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
Objectives
The genetic architecture of Alzheimer disease (AD) is largely unknown. AD is a heterogeneous condition that presents with variable age of onset and speed of progression. GWA studies have revealed associations between many genetic loci and AD. It has been suggested that multiple loci working in concert contribute to AD phenotype through epistatic interactions.

We hypothesize that the genetic markers within the interacting loci identified by GWASs can be used to develop risk prediction models for AD.

Testing of this hypothesis will require access to the genotype-phenotype databases that contain genetic and clinical information such as age, gender, age of diagnosis and death of the individuals with and without AD.

Study design
We will partition the subjects into the phenotypic categories and apply analytical, machine-learning (ML) methods to develop risk-prediction algorithms and evaluate them using Monte-Carlo cross-validation.

Analysis plan
First, we will divide the cohort by age of onset, rate of progression, length of survival and any other recorded phenotype. After the inspection of statistical distributions, we will do manual classification or k-means clustering and then apply ML for each category.

ML methods will include classification trees, random forests, support vector machines, neural networks, latent class analysis and boosting algorithms, which iteratively weight mis-classified observations.

The discriminatory performance of the algorithms will be assessed using Monte-Carlo cross-validation. These algorithms will be trained on a random subset of the data and tested on the remainder, with results summarized by the C-statistic. Multiple iterations will be used to develop empirical confidence intervals around the validation C-statistic. We will also compare the performance of standard logistic regression and generalizations like elastic net regression to assess whether the ML methods are worth their associated computational demands.

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.

Non-Technical Summary
Much data have been collected in recent years in attempt to explain the genetic underpinnings of Alzheimer disease. However, it remains a puzzle. In general, understanding the relationship between disease phenotype and its underlying genotype is one of the difficult problems in modern biomedicine. Among the more promising strategies for addressing it is Systems Biology. This approach is distinguished by its holistic views rather than more traditional reductionist methods of investigating one part of a system at a time. In adapting Systems Biology tactics application of Machine Learning methods is the next logical step.