AJHG, Volume *86*

Supplemental Data

Pooled Association Tests for Rare Variants in Exon-Resequencing Studies

Alkes L. Price, Gregory V. Kryukov, Paul I. W. de Bakker, Shaun M. Purcell, Jeff Staples, Lee-Jen Wei, and Shamil R. Sunyaev

Supplemental Tables and Figures

Supplemental Tables

Table S1. Power of Various Approaches Using Quantitative Phenotypes

We display results for T1 (1% allele frequency threshold), T5 (5% allele frequency threshold), WE (weighted) and VT (variable threshold) analyses for various values of s and δ. Results are displayed for α=0.001 and α=0.05 significance levels, based on 10,000 independent simulations. The best-performing method(s) are indicated in bold. We note that of 54 pairwise comparisons between VT and one other method, VT performs worse in 6/54 comparisons, with an average reduction in power of 0.003 in those 6 comparisons.

	T1	T5	WE	VT
s=0.0001,δ=0.125, α=0.001	0.031	0.081	0.030	**0.109**
s=0.0001,δ=0.125, α=0.05	0.299	0.386	0.310	**0.432**
s=0.0001,δ=0.25, α=0.001	0.230	0.368	0.245	**0.487**
s=0.0001,δ=0.25, α=0.05	0.674	0.707	0.735	**0.803**
s=0.0001,δ=0.5, α=0.001	0.787	0.778	0.888	**0.926**
s=0.0001,δ=0.5, α=0.05	0.954	0.926	0.983	**0.985**
s=0.001,δ=0.125, α=0.001	0.017	**0.036**	0.013	0.029
s=0.001,δ=0.125, α=0.05	0.233	**0.253**	0.220	0.249
s=0.001,δ=0.25, α=0.001	0.137	0.182	0.098	**0.204**
s=0.001,δ=0.25, α=0.05	0.547	0.503	0.543	**0.600**
s=0.001,δ=0.5, α=0.001	0.569	0.493	0.595	**0.697**
s=0.001,δ=0.5, α=0.05	0.893	0.792	0.941	**0.948**
s=0.01,δ=0.125, α=0.001	0.001	0.001	0.001	**0.002**
s=0.01,δ=0.125, α=0.05	0.061	0.055	**0.066**	**0.066**
s=0.01,δ=0.25, α=0.001	0.002	0.002	**0.003**	0.002
s=0.01,δ=0.25, α=0.05	0.075	0.068	**0.083**	0.080
s=0.01,δ=0.5, α=0.001	0.004	0.003	**0.006**	0.005
s=0.01,δ=0.5, α=0.05	0.111	0.090	**0.132**	0.128

Table S2

(a) Power of various approaches using quantitative phenotypes, for a simulation in which phenotypes are independent of selection coefficient. The fraction of functional mutations, i.e. mutations that affect phenotype, was equal to the fraction of mutations with selection coefficient above 10^{-3} (48%), however the functionality of each mutation was independent of the selection coefficient. We display results for T1 (1% allele frequency threshold), T5 (5% allele frequency threshold), WE (weighted), and VT (variable threshold) analyses for α=0.001 and α=0.05 significance levels, based on 10,000 independent simulations.

	T1	T5	WE	VT
α=0.001	0.099	0.226	0.156	0.405
α=0.05	0.404	0.501	0.520	0.631

(b) Value of incorporation of PolyPhen-2 predictions, as a function of the (known) distribution of PolyPhen-2 scores associated to either deep or shallow multiple sequence alignments. Alignments with less than 10 sequences (after PolyPhen-2 filtering for very close homologs) were classified as "shallow". Alignments with 100 or more sequences were classified as "deep". We display results for T1 (1% allele frequency threshold), T5 (5% allele frequency threshold), WE (weighted), VT (variable threshold), VTP (VT plus PolyPhen-2), VTP-deep (VT plus PolyPhen-2, deep alignments) and VTP-shallow (VT plus PolyPhen-2, shallow alignments) analyses for α=0.001 and α=0.05 significance levels, based on 10,000 independent simulations. All columns except VTP-deep and VTP-shallow are identical to Table 1. As expected, VTP-deep outperforms VTP-shallow, but even VTP-shallow is substantially better than VT.

	T1	T5	WE	VT	VTP	VTP-hi	VTP-lo
α=0.001	0.137	0.182	0.098	0.204	0.259	0.269	0.230
α=0.05	0.547	0.503	0.543	0.600	0.686	0.696	0.639

Table S3. Power of Various Approaches Using Dichotomous Phenotypes

We display results for T1 (1% allele frequency threshold), T5 (5% allele frequency threshold), WE (weighted) and VT (variable threshold) analyses for various values of s and δ. Results are displayed for for α=0.001 and α=0.05 significance levels, based on 10,000 independent simulations. The best-performing method(s) are indicated in bold. We note that of 54 pairwise comparisons between VT and one other method, VT performs worse in 6/54 comparisons, with an average reduction in power of 0.002 in those 6 comparisons.

