

Questions on Wu et al (2011) *Am J Hum Genet* 89:82-93

Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test.

1. The authors discuss the potential contributions of rare variants to missing heritability. Why else is it important to study rare variants in human genetic association studies?
2. Many rare variant association tests, score the total count or presence versus absence of rare alleles in a gene of interest and then apply a univariate test. What are some weaknesses of these approaches?
3. Consider the following thought experiment: imagine a large sample with equal numbers of cases and controls and a series of variants that each present in exactly two individuals in your sample. What are the possible splits of case-control counts for each of these variants? How likely is each split under the null, when the variants are not associated with the trait? What about in a situation where these variants increase disease risk? What about in a situation where some variants increase and some decrease risk?
4. The paper suggests aggregating rare variants by gene or in sliding windows across the genome. What are some of the advantages and challenges in each approach?
5. The proposed SKAT approach test the hypothesis that $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_N = 0$ using a single degree of freedom. How does it accomplish this? Why is that important?
6. Write out the definition of the kernel function $K(\mathbf{G}_i, \mathbf{G}_j)$. What does it measure?
7. Adjusting for relatedness and population structure is important in any genetic association test. How would you combine the linear mixed model approach described by Kang and colleagues (2010) to calculate single variant score statistics that are adjusted for population structure with this Sequence Kernel Association Test?
8. Empirically evaluating type 1 error for very small α can be computationally challenging. How did the authors increase the computational efficiency of their simulations? What computational savings do you expect from their approach?
9. When does SKAT outperform previous burden tests? When does it underperform?
10. What struck you most about the paper?