Section 3. General Biology

https://doi.org/10.29013/ELBLS-21-4-34-43

Ou Rachel,
Bridgewater-Raritan High School
Bridgewater, NJ 08807, USA
E-mail: rachelo1500@gmail.com

IDENTIFICATION OF AUTISM SPECTRUM DISORDER RELEVANT GENES USING GENE EXPRESSION AND GENETIC INFORMATION

Abstract. Autism Spectrum Disorder (ASD) refers to a broad range of neurodevelopmental conditions that cause significant communication and behavioral issues. It is mostly a heritable condition, with multiple genes playing pathogenic roles. Although extensive studies have been conducted to establish the genic connections in ASD patients, some gene connections may have been overlooked due to the vast number of genes and the vast amount of information available. In this research, we used statistical tools to analyze public databases in an effort to discover novel genes associated with autism. By comparing DNA and RNA from autistic population and the general population, we identified 31 down-regulated and up-regulated genes in the cerebellum, frontal cortex, and temporal cortex which have a positive connection with ASD. Of these genes, 6 genes GGNBP2, TUBGCP5, ZDHHC8, DHRS11, RABL2B, and PANX2 have been identified to be tightly associated with ASD at RNA and genetic levels. All six genes are novel and they may play a role in the development of ASD.

Keywords: Autism Spectrum Disorder (ASD), differentially expressed genes (DEGs), up-regulated, down-regulated, cerebellum, frontal cortex, temporal cortex.

Introduction:

Autism Spectrum Disorder (ASD) refers to a broad range of neurodevelopmental conditions that cause significant communication and behavioral issues. Symptoms include abnormal facial expressions, repetitive behavior, and delayed language skills. ASD is a heritable condition, with multiple genes playing pathogenic roles. Compared with normal individuals, those that display phenotypes associated with ASD have different genes which have abnormal regulations.

The occurrence of ASD is 62/10,000. Males are four times more likely to suffer than females. A large number of de novo copy-number variations and single-base-pair mutations have been found in ASD patients. In addition to genetic influences, many environmental risk factors have been connected to ASD, such as various pharmaceutical drugs and toxicants. In recent years, ASD research has progressed beyond genes and molecules to circuits and neural connectivity. Structural neuroimaging studies have revealed differences in brain volume and connectivity in those with ASD and those without [1].

Although there is no cure for ASD, many interventions have been developed. These interventions

seek to improve behavioral, cognitive, and social skills. Since each individual with ASD has his/her unique strengths and weaknesses, treatment plans are specific to an individual's needs. Some examples of treatments include behavior analysis, speech therapy, social skills training, and the use of assistive technology [2].

This study seeks to: 1. Identify novel differentially expressed genes (DEGs), which are dysregulated in terms of gene expression in brain tissues of ASD patients; 2. Identify the brain regions in which these DEGs are expressed; and 3. Identify the cell types in which these DEGs are expressed. By understanding the function of these genes, their expression in different brain regions, researchers will be better able to address the genes' roles in ASD, and new methods for diagnosing ASD early on may be revealed. In addition, these abnormal genes may be used as targets for drug development.

Methods:

Data was collected from public databases and analyzed using statistical tools. First, preliminary data was collected from the Gene Expression Omnibus (GEO) database (GSE28521) [3], which is a public functional genomics data repository. Next, a list of the down-regulated and up-regulated DEGs was compiled based on statistical significance (P) and fold change (FC) value. Then, in order to identify key pathways and biological processes involved in ASD, we performed functional enrichment analysis for the up-regulated and down-regulated genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [4], a database including functional annotation tools for large lists of genes. Next, the Genotype-Tissue Expression (GTEx) Portal [5], a public resource for the study of tissue-specific gene expression and regulation, was utilized in order to investigate the brain regions in which these genes are expressed. Then, expression profile analysis using single cell RNA sequencing data was performed to investigate what cell types are mainly associated with the DEGs. Finally, data was compared to the Clinvar database [6], a public archive of reports of the relationships among human variations and phenotypes, to identify overlapping DEGs in order to correlate the data with genetic evidence.

Final DEGs were annotated using annotation tool SOURCE by Princeton (https://source-search.princeton.edu/). The annotations include the gene symbol, full name, Entrez Gene ID, chromosome number, protein subcellular localization, UniProt ID, and a brief functional description.

