Section 1. Life Sciences

https://doi.org/10.29013/ELBLS-22-1-3-13

Wenqiao Yu,
School: St. Andrew' DE
United State of America
E-mail: danny.yu331@gmail.com

DYSREGULATION OF ALTERNATIVE SPLICING IS INVOLVED IN MAJOR DEPRESSIVE DISORDER

Abstract. Major Depressive Disorder (MDD) is a mental disorder caused by brain malfunction. It's one of the most common mental disorder, effecting over 7% of the population. Most studies agree that MDD is most likely caused by many different factors, some genetic and some environmental, instead of one specific single nucleotide polymorphism (SNP). This study collected numerous SNPs related to MDD and looked at the function of these genes. We selected the most significant one, alternative splicing for further analysis. By taking a closer look at the three genes that are directly involved in alternative splicing, which are CELF4, RSRC1, and RBFOX1, we were able to understand how it effect the brain. The gene RSRC1 increased the gene expression in the brain and nerves. Data from other studies also showed that CELF4 and RBFOX1 and decrease the gene expression level in other organs. To sum those data up, the dysregulation of alternative splicing can cause the gene expression in different organs to alter in a way that increase the risk of MDD. The systematic abnormity of the alternative splicing process plays an essential role in MDD. Using this study, scientist will be able to better understand how alternative splicing can contribute to MDD and some genes associated with this process.

Keywords: Major Depressive Disorder, single nucleotide polymorphism, SNP, alternative splicing, gene expression level.

Introduction to Major Depressive Disorder

Major Depressive Disorder (MDD) is a serious medical and mental illness caused by brain malfunction. Life experiences, especially chronic pressure and physical/sexual abuse, can trigger brain malfunction, ultimately causing depression. Stress can cause epigenetic marks on the DNA and the chromatins. These marks can toggle genes in a way that trigger depression [5]. The exact neutral basis for depression is still unclear. Possible proposals for the cause include short gene alleles or the chromosome 3p25–26 [20]. However, studies have showed that a depressed brain has some important differences including grey matter abnormalities, brain shrinkage, and a more active amygdala [3].

Because it is a mental disorder, there are few to no external signs of depression. However, depression is more than sadness, also grief is one of the symptoms. Furthermore, other important symptoms include loss of interest in daily activities for a stretch

of time, feelings of worthlessness and self-loathing, and hopelessness. Not only does MDD effect one's emotion, but it can also have effect on the victim's physical appearance as depression can cause its victim to loss appetite [21].

Many studies have confirmed that depression is heritable by about 40 to 50 percent, which meant that about half the cause for MDD is genetic while the other half is due to life experiences. It has been suggested that there is no definite way to avoid MDD, as some may get it without experiencing drawback or pressure in life, whoever studies have demonstrated that the heritability of MDD caused the offspring of those with depression is more likely to get it. Someone with a family member suffering from depression has a 2- or 3-times greater risk, while those with a parent or sibling with recurring depression have a 4 to 5 times greater risk [1].

As previously stated, the exact genetic cause of MDD is unclear, however most studies agree that MDD is unlikely caused by a single gene. On the contrary, it's more likely caused by a combination of change of amino acid in single nucleotide polymorphism (SNP) that are related to MDD. Those SNPs and genes work collectively to affect the brain and the neurons which lead to MDD.

This study performed an analysis of the SNPs that are related to MDD. A total of 440 SNPs, which represent 145 different genes, are collected and analysis to determine their functions and how they contribute to the cause of depression. Then the study focused on one important function that is affected by these genes and looked closely at the genes that directly effect this function. The expression level of these is then measured and compared to other depression cases. Lastly this study looked at the relationship between the numerous genes that are believed to contribute to MDD.

