Stability of bivariate GWAS biomarker detection

Justin Bedő[1,2], David Rawlinson[1,3], Cheng Soon Ong[1,3]

[1] NICTA Victoria Research Laboratory, University of Melbourne, Victoria 3010, Australia

[2] Department of Computing and Information Systems, University of Melbourne, Victoria 3010, Australia

[3] Department of Electrical & Electronic Engineering, University of Melbourne, Victoria 3010, Australia

E-mail: {justin.bedo, david.rawlinson, cheng-soon.ong}@nicta.com.au

1 Spearman's ρ

The results from measuring Spearman's ρ against the whole list using 2-fold cross-validation repeated 10 times on seven WTCCC case–control datasets.

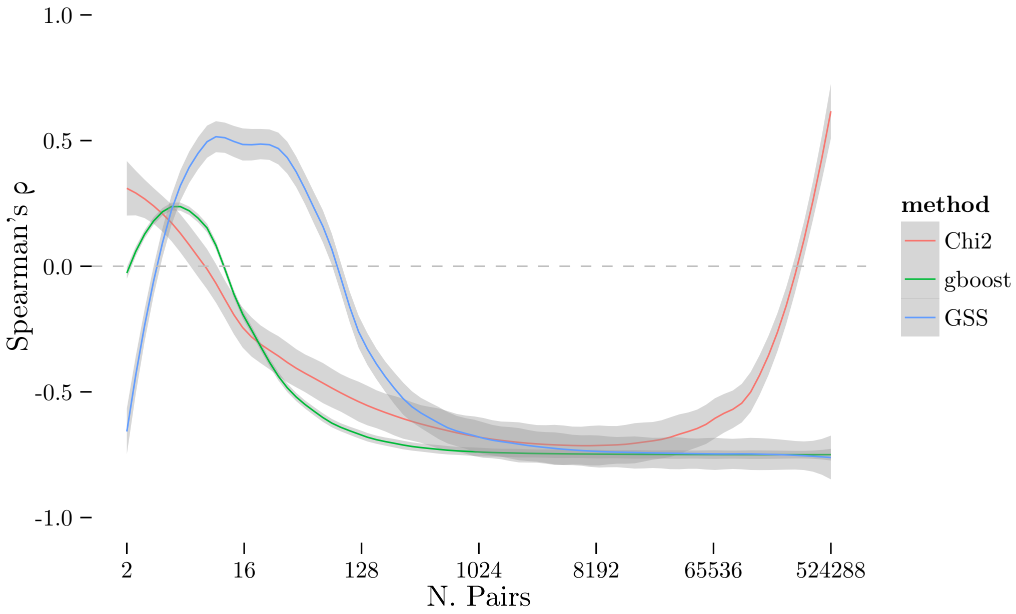

Figure S1. CD

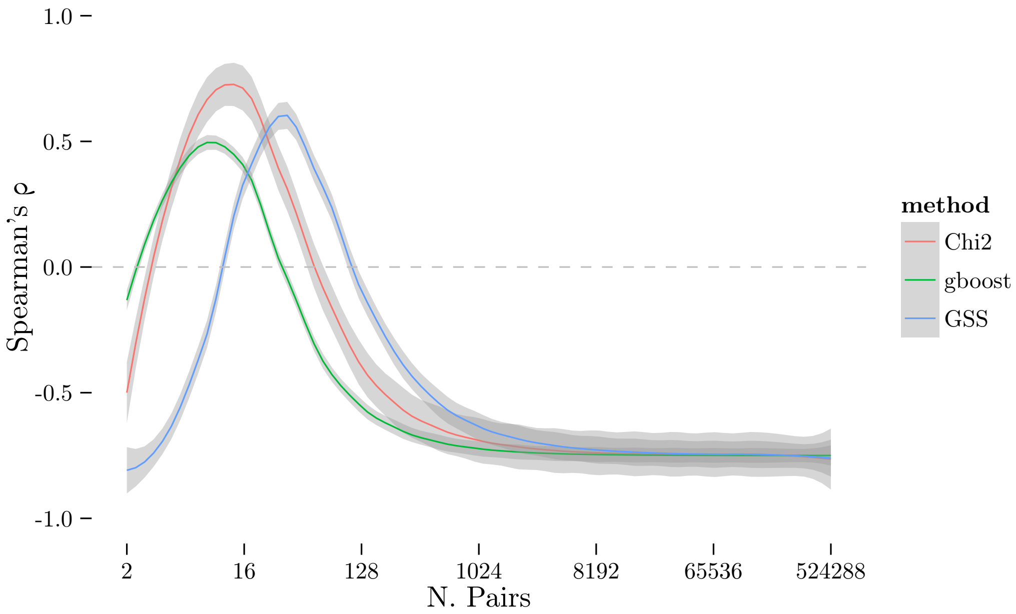

Figure S2. HT

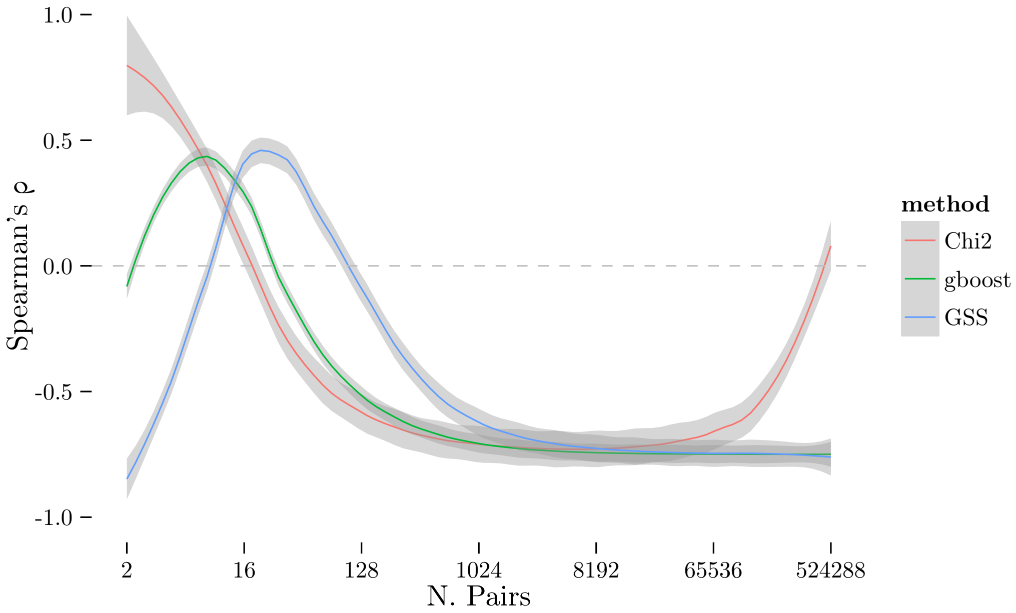

Figure S3. T1D

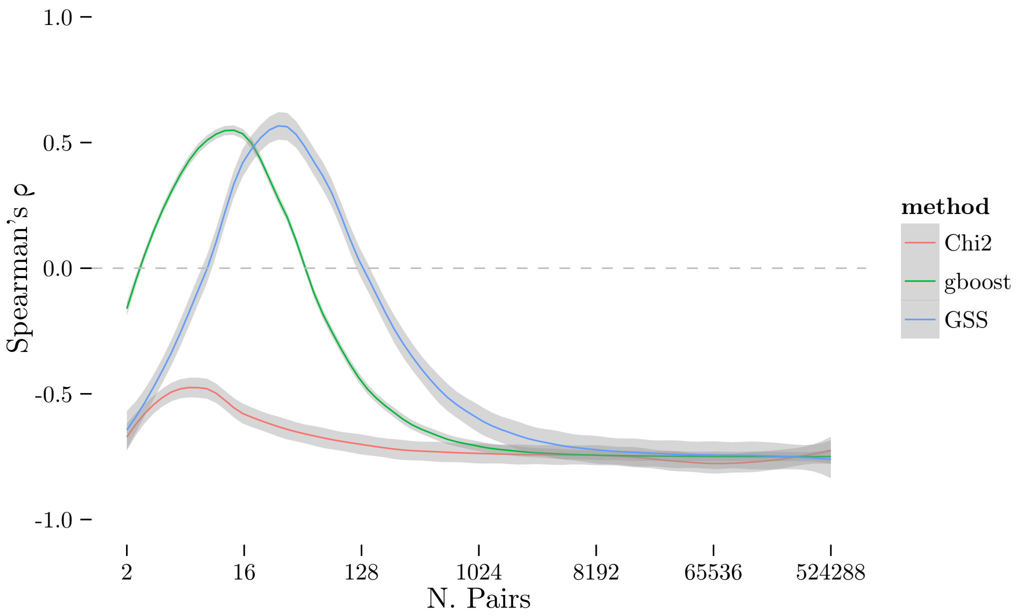

Figure S4. T2D

2 Spearman's ρ – pruned

The results from measuring Spearman's ρ against the whole list using 2-fold cross-validation repeated 10 times on seven WTCCC case–control datasets, with univariately significant SNPs removed.

