The American Journal of Human Genetics

Supplemental Data

Testing in Microbiome-Profiling Studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test

Ni Zhao, Jun Chen, Ian M. Carroll, Tamar Ringel-Kulka, Michael P. Epstein, Hua Zhou, Jin J. Zhou, Yehuda Ringel, Hongzhe Li, and Michael C. Wu

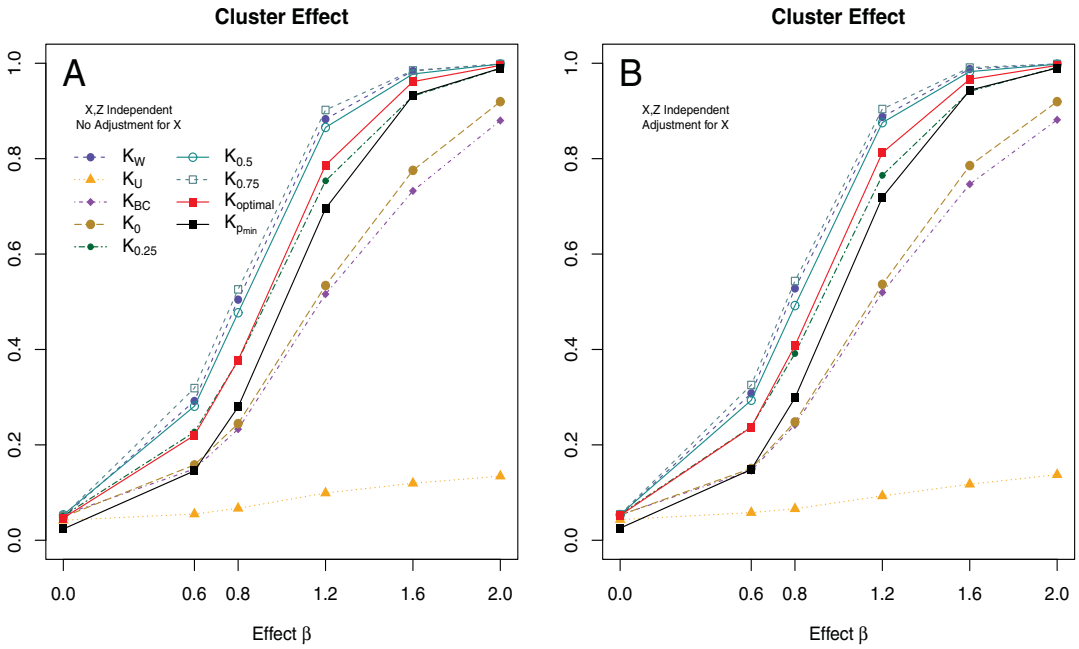

Figure S1: **Type I error and Power of MiRKAT Based on Different Kernels for *Simulation Scenario 1* with Dichotomous Outcome when X and Z are Independent**: a selected phylogenetic cluster of the OTUs are associated with the outcome. Additional covariates X and microbiome effect Z were simulated independently. Panel A shows the results for tests that do not adjust for **X** and panel B shows results that adjust for **X**. K_w, K_u, K_{BC}, K_0, $K_{0.25}$, $K_{0.5}$ and $K_{0.75}$ represents MiRKAT results using different individual kernels respectively: weighted UniFrac, unweighted UniFrac, Bray-Curtis, and generalized UniFrac kernels with $\alpha = 0$, 0.25, 0.5 and 0.75. $K_{optimal}$ represents the simulation results for optimal MiRKAT considering all seven kernels and K_{minP} shows the results using a naive Bonferroni adjusted test. Results were presented at $n = 200$.