Material and Method

SNP Collection

First, in order to gather SNPs that are associated with MDD, the tool GWAS Central (Genome-Wide Association Study) (https://www.gwascentral.org) was used. Using the search bar under "Phenotypes", depression was searched. Multiple studies with a high amount of "Total p-value in study" were located. All of these SNPs have a OR value that is bigger than 1, which meant that the occurrence of this SNP will increase the chance of depression. The OR value is the ration between two conditions (control and depression. All of the SNP chosen also have small p-values. P-value is the number that indicate whether the result is likely caused by chance. A p-value of less that 0.05 meant that the result is significant and reliable, unlikely due to chance. Most, if not all of the SNPs chosen for this study have a p-value that is less than 0.05, which strengthened the link between the case of depression and the individual SNP markers.

SNP Annotation

The tool wANNOVAR (https://wannovar.wglab.org/), a website that functionally annotate genetic variants, was used to annotate all 440 of the SNPs. wANNOVAR yielded information such as which chromosome the SNPs are on, which SNP is an exon, and the Official Gene Symbol for each SNP.

Enrichment Analysis

Thirdly, DAVID (https://david-d.ncifcrf.gov/) was used for functional enrichment analysis. It was used to better understand the major function of these genes. It also serves to provide the Gene Ontology terms to help understand the biological meaning of each gene. This website also provided links that are helpful in learning about each function effected by the genes.

Gene Information

Then, SOURCE search (https://source-search.princeton.edu/) was used to find the genes that directly affect the alternative splicing process. By enter all of the Official Genes Symbols into the website and then searching for "splic".

Then NCBI (www.ncbi.nlm.nih.gov) and DAVID was used again to get a deeper understanding for each of these genes. NCBI is an extreme helpful source as it has numerous sources for different pur-

poses. It is a website that was utilized a lot later in this study. Then the document that is annotated by wANNOVAR was used to identify the specific SNPs in each gene.

To better understand the structure of each gene and the position of the SNPs on each gene, the UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgGateway) was utilized. This website provides a tool that display the position of each gene on the chromosome and which exon or intron each SNP sit on.

Expression Level Analysis

After that, specific SNP was studied for their potential effects on gene expression using GTex Portal (https://gtexportal.org/home/). This website showed whether a specific SNP will cause the expression level in certain areas to increase or decrease. Because the illness studied is depression, this study focused on the expression level change in the brain and nerves. All of the variation in expression level has a low p-value to ensure that they will have potential effect on the cause of depression.

Then, NCBI was utilized again to search for other studies that contain the expression level of patients with MDD and the control. By analyzing with GEO2R, the average expression level was collected. In addition to calculating the average, a student's t-test was performed on each comparison set to get the p-value. Several comparison cases have a low p-value, ensuring that the difference between the control and MDD cases was not caused by chance.

Network Analysis

Lastly, STRING (https://string-db.org/) was used in attempt to find connections between all of the genes that contribute to depression. This tool gave a clear diagram displaying the relation of each gene related to depression on a web.

Results

Data and Statistics

In the study, first the data from GWAS Central database was used to identify SNPs that are associated with the occurrence of MDD. We found 9 different studies with a high amount of "total p-value in study". These 9 studies contain a total of 440 SNPs. All of these SNPs have an OR value that is bigger than one, which means that they all have a positive relationship with the occurrence of MDD. Then the list of SNPs was into a vcf file and submitted that to wANNOVAR. Out of the 440 SNPs collect, only one is exonic. This SNP belong to the gene DENND1B (DENN domain containing 1B). This gene functions as guanine nucleotide exchange factors (GEFs) for the early endosomal small GTPase RAB35 and bind to clathrin as well as clathrin adaptor protein-2.

These SNPs account for a total of 140 different genes. Some genes show up multiple times and have multiple SNPs. Below is a table of all the genes that show up repeatedly for 5 or more times.

