Computational Biology Approach to Analyzing the Genes Affecting Neurodegenerative Diseases

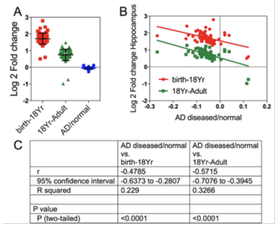
Amulya Garimella, Fox Chapel Area High School, Pittsburgh, PA

Background

- **Neurodegenerative diseases**, such as Alzheimer's, Frontotemporal Dementia and Huntington's kill neurons over time, resulting in the loss of a specific function
- They are **irreversible and can be devastating**. Many of the root causes underlying neurodegenerative disease are unknown, and currently there is no cure for the vast majority. They overwhelmingly effect the aging population; due to **increased life expectancy**, millions of people are affected and there is an **urgent need** to discover the root of these diseases and find potential cures
- Over **5.5 million Americans** of all ages had Alzheimer's Disease, as of 2016; Estimated 5.3 million people aged 65+
- If we could find genes that affect the development of the function that is lost and ALSO play a role in the disease, **we could find the causes and thus the cures of these diseases**

Previous Work (2017)

- Goal - Find if there are **genes that affect both the function of memory as the hippocampus develops, and Alzheimer's**; if they exist, find what patterns those genes exhibit in healthy and diseased brains.
- Procedure - I used the Hippocampus as the experimental data and the Cerebellum as the control data
- Results - **I found 71 significant genes. Genes whose expressions changed in the hippocampus during development are deregulated in the Alzheimer's affected hippocampus.**



C

	AD diseased/normal vs. birth-18Yr	AD diseased/normal vs. 18Yr-Adult
r	-0.4785	-0.5715
95% confidence interval	-0.6373 to -0.2807	-0.7076 to -0.3945
R squared	0.229	0.3266
P value		
P (two-tailed)	<0.0001	<0.0001

A: The x-axis of this graph shows the hippocampus in various stages: before birth/ 18yrs (red), 18yrs/adult (green), and between Alzheimer's-affected and normal hippocampi (blue).

B: A graph showing correlation of the genes that change in development (before birth/18yrs, red) and in adulthood (18 yrs/adult, green) with the x-axis showing the genes that changed in an Alzheimer's affected hippocampus. This graph reveals an inverse relationship.

C: Tabular results of correlation analysis.

Objective

- **Inspired by my previous findings, I expanded my research question and formulated a global hypothesis.**
- **Research Question:** Will neurodegenerative diseases have developmental genes that are deregulated?
- **Hypothesis:** Genes that change in a specific region of the brain but not in the cerebellum (the brain region least affected by neurodegenerative disease), during development will be deregulated in that same region when it is affected by a certain neurodegenerative disease.

Disease	Experimental	Control
Alzheimer's	Hippocampus	Cerebellum
Frontotemporal Dementia	Frontal Cortex	Cerebellum
Huntington's	Striatum	Cerebellum

