

# Determining Candidate Genes for OCD in the Basal Ganglia

Naila Ahmed, Newton North High School, Newton, MA, 02460 & BioScience Project, Wakefield, MA, 01880

## Introduction

- Obsessive-compulsive disorder is a mental disorder which compels people to have repeated habits, feelings, thoughts, or ideas. OCD can also be accompanied by depression, anxiety disorders, or eating disorders.
- The basal ganglia or the basal nuclei is a region located in the base of the brain that consists of three clusters of neurons. These clusters are responsible for involuntary movements and are associated with OCD.
- In this project, our goal was to determine candidate genes related to OCD in the basal ganglia by exploring and analyzing gene expression patterns.

## Methods

Gene expression data for the basal ganglia was collected from the The Allen Brain Atlas (<http://www.brain-map.org>) using the differential search option. Data was collected from four available donors: H0351.2001, H0351.2002, H0351.1009, H0351.1012

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.

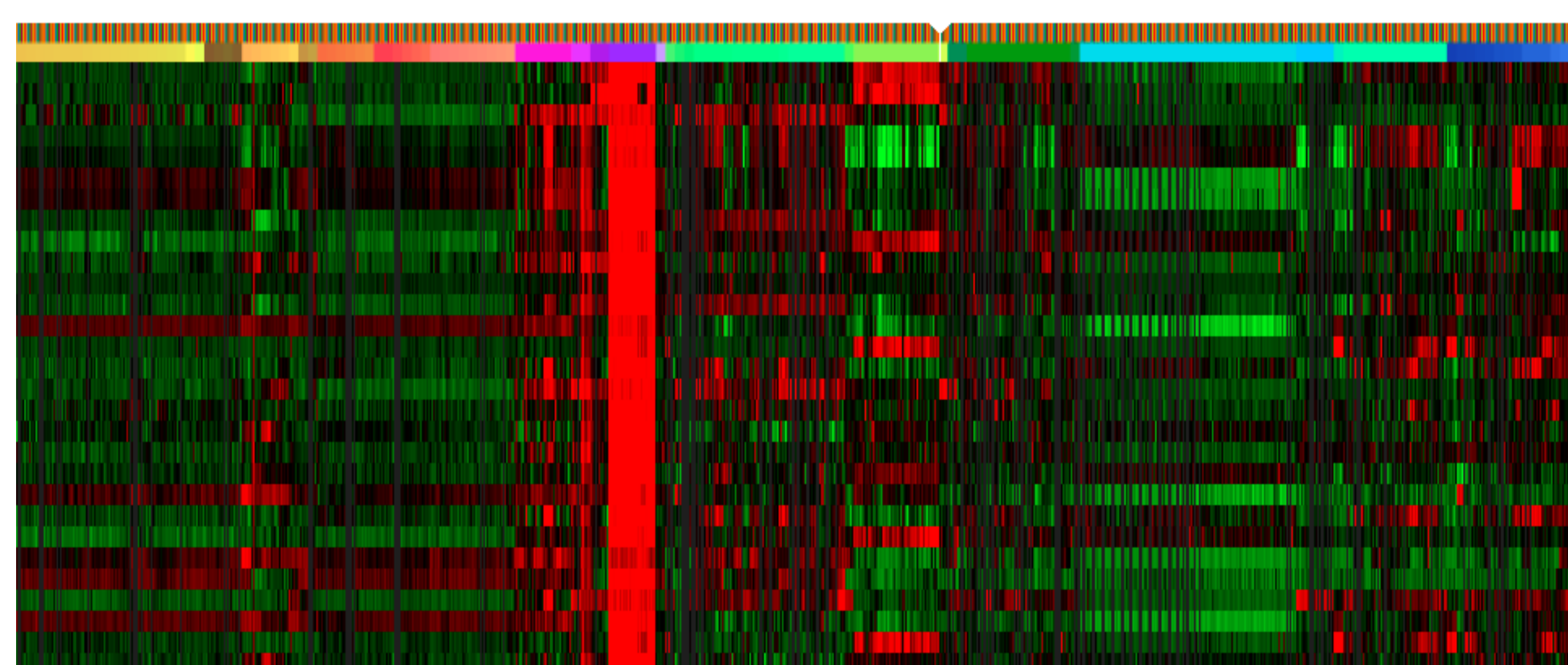
Statistical analysis was done in Python Anywhere (<https://www.pythonanywhere.com>) an online programming tool

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.

