Assessing methods for assigning SNPs to genes for gene-based tests of association with common variants

Carolina Alvarez[1], Allison Hainline[2], Andrew Beck[3], Brian Greco[4], Nathan Tintle[5*]

[1]Department of Statistics, Florida International University, Miami, Florida [2]Department of Statistics, Baylor University, Waco, Texas [3]Department of Mathematics, Loyola University, Chicago, Illinois [4] Department of Mathematics and Statistics, Grinnell College, Grinnell Iowa [5] Department of Mathematics, Statistics and Computer Science, Dordt College, Sioux Center, Iowa

Gene-based tests are being applied with increasing frequency to common SNPs (MAF>5%) measured by SNP microarrays in GWAS as an alternative to single-marker tests. While it is standard to apply gene-based tests to all SNPs within the gene, two main approaches are used to assign intergenic SNPs to genes before testing. Option 1 is a window-based approach that assigns all intragenic SNPs within a certain distance of the start and stop positions of the gene and an LD approach that assigns all intergenic SNPs in LD (at some threshold) with SNPs inside the gene. In this poster we present the results of simulation studies in which we have explored factors related to the power of gene-based tests of association. In addition to expected trends relating to increased sample size, increased relative risk and increased MAF (all of which increased power), a substantial loss in power was observed from the inclusion of non-causal variants. This important result applied to all tests, though certain methods showed modestly better resistance to the inclusion of non-causal variants. The inclusion of non-causal variants is mitigated to some extent by LD structure within the non-causal variants (LD within non-causal variants mitigates some of the observed power loss due to the inclusion of non-causal variants). LD structure further mitigates power loss when the non-causal SNPs included in the test are in LD with causal variants. Due to the sensitivity of most tests to the inclusion of non-causal SNPs, we conclude that the window approach is typically contributing to substantial power loss when utilized in gene-based tests of association.