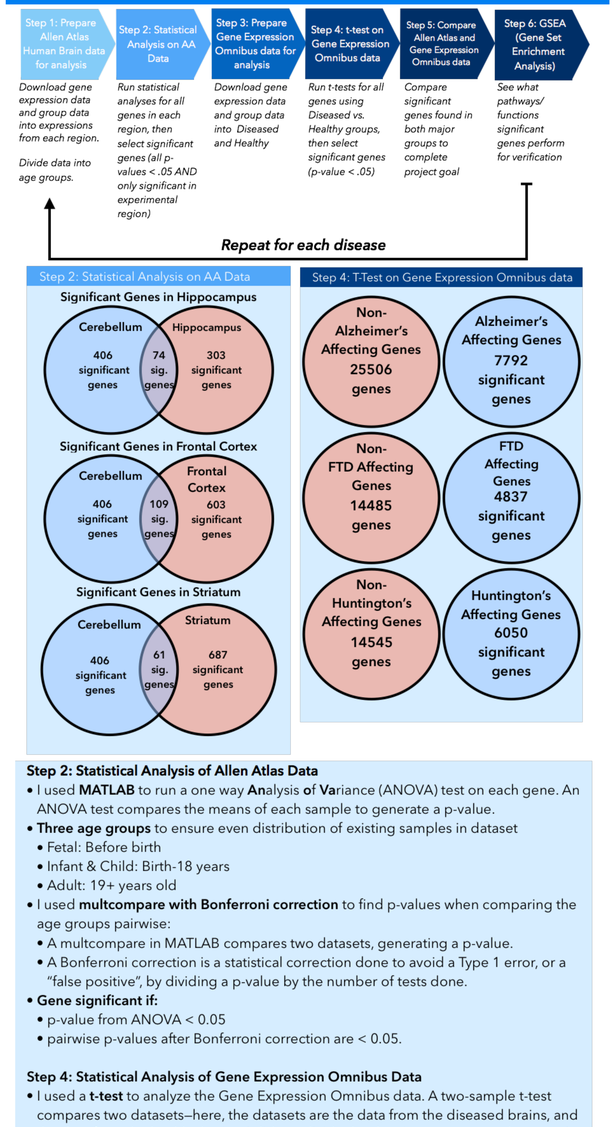
Abstract

- Neurodegenerative diseases like Alzheimer's, Frontotemporal Dementia (FTD) and Huntington's **affect distinct brain regions**, but **it is unknown how this specificity for neuronal toxicity occurs.**
- Because different regions have distinct developmental patterns, I hypothesized that genes that change in specific brain regions during development are deregulated in that same region afflicted by the corresponding disease. **I explored a novel association between developmental genes and neurodegenerative diseases.**
- Results: For Alzheimer's, my hypothesis was proved - developmental genes in the hippocampus decreased in the disease. In contrast to Alzheimer's, FTD and Huntington's did not exhibit deregulation of developmental genes.

Research and Related Work

- Several hypotheses have been suggested to explain the predilection for specific brain regions being affected in different diseases, but no connection has been made between neurodegenerative diseases and the development of the structures that they affect.
- **A genetic correlation between neurodegenerative disease and brain development is a novel idea, and is the basis of my work.**
- **Neurodevelopmental vs. Neurodegenerative Diseases.** In general, neurodevelopmental diseases (caused by issues in the regulation of fundamental processes in the development of the brain, such as ADHD and autism) and neurodegenerative diseases (caused by neuron death later in life) have been seen as two distinct categories of brain disorders. There has been no correlation between the genes that change during brain development and during disease.

Materials and Methods



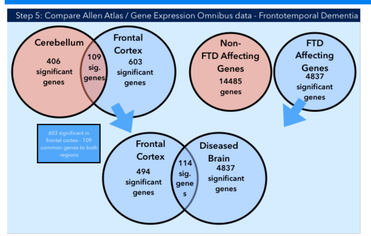
Step 2: Statistical Analysis of Allen Atlas Data

- I used **MATLAB** to run a one way **An**alysis **of Va**riance (ANOVA) test on each gene. An ANOVA test compares the means of each sample to generate a p-value.
- **Three age groups** to ensure even distribution of existing samples in dataset
  - Fetal: Before birth
  - Infant & Child: Birth-18 years
  - Adult: 19+ years old
- I used **multcompare with Bonferroni correction** to find p-values when comparing the age groups pairwise:
  - A multcompare in MATLAB compares two datasets, generating a p-value.
  - A Bonferroni correction is a statistical correction done to avoid a Type 1 error, or a "false positive", by dividing a p-value by the number of tests done.
- **Gene significant if:**
  - p-value from ANOVA < 0.05
  - pairwise p-values after Bonferroni correction are < 0.05.

Step 4: Statistical Analysis of Gene Expression Omnibus Data

- I used a **t-test** to analyze the Gene Expression Omnibus data. A two-sample t-test compares two datasets—here, the datasets are the data from the diseased brains, and the data from the healthy brains. I did not divide the data into age groups because I only wanted to see how genes were changed through disease. Thus, a t-test was sufficient to analyze this data, unlike the statistical analysis for the Allen Atlas data, where I used an ANOVA test to compare three age groups.