Results:

Dataset and Workflow

To identify genes tightly associated with ASD, we performed analyses according to the workflow chart shown in Figure 1. Briefly, we used GEO database to retrieve the dataset with GEO accession number GSE28521. The dataset consists of RNA from approximately 100mg of postmortem brain tissue, including cerebellum, frontal cortex, and temporal cortex of autistic and control groups. We identified DEGs in these tissues and selected all down-regulated and up-regulated genes for further enrichment analysis. The purpose of this analysis is to identify potential signaling pathway, metabolic pathway and other enriched physiological processes or molecular functions that are tightly associated with the DEGs. Because differential expression analysis is based on a microarray dataset of which cross-hybridization might cause false-positive results, we further used RNA-sequencing (RNA-seq) data to confirm the expression of DEGs in human brain regions.

Besides gene expression analysis mentioned above, we also used the ClinVar database to investigate ASD relevant genes that were altered by gene variation such as mutation and SNP (Single Nucleotide Polymorphism). Genes that were common from both genetic analysis (ClinVar) and expressional analysis are more likely to be connected to ASD. Therefore, we paid more attention to these genes and made detailed annotations.

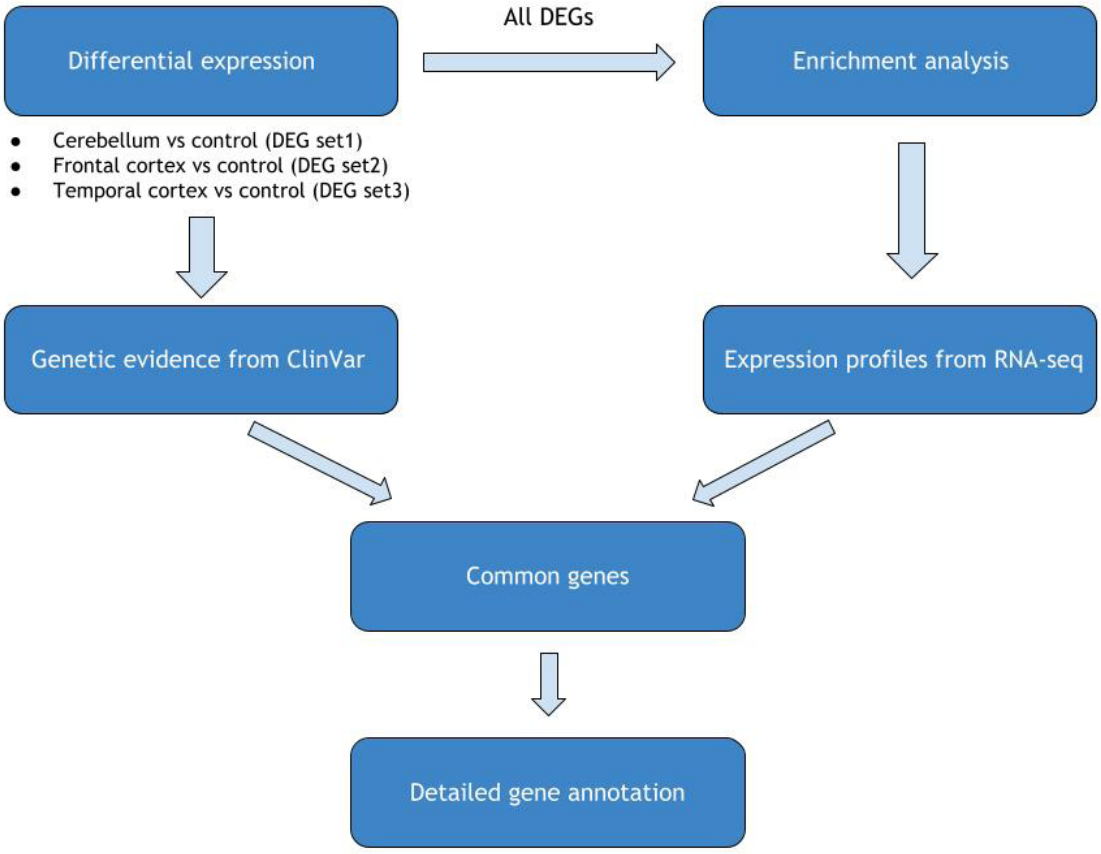

Figure 1. Workflow of the current study

Identification of DEGs Using Micro Arraying Data

A total of 31 down-regulated and up-regulated genes were identified in the cerebellum, frontal cortex, and temporal cortex. For all analyses, the P-value was <= 0.05, indicating statistical significance. The criteria for up-regulated genes is Fold Change (FC) value >= 1.5, which means that the average gene expression level of a gene in ASD group is at least 1.5 times higher than that in the control group. For down-regulated genes, the FC value was set <= 0.6. All the DEGs are listed in Table 1 (cerebellum), Table 2 (frontal cortex) and Table 3 (temporal cortex).

