GWIS: Online exhaustive bivariate GWAS in minutes…

David Rawlinson, Herman Ferra, Andrew Kowalczyk, Qiao Wang, Ben Goudey, Adam Kowalczyk

Correspondance: gwis@nicta.com.au

GWIS (Genome-Wide Interaction Search) is a fast method for detecting bivariate association between genotype and phenotype in GWAS data. The algorithms used in GWIS were recently evaluated against conventional methods [1] on 7 Welcome-Trust Case-Control Consortium datasets [2].

Not only was it shown that GWIS methods were faster than all other exhaustive algorithms, but they explicitly search for a well-defined proxy of epistasis: An improvement in association for SNP pairs, over the association for each individual SNP.

Web Service

We have now developed a free online interface to GWIS, based on an instance of the GalaxyProject server [3]. Users can upload GWAS datasets in PLINK format [4] for processing with a battery of popular conventional tests for association, or using the 3 tests specific to GWIS (SS, DSS and GSS).

The server is free for public use as a demonstration of our methods. It is hosted on only a single desktop machine, yet exhaustive bivariate analysis for e.g. 3 tests can be completed in 15 minutes on a dataset of equal dimension to the WTCCC examples.

For each statistical test, the server returns a separate list of the most significant N SNP pairs along with the score computed. Up to 1 million pairs can be ranked.

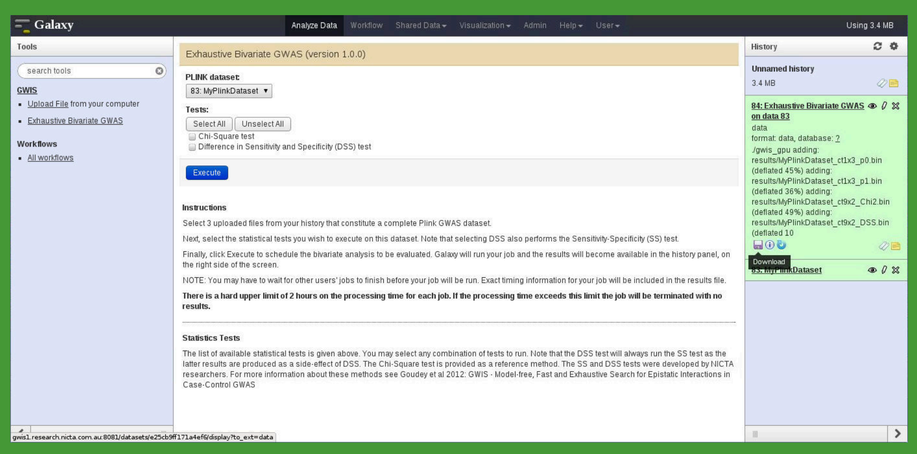
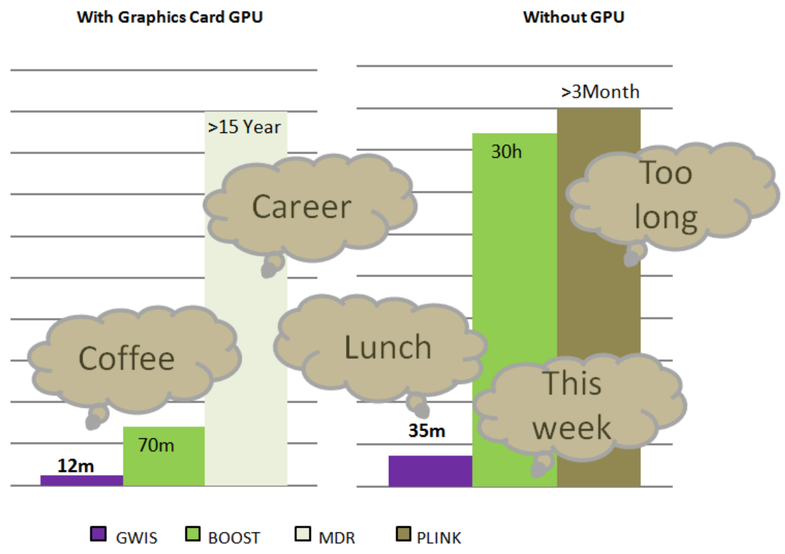


Figure 1: Screenshot of the GalaxyProject web server running GWIS

Timing Data

Execution times for exhaustive bivariate analysis using the GWIS demo server are reported below. For comparison, we include timing figures reported for BOOST / GBOOST[5] and PLINK[4], the fastest and most widely cited alternative methods reported in the literature.



