Finding Genes Related to ADHD

Aaron Mink, Saint Dominic Savio Catholic High School, Austin, TX, 78726
BioScience Project, Wakefield, MA, 01880

Background

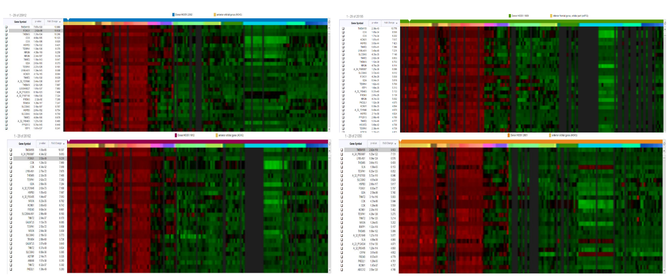
- Attention-Deficit/Hyperactivity Disorder or “AD/HD” develops from an imbalance or deficiency in certain neurotransmitters.
- This leads to symptoms such as inattention, hyperactivity, and impulsivity.
- Recent studies hypothesize a connection between inefficient use of dopamine in the frontal lobe and AD/HD.
- In this study we used an integrative genomics approach to profile gene expression patterns that may be related to ADHD

Methods

- The Allen Brain Atlas (http://human.brain-map.org/) is a database used to collect gene expression data for the chosen brain regions in comparison with the grey matter using a differential search. Data for the heat maps was collected from H0351.2001, H0351.2002, H0351.1009, H0351.1012.
- Venny 2.1.0 (http://bioinfogp.cnb.csic.es/tools/venny/) was used to compare the data from the four donors in order to find common genes between the donors. A venn diagram is created to visualize the data.
- Python Anywhere (https://www.pythonanywhere.com/) is a programming tool used to find statistics and variance of the fold change values in the four gene lists. The program also creates histograms for visual representation.
- DAVID (https://david.ncifcrf.gov) is a bioinformatics tool that subdivided the gene lists based on varying criteria related to function. Gene lists were sorted using the “official gene symbol” identifier and limited to annotations of “homo sapiens”. Functional annotation tools were used to analyze the results.
- Relevant genes were entered into the STRING database (http://string-db.org) to identify potential interacting partners, pathways, and other genes relating to ADHD/ADD. The database consists of networks with experimentally validated interactions.
- Gene Weaver (http://geneweaver.org/) combines cross-species data and gene entity integration, scalable hierarchical analysis of user data with a community-built and curated data archive of gene sets and gene networks, and tools for data driven comparison of user-defined biological, behavioral and disease concepts.

Results

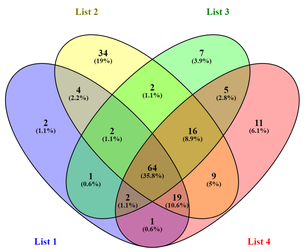
1. Gene Expression Profile



The heat maps show microarray data for four donors, in which gene expression in the frontal lobe is compared to grey matter. Each colored column represents a tissue sample. The microarray data is collected from mRNA and is then copied into cDNA and labeled and hybridized to an array containing all human genes. Two different samples are used for comparison: the sample under study and the control. The genes are mapped by color: Red indicates that the gene is expressed more in the subject under study than the control, green indicates it is expressed less, and black means the genes are expressed the same in the subject under study and the control.

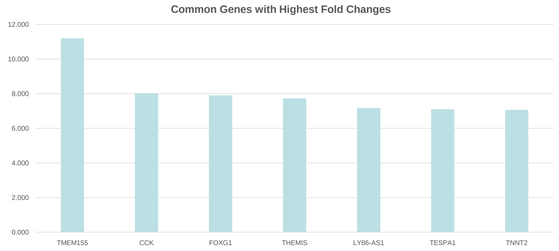
Genes with fold changes of 3 or more were downloaded for analysis. From donor H0351.2001, 139 genes were collected. 248 genes were collected from donor H0351.2002 , 145 genes from donor H0351.1009, and 183 genes from H0351.1012.