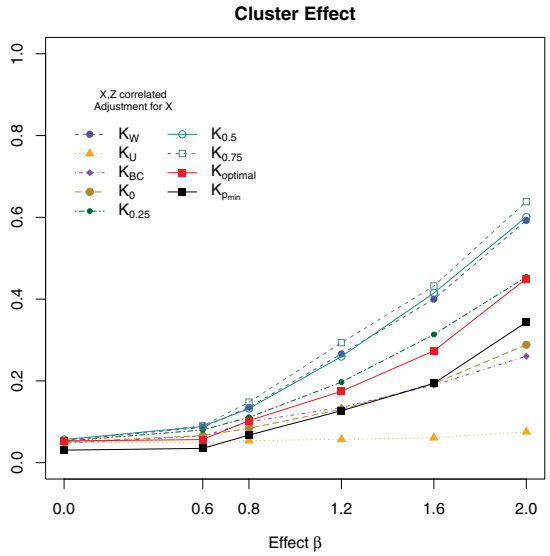

Figure S2: **Type I error and Power of MiRKAT Based on Different Kernels for *Simulation Scenario 1* with Dichotomous Outcome when X and Z are Correlated**: a selected phylogenetic cluster of the OTUs are associated with the outcome. Additional covariates X and microbiome composition Z are correlated through $X_{2i} = \text{scale}\left(\sum_{j\in\mathcal{A}} Z_{ij}\right) + N(0,1)$. We only considered MiRKAT with **X** adjustment because unadjusted tests give seriously inflated type I error. K_w, K_u, K_{BC}, K_0, $K_{0.25}$, $K_{0.5}$ and $K_{0.75}$ represents MiRKAT results using different individual kernels respectively: weighted UniFrac, unweighted UniFrac, Bray-Curtis, and generalized UniFrac kernels with $\alpha = 0$, 0.25, 0.5 and 0.75. $K_{optimal}$ represents the simulation results for optimal MiRKAT considering all seven kernels and K_{minP} shows the results using a naive Bonferroni adjusted test. Sample Size $n = 200$.

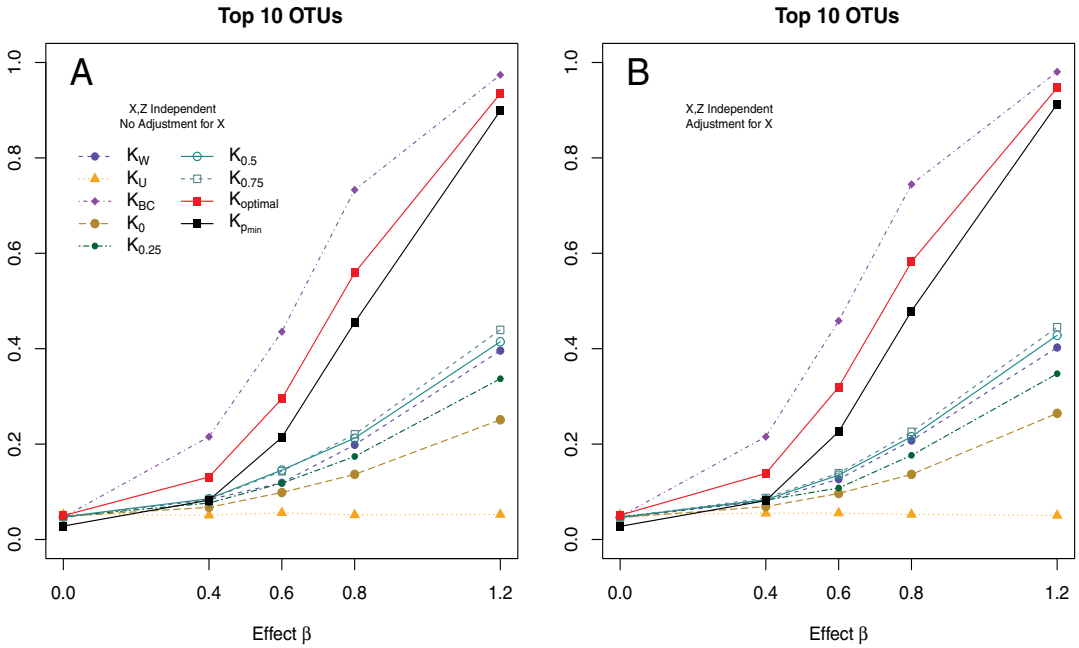

Figure S3: **Type I error and Power of MiRKAT Based on Different Kernels for *Simulation Scenario 2* with Dichotomous Outcome when X and Z are Independent**: the 10 most abundant OTUs are associated with the outcome. Additional covariates X and microbiome effect Z were simulated independently. Panel A shows the results for tests that do not adjust for **X** and panel B shows results that adjust for **X**. K_w, K_u, K_{BC}, K_0, $K_{0.25}$, $K_{0.5}$ and $K_{0.75}$ represents MiRKAT results using different individual kernels respectively: weighted UniFrac, unweighted UniFrac, Bray-Curtis, and generalized UniFrac kernels with $\alpha = 0$, 0.25, 0.5 and 0.75. $K_{optimal}$ represents the simulation results for optimal MiRKAT considering all seven kernels and K_{minP} shows the results using a naive Bonferroni adjusted test. Results were presented at $n = 200$.