Table 1.– MDD related genes that repeatedly show up and their SNPs

GENE	SNPs
DCC	rs149735550; rs62097899; rs11663393; rs4277413; rs7505145; rs8084351; rs7227069; rs4632195; rs12968428; rs8099160; rs62100776; rs1431181; rs8089865
RBFOX1	rs8063603; rs7193263; rs7198928; rs7198928; rs3785234; rs2191130; rs2191130;
SORCS3	rs61867293; rs1021363; rs2496022; rs1961639; rs7074335
CNTN5	rs1690818; rs1690816; rs2458167; rs586533
GRM5	rs1150313; rs7932640; rs7126679; rs10741299; rs10830220
S0X5	rs4074723; rs78337797; rs17487383
TCF4	rs1262465; rs1261070; rs12967143; rs12958048; rs1452788; rs4801157

Showing up repeatedly meant that these genes are included in many studies. Being found in many studies proved their significance and connection to MDD.

Alternative Splicing is One of the Most Important Function

Then we investigated the main function effected by these genes with DAVID. Using DAVID's functional annotation tool, we were able to get the enrichment annotation for these genes. The main effects of these genes include alternative splicing and changes to phosphoprotein, dendrite, and cell junction.

Table 2. – Top 10 Terms from DAVID's Functional Annotation Chart and their corresponding statistic. Gene Count is the number of gene involved in this term and% of genes involved is the ratio between the gene count and the total amount of genes entered

Terms	Gene Count	% of genes involved	P-value
splice variant	78	55.7	1.30E–08
Alternative splicing	88	62.9	9.80E–06
Membrane	68	48.6	3.20E–05
Phosphoprotein	70	50.0	2.30E–04
Dendrite	10	7.1	3.10E–04
Learning	5	3.6	4.90E–04
Cell Junction	13	9.3	7.30E–04
Synapse	9	6.4	1.40E–03
dendrite morphogenesis	4	2.9	1.70E–03
Guanine-nucleotide releasing factor	6	4.3	2.00E–03

Out of every term, alternative splicing involves the most genes out of the list in this study, while it also has the lowest p-value. Alternative Splicing is a process during gene expression that allows a single gene to code for multiple proteins, thus allowing for gene diversity. The statistics in (table 2) meant that a lot of genes go through the alternative splicing process and is significantly affected by it. Alternative splicing is significantly enriched by these genes and its likely to effect MDD. Therefore, this study is going to focus on genes associated from alternative splicing from this point on.

Three Genes are Directly Involved in Alternative Splicing

In order to find the genes that are directly involved in the process of alternative splicing, SOURCE search developed by Princeton University was used. By entering the official gene symbols, we receive the function and classification of each gene. We found that three genes are directly involved in the process of alternative splicing. Those genes are CELF4 (CUGBP, Elav-like family member 4), RSRC1(arginine and serine rich coiled-coil 1), and RBFOX1(RNA binding protein, fox-1 homolog 1). In addition to being directly involved in alternative splicing, these 3 genes also go through the process themselves.

In order to learn more about each gene, the database on NCBI was utilized. Meanwhile the links on DAVID are also helpful in enhancing my understanding of these genes.

CELF4, also known as BRUNOL4, is a member of the CELF/BRUNOL protein family. This protein family is mainly responsible for regulating pre-mRNA alternative splicing. It can also take part in in mRNA editing and translation. More specifically, CELF4 regulates translation and local abundance of numerous mRNAs, including those associated with regulation of synaptic function. This gene can cause a biased expression in brain and adrenal [17].

RSRC1 encodes for a member of the serine and arginine-rich related protein family. It's involved in both the constitutive and alternative mRNA splicing process, which can lead to multiple transcript variants encoding different isoforms. This gene might be associated with schizophrenia and has been implicated in various neurological disorders [18].

RBFOX1 belong to the FOX-1 family of RNA-binding proteins. It's an RNA-binding protein that regulates alternative splicing events through binding with 5'-UGCAUGU-3' elements. RBFOX1 isoforms specifically activate splicing of neuronally regulated exons, which requires UGCAUG enhancer elements. This gene may cause biased expression in the brain or heart [19].

