File S1

Supplementary Materials: A powerful and adaptive association test for rare variants

Wei Pan[1,*] , Junghi Kim[1] , Yiwei Zhang[1] , Xiaotong Shen[2], Peng Wei[3,*]

[1] *Division of Biostatistics, School of Public Health,* [2] *School of Statistics, University of Minnesota, Minneapolis, MN 55455*

[3] *Division of Biostatistics and Human Genetics Center, University of Texas School of Public Health, Houston, TX 77030*

[*] Co-correspondence authors

Correspondence to: Wei Pan

Telephone: (612) 626-2705

Fax: (612) 626-0660

Email: weip@biostat.umn.edu

Address: MMC 303, A460 Mayo,

Division of Biostatistics, School of Public Health,

University of Minnesota,

Minneapolis, Minnesota 55455–0392, USA.

Simulation results with independent RVs

When all the RVs were in linkage equilibrium (with $\rho = 0$ in simulations), all the tests seemed to have satisfactory Type I error rates that were well controlled around the specified nominal level $\alpha = 0.05$ (Table 1). Next we investigated their power properties.

First, we considered a situation with a common association effect: all the 8 causal RVs had an equal odds ratio $OR = \exp(\beta_j) = 2$ associated with the binary trait, which was ideal to the pooled association tests. As shown in Table 2, among the SPU tests, when the number of non-associated RVs was small, the SPU(1) (i.e. Sum) test was most powerful; however, as the number of non-associated RVs increased, SPU(3) became most powerful. This observation is in agreement with Basu and Pan (2011), showing the deteriorating performance of the Sum (and other similar pooled association tests) in the presence of many non-associated RVs. The reason for better performance of the SPU(3) test, or more generally of a SPU(γ) test with a large value of γ, in the presence of many non-associated RVs is the following: as the number of non-associated RVs increased, more and more components of the score vector U were just noises; using a larger value of γ corresponds to down-weighting those smaller, and likely noisy, components of U. However, there is a trade-off: a too large value of γ will also down-weight and thus diminish those smaller signals in U; an extreme is that, the SPU(∞) only uses the largest component of $|U|$, ignoring the signals contained in $|U|$ for other causal RVs. We also note that, although the SPU(2) (i.e. SSU) test performed well, it was always less powerful than SPU(3), and their power difference was large in the presence of many non-associated RVs. A SPU(γ) test with a large value of γ, e.g. $\gamma > 8$, performed similarly to SPU(∞).

Among the adaptive tests, the KBAC test was most powerful with no or few non-associated RVs, but overall the aSum+ test performed best because, as the Sum test, the aSum+ test used the common OR assumption while having the capability of RV selection to deal with non-associated RVs. Interestingly, as the number of the non-associated RVs increased, the aSPU test gradually caught up with power almost the same as that of the aSum+ test. The EREC test was also high powered with no or few non-associated RVs, but not in the presence of many non-associated RVs. In particular, as the number of non-associated RVs increased, the aSPU test was much more powerful than the EREC, PWST, KBAC and other adaptive tests (except aSum+). It is noted that the aSPU test maintained high power close to the winner in the class of the SPU tests.

Second, we considered a more realistic situation: there was no common association strength but only a common association direction among the 8 causal RVs; the ORs for the 8 causal RVs were randomly drawn from a uniform distribution between 1 and 3, $U(1,3)$, in each simulation (Table 3). Many of the earlier conclusions held. For example, among the SPU tests, the SPU(1) (i.e. Sum) test was most powerful in the absence of non-associated RVs; otherwise, the SPU(3) test was most powerful, though several other SPU(γ) tests with $\gamma > 3$ were similarly powerful in the presence of 128 non-associated RVs. Again the SPU(16), SPU(32) and SPU(∞) behaved similarly. However, there were also some deviations. Overall, the aSum+ test was most powerful only for ≤ 32 non-associated RVs; otherwise, the aSPU test was most powerful. For ≥ 64 non-associated RVs, the aSSU test performed as well as the aSum+ test, much more powerful than the KBAC, aSum, PWST and EREC tests, though much less powerful than the aSPU test.