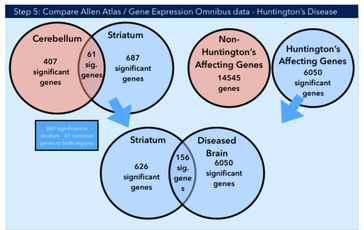
Results – FTD & Huntington's



C

	FTD diseased/normal vs. birth-18Yr	FTD diseased/normal vs. 18Yr-Adult
Pearson r		
r	0.1751	0.1535
95% confidence interval	-0.009061 to 0.3479	-0.03135 to 0.3281
R squared	0.03068	0.02355
P value		
P (two-tailed)	0.0623	0.1031

- **Genes whose expressions changed in the frontal cortex during development show no correlation to those in the FTD affected frontal cortex.**
  - **A:** The x-axis of this graph shows three groups of Log2 fold changes: the frontal cortex before birth/18yrs (red), 18yrs/adult (green), and between FTD-affected and normal frontal cortices (blue).
  - **B:** A graph showing correlation of the genes that change in development (before birth/18yrs, red) and in adulthood (18 yrs/adult, green) with the x-axis showing the genes that changed in an FTD affected frontal cortex. Again, the Log2 fold ratios are shown here.
  - **C:** Tabular results of correlation analysis.
- Graph B plots the fold ratios of each significant gene in this disease. The green points represent the fold ratio for the expressions in the adult brain, and the red for the childhood and fetal brains. They are plotted against the fold ratios in the diseased brains. Here, **no strong correlation is observed**. Unlike in the Alzheimer's graph, there is no strong upward or downward trend, showing statistical insignificance.



C

	HD diseased/normal vs. birth-18Yr	HD diseased/normal vs. 18Yr-Adult
Pearson r		
r	0.0377	0.1002
95% confidence interval	-0.1202 to 0.1937	-0.05794 to 0.2534
R squared	0.001421	0.01004
P value		
P (two-tailed)	0.6403	0.2133

- **Genes whose expressions changed in the striatum during development show no correlation to those in the Huntington's affected striatum.**
  - **A:** The x-axis of this graph shows three groups of x Log2 fold changes: the striatum before birth/18yrs (red), 18yrs/adult (green), and between Huntington's-affected and normal striatum (blue).
  - **B:** A graph showing correlation of the genes that change in development (before birth/18yrs, red) and in adulthood (18 yrs/adult, green) with the x-axis showing the genes that changed in an Huntington's-affected striatum. Again, the Log2 fold ratios are shown here.
  - **C:** Tabular results of correlation analysis.
- Similar to Frontotemporal Dementia, there is no major trend observed in Graph B, which plots the fold ratios of each significant gene in this disease. Again, the green points represent the fold ratio for the expressions in the adult brain, and the red for the childhood and fetal brains. They are plotted against the fold ratios in the diseased brains.

Conclusions and Discussion

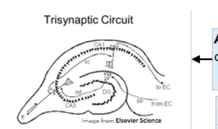
The genes that changed the most in development were decreased in Alzheimer's, but not in FTD or Huntington's. This leads to the following significant conclusions:

- All neurodegenerative diseases do not exhibit deregulation of developmental genes
- Alzheimer's is different from other neurodegenerative diseases (if not unique) and shows a correlation between the neurodegenerative disease and hippocampus development
- Different structures of the brain must have fundamentally different patterns of development in order for these findings to be true.

My findings show that in Alzheimer's (but not necessarily other neurodegenerative diseases**), the same genes important to the childhood and post-conception development of the hippocampus in the brain are affected and actually deregulated in Alzheimer's disease.**

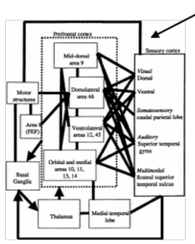
These findings are made even more significant, because other neurodegenerative diseases, specifically FTD and Huntington's **do not exhibit this pattern.** This finding merits far more study. It not only sheds light on the development of Alzheimer's disease, but on that of other neurodegenerative diseases as well.

