Simulations of GWAS datasets harboring patterns of epistasis

1. General description of the four models used in our simulations

Each table below describes a penetrance table. Put in other words, each table cell depicts the relative disease risk given the corresponding genotype combination. With models involving two loci, one will test epistasis as the significant association between the phenotype and one of the two modeled disease susceptibility loci (locus A in model 1, for instance), conditional on the status of the other modeled disease susceptibility locus (locus B in model 1, for instance). This consists in testing association of locus A with the phenotype for a fixed genotype value at locus B, and repeating that operation for each possible genotype value at locus B. In all listed penetrance tables, α is the baseline effect, while β is the interaction effect.

Model 1 is the multiplicative model. This model simulates a pairwise epistatic interaction model. The disease risk increases in a non-linear way with the number of disease susceptibility alleles at each locus.

Table 1. Model 1: multiplicative model.

Model 1	BB	Bb	bb
AA	α	α	α
Aa	α	$\alpha(1+\beta)^2$	$\alpha(1+\beta)^3$
aa	α	$\alpha(1+\beta)^3$	$\alpha(1+\beta)^4$

Model 2 is the threshold model. This model also simulates a pairwise epistatic interaction model. The disease risk is increased when each genetic variant implicated in epistatic interaction presents at least one disease susceptibility allele. In this model, additional disease susceptibility allele (*i.e.* recessive homozygous) does not further increase the relative disease risk.

Table 2. Model 2: threshold model.

Model 2	BB	Bb	bb
AA	α	α	α
Aa	α	$\alpha(1+\beta)$	$\alpha(1+\beta)$
aa	α	$\alpha(1+\beta)$	$\alpha(1+\beta)$

Model 3 simulates a three-way epistatic interaction model. The disease risk is increased when all interacting loci are heterozygous. The disease risk is also increased when the three interacting loci respectively have a different genotype. Put in other words, if one locus is dominant homozygous, another one is heterozygous and the last one is recessive homozygous, then the relative disease susceptibility is increased (but less increased than with three heterozygous interacting loci).

Table 3. Three-way epistatic interaction model.

Model 3	CC			Cc			cc		
	BB	**Bb**	**bb**	**BB**	**Bb**	**bb**	**BB**	**Bb**	**bb**
AA	α	α	α	α	α	$\alpha(1+\beta)$	α	$\alpha(1+\beta)$	α
Aa	α	α	$\alpha(1+\beta)$	α	$\alpha(1+a\beta)$	α	$\alpha(1+\beta)$	α	α
aa	α	$\alpha(1+\beta)$	α	$\alpha(1+\beta)$	α	α	α	α	α

To generate simulated data, we used the models already used by Marchini and co-workers (Marchini *et al*.,1999) (models 1 and 2), and Zhang and Liu (Zhang,Y. and Liu,J.S., 2007) (model 3). In all three models, the prevalence is $P(D)$ = 0.1. Heritability is h^2 = 0.005 in model 1, and h^2 = 0.02 in models 2 and 3. For each experiment, the 2 to 3 SNPs involved in the simulated interaction share the same minor allele frequency (MAF). Across the experiments, the common MAF of modeled epistatic disease susceptibility alleles vary between four possible values: $MAF \in \{0.05, 0.1, 0.2, 0.5\}$.

The Java package GAMETES (Urbanowicz,R.J. *et al*., 2012), version 2.1, was used to simulate these 2- and 3-order epistatic interactions. For 2- and 3-order interactions, GAMETES requires the following input parameters: penetrance table, total number of SNPs, number of SNPs involved in the epistatic interaction (2 or 3), MAF for each such SNP, interval of MAFs for non-causal SNPs, number of cases and number of controls, number of replicates (i.e. datasets generated under these conditions). In this usage, GAMETES automatically computes the heritability h^2 and the disease prevalence $P(D)$.

