Leveraging Family History in Case-Control Analyses of Rare Variation

Claudia R. Solis-Lemus[1*], *S. Taylor Fischer*[2*], *Andrei Todor*[1], *Cuining Liu*[3], *Elizabeth J. Leslie*[1], *David J. Cutler*[1], *Debashis Ghosh*[3], *Michael P. Epstein*[1]

∗ Joint first author
[1] Department of Human Genetics, Emory University, Atlanta, GA
[2] Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA
[3] Department of Biostatistics and Informatics, University of Colorado, Aurora, CO

Supplementary Material

Table 1: Empirical type I error rates for burden, SKAT, and ordinal GAMuT methods from 10,000 null simulations assuming 750 subjects per group, a 10kb region, and rare variants defined as those with $MAF < 3\%$.

		$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$			$\alpha = 0.001$		
λ	Prevalence	Burden	SKAT	GAMuT	Burden	SKAT	GAMuT	Burden	SKAT	GAMuT	Burden	SKAT	GAMuT
2	0.01	0.1011	0.1126	0.1079	0.0500	0.0575	0.0556	0.0103	0.0134	0.0130	0.0006	0.0020	0.0015
	0.05	0.1084	0.1014	0.1089	0.0534	0.0530	0.0581	0.0107	0.0111	0.0111	0.0009	0.0017	0.0015
4	0.01	0.1038	0.1161	0.1078	0.0510	0.0615	0.0568	0.0109	0.0136	0.0135	0.0008	0.0017	0.0014
	0.05	0.1061	0.1092	0.1083	0.0553	0.0536	0.0542	0.0115	0.0119	0.0100	0.0012	0.0015	0.0015

Abbreviations: λ, conditional recurrence risk ratio; α, significance threshold; MAF, minor allele frequency

Table 2: Average computing time for one test from 1,000 null simulations assuming $\lambda = 8$ and target disease prevalence 0.05.

Region Size (kb)	Sample Size (Per Group)	Ordinal GAMuT Time (sec, mean [SD])	SKAT Time (sec, mean [SD])
10	750	0.40 (0.08)	0.10 (0.02)
	1000	0.56 (0.13)	0.15 (0.03)
	1500	1.12 (0.32)	0.28 (0.05)
	2500	3.15 (0.8)	0.52 (0.08)
50	750	1.30 (0.22)	0.59 (0.1)
	1000	2.23 (0.41)	0.90 (0.16)
	1500	5.10 (0.96)	1.83 (0.37)
	2500	13.27 (2.64)	3.51 (0.59)

Abbreviations: λ, conditional recurrence risk ratio

Enrichment of Causal Variants

In Figure 1, we show that, as expected, the average number of causal rare variants is greater for the cases with family history, followed by cases without family history, and lastly for controls. This simulated dataset comprises of 1000 controls, 1000 cases without family history, and 1000 cases with family history for three levels of conditional recurrence risk ratios (columns: $\lambda = 2, 4, 8$) and 2 siblings as family history. The effect size was set as $C = 2$.

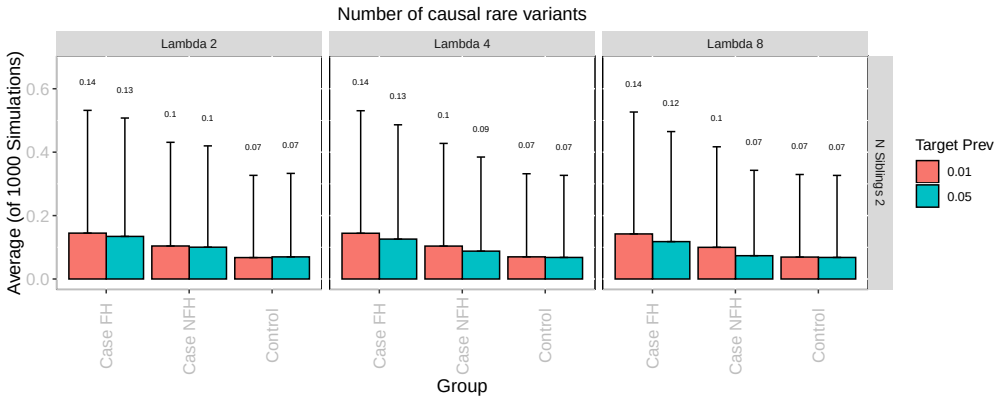

Figure 1: Average of 1000 simulations of number of causal rare variants (left) and probability of disease (right) in proband for three groups: controls, cases without family history, and cases with family history under two disease prevalences (red=0.01, blue=0.05), with one (top) or two (bottom) siblings, and three conditional recurrence risk ratios as columns.