## Results

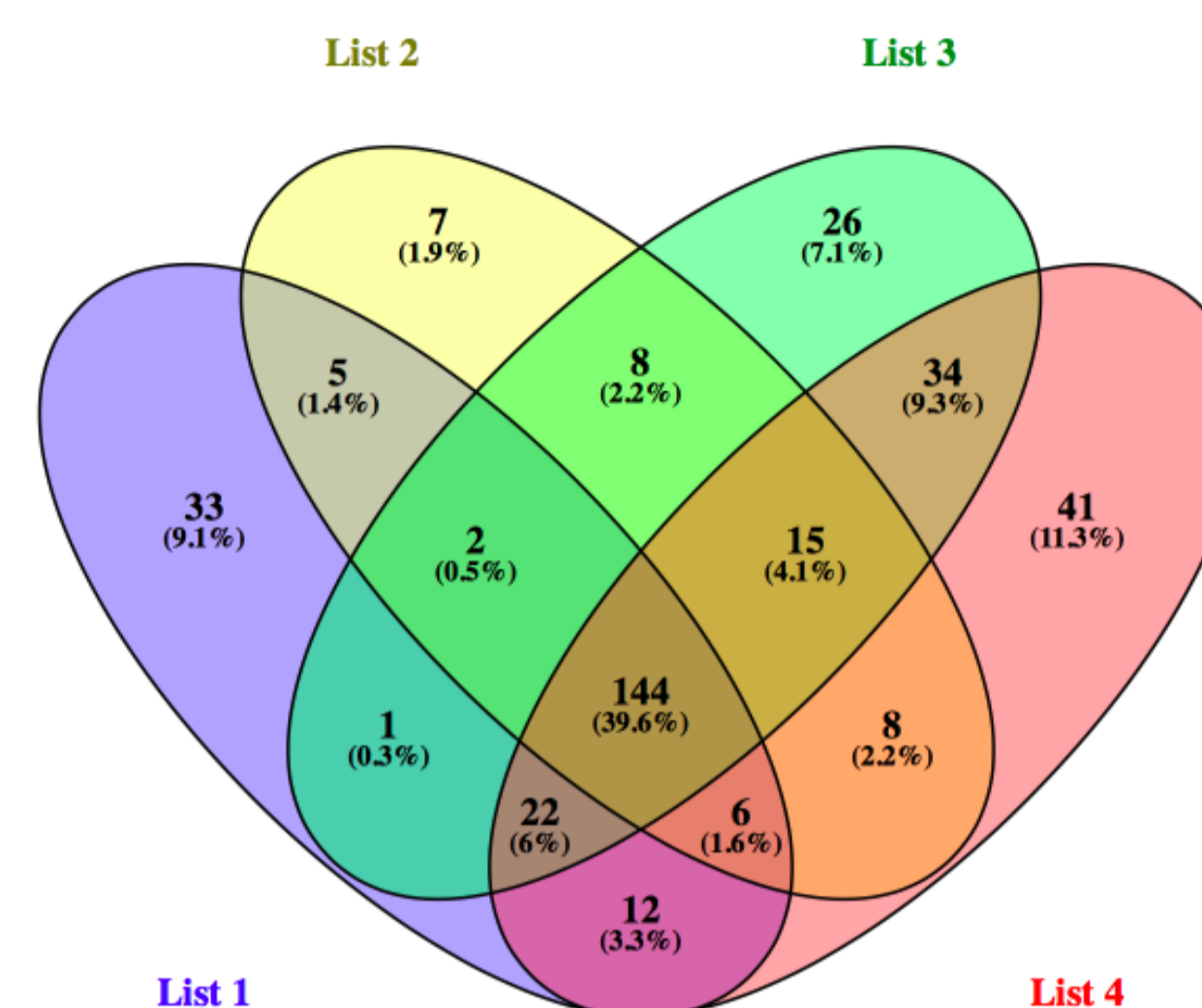
### Gene Expression Profiling



The microarray heat map displays gene expression data in the basal ganglia in contrast to gray matter of 6 donor brains. The top bar of the heat map represents each of the 6 donors and the bar below depicts subregions of the basal ganglia. In the microarray data, the sample under study versus the control are compared. Each color displayed in the heat map represents the expression of the gene relative to the control: red = expression greater than control, green = expression less than control, black = expression equal to control. Data that had a fold change of 3 and above was used in the analysis.

There is high gene expression in the tail of the caudate nucleus. This could be due to the fact that the caudate nucleus is responsible for many behavioral functions such as learning and executive functions. All donors have similar gene expression.

## Common Genes



There are 177 (39.6%) common genes in the basal ganglia between 4 donors.

The Venn diagram compares the genes of each donor and identifies the similarities and the differences between the data.

