Researching Genes Associated with Huntington's Disease in the Striatum

Bobby Bourque, Woburn High School, Woburn, MA, 01801
BioScience Project, Wakefield, MA 01880

Background

- Huntington's Disease (HD) is a genetic disorder
- It is a dominant trait so it is easily passed on through families
- The disease breaks down nerve cells which leads to involuntary locomotion issues
- Early signs include behavioral changes associated with eating, disrupted sleeping cycles and depression
- Juvenile HD is less common but still as damaging

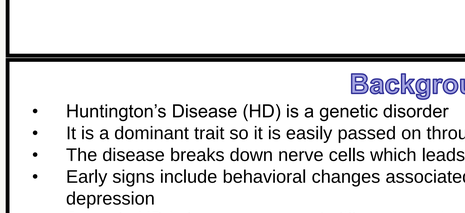
Methods

1. The Allen brain Atlas (http://human.brainn-map.org/) is a database used to collect data using different search options such as a differential search between the structure(s) of interest and the rest of the brai (grey matter) or a correlation search with a specific gene in the structure of interest. Data for the heat maps was collected from donors H0351.2001, H0351.2002, H0351.1009, and H0351.1012.
2. Venny 2.1.0 (http://bioinfogp.cnb.csic.es/tools/venny/) was used to generate a Venn diagram showing the unique and common genes between the four donors.
3. Python Anywhere (https://www.pythonanywhere.com/) was used to create a script that would generate valuable statistics about the data entered from the donors. It also represented the data as a histogram.
4. DAVID (https://david.ncifcrf.gov) was used to identify genes of interest based on phenotypic traits that are responsible and/or implicated in HD. The genes were entered and identified using the official gene symbol and then the functional annotation table was analyzed using key terms such as "behavior" and "cycle".
5. The STRING database (http://string-db.org) was used to find genes that associate with the genes of interest determined after identification with DAVID. STRING creates a protein interaction network related to the central gene based on experimental evidence. The maximum amount of interactions was set to 1000 in order to include all possible interactions.
6. GeneWeaver (http://geneweaver.org/) was used to check if the genes discovered with the STRING database were linked to HD by checking an expansive library of community submitted gene sets from a variety of experimental data submitted by many different laboratories.

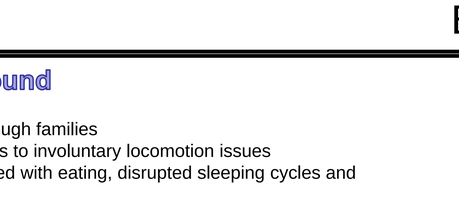
Results

1. Gene Expression Profile

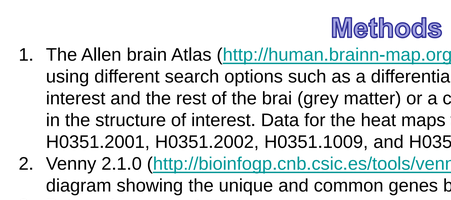
Donor .1009



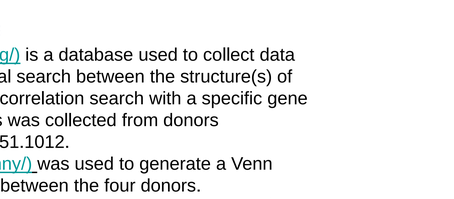
Donor .1012



Donor .2001

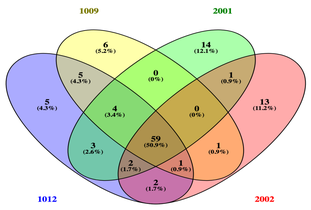


Donor .2002



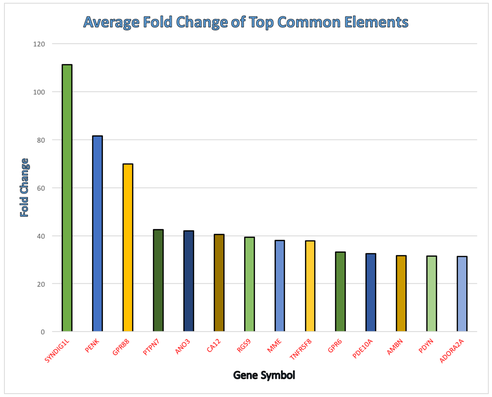
The 4 donor's heat maps show the gene and fold change from highest to lowest using a differential search between the Striatum and grey matter.

