Supplementary Materials for "Functional Linear Models for Association Analysis of Quantitative Traits"

Application to The SNP Data of The Trinity Students Study

Table S.1 presents extra results of SNP data of the enzyme gene of the Trinity Students Study. The results of functional kernel score tests of mixed effect models are somewhat mixed: for additive genetic variant function and Fourier basis, the functional kernel score tests perform better than SKAT. For dominant and recessive genetic variant functions and Fourier basis, the functional kernel score tests may perform poorer than SKAT. For B-spline basis, the functional kernel score tests may sometimes perform poorer, or sometimes better than SKAT.

Table S.1: **Results of Three Traits of the Trinity Students Study in the Region of An Enzyme Gene.** The p-values of "Basis of both GVF and $\beta(t)$" Were Based on the F-distributed Test Statistics of the Fixed Effect Model (3), the p-values of "Basis of FKST" Were Based on the Variance-Component Functional Kernel Score Test (7), and the p-values of SKAT Were Based of R Package SKAT. When Type of GVF = Add, Dom, and Rec, the Genetic Variant Function Was Taken as Additive, Dominant, and Recessive, Respectively. Abbreviation: GVF = Genetic Variant Function, FKST = Functional Kernel Score Tests.

Type of GVF	**Trait**	**P-values of the Proposed Tests**				**P-values of SKAT**
		Basis of both GVF and $\beta(t)$		**Basis of FKST**		
		B-spline Basis	**Fourier Basis**	**B-spline Basis**	**Fourier Basis**	
Add	A	1.73E-13	7.89E-13	3.98E-10	5.05E-13	7.13E-08
	B	3.44E-13	1.80E-11	0.180845	5.46E-08	4.34E-05
	C	1.11E-11	9.91E-10	0.008551	3.10E-07	5.80E-05
Dom	A	1.83E-10	1.71E-10	1.04E-08	3.58E-07	3.10E-07
	B	4.66E-09	3.35E-07	0.150001	2.85E-05	0.000152
	C	1.54E-08	5.93E-07	0.003514	0.000165	8.99E-05
Rec	A	5.26E-11	9.92E-11	5.23E-08	0.009039	7.28E-06
	B	2.75E-06	4.38E-06	0.566176	0.057826	0.063910
	C	2.26E-07	5.19E-08	0.251738	0.103260	0.075480

Simulation Results

In this **Supplementary Materials**, we report the simulation results in addition to those reported in the main text.

Type I Errors. Table S.2 reports results of type I error rates of the four tests based on sequence data generated by COSI. In Tables S.3 and S.4, we used dominant and recessive genetic variant functions, respectively, based on the enzyme gene SNP data of the Trinity Students Study, to calculate the empirical type I error rates.

Statistical Power Based on the Enzyme Gene SNP Data of the Trinity Students Study. Figures S.7, S.8, S.9, S.10, and S.11 report the results of power simulations based on the enzyme gene SNP data of the Trinity Students Study and model (11), when one SNP is the only causal variant for the modes of additive, dominant, and recessive inheritance. The Figures S.7 and S.8 correspond to small and moderate sample sizes of $n = 250$ and 500, respectively. The Figures S.9, S.10, and S.11 provide power performance of larger sample sizes of $n = 1,000, 1,500$, and $2,000$, respectively.

As expected, single causal SNP regression model provides the highest power compared with the proposed tests, and SKAT and SKAT-O. For the mode of additive inheritance, the proposed test statistics have substantially higher power than that of SKAT except for small sample sizes $n = 250$ (when $n = 250$, both SKAT/SKAT-O and the functional kernel score test of Fourier basis have minimal power as shown in graphs (a1), (a2), and (a3) of Figure S.7). Interestingly, the functional kernel score test of B-spline offers the second highest power for small and moderate sample sizes of $n = 250$ and 500 as shown in graphs (a1), (a2), and (a3) of Figures S.7 and S.8. Hence, it can be useful when the mode of additive inheritance can be assumed. For the mode of dominant inheritance, the two F-distributed tests of the fixed effect models and the functional kernel score test of B-spline have higher power than that of SKAT and SKAT-O, and functional kernel score test of Fourier basis and SKAT and SKAT-O have minimal power except for large sample size case of $n = 2,000$. For the mode of recessive inheritance, only the two F-distributed tests of fixed effect models offer good power when the sample size $n \geq 1,000$. Both the two functional kernel score tests of mixed effect models and SKAT/SKAT-O have minimal power.

Table S.2: **Simulation Results of Type I Error Rates of Four Tests Based on Sequence Data Generated by COSI, When The Genetic Variant Functions Were Taken as Additive.** The Results of "Basis of both GVF and $\beta(t)$" Were Based on the F-distributed Test Statistics of the Fixed Effect Model (3) and the Results of "Basis of FKST" Were Based on the Variance-Component Functional Kernel Score Test (7) of Mixed Effect Models, for B-spline Basis and Fourier Basis, Respectively. Abbreviation: GVF = Genetic Variant Function, FKST = Functional Kernel Score Tests.

Nominal Level α	**Sample Size**	**Basis of both GVF and $\beta(t)$**		**Basis of FKST**	
		B-spline Basis	**Fourier Basis**	**B-spline Basis**	**Fourier Basis**
0.05	250	0.049816	0.049839	0.049518	0.049375
	500	0.049955	0.050035	0.050154	0.049880
	1000	0.050112	0.050188	0.049498	0.049729
	1500	0.050000	0.049993	0.049946	0.049729
	2000	0.050186	0.050196	0.049725	0.049643
0.01	250	0.009757	0.009925	0.010495	0.010543
	500	0.009930	0.010094	0.011025	0.010312
	1000	0.010164	0.010005	0.011403	0.010347
	1500	0.010012	0.010058	0.011882	0.010829
	2000	0.009937	0.010081	0.011985	0.011031
0.001	250	0.000987	0.000995	0.001152	0.001227
	500	0.001018	0.001007	0.001372	0.001174
	1000	0.001037	0.000988	0.001553	0.001149
	1500	0.000949	0.000998	0.001699	0.001265
	2000	0.000935	0.000978	0.001766	0.001317
0.0001	250	0.000102	0.000109	0.000126	0.000143
	500	0.000106	9.80E-05	0.000182	0.000134
	1000	0.000108	8.80E-05	0.000213	0.000129
	1500	0.000110	8.60E-05	0.000272	0.000142
	2000	9.20E-05	0.000107	0.000256	0.000167

Table S.3: **Simulation Results of Type I Error Rates of Four Tests Based on 36 SNPs of the Trinity Students Study When the Dominant Genetic Variant Functions Were Used.** The Results of "Basis of both GVF and $\beta(t)$" Were Based on the F-distributed Test Statistics of the Fixed Effect Model (3) and the Results of "Basis of FKST" Were Based on the Variance-Component Functional Kernel Score Test (7) of Mixed Effect Models, for B-spline Basis and Fourier Basis, Respectively. Abbreviation: GVF = Genetic Variant Function, FKST = Functional Kernel Score Tests.