Following the indications in (Marchini *et al*.,1999), and in (Zhang,Y. and Liu,J.S., 2007), we instantiated four penetrance tables per model (one per MAF shared by the SNPs involved in the epistatic interaction). These 12 penetrance tables are shown below in Section 2. These settings allow to constrain the heritability and the disease penetrance to the wished values.

Model 4 simulates a 5-order epistatic interaction model. The heritability is h^2 = 0.002. The MAF of all five SNPs involved in the simulated epistatic interaction is 0.2. The MAFs of the SNPs which are not risk factors for the disease vary in the range [0.05, 0.5].

To simulate epistatic interactions strictly above 3-order, the software program GAMETES proceeds in a way different than for 2- and 3-orders. The input parameters are the same as for 2- and 3-orders, except that the heritability is now requested (the prevalence is an optional parameter in this usage), and that this time, a penetrance table is computed.

Given the above constraints, GAMETES found 76 potential simulated models for model 4, corresponding to the parameters *quantile population size* in GAMETES. The model with the median detection difficulty was selected out of these 76 models, with the GAMETES parameter *quantile count* = 1. The resulting penetrance table is presented in Section 2.

2. Detailed description of the four models used in our simulations

Model 1	P(D)=0.1, h2=0.005, MAF=0.05		
	BB	**Bb**	**bb**
AA	0.098	0.098	0.098
Aa	0.098	0.2989	0.5222
aa	0.098	0.5222	0.9121

Model 1	P(D)=0.1, h2=0.005, MAF=0.1		
	BB	**Bb**	**bb**
AA	0.096	0.096	0.096
Aa	0.096	0.1971	0.2824
aa	0.096	0.2824	0.4047

Model 1	P(D)=0.1, h2=0.005, MAF=0.2		
	BB	**Bb**	**bb**
AA	0.0921	0.0921	0.0921
Aa	0.0921	0.1445	0.181
aa	0.0921	0.181	0.2266

Model 1	P(D)=0.1, h2=0.005, MAF=0.5		
	BB	**Bb**	**bb**
AA	0.0782	0.0782	0.0782
Aa	0.0782	0.1054	0.1223
aa	0.0782	0.1223	0.142

Model 2	P(D)=0.1, h2=0.02, MAF=0.05		
	BB	**Bb**	**bb**
AA	0.0958	0.0958	0.0958
Aa	0.0958	0.5331	0.5331
aa	0.0958	0.5331	0.5331

Model 2	P(D)=0.1, h2=0.02, MAF=0.1		
	BB	**Bb**	**bb**
AA	0.0918	0.0918	0.0918
Aa	0.0918	0.3192	0.3192
aa	0.0918	0.3192	0.3192

Model 2	P(D)=0.1, h2=0.02, MAF=0.2		
	BB	**Bb**	**bb**
AA	0.0836	0.0836	0.0836
Aa	0.0836	0.2099	0.2099
aa	0.0836	0.2099	0.2099

Model 2	P(D)=0.1, h2=0.02, MAF=0.5		
	BB	**Bb**	**bb**
AA	0.0519	0.0519	0.0519
Aa	0.0519	0.1374	0.1374
aa	0.0519	0.1374	0.1374

Model 3	P(D)=0.1, h2=0.02, MAF=0.05								
	CC			**Cc**			**cc**		
	BB	**Bb**	**bb**	**BB**	**Bb**	**bb**	**BB**	**Bb**	**bb**
AA	0.0942	0.0942	0.0942	0.0942	0.0942	0.6535	0.0942	0.6535	0.0942
Aa	0.0942	0.0942	0.6535	0.0942	0.1469	0.0942	0.6535	0.0942	0.0942
aa	0.0942	0.6535	0.0942	0.6535	0.0942	0.0942	0.0942	0.0942	0.0942

Model 3	P(D)=0.1, h2=0.02, MAF=0.1								
	CC			**Cc**			**cc**		
	BB	**Bb**	**bb**	**BB**	**Bb**	**bb**	**BB**	**Bb**	**bb**
AA	0.099	0.099	0.099	0.099	0.099	0.549	0.099	0.549	0.099
Aa	0.099	0.099	0.549	0.099	0.1436	0.099	0.549	0.099	0.099
aa	0.099	0.549	0.099	0.549	0.099	0.099	0.099	0.099	0.099

