX29)	(0.037,0.955)	3.331×10^{-16}
	(rs901064,rs7503807)	Chr17	(KIAA1303,KIAA1303)	(0.034,0.016)	8.831×10^{-38}
	(rs2343244,rs9908879)	Chr17	(RAB11FIP4,RAB11FIP4)	(0.952,0.044)	1.552×10^{-13}
	(rs11867262,rs8065239)	Chr17	(AATF,AATF)	(0.071,0.836)	8.588×10^{-30}
	(rs3850528,rs6141601)	Chr20	(EPB41L1,EPB41L1)	(0.001,0.235)	4.642×10^{-25}
	(rs1006746,rs6016951)	Chr20	(PTPRT,PTPRT)	(0.178,0.003)	3.997×10^{-15}
	(rs4911287,rs2070320)	Chr20	(BPIL3,BPIL3)	(0.522,0.333)	2.620×10^{-14}
	(rs2836318,rs2836325)	Chr21	(LOC100131955,LOC100131955)	(0.004,0.027)	5.285×10^{-14}
	(rs11090066,rs739296)	Chr22	(WBP2NL,SEPT3)	(0.499,0.006)	2.839×10^{-26}

Table 9: Some significant SNP groups identified by SNPRuler on WTCCC data.

4 The deduction of utility function

	$\zeta = 0$	$\zeta \neq 0$	$Total$
$\mathcal{N}(r)$	a	b	$a+b$
$\mathcal{N}(\neg r)$	c	d	$c+d$
$Total$	$a+c$	$b+d$	$a+b+c+d$

Table 10: The contingency table of a given rule (r, ζ) from observations. There are a controls and b cases which satisfy r, and c controls and d cases which satisfy $\neg r$ (complementary of r). The total number of observations is $a+b+c+d$.

$$\chi^2(r,\zeta) = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \quad (1)$$

$$= \frac{(a(b+d)-b(a+c))^2}{(a+b)(c+d)} \cdot \frac{(a+b+c+d)}{(a+c)(b+d)} \quad (2)$$

$$= \frac{(\frac{b+d}{a+c} - b/a)^2}{(1+b/a)((a+b+c+d)/a - b/a - 1)} \cdot \frac{(a+b+c+d)(a+c)}{(b+d)} \quad (3)$$

$$= \frac{(R-\delta)^2}{(1+\delta)(\gamma-\delta-1)} \cdot \frac{(a+b+c+d)(a+c)}{(b+d)} \quad (4)$$

$$= U(r,\zeta) \cdot \frac{(a+b+c+d)(a+c)}{(b+d)} \quad (5)$$

$$(6)$$

where $R = \frac{(b+d)}{(a+c)}$, $\delta = b/a$ and $\gamma = \frac{a+b+c+d}{a}$

5 Software instruction

- Input file format. The genotype data can either be stored in a single file or two files for case data and control data, respectively. The input file format is illustrated as below,

$$D = \begin{array}{c} \\ T_1 \\ T_2 \\ \cdots \\ T_{m-1} \\ T_m \end{array} \begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & \cdots & S_n & CLASS \end{array} \\ \left[\begin{array}{ccccc} 1 & 2 & \cdots & 2 & 1 \\ 0 & 1 & \cdots & 1 & 0 \\ \cdots & \cdots & & \cdots & \cdots \\ 2 & 0 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 0 & 0 \end{array} \right] \end{array},$$

  where the genotype of SNP is represented with $\{0, 1, 2\}$ corresponding to homozygous reference genotype (AA), heterozygous genotype (Aa), and homozygous variant genotype (aa), respectively. The class label is either 0 (control) or 1 (case).

- Input parameters. To run the program, input the command “java -jar rule.jar listSize depth updateRatio dataFile” or “java -jar rule.jar listSize depth updateRatio controlDataFile caseDataFile”. Here **listSize** is the expected number of interactions, **depth** is the order of interaction, and **updateRatio** is the step size of updating a rule (which usually takes a value between 0 and 1. The default value is 0.2).

References

[1] Marchini J., Donnelly P., Cardon L.R. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases, *Nature Genetics*, **37**,413-417.

[2] Velez D.R., White B.C., Motsinger A.A., Bush W.S., Ritchie M.D., Williams S.M. and Moore J.H. Moore (2007) A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction, *Genet Epidemiol*, **31**, 306-315.

[3] Wang W.Y., Barratt B.J., Clayton D.G. and Todd J.A. (2005) Genome-wide association studies: theoretical and practical concerns, *Nat. Rev. Genet.*, **6**, 109-118.

[4] Zhang Y., Liu J.S. (2007) Bayesian inference of epistatic interactions in case-control studies, *Nature Genetics*, **39(9)**,1167-1173.