2. Gene Overlaps Among Donors



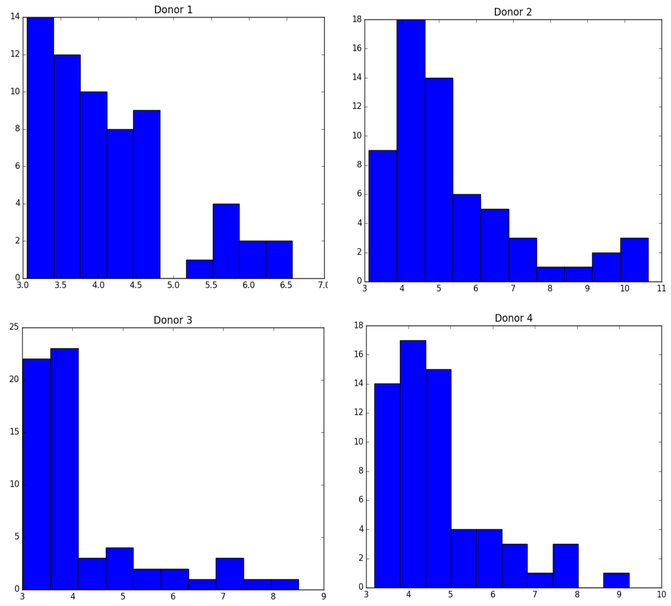
The Venn Diagram shows that there are 64 common genes between the four donors. This means that 35.8% of the genes that were initially downloaded were common among donors H0351.2001, H0351.2002, H0351.1009, and H0351.1012.

3. Top 20 Genes with Highest Fold Changes



This graph represents the 9 genes with the highest fold changes, averaged from the 4 donors. The genes are TMEM155, CCK, FOXG1, THEMIS, LY86-AS1, TESPA1, and TNNT2. TMEM155 is a protein-coding gene. CCK encodes a preprotein in the gastrin/cholecystokinin family of proteins. FOXG1 encodes a member of the forked-head transcription factor family. The encoded protein, which functions as a repressor, may play a role in brain development. THEMIS encodes a protein that plays a regulatory role in both positive and negative T-cell selection during late thymocyte development. LY86-AS1 is an ncRNA gene (non-coding RNA), meaning its function is to regulate gene expression. TESPA1 is a protein-coding gene. TNNT2 encodes the protein that is the tropomyosin-binding subunit of the troponin complex

4. Statistics and Variance Changes




```
stats
          Donor 1    Donor 2    Donor 3    Donor 4
count  62.000000  62.000000  62.000000  62.000000
mean    4.159855   5.381387   4.190097   4.700677
std     0.905988   1.795663   1.284875   1.296255
min     3.058000   3.116000   3.015000   3.201000
25%     3.480500   4.276500   3.424500   3.857250
50%     3.947500   4.735000   3.664000   4.388500
75%     4.627500   5.993000   4.308500   5.011250
max     6.576000  10.639000   8.500000   9.230000
variance
Donor 1    0.820814
Donor 2    3.224406
Donor 3    1.650903
Donor 4    1.680277
dtype: float64
>>>
```

Histograms (above): In each of the 4 donors, the frequency of genes (y-axis) is heavily skewed for smaller fold change (x-axis). The mean fold change values are very close for Donors 1,2, and 3, while Donor 4 has a slightly higher mean fold change. Statistics table (below): For donors 1,3, and 4 the means are comparable whereas for Donor 2, the mean is higher indicating greater spread in the dataset.

