Similarity-Based Multimodal Regression Supplementary Materials

Andrew A. Chen[a,b,*], Sarah M. Weinstein[a], Azeez Adebimpe[c,d], Ruben C. Gur[d,e], Raquel E. Gur[d,e], Kathleen Ries Merikangas[f], Theodore D. Satterthwaite[a,b,c,d], Russell T. Shinohara[a,b,1], and Haochang Shou[a,b,1]

[a] *Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA*

[b] *Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA*

[c] *Penn Lifespan Informatics & Neuroimaging Center, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA*

[d] *Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA*

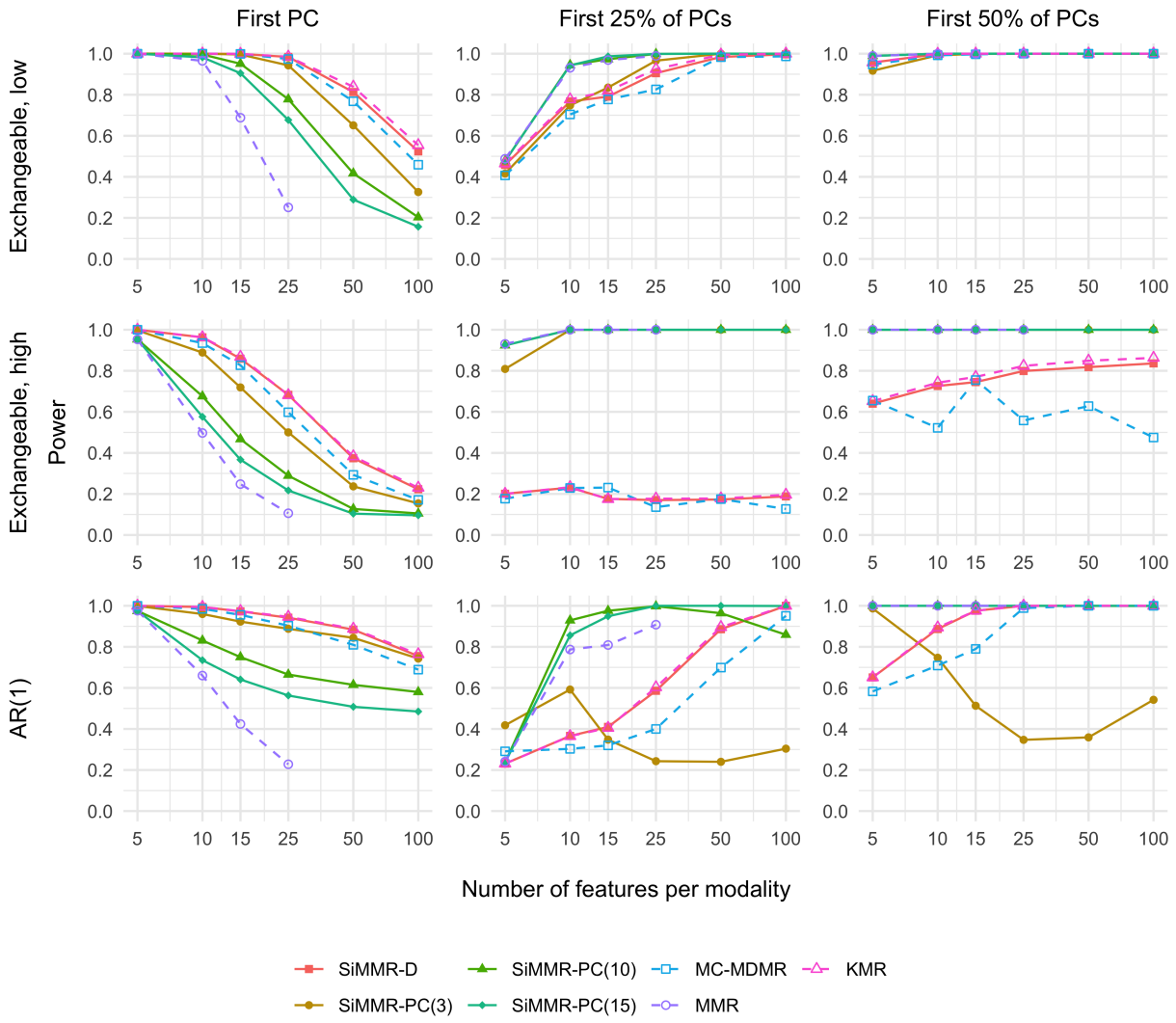
[e] *Lifespan Brain Institute Penn Medicine and CHOP, University of Pennsylvania, Philadelphia, PA 19104, USA*

[f] *Genetic Epidemiology Research Branch, Intramural Research Program, National Institute of Mental Health, Bethesda, MD 20892, USA*