Table 1 lists DEGs in cerebellum of ASD patents. Cerebellum is a brain region that coordinates voluntary movements such as posture, coordination, speech, and smooth muscular activity. A total of 9 DEGS were identified: 7 up-regulated genes, and 2 down-regulated genes.

Table 1.– DEGs of cerebellum in ASD patients

Gene Symbol	Gene Title	Fold Change	P value	Direction
1	*2*	*3*	*4*	*5*
EPHB6	EPH receptor B6	1.616	0.00227	Up
NES	Nestin	1.502	0.00456	Up
BCHE	Butyrylcholinesterase	1.583	0.0102	Up
LRRC37A4P	leucine rich repeat containing 37 member A4, pseudogene	1.661	0.0114	Up
NDRG2	NDRG family member 2	1.641	0.0173	Up

1	*2*	*3*	*4*	*5*
FAM181B	family with sequence similarity 181 member B	1.641	0.0230	Up
NTM	Neurotrimin	1.510	0.0321	Up
MYH11	myosin, heavy chain 11, smooth muscle	0.562	0.0180	Down
CPLX3	complexin 3	0.577	0.0276	Down

Table 2 lists the down regulated DEGs in the frontal cortex of ASD patients. Frontal cortex is a brain region in control of memory, emotions and impulse control. 10 down-regulated genes were identified.

Table 2.– Down-regulated DEGs of frontal cortex in ASD patients

Gene Symbol	Gene Title	Fold Change	P value
STAT4	signal transducer and activator of transcription 4	0.504	0.000116
HAPLN4	hyaluronan and proteoglycan link protein 4	0.489	0.000119
VAMP1	vesicle associated membrane protein 1	0.582	0.000436
PVALB	Parvalbumin	0.424	0.000691
TUBGCP5	tubulin, gamma complex associated protein 5	0.510	0.000810
SCN1B	sodium voltage-gated channel beta subunit 1	0.563	0.00273
GABRG2	gamma-aminobutyric acid type A receptor gamma2 subunit	0.599	0.00397
GAD2	glutamate decarboxylase 2	0.594	0.0122
RTN4	reticulon 4	0.572	0.0319
RTN1	reticulon 1	0.525	0.0362

Table 3.– Down-regulated DEGs of temporal cortex in ASD patients

Gene Symbol	Gene Title	Fold Change	P value
NEFH	neurofilament heavy polypeptide	0.441	0.0000916
PCSK1	proprotein convertase subtilisin/kexin type 1	0.533	0.000108
HAPLN4	hyaluronan and proteoglycan link protein 4	0.407	0.000110
CCDC184	coiled-coil domain containing 184	0.575	0.000206
VAMP1	vesicle associated membrane protein 1	0.585	0.00113
ADCYAP1	adenylate cyclase activating polypeptide 1	0.589	0.00159
STAT4	signal transducer and activator of transcription 4	0.539	0.00267
VGF	VGF nerve growth factor inducible	0.558	0.00302
NGEF	neuronal guanine nucleotide exchange factor	0.592	0.00360
PVALB	Parvalbumin	0.340	0.00433
SCN1B	sodium voltage-gated channel beta subunit 1	0.438	0.00560
NSG1	neuron specific gene family member 1	0.551	0.00619
CRH	corticotropin releasing hormone	0.493	0.0189
GAD1	glutamate decarboxylase 1	0.536	0.0194
GAD2	glutamate decarboxylase 2	0.226	0.0200

In the temporal cortex, a brain region responsible for auditory stimuli, memory, and speaking, 15 down-regulated genes were identified (Table 3).

These results suggest that different brain regions of ASD patients are affected and the dysregulated genes are not exactly the same in different brain regions, which further suggests the complexity of

the disease. On the other hand, we found that there are six co-downregulated genes *STAT4*, *HAPLN4*, *VAMP1*, *PVALB*, *SCN1B* and *GAD2* in the frontal cortex and temporal cortex regions. These genes should play an important role in the development of ASD disease.