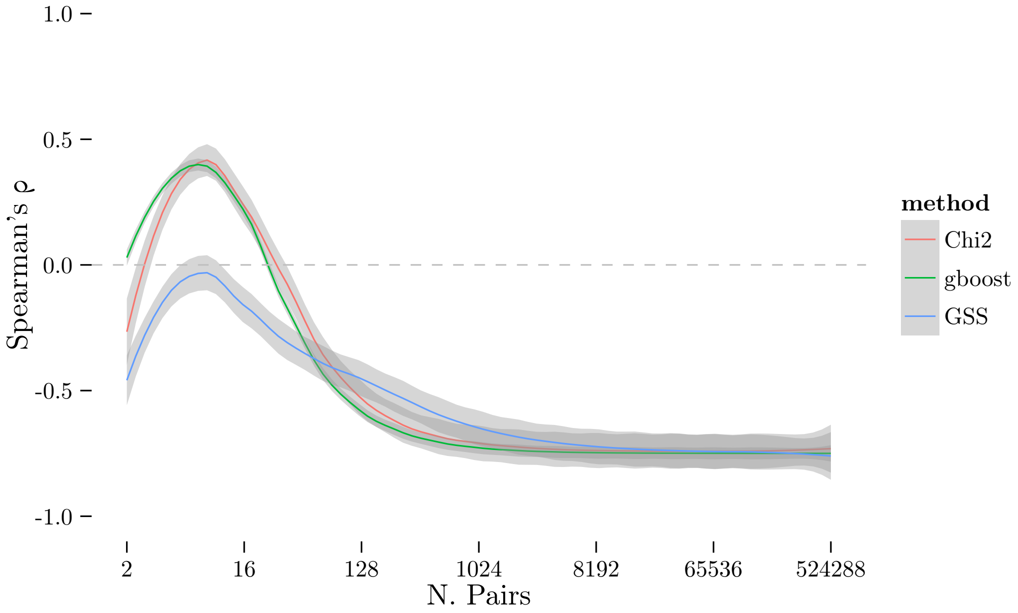

Figure S5. BD

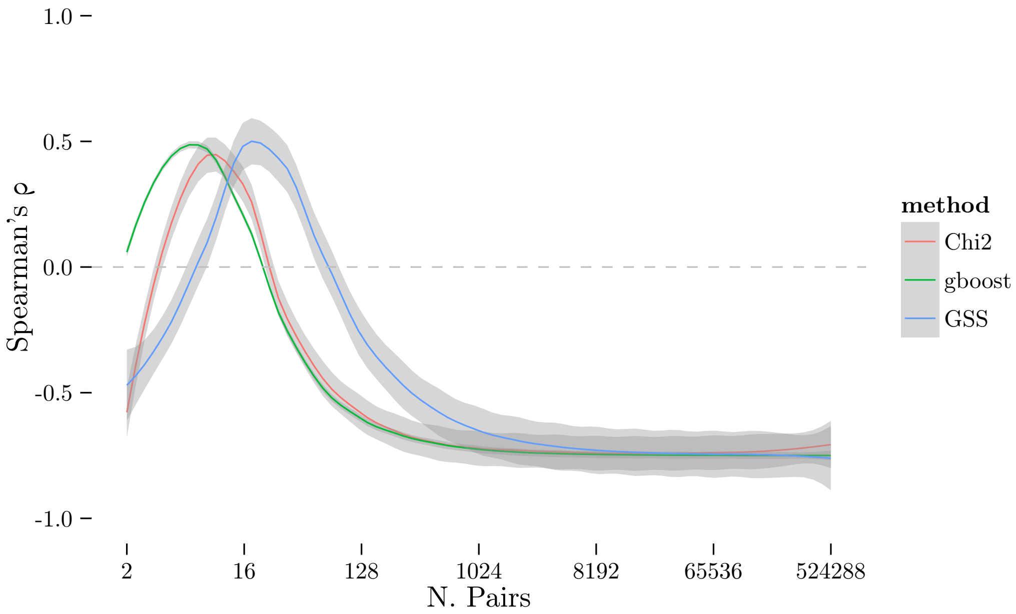

Figure S6. CAD

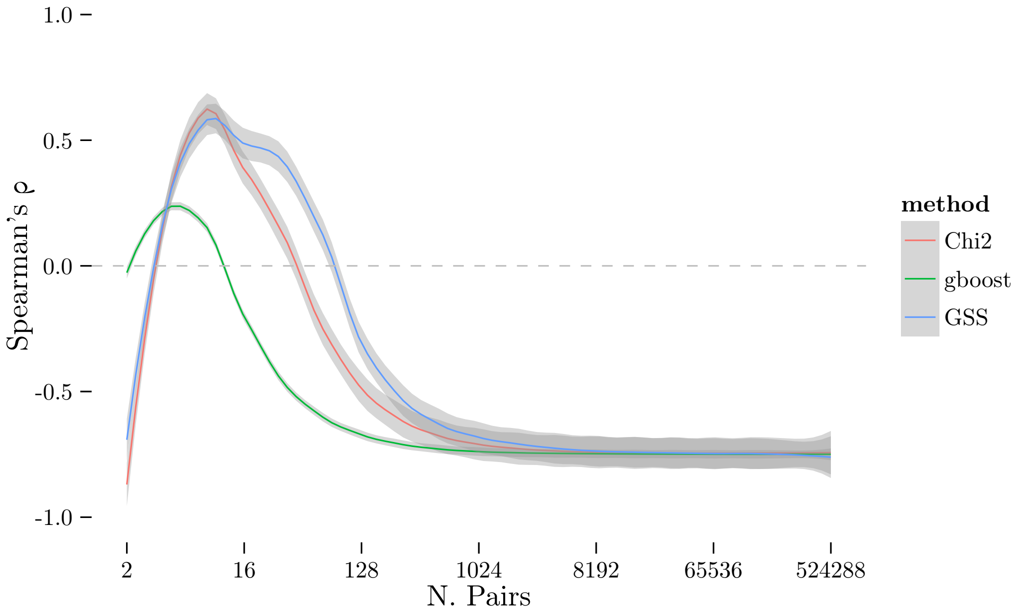

Figure S7. CD

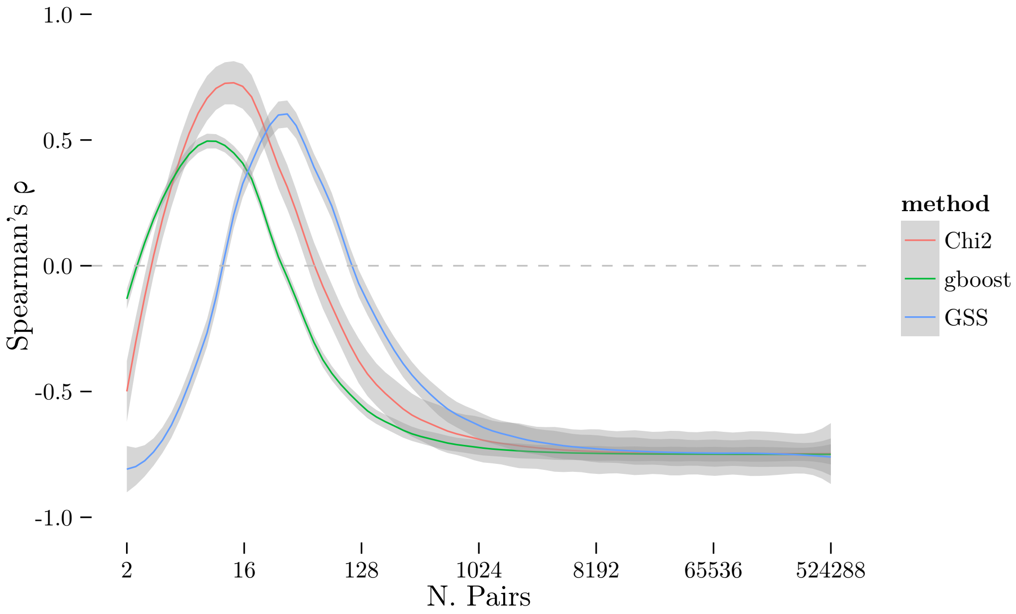

Figure S8. HT

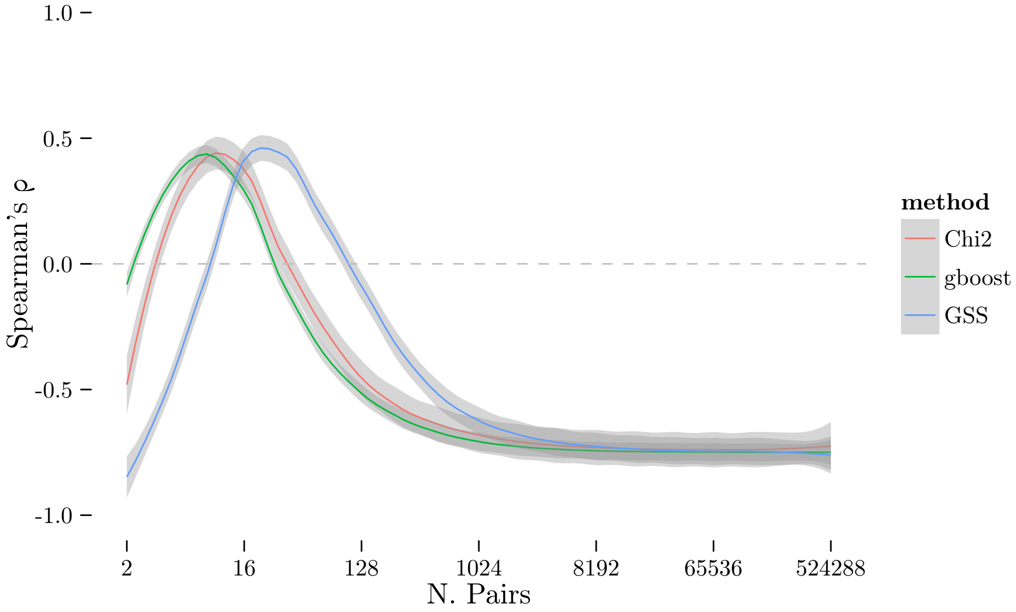

Figure S9. T1D

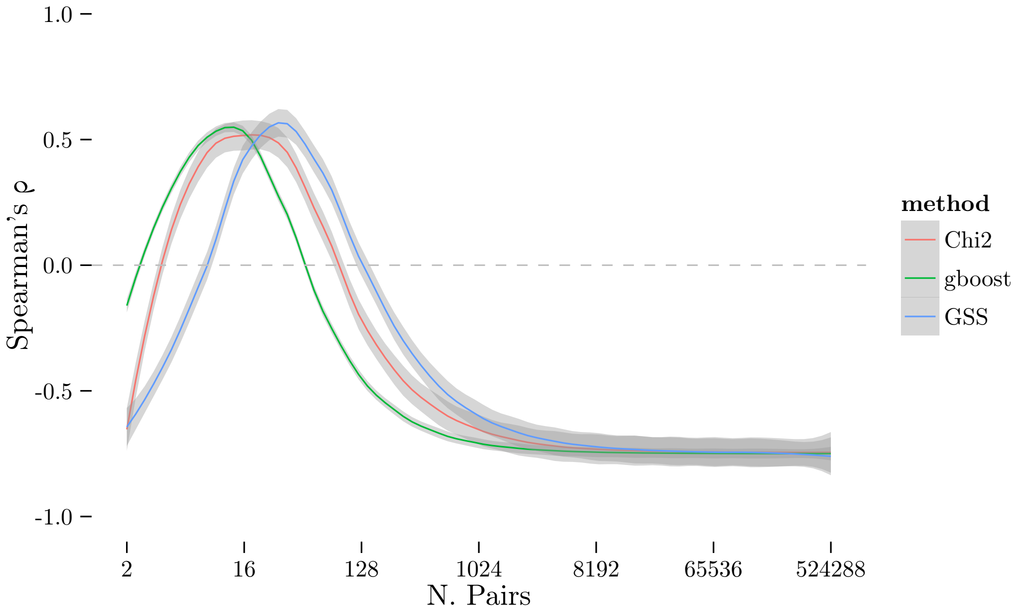

Figure S10. T2D

3 Jaccard

The results from measuring Jaccard against the whole list using 2-fold cross-validation repeated 10 times on seven WTCCC case–control datasets.