Third, we examined a case with both varying association strengths and varying association directions for the 8 causal RVs (Table 4). As expected, the SPU(1) (i.e. Sum) test performed terribly. Among the SPU tests, with a smaller number (≤ 32) of non-associated RVs, the SPU(2) (i.e. SSU) test was most powerful; otherwise the SPU(4) was the winner. Among the adaptive tests, with only a smaller number of non-associated RVs, the SKAT was most powerful, closely followed by the PWST, EREC, aSPU and aSSU tests; otherwise, the aSPU and aSSU tests performed similarly and were winners. Although the aSum+ test dramatically improved over the Sum test, it still had deteriorating performance in the presence of many non-associated RVs as compared to the aSSU test. Surprisingly, although the aSum2d was designed to take account of both positive and negative associations, it did not perform better than the aSum+ test. The reason was that, it was much difficult to detect negative associations for RVs, unlike for CVs as shown in Pan et al (2011): the aSum- test aiming to detect negative associations was consistently low powered across all the scenarios (not shown). It is also noted that both BhGLM and KBAC did not perform well.

Fourth, we investigated a more extreme case: there was only one causal RV with a large effect with $OR = 5$, for which case the SPU(∞) was expected to perform best due to its selecting only one RV with the largest $|U_j|$ (Table 5). Interestingly, any SPU(γ) test with $\gamma > 4$ performed similarly to each other, and were winners. Again the aSPU test maintained high power close to the winners in the SPU test family, and had a clear edge over other adaptive tests, especially in the presence of many non-associated RVs; the aSSU test also performed well with no or only few non-associated RVs.

Simulation results with higher signficance levels

We also considered using higher nominal significance levels α. The simulation set-ups were the same as those presented in the paper; in particular, the RVs were correlated with $\rho = 0.9$ being used for latent variables. As for the GAW17 data analysis, we started with $B = 10^3$, then gradually increased B: if an estimated p-value was less than $5/B$, we increased B to ten times of its current value to re-estimate the p-value, and the process was repeated until no estimated p-value was less than $5/B$; we used B up to $B = 10^5$.

Table 6 shows the estimated Type I error rates with $k - k_1 = 96$ null RVs at various values of α based on 10^5 simulation replicates. It is clear that the SPU and aSPU tests could control their Type I error rates satisfactorily. Table 7 shows the estimated power based on 10^3 simulation replicates, again with $k - k_1 = 96$ null RVs but $k_1 = 8$ causal RVs with their association ORs randomly drawn from $U(1, 2)$. As for $\alpha = 0.05$, with more significant α levels the aSPU test was more powerful than SKAT and SKAT-O; more interestingly, the advantage of the aSPU test was more dramatic with a more significant α.

Simulation results with higher signficance levels and a covariate

We considered a new simulation set-up with a single covariate. The correlated RVs were generated as before with $\rho = 0.9$ being used for latent variables. A single covariate was generated from a normal distribution $N(0, 10)$; it was associated with the binary trait with regression coefficient 1 in the logistic regression model. We used 10^5 simulation replicates. As for the GAW17 data analysis, we started with $B = 10^3$, then gradually increased B: if an estimated p-value was less than $5/B$, we increased B to ten times of its current value to re-estimate the p-value, and the process was repeated until no estimated p-value was less than $5/B$; we used B up to $B = 10^5$. We used the permutation method based on permuting residuals to calculate the p-values for the SPU and aSPU tests. As shown in Table 7, the SPU and aSPU tests could control Type I error rates satisfactorily at the various values of the significance level α.

Table 1: Empirical Type I error rates of various tests for the cases with 8 RVs plus various numbers of non-associated RVs; all RVs were independent; all results were based on 1000 simulation replicates.

