Title: Rare Variant Analysis Using the Multi-Kernel Sequence Kernel Association Test

Abstract:

Rare variants are thought to have a strong effect on complex traits and can now be detected using sequencing technology. Since testing strategies for common variants are underpowered for rare variants, region based testing for the cumulative effect of a group of rare variants on a trait has been established as an effective strategy and several tests have been developed.

Two important practical challenges have emerged for researchers interested in rare variant analysis. First, it is difficult to choose which test to use in practice, since each of the developed methods is powerful under different structures of the genotype/phenotype relationship. Second, it is unclear which group of variants within a region should be tested, e.g. all rare variants or only the nonsynonymous variants. The answer to both problems depends on knowing the underlying trait architecture and which variants are causal. This information is unavailable and would preclude need for analysis yet resolution of these challenges is keenly needed.

Therefore, we propose a new pragmatic testing approach that searches across a range of particular tests and variant grouping strategies. We demonstrate that several popular rare variant tests including burden based tests, the C-alpha test, and others are all special cases of the sequence kernel association test (SKAT). SKAT is a similarity based test wherein pair-wise similarity in trait value between subjects is compared to pair-wise similarity in the rare variant genotypes between subjects as measured through a kernel function. Choosing a particular test is equivalent to choosing a particular kernel function. Similarly, we can show that the choice of which group of variants to test also reduces to a kernel choice problem. Thus, we develop the Multi-Kernel SKAT (MK-SKAT), a statistical framework based on perturbation, which tests across a range of kernel choices that correspond to a range of rare variant tests and variant groupings while protecting type I error.

Real data analyses and simulations show that MK-SKAT tends to have high power across all settings, losing a little bit of power when compared to the optimal test for a particular scenario but having much greater power than poor choices of test or grouping strategy. Additional features of the method are its ability to effectively control for confounding variables, as well as its computational efficiency.