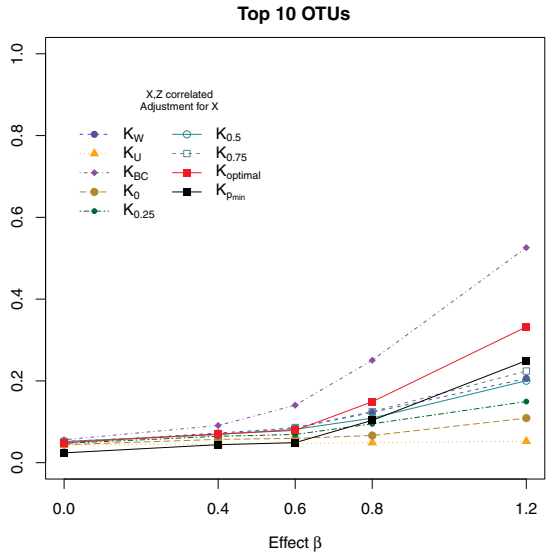

Figure S4: **Type I error and Power of MiRKAT Based on Different Kernels for *Simulation Scenario* 2 with Dichotomous Outcome when X and Z are Correlated**: the 10 most abundant OTUs are associated with the outcome. Additional covariates X and microbiome composition Z are correlated through $X_{2i} =$ scale $(\sum_{j\in\mathcal{A}} Z_{ij}) + N(0,1)$. We only considered MiRKAT with **X** adjustment because unadjusted tests give seriously inflated type I error. K_w, K_u, K_{BC}, K_0, $K_{0.25}$, $K_{0.5}$ and $K_{0.75}$ represents MiRKAT results using different individual kernels respectively: weighted UniFrac, unweighted UniFrac, Bray-Curtis, and generalized UniFrac kernels with $\alpha = 0$, 0.25, 0.5 and 0.75. $K_{optimal}$ represents the simulation results for optimal MiRKAT considering all seven kernels and K_{minP} shows the results using a naive Bonferroni adjusted test. Results were presented at $n = 200$.

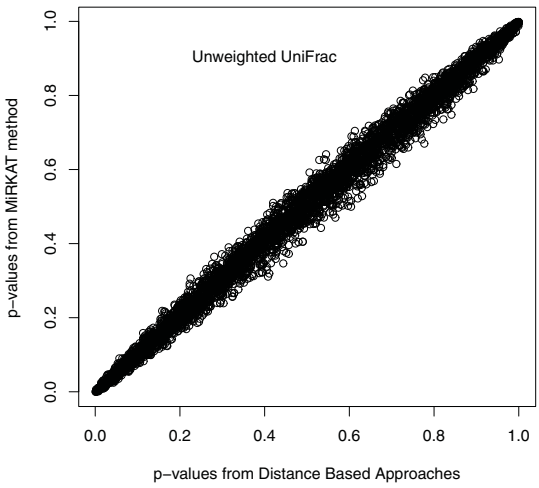

Figure S5: Example plot of the p-value correlation using distance based approach and MiRKAT when no additional covariates are considered. 5000 simulations are plotted at sample size n = 200 for continuous outcome. Unweighted UniFrac distance and kernel were used for the distance based approach and MiRKAT respectively.