	# non-associated RVs						
Test	0	8	16	32	64	96	128
UminP	.031	.024	.021	.009	.008	.011	.009
SPU(1)	.045	.051	.056	.059	.046	.042	.049
SPU(2)	.047	.047	.052	.040	.042	.036	.043
SPU(3)	.043	.038	.046	.031	.033	.030	.033
SPU(4)	.042	.045	.053	.029	.033	.036	.029
SPU(5)	.042	.033	.048	.027	.040	.044	.031
SPU(6)	.042	.039	.051	.030	.039	.033	.029
SPU(7)	.041	.033	.049	.030	.037	.043	.031
SPU(8)	.042	.037	.049	.033	.041	.042	.031
SPU(16)	.041	.036	.046	.028	.042	.040	.030
SPU(32)	.041	.035	.046	.028	.041	.041	.032
SPU(∞)	.040	.035	.046	.028	.041	.041	.033
aSPU	.046	.056	.054	.042	.042	.049	.048
aSum+	.052	.055	.054	.041	.041	.041	.070
aSum2d	.053	.052	.056	.047	.039	.043	.033
aSSU	.050	.047	.059	.040	.041	.054	.055
KBAC	.060	.051	.056	.047	.046	.043	.050
aSum	.054	.046	.060	.046	.047	.049	.049
PWST	.061	.051	.053	.046	.042	.047	.057
EREC	.062	.048	.056	.044	.039	.045	.051
BhGLM	.052	.056	.056	.059	.043	.042	.055
SKAT	.060	.047	.056	.050	.050	.050	.055

Table 2: Empirical power of various tests for the cases with 8 causal RVs with ORs=(2, 2, 2, 2, 2, 2, 2, 2); all RVs were independent; all results were based on 1000 simulation replicates. The highest powered non-adaptive and adaptive tests in each case are bold-faced.

	# non-associated RVs						
Test	0	8	16	32	64	96	128
UminP	.421	.281	.230	.156	.115	.076	.076
SPU(1)	**.953**	**.790**	.654	.461	.269	.211	.183
SPU(2)	.742	.673	.615	.516	.418	.306	.278
SPU(3)	.769	.718	**.666**	**.557**	**.472**	**.391**	**.361**
SPU(4)	.632	.594	.550	.460	.411	.352	.324
SPU(5)	.629	.588	.531	.444	.415	.357	.337
SPU(6)	.573	.529	.488	.397	.383	.318	.295
SPU(7)	.574	.535	.487	.397	.383	.333	.301
SPU(8)	.546	.515	.465	.380	.368	.299	.283
SPU(16)	.514	.482	.427	.354	.346	.286	.270
SPU(32)	.508	.470	.419	.349	.337	.281	.265
SPU(∞)	.506	.464	.419	.347	.338	.279	.265
aSPU	.914	.767	.697	.571	.458	.381	.351
aSum+	.912	.834	**.776**	**.661**	**.522**	**.396**	**.354**
aSum2d	.867	.758	.683	.557	.415	.307	.174
aSSU	.632	.582	.526	.442	.387	.293	.281
KBAC	**.953**	**.858**	.765	.596	.392	.225	.183
aSum	.937	.765	.644	.485	.348	.246	.221
PWST	.764	.641	.558	.413	.319	.229	.195
EREC	.915	.805	.734	.571	.411	.299	.265
BhGLM	.952	.808	.671	.469	.285	.215	.184
SKAT	.773	.676	.615	.512	.413	.284	.275

Table 3: Empirical power of various tests for the cases with 8 causal RVs with ORs randomly chosen from U(1,3); all RVs were independent; all results were based on 1000 simulation replicates. The highest powered non-adaptive and adaptive tests are bold-faced.

	# non-associated RVs						
Test	0	8	16	32	64	96	128
UminP	.552	.427	.336	.281	.199	.178	.146
SPU(1)	**.900**	.749	.593	.442	.270	.232	.177
SPU(2)	.795	.737	.655	.607	.501	.428	.357
SPU(3)	.818	**.765**	**.684**	**.645**	**.571**	**.537**	.437
SPU(4)	.730	.688	.620	.599	.542	.492	.442
SPU(5)	.732	.682	.618	.577	.548	.510	**.456**
SPU(6)	.696	.652	.587	.560	.527	.475	.446
SPU(7)	.688	.639	.579	.544	.518	.475	.449
SPU(8)	.670	.630	.570	.533	.507	.464	.434
SPU(16)	.648	.609	.552	.516	.487	.443	.412
SPU(32)	.646	.599	.544	.507	.483	.429	.407
SPU(∞)	.641	.597	.542	.507	.480	.423	.408
aSPU	.879	.783	.692	.643	**.565**	**.523**	**.451**
aSum+	.905	**.846**	**.760**	**.670**	.517	.465	.377
aSum2d	.863	.778	.693	.580	.408	.366	.193
aSSU	.721	.665	.594	.535	.483	.433	.379
KBAC	**.925**	.837	.719	.593	.330	.255	.171
aSum	.892	.746	.613	.483	.329	.277	.217
PWST	.785	.682	.583	.464	.344	.287	.230
EREC	.902	.816	.701	.578	.408	.358	.273
BhGLM	.905	.766	.619	.465	.280	.235	.179
SKAT	.798	.722	.641	.562	.467	.401	.318

Table 4: Empirical power of various tests for the cases with 8 causal RVs with ORs=(3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2); all RVs were independent; all results were based on 1000 simulation replicates. The highest powered non-adaptive and adaptive tests are boldfaced.