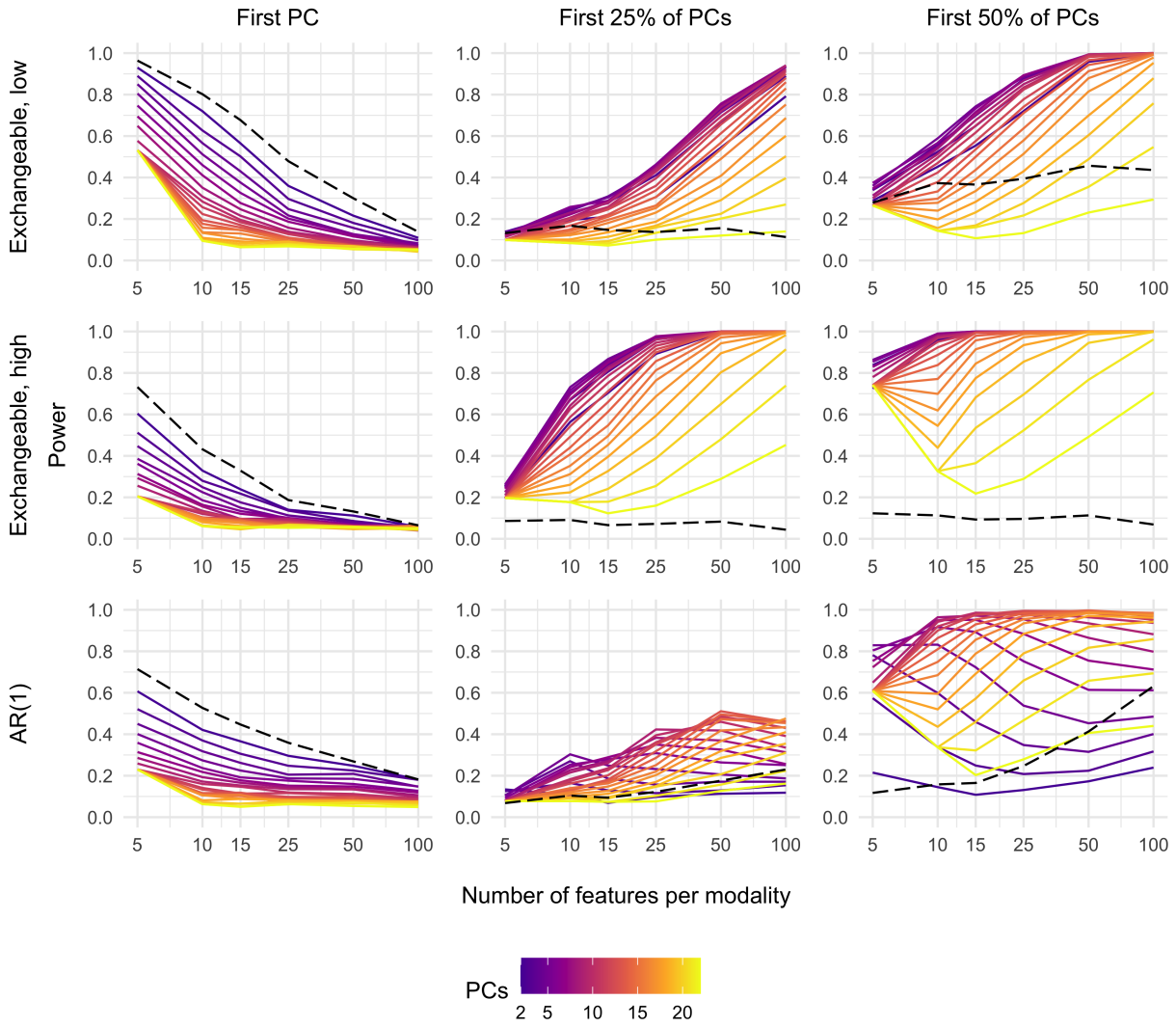
[1] *Equal contribution*

***Correspondence: Andrew A. Chen**

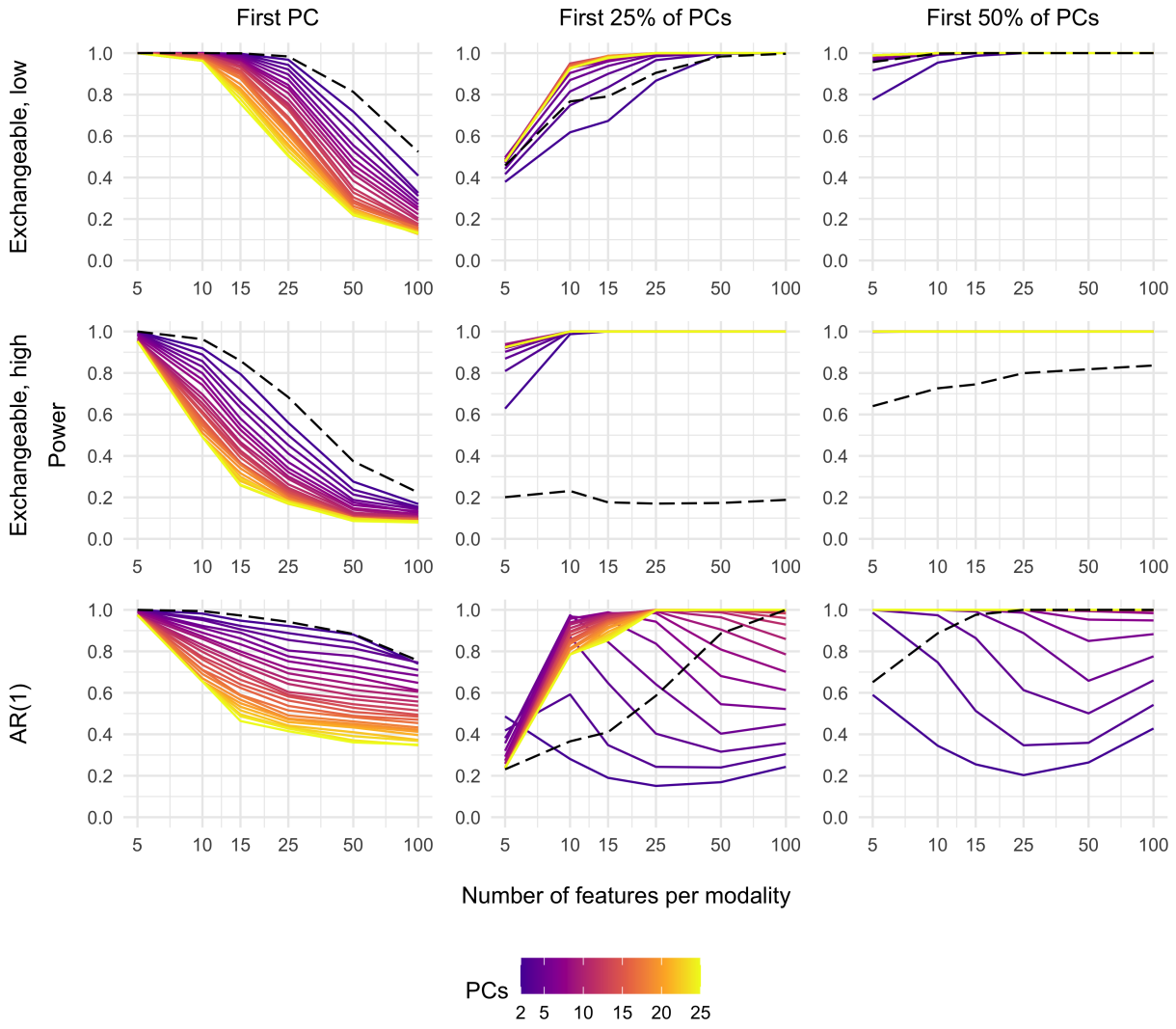
andrewac@pennmedicine.upenn.edu



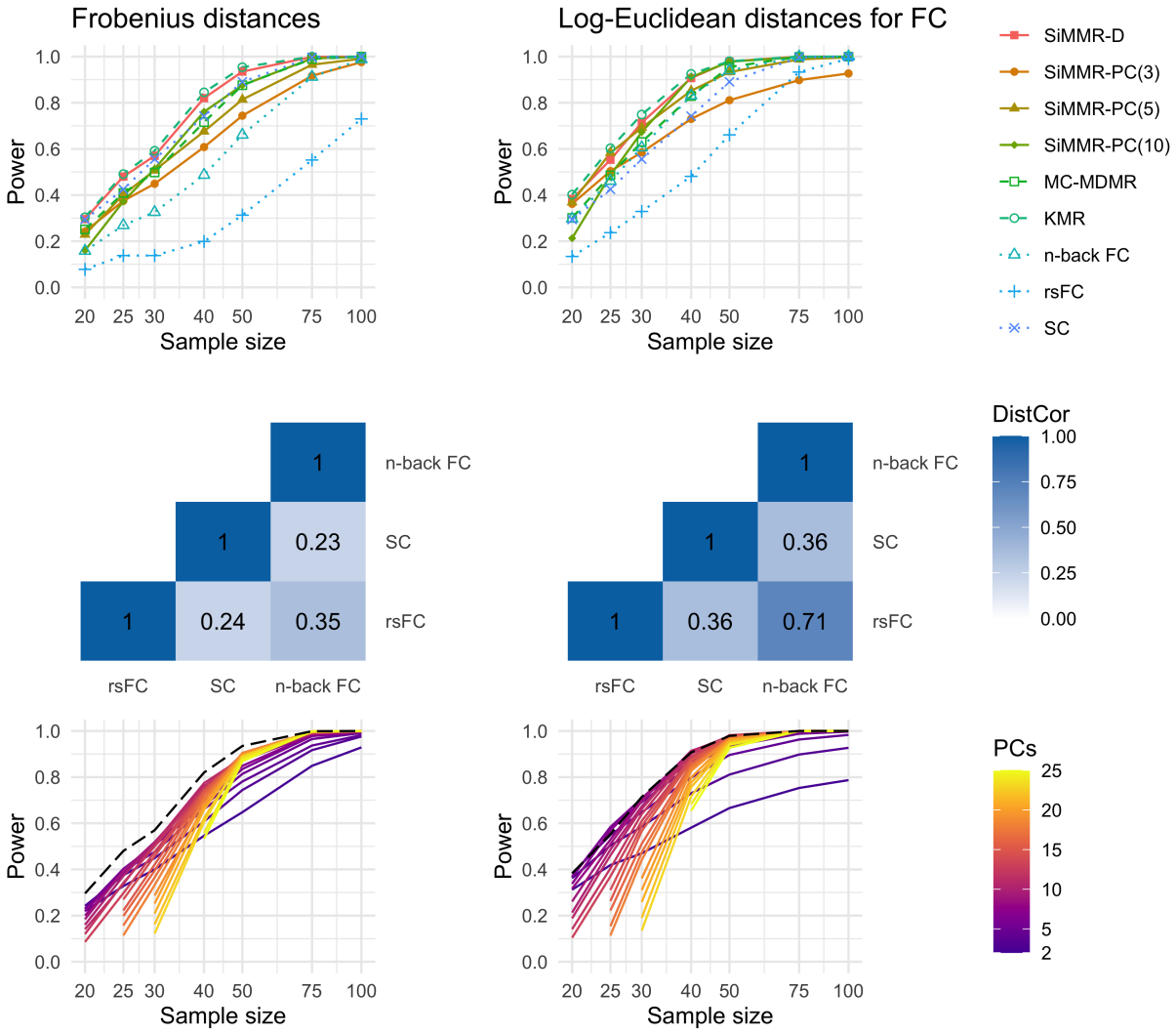
Supplementary Figure 1: **Results in simulations with exchangeable and AR(1) correlation structures for a sample size of 100.** Each trace represents a different test statistic. Different simulation settings are distinguished by correlation structure across rows and by rank of the binary covariate effect across columns. Abbreviations: MDMR, multivariate distance matrix regression; MC-MDMR, multiple MDMR statistics after Bonferroni correction; MMR, multivariate multiple regression using Pillai's trace.



