**Systems ML Presentation Notes:**

1. Zeal, Henry, and I, Kristen analyzed Alzheimer’s Disease gene expression for diagnostic applications.
2. There are 50 million AD patients worldwide and these numbers are projected to double every 5 years until 152 million people are affected by 2050. AD is the single largest healthcare cost, costing the us 1 trillion dollars annually yet despite decades of research there is still a large debate about the cause and no drug treatments that halt or cure its progression. Clinicians even disagree about what defines AD, the presence of plaques or behavior, because not all AD patients have the same biomarkers. This provides motivation to find the omic definition of AD, or the ability to diagnose AD from gene expression. In the figure you can see a comparison between a healthy brain and the brain of a patient with AD showing shrinkage and enlarged ventricles.
3. AD is characterized by several molecular and cellular such as Amyloid Plaques and Tau Neurofibrillary Tangles or NFTs. Amyloid plaques are extracellular deposits of amyloid beta proteins. These insoluble plaques play a role in neurotoxicity and cause damage to axons leading to neuron death. NFTs are large hyperphosphorylated microtubule proteins that accumulate in the end of a neuron leading to cell death. Amyloid plaques and NFTs are just some of the many contributions to synaptic loss, reducing neuron connections in the brain. Figure to the right shows a healthy neuron and a neuron from a patient with AD, demonstrating the effects and location of NFTs and Amyloid plaques.


7. For our project we aimed to better understand the omic definition of AD. Find differentially expressed genes between the control and AD subjects. Find the functional annotations to understand the biological meaning behind large lists of genes. And predict if a patient has AD based on gene expression alone.


11. For our pathway analysis we took the terms from the GO biological process library and grouped them using rrvgo R package. This placed the common terms together, so the larger the box the more terms appeared in that area. For example, the blood pathway analysis on the left shows translation as a larger box with subgroups of positive regulation of translation, cytoplasmic translation and many more. When comparing the pathway analysis of the blood vs the brain you can see how different terms appeared based on the different data sets. Since the terms align with the pathways that occur in the blood or the brain we continue with our analysis.
12. Looking at all the differentially expressed genes from the KEGG library you can see how the terms associated with the data align with degenerative diseases such as Parkinson disease, Prion disease, multiple neurogenerative diseases and Alzheimer’s disease.

Paper uses Random forests-internal feature selction

Change eval to use instead of svm use ababoost