## Top 20 Genes with Highest Expression Values



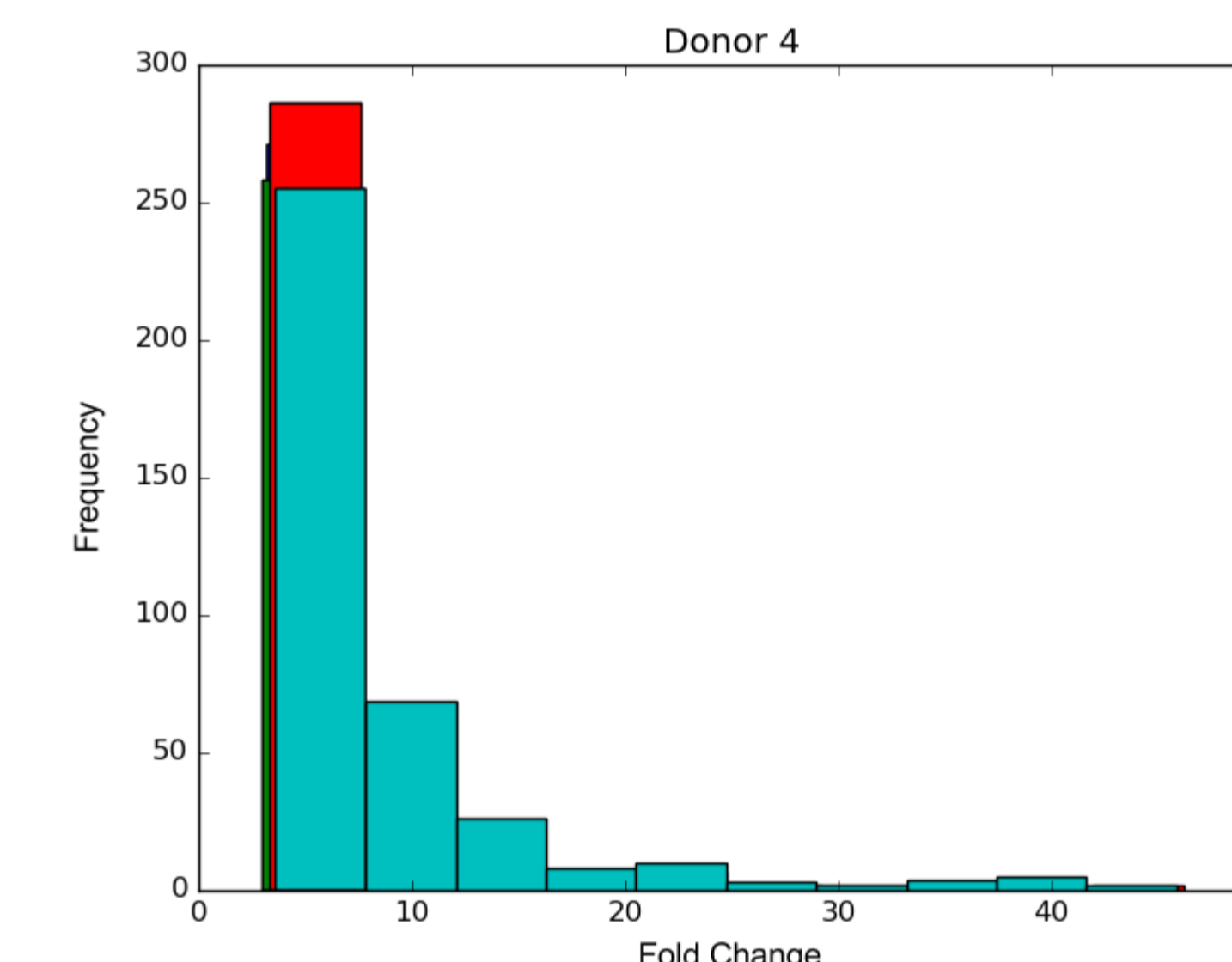
The top 20 genes with the highest gene expression values of each donor are shown on the graphs above. Red bars = common genes in all 4 donors, blue bars = genes that are not common in all four donors.

10 (50%) of the genes are common between the four donors. The gene SYNGIG1L has a notably higher gene expression value in all four donors. The gene expression for the rest of the genes are relatively close in value.