5. Genes of Interest

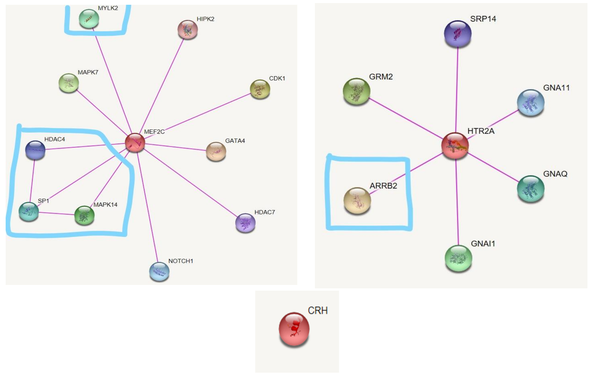
CRH	corticotropin releasing hormone(CRH)	Related Genes	Homo sapiens
BBID	110.CRHandLC-NA.		
GOTERM_BP_DIRECT	positive regulation of protein phosphorylation, synaptic transmission, dopaminergic, glucocorticoid biosynthetic process, inflammatory response, signal transduction, chemical synaptic transmission, female pregnancy, parturition, learning or memory, positive regulation of cell proliferation, associative learning, hormone-mediated apoptotic signaling pathway, positive regulation of gene expression, negative regulation of gene expression, negative regulation of norepinephrine secretion, positive regulation of circadian sleep/wake cycle, wakefulness, positive regulation of cell death, regulation of serotonin secretion, determined metabolic process, hypothalamus development, lung development, adrenal gland development, positive regulation of cAMP biosynthetic process, negative regulation of epinephrine secretion, negative regulation of luteinizing hormone secretion, locomotory exploration behavior, positive regulation of insulin secretion involved in cellular response to glucose stimulus, response to immobilization stress, negative regulation of circadian sleep/wake cycle, REM sleep, response to drug, response to estrogen, response to ethanol, response to ether, negative regulation of blood pressure, response to pain, ion homeostasis, response to corticosterone, positive regulation of corticotropin secretion, positive regulation of cortisol secretion, long term synaptic potentiation, positive regulation of digestive system process, negative regulation of cell death, negative regulation of glucagon secretion, cellular response to cocaine, cellular response to dexamethasone stimulus, positive regulation of calcium ion import, regulation of N-methyl-D-aspartate selective glutamate receptor activity, positive regulation of corticosterone secretion, positive regulation of behavioral fear response.		
GOTERM_CC_DIRECT	extracellular region, extracellular space, cytoplasm, varicosity, perikaryon.		
GOTERM_MF_DIRECT	receptor binding, hormone activity, neuropeptide hormone activity, protein binding, corticotropin-releasing hormone activity, corticotropin-releasing hormone receptor 1 binding, corticotropin-releasing hormone receptor 2 binding.		
INTERPRO	Corticotropin-releasing factor, CRF, Urocortin/corticotropin-releasing factor, Corticotropin-releasing factor conserved site.		
KEGG_PATHWAY	Long-term depression, Alcoholism.		
SMART	CRF.		
UP_KEYWORDS	3D-structure, Amidation, Cleavage on pair of basic residues, Complete proteome, Direct protein sequencing, Hormone, Polymorphism, Reference proteome, Secreted, Signal.		
UP_SEQ_FEATURE	helix, modified residue, peptide:Corticoliberin, signal peptide,		

HTR2A	5-hydroxytryptamine receptor 2A(HTR2A)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	temperature homeostasis, cellular calcium ion homeostasis, smooth muscle contraction, adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway, phospholipase C-activating G-protein coupled receptor signaling pathway, activation of phospholipase C activity, phospholipase C-activating serotonin receptor signaling pathway, serotonin receptor signaling pathway, chemical synaptic transmission, going, primary, cell death, positive regulation of cell proliferation, positive regulation of phosphatidylinositol biosynthetic process, regulation of dopamine secretion, phosphatidylinositol 3-kinase signaling, artery smooth muscle contraction, urinary bladder smooth muscle contraction, sleep, response to drug, negative regulation of potassium ion transport, positive regulation of MAP kinase activity, protein localization to cytoskeleton, positive regulation of fat cell differentiation, positive regulation of glycolytic process, positive regulation of vasoconstriction, viral entry into host cell, regulation of hormone secretion, behavioral response to cocaine, positive regulation of peptidyl-tyrosine phosphorylation, regulation of behavior, detection of temperature stimulus involved in sensory perception of pain, detection of mechanical stimulus involved in sensory perception of pain, release of sequestered calcium ion into cytosol, negative regulation of synaptic transmission, glutamatergic, positive regulation of ERK1 and ERK2 cascade,		
GOTERM_CC_DIRECT	cytosol, plasma membrane, integral component of plasma membrane, caveola, cytoplasmic, membrane-bounded vesicle, axon, dendrite,		