Nominal Level α	**Sample Size**	**Basis of both GVF and $\beta(t)$**		**Basis of FKST**	
		B-spline Basis	**Fourier Basis**	**B-spline Basis**	**Fourier Basis**
0.05	250	0.050123	0.049815	0.049808	0.050213
	500	0.050127	0.050108	0.048721	0.049909
	1000	0.049455	0.049659	0.049098	0.048967
	1500	0.050021	0.049792	0.049307	0.050281
	2000	0.049848	0.049857	0.048997	0.050328
0.01	250	0.009966	0.010106	0.011262	0.009870
	500	0.010004	0.010067	0.011786	0.009889
	1000	0.009845	0.009871	0.012135	0.011218
	1500	0.010037	0.010103	0.012050	0.010544
	2000	0.010006	0.010003	0.012012	0.011956
0.001	250	0.000984	0.001007	0.001380	0.000968
	500	0.001005	0.000967	0.001793	0.000928
	1000	0.000954	0.000986	0.001721	0.001509
	1500	0.000959	0.000926	0.001827	0.001136
	2000	0.000954	0.001005	0.001774	0.001811
0.0001	250	0.000104	0.000093	0.000187	0.000084
	500	0.000100	0.000100	0.000285	8.70E-05
	1000	9.60E-05	0.000108	0.000255	0.000196
	1500	9.20E-05	8.20E-05	0.000268	0.000112
	2000	0.000102	0.000115	0.000274	0.000303

Table S.4: **Simulation Results of Type I Error Rates of Four Tests Based on 36 SNPs of the Trinity Students Study When the Recessive Genetic Variant Functions Were Used.** The Results of "Basis of both GVF and $\beta(t)$" Were Based on the F-distributed Test Statistics of the Fixed Effect Model (3) and the Results of "Basis of FKST" Were Based on the Variance-Component Functional Kernel Score Test (7) of Mixed Effect Models, for B-spline Basis and Fourier Basis, Respectively. Abbreviation: GVF = Genetic Variant Function, FKST = Functional Kernel Score Tests.

Nominal Level α	**Sample Size**	**Basis of both GVF and $\beta(t)$**		**Basis of FKST**	
		B-spline Basis	**Fourier Basis**	**B-spline Basis**	**Fourier Basis**
0.05	250	0.050189	0.050024	0.050066	0.049615
	500	0.050128	0.049874	0.049541	0.048644
	1000	0.050030	0.049886	0.049952	0.049989
	1500	0.049813	0.049779	0.049151	0.048689
	2000	0.049927	0.049869	0.049618	0.048787
0.01	250	0.010093	0.009978	0.009780	0.010726
	500	0.010154	0.010088	0.011453	0.010967
	1000	0.009934	0.009967	0.011567	0.009993
	1500	0.010027	0.009811	0.011415	0.011001
	2000	0.010023	0.009939	0.011747	0.011215
0.001	250	0.001012	0.001042	0.000957	0.001221
	500	0.000977	0.000971	0.001467	0.001480
	1000	0.000971	0.000962	0.001461	0.000990
	1500	0.000965	0.001001	0.001523	0.001393
	2000	0.000994	0.000939	0.001662	0.001470
0.0001	250	0.000115	9.90E-05	7.80E-05	0.000139
	500	9.60E-05	0.000105	0.000175	0.000192
	1000	9.70E-05	9.70E-05	0.000179	9.70E-05
	1500	8.60E-05	0.000105	0.000228	0.000186
	2000	9.90E-05	9.30E-05	0.000262	0.000173

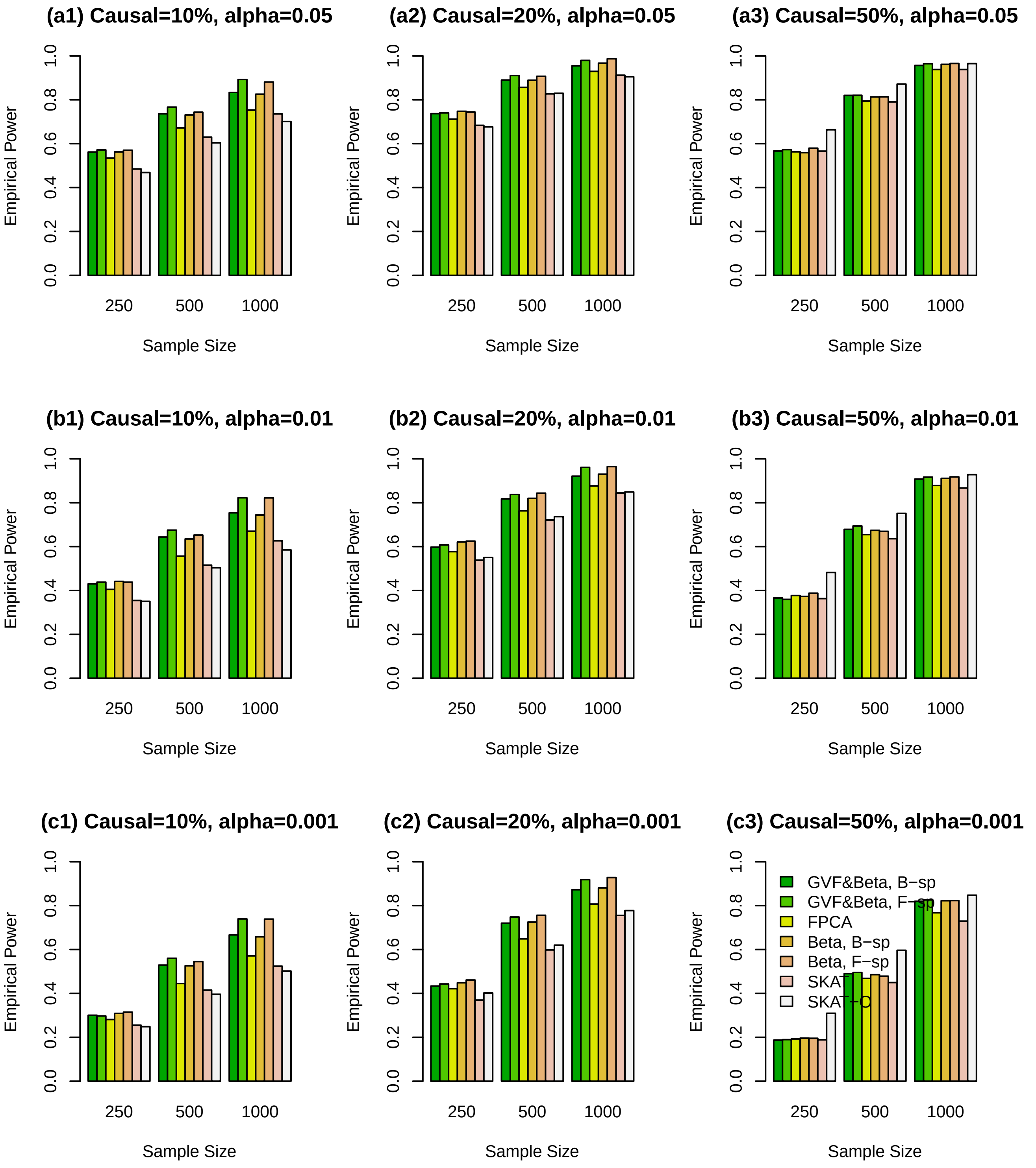

Figure S.1: The Empirical Power of the F-test Statistics of the Fixed Effect Models (3), (4), and (6), and SKAT and SKAT-O Using Rare Variants in Analysis, When Causal Variants Were Both Rare and Common, and All Causal Variants Had Positive Effects. The Simulations Were Based on COSI Sequence Data.

