

Genetics and population analysis

Response: an empirical comparison of several recent epistatic interaction detection methods

Yue Wang^{1,*}, Guimei Liu², Mengling Feng³ and Limsoon Wong²

¹NUS Graduate School for Integrative Sciences and Engineering, ²School of Computing, National University of Singapore and ³Data Mining Department, Institute for Infocomm Research, Singapore

Associate Editor: Jonathan Wren

Contact: fayue1015@gmail.com

Received and revised on October 6, 2011; accepted on October 22, 2011

We thank Wan *et al.* (2011) for their useful comments on our recently published paper in *Bioinformatics* (Wang *et al.*, 2011). We are grateful also to the editor for providing us this opportunity to clarify the issues.

As mentioned by Wan *et al.* (2011), the two common procedures for testing statistical epistasis are as follows: (i) ‘two-locus interaction’ as used in BOOST (Wan *et al.*, 2010a); and (ii) ‘two-locus association allowing for interaction’ as used in SNPRuler (Wan *et al.*, 2010b), SNPHarvester (Yang *et al.*, 2009), TEAM (Zhang *et al.*, 2010) and Screen and Clean (Wu *et al.*, 2010). In our paper, we tested BOOST, TEAM, SNPRuler, SNPHarvester, and Screen and Clean using both types of epistatic interactions. We believe this is a more realistic and more comprehensive treatment because, in practice, researchers may have no idea whether the underlying epistatic interactions in their data have main effects or not. Our results suggest that researchers should apply both TEAM and BOOST to their data. Given the same nominal threshold, if BOOST outputs more significant pairs, this hints that pure epistatic interaction effects hidden in the data dominate over marginal effects. A similar logic follows for TEAM.

Wan *et al.* (2011) queried our experiments on the completeness of BOOST. Upon a careful review of our experiments, we discovered a bug in our program script. In particular, we missed out in our program script a step to sort the output of BOOST. This caused the top interacting SNP pair to be not always chosen in each test. Consequently, we misreported in Section 6.5 of our paper that BOOST wrongly pruned the most significant SNP pairs in 4195 datasets without main effect and 756 datasets with main effect. After correcting the program script, BOOST was verified to be complete and did not mis-prune any most significant SNP pairs. This point is corrected in our published paper. Thus, Figure 5 of our original paper should be revised as per Figure 1 here.

Wan *et al.* (2011) also commented on our experiments on ‘type-1 error rate’ of BOOST. The type-1 error rate we observed is relatively high compared with the theoretical nominal significance level (0.05). However, this is the empirical result obtained from our simulated

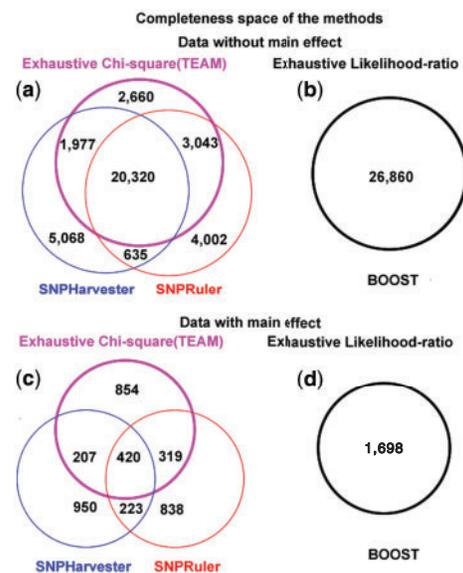


Fig. 1. The completeness space for the four methods. As there are two types of datasets and two types of test statistics, four venn diagrams are drawn respectively. In (a), all three methods—TEAM, SNPRuler and SNPHarvester—use chi square test. TEAM’s outputs represent the 28 000 (20 320 + 1977 + 2660 + 3043) top significant SNP pairs in 28 000 datasets. SNPRuler can identify 22 297 (20 320 + 1977) of them. Among the 28 000 top SNP pairs, 20 320 of them can be identified by all three methods. (b), (c) and (d) follow similar explanations.

datasets. We double-checked our experiments and results; and we also investigated those false positives identified by BOOST. We think there are two reasons for this result. First, with the nominal significance threshold set to 0.05, the test value of the likelihood ratio test with 4 degrees of freedom is 38.24 (after bonferroni correction), while the test value of the chi-square test with 8 degrees of freedom is 47.87 (after bonferroni correction). For example, BOOST reports the top SNP pair (SNP669 and SNP723) with likelihood ratio test statistic value 39.173 845 in the 19th dataset of our simulated datasets. This value exceeds the nominal likelihood ratio test threshold of 38.24. Thus, we report it as a type-1 error of BOOST. However, the same pair using chi-square test gives the test value of 37.2438, which is below the nominal chi-square test threshold of 47.87. Thus TEAM, SNPRuler and SNPHarvester report nothing on this dataset. Second, we increase the number of

*To whom correspondence should be addressed.

simulations from 1000 to 10 000, the type-1 error we obtained is 0.0545, which is close to the nominal significance level (0.05).

Funding: This research is funded in part by a National University of Singapore NGS scholarship (to Y.W.) and by a Singapore Agency for Science Technology & Research grant SERC 102 101 0030 (to L.W. and G.L.).

Conflict of Interest: none declared.

REFERENCES

- Wan,X. et al. (2010a) BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am. J. Hum. Genet.*, **87**, 325–340.
- Wan,X. et al. (2010b) Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics*, **26**, 30–37.
- Wan,X. et al. (2011) Comments on 'An empirical comparison of several recent epistatic interaction detection methods'. *Bioinformatics*, this issue.
- Wang,Y. et al. (2011) An empirical comparison of several recent epistatic interaction detection methods. *Bioinformatics*, **27**, 2936–2943.
- Wu,J. et al. (2010) Screen and Clean: a tool for identifying interactions in genome-wide association studies. *Genet. Epidemiol.*, **34**, 275–285.
- Yang,C. et al. (2009) SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies. *Bioinformatics*, **25**, 504–511.
- Zhang,X. et al. (2010) TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics*, **26**, i217–i227.