SNP Set Association Analysis for Familial Data

Supporting Information

Running Title: SNP Set Analysis for Familial Data

Elizabeth D. Schifano[1*], Michael P. Epstein[2], Lawrence F. Bielak[3], Min A. Jhun[3], Sharon L.R. Kardia[3], Patricia A. Peyser[3], Xihong Lin[1]

[1]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA

[2]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

[3]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA

[*]Address for Correspondence: Elizabeth D. Schifano, Ph.D.

Department of Biostatistics, Harvard School of Public Health

655 Huntington Avenue, Boston, MA 02115

eschifan@hsph.harvard.edu

617 432 6088

Table 1: Empirical size for different kernels, polygenic effects for $n = 300$ sib trios and their parents ($n_i = 5$), using Satterthwaite (top) and Davies (bottom) methods to approximate the null distribution.

Results using Satterthwaite Method

Kernel	LIN			wLIN			IBS			wIBS		
v_{0b}	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
FGFR2	0.047	0.048	0.046	0.049	0.048	0.052	0.045	0.049	0.050	0.048	0.046	0.050
ASAH1	0.048	0.049	0.051	0.045	0.049	0.053	0.048	0.049	0.051	0.050	0.050	0.052

Results using Davies Method

Kernel	LIN			wLIN			IBS			wIBS		
v_{0b}	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
FGFR2	0.048	0.049	0.046	0.048	0.046	0.051	0.044	0.047	0.047	0.046	0.049	0.048
ASAH1	0.048	0.048	0.051	0.046	0.050	0.053	0.047	0.049	0.051	0.049	0.048	0.051

$*$ $v_{0b} = \sigma_b^2[\sigma_b^2 + \sigma_e^2]^{-1}$ is the heritability due to polygenic effects for within-family correlation.

Figure 1: Top: Power to detect causal SNP for FGFR2 using additive (left) and dominant (right) genetic models under MED heritability due to SNP (max(h^2)=1%) for $n = 300$ sib trios and their parents ($n_i = 5$). SNPs are ordered according to genomic location. Lines in shades of gray and blue correspond to KM-based methods, wheres different line types and widths differentiate between types of kernels; the red solid line corresponds to the multiple testing adjusted individual-SNP based approach. The typed SNPs, indicated by an 'x' along the bottom of the plot, compose the SNP set. Middle: Corresponding MAF for SNPs plotted above. Bottom: Corresponding LD plot for SNPs plotted above (grayscale for squared correlation R^2: white - $R^2 = 0$, black - $R^2 = 1$). Due to the increased effective sample size, power is higher for all testing methods in this simulation as compared to the sib-trio only simulation. However, the trends in power (i.e., the relative ordering of the power curves) remain the same.