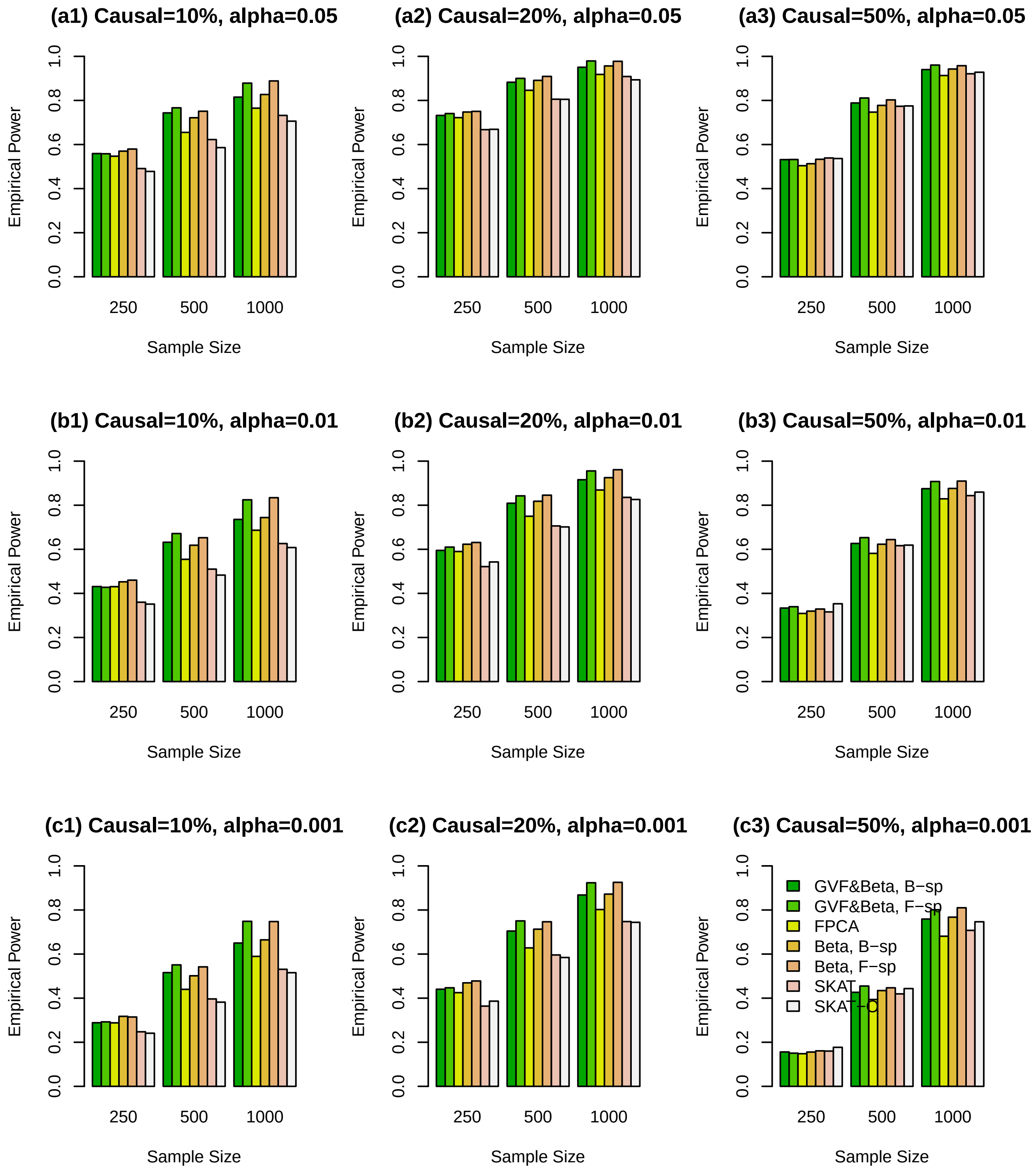

Figure S.2: The Empirical Power of the F-test Statistics of the Fixed Effect Models (3), (4), and (6), and SKAT and SKAT-O Using Rare Variants in Analysis, When Causal Variants Were both Rare and Common, and 20%/80% Causal Variants Had Negative/Positive Effects. The Simulations Were Based on COSI Sequence Data.

