S4 Capitalizing on the presence of both risk and protective variants

In the presence of both risk and protective variants, a more powerful strategy might be to combine the signals from both types of variants. Our framework can easily be extended to accommodate both types of signals, as follows. Let

$$S_+ = \sum_{k=0}^{N_r} \sum_{k'>k} -n_k^{k'} \log[p(k,k')],$$

as defined in the Methods Section, where for a variant k is the number of occurrences of the minor allele in controls, and k' is the number of occurrences of the minor allele in cases. We define S_- similarly, with the roles of cases and controls reversed. More precisely, if we now denote by k and k' the counts in *cases* and *controls*, respectively, then:

$$S_- = \sum_{k=0}^{N_r} \sum_{k'>k} -n_k^{k'} \log[p(k,k')].$$

Then the statistic $S_c = S_+ + S_-$ can be used to perform a test that combines signals from both risk and protective variants. We have compared this combined test ($\alpha = 0.05$) with the two-sided test based on the max-statistic, i.e., $\max(S_+, S_-)$ ($\alpha = 0.05$), and the results are in Supplementary Table S-2. As can be observed, there is indeed substantial gain in power when both types of variants are present, e.g., when there are 20 risk and 20 protective variants with total PAR of 0.05, the power for the combined test is 0.844, while that of the test based on the max statistic is 0.478. However, this advantage does come at a loss in power when only one type of variants is present, e.g., the power for the combined test vs. max-statistics test is 0.400 vs. 0.718 when only risk variants are present, and the total PAR is 0.05.

			PAR=0.03		PAR=0.05	
Sim. Model	#Risk	#Protective	$R_{\max}$	R_C	$R_{\max}$	R_C
1	20	0	0.334	0.162	0.718	0.400
		5	0.262	0.222	0.624	0.608
		10	0.214	0.268	0.446	0.646
		20	0.172	0.378	0.478	0.844

Table S-2: Power Estimates for the max-statistic test ($R_{\max}$) vs. the combined test (R_C) at $\alpha = 0.05$, when both risk and protective variants may be present in a region of interest. Simulation model 1 corresponds to the neutral Wright-Fisher model. The total sample size is 1000 cases and controls.