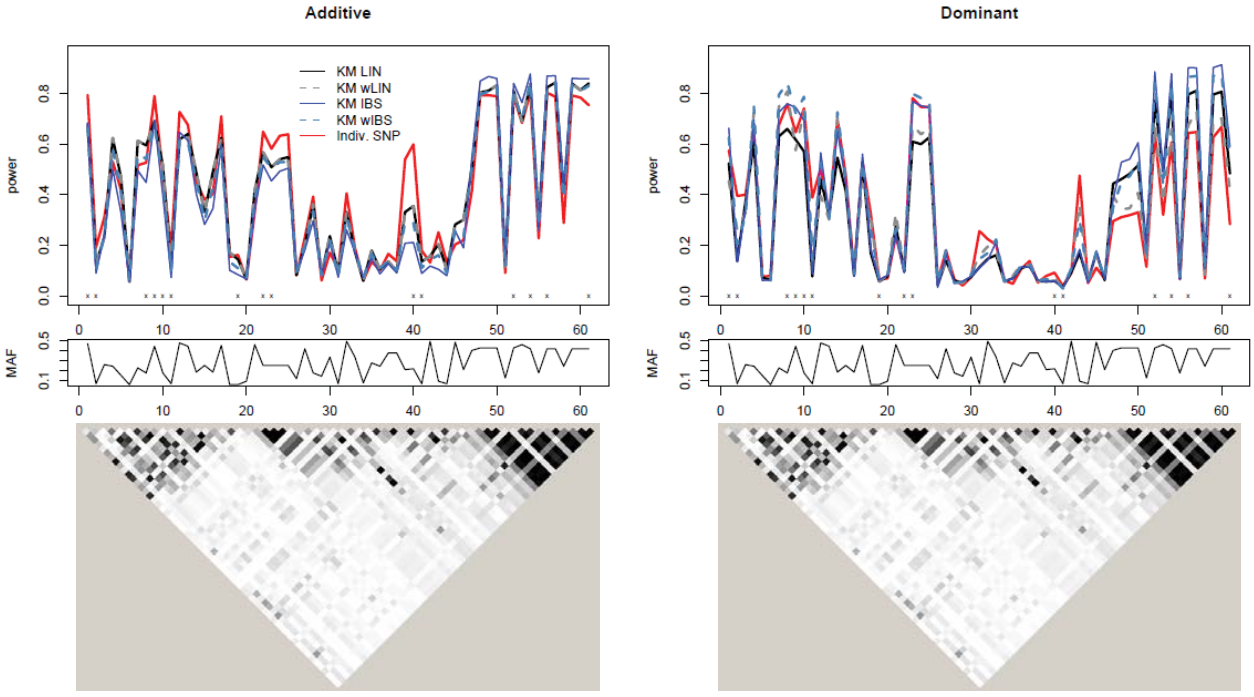

Figure 2: Power to detect causal SNP for ASAH1 using additive (left) and dominant (right) genetic models under MED heritability due to SNP (max(h^2)=1%) for n = 300 sib trios and their parents (n_i = 5). Legend is the same as Supplemental Figure 1.