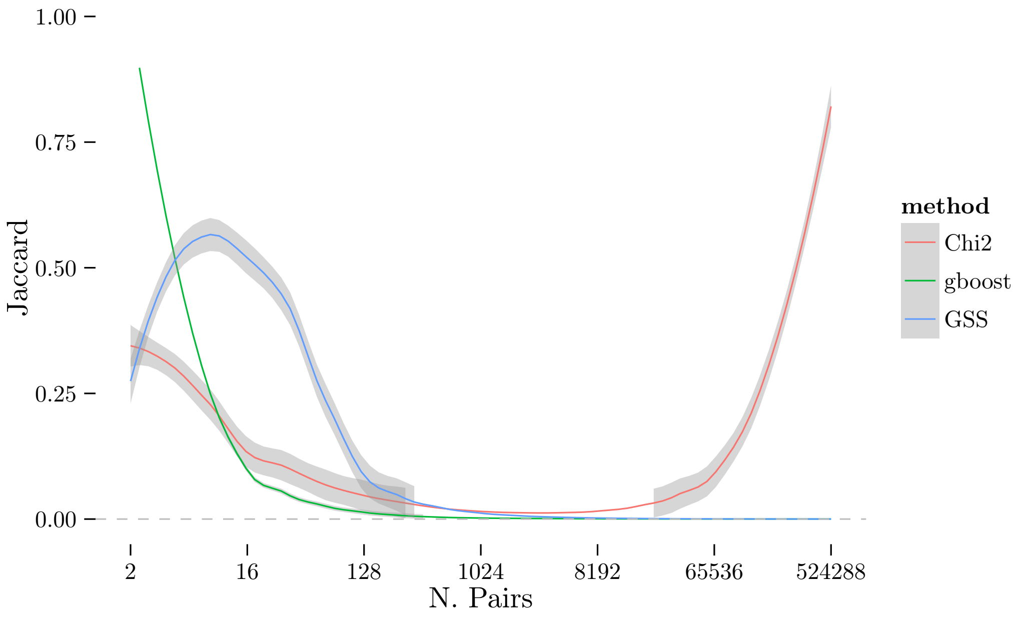

Figure S11. CD

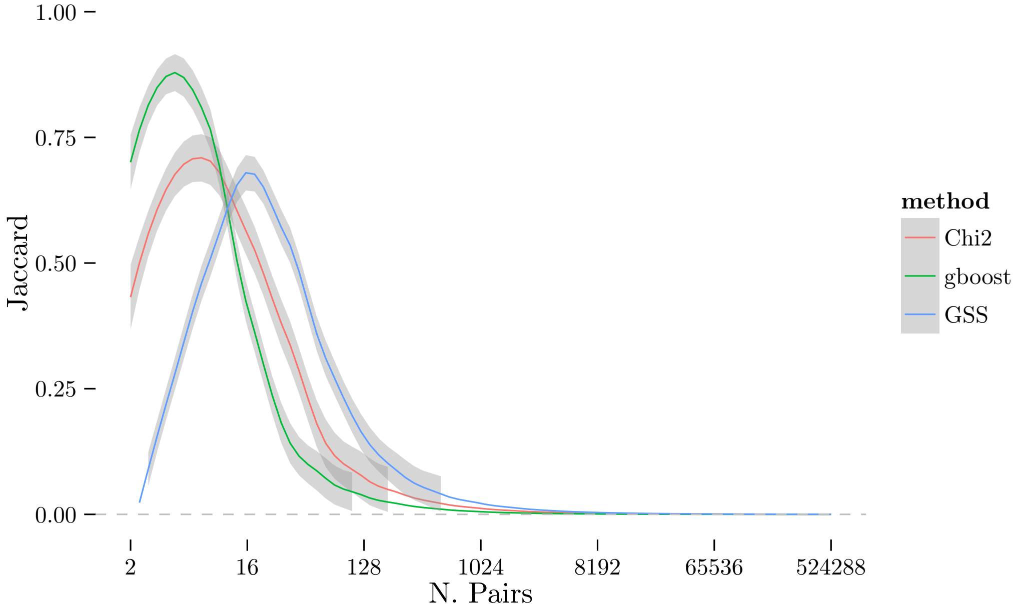

Figure S12. HT

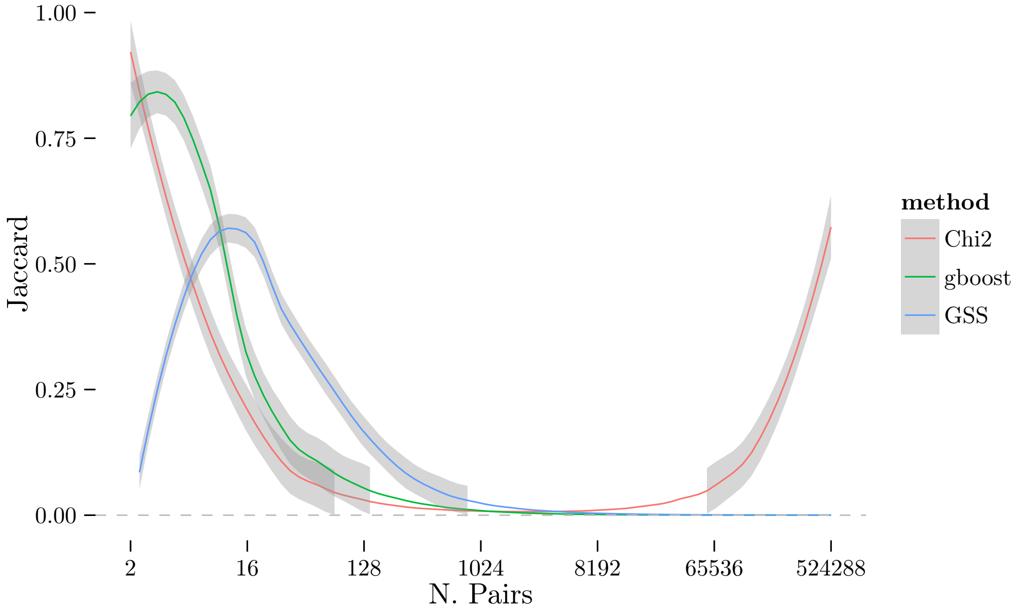

Figure S13. T1D

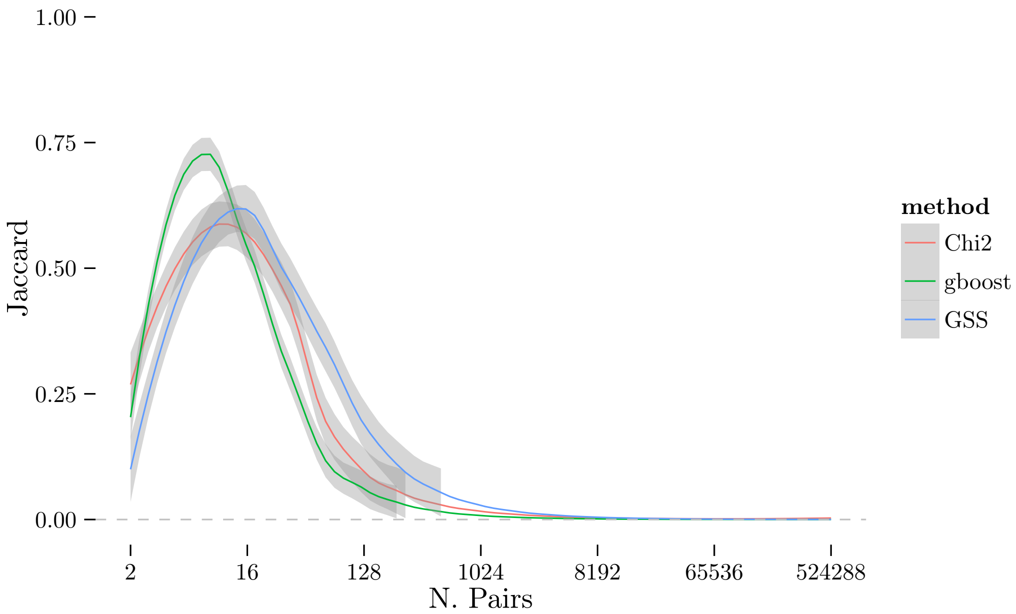

Figure S14. T2D

4 Jaccard – pruned

The results from measuring Jaccard against the whole list using 2-fold cross-validation repeated 10 times on seven WTCCC case–control datasets, with univariately significant SNPs removed.