One possible reason for differences between these diseases is **structural differences.**

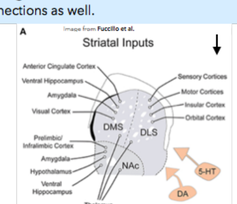


Alzheimer's Disease: The hippocampus has a simple, self-contained synaptic circuit consisting of three major connections.

Frontotemporal Dementia: In contrast, the frontal cortex has a complex synaptic circuit consisting of several major connections to other parts of the brain.



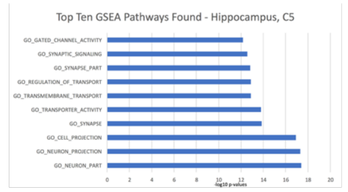
Huntington's Disease: The striatum makes very complex connections as well.



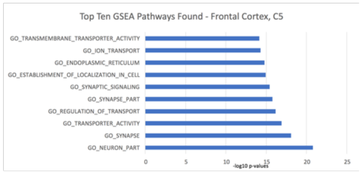
Results - GSEA

Gene Set Enrichment Analysis (GSEA) is a way to **retrieve a functional profile of a given gene set** in order to better understand the **underlying biological processes.** For each structure, I used two gene sets to categorize the pathways of the significant genes identified in Step 5: **C2 Curated gene sets** and **C5 Gene Ontology gene sets**. Shown below are some of the top pathways found in the **C5 Gene Ontology gene sets.**

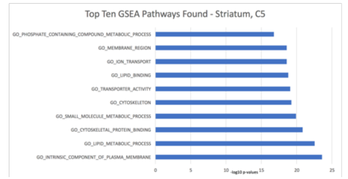
Step 6: GSEA - Hippocampus. The genes that were significant through age in the hippocampus **participate in pathways that are directly related to the brain and synaptic function.**
Nearly all pathways found were related to synaptic activity.



Step 6: GSEA - Frontal Cortex. The genes that were significant through age in the frontal cortex were involved in the same pathways as the hippocampus, but did not exhibit the same pattern correlating development to disease .
Some pathways were involved in synaptic function.

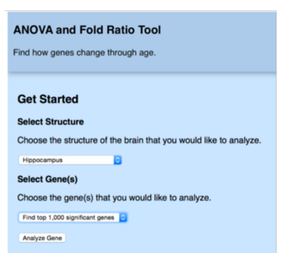


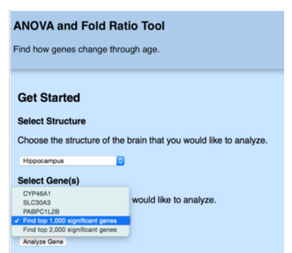
Step 6: GSEA - Striatum. The genes that were significant through age in the striatum played no part in its function. These pathways did not overlap with movement pathways associated with Huntington's Disease.



Biotechnology Applications

- **My findings fundamentally change the way that neurodegenerative diseases, and by extension, the regions that they affect, are viewed. This could lead to more effective treatments for these diseases, as well as an expanded knowledge of the brain.**
- While completing my project, I found much of the manual statistical analysis time consuming and potentially error prone. I created a Allen Brain Atlas API that finds how genes change through age, this will be useful to other researchers.
- **This API is unique – no such API currently exists! This would help researchers to see how genes during development of any part of the brain.**
- I used HTML, CSS, and JavaScript to code the frontend of my API, while I used Python and Flask for the backend to code the backend of my API.
- The libraries that I used included the scipy.stats package for ANOVA and t-test functionality. I stored the data that the API used in MongoDB, due to its support for heterogeneous data.





Acknowledgements

The work I did and the conclusions I reached would not be possible without the help of Dr. Sriram Venneti and Dr. Pooja Panwalkar at the University of Michigan. I received essential guidance on my project from both of them. I would also like to thank my science fair sponsor, Mr. Gregory Schubert, for providing me vital help with the application process. Finally, a huge thank you to my family and my friends! They gave me so much motivation and support along the way.