

Results for the simulated data harboring a 5-order epistasis interaction

In this experiments, the model used to generate the datasets is model 4. This model is described in additional file 3 (Supplementary data).

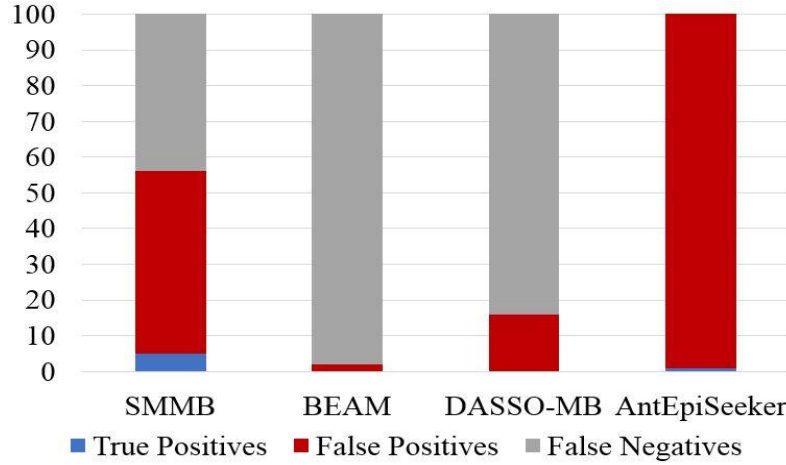


Figure 1. Comparison of the percentages of true positives, false positives and false negatives between SMMB, BEAM, DASSO-MB and AntEpiSeeker, for the simulated datasets harboring a 5-order epistasis interaction. Simulated data: model 4 (5-order interaction), 2000 cases, 2000 controls, 100 SNPs, common MAF of the SNPs involved in the 5-order interaction: $MAF = 0.2$, other MAFs in $[0.05, 0.5]$, number of datasets replicated = 100, heritability = 0.002, disease prevalence = 0.1. Parameters for SMMB: $t = 1000$, $r = 100$, $m = 30$, $K = 10$, $k = 3$, $\alpha = 0.05$. Parameters for BEAM: *nb iterations for burn-in phase: 1000, nb iterations for stationary phase: 10000*. Parameters for DASSO-MB: $\alpha = 0.05$. Parameters for AntEpiSeeker: 450 iterations, 1000 ants, $\alpha = 0.01$.

Table 1. Number of simulated epistatic interactions retrieved over 100 datasets each harboring a simulated interaction. For further details, see Figure 1.

Number of simulated interactions retrieved (partially or totally)		Software			
		SMMB	BEAM	DASSO-MB	AntEpiSeeker
Size of the subset retrieved	5	0	0	0	0
	4	1	0	0	0
	3	1	0	0	0
	2	3	0	0	1

Comments

Figure 1 indicates that in this experiment, all the methods show practically null power. SMMB is nonetheless able to detect a subset for few true interactions. AntEpiSeeker is only able to detect a subset of size 2 for one interaction.

The difference between SMMB and AntEpiSeeker is that the former misses simulated interactions for half of the cases, whereas the latter always yields false interactions.

It is worth noting that existing published methods have been evaluated for 3-way epistasis models at the most, and have so far retrieved 3-way epistatic patterns at the best, on real GWAS datasets. An exception is the recently published method HiSeeker (Liu *et al.*, 2017), which was exceptionally tested for a 6-way epistatic model. However, this model was a model with *marginal* effect for each SNP, which does not correspond to the hypothesis SMMB deals with. Moreover, HiSeeker is a two-stage approach which first attempts to identify 2-way interactions having significant or intermediate association with the phenotype, to further extend them if possible. Thus HiSeeker tackles the problem of detecting embedded epistasis, whereas SMMB copes with pure epistasis detection.

References

Wu, T.T *et al.* (2009) Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6), 714-721. doi: 10.1093/bioinformatics/btp041.