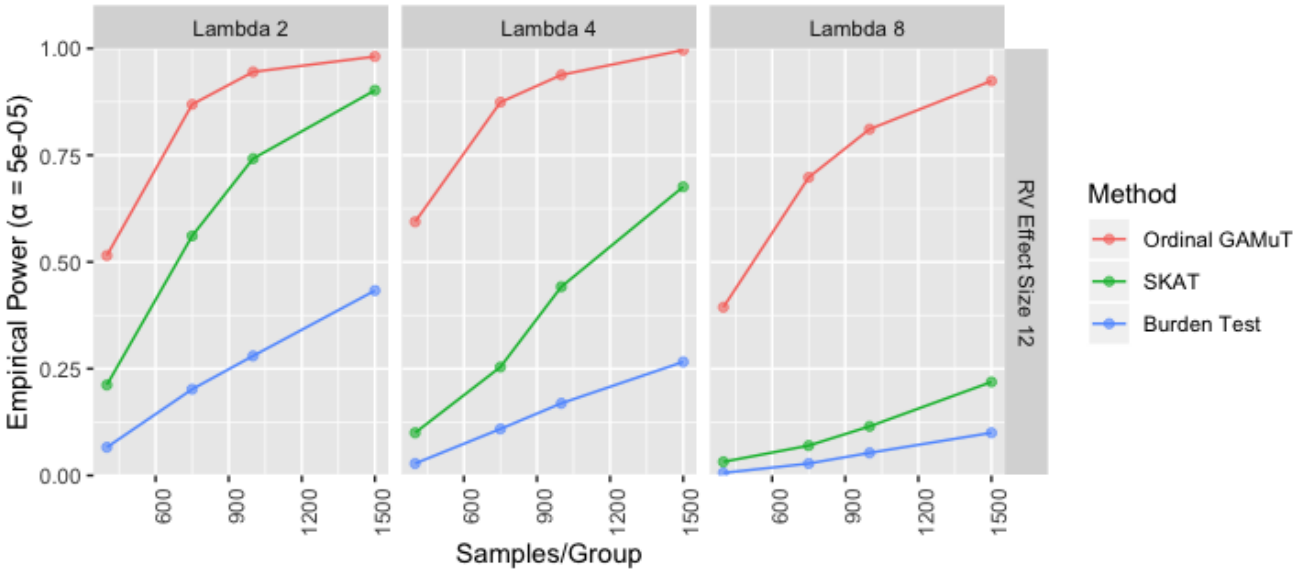

Figure 2: Power across 1000 simulations in which there were 4 causal rare variants per region with MAF between 0.01 and 0.001 with a large effect size ($C = 12$) at a target disease prevalence of 5%.