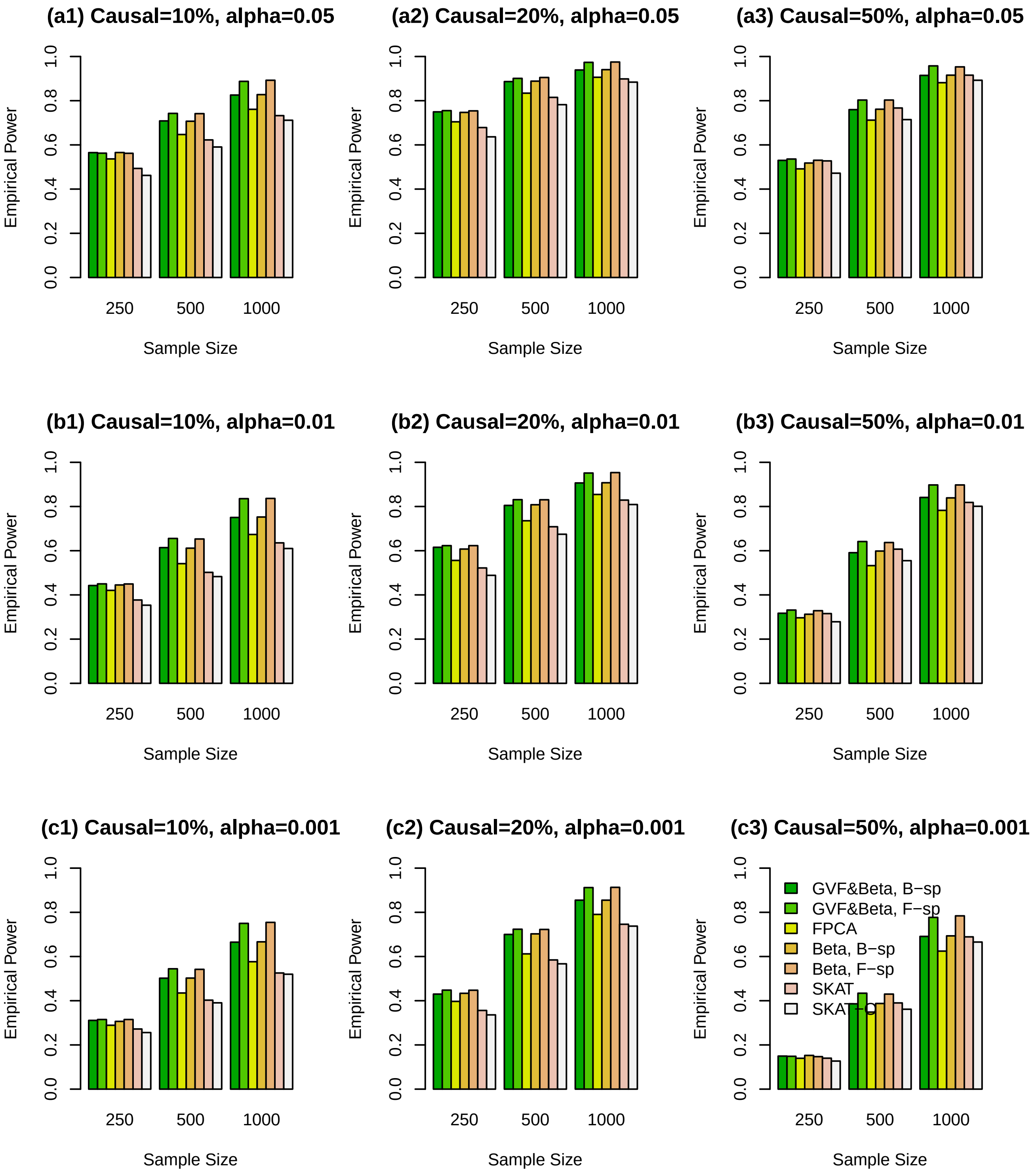

Figure S.3: The Empirical Power of the F-test Statistics of the Fixed Effect Models (3), (4), and (6), and SKAT and SKAT-O Using Rare Variants in Analysis, When Causal Variants Were both Rare and Common, and 50%/50% Causal Variants Had Negative/Positive Effects. The Simulations Were Based on COSI Sequence Data.

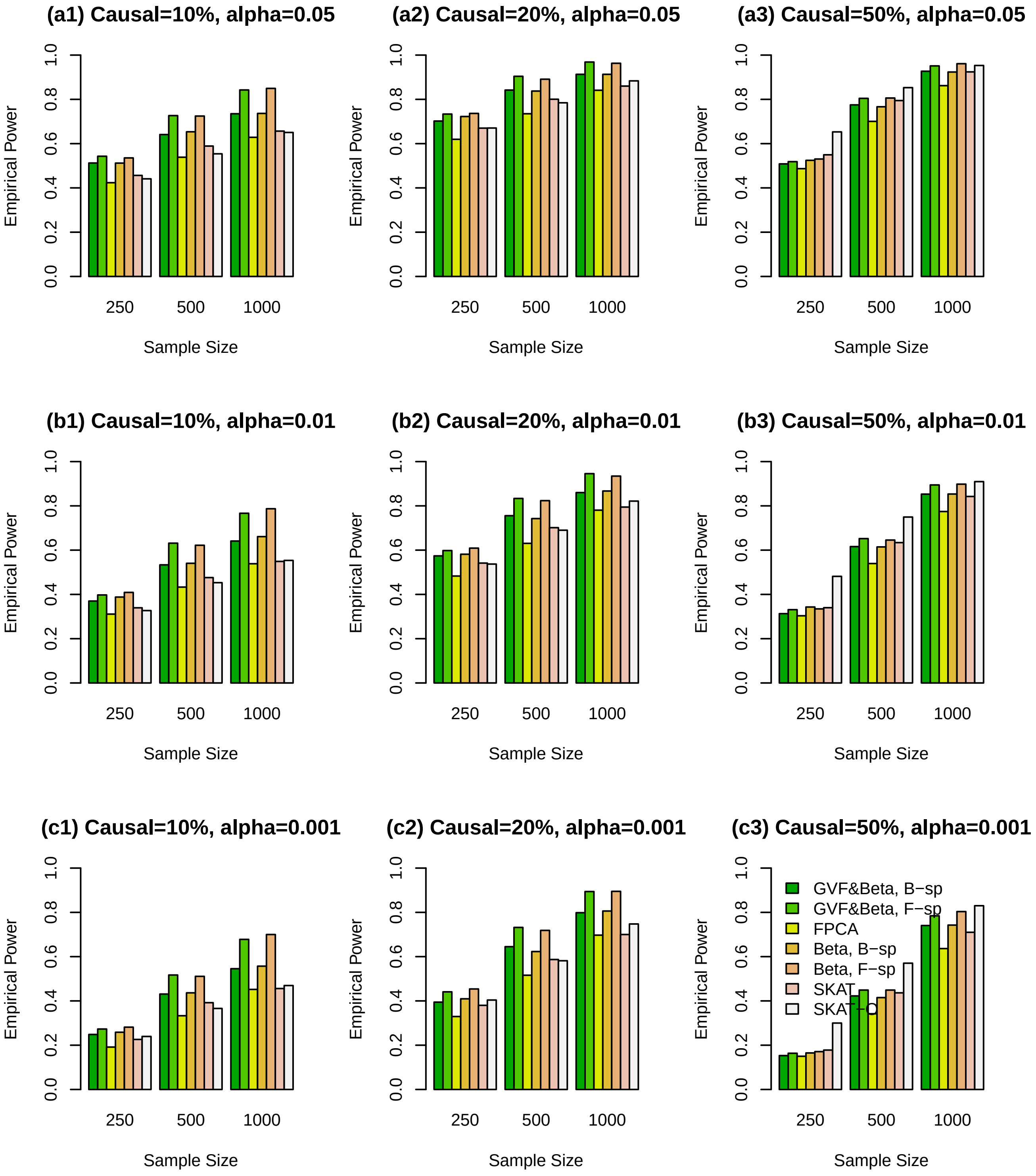

Figure S.4: The Empirical Power of the F-test Statistics of the Fixed Effect Models (3), (4), and (6), and SKAT and SKAT-O Using Both Rare and Common Variants in Analysis, When Causal Variants Were only Rare, and All Causal Variants Had Positive Effects. The Simulations Were Based on COSI Sequence Data.

