Research Use Statement for Application for Genomic Data from NIAGADS

Please limit to 2,200 characters max. The statement should include the following components:

- Objectives of the proposed research;
- Study design;
- Analysis plan, including the phenotypic characteristics that will be evaluated in association with genetic variants

Research Use Statement:

While GWAS studies are an efficient way of identifying loci associated with disease, more work is typically necessary to identify the underlying gene and process driving the association. The purpose of this analysis is to identify genes responsible for Alzheimer's risk by incorporating brain gene expression data from external cohorts along with GWAS summary statistics from *IGAP Rare Variant Summary Statistics* (Kunkle et al. (2019)).

The first approach will involve using eQTL associations generated from multiple post-mortem brain expression profiling cohorts from the AMP-AD and CommonMind Consortia to aid in the interpretation of the summary statistics from Kunkle et al. (2019). We will use cerebral cortical eQTL generated from more than 1,400 individuals for this purpose. This will be done with two different approaches: (1) colocalization analysis, which seeks to identify genes whose eQTL association patterns correspond with the risk association patterns for the disease of interest (Alzheimer's), and (2) transcriptwide-association study (TWAS), which seeks to "impute" gene expression into a larger GWAS cohort from models built in eQTL cohorts in order to identify genes whose genetic component of expression is associated with disease or protection. In both cases, we will use the genomewide summary statistics from Kunkle et al. (2019) to develop gene-level summary statistics such as posterior probability of colocalization[1] or z-scores and p-values using a MetaXcan approach[2] to identify genes putatively associated with risk for Alzheimer's disease.

The second approach will use these summary statistics as a form of validate for genes predicted to be risk variants from network[3] and lineage-based[4] models built from gene expression and/or proteomics data available from the AMP-AD Consortium to show that our predicted genes are truly associated with disease risk.

1. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10,** e1004383 (2014).
2. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *bioRxiv* 45260 (2017). doi:10.1101/045260
3. Mukherjee, S. *et al.* Identifying and ranking potential driver genes of Alzheimer's Disease using multi-view evidence aggregation. *bioRxiv* 534305 (2019). doi:10.1101/534305
4. Mukherjee, S. *et al*. Molecular estimation of late-onset Alzheimer's disease stage. *In prep.* (2019).

Non-Technical Summary for Application for Genomic Data from NIAGADS

Investigators will provide a non-technical summary of their proposed research. If

the project is approved, this statement will be publicly available for lay audiences to read the purpose and objectives of the research. Please limit to 1,100 characters.

While GWAS studies are an efficient way of identifying loci associated with disease, more work is typically necessary to identify the underlying gene and process driving the association. We propose to identify genes associated with Alzheimer's disease by incorporate complementary brain gene expression data from external cohorts such as those available from the AMP-AD and CommonMind Consortia to draw inference about the genes responsible for risk to Alzheimer's disease as identified in the *IGAP Rare Variant Summary Statistics* (Kunkle et al. (2019)) data set. We also propose to use these Alzheimer's disease risk associations to validate risk gene predictions made from network and lineage-based models of gene expression.