Processing Pipeline

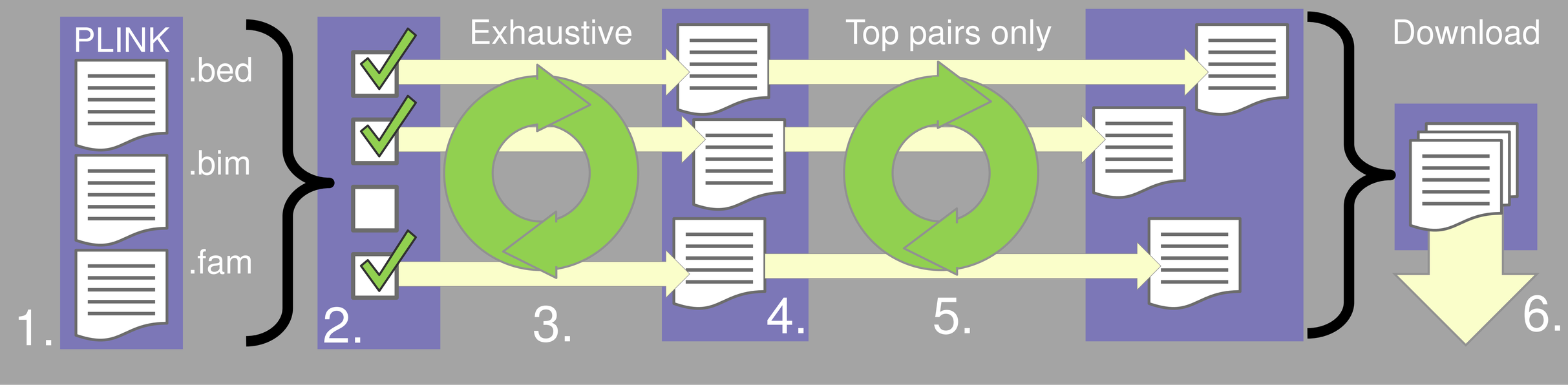


Figure 2: The GWIS web-server pipeline. Users upload PLINK datasets via the web interface (1). Users then select a set of statistical tests (2). Galaxy queues the job until the machine is available, and then triggers GWIS (3) to execute an exhaustive bivariate analysis using the specified tests.

Each test produces a list of SNP pairs ranked by test score (4). Optionally, (5), the GSS test can be run on the ranked SNP pairs produced by exhaustive analysis. Finally, users can download a zip file containing all results (6).

Statistical Tests

The DSS and GSS statistics featured in GWIS are explicitly designed to detect epistatic effects by measuring bivariate Gain in Sensitivity and Specificity over univariate models. DSS and GSS tests construct ROC-based classifiers for all combinations of genotypes, at each probe, univariately and bivariately.

Any improvement in sensitivity and specificity between univariate and bivariate classifiers is quantified as a Binomial selection experiment. For speed, the DSS test uses a constant univariate classifier based on the null hypothesis that there is no association between genotype and phenotype. The ROC curve of this classifier is a straight diagonal line (figure 4, left).

The more computationally demanding GSS test creates a null hypothesis for each SNP-pair, by merging optimal ROC classifiers for univariate contingencies of the two SNPs (figure 4, right).

New statistics to be added

A number of new statistics will be added to the GWIS web server in coming weeks. These include regression based likelihood ratio tests (equivalent to GBOOST).