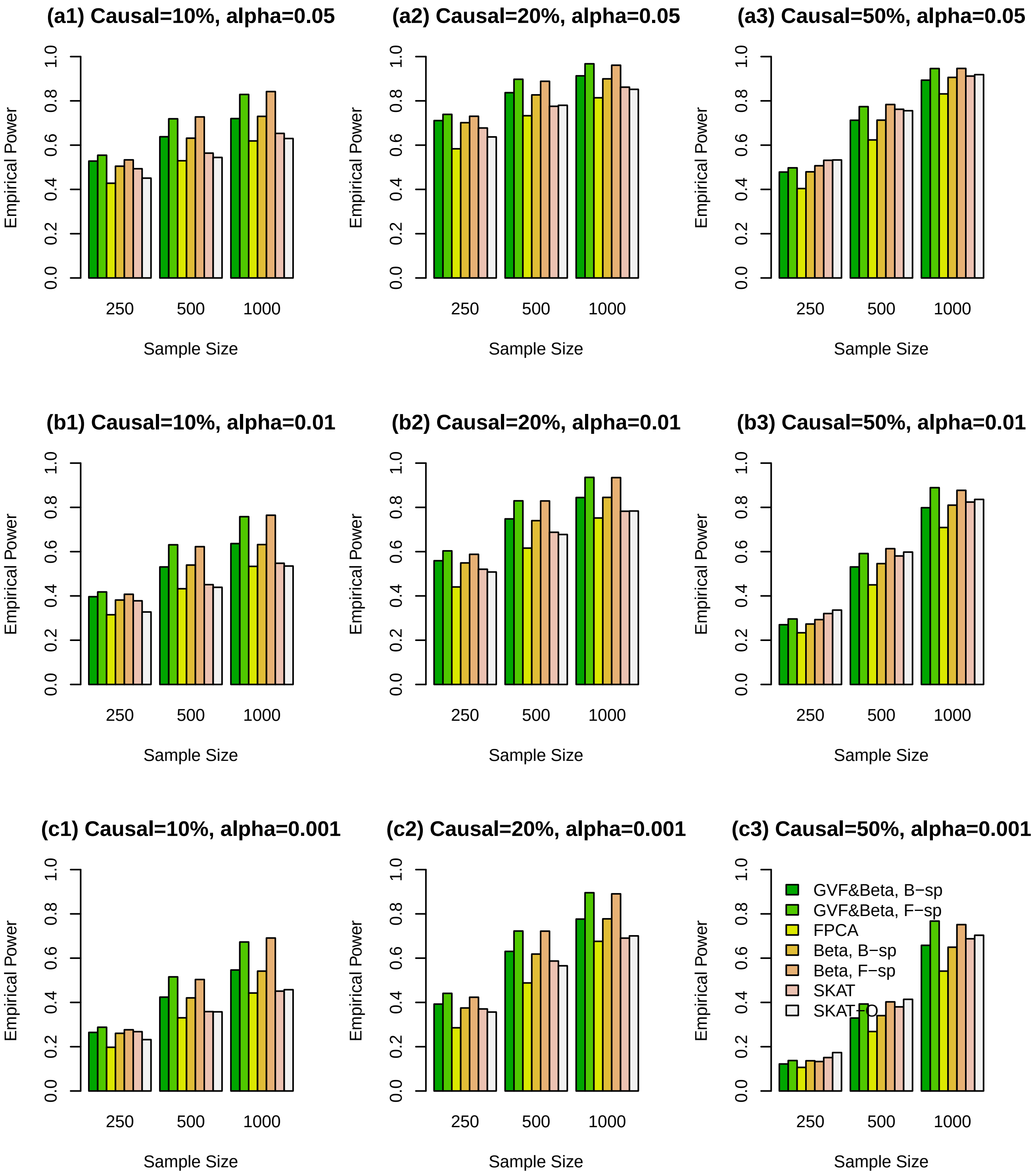

Figure S.5: The Empirical Power of the F-test Statistics of the Fixed Effect Models (3), (4), and (6), and SKAT and SKAT-O Using Both Rare and Common Variants in Analysis, When Causal Variants Were only Rare, and 20%/80% Causal Variants Had Negative/Positive Effects. The Simulations Were Based on COSI Sequence Data.

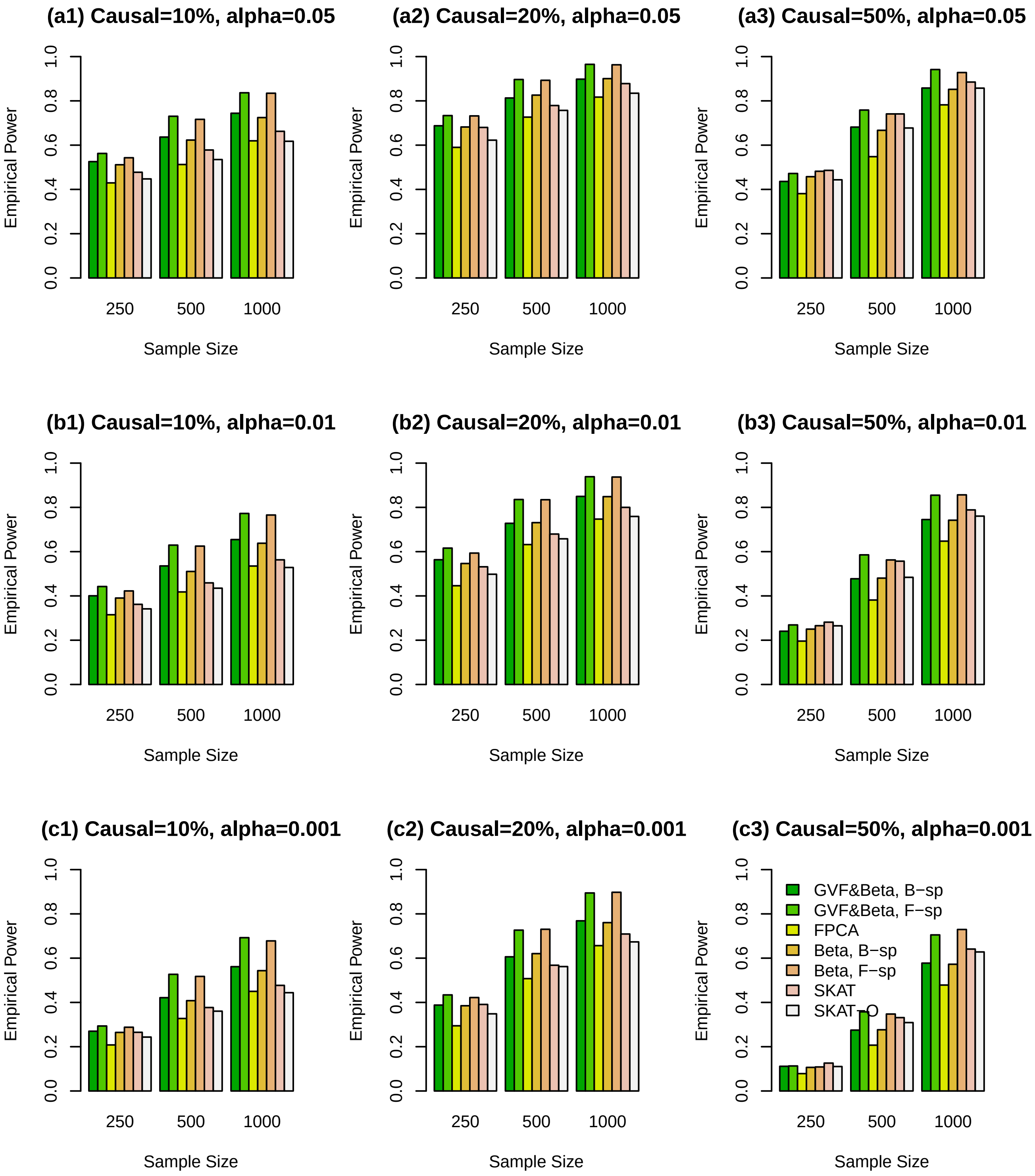

Figure S.6: The Empirical Power of the F-test Statistics of the Fixed Effect Models (3), (4), and (6), and SKAT and SKAT-O Using Both Rare and Common Variants in Analysis, When Causal Variants Were Only Rare, and 50%/50% Causal Variants Had Negative/Positive Effects. The Simulations Were Based on COSI Sequence Data.