Functional Enrichment Analysis of DEGs

Next, we performed functional enrichment analysis using the DAVID tool on the up- and down-regulated DEGs to reveal the key pathways and biological processes associated with the genes. We did not obtain any significant enrichment results based on the up-regulated genes in Table 1 when p value was set less than or equal to 0.05. This may be the result from the limited gene counts. However, when we used all of the down-regulated genes for analysis, we found that a lot of biological processes and pathways are involved in these genes (Figure 2). For example, the most significant process is neurogenesis, which is the process in which new neurons are formed in the brain. This process is crucial for brain development in embryos, but it also continues after birth in certain brain regions. The involved genes related to neurogenesis include ADCYAP1, RTN4 and NGEF, and so on. Therefore, the enriched process reveals that the ability to form new neurons in ASD patients is damaged because of down-regulation of relevant genes.

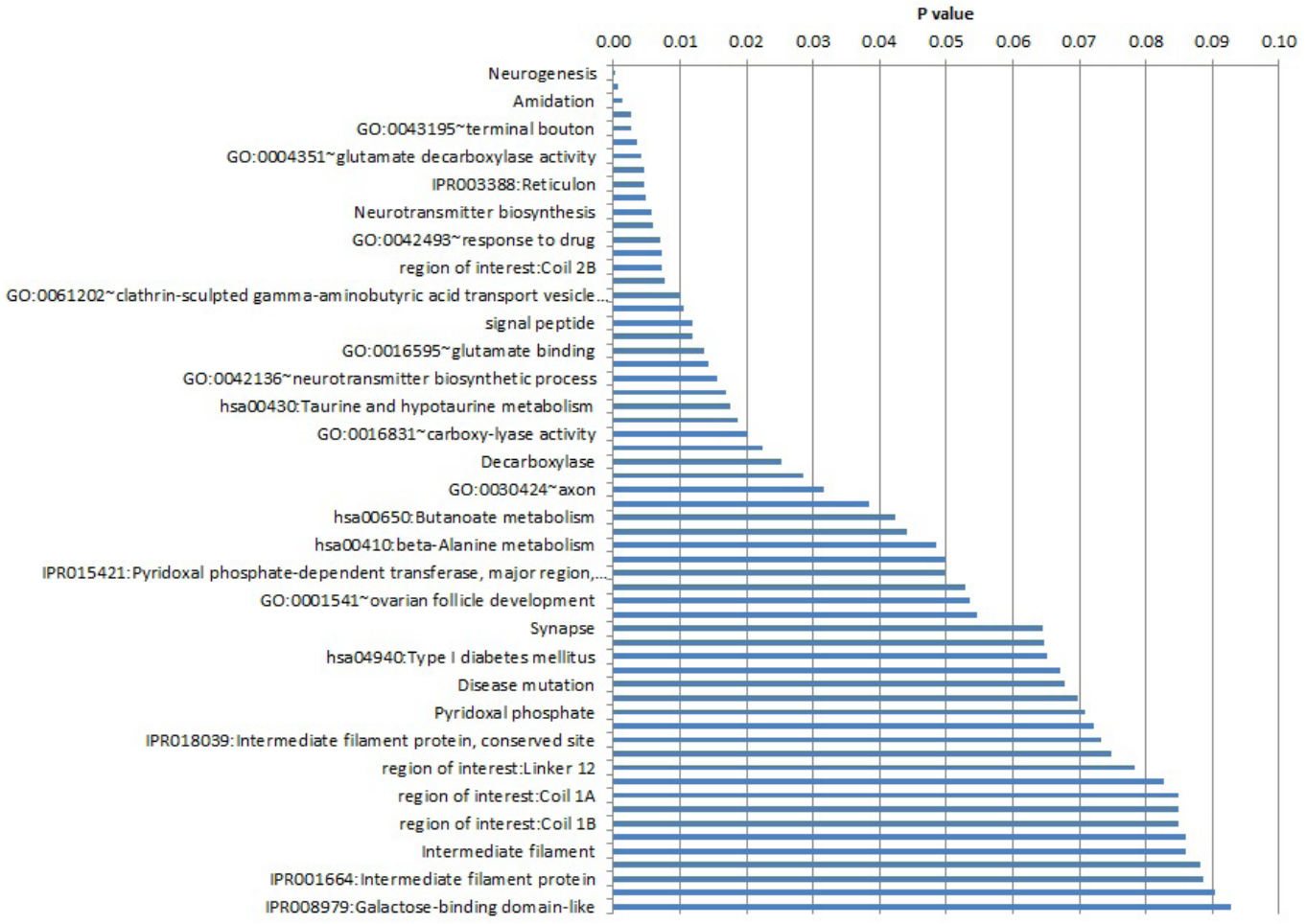

Figure 2. Functional enrichment analysis using DAVID tool.

