***Article***: The Alzheimer’s Disease Sequencing Project (ADSP): Data Production, Management, and Availability

Partch, A. B., DeStefano, A. L., Cantwell, L., Faber, K., Feolo, M., Stine, A., ... & Bis, J. C. (2016). THE ALZHEIMER’S DISEASE SEQUENCING PROJECT (ADSP): DATA PRODUCTION, MANAGEMENT, AND AVAILABILITY. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *12*(7), P648.

<http://www.alzheimersanddementia.com/article/S1552-5260(16)31605-3/fulltext>

Suggestions After Presentation:

None. Seems reasonable.

***Article Presentation*** Tuesday, March 28th:

**Goal:** Code their pipeline for generating their data and conducting their analysis

This data set is available in multiple locations (both publically and though request). This article does not specifically lay out the pipelines used, but the steps are known in the field in general. We plan to use this previous knowledge to re-run this data with proper documentation and replication. For generating our data we plan on using bash (this has already been accomplished) for statistical analysis we plan to use R, but are open to using Python or PILNK depending on what has the most utility. There will be three main components (sets of code) of our project/presentation: Data Processing, Statistical Analysis, and Graphing.

Summary (Ali’s very basic understanding of the article): *FRMD6* has been shown to be implicated in the development AD. The authors took seven SNPs within *FRMD6* are specifically associated with AD and did a whole brain surface analysis to see which brain regions’ degradation was associated with each of the aforementioned SNPs. 1,239 Caucasian participants were MRI scanned and genotyped. Then they used general linear modeling in order to find associations with various variables and the SNPs in question. Of the 7 SNPs, 4 were significantly associated with brain atrophy (p < .05). Rs12885443 was associated with cortical thickness in the left temporal lobe. Polygenic risk was scored with the 4 SNPs (sum of number of risk alleles, not controlling for effect size, associated with atrophy in multiple locations). “ Further analysis of the *FRMD6* gene revealed that rs4898698 was significantly associated with CSF Aβ1-42 (*p*=0.00063) and global cortical Aβ load (measured by [18F]Florbetapir PET) (*p*=0.015; Fig. 3). In addition, rs4898698 is associated with subcortical brain structures, volumes of the pallidum (*p*=0.019) and amygdala (*p*=0.038), in 9,333 normal individuals (ENIGMA).” These SNPs were, in fact, associated with AD and changes in the brain structure. More research is warranted.

***Proposed Division of Tasks***:

**1) Data Processing**

Impute genetic data

Code is pretty much done and just needs outlined into a single file

Even if we used all of the HPC resources at Iowa State, we do not have enough time in the next month to generate all of the brain data for 1200 subjects.

So since I am working on this in the background, we can replicate the analyses in the subjects that I currently have done.

**2) Statistical Analysis**

GLMs

SNP \* age

SNP \* gender

SNP \* years of education

SNP \* intracranial volume (ICV)

SNP \* MRI field strength

Examine SNPs associated with total brain atrophy

Replicate analyses of 4 SNPs individually and

Replicate analyses of sum of number of risk alleles

**3) Graphing**

Graph association of SNPs with Amyloid beta values - manhattan plot

Graph association of 4 SNPs associated with brain atrophy (p < .05).

Graph of a “polygenic risk score” (Additive number of risk alleles each subject has) for association with

***Project Guidelines Via Class Repository Instructions:***

~~Form group by March 3rd~~

~~Select article with available data repository~~

**Document the study**

1. Downloading, inspecting, and describing the data utilized in the study.

2. Processing the data if necessary to format them for the analysis the group has chosen to reproduce.

3. Rerunning the analysis described in the manuscript using your personal computers or ISU HPC resources.

4. Providing visual summaries (e.g., ggplot ﬁgures) of your results.

**Submitting Items for Grading**

Create Github Repository for all files (include "BCB546X-Spring2017" in the name of your repository.)

1. A README ﬁle in Markdown format that describes the workﬂow throughout the project and lists initials by tasks undertaken by each group member.

2. Files or a folder with any scripts written to process data, run an analysis, and produce ﬁgures.

3. pdfs or jpegs generated to help visualize results.

4. Slides for the group's in-class presentation

***Presentation Guidelines Via Class Repository Instructions:***

Each group will have ~20 minutes to present their work on either April 25th or 27th.

Each presentation should include:

1. Background on the biological question being investigated.

2. A description of the workﬂow carried out by the group.

3. An overview of the group's documentation.

4. Presentation of results including comparison to results from the published paper.

Action Items:

1. **Genetic Processing: - Joe**
   * + 1. Upload all existing genetic processing scripts into a google doc and then combine them into a cohesive single script.
       2. Maybe find hyperlinks to places to be able to curl in reference genome and other external resources needed including
       3. Add in sections to download the software that we will need - impute2, Shapeit, Plink, LiftOver,
       4. Attempt to replicate analysis between PLNK and R for processes genomic data (<https://atgu.mgh.harvard.edu/plinkseq/r-intro.shtml>).
2. A
3. Each of the folders on our shared drive will contain all of the following pieces:
   * + 1. Result file from plink analysis with brain volume
       2. File with all of the subject numbers and family ID, and Gender and Alleles at each specific snp
       3. A file with the SNP name, Chromosome Number, a useless column and the base pair position
       4. A file with the phenotype “brain volume” data for each subject organized in the same sorting order as the genetic file with 3 columns: a unique number for each subject, the subject ID, and the phenotype
4. **Analysis - Elliot**
   * + 1. Snp 1 - association with brain volume
       2. Snp 2 - association with brain volume
       3. Snp 3 - association with brain volume
       4. Snp 4 - association with brain volume
       5. Aggregate score - association with brain volume
5. **Graphing: (Catter-plotting) - Ali**
   * + 1. Snp 1 - association with brain volume
       2. Snp 2 - association with brain volume
       3. Snp 3 - association with brain volume
       4. Snp 4 - association with brain volume
       5. Aggregate score - association with brain volume
6. Each of the folders on our shared drive will contain all of the following pieces:
   1. Result file from plink analysis with brain volume
   2. File with all of the subject numbers and family ID, and Gender and Alleles at each specific snp
   3. A file with the SNP name, Chromosome Number, a useless column and the base pair position
   4. A file with the phenotype “brain volume” data for each subject organized in the same sorting order as the genetic file with 3 columns: a unique number for each subject, the subject ID, and the phenotype