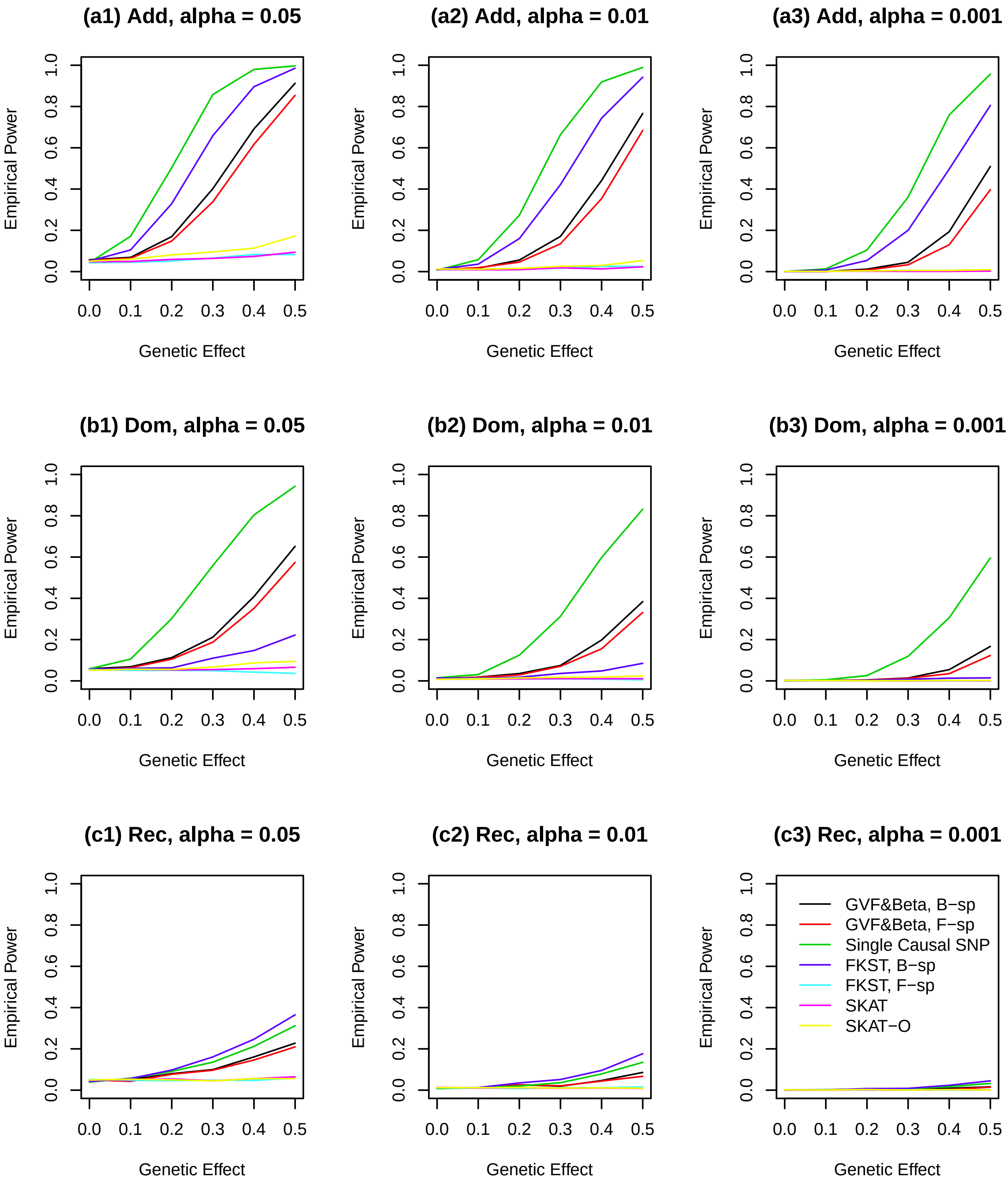

Figure S.7: The Empirical Power of Single Causal SNP Regression Model, Functional Linear Models, and SKAT and SKAT-O Based on Trinity Students Study SNP Data for Modes of Additive, Dominant, and Recessive Inheritance, Respectively, and a Sample Size of 250. Abbreviations: Add = Additive, Dom = Dominant, and Rec = Recessive, Respectively.

