Finding Candidate Genes for Alzheimer's Disease

Kitty Moy, Anna Neumann, Midori Sangiolo, Newton South High School, Newton, MA 02459 US & BioScience Project, Wakefield, MA, 01880

Introduction

- Alzheimer's disease is a progressive neurodegenerative disease that affects many elderly individuals.
- Symptoms include extreme memory loss and confusion.
- The hippocampus, amygdala, and frontal lobe are all regions of the brain associated with memory and may be affected by the disease.
- In this project a systems biology approach was used to identify candidate genes linked to Alzheimer's disease.

Methods

Gene expression data for the hippocampus, amygdala, and frontal lobe were collected from the The Allen Brain Atlas (http://www.brain-map.org) using the differential search option. Data was collected from four available donors: H0351., H0351., H0351, H0351. Data with a fold-change score greater than 3 were considered in this study.

Venny 2.1.0 (http://bioinfogp.cnb.csic.es/tools/venny/) was used to compare the gene lists from four chosen brain donors to identify genes that are common and different across each donor.

Statistical analysis was done in Python Anywhere (https://www.pythonanywhere.com) an online programming tool.

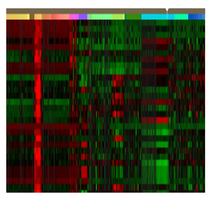
Cluster analysis and Gene Ontology classifications were obtained with DAVID (https://david.ncifcrf.gov).

The STRING database (http://string-db.org) was used to identify potential interacting partners, pathways, and other genes relating to learning.

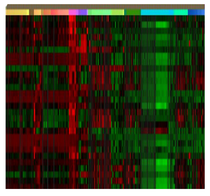
.

Gene Expression Profiling

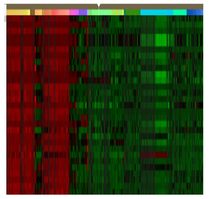
Hippocampus



Amygdala



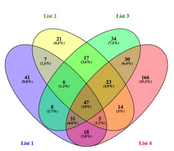
Frontal Lobe



- The above heat maps represent gene expression data for the three brain regions considered in this study. Data was collected from four human donor brains.
- Each column represents a sub-structure within the hippocampus, amygdala, or frontal lobe, respectively.
- Red regions indicate high gene expression relative to gray matter, whereas green regions indicate a lower expression. Black regions indicate equal expression between the respective brain region and gray matter.
- The CA2 (hippocampus), basomedial nucleus (amygdala), and inferior gyrus, occipital lobe, parietal lobe, temporal lobe (frontal lobe) have areas of high gene expression that is conserved across donors.
- From these data, the top 20 most expressed within each donor and common genes all donors were isolated and studied.

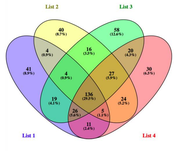
Common Genes

Hippocampus



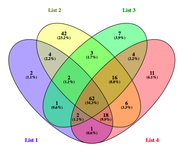
29.5% of the genes are common between all 4 donors

Amygdala



10% of the genes are common between all 4 donors

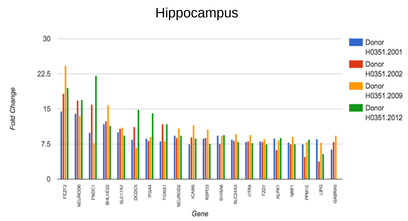
Frontal Lobe



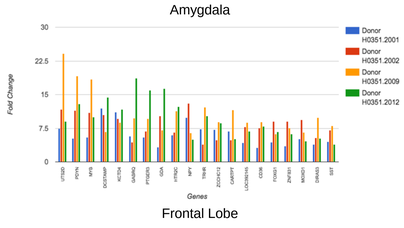
34.3% of the genes are common between all 4 donors

Top 20 Genes with Highest Expression Values

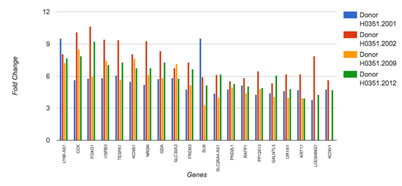
Hippocampus



Amygdala



Frontal Lobe



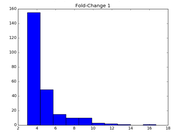
These graphs illustrate the top 20 genes with the highest fold-change expression for each donor. On the x axis are the gene names and on the y are their fold-change expressions.Genes were ordered by the highest average fold-change across all four donors to the lowest average.