Besides neurogenesis, another important process is amidation, which represents a type of protein modification. The decreased process indicates that a decreased neuropeptide amidation could occur in ASD patients. The third process is terminal bouton, also known as axon terminal or synaptic bouton. It is

the far end of a neuron's axon and is crucial for neural communication. This process may also be damaged in ASD patients. Other processes such as reticulon, neurotransmitter biosynthesis and glutamate binding, and so on, are also enriched in the downregulated genes, suggesting an extensive affection of brain function in ASD patients.

Expression Profile of DEGs Using RNA Sequencing Data

The above DEGs were identified using microarray data. Next, RNA sequencing data was analyzed to validate the expression of DEGs in human tissue. In particular we wanted to see whether these genes show specific expression in human brain regions. Results from analyzing the data using the GTEx portal reveal the brain regions in which these DEGs are expressed: cerebellum, cerebellar hemipster, hypothalamus, anterior cingulate cortex, frontal cortex, and so on. As shown in Figure 3, we can see all of these genes can be expressed in human brain tissue. However, DEGs vary in their expression levels. For example, NDRG2, RNT4 and RNT1 are highly expressed in all regions compared with other DEGs. However, we found that MYH11, BCHE, CRH and ADCYAP1 show very low expression levels in the brain tissues. BCHE is upregulated in the cerebellum of ASD patients (Table 1), whereas the other three genes are further downregulated. The mechanism and significance still await further investigation.