	T1	T5	WE	VT
s=0.0001,δ=0.125, α=0.001	0.021	0.062	0.029	**0.087**
s=0.0001,δ=0.125, α=0.05	0.266	0.346	0.302	**0.399**
s=0.0001,δ=0.25, α=0.001	0.156	0.319	0.219	**0.439**
s=0.0001,δ=0.25, α=0.05	0.616	0.664	0.706	**0.773**
s=0.0001,δ=0.5, α=0.001	0.691	0.722	0.836	**0.885**
s=0.0001,δ=0.5, α=0.05	0.933	0.906	0.984	**0.985**
s=0.001,δ=0.125, α=0.001	0.012	**0.026**	0.011	0.021
s=0.001,δ=0.125, α=0.05	0.207	0.227	0.203	**0.233**
s=0.001,δ=0.25, α=0.001	0.089	0.150	0.078	**0.161**
s=0.001,δ=0.25, α=0.05	0.482	0.458	0.488	**0.533**
s=0.001,δ=0.5, α=0.001	0.443	0.427	0.474	**0.583**
s=0.001,δ=0.5, α=0.05	0.843	0.744	0.890	**0.902**
s=0.01,δ=0.125, α=0.001	0.001	**0.002**	0.001	0.001
s=0.01,δ=0.125, α=0.05	0.060	0.059	0.061	**0.063**
s=0.01,δ=0.25, α=0.001	**0.003**	**0.003**	0.002	0.002
s=0.01,δ=0.25, α=0.05	0.072	0.066	**0.077**	0.074
s=0.01,δ=0.5, α=0.001	0.003	0.003	**0.004**	**0.004**
s=0.01,δ=0.5, α=0.05	0.099	0.084	**0.111**	0.109

Table S4

(a) Effect of PolyPhen-2 on power of various approaches using quantitative phenotypes. We display results for T1 (1% allele frequency threshold), T5 (5% allele frequency threshold), WE (weighted), and VT (variable threshold) analyses (each with and without PolyPhen-2) for α=0.001 and α=0.05 significance levels, based on 10,000 independent simulations.

	T1		T5		WE		VT	
	-PolyPhen	+PolyPhen	-PolyPhen	+PolyPhen	-PolyPhen	+PolyPhen	-PolyPhen	+PolyPhen
α=0.001	0.14	0.21	0.18	0.24	0.10	0.15	0.20	0.26
α=0.05	0.55	0.69	0.50	0.62	0.54	0.65	0.60	0.69

(b) Effect of PolyPhen-2 on power of various approaches using dichotomous phenotypes. We display results for T1 (1% allele frequency threshold), T5 (5% allele frequency threshold), WE (weighted), and VT (variable threshold) analyses (each with and without PolyPhen-2) for α=0.001 and α=0.05 significance levels, based on 10,000 independent simulations.

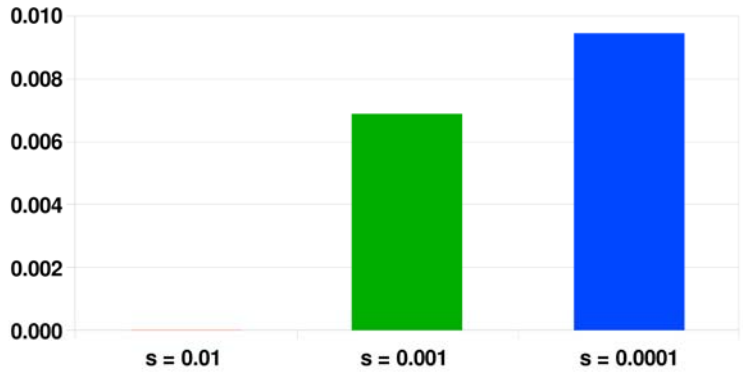
	T1		T5		WE		VT	
	-PolyPhen	+PolyPhen	-PolyPhen	+PolyPhen	-PolyPhen	+PolyPhen	-PolyPhen	+PolyPhen
α=0.001	0.09	0.16	0.15	0.19	0.08	0.12	0.16	0.21
α=0.05	0.48	0.62	0.46	0.56	0.49	0.60	0.53	0.63

Supplemental Figures

Figure S1. Prevalence of Goldilocks Alleles in Simulations at Different Values of Selection Parameter *s*

We consider s=0.01 ($\log_{10}s$=-2), s=0.001 ($\log_{10}s$=-3), or s=0.0001 ($\log_{10}s$=-4). We plot (a) the fraction of all functional SNPs that have frequency between 0.5% and 2%, and (b) the probability that a gene has at least one functional SNP with frequency between 0.5% and 2%, as a function of s.

(a)



(b)