For the frontal lobe, genes LOC646627 and KCNV1 were only common across three donors, indicating that they are not as common in this area of the brain. Similarly, in the hippocampus, GABRA5 was only found across three donors.

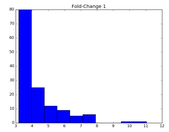
FOXG1 was found across all regions of the brain for all donors, expressing a high fold-change, and GDA had a high fold-change across the frontal lobe and amygdala region of the brain.

Distributions

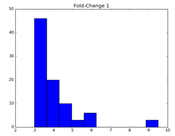
Hippocampus



Amygdala



Frontal Lobe



Above are the data distributions for the hippocampus, amygdala, and frontal lobe respectively.
A commonality between all the donors is seen in the their pattern of gene expression; these histograms are all right-skewed. All of the donors' histograms show that most genes have lower expression and few genes are expressed highly.

Statistical Analysis

Hippocampus

```
       Fold-Change 1  Fold-Change 2  Fold-Change 3  Fold-Change 4
count     246.000000     246.000000     246.000000     246.000000
mean        4.651492       4.668711       5.057817       5.452402
std         2.016637       2.184495       2.056869       2.292573
min         3.000000       3.000000       3.240000       3.160000
25%         3.391500       3.335250       3.630000       3.539750
50%         3.940000       3.865000       4.265500       4.167000
75%         5.117500       5.040250       5.790750       5.571000
max        16.717000      17.261000      15.858000      19.497000
variance
Fold-Change 1    4.066826
Fold-Change 2    4.772817
Fold-Change 3    4.230710
Fold-Change 4    5.724405
dtype: float64
>>> 
```

Amygdala

```
       Fold-Change 1  Fold-Change 2  Fold-Change 3  Fold-Change 4
count     139.000000     139.000000     139.000000     139.000000
mean        4.308453       4.582712       5.426096       5.622647
std         1.347854       1.887051       2.611831       2.341845
min         3.186000       2.967000       3.410000       3.820000
25%         3.400500       3.400000       3.863000       4.115500
50%         3.780000       3.838000       4.400000       4.805000
75%         4.360500       5.027500       6.124500       6.311000
max        11.046000      11.725000      19.117000      16.289000
variance
Fold-Change 1    1.816711
Fold-Change 2    3.560962
Fold-Change 3    6.821661
Fold-Change 4    5.484238
dtype: float64
>>> 
```

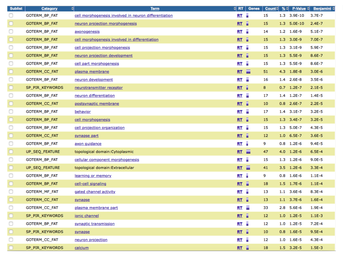
Frontal Lobe

```
       Fold-Change 1  Fold-Change 2  Fold-Change 3  Fold-Change 4
count      88.000000      86.000000      88.000000      88.000000
mean        4.029271       5.221545       4.102739       4.471364
std         1.298486       1.672612       1.188622       1.226299
min         2.999000       3.668000       3.142000       3.298000
25%         3.224000       4.056250       3.374250       3.650750
50%         3.600000       4.720000       3.664000       3.999500
75%         4.293500       5.433750       4.204250       4.671250
max         9.503000      10.639000       8.500000       9.230000
variance
Fold-Change 1    1.680668
Fold-Change 2    2.797632
Fold-Change 3    1.412823
Fold-Change 4    1.503810
dtype: float64
>>> 
```

The means are comparable across donors for each brain region. The standard deviations are also in the same range, except for the amygdala: Donors 3 and 4 are around 2-fold higher. These similar deviations indicate a large spread in the data. Maximum and minimum values are also comparable for each brain region.

