**Team 1**

**Short Description of the Analysis**

**Step-by-Step Analysis:**

1. We did research on Alzheimer's disease. This included reading the readings in the Reading Materials for Midterm Project folder and additional articles. These articles gave us an idea about Alzheimer’s symptoms, potential causes, and specifics about the beginning stage of the disease.
2. First, we found distinguishable variations in gene expression among subjects in the early stages of Alzheimer's disease when compared to those in a healthy control group. We did this using R analysis and a significance level of 0.01. This gave us a list of 289 differentiated expressed genes. Here are the specific steps of this analysis:
   1. We are given an Alzheimer's data set with 30 patients. There are 8 patients in the control group and 7 patients in the incipient group.
   2. Read in clinical data and processed normalized gene expression data in log2 scale
   3. Goal: the outcome variable is the "DISEASE\_STATUS" column has values "Control" (baseline group) and "Incipient" (comparison group)
   4. Clean and filter clinical data: clinical data has one row per patient.
   5. Filter gene expression data: good to go because rows are features (genes) and columns are patients.
   6. Identify sample identifiers: For the clinical data column of "BIOSPECIMEN\_ID" and for gene expression data the names of the columns are BIOSPECIMEN\_IDs
   7. We identified the groups to be compared (Baseline of control and Comparison Groups of incipient)
   8. Did a sanity check: confirmed filtered clinical data in R matches filtered clinical data in Excel, confirmed sample IDs in clinical data match sample IDs in gene exp data, verified there are the correct number of samples in baseline and comp groups, and exported the column names from gene expression data to confirm it contains only probe/gene names and no other garbage
   9. Prepped for t-test: confirmed data is numeric, and genes are rows and samples are columns
   10. Performed t-test and go our DEGs
3. Perform a systems biology analysis on the set of differentially expressed genes (DEGs), including pathway enrichment analysis and Gene Ontology enrichment analysis. This will aid in creating a biological understanding of the outcomes and refining the quest for potential biomarkers through biologically meaningful interpretation of the data. Further steps are below:
   1. Imported T-test results (from the step above) to shorten it further.
   2. Short-listed results based on p-value cut-off (p-value <= 0.01)
   3. Exported the short-listed results for reference (into the output folder)
   4. Cleaned gene name. Split the gene names on |
   5. Loaded Databases for the Enrichr R package
   6. Called function to run Enrichment
   7. Got results from enrichment analysis
4. We used our enrichment analysis to find which pathways/genes are significant. Then we did research to find which pathways and genes are connected to Alzheimer’s disease. This could be through impacts on memory, brain function, neurons, etc.
5. We found most of the significant pathways and genes fell into four categories.
   1. Regulation/signaling
   2. Inflammation/Immune Response
   3. Cytoskeleton
   4. DNA Damage

Each person in the group took a group and studied what these pathways and gene functions are and how they relate to Alzheimer’s.

1. After research we came up with “our story” which goes through environmental changes, inflammation and immune response, blood-brain barrier, and DNA Damage. We picked pathways and genes that fit into these categories and charted their functions and connection to Alzheimer’s disease.
2. We compared our DEGs with the severe and moderate groups of DEGS. We particularly investigated genes that we found important that were also in the severe and moderate groups.
3. We created our presentation of our analysis

**Summary:** Alzheimer’s disease is a progressive and irreversible brain disorder that primarily affects memory, thinking, and behavior. We completed a gene expression analysis and enrichment analysis in R to compare the incipient group and control group of Alzheimer’s disease. We found significantly differentiated pathways in genes between the control and incipient groups. We discovered pathways and genes that impact environmental changes, the immune/inflammation system, the blood-brain barrier, and DNA damage. These impacts potentially impact why a person gets Alzheimer’s disease.

**Conclusion:** In conclusion, we have found that significantly differentiated pathways and genes that function may impact Alzheimer’s disease. First, the genes HHIP, MAZ, INPPL1, and CD47 impact how a person responds to environmental changes. Environmental changes are important because researchers attribute Alzheimer's pathology to the brain's innate immune response, which is triggered when these pathogens enter. Second, the genes CD47, JAK2, MEFV, and WASF1 play roles in regulating immune responses and inflammation. Alzheimer’s is associated with chronic neuroinflammation which potentially contributes to neuronal damage. The genes CD47, APCDD1, and DCN impact the Blood Brain Barrier and hyperpermeability of blood vessels. This is key because the blood-brain barrier serves as a protective barrier that prevents harmful plasma-derived components, cells, and pathogens from entering the brain. Finally, the genes HIPK2, SUMO1, and KCNAB1 impact DNA Damage. In Alzheimer's a common occurrence is DNA damage, with double-strand DNA breaks being especially problematic for neurons.

‌

‌

‌

‌