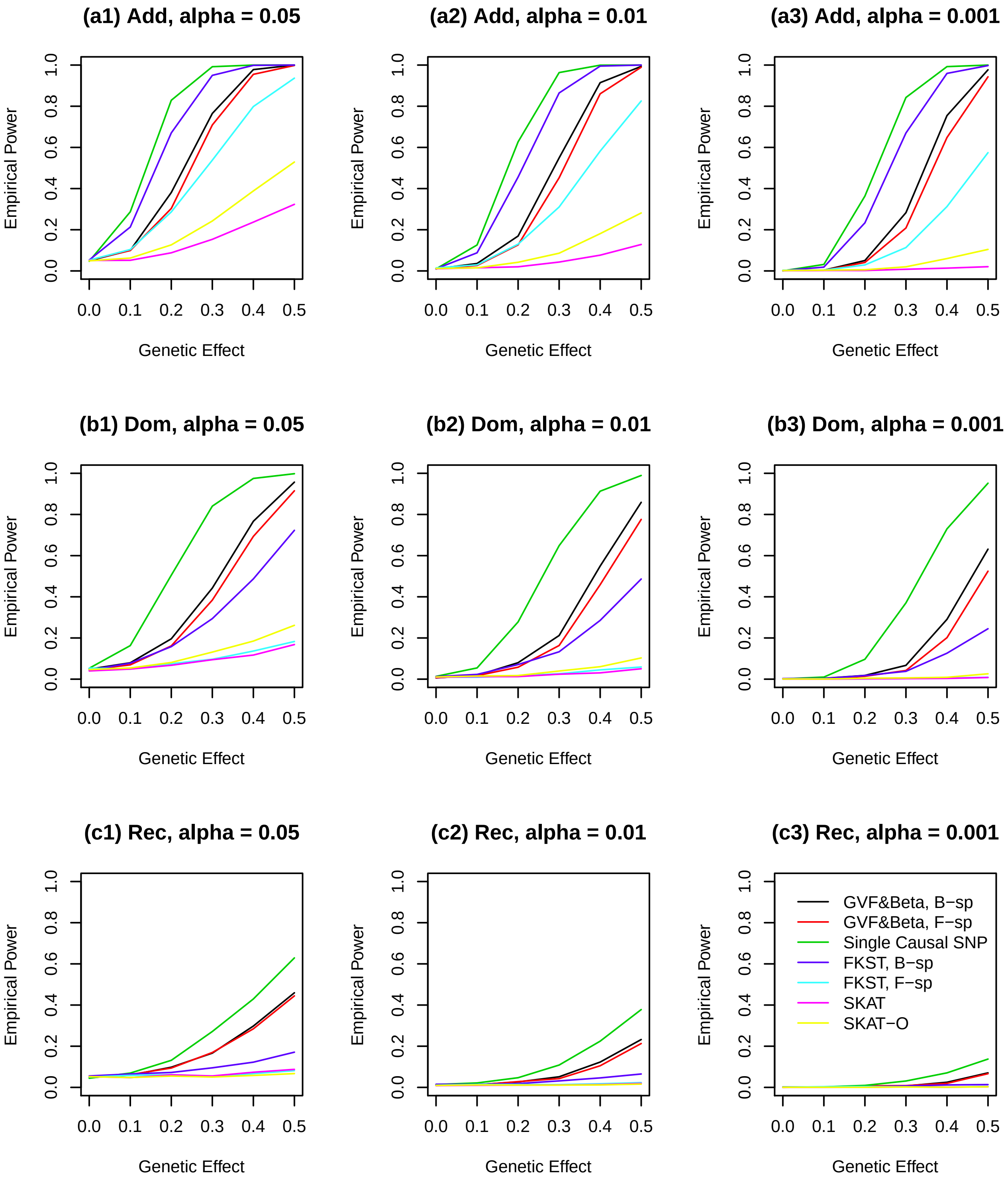

Figure S.8: The Empirical Power of Single Causal SNP Regression Model, Functional Linear Models, and SKAT and SKAT-O Based on Trinity Students Study SNP Data for Modes of Additive, Dominant, and Recessive Inheritance, Respectively, and a Sample Size of 500. Abbreviations: Add = Additive, Dom = Dominant, and Rec = Recessive, Respectively.

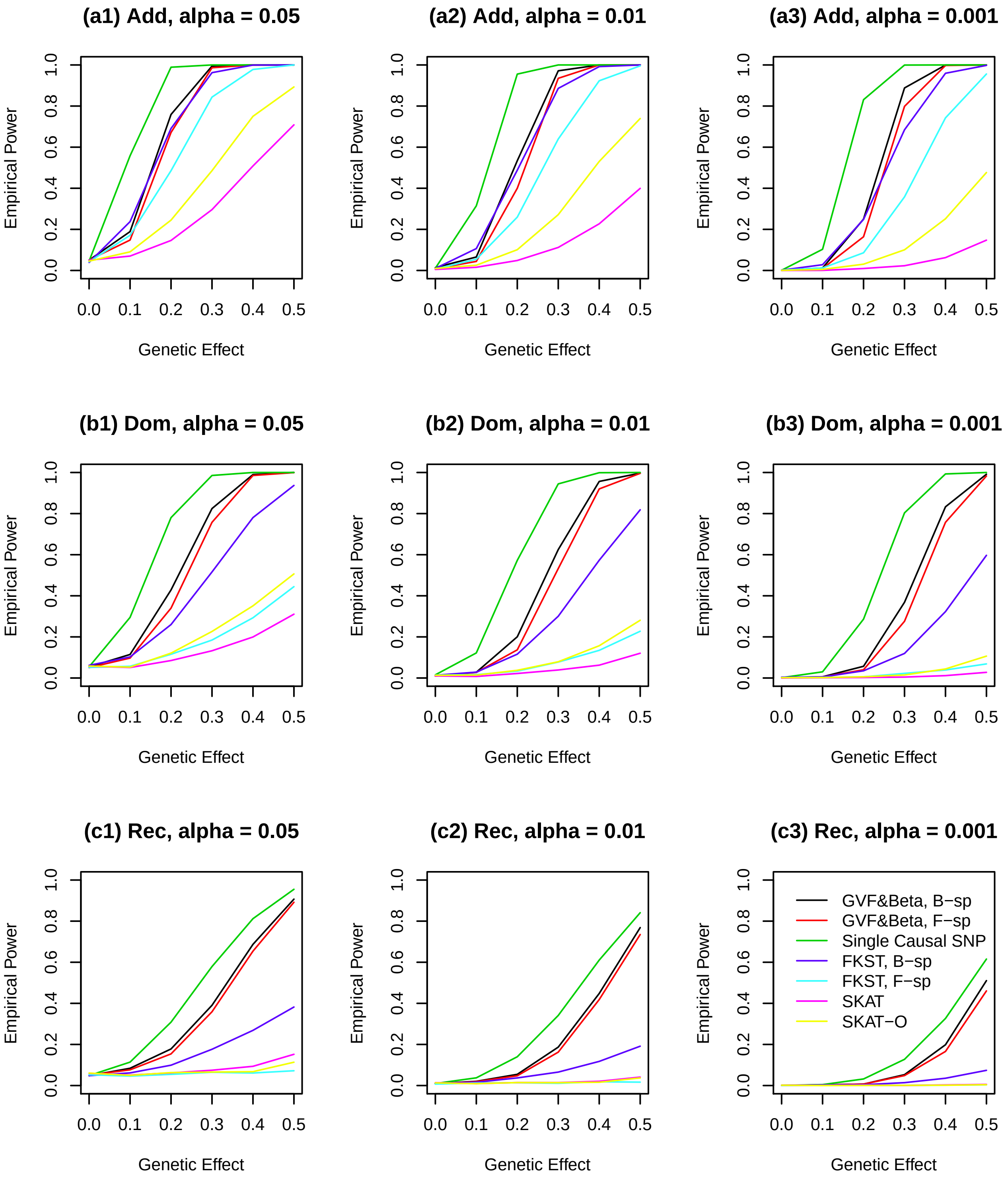

Figure S.9: The Empirical Power of Single Causal SNP Regression Model, Functional Linear Models, and SKAT and SKAT-O Based on Trinity Students Study SNP Data for Modes of Additive, Dominant, and Recessive Inheritance, Respectively, and a Sample Size of 1,000. Abbreviations: Add = Additive, Dom = Dominant, and Rec = Recessive, Respectively.

