

ORAL PRESENTATION

Open Access

Replication of epistatic DNA loci in two case-control GWAS studies using OPE algorithm

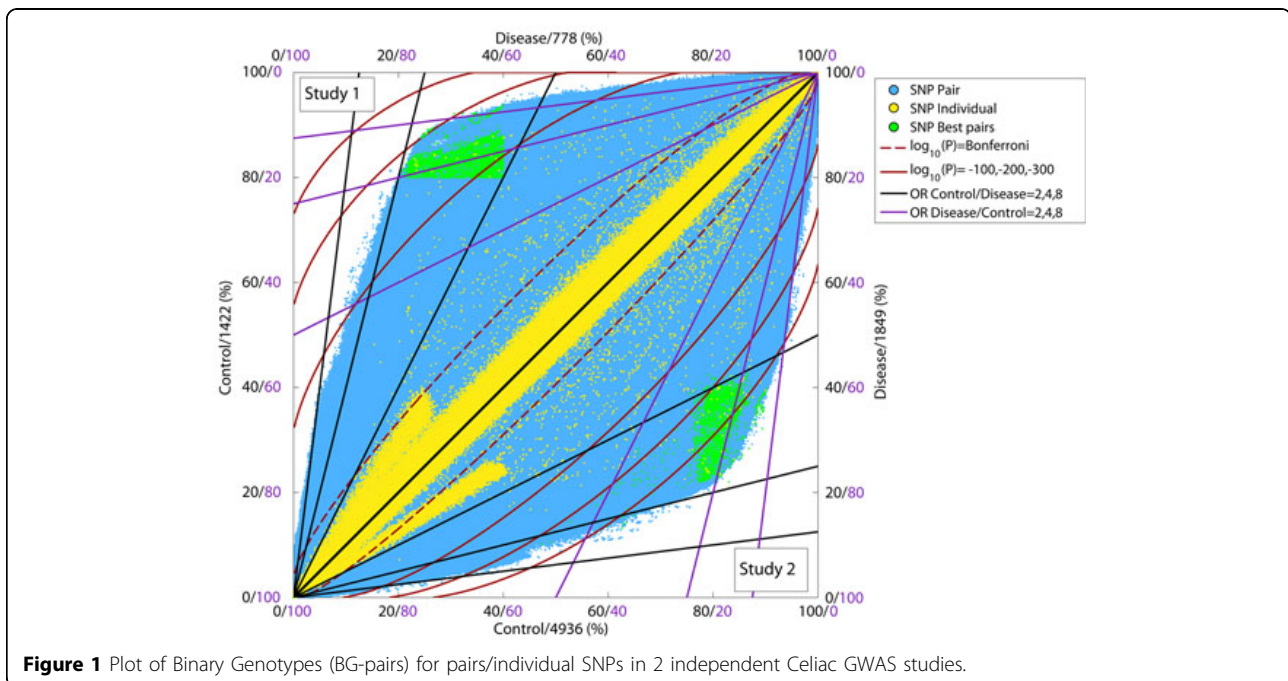
Benjamin Goudey^{1,2*}, Qiao Wang², Dave Rawlinson², Armita Zarnegar², Eder Kikianty², John Markham², Geoff Macintyre^{1,2}, Gad Abraham^{1,2}, Linda Stern¹, Michael Inouye^{3,5}, Izhak Haviv^{2,4}, Adam Kowalczyk²

From Seventh International Society for Computational Biology (ISCB) Student Council Symposium 2011 Vienna, Austria. 15 July 2011

Background

One of the limiting factors of current genome-wide association studies (GWAS) is the inability of current methods to comprehensively examine SNP interactions for a reasonable sized dataset. It is hypothesised that this limitation is one of the reasons that GWAS studies have not been able to have a greater impact [1,2]. Many current methods for handling interactions are

computationally expensive and do not scale to entire studies. Those methods that do scale often achieve this by pruning their datasets in some manner. This is commonly done by considering only those SNPs that show strong marginal effects, despite the fact that a strongly interacting pair may consist of SNPs with low effects individually.



* Correspondence: bwgoudey@csse.unimelb.edu.au

¹Department of Software Engineering and Computer Science, The University of Melbourne, Parkville, Victoria 3010, Australia

Full list of author information is available at the end of the article

Material and methods

In this presentation, we validate the robustness of a novel algorithm known as Optimal Pairwise Epistasis (OPE) for exhaustively examining all pairwise interactions in GWAS data. This method is based on the systematic evaluation of “binary genotype pairs” (BG-pairs), i.e. the pairs of complementary binary classification of genotype calls for an individual SNP, or a pair of SNPs. We can quantify the discrimination potential of BG-pairs using a family of statistics based on odds ratios.

Results and conclusion

The approach is computationally efficient: the dataset reported here as Study 1 (consisting of ~310K SNPs and 2200 samples [3]) takes 12 hour to process on a single CPU (compared to 149 hours of the recent BOOST algorithm [4]). The method can be highly parallelised with a recent GPU implementation reducing this processing time to less than 15 minutes.

We have tested our approach over 2 independent GWAS studies of Celiac disease: the first (Study 1 mentioned above, [3]) with 778/1422 and the second (Study 2, [5]) with 1849/4936 of case/control samples, respectively. Each point in the figure 1 below shows the observed frequency of the BG carriers for the case and control subpopulations: in blue for a pair of SNPs or in yellow for an individual SNP. Every BG-pair can be evaluated with respect to the two sets of axes labels: purple labels for the protective BG and black labels for the risk BG. The resulting figure shows both studies related by symmetry in the main diagonal and indicates replication of results across studies. We emphasise the replicability of our approach by showing in green the same subset of SNP pairs in both studies. We also show in red contours for p-values and plot in black / purple solid diagonal lines to indicate different odds ratios.

Author details

¹Department of Software Engineering and Computer Science, The University of Melbourne, Parkville, Victoria 3010, Australia. ²National ICT Australia (NICTA) Victoria Research Laboratories, The University of Melbourne, Parkville, Victoria 3010, Australia. ³The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia. ⁴Baker IDI Heart and Diabetes Institute, Melbourne, Victoria 3004, Australia. ⁵Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia.

Published: 21 November 2011

References

1. Cordell HJ: Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 2009, **10**:392-404.
2. Moore JH, Asselbergs FW, Williams SM: Bioinformatics challenges for genome-wide association studies. *Bioinformatics* 2010, **26**:445-455.
3. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, et al: A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007, **39**:827-829.

4. Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NL, Yu W: BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet* 2010, **87**:325-340.
5. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GA, Adány R, Aromaa A, et al: Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010, **42**:295-302.

doi:10.1186/1471-2105-12-S11-A5

Cite this article as: Goudey et al.: Replication of epistatic DNA loci in two case-control GWAS studies using OPE algorithm. *BMC Bioinformatics* 2011 **12**(Suppl 11):A5.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