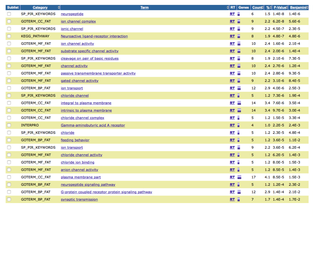
Common Gene Analysis

Hippocampus



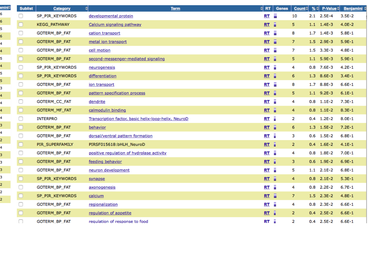
Subfilter	Category	#	Term	FDR	Genes	Count	%	P-Value	Benjamini
☐	GOTERM_BP_FAT		cell morphogenesis involved in neuron differentiation	RT	22	15	1.3	3.9E-10	5.7E-7
☐	GOTERM_BP_FAT		neuron projection morphogenesis	RT	22	15	1.3	5.0E-10	2.4E-7
☐	GOTERM_BP_FAT		axonogenesis	RT	22	14	1.2	1.6E-9	5.1E-7
☐	GOTERM_BP_FAT		cell morphogenesis involved in differentiation	RT	22	15	1.3	5.0E-9	7.0E-7
☐	GOTERM_BP_FAT		cell projection morphogenesis	RT	22	15	1.3	3.1E-9	5.9E-7
☐	GOTERM_BP_FAT		neuron projection development	RT	22	15	1.3	5.5E-9	8.0E-7
☐	GOTERM_BP_FAT		cell part morphogenesis	RT	22	15	1.3	5.5E-9	8.6E-7
☐	GOTERM_CC_FAT		plasma membrane	RT	22	51	4.3	1.8E-8	3.1E-6
☐	GOTERM_BP_FAT		neuron development	RT	22	16	1.4	2.6E-8	3.5E-6
☐	SP_PIR_KEYWORDS		neurotransmitter receptor	RT	7	8	0.7	1.2E-7	2.1E-5
☐	GOTERM_BP_FAT		neuron differentiation	RT	22	17	1.4	1.2E-7	1.4E-5
☐	GOTERM_CC_FAT		endosomal membrane	RT	22	10	0.8	2.6E-7	2.2E-5
☐	GOTERM_BP_FAT		behavior	RT	22	17	1.4	3.1E-7	3.2E-5
☐	GOTERM_BP_FAT		cell morphogenesis	RT	22	15	1.3	3.4E-7	3.2E-5
☐	GOTERM_BP_FAT		cell projection organization	RT	22	15	1.3	8.0E-7	4.3E-5
☐	GOTERM_BP_FAT		synapse part	RT	22	12	1.0	6.5E-7	3.6E-5
☐	GOTERM_BP_FAT		axon guidance	RT	22	9	0.8	1.2E-6	9.4E-5
☐	UP_SEQ_FEATURE		transmembrane domain:Cytoplasmic	RT	22	47	4.0	1.2E-6	6.5E-4
☐	GOTERM_BP_FAT		cellular component morphogenesis	RT	22	15	1.3	1.2E-6	9.0E-5
☐	UP_SEQ_FEATURE		transmembrane domain:Extracellular	RT	22	41	3.5	1.2E-6	3.3E-4
☐	GOTERM_BP_FAT		learning or memory	RT	22	9	0.8	1.6E-6	1.1E-4
☐	GOTERM_BP_FAT		cell-cell signaling	RT	22	18	1.5	1.7E-6	1.1E-4
☐	GOTERM_MF_FAT		gated channel activity	RT	22	13	1.1	3.6E-6	8.3E-4
☐	GOTERM_CC_FAT		synapse	RT	22	13	1.1	3.7E-6	1.6E-4
☐	GOTERM_CC_FAT		plasma membrane part	RT	22	33	2.8	5.6E-6	1.9E-4
☐	SP_PIR_KEYWORDS		ionic channel	RT	22	12	1.0	1.3E-5	1.1E-3
☐	GOTERM_BP_FAT		synaptic transmission	RT	22	12	1.0	1.2E-5	7.2E-4
☐	SP_PIR_KEYWORDS		synapse	RT	22	10	0.8	1.6E-5	9.5E-4
☐	GOTERM_CC_FAT		neuron projection	RT	22	12	1.0	1.6E-5	4.3E-4
☐	SP_PIR_KEYWORDS		calcium	RT	22	18	1.5	3.2E-5	1.9E-3