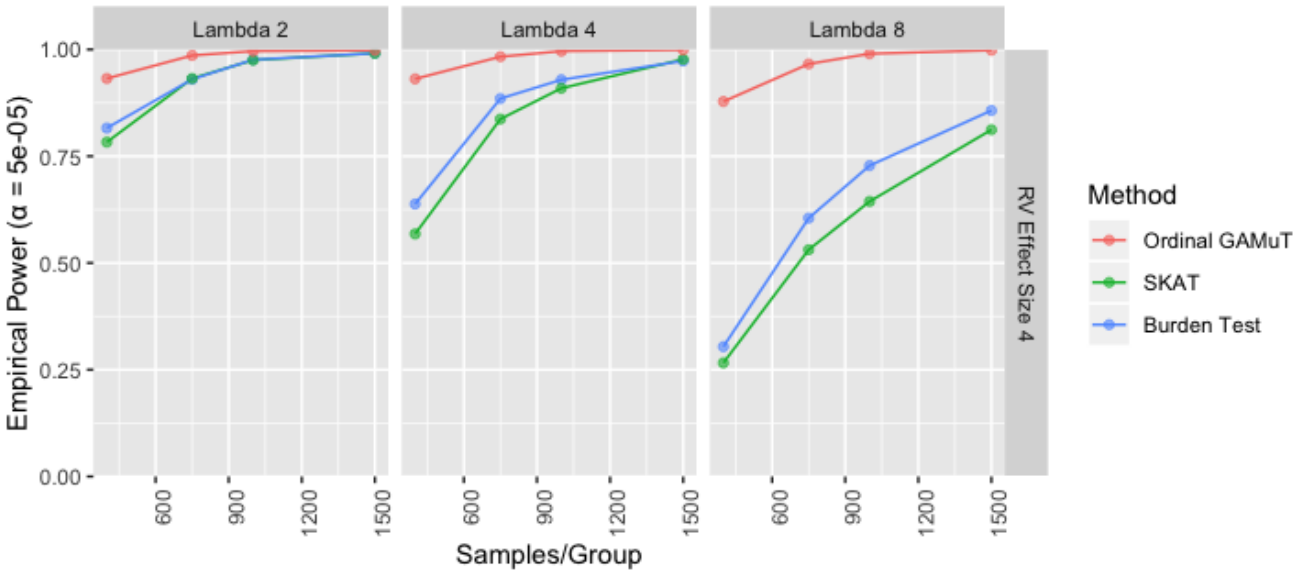

Figure 3: Power across 1000 simulations using a 50kb region, effect size C=4, and a target disease prevalence of 0.05.

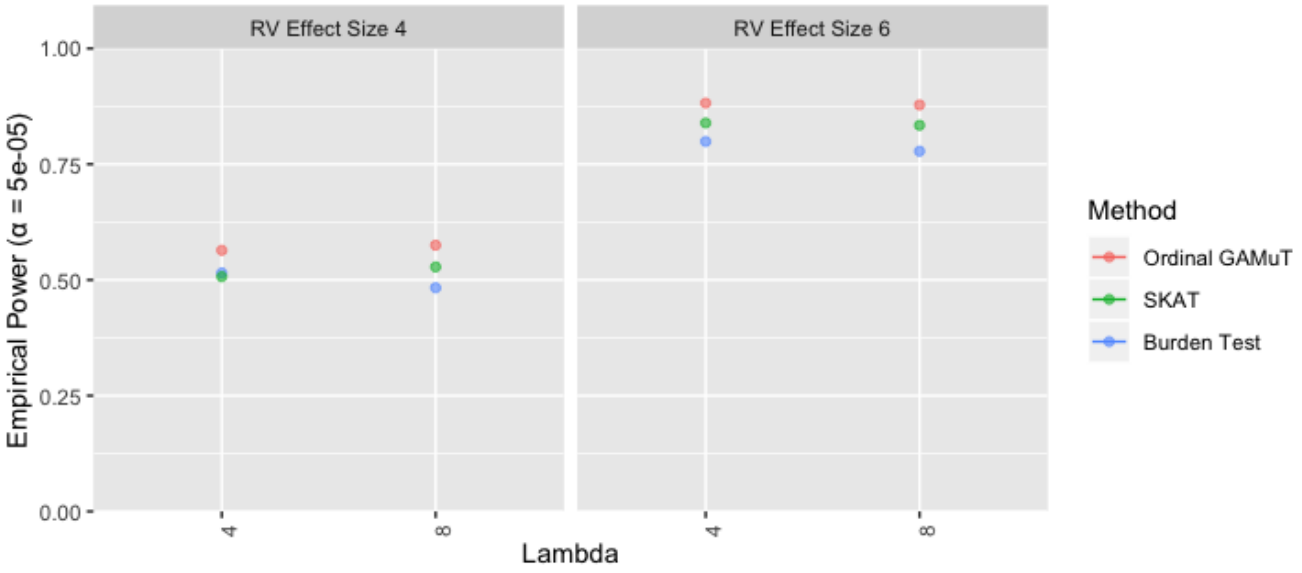

Figure 4: Power across 1000 simulations at a reduced target disease prevalence of 0.001 with causal rare variants of moderate effect size ($C = 4, 6$) and $\lambda = 4, 8$ assuming 750 subjects per group.