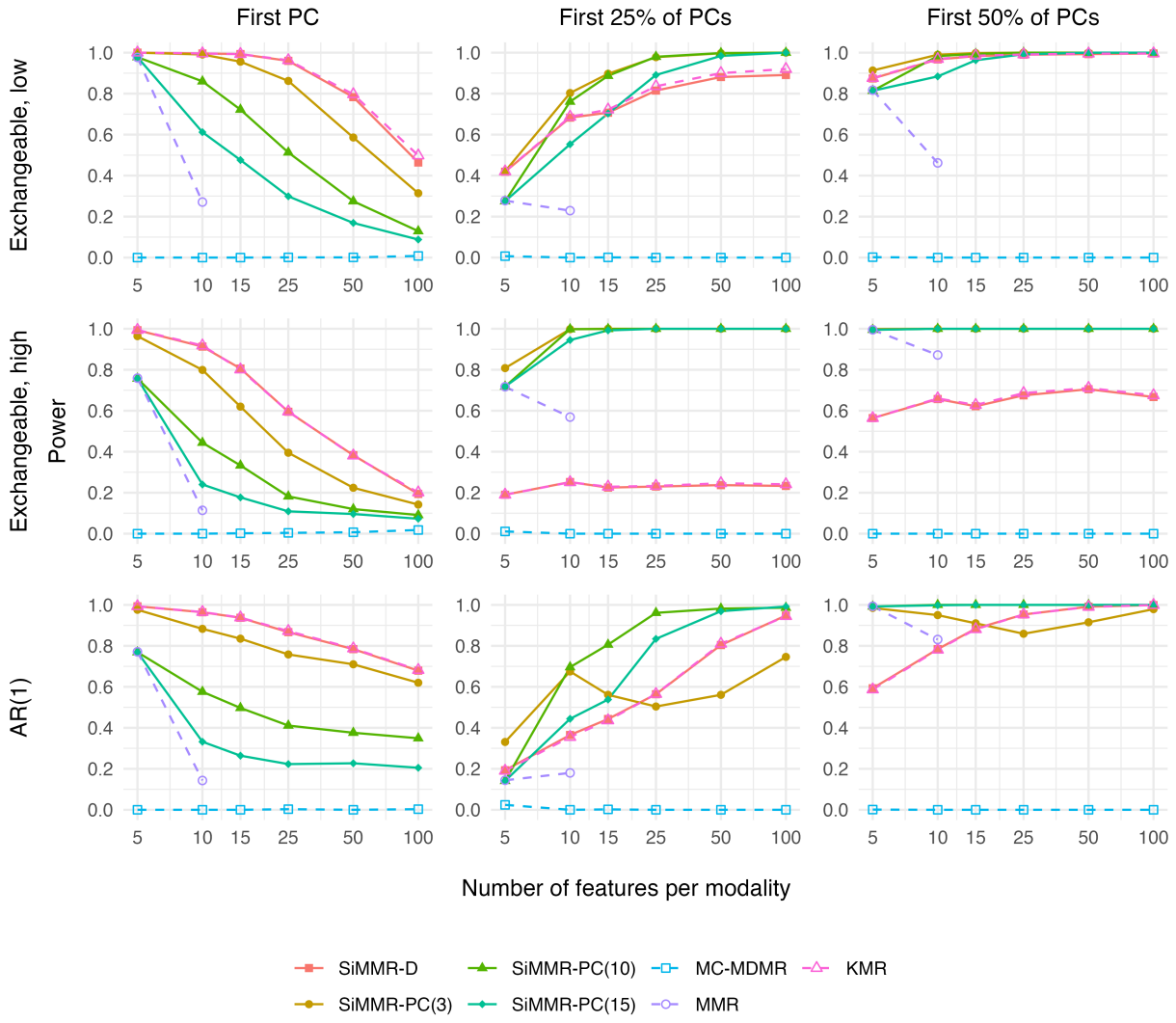
Supplementary Figure 2: **SiMMR-PC power results in simulations with exchangeable and AR1 correlation structures with a sample size of 25.** Results are shown for SiMMR-PC with PCs ranging from 2 to 25. The dashed line represents the power of SiMMR-D for comparison.