## Statistical Analysis

	6.221078	5.717883	6.939385	8.505982
mean	6.221078	5.717883	6.939385	8.505982
std	4.177637	3.759380	5.288894	7.034283
min	3.218000	3.002000	3.357000	3.609000
25%	3.764750	3.386500	4.018500	4.432750
50%	4.623000	4.309500	5.159500	5.870000
75%	6.939500	6.736250	7.869500	9.748500
max	35.656000	27.877000	46.236000	45.878000
variance				
Donor 1	17.452648			
Donor 2	14.132940			
Donor 3	27.972403			
Donor 4	49.481133			
dtype:	float64			

The minimum and maximum for all the donors are very similar. However, the mean fold-change and the standard deviation for donors 1-3 are relatively close in value while donor 4 has a slightly higher mean fold change and standard deviation. The standard deviation for all of the donors are close in value to the means indicating a large spread in the data. Donor 4's differences are shown again in the variance where the data ranges from 17.45 to 49.48.



The distribution of the gene expression data for all 4 donors is right-skewed. The majority of the data falls below ~8 (fold-change).

## Gene Classification

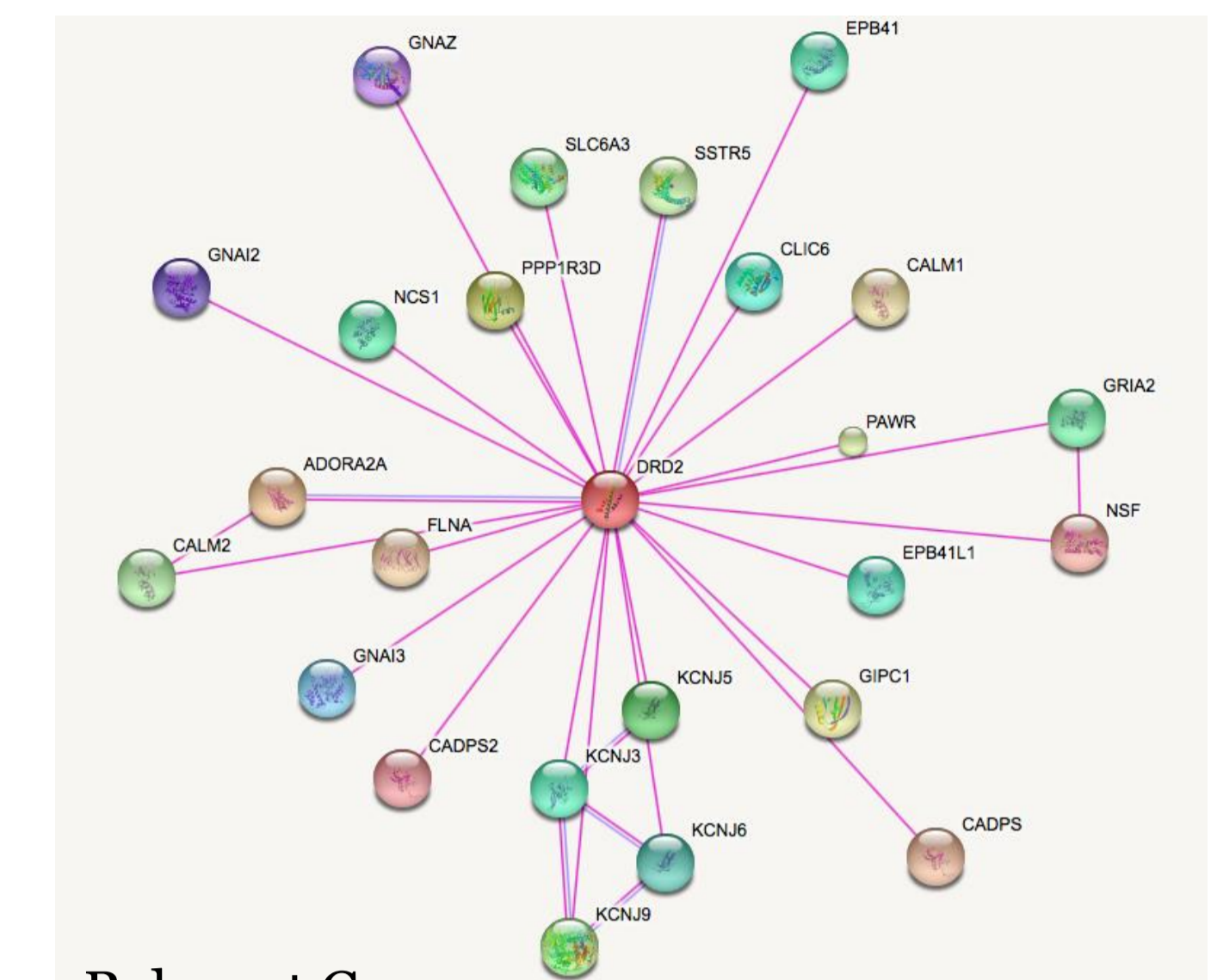
DRD2	dopamine receptor D1	Related Genes	Homo sapiens
BIOCARTA	Regulation of cdk1/cdk5 by type 1 glutamate receptors		
GOTERM_BP_FAT	regulation of action potential, temperature homeostasis, conditioned taste aversion, behavioral fear response, synaptic transmission, dopaminergic startle response, response to amphetamine, behavioral defense response, muscle system process, regulation of nucleotide metabolic process, negative regulation of protein kinase activity, cellular amino acid derivative metabolic process, biogenic amine metabolic process, catecholamine metabolic process, protein targeting, protein import into nucleus, phospholipid metabolic process, glucocorticoid metabolic process, cellular ion homeostasis, cellular calcium ion homeostasis, cellular metal ion homeostasis, intracellular protein transport, nucleocytoplasmic transport, cell motion, muscle contraction, smooth muscle contraction, defense response, cell surface receptor linked signal transduction, G-protein coupled receptor protein signaling pathway, G-protein signaling, coupled to cyclic nucleotide second messenger, G-protein signaling, coupled to cAMP nucleotide second messenger, activation of adenylate cyclase activity by G-protein signaling pathway, activation of adenylate cyclase activity, activation of adenylate cyclase activity by dopamine receptor signaling pathway, activation of phospholipase C activity by G-protein coupled receptor protein signaling pathway coupled to IP3 second messenger, activation of phospholipase C activity, elevation of cytosolic calcium ion concentration, dopamine receptor signaling pathway, intracellular signaling cascade, cell-cell signaling, synaptic transmission, nerve synaptic transmission, synaptogenesis, behavior learning or memory, learning, memory, mating behavior, mating, rhythmic behavior, circadian rhythm, grooming behavior, locomotory behavior, adult walking behavior, feeding behavior, visual behavior, protein localization, associative learning, adult locomotory behavior, visual learning, carbohydrate transport,ucose transport,		