Model 3	P(D)=0.1, h2=0.02, MAF=0.2								
	CC			**Cc**			**cc**		
	BB	**Bb**	**bb**	**BB**	**Bb**	**bb**	**BB**	**Bb**	**bb**
AA	0.0916	00916	0.0916	0.0916	0.0916	0.3141	0.0916	0.3141	0.0916
Aa	0.0916	0. 0916	0.3141	0.0916	0.1014	0.0916	0.3141	0.0916	0.0916
aa	0.0916	0.3141	0.0916	0.3141	0.0916	0.0916	0.0916	0.0916	0.0916

Model 3	P(D)=0.1, h2=0.02, MAF=0.5								
	CC			Cc			cc		
	BB	Bb	bb	BB	Bb	bb	BB	Bb	bb
AA	0.0842	0.0842	0.0842	0.0842	0.0842	0.213	0.0842	0.213	0.0842
Aa	0.0842	0.0842	0.213	0.0842	0.095	0.0842	0.213	0.0842	0.0842
aa	0.0842	0.213	0.0842	0.213	0.0842	0.0842	0.0842	0.0842	0.0842

Model 4	P(D)=0.1, h2=0.002, MAF=0.2		
	CC, DD, EE		
	BB	Bb	bb
AA	0.45754564	0.45372863	0.45679482
Aa	0.45375128	0.46118589	0.45784658
aa	0.45661364	0.45820894	0.43586443
	Cc, DD, EE		
	BB	Bb	bb
AA	0.45406104	0.46087487	0.45537852
Aa	0.46075669	0.44778968	0.45292959
aa	0.45632393	0.45103877	0.49786115
	cc, DD, EE		
AA	0.45413551	0.46069713	0.45560895
Aa	0.46128017	0.4460624	0.45837211
aa	0.45094456	0.46770089	0.4506341
	CC, Dd, EE		
	BB	Bb	bb
AA	0.45345118	0.46209259	0.45539462
Aa	0.46203869	0.44521851	0.45298704
aa	0.45582582	0.45212464	0.49714395
	Cc, Dd, EE		
	BB	Bb	bb
AA	0.46144451	0.44578571	0.45795625
Aa	0.44605222	0.47589497	0.46335891
aa	0.45582422	0.46762298	0.37318287
	cc, Dd, EE		
	BB	Bb	bb
AA	0.46057924	0.447587	0.45739028
Aa	0.44631737	0.47669824	0.45269021
aa	0.46754732	0.43237612	0.46758795

	CC, dd, EE		
	BB	**Bb**	**Bb**
AA	0.45901443	0.45095538	0.45548014
Aa	0.45102422	0.46663176	0.45791252
aa	0.45492942	0.45901396	0.45637169
	Cc, dd, EE		
	BB	**Bb**	**bb**
AA	0.45082139	0.46707049	0.45764789
Aa	0.46682928	0.43555953	0.45360939
aa	0.45957759	0.44974999	0.45611276
	cc, dd, EE		
	BB	**Bb**	**bb**
AA	0.45655205	0.45550408	0.45848876
Aa	0.45633234	0.45676975	0.45187864
aa	0.45186264	0.46513088	0.45650491
	CC, DD, Ee		
	BB	**Bb**	**bb**
AA	0.45104066	0.46681377	0.45619344
Aa	0.46689535	0.43553368	0.45275906
aa	0.45554074	0.45406445	0.48618665
	Cc, DD, Ee		
	BB	**Bb**	**bb**
AA	0.46692468	0.43502656	0.45634682
Aa	0.43478401	0.49837067	0.46384458
aa	0.4582872	0.45996381	0.39504841
	cc, DD, Ee		
	BB	**Bb**	**bb**
AA	0.45530614	0.45812142	0.45748457
Aa	0.45875641	0.45184987	0.45245257
aa	0.45240461	0.46261249	0.46798049
	CC, Dd, Ee		
	BB	**Bb**	**bb**
AA	0.46748064	0.43392094	0.45629641
Aa	0.43377367	0.50039806	0.46379101
aa	0.45747458	0.46143467	0.39628351
	Cc, Dd, Ee		
	BB	**Bb**	**bb**
AA	0.43368486	0.50146345	0.4566887
Aa	0.50191076	0.36707822	0.44015624
aa	0.45311024	0.44731316	0.57908507