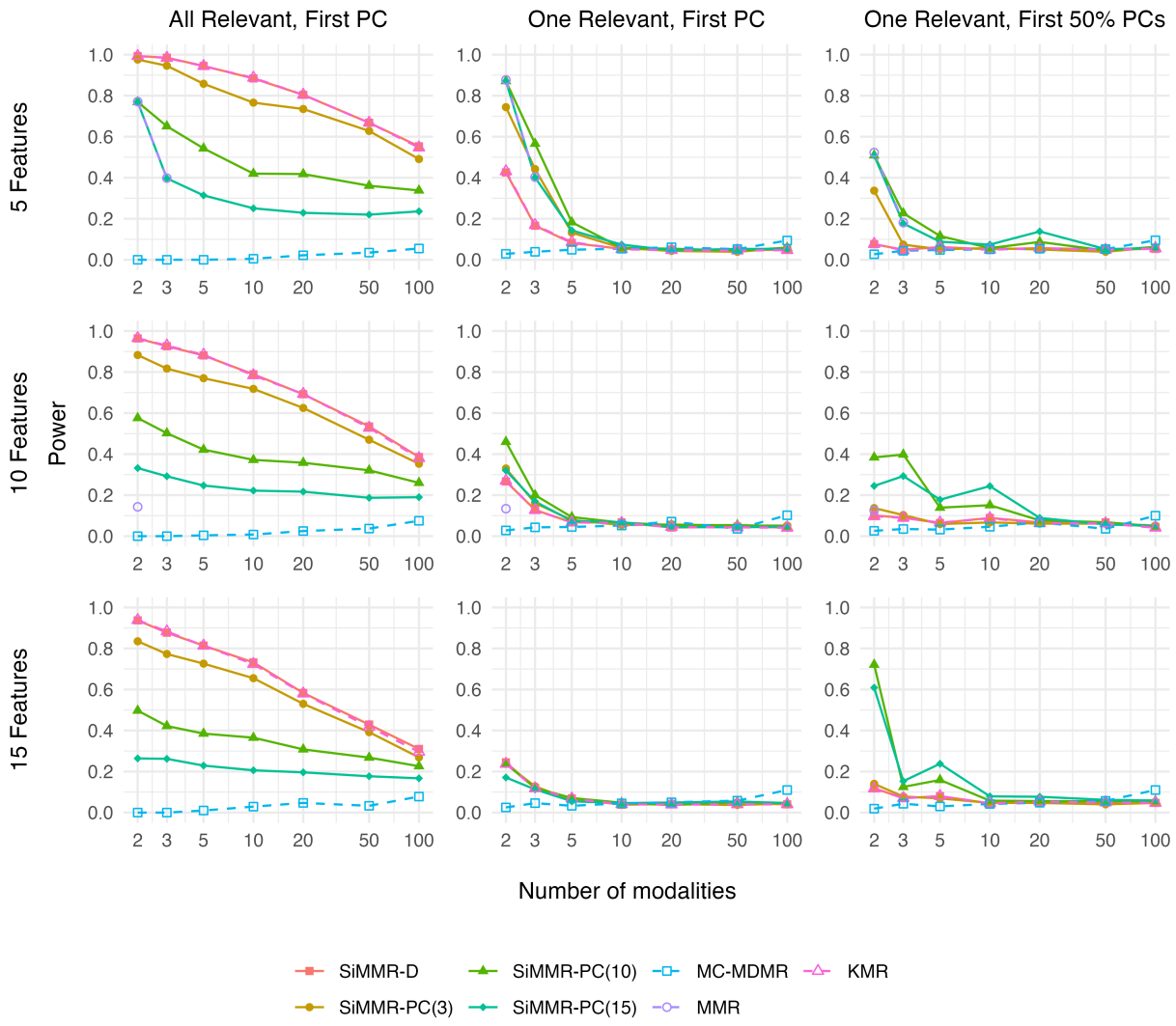
Supplementary Figure 3: **SiMMR-PC power results in simulations with exchangeable and AR1 correlation structures with a sample size of 100.** Results are shown for SiMMR-PC with PCs ranging from 2 to 25. The dashed line represents the power of SiMMR-D for comparison.