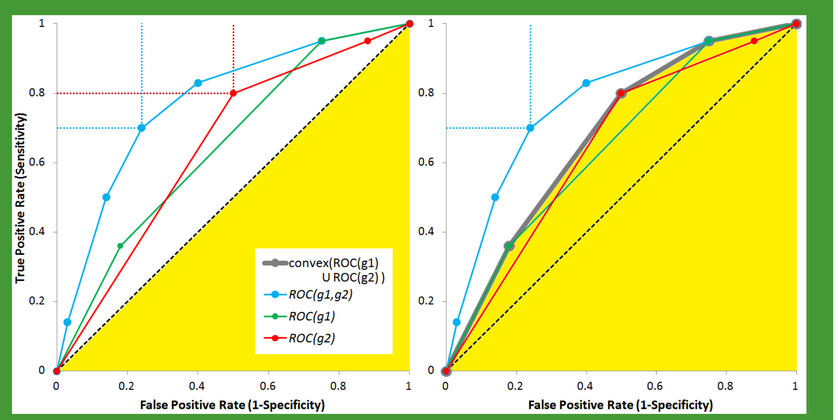


Figure 4: Left: Optimal ROC classifiers for univariate and bivariate contingencies. For the DSS test, the null hypothesis is the area under the diagonal line (shaded yellow). Right: For the GSS test, the null hypothesis is derived from the optimal classifier produced by both SNPs' univariate contingencies.

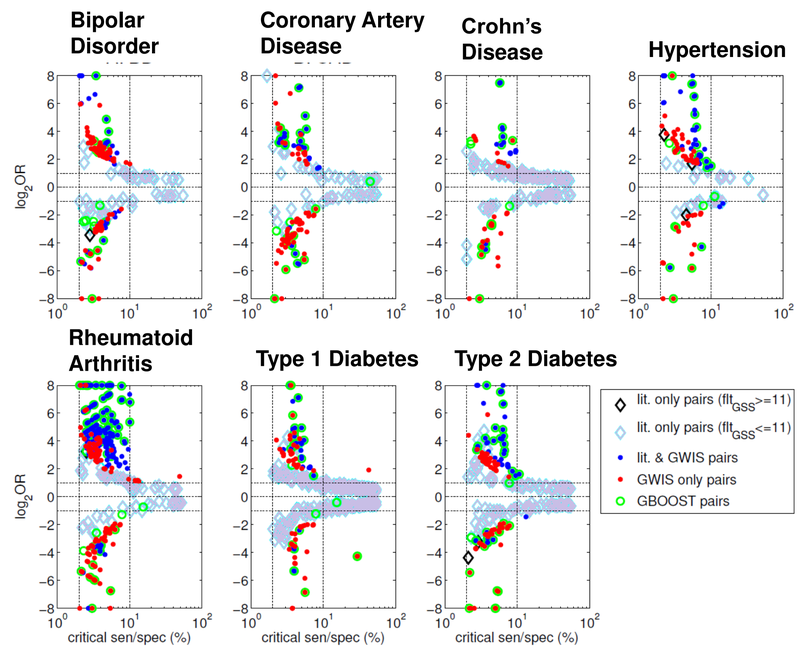


Figure 3: Odds-ratio and coverage plots of top-ranked SNP pairs picked by GWIS and literature methods on 7 WTCCC datasets. GWIS detects many pairs with comparatively high odds-ratios and coverage that were not reported by other methods from the literature.

[1] B. Goudey, D. Rawlinson, Q. Wang, F. Shi, H. Ferra, R. M. Campbell, L. Stern, M. T. Inouye, C. Soon Ong, A. Kowalczyk, "GWIS - Model-free, Fast and Exhaustive Search for Epistatic Interactions in Case-Control GWAS", BMC Genomics 2013.
[2] Wellcome Trust Case-Control Consortium http://www.wtccc.org.uk
[3] GalaxyProject web server software http://galaxyproject.org
[4] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC "PLINK: a toolset for whole-genome association and population-based linkage analysis." American Journal of Human Genetics. 2007
[5] X. Wan*, C. Yang*, Q. Yang, H. Xue, X. Fan, N. Tang and W. Yu. *Joint first author.
"BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies", American Journal of Human Genetics. 2010

Research Excellence in ICT

Wealth Creation for Australia

This work was supported by the Australian Research Council (ARC), and by the NICTA Victorian Research Laboratory. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications, and the Digital Economy, and the Australian Research Council through the ICT Centre of Excellence program.

This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475.