Robert A. Sweet, MD
Request for NIAGADS Dataset NG00076

Dataset to be analyzed
We are requesting NIAGADs dataset NG00076- "ADGC case-control summary statistics on 7050 samples not included in the IGAP-2013 discovery stage".

Variable(s) to be requested and analyzed
Association statistics (AD vs Control) and allele frequencies.

Analyses to be performed (1 ½ pages)
We are currently funded by NIA and approved by ADGC to conduct the following analyses of the genetic correlates of psychotic symptoms, defined as the occurrence of delusions or hallucinations, in individuals with Alzheimer Disease (AD+Psychosis, AD+P):

> **Aim 1. To perform an unbiased genome-wide association study (GWAS) of AD+P risk**. Some subjects will have existing genome-wide SNP array data (**including the ADGC ADC data sets**), all others will be genotyped on the Illumina GSA chip. All subjects will be imputed to a common set of SNPs. *We hypothesize that multiple SNPs are associated with AD+P risk.*
>
> **Aim 2. To evaluate Schizophrenia and AD risk alleles for association with AD+P.** Identification of common and rare variants contributing to schizophrenia and AD risk continues at a rapid pace. We will use these discoveries to expand our current schizophrenia and AD polygenic risk scores and test for their independent and interactive associations with AD+P risk in the cohort described in **Aim 1**. *We hypothesize that polygenic* schizophrenia *risk score is inversely, and polygenic AD risk score directly, related to increasing AD+P risk.*

We have now completed our initial analysis of Aim 1 in ~7400 AD subjects. We are requesting access to the new data set to facilitate our Aim 2 analysis. We will evaluate the genetic score and components of that score for their association with AD+P (versus AD without Psychosis). Our ADGC-approved analysis plan for Aim 2 is as follows:

Aim 2 Analysis: Sz and AD risk scores will utilize the most current available consortium data (PGC, ADGC, IGAP). AD+P risk score calculations will be generated using and confirmed in separate subsets of subjects. Weighted and unweighted genetic scores will be produced from the count of risk alleles and weight being a function of the SNP impact on risk (odds ratio). We will also investigate performance of a novel method for producing scores, which involves techniques of multivariate analysis, mixed models, and the statistical concept of shrinkage estimators. Ancestry for cases and controls will be determined by using SNPs with a call rate > 99.5% that are independent (pairwise $r^2 < 0.1$) and determined using SpectralGEM. **Aim 2 Power Analysis:** Determining the association between a score and AD+P status involves a 1 degree-of-freedom test. Multiple testing, beyond the 3 scores, is not involved. We anticipate predicting a greater proportion of AD+P risk than in our preliminary studies with the inclusion of greater numbers of SNPs and larger sample sizes. Nevertheless, our conservative assumption of 4000 AD+P and 4000 AD-P subjects passing QC will provide power > 0.8 to detect effects similar to, or even smaller than, those observed in our published studies.

Members of the proposed analysis teams: Robert A. Sweet MD, Bernie Devlin PhD, M. Ilyas Kamboh PhD, Lambertus Klei PhD, Mary Ann DeMichele-Sweet PhD, Elise Weamer MPH.