2. Venn Diagram Showing Gene Overlap



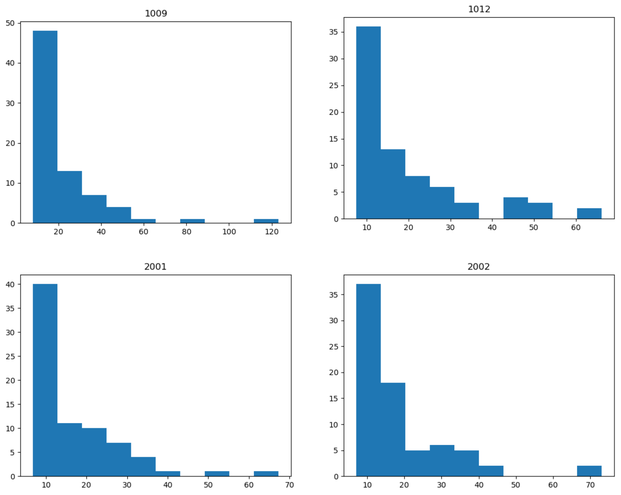
Venn Diagram shows that there are 59 common genes (50.9%) between the four data sets from the four donors.

3. Top Genes with Average Fold Change



The top expressing genes were selected by finding the average fold change values for each of the 4 donors. Some of the genes were common in the top ten across all four donors, like SYNDIG1L. Based on gene ontology description, several of these genes are relevant to HD. ADORA2A : sleep/wake cycles, prepulse inhibition, locomotion, PENK: aggressive behavior, addiction, fear response, ANO3: genetic neuromuscular disorder Dystonia24-, repetitive muscle contractions and movements, freezing postures.

4. Statistics and Variance Histograms




```
stats
              1009          1012          2001          2002
count    75.000000     75.000000     75.000000     75.000000
mean     21.710707     19.492267     16.411493     17.843907
std      18.683959     13.953145     11.184669     13.124141
min       7.776000      7.465000      6.735000      7.013000
25%      10.364500     10.124500      8.666000      8.644500
50%      15.819000     13.801000     11.815000     13.757000
75%      26.964000     24.115500     21.388500     20.440500
max     123.312000     66.174000     67.335000     73.076000
variance
1009     349.090309
1012     194.690265
2001     125.096811
2002     172.243072
dtype: float64
```

All of the charts, with the y-axis being the number of genes and the x axis being the fold change, are skewed right because there are low fold changes corresponding to many of the genes. There are also outliers of genes that have an unusually high fold change and are isolated on the right side of the chart. One can also see that there is usually only one peak in the data within the first bin. The mean fold changes of the data is mostly similar with Donor 1009 having the highest, most likely due to its maximum fold change being the highest as well.

5. Genes of Interest

GPR88	G protein-coupled receptor 88(GPR88)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	cell surface receptor signaling pathway, G-protein coupled receptor signaling pathway, locomotory behavior, neuronal action potential, neuromuscular process controlling balance, motor learning,		

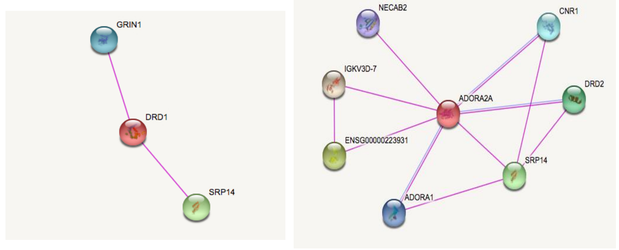
ADORA2A	adenosine A2a receptor(ADORA2A)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	synaptic transmission, dopaminergic, adenosine receptor signaling pathway, response to amphetamine, cAMP biosynthetic process, regulation of transcription, DNA-templated, negative regulation of protein kinase activity, phagocytosis, apoptotic process, inflammatory response, cellular defense response, adenylate cyclase-modulating G-protein coupled receptor signaling pathway, adenylate cyclase-activating G-protein coupled receptor signaling pathway, protein kinase C-activating G-protein coupled receptor signaling pathway, cell-cell signaling, synaptic transmission, cholinergic, central nervous system development, blood coagulation, sensory perception, locomotory behavior, blood circulation, negative regulation of cell proliferation, positive regulation of adenylate cyclase activity involved in G-protein coupled receptor signaling pathway, positive regulation of glutamate secretion, positive regulation of acetylcholine secretion, neurotransmission, regulation of norepinephrine secretion, positive regulation of cAMP biosynthetic process, response to caffeine, positive regulation of synaptic transmission, GABAergic, synaptic transmission, glutamatergic, positive regulation of urine volume, positive regulation of renal sodium excretion, negative regulation of locomotion, vasodilation, response to drug, eating behavior, negative regulation of vascular permeability, negative regulation of cysteine-type endopeptidase activity involved in apoptotic process, negative regulation of neuron apoptotic process, cellular protein metabolic process, positive regulation of circadian sleep/wake cycle, sleep, negative regulation of alpha-beta T cell activation, astrocyte activation, regulation of synaptic plasticity, neuron projection morphogenesis, positive regulation of protein secretion, negative regulation of inflammatory response, regulation of mitochondrial membrane potential, membrane depolarization, regulation of calcium ion transport, positive regulation of synaptic transmission, glutamatergic, excitatory postsynaptic potential, inhibitory postsynaptic potential, prepulse inhibition, positive regulation of apoptotic signaling pathway,		