Table 3. – Information on the 3 genes that are directly involved in alternative splicing. Corresponding SNPs are listed in the last column

Symbol	Name	GeneID	Cytoband	SP Local	UniProt	SNP
CELF4	CUGBP, Elav-like family member 4	56853	18q12	nucleus	Q9BZC1	rs4799936; rs1557341; rs12967855; rs11082011; rs11665070
RSRC1	arginine/serine-rich coiled-coil 1	51319	3q25.32	nucleus	Q96IZ7	rs6441175; rs1095626; rs7430565
RBFOX1	RNA binding protein, fox-1 homolog (C. elegans) 1	54715	16p13.3	nucleus	Q9NWB1 B7Z1U7 Q59HD3	rs8063603; rs7193263; rs7198928; rs7198928; rs3785234; rs2191130; rs2191130

Each of these genes include multiple SNPs from the initial data of all 440 SNPs. Each of them contains at least 3 or more SNPs, meaning that multiple studies have also identified these genes.

All of the MDD related SNPs are Located in Non-exon Regions

To better understand the structure of the genes where these SNPs sit on, and each individual SNPs specific position on the gene, the UCSC Genome Browser was used as it clearly displayed the position of each gene. All of the SNPs are intronic, however they are still important as they can have effect on gene expressions and importantly impact the gene. The figure below indicates the specific location of each SNP on their gene.

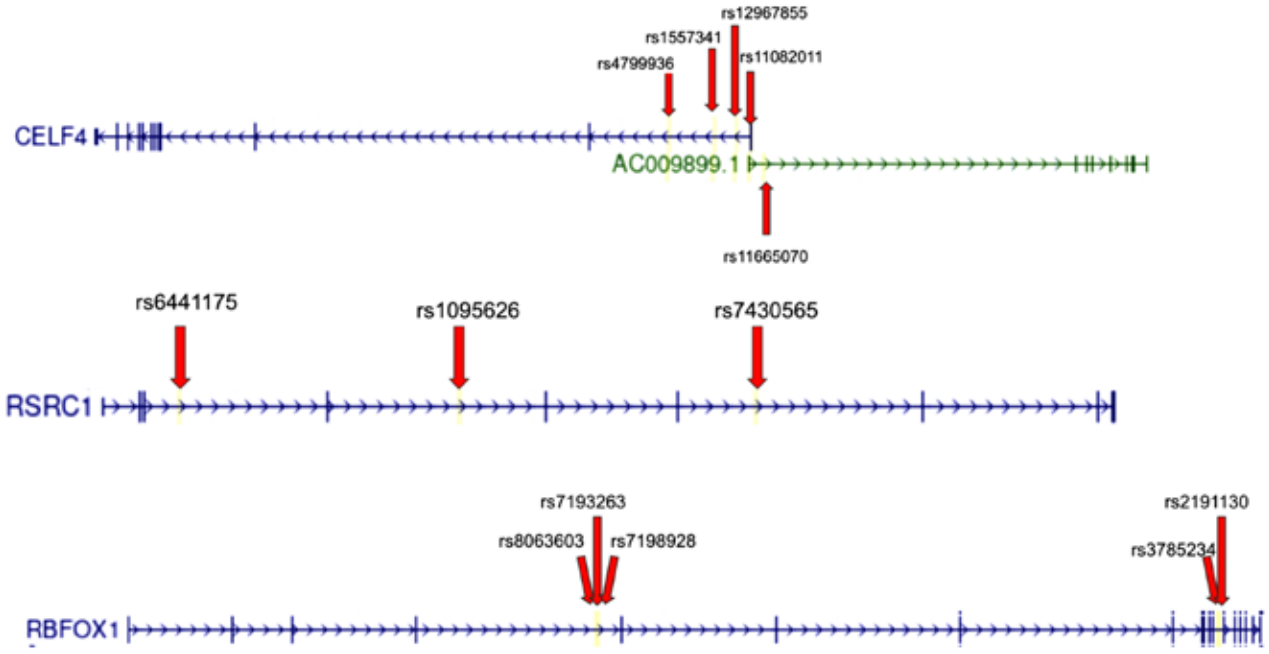

Figure 1. CELF4, RSRC1, and RBFOX1 gene structure and SNP position

On the horizontal blue line that symbolize a gene, each blue vertical dash represents one exon while the part between each exon are introns. Because all of the red arrows, which points to the position of a SNP in the gene, are located between the vertical dashes, they are all intronic. The arrows on each gene represent the direction of the gene and whether it's up or down stream. If the arrow is pointing to the left, then the gene is downstream. If the arrow is to the right, then the gene is upstream.