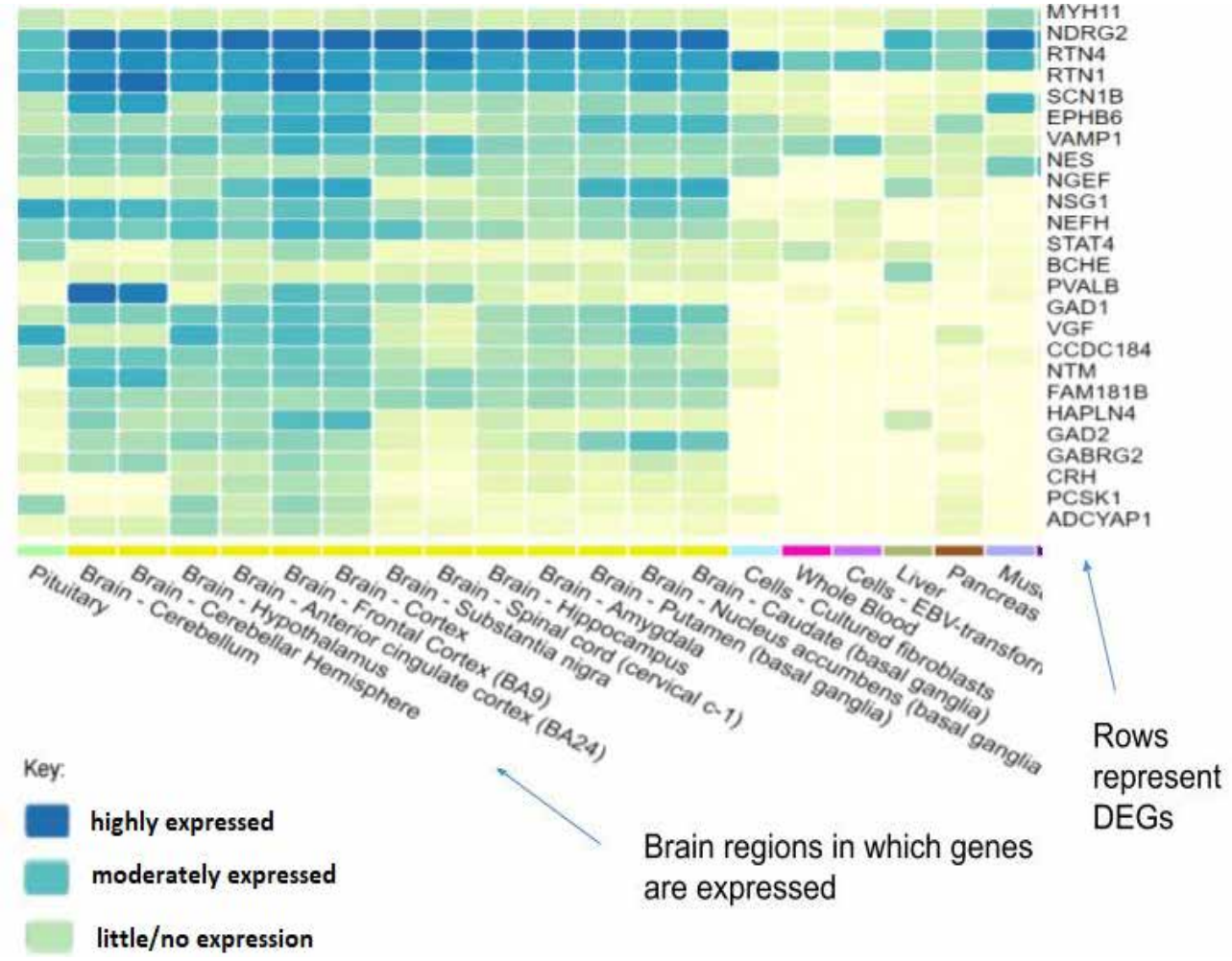

Figure 3. Expression levels of selected DEGs in brain regions using GTEx RNA-seq data. Each block represents gene expression level indicated by the color bar

Identification of gene variation relevant to ASD

The above analyses were based on gene expression at RNA level, and we don't know whether there is genetic change of these genes that also represents the potential mechanism to cause ASD. Therefore, we searched the ClinVar database, a public archive of reports of the relationships among human variations and phenotypes, to find genomic variations related to ASD. If a gene mutation happens to be located on the above differential genes, it means that the likelihood of this gene playing a more important role in the etiology of ASD is strong based on both the heredity factor and the gene expression factor. Consequently, we compared data used in this research and data from ClinVar. We found that there are 184 non-redundant genes and several chromosome regions recorded in the database to be associated with ASD. Surprisingly, there was no overlap between these genes and the DEGs in Tables 1, 2 and 3. The count of DEGs mainly depends on filter condition, such as fold change. If we change the filter conditions during DEG identification, some differential genes may have genetic changes at the same time.

As shown in Table 4, 6 downregulated genes including *GGNBP2, TUBGCP5, ZDHHC8, DHRS11, RABL2B, PANX2* (Table 4) were identified to be overlapped when only p value was used as filter condition. The gene *TUBGCP5* showed the maximum reduction with a fold change 0.51.

Table 4. – Down-regulated DEGs with simultaneous genetic variation in ASD.

Gene Symbol	Gene Title	Fold Change	P value	Region	Protein subcellular localization	Functional description
GGNBP2	gametogenetin binding protein 2	0.788	0.000174	frontal cortex	cytoplasmic vesicle; associated with vesicular structures	may be involved in spermatogenesis
TUBGCP5	tubulin, gamma complex associated protein 5	0.51	0.00081	frontal cortex	cytoplasm, cytoskeleton, microtubule organizing center, centrosome.	gamma-tubulin complex is required for microtubule nucleation at the centrosome.
ZDHHC8	zinc finger, DHHC-type containing 8	0.722	0.00792	frontal cortex	cytoplasmic vesicle membrane; multi-pass membrane protein	palmitoyltransferase involved in glutamatergic transmission. mediates palmitoylation of abca1
RABL2B	RAB, member of RAS oncogene family-like 2B	0.732	0.00757	temporal cortex	secreted	oxidoreductase activity, oxidation-reduction process
PANX2	pannexin 2	0.987	0.00446	temporal cortex	intracellular	GTP Binding, GTPase activity, Rab protein signal transduction, intracellular protein transport
DHRS11	dehydrogenase/reductase (SDR family) member 11	0.796	0.000216	temporal cortex	cell membrane, multi-pass membrane protein, cell junction, gap junction	structural component of the gap junctions and the hemichannels

This result from both gene expression and genetic variation suggests that these six genes should play an important role in the etiology of ASD.

Identification of gene expression of ASD-related genes at single cell level

The above gene expression analysis from GEO microarray and GTEX RNA sequencing was based on bulk tissue samples which contained a wide variety of cells. The expression data represented an averaged effect of gene expression across thousands to millions of cells, which might obscure biologically relevant and critical differences between cells. Unlike these traditional methods, single cell RNA sequencing (scRNA-Seq) examines gene expression at single cell resolution, thus providing a better understanding of the gene function in different cell types in the context of tissue microenvironment. In order to determine the specific cell types in which the genes in Table 4 are expressed, a dataset from the UCSC cell browser was used (http://cells.ucsc.edu/?ds=autism). The dataset analyzed single cells in the cortex of ASD patients using single-nucleus RNA sequencing data to identify autism-associated transcriptomic changes in specific cell types [7].