Gene ontology classification was used to determine candidate and relevant genes out of the common genes of all 4 donors. Some of the many genes of interest found are: DRD1, GNG7, PENK, DRD2, ADORA2A, NPY, PBX3. These genes are associated with many neural processes related to OCD such as, behavioral fear response, rhythmic behavior, grooming behavior, locomotory behavior, and feeding behavior.

Annotation Cluster 2	Enrichment Score: 2.79	Count	P-Value	Benjamini
GOTERM_BP_FAT	neurological system process	24	4.3E-5	8.4E-3
GOTERM_BP_FAT	cognition	16	4.3E-3	8.5E-2
GOTERM_BP_FAT	sensory perception	13	2.3E-2	2.4E-1

There are 24 genes involved with neurological system process, 16 genes involved with cognition, and 13 genes involved with sensory perception.

## Network Analysis



Relevant Genes:

DRD2 is a dopamine receptor gene which has been associated with diseases similar to OCD such as Parkinson's and Schizophrenia where repetitive words/phrases and behaviors are common.

SLC6A3 is a dopamine receptor gene that has also been associated with a disease similar to OCD which is Parkinson's

ADORA2A is an adenosine receptor gene that has also been associated with diseases similar to OCD such as Alzheimer's, Parkinson's, and Huntington's disease. ADORA2A is also a common gene of all 4 donors.

## Conclusions

- OCD is an interesting yet complicated disease due to the fact that it can be diagnosed with other complex diseases and also because different factors like gender and environment can cause OCD. Considering this, OCD should be studied and researched more.
- Diseases like Alzheimer's and Parkinson's are related to OCD because they possess OCD-like symptoms like repeating words and behaviors and also because all three diseases share associations with regions of the brain.
- There is a network between ADORA2A and DRD2.
- DRD2 and ADORA2A are the most strongly linked genes to OCD because the genes are associated to many neural processes related to OCD and diseases that are related to OCD.
- DRD2 and ADORA2A may also be strongly linked to OCD due to the fact that OCD can be caused by irregular amounts of dopamine, and both candidate genes are dopamine receptors. This can also explain the association with OCD and addiction since both rely on dopamine and have similar candidate genes.
- There is a high amount of gene expression for all donors in the caudate nucleus which plays a role in motor processes and is associated with Parkinson's disease
- Considering that SLC6A3 and ADORA2A's inclusion in the network is based on experimental evidence, the genes are good candidates for future study.

## Key References

[https://en.wikipedia.org/wiki/Obsessive-compulsive\\_disorder#Associated\\_conditions](https://en.wikipedia.org/wiki/Obsessive-compulsive_disorder#Associated_conditions)