camh
Centre for Addiction and Mental Health

Centre for Addiction and Mental Health
250 College Street
Toronto, Ontario
January 19, 2016

Canadian Institute for Health Research
Institutes for Genetics
Ottawa, Ontario

To Whom it May Concern:

With the recent radical decrease in genotyping cost, genome wide association studies (GWAS) have become a commonly used tool for the hypothesis free discovery of disease associated loci. Loci identified through GWAS may be validated in biological systems and become potential targets of clinical intervention. Crucial to these studies, which incorporate million of simultaneous statistical tests with a complex depedency structure, is the decision whether to accept or reject the null hypothesis H_0 of no association. As the number of simultaneous statistical tests increases, so too does the number of false positives identified. This issue is denoted as the "multiple comparissons problem" and methods arising from this area are becoming critical to differentiating between statistical coincidence and biological relevance.

One such method which has been gaining traction among the statistical genetics community is "stratefied false disovery rate" (sFDR), which incorporates biologically relevant strata before correcting association P values to control the false discovery rate $(\frac{\text{Falsely rejected Null Hypothesis}}{\text{Total rejected Null Hyptothesis}})$ to an acceptable level. Despite much controversy, the applicability of this methodology to the human gneome remains and open question in the field. It is unknown whether or notthis method adequetely controls the false disovery rate in a system as complex as the human geome, and it is unknown whether or not this method is demonstrably superior to false disovery rate (FDR) in terms of true asocciated loci that it uncovers.

Under my supervision, Christopher Cole will investigate this open problem through a simulation study. Using the reference panel collected by the 1000 Genomes Consortium and the UK10K Consortium, Christopher will simulate novel human genomes while maintaining linkage disequilibrium (LD) structure with HAPGEN2. He will then investigate the ability of various multiple testing correction methodologies to differentiate between random noise and simulated biological signal. He will additionally research common scenarios where researchers use sFDR and FDR and examine how these situations could impact the methodology's accuracy. Christopher will also question under which circumstances (LD, minor allele frequence of variants, etc.) the statistical assumptions and ability to control the FDR are conserved.

Cordially yours,

Joanne Knight, PhD