DRD1	dopamine receptor D1(DRD1)	Related Genes	Homo sapiens
BIOCARTA	Regulation of ck1/cdk5 by type 1 glutamate receptors,		
GOTERM_BP_DIRECT	temperature homeostasis, conditioned taste aversion, behavioral fear response, synaptic transmission, dopaminergic, response to amphetamine, protein import into nucleus, G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger, adenylate cyclase-activating G-protein coupled receptor signaling pathway, activation of adenylate cyclase activity, adenylate cyclase-activating dopamine receptor signaling pathway, synapse assembly, learning, memory, mating behavior, grooming behavior, adult walking behavior, visual learning, positive regulation of adenylate cyclase activity involved in G-protein coupled receptor signaling pathway, astrocyte development, dopamine transport, transmission of nerve impulse, neuronal action potential, dentate gyrus development, striatum development, cerebral cortex GABAergic interneuron migration, positive regulation of cell migration, peristalsis, positive regulation of cAMP biosynthetic process, operant conditioning, regulation of dopamine metabolic process, vasodilation, dopamine metabolic process, response to drug, maternal behavior, positive regulation of potassium ion transport, glucose import, habituation, sensitization, behavioral response to cocaine, positive regulation of release of sequestered calcium ion into cytosol, positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G-protein coupled signaling pathway, regulation of dopamine uptake involved in synaptic transmission, positive regulation of synaptic transmission, glutamatergic, prepulse inhibition, phospholipase C-activating dopamine receptor signaling pathway, long-term synaptic potentiation, long term synaptic depression, cellular response to catecholamine stimulus,		

PENK	proenkephalin(PENK)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	osteoblast differentiation, behavioral fear response, response to hypoxia, startle response, aggressive behavior, signal transduction, neuropeptide signaling pathway, chemical synaptic transmission, lacrimation, sensory perception, response to radiation, response to toxic substance, glial cell proliferation, sensory perception of pain, response to estradiol, response to lipopolysaccharide, cellular response to oxidative stress, response to nicotine, locomotory exploration behavior, response to morphine, response to ethanol, response to calcium ion, general adaptation syndrome, behavioral process, cellular response to vitamin D, cellular response to cAMP, cellular response to transforming growth factor beta stimulus, response to epinephrine, cellular response to virus, positive regulation of behavioral fear response,		
GOTERM_CC_DIRECT	extracellular region, plasma membrane, dendrite, symmetric synapse, neuronal cell body, perikaryon, axon terminus, cell body fiber		

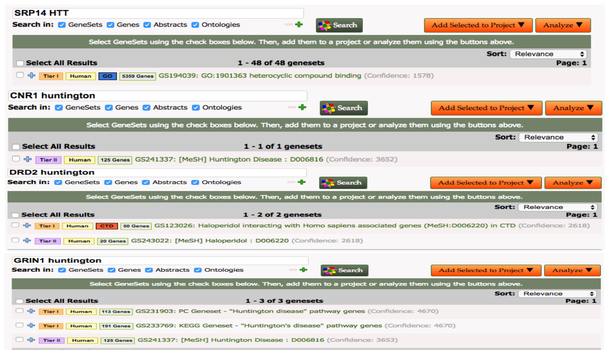
The genes of interest: GPR88, ADORA2A, DRD1, and PENK, were selected based on the gene ontology GO terms for biological process processes which gave support for their relation to HD. Some of the more interesting processes were the responses to drugs and other behavioral descriptions. It was interesting to see "negative regulation of locomotion" because that prevents locomotion in an organism and HD can be exhibited through muscle spasms and reduced control in the early stages.

6. Protein Interaction Networks



The STRING parameters for the network analysis were set to Homo sapiens and experimental interactions. There was a maximum interaction limit of one thousand to set include all of the possible interactions. The interactions of the genes were interesting. PENK had none (PENK network not shown), ADORA2A had seven and high degree of interconnectivity between DRD1 had 2 and GPR88 had 1 (GPR88 network not shown).

7. Checking Genes Related to Huntington's Disease



All of the genes were checked to see if they had been associated with HD in earlier experiments. Four of the gene were found in genesets verifying the relation. It was interesting that, while SRP14 had no gene sets relating it to "HD", there were gene sets linking it to HTT., the HD gene.

Conclusions

- This project was conducted using data mining, data wrangling, statistics, gene profiling, and protein networks genes.
- Genes that were prevalent in the striatum in comparison to the rest of the brain were identified using the Allen Brain Atlas and cleaned up in Excel with their fold change values.
- DAVID was used to identify genes of interest: ADORA2A, PENK, GPR88, and DRD1. These genes were selected due to the relevance to HD.
- The genes were then entered into the STRING database to discover other genes based on experimental interactions. DRD1 and GRP88 were linked only to genes related to HD, and PENK had no interactions.
- Evidence that these genes were related to HD was found in different GeneWeaver genesets.
- SRP14, GRIN1, CNR1, and DRD2 should be further investigated for their potential role in HD.
- SRP14 appears the most frequently and so it should be investigated also due to its prevalence in the protein interaction networks, being connected to ADORA2A, DRD1, and GPR88.