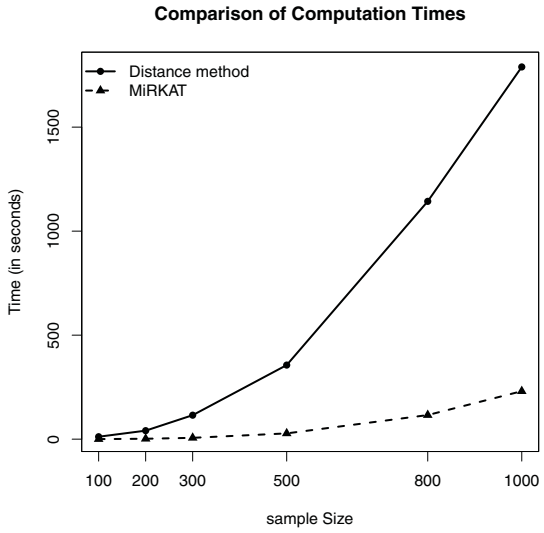

Figure S6: Computation times of MiRKAT and distance based test as a function of the sample size for continuous outcome. The figure presents the total computation time for 100 repeated tests with each sample size. 999 permutations (the default number) were used in distance based approaches.

Table S1: Empirical type I errors for MiRKAT and "optimal" MiRKAT with dichotomous outcome. Type I error was evaluated for scenarios when additional covariates are independent with the OTUs ($X \perp Z$) and scenarios when covariates are related to the OTUs ($X \not\perp Z$) using 5000 simulated data sets. K_w, K_u, K_{BC}, K_0, $K_{0.25}$, $K_{0.5}$ and $K_{0.75}$ represents MiRKAT results using different individual kernels respectively: weighted UniFrac, unweighted UniFrac, Bray-Curtis, and generalized UniFrac kernels with $\alpha = 0$, 0.25, 0.5 and 0.75. $K_{optimal}$ represents the simulation results for optimal MiRKAT considering all seven kernels and K_{minP} shows the results using a naive Bonferroni adjusted test. P-values for "optimal" MiRKAT were obtained by 1000 permutations. Numbers in **bold** show inflated type I error.

Simulation scenario 1: Clustered OTUs									
$X \perp Z$	Unadjust for **X**								
n	K_W	K_U	K_{BC}	K_0	$K_{0.25}$	$K_{0.5}$	$K_{0.75}$	K_{opt}	K_{minP}
200	0.051	0.049	0.049	0.051	0.052	0.054	0.051	0.049	0.025
500	0.046	0.049	0.054	0.056	0.053	0.054	0.053	0.053	0.028
$X \perp Z$	Adjust for **X**								
200	0.054	0.051	0.050	0.051	0.053	0.054	0.054	0.053	0.028
500	0.047	0.048	0.051	0.053	0.055	0.051	0.049	0.055	0.029
$X \not\perp Z$	Unadjust for **X**								
200	**0.105**	**0.054**	**0.075**	**0.081**	**0.099**	**0.116**	**0.123**	**0.092**	**0.057**
500	**0.156**	**0.056**	**0.092**	**0.149**	**0.210**	**0.260**	**0.285**	**0.214**	**0.138**
$X \not\perp Z$	Adjust for **X**								
200	0.048	0.054	0.049	0.050	0.050	0.053	0.052	0.051	0.028
500	0.045	0.051	0.050	0.051	0.048	0.049	0.049	0.048	0.024
Simulation scenario 2: top 10 OTUs									
$X \perp Z$	Unadjust for **X**								
n	K_W	K_U	K_{BC}	K_0	$K_{0.25}$	$K_{0.5}$	$K_{0.75}$	K_{opt}	K_{minP}
200	0.046	0.052	0.047	0.048	0.048	0.047	0.047	0.050	0.028
500	0.058	0.044	0.045	0.051	0.050	0.052	0.053	0.048	0.025
$X \perp Z$	Adjust for **X**								
200	0.045	0.052	0.048	0.046	0.048	0.046	0.046	0.051	0.028
500	0.052	0.045	0.040	0.048	0.052	0.052	0.050	0.042	0.022
$X \not\perp Z$	Unadjust for **X**								
200	**0.066**	**0.051**	**0.201**	**0.064**	**0.069**	**0.070**	**0.073**	**0.125**	**0.077**
500	**0.123**	**0.049**	**0.544**	**0.101**	**0.104**	**0.123**	**0.126**	**0.378**	**0.307**
$X \not\perp Z$	Adjust for **X**								
200	0.047	0.056	0.052	0.044	0.047	0.052	0.052	0.049	0.024
500	0.051	0.047	0.056	0.051	0.050	0.046	0.049	0.054	0.024