Amygdala



Subfilter	Category	#	Term	FDR	Genes	Count	%	P-Value	Benjamini
☐	SP_PIR_KEYWORDS		nucleotide	RT	9	6	1.5	1.4E-6	3.6E-4
☐	GOTERM_CC_FAT		ion channel complex	RT	9	5	2.2	4.2E-6	3.8E-4
☐	SP_PIR_KEYWORDS		ionic channel	RT	9	6	2.2	4.5E-7	2.3E-5
☐	KEGG_PATHWAY		Neuroactive ligand-receptor interaction	RT	9	8	1.9	4.8E-7	4.8E-6
☐	GOTERM_MF_FAT		ion channel activity	RT	10	10	2.4	1.4E-6	2.1E-4
☐	GOTERM_MF_FAT		substrate-specific channel activity	RT	10	10	2.4	2.0E-6	1.4E-4
☐	SP_PIR_KEYWORDS		channel or pair of basic residues	RT	9	8	1.9	2.1E-6	7.3E-5
☐	GOTERM_MF_FAT		channel activity	RT	10	10	2.4	2.7E-6	1.2E-4
☐	GOTERM_MF_FAT		passive transmembrane transporter activity	RT	10	10	2.4	2.8E-6	9.3E-5
☐	GOTERM_MF_FAT		gated channel activity	RT	9	9	2.2	5.2E-6	5.8E-5
☐	GOTERM_BP_FAT		ion transport	RT	12	12	2.9	4.9E-6	3.5E-3
☐	SP_PIR_KEYWORDS		chloride channel	RT	5	5	1.2	7.5E-6	1.9E-4
☐	GOTERM_CC_FAT		integral to plasma membrane	RT	14	14	3.4	9.7E-6	3.6E-4
☐	GOTERM_CC_FAT		intrinsic to plasma membrane	RT	14	14	3.4	9.7E-6	3.6E-4
☐	GOTERM_CC_FAT		chloride channel complex	RT	5	5	1.2	1.9E-5	3.3E-4
☐	INTERPRO		Gamma-aminobutyric acid A receptor	RT	4	4	1.0	3.2E-5	2.4E-3
☐	SP_PIR_KEYWORDS		chloride	RT	5	5	1.2	2.3E-5	4.8E-4
☐	GOTERM_BP_FAT		feeding behavior	RT	5	5	1.2	3.6E-5	2.1E-3
☐	SP_PIR_KEYWORDS		ion transport	RT	9	9	2.2	3.6E-5	6.2E-4
☐	GOTERM_MF_FAT		chloride channel activity	RT	5	5	1.2	4.3E-5	1.4E-3
☐	GOTERM_MF_FAT		chloride ion binding	RT	5	5	1.2	8.0E-5	1.9E-3
☐	GOTERM_MF_FAT		anion channel activity	RT	5	5	1.2	8.9E-5	1.4E-3
☐	GOTERM_CC_FAT		plasma membrane part	RT	17	17	4.1	9.5E-5	1.9E-3
☐	GOTERM_BP_FAT		neuropeptide signaling pathway	RT	5	5	1.2	1.2E-4	2.3E-2
☐	GOTERM_MF_FAT		G-protein coupled receptor protein signaling pathway	RT	12	12	2.9	1.4E-4	3.1E-3
☐	GOTERM_BP_FAT		synaptic transmission	RT	7	7	1.7	1.4E-4	1.7E-2