As shown in Figure 4, genes *GGNBP2*, *TUBGCP5*, *ZDHHC8*, and *RABL2B* have similar expression profiles. All of these genes are highly expressed in excitatory neurons, cortico-cortical projection neurons, and somatostatin interneurons. Expression of these genes in endothelial cells, microglia, oligodendrocyte precursor cells (OPCs), and oligodendrocytes is relatively low. Excitatory neurons aid in the electoral transmission of neuronal signals. Cortico-cortical projection neurons communicate by sending action potentials that release glutamate. Somatostatin interneurons facilitate synapses. The results suggest these cell types should be directly involved in the pathogenesis of ASD.

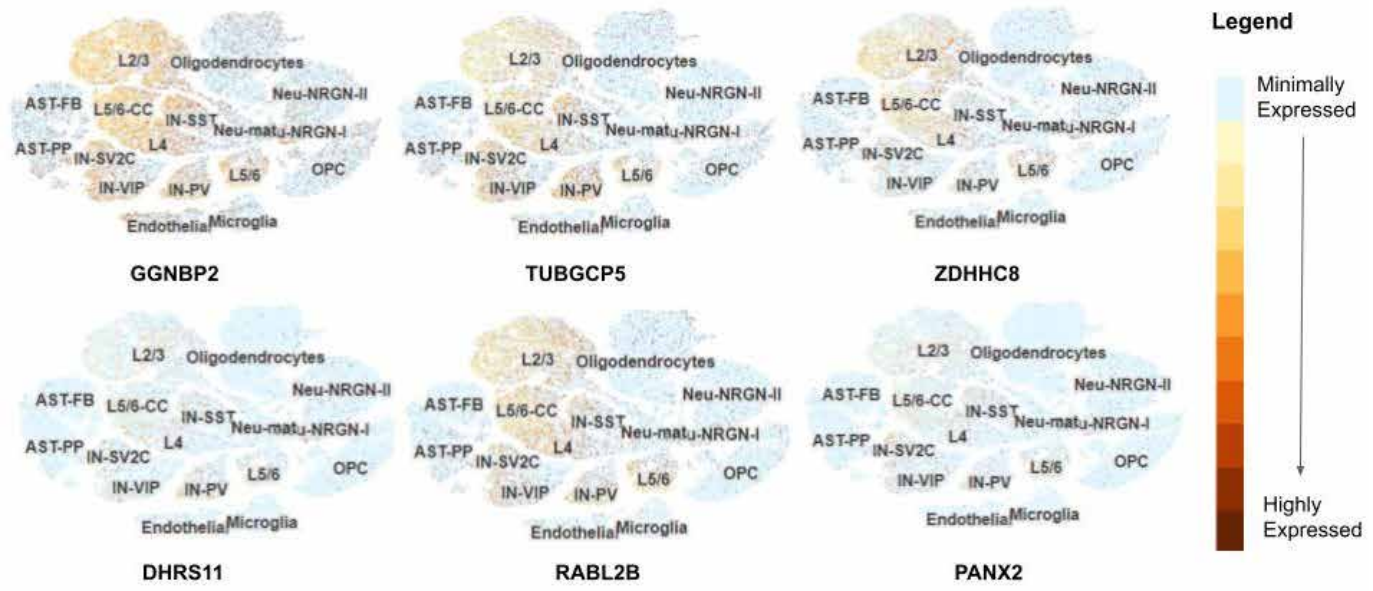

Figure 4. Expression profile of the six genes at single cell resolution in ASD brain tissue. Each dot corresponds to a single cell and each cluster is labeled with the cell type. The clusters include L2/3, IN-SST, IN-PV, IN-VIP and L2/3 represent excitatory neurons. L5/6 represents cortico-cortical projection neurons and IN-SST represents somatostatin interneurons

Discussion:

In the current study, we aimed to identify Autism Spectrum Disorder relevant genes using gene expression and genetic information. Evidences from both analyses would make the results more reliable. Under our original strict filter conditions to identify DEGs, there were no overlapped genes with simultaneous genetic change associated with ASD. When we used

a different strategy to see which genes were dysregulated in gene expression among all the genes with genetic association with ASD, 6 genes were identified to get both supports.