	# non-associated RVs						
Test	0	8	16	32	64	96	128
UminP	.486	.351	.295	.208	.171	.145	.133
SPU(1)	.276	.190	.142	.101	.072	.051	.068
SPU(2)	**.797**	**.690**	**.638**	**.513**	.409	.336	.292
SPU(3)	.603	.515	.495	.418	.347	.307	.288
SPU(4)	.706	.602	.569	.478	**.452**	**.403**	**.380**
SPU(5)	.634	.522	.498	.430	.404	.372	.343
SPU(6)	.674	.565	.535	.455	.444	.399	.369
SPU(7)	.624	.527	.492	.423	.411	.380	.351
SPU(8)	.655	.551	.512	.430	.437	.390	.356
SPU(16)	.637	.532	.494	.421	.420	.384	.348
SPU(32)	.628	.526	.488	.413	.419	.380	.347
SPU(∞)	.626	.522	.485	.413	.418	.377	.349
aSPU	.718	.598	.564	.469	**.421**	.360	**.339**
aSum+	.689	.562	.509	.377	.283	.228	.196
aSum2d	.694	.523	.470	.330	.240	.201	.106
aSSU	.692	.597	.557	.484	.418	**.368**	.320
KBAC	.699	.485	.389	.250	.163	.116	.096
aSum	.670	.505	.402	.284	.241	.173	.137
PWST	.784	.645	.579	.421	.328	.256	.197
EREC	.769	.630	.518	.376	.277	.215	.202
BhGLM	.490	.303	.219	.134	.080	.060	.071
SKAT	**.810**	**.685**	**.625**	**.504**	.404	.318	.291

Table 5: Empirical power of various tests for the cases with only one causal RV with OR=5; all RVs were independent; all results were based on 1000 simulation replicates. The empirical power for all the tests was around 0.850 in the absence of non-associated RVs. The highest powered non-adaptive and adaptive tests are bold-faced.

	# non-associated RVs					
Test	8	16	32	64	96	128
UminP	.696	.629	.556	.496	.479	.461
SPU(1)	.365	.263	.160	.096	.088	.086
SPU(2)	.710	.664	.580	.520	.470	.427
SPU(3)	.717	.664	.634	.585	.569	.541
SPU(4)	.731	.697	.653	.633	.605	.574
SPU(5)	.727	.692	.654	.627	.622	.593
SPU(6)	**.732**	**.701**	.651	.637	.620	.598
SPU(7)	.731	.696	.652	.634	.621	.596
SPU(8)	.730	.699	**.656**	.634	.623	**.600**
SPU(16)	.729	.700	.653	.638	.624	.594
SPU(32)	.730	.700	.652	.638	.626	.594
SPU(∞)	.730	.700	.651	**.640**	**.627**	.594
aSPU	.707	.683	**.645**	**.615**	**.592**	**.571**
aSum+	.731	.627	.512	.329	.278	.256
aSum2d	.668	.561	.432	.263	.202	.187
aSSU	**.736**	**.685**	.628	.561	.518	.481
KBAC	.629	.483	.330	.193	.128	.103
aSum	.447	.314	.215	.152	.130	.126
PWST	.665	.533	.405	.280	.211	.174
EREC	.685	.545	.424	.272	.197	.184
BhGLM	.480	.385	.257	.157	.127	.121
SKAT	.713	.638	.544	.436	.379	.333

Table 6: Empirical Type I error rates based on 10^5 simulation replicates with $96 + 8$ null RVs and $\rho = 0.9$. For comparison, the results for resampling-based SKAT and SKAT-O and asymptotics-based SKAT and SKAT-O (A-SKAT and A-SKAT-O) are also included.