Frontal Lobe

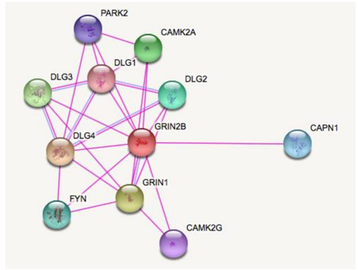


Subfilter	Category	#	Term	FDR	Genes	Count	%	P-Value	Benjamini
☐	SP_PIR_KEYWORDS		developmental protein	RT	10	10	2.1	2.5E-6	3.5E-3
☐	KEGG_PATHWAY		Calcium signaling pathway	RT	5	5	1.1	1.4E-3	4.0E-2
☐	GOTERM_BP_FAT		cation transport	RT	8	8	1.7	1.4E-3	5.8E-1
☐	GOTERM_BP_FAT		metal ion transport	RT	7	7	1.5	2.8E-3	5.9E-1
☐	GOTERM_BP_FAT		cell motion	RT	7	7	1.5	3.3E-3	4.8E-1
☐	GOTERM_BP_FAT		second-messenger-mediated signaling	RT	5	5	1.1	5.9E-3	5.9E-1
☐	SP_PIR_KEYWORDS		neurogenesis	RT	6	6	1.3	7.6E-3	4.2E-1
☐	SP_PIR_KEYWORDS		differentiation	RT	6	6	1.3	8.6E-3	3.4E-1
☐	GOTERM_BP_FAT		ion transport	RT	9	9	1.9	8.8E-3	8.6E-1
☐	GOTERM_BP_FAT		pattern specification process	RT	5	5	1.1	9.2E-3	8.1E-1
☐	GOTERM_CC_FAT		dendrite	RT	4	4	0.9	1.1E-2	7.3E-1
☐	GOTERM_MF_FAT		calmodulin binding	RT	4	4	0.9	1.1E-2	8.3E-1
☐	INTERPRO		Transcription factor, basic helix-loop-helix, NeuroD	RT	2	2	0.4	1.2E-2	8.9E-1
☐	GOTERM_BP_FAT		behavior	RT	6	6	1.3	1.5E-2	7.2E-1
☐	GOTERM_BP_FAT		dorsal/ventral pattern formation	RT	3	3	0.6	1.5E-2	8.8E-1
☐	PIR_SUPERFAMILY		PIRSF015618:bHLH_NeuroD	RT	2	2	0.4	1.6E-2	4.1E-1
☐	GOTERM_BP_FAT		positive regulation of hydrolase activity	RT	4	4	0.9	1.8E-2	7.6E-1
☐	GOTERM_BP_FAT		feeding behavior	RT	3	3	0.6	1.9E-2	8.9E-1
☐	GOTERM_BP_FAT		neuron development	RT	5	5	1.1	2.1E-2	8.8E-1
☐	SP_PIR_KEYWORDS		synapse	RT	4	4	0.9	2.1E-2	5.3E-1
☐	GOTERM_BP_FAT		axonogenesis	RT	4	4	0.9	2.2E-2	8.7E-1
☐	SP_PIR_KEYWORDS		calcium	RT	7	7	1.5	2.3E-2	4.8E-1
☐	GOTERM_BP_FAT		regeneration	RT	4	4	0.9	2.3E-2	8.6E-1
☐	GOTERM_BP_FAT		regulation of peptide	RT	3	3	0.6	2.5E-2	8.6E-1
☐	GOTERM_BP_FAT		regulation of response to food	RT	2	2	0.4	2.5E-2	8.6E-1

By partitioning the common genes based on functional classification (Gene Ontology), genes relevant to Alzheimer's phenotype/symptoms were found in the hippocampus, amygdala, and frontal lobe. The hippocampus genes were used in the subsequent network analysis due to its unique relevance to memory in order to identify additional genes that may be involved in Alzheimer's. Sections highlighted* show genes that are connected with learning, memory, aging, and neurological systems. Genes relevant to Alzheimer's specifically were found in all three regions of the brain: GRIN2B and PRKCA (hippocampus) HTR2A and NEUROD2 (frontal lobe) CDH4 and FOXG1 (Amygdala)

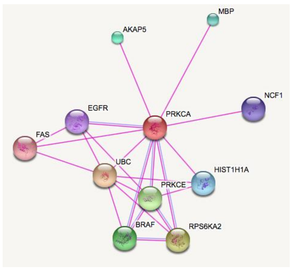
Network Analysis

GRIN2B



- GRIN2B triggers stroke damage by inducing Ca^{2+} influx through NMDA receptor, causing neuronal death. This gene contains an NR2B subunit whose misregulation is commonly linked with Alzheimer's disease. It is also linked with Amyotrophic Lateral Sclerosis and Huntington's Disease
- DLG3--discs, large homolog 3--is essential for learning through synaptic plasticity activity following NMDA receptor signaling. This gene is also related with the cytoskeleton/synapse

PRKCA



- Protein Kinase C is involved in cell growth and differentiation and relates to work recognition and memory. It activates signaling in MAPK1/3 (ERK1/2) and RAP1GAP.
- EGFR helps control cell growth, playing a role in brain development. EGFR inhibitors have been experimentally linked with reversing the symptoms of Alzheimer's disease in animal test subjects.

Conclusions

The GRIN2B and PRKCA networks have a high degree of interconnectivity.

Both networks contain genes that are already associated with Alzheimer's and other neurological disorders such as Huntington's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis and Prion disease.

The GRIN2B network has a concentration of genes (FAS, EGFR, PRKCA, RPS6KA2,BRAF, associated with MAP kinase signaling pathway which relays information from cell surface receptors to the cell nucleus

Other candidate genes are associated with cytoskeletal/synapse function as well as cell cycle and cell death.