The data we used in this analysis were from bulk mRNAs, which were extracted from the whole tissues and represented a mixture of various cell types. We further used single cell data to identify the cell source of the final six genes and found that they could be expressed in neurons (data not shown).

The study reveals novel key genes that are tightly associated with ASD on genetic and RNA levels. Researchers have not studied in great detail the 6 identified DEGs and their influence in the pathogenesis of ASD. However, this study examines the brain regions and the exact cell types that express these genes.

To be noted, however, is that the data used to determine the specific cell types in which the genes are expressed come from PFC and ACC brain regions, whereas the genes studied are from the frontal cortex and temporal cortex.

Functional annotations of the 6 overlapped DEGs are provided in the results section. However, this study only seeks to identify the DEGs tightly associated with ASD, and the brain regions and cell types in which these DEGs are expressed. Future studies will be directed toward utilizing the annotations to understand the functional associations between these genes and the clinical symptoms of ASD.

The 6 genes identified may prove to be vital in animal research. Advanced knowledge about these genes and their pathways, along with technological advancements in genome engineering, can allow scientists to better understand the neurological basis of ASD and develop potential treatments using molecular-engineered animal models. For example, the zebrafish is a promising model for studying ASD, as it is highly social and well-characterized genetically. The model has the capacity to accomplish the dissection of molecular pathways related to synaptogenesis and therapeutic investigations. With the discovery of the above 6 genes, scientists will be able to better understand the neurobiology of ASD from the perspective of animal models and develop pharmacological interventions [8].

Future studies directed toward understanding the functional associations between these genes and the clinical symptoms of ASD will help shed light on the molecular mechanisms behind the occurrence of ASD and help find nutritional and medical interventions for ASD patients and mitigate the adverse effects resulting from ASD. Although controversial, the discovery of these genes may also be helpful in the field of molecular engineering, as these genes can be targeted in genetic modifying techniques, such as CRISPR, to prevent the development of ASD.

Conclusion:

In this study, statistical tools were used to analyze public databases in an effort to discover novel genes associated with autism by comparing DNA and RNA from autistic population and the general population. 31 down-regulated and up-regulated genes were identified in the cerebellum, frontal cortex, and temporal cortex which have a positive connection. Of these genes, 6 genes GGNBP2, TUBGCP5, ZDHHC8, DHRS11, RABL2B, and PANX2 have been identified to be tightly associated with ASD at RNA and genetic levels, and these genes may play a role in the pathogenesis of ASD.

Acknowledgements:

The author would like to thank Professor Pingzhang Wang of Peking University for guidance in this research.

References:

1. Yu X., Qiu Z., Zhang D. Recent Research Progress in Autism Spectrum Disorder // Neurosci. Bull.– No. 33(2). 2017.– P. 125–129.

2. Centers for Disease Control and Prevention, 2019. Treatment and Intervention Services for Autism Spectrum Disorder. URL: http://www.cdc.gov/ncbddd/autism/treatment.html.
3. Voineagu I., Wang X., Johnston P., Lowe J. K., Tian Y., Horvath S., Mill J., Cantor R. M., Blencowe B. J., Geschwind D. H. Transcriptomic analysis of autistic brain reveals convergent molecular pathology//Nature.– No. 474. 2011.– P. 380–384.
4. Dennis G. Jr., Sherman B. T., Hosack D. A., Yang J., Gao W., Lane H. C., Lempicki R. A. DAVID: Database for Annotation, Visualization, and Integrated Discovery // Genome Biol.– No. 4(5). 2003.– 60 p.
5. Lonsdale J., Thomas J., Salvatore M. et al. The Genotype-Tissue Expression (GTEx) project // Nat. Genet.– No. 45(6). 2013.– P. 580–585.
6. Landrum M. J., Chitipiralla S., Brown G. R., Chen C., Gu B., Hart J., Hoffman D., Jang W., Kaur K., Liu C. et al. ClinVar: improvements to accessing data // Nucleic Acids Res.– No. 48(D1). 2020.– P. D835-D844.
7. Velmeshev D., Schirmer L., Jung D., Haeussler M., Perez Y., Mayer S., Bhaduri A., Goyal N., Rowitch D. H., Kriegstein A. R. Single-cell genomics identifies cell type-specific molecular changes in autism // Science.– No. 364(6441). 2019.– P. 685–689.
8. Pardo C. A., Meffert M. K. Animal models in autism research: The legacy of Paul H. Patterson // Exp. Neurol.– No. 299(Pt A). 2018.– P. 197–198.