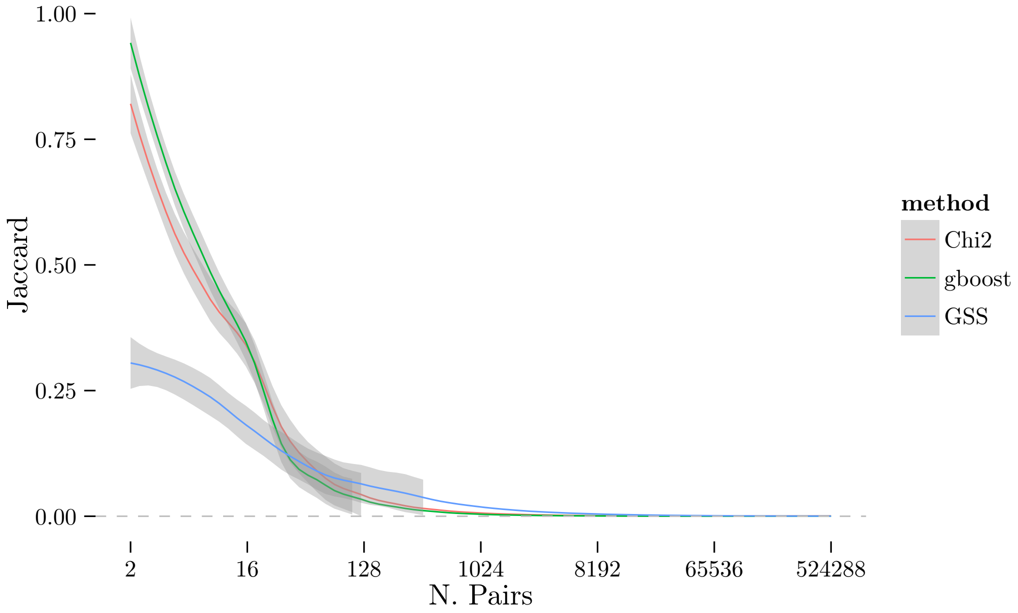

Figure S15. BD

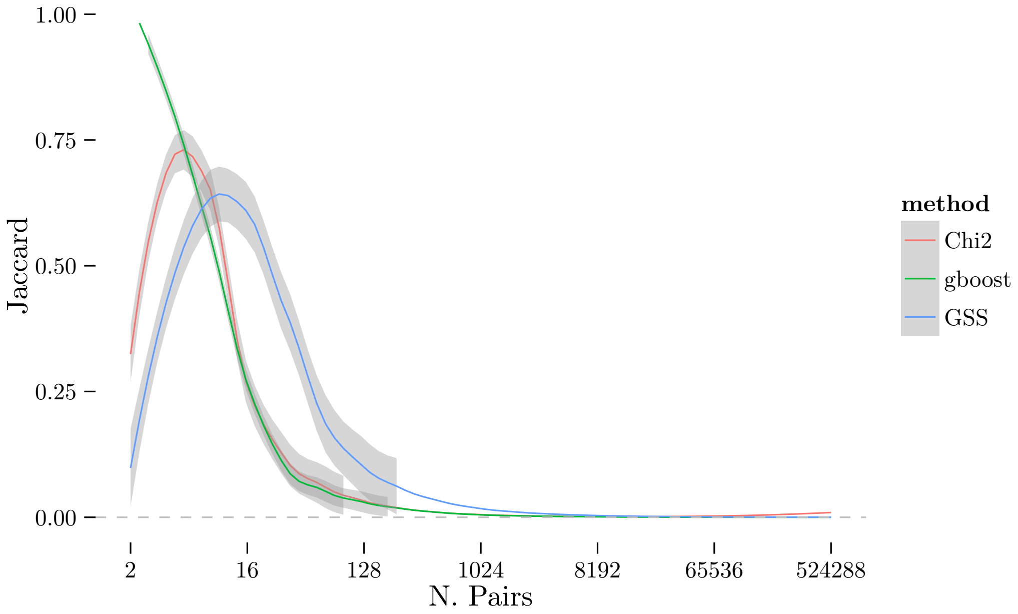

Figure S16. CAD

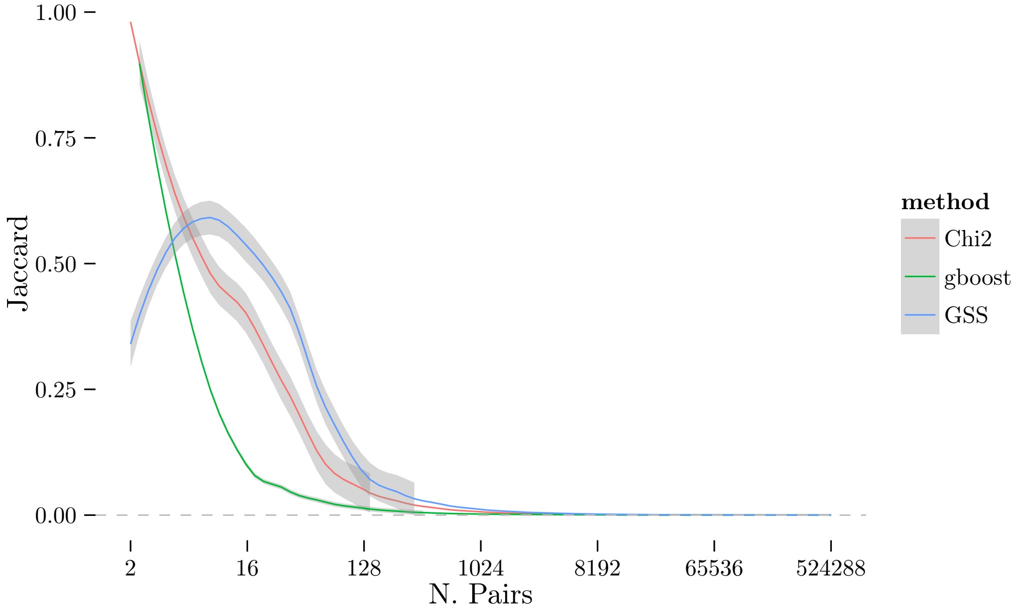

Figure S17. CD

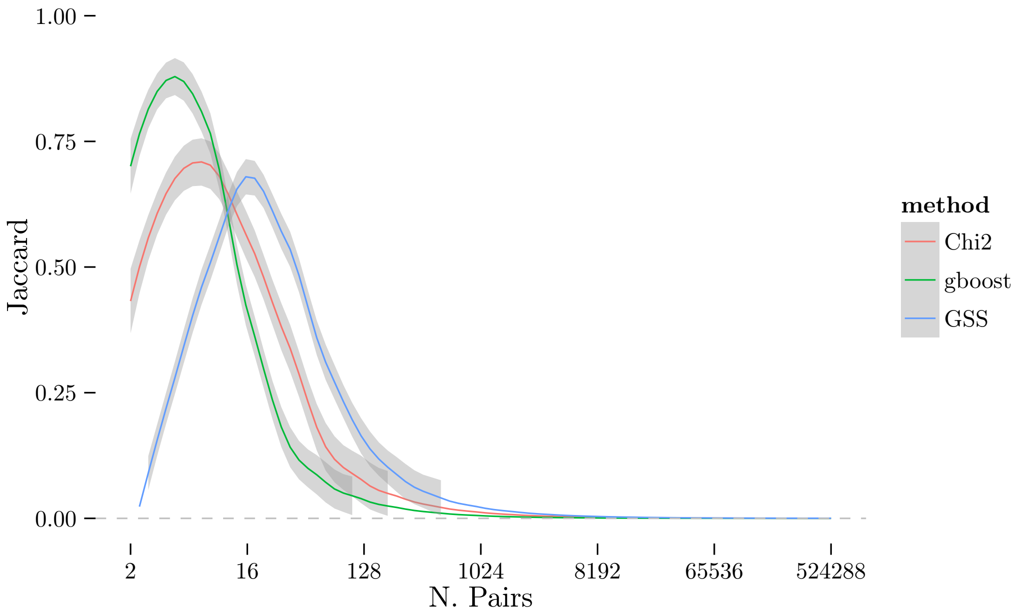

Figure S18. HT

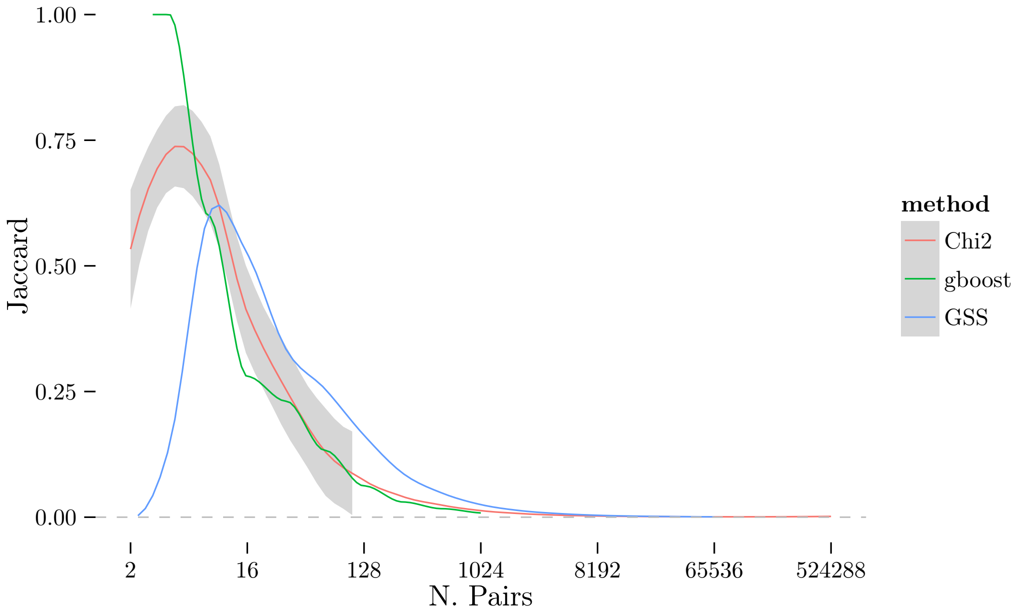

Figure S19. T1D

Figure S20. T2D

5 Overlap between the pairs found by GSS and GBOOST

The overlap between the putative SNP pairs found by GSS and GBOOST is plotted for various values of k. The vertical axis is scaled by the size of the union of both sets, and the blue, green and red sections show the percentage of pairs which are found by GSS only, common to both methods, and found by GBOOST only. The vertical dashed red and blue lines are the ZIC values for GSS and GBOOST respectively.