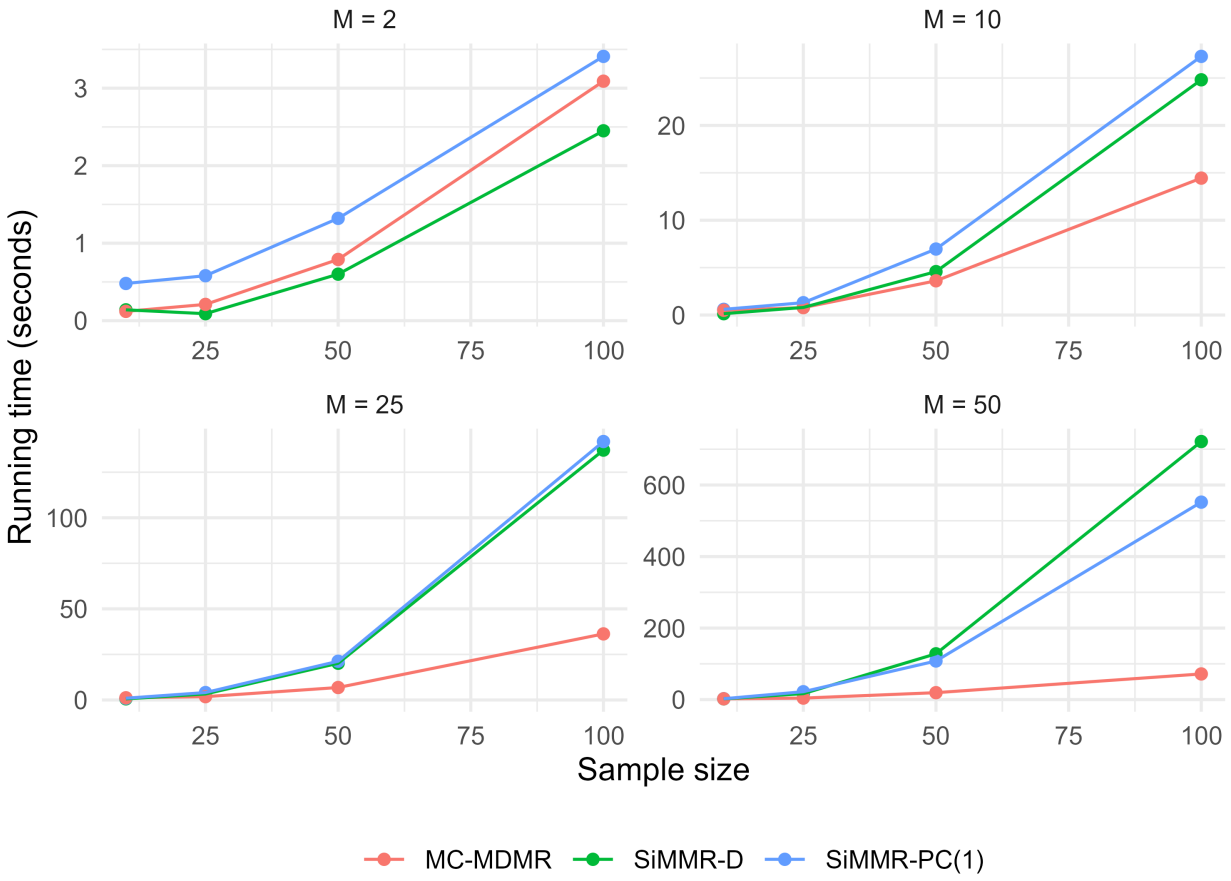
Supplementary Figure 4: **Comparison of Frobenius and log-Euclidean distances for PNC functional connectivity analyses.** Power for each test statistic and power across SiMMR-PC test statistics are shown for analyses using Frobenius versus log-Euclidean distances for rsFC and n-back FC dissimilarity matrices. Distance correlations are also displayed for the two choices of metric. Abbreviations: PNC, Philadelphia Neurodevelopmental Cohort; FC, functional connectivity; rsFC, resting-state functional connectivity; SC, structural connectivity.



Supplementary Figure 5: **Power results in continuous covariate simulations with exchangeable and AR(1) correlation structures for a sample size of 25.** Each trace represents a different test statistic. Different simulation settings are distinguished by correlation structure across rows and by rank of the binary covariate effect across columns. Exchangeable refers to an exchangeable correlation structure with low or high correlation and AR(1) refers to a first-order autoregressive structure. Abbreviations: MDMR, multivariate distance matrix regression; MC-MDMR, multiple MDMR statistics after Bonferroni correction; MMR, multivariate multiple regression using Pillai's trace.



Supplementary Figure 6: **Power results in AR(1) simulations where only one modality is relevant for a sample size of 25.** Each trace represents a different test statistic. Different simulation settings are distinguished by number of features per modality on the rows and simulation settings on the columns. All Relevant refers to the setting where every modality is related to the covariate of interest while One Relevant refers to a setting where only one modality is related, across all numbers of modalities. Abbreviations: MDMR, multivariate distance matrix regression; MC-MDMR, multiple MDMR statistics after Bonferroni correction; MMR, multivariate multiple regression using Pillai's trace.



Supplementary Figure 7: **Running time in seconds for SiMMR test statistics and MC-MDMR across settings with exchangeable, high correlation.** Running time is reported for one simulated dataset in each setting. Settings vary in their number of observations (N) and number of modalities (M). All tests were run on a laptop with an Intel Core i9-13900H (24M Cache, up to 5.40 GHz) and 32 GB of RAM. Abbreviations: MDMR, multivariate distance matrix regression; MC-MDMR.