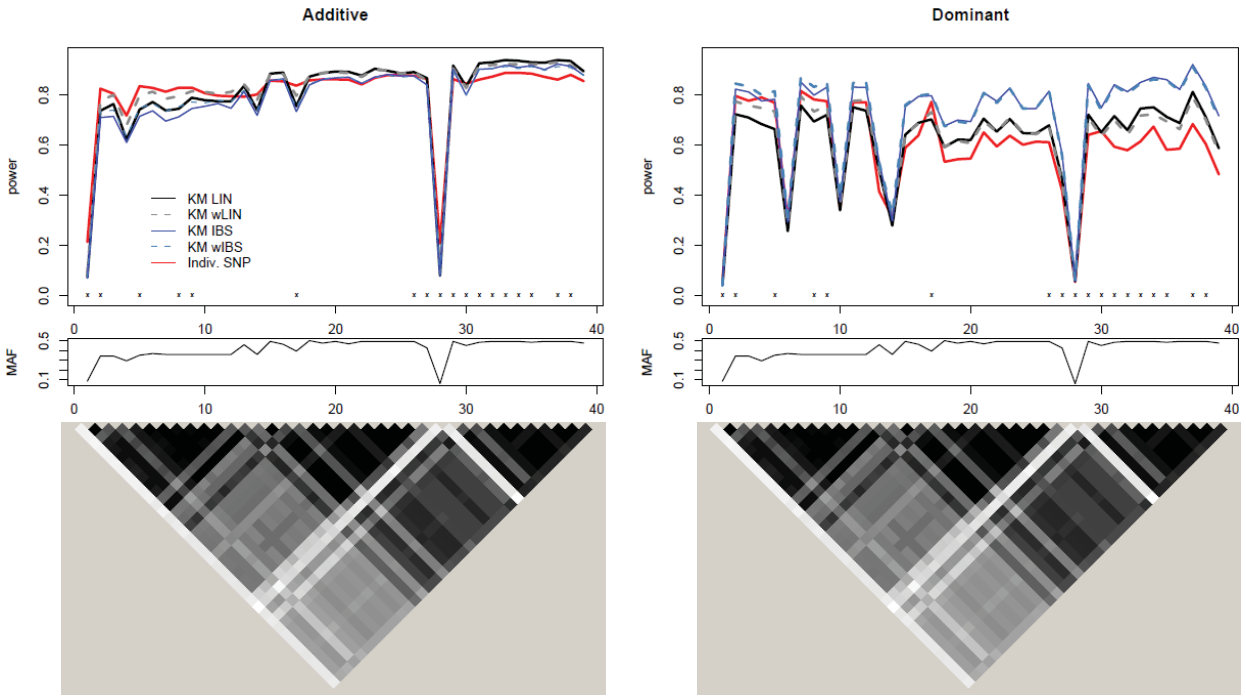

Figure 3: Top: Power to detect causal SNP for ASAH1 using additive (left) and dominant (right) genetic models under MED heritability due to SNP (max(h^2)=1%) for n = 300 sib trios and their parents (n_i = 5). SNPs are ordered according to median R^2. Legend is the same as Supplemental Figure 1. Bottom: Corresponding MAF for SNPs plotted above.

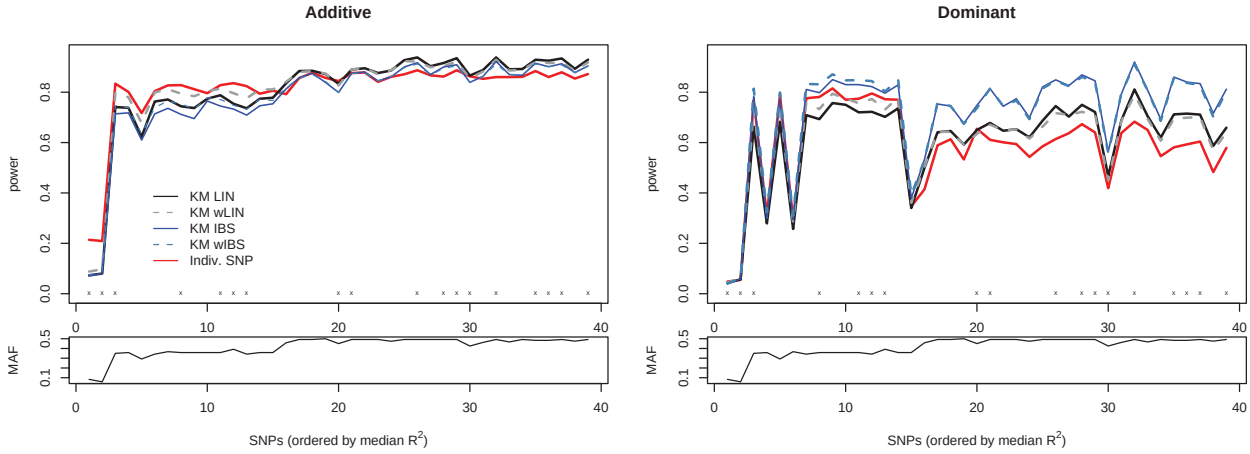

Figure 4: Effect of SNP Set Size: Analogous simulations were performed to those described in Section 3.1, but including *all* HapMap SNPs in the SNP set as opposed to just including the *typed* SNPs in the SNP set. Row 1: Power to detect causal SNP for FGFR2 (left) and ASAH1 (right) using additive genetic models under MED heritability due to SNP (max(h^2)=1%) for $n = 300$ sib trios. SNPs are ordered according to genomic location. Dotted lines in shades of gray correspond to linear KM-based methods and solid lines in shades of red correspond to multiple testing adjusted individual-SNP based approach; shades of color differentiate between SNP sets with only typed SNPs (lighter: 'Typed') in the set and all HapMap SNPs in the region (darker: typed and non-typed; 'All'). The typed SNPs are indicated by an 'x' along the bottom of the plot. Row 2: Corresponding difference in power comparing 'All' SNP set with 'Typed' SNP set for KM (black) and multiple testing adjusted individual-SNP based approach (red). Row 3: Corresponding MAF for SNPs plotted above. Row 4: Corresponding LD plot for SNPs plotted above (grayscale for squared correlation R^2: white - $R^2 = 0$, black - $R^2 = 1$). NOTE: empirical size for kernel-based methods is not affected by increasing SNP set size: ASAH1 – 0.048, FGFR2 – 0.052.