	cc, Dd, Ee		
	BB	**Bb**	**bb**
AA	0.45818501	0.4529115	0.45310197
Aa	0.45168938	0.46435949	0.46544815
aa	0.46287898	0.44589414	0.43413737

	CC, dd, Ee		
	BB	**Bb**	**bb**
AA	0.45085846	0.46696638	0.45788784
Aa	0.46683915	0.43563082	0.45288113
aa	0.4589056	0.45084561	0.45809967

	Cc, dd, Ee		
	BB	**Bb**	**bb**
AA	0.46708037	0.43522159	0.45229543
Aa	0.43552389	0.4967976	0.46459102
aa	0.44987699	0.46942792	0.453899

	cc, dd, Ee		
	BB	**Bb**	**bb**
AA	0.45697584	0.45411943	0.46278526
Aa	0.45373651	0.46288028	0.44452783
aa	0.46584863	0.43840107	0.44656739

	CC, DD, ee		
	BB	**Bb**	**bb**
AA	0.47829855	0.41318596	0.44908958
Aa	0.41217091	0.5441104	0.45973635
aa	0.45721001	0.44349548	0.54403011

	Cc, DD, ee		
	BB	**Bb**	**bb**
AA	0.40698006	0.55314374	0.47052331
Aa	0.55697494	0.25575388	0.44972409
aa	0.4398737	0.51102331	0.28118847

	cc, DD, ee		
	BB	**Bb**	**bb**
AA	0.49873684	0.37122878	0.45773445
Aa	0.35682004	0.65555671	0.45377979
aa	0.57300439	0.22323991	0.4533646

	CC, Dd, ee		
	BB	**Bb**	**bb**
AA	0.41229022	0.54250521	0.47066897
Aa	0.54454578	0.28067353	0.44923348
aa	0.45434442	0.48188256	0.28278284

	Cc, Dd, ee		
	BB	**Bb**	**bb**
AA	0.55476305	0.2630751	0.4265446
Aa	0.25523257	0.85640891	0.47236162
aa	0.48928478	0.34688127	0.80374759
	cc, Dd, ee		
	BB	**Bb**	**bb**
AA	0.37260625	0.62267002	0.46429387
Aa	0.65276111	0.06530636	0.44072553
aa	0.22356513	0.92218299	0.452848
	CC, dd, ee		
	BB	**Bb**	**bb**
AA	0.45625555	0.45633702	0.45656917
Aa	0.45625334	0.45619949	0.45770472
aa	0.45723857	0.45636591	0.44060964
	Cc, dd, ee		
	BB	**Bb**	**bb**
AA	0.45756883	0.45245355	0.46662442
Aa	0.45389451	0.46402079	0.43287568
aa	0.45509676	0.45593101	0.47835791
	cc, dd, ee		
	BB	**Bb**	**bb**
AA	0.44671466	0.48633344	0.369252
Aa	0.47614458	0.39599595	0.6210732
aa	0.45076286	0.45805149	0.53073653

Bibliographical references

Marchini,J. *et al*. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet*. 37,413–417. doi:10.1038/ng1537.

Urbanowicz,R.J. *et al*. (2012) GAMETES: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures. *BioData Min.*, 5, 16. doi: 10.1186/1756-0381-5-16.

Zhang,Y. and Liu,J.S. (2007) Bayesian inference of epistatic interactions in case-control studies. *Nat. Genet.* 39,1167–1173. doi:10.1038/ng2110.