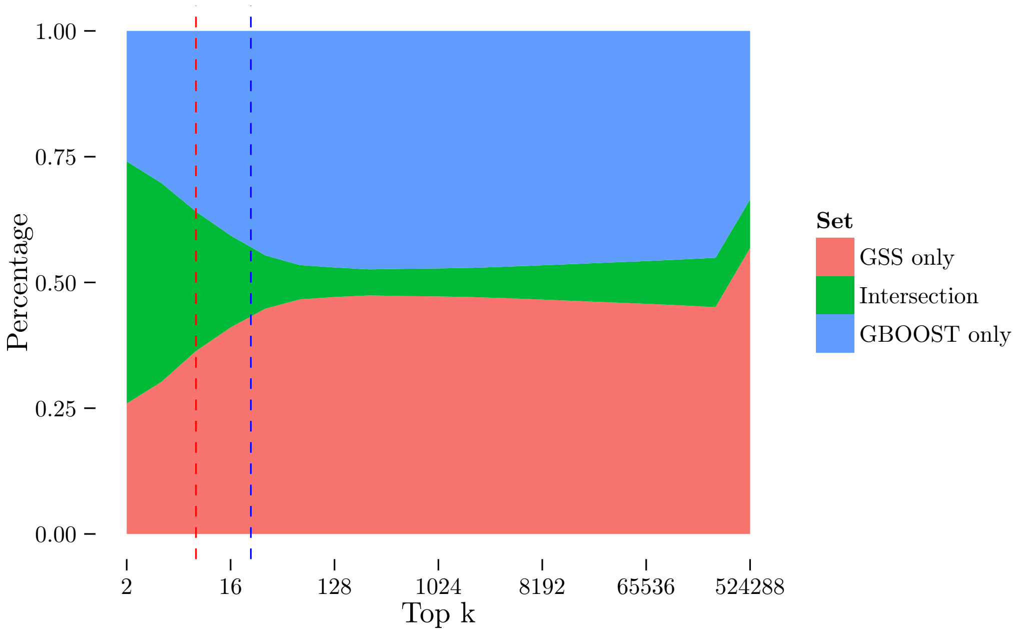

Figure S21. BD

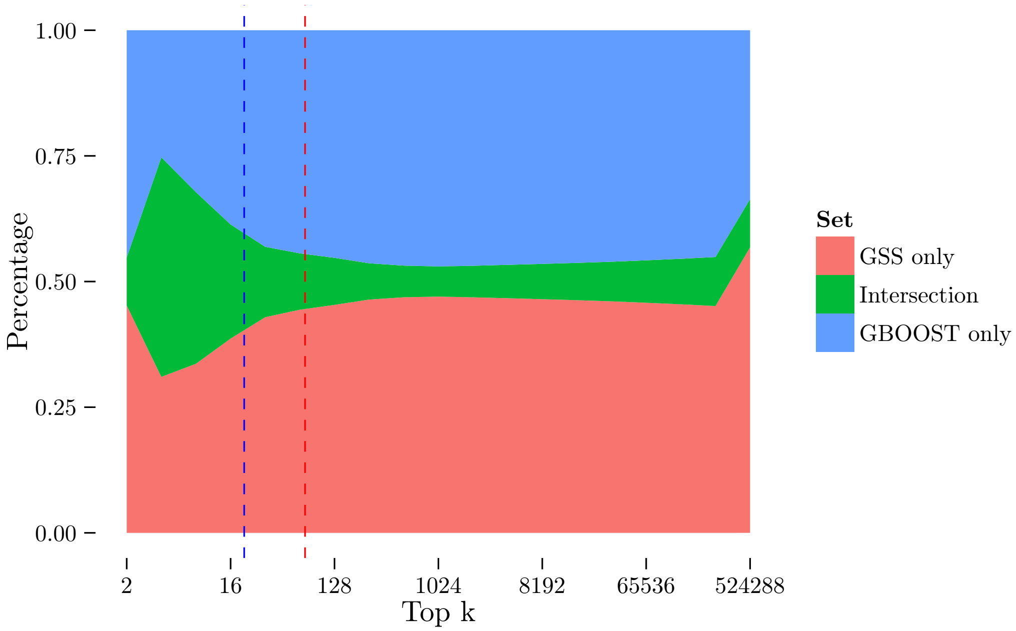

Figure S22. CAD

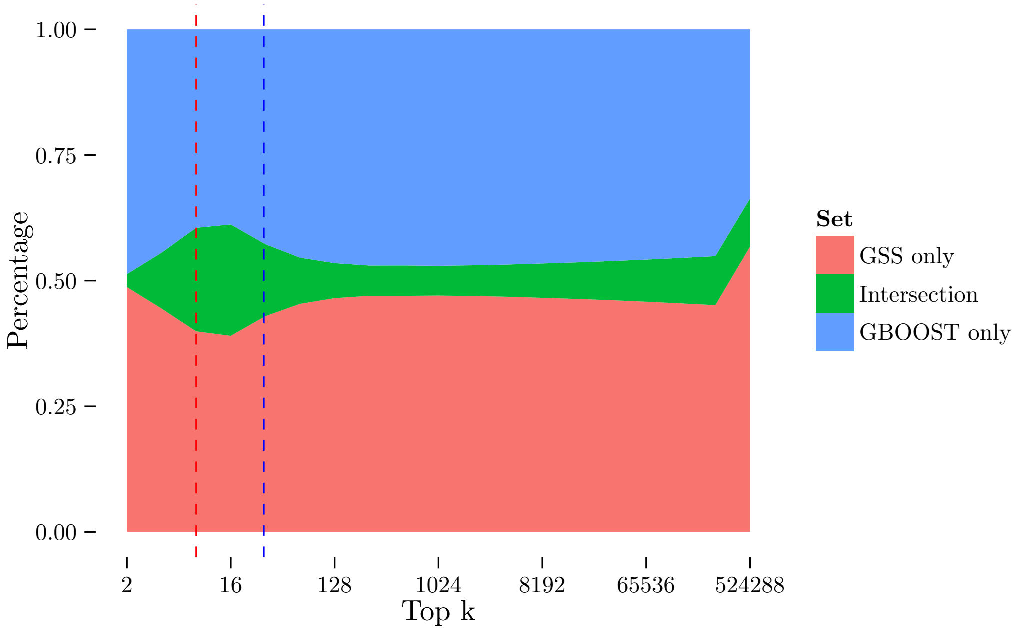

Figure S23. HT

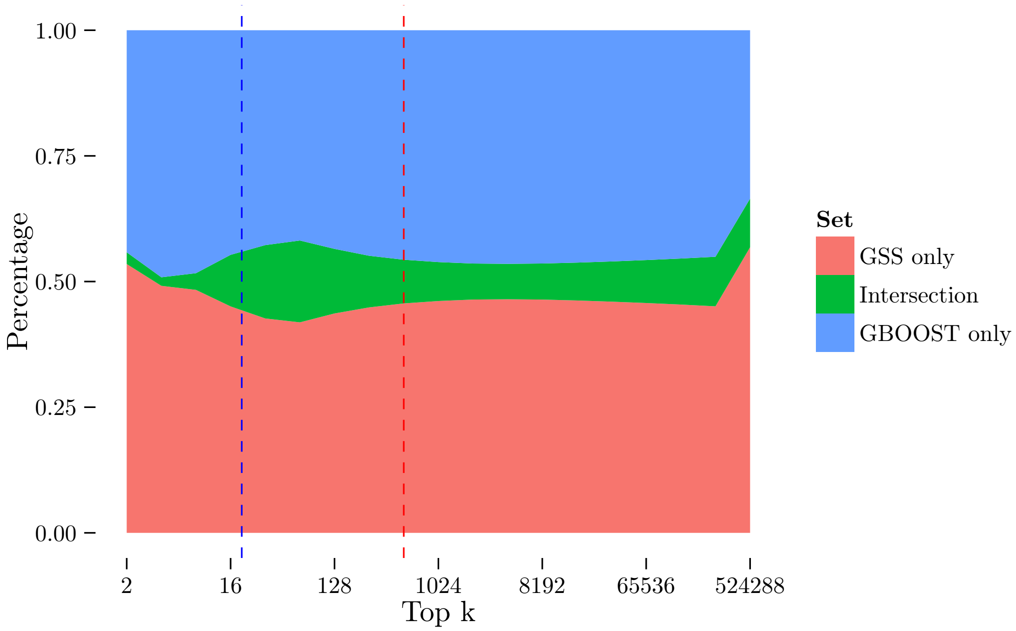

Figure S24. RA

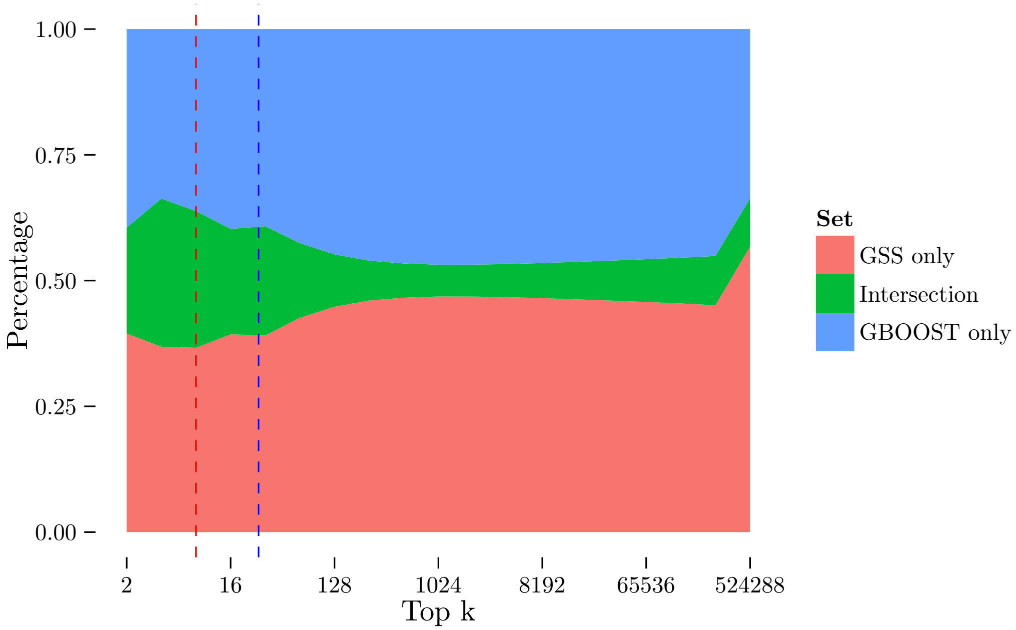

Figure S25. T1D

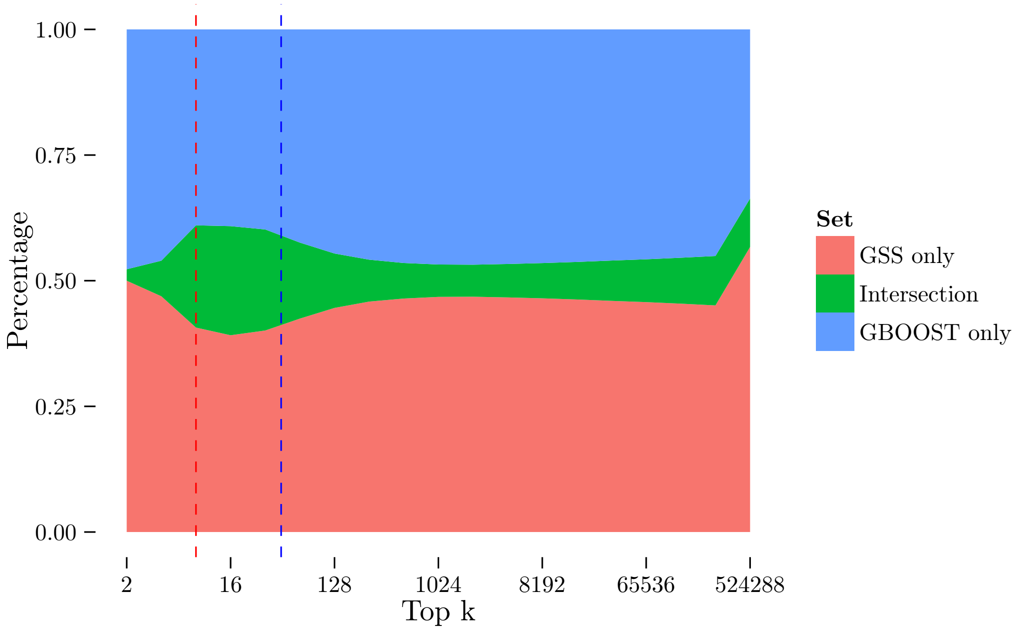

Figure S26. T2D

6 Comparing multiple testing correction with ZIC

Scatter plots for the values in each split, showing ZIC on the vertical axis, and family wise error rate (FWER) and false discovery rate (FDR) threshold values on the horizontal axis. The FWER rank is obtained by the number of pairs above the Bonferonni correction level at 95% level of significance, and the FDR rank is obtained by the number of pairs satifying 5% FDR with the Benjamini–Hochberg procedure. The horizontal green dashed line is at $\tau = 16$, the value where we start searching for ZIC. The blue dashed diagnonal line denotes values where the horizontal and vertical axes are equal.