Alternative Splicing Associated Genes Increase Expression Level in the Brain and Nerves

These SNPS and their variations can change the expression level of genes. To figure out the potential expression level change, we used GTex Portal to search for each SNP. The table below indicate the potential change in expression level caused by these SNPs. Although all of the SNPs listed in the above table are search, only two yielded reliable result (rs1095626 and rs7430565). Both SNP is a part of the gene RSRC1. It doesn't mean the other SNPs have no effect; however, they need to be further investigated. The first two figures on the top come from the SNP rs1095626. The risk allele for this SNP is C, which is displayed to the right of each figure. Therefore, the gene expression goes from control on the left to SNP case to the right. In the bottom two figure, which come from the SNP rs7430565, the risk allele is G. Because G is to the left of the two figures in the bottom, the figure goes from controlled to the right to SNP to the left. In all four variations, the gene expression amount has been increased. It's also worth noting that all of these variations have a low p-value, making them significant and not by chance.

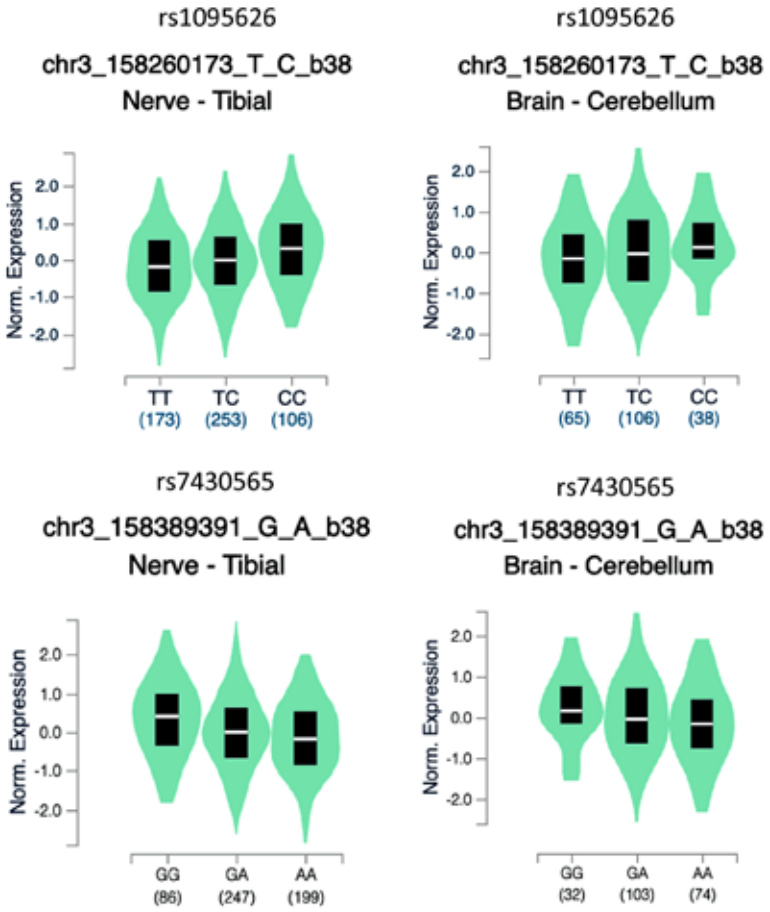

Figure 2. MDD associated genetic variations' effect on gene expression. All graphs come from either variation in the brain or nerves. All variations come from the gene RSRC1