α	SPU(1)	SPU(2)	SPU(3)	SPU(4)	SPU(5)	SPU(6)	SPU(7)	SPU(8)	SPU(15)	SPU(16)	SPU(31)	SPU(32)	SPU(∞)	aSPU	A-SKAT	SKAT	A-SKAT-O	SKAT-O
0.05	0.04875	0.04964	0.04976	0.05028	0.04970	0.04956	0.04931	0.04961	0.04968	0.04957	0.04967	0.04956	0.03359	0.04862	0.05050	0.05070	0.05300	0.05119
0.01	0.00904	0.00935	0.00917	0.00937	0.00922	0.00924	0.00933	0.00913	0.00925	0.00915	0.00926	0.00915	0.00587	0.00882	0.00978	0.01023	0.01085	0.00982
0.005	0.00362	0.00396	0.00421	0.00442	0.00431	0.00446	0.00419	0.00427	0.00408	0.00416	0.00409	0.00416	0.00274	0.00445	0.00520	0.00418	0.00547	0.00442
0.001	0.00086	0.00084	0.00082	0.00088	0.00089	0.00085	0.00086	0.00087	0.00078	0.00085	0.00078	0.00085	0.00056	0.00083	0.00104	0.00095	0.00117	0.00083
0.0005	0.00034	0.00037	0.00039	0.00051	0.00048	0.00042	0.00035	0.00036	0.00035	0.00035	0.00035	0.00035	0.00025	0.00045	0.00060	0.00041	0.00050	0.00038

Table 7: Empirical power based on 10^3 simulation replicates with 8 causal RVs, 96 null RVs and $\rho = 0.9$. For comparison, the results for resampling-based SKAT and SKAT-O and asymptotics-based SKAT and SKAT-O (A-SKAT and A-SKAT-O) are also included.

α	SPU(1)	SPU(2)	SPU(3)	SPU(4)	SPU(5)	SPU(6)	SPU(7)	SPU(8)	SPU(15)	SPU(16)	SPU(31)	SPU(32)	SPU(∞)	aSPU	A-SKAT	SKAT	A-SKAT-O	SKAT-O
0.05	0.29700	0.83300	0.81900	0.86000	0.84600	0.85400	0.84600	0.84800	0.83800	0.84100	0.83800	0.84100	0.79600	0.84400	0.79600	0.79900	0.76800	0.76200
0.01	0.14600	0.70300	0.72600	0.77400	0.76100	0.75700	0.74700	0.74100	0.72800	0.72500	0.72600	0.72400	0.66700	0.73300	0.65000	0.64500	0.59600	0.57200
0.005	0.09900	0.65500	0.68200	0.73400	0.72100	0.71300	0.70000	0.69100	0.65900	0.66200	0.65800	0.66100	0.58700	0.68600	0.58400	0.55200	0.53400	0.51200
0.001	0.04500	0.52200	0.58600	0.61900	0.60400	0.58700	0.57000	0.55600	0.52700	0.52600	0.52500	0.52500	0.46200	0.57100	0.45300	0.44600	0.42000	0.39800

Table 8: Empirical Type I error rates based on 10^5 simulation replicates with a covariate, $96 + 8$ null RVs and $\rho = 0.9$.

α	SPU(1)	SPU(2)	SPU(3)	SPU(4)	SPU(5)	SPU(6)	SPU(7)	SPU(8)	SPU(15)	SPU(16)	SPU(31)	SPU(32)	SPU(∞)	aSPU
0.05	0.05024	0.05037	0.05007	0.04938	0.05014	0.04949	0.04939	0.04917	0.04895	0.04886	0.04888	0.04883	0.04889	0.04738
0.01	0.00993	0.00961	0.00999	0.00966	0.00982	0.00993	0.00994	0.01017	0.01001	0.00996	0.00979	0.00980	0.00985	0.00865
0.005	0.00460	0.00459	0.00494	0.00487	0.00490	0.00478	0.00477	0.00472	0.00457	0.00458	0.00450	0.00451	0.00451	0.00471
0.001	0.00092	0.00102	0.00108	0.00101	0.00121	0.00107	0.00107	0.00112	0.00114	0.00112	0.00113	0.00113	0.00111	0.00093
0.0005	0.00037	0.00049	0.00043	0.00052	0.00053	0.00049	0.00050	0.00048	0.00051	0.00053	0.00051	0.00051	0.00051	0.00040