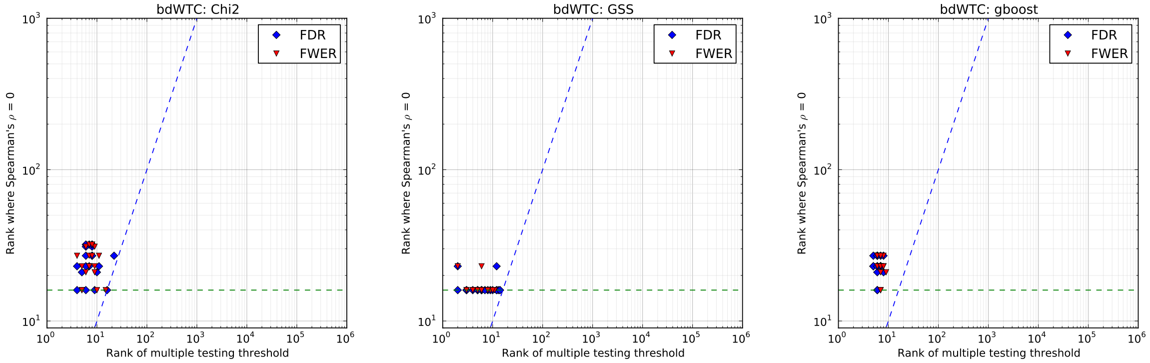

Figure S27. Multiple testing correction vs ZIC: bdWTC

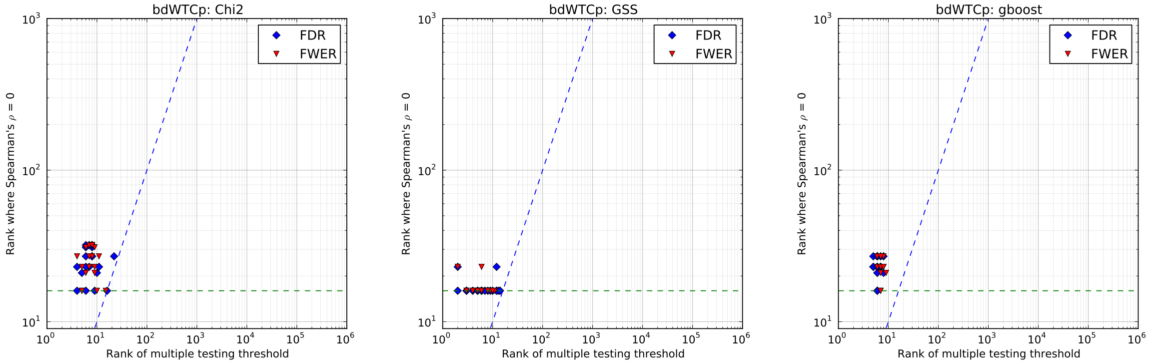

Figure S28. Multiple testing correction vs ZIC: bdWTCp

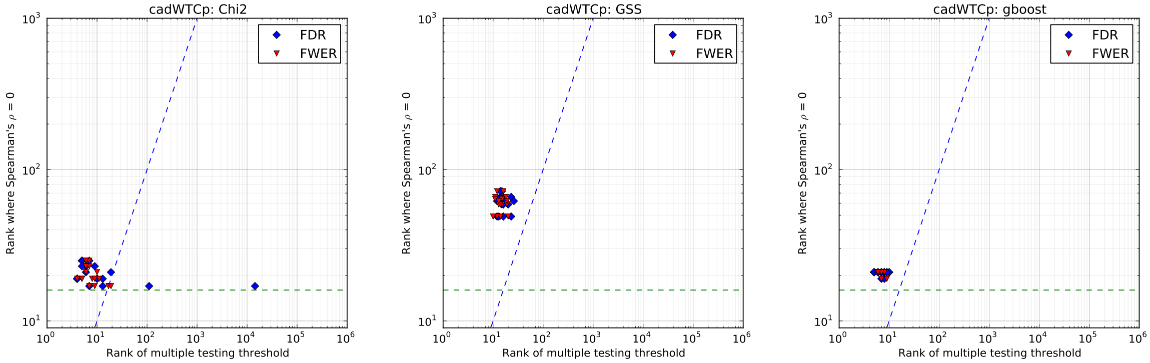

Figure S29. Multiple testing correction vs ZIC: cadWTCp

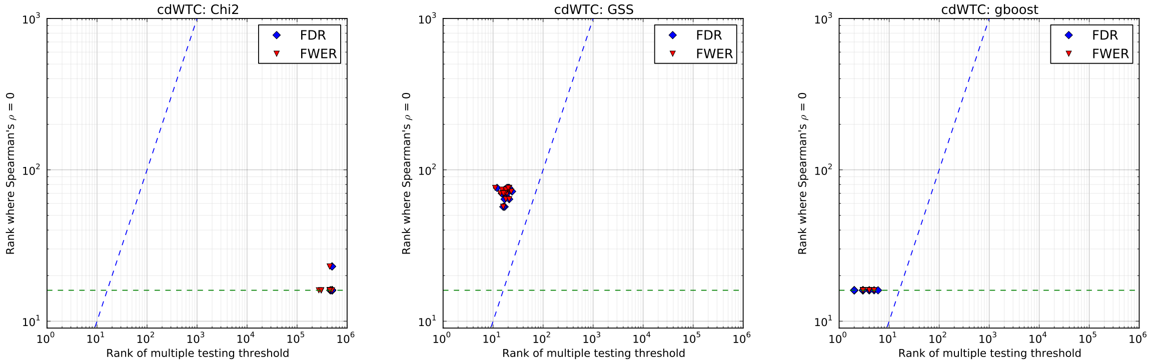

Figure S30. Multiple testing correction vs ZIC: cdWTC

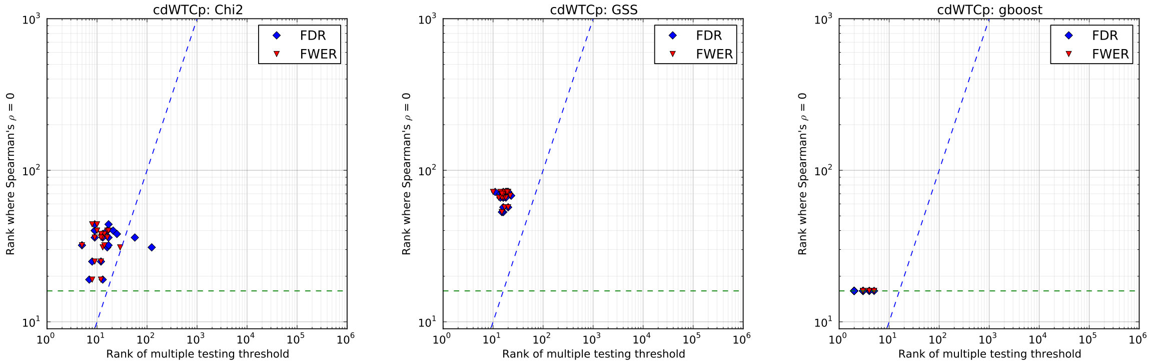

Figure S31. Multiple testing correction vs ZIC: cdWTCp

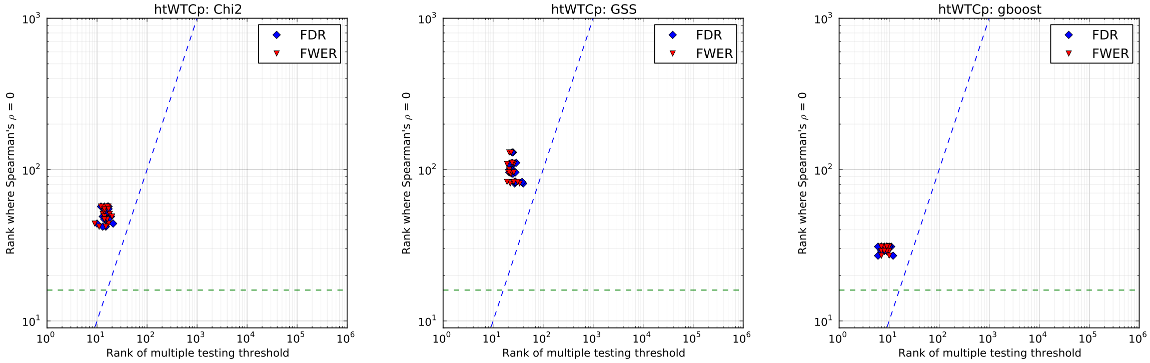

Figure S32. Multiple testing correction vs ZIC: htWTCp