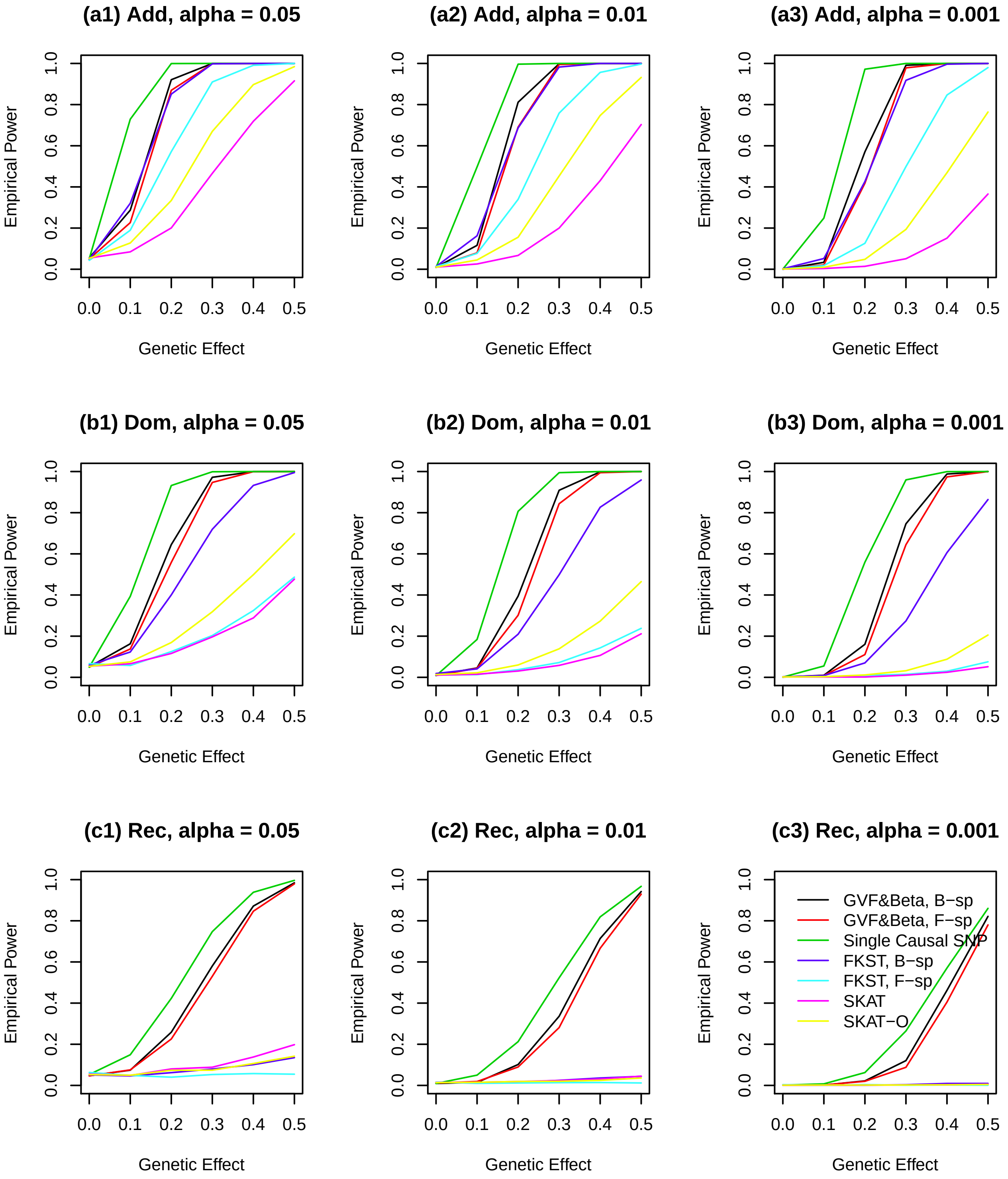

Figure S.10: The Empirical Power of Single Causal SNP Regression Model, Functional Linear Models, and SKAT and SKAT-O Based on Trinity Students Study SNP Data for Modes of Additive, Dominant, and Recessive Inheritance, Respectively, and a Sample Size of 1,500. Abbreviations: Add = Additive, Dom = Dominant, and Rec = Recessive, Respectively.

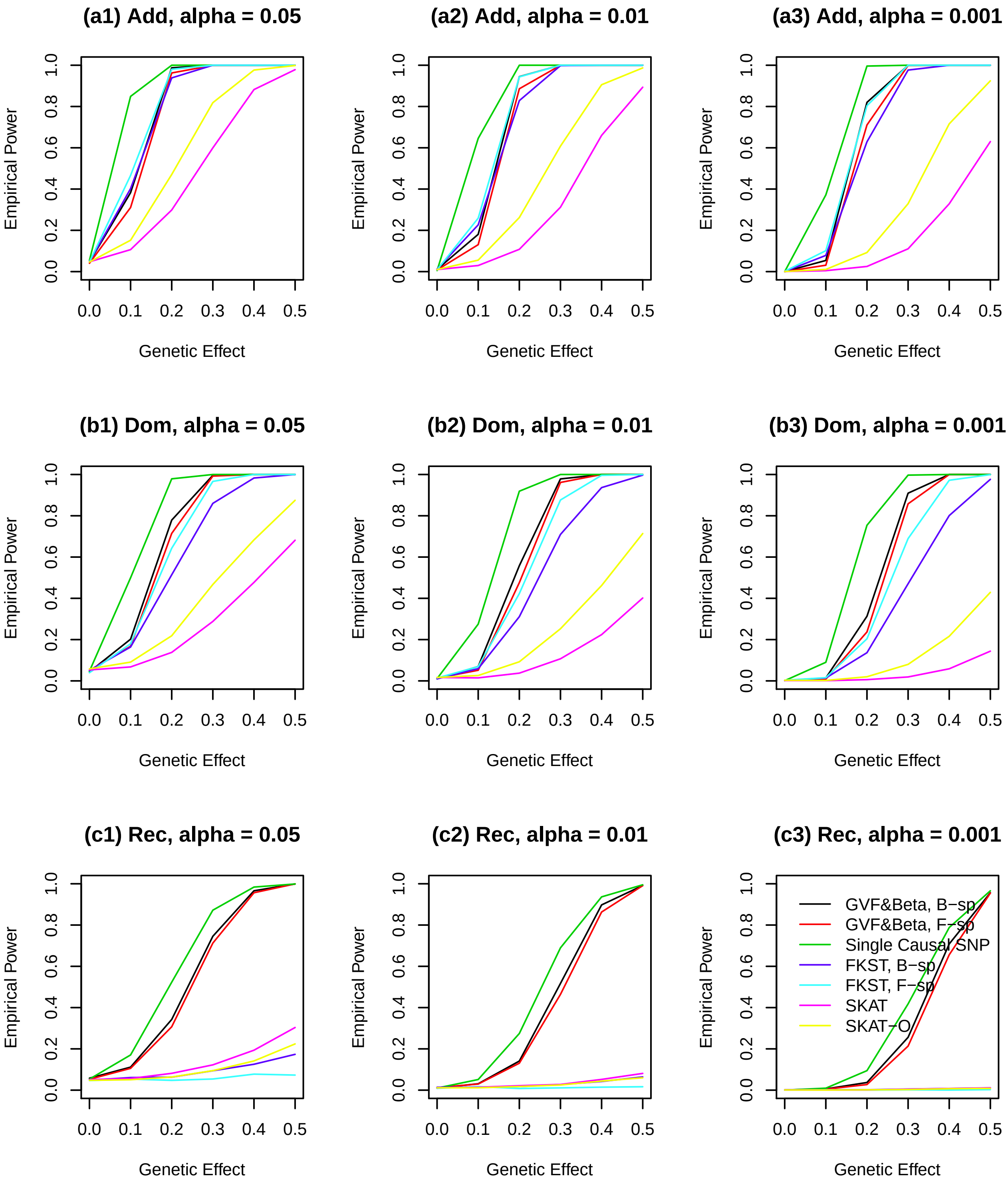

Figure S.11: The Empirical Power of Single Causal SNP Regression Model, Functional Linear Models, and SKAT and SKAT-O Based on Trinity Students Study SNP Data for Modes of Additive, Dominant, and Recessive Inheritance, Respectively, and a Sample Size of 2,000. Abbreviations: Add = Additive, Dom = Dominant, and Rec = Recessive, Respectively.