PC	N	M	Q	MMR	MCM	KMR	SiMMR-D	SiMMR-PC(3)	SiMMR-PC(10)	SiMMR-PC(15)
None	25	2	5	0.054	0.042	0.048	0.048	0.051	0.052	0.052
			25		0.048	0.043	0.044	0.051	0.049	0.045
			100		0.036	0.037	0.037	0.042	0.038	0.045
		10	5		0.032	0.05	0.048	0.033	0.035	0.037
			25		0.045	0.045	0.044	0.037	0.053	0.056
			100		0.05	0.042	0.041	0.04	0.041	0.054
	100	2	5	0.051	0.052	0.057	0.055	0.05	0.048	0.048
			25	0.053	0.045	0.046	0.042	0.047	0.048	0.046
			100		0.044	0.047	0.046	0.038	0.048	0.052
		10	5	0.04	0.029	0.05	0.05	0.06	0.039	0.048
			25		0.051	0.046	0.042	0.042	0.044	0.042
			100		0.051	0.055	0.055	0.045	0.048	0.05
25%	25	2	5	0.078	0.081	0.065	0.068	**0.125**	0.077	0.077
			25		0.1	0.124	0.122	0.116	**0.395**	0.251
			100		0.151	0.23	0.23	0.154	0.391	**0.454**
		10	5		0.09	0.114	0.115	0.106	**0.373**	0.245
			25		0.134	0.295	0.305	0.202	0.395	**0.493**
			100		0.227	0.787	**0.796**	0.517	0.718	0.693
	100	2	5	0.241	0.291	0.231	0.231	**0.418**	0.237	0.237
			25	0.908	0.4	0.603	0.586	0.243	0.998	**1**
			100		0.951	**1**	**1**	0.304	0.859	0.999
		10	5	0.9	0.409	0.576	0.568	0.222	**1**	**1**
			25		0.68	**1**	0.999	0.318	0.843	0.99
			100		**1**	**1**	**1**	0.92	0.999	**1**
50%	25	2	5	**0.616**	0.125	0.117	0.117	0.574	0.609	0.609
			25		0.167	0.238	0.247	0.208	**0.983**	0.971
			100		0.393	0.635	0.632	0.317	0.937	**0.985**
		10	5		0.136	0.238	0.233	0.187	**0.982**	0.975
			25		0.209	0.789	0.793	0.45	0.906	**0.985**
			100		0.554	**1**	**1**	0.954	0.999	0.996
	100	2	5	**1**	0.583	0.651	0.651	0.987	**1**	**1**
			25	**1**	0.989	**1**	**1**	0.347	**1**	**1**
			100		**1**	**1**	**1**	0.542	0.999	**1**
		10	5	**1**	0.959	**1**	**1**	0.322	**1**	**1**
			25		0.995	**1**	**1**	0.648	**1**	**1**
			100		**1**	**1**	**1**	**1**	**1**	**1**

Supplementary Table 1: **Simulation results for the AR(1) correlation setting.** Rejection rates are shown across varying number of subjects (N), number of modalities (M), number of features per modality (Q), and number of principal components included in the binary covariate effect (PC). The highest power in each row is bolded.

				Exchangeable, low correlation					Exchangeable, high correlation					AR(1) correlation				
PC	N	M	Q	MMR	MCM	KMR	SiMMR-D	SiMMR-PC(3)	MMR	MCM	KMR	SiMMR-D	SiMMR-PC(3)	MMR	MCM	KMR	SiMMR-D	SiMMR-PC(3)
None	25	2	5	0.038	0.044	0.05	0.049	0.05	0.038	0.051	0.047	0.046	0.049	0.038	0.051	0.045	0.045	0.043
			100		0.032	0.047	0.038	0.05		0.039	0.041	0.043	0.049		0.038	0.041	0.043	0.051
		10	5		0.06	0.053	0.047	0.047		0.053	0.048	0.045	0.043		0.052	0.039	0.039	0.042
			100		0.05	0.064	0.048	0.042		0.038	0.054	0.048	0.04		0.052	0.052	0.057	0.046
	100	2	5	0.053	0.054	0.05	0.05	0.057	0.053	0.052	0.051	0.051	0.056	0.053	0.054	0.053	0.052	0.056
			100		0.046	0.058	0.049	0.063		0.041	0.048	0.05	0.061		0.038	0.063	0.061	0.054
		10	5	0.057	0.048	0.039	0.039	0.052	0.057	0.047	0.044	0.043	0.051	0.057	0.054	0.042	0.041	0.047
			100		0.045	0.064	0.059	0.046		0.043	0.059	0.06	0.05		0.041	0.065	0.066	0.053
1	25	2	5	0.977	0	**1**	**1**	**1**	0.759	0	**0.994**	0.993	0.964	0.771	0	**0.993**	**0.993**	0.976
			100		0.008	**0.498**	0.464	0.314		0.018	**0.201**	0.194	0.142		0.003	**0.682**	0.679	0.62
		10	5		0.001	**0.96**	0.951	0.847		0.004	**0.602**	**0.602**	0.38		0.005	**0.887**	0.884	0.766
			100		0.035	**0.156**	0.137	0.1		0.036	**0.073**	0.071	0.06		0.037	0.37	**0.379**	0.324
	100	2	5	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	**1**
			100		0	**0.985**	0.984	0.942		0.003	**0.676**	0.668	0.5		0	**1**	**1**	**1**
		10	5	0.919	0	**1**	**1**	**1**	0.375	0	**0.995**	0.994	0.986	0.843	0	**1**	**1**	**1**
			100		0.006	**0.506**	0.48	0.318		0.023	**0.204**	0.2	0.129		0.009	**0.991**	0.99	0.943
25%	25	2	5	0.278	0.007	0.42	0.419	**0.421**	0.717	0.011	0.19	0.19	**0.808**	0.144	0.024	0.19	0.193	**0.331**
			100		0	0.92	0.891	**1**		0	0.241	0.233	**1**		0	0.945	**0.948**	0.746
		10	5		0.01	0.811	0.786	**0.969**		0.007	0.218	0.208	**1**		0.007	**0.538**	0.534	0.456
			100		0	0.934	0.907	**1**		0	0.266	0.258	**1**		0	**1**	**1**	0.996
	100	2	5	0.992	0	**0.994**	0.993	0.964	**1**	0.001	0.833	0.826	**1**	0.839	0	0.785	0.779	**0.967**
			100		0	**1**	**1**	**1**		0	0.999	0.999	**1**		0	**1**	**1**	0.989
		10	5	**1**	0.001	**1**	**1**	**1**	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	0.837
			100		0	**1**	**1**	**1**		0	**1**	**1**	**1**		0	**1**	**1**	**1**
50%	25	2	5	0.817	0.002	0.876	0.874	**0.913**	0.995	0	0.564	0.564	**0.999**	**0.993**	0.001	0.587	0.593	0.985
			100		0	0.996	0.996	**1**		0	0.675	0.667	**1**		0	**0.999**	**0.999**	0.979
		10	5		0.002	0.993	0.991	**1**		0.002	0.657	0.651	**1**		0.001	**0.956**	0.952	0.848
			100		0	0.998	0.996	**1**		0	0.692	0.684	**1**		0	**1**	**1**	**1**
	100	2	5	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	**1**
			100		0	**1**	**1**	**1**		0	**1**	**1**	**1**		0	**1**	**1**	**1**
		10	5	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	**1**	**1**	0	**1**	**1**	0.984
			100		0	**1**	**1**	**1**		0	**1**	**1**	**1**		0	**1**	**1**	**1**