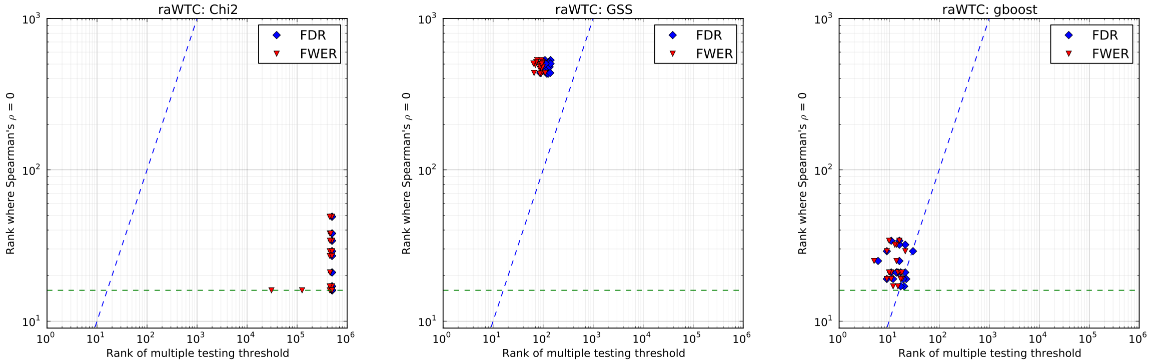

Figure S33. Multiple testing correction vs ZIC: raWTC

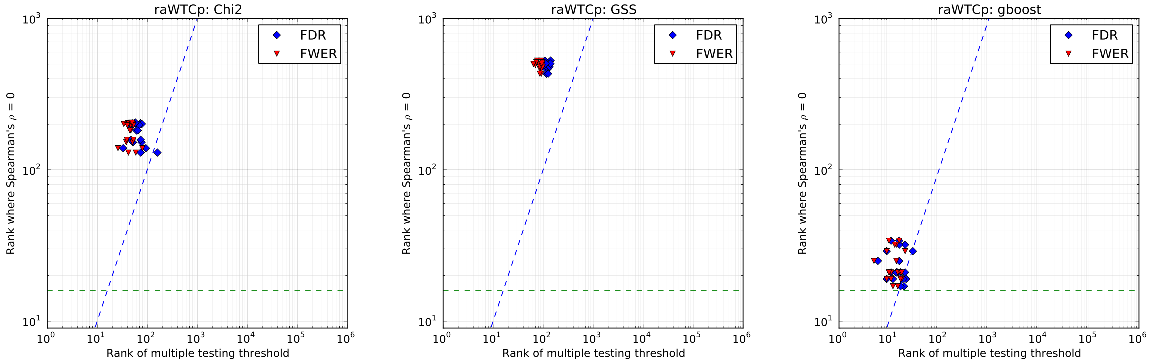

Figure S34. Multiple testing correction vs ZIC: raWTCp

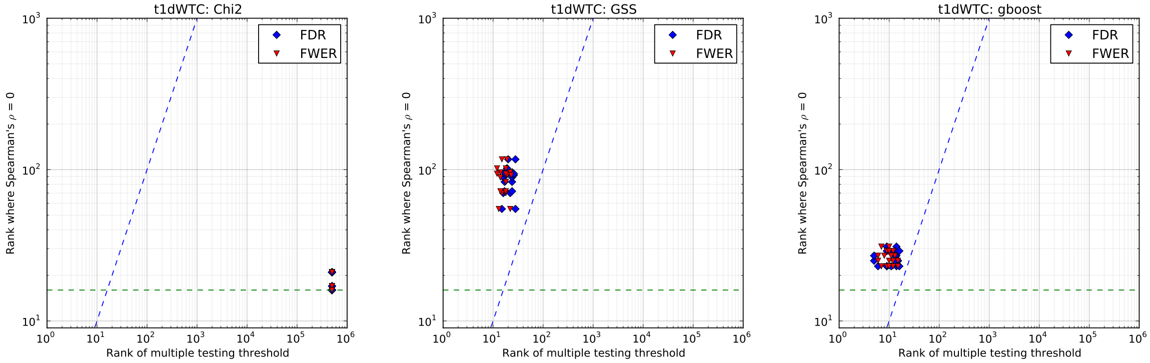

Figure S35. Multiple testing correction vs ZIC: t1dWTC

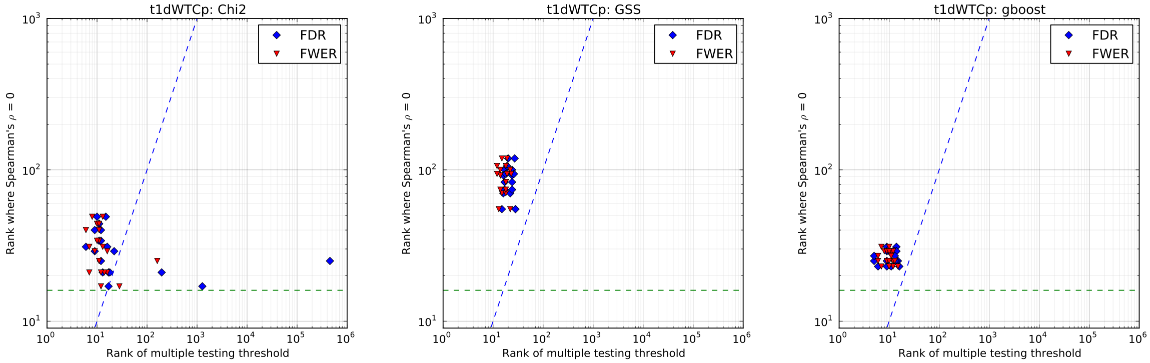

Figure S36. Multiple testing correction vs ZIC: t1dWTCp

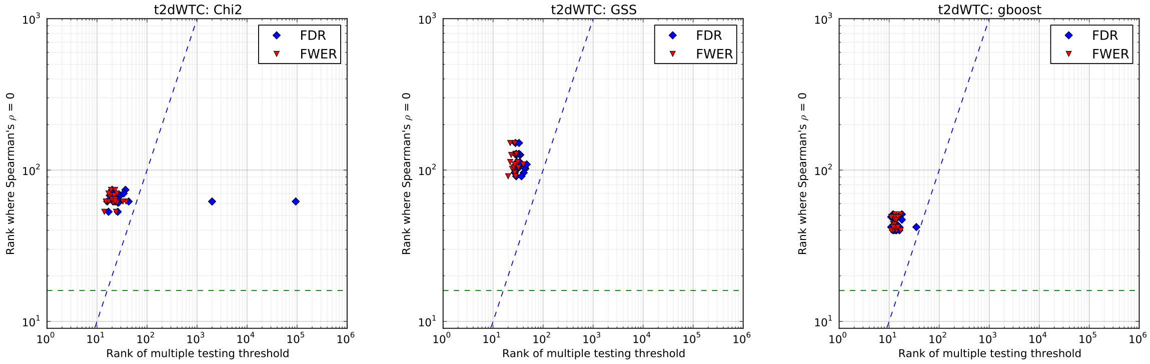

Figure S37. Multiple testing correction vs ZIC: t2dWTC

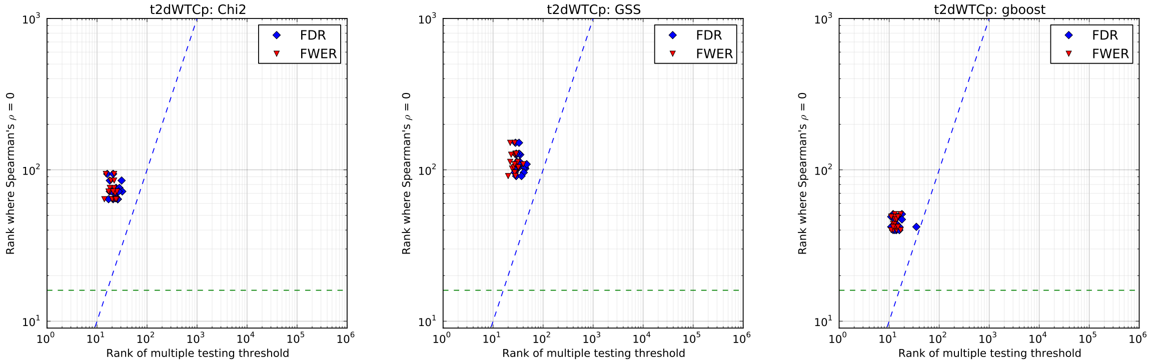

Figure S38. Multiple testing correction vs ZIC: t2dWTCp