MEF2C	myocyte enhancer factor 2C(MEF2C)	Related Genes	Homo sapiens
BIOCARTA	ALK in cardiac myocytes, Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling, CARM1 and Regulation of the Estrogen Receptor, Role of EAG in Neuronal Survival, Control of skeletal myogenesis by HDAC & calcium/calmodulin-dependent kinase (CaMK), MAPKinase Signaling Pathway, Signal Dependent Regulation of Myogenesis by Coregressor MITR, NFAT and Hypertrophy of the heart (Transcription in the broken heart), p38 MAPK Signaling Pathway, Regulation of PGC-1a.		
COG_ONTOLOGY	Transcription.		
GOTERM_BP_DIRECT	negative regulation of transcription from RNA polymerase II promoter, MAPK cascade, blood vessel development, osteoblast differentiation, neuron migration, B cell homeostasis, heart looping, endochondral ossification, blood vessel remodeling, chondrocyte differentiation, germinal center formation, regulation of germinal center formation, response to ischemia, primary heart field specification, secondary heart field specification, outflow tract morphogenesis, sinoatrial valve morphogenesis, cardiac ventricle formation, transcription, DNA-templated, regulation of transcription, DNA-templated, transcription from RNA polymerase II promoter, apoptotic process, humoral immune response, nervous system development, heart development, muscle organ development, skeletal muscle tissue development, muscle cell fate determination, learning or memory, response to virus, positive regulation of gene expression, negative regulation of gene		

All the common genes were entered into DAVID, and 3 genes of interest were found in the Functional Annotation Table. CRH and MEF2C were of interest due to their association with learning and memory. HTR2A is interesting because of its association with the regulation of dopamine, regulation of behavior, and regulation of hormone secretion.

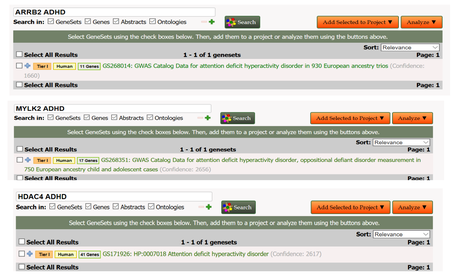
6. Protein Interaction Networks



MEF2C, HTR2A, and CRH were entered into the STRING database, and selected for interactions in homosapiens. The sources for interactions were then limited to experiments, and the max interaction number was set to 1000. MEF2C and HTR2A showed small protein interaction networks. CRH had no connections, however. All the genes connected to MEFC2 and HTR2A were entered into Gene Weaver, in order to find which genes might have a direct relation to ADHD.

- MEF2C is a transcription activator which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. It is also plays an essential role in hippocampal-dependent learning and memory by suppressing the number of excitatory synapses.
- HTR2A is a serotonin receptor that is coupled with G proteins, which activate a phosphatidylinositol-calcium second messenger system.
- CRH regulates the release of corticotropin from pituitary gland. Corticotropin, also known as adrenocorticotropic hormone, is released as a response to stress or adrenal insufficiency.

7. Genes Related to ADHD



The GeneWeaver database was queried to determine if the genes found in the HTR2A and MEF2C interaction networks have been previously linked with ADHD. ARRB2, MYLK2, and HDAC4 had in fact been reported in studies on ADHD. Each of the genes was identified in distinct experiments.

- ARRB2 is protein-coding gene and a member of the arrestin/beta-arrestin protein family, which is are thought to participate in agonist-mediated desensitization of G-protein-coupled receptors and cause specific dampening of cellular responses to stimuli such as hormones, neurotransmitters, or sensory signals.ARRB2 was shown to inhibit beta-adrenergic receptor function in vitro. It is expressed at high levels in the central nervous system and may play a role in the regulation of synaptic receptors.
- HDAC4, or histone deacetylation 4, is a protein-coding gene. Histone deacetylation gives a tag for epigenetic repression and plays an important role in transcriptional regulation, cell cycle progression and developmental events.
- MYLK2 is a protein-coding gene that only encodes a myosin light chain kinase. The chain kinase is an calcium-dependent enzyme that is only found in adult skeletal muscle.

Conclusion

- An integrative genomic approach was used for profiling gene expression patterns to identify genes that may be related to ADHD.
- After the genes common to four donor frontal lobes were analyzed, and entered into the DAVID database, three genes of interest were found: CRH, HTR2A and MEF2C.
- These genes were then entered into the STRING database, which showed small protein interaction networks for HTR2A and MEF2C (no interactions in CRH) using experimental evidence as a parameter.
- ARRB2, MYLK2 and HDAC4 were linked to ADHD in three separate experiments. This supports that the results are meaningful and did not occur by chance.
- These results indicate that the other genes in the MEF2C and HTR2A protein interaction networks should be further investigated, since there is a high likelihood that they are related to ADHD. Especially SP1 and MAPK14, since they are not only connected to MEF2C, but also directly connected to HDAC4.