Supplementary Table 2: **Continuous covariate simulation results across varying sample size, number of features, covariate effect, and correlation structure.** Rejection rates are shown across varying number of subjects (N), number of modalities (M), number of features per modality (Q), and number of principal components included in the binary covariate effect (PC). The highest power among tests within each simulation setting is bolded. Abbreviations: MMR, multivariate multiple regression using Pillai's trace; MCM, multiple MDMR statistics after Bonferroni correction; KMR: kernel machine regression.

PC	N	M	Q	MMR	MCM	KMR	SiMMR-D	SiMMR-PC(3)	SiMMR-PC(10)	SiMMR-PC(15)
None	25	2	5	0.038	0.051	0.045	0.045	0.043	0.037	0.037
			25		0.044	0.044	0.045	0.056	0.046	0.046
			100		0.038	0.041	0.043	0.051	0.043	0.047
		10	5		0.052	0.039	0.039	0.042	0.045	0.043
			25		0.05	0.05	0.051	0.048	0.05	0.046
			100		0.052	0.052	0.057	0.046	0.057	0.049
	100	2	5	0.053	0.054	0.053	0.052	0.056	0.052	0.052
			25	0.058	0.043	0.06	0.058	0.064	0.056	0.061
			100		0.038	0.063	0.061	0.054	0.049	0.05
		10	5	0.057	0.054	0.042	0.041	0.047	0.05	0.057
			25		0.058	0.044	0.038	0.032	0.041	0.061
			100		0.041	0.065	0.066	0.053	0.061	0.065
25%	25	2	5	0.144	0.024	0.19	0.193	**0.331**	0.141	0.141
			25		0	0.565	0.564	0.504	**0.961**	0.834
			100		0	0.945	0.948	0.746	0.986	**0.992**
		10	5		0.007	0.538	0.534	0.456	**0.951**	0.816
			25		0	0.972	0.971	0.858	0.977	**0.989**
			100		0	**1**	**1**	0.996	**1**	**1**
	100	2	5	0.839	0	0.785	0.779	**0.967**	0.836	0.836
			25	**1**	0	**1**	**1**	0.9	**1**	**1**
			100		0	**1**	**1**	0.989	**1**	**1**
		10	5	**1**	0	**1**	**1**	0.837	**1**	**1**
			25		0	**1**	**1**	0.992	**1**	**1**
			100		0	**1**	**1**	**1**	**1**	**1**
50%	25	2	5	**0.993**	0.001	0.587	0.593	0.985	0.992	0.992
			25		0	0.953	0.953	0.859	**1**	**1**
			100		0	0.999	0.999	0.979	**1**	**1**
		10	5		0.001	0.956	0.952	0.848	**1**	**1**
			25		0	**1**	**1**	0.993	**1**	**1**
			100		0	**1**	**1**	**1**	**1**	**1**
	100	2	5	**1**	0	**1**	**1**	**1**	**1**	**1**
			25	**1**	0	**1**	**1**	0.992	**1**	**1**
			100		0	**1**	**1**	**1**	**1**	**1**
		10	5	**1**	0	**1**	**1**	0.984	**1**	**1**
			25		0	**1**	**1**	**1**	**1**	**1**
			100		0	**1**	**1**	**1**	**1**	**1**

Supplementary Table 3: **Continuous covariate simulation results for the AR(1) correlation setting.** Rejection rates are shown across varying number of subjects (N), number of modalities (M), number of features per modality (Q), and number of principal components included in the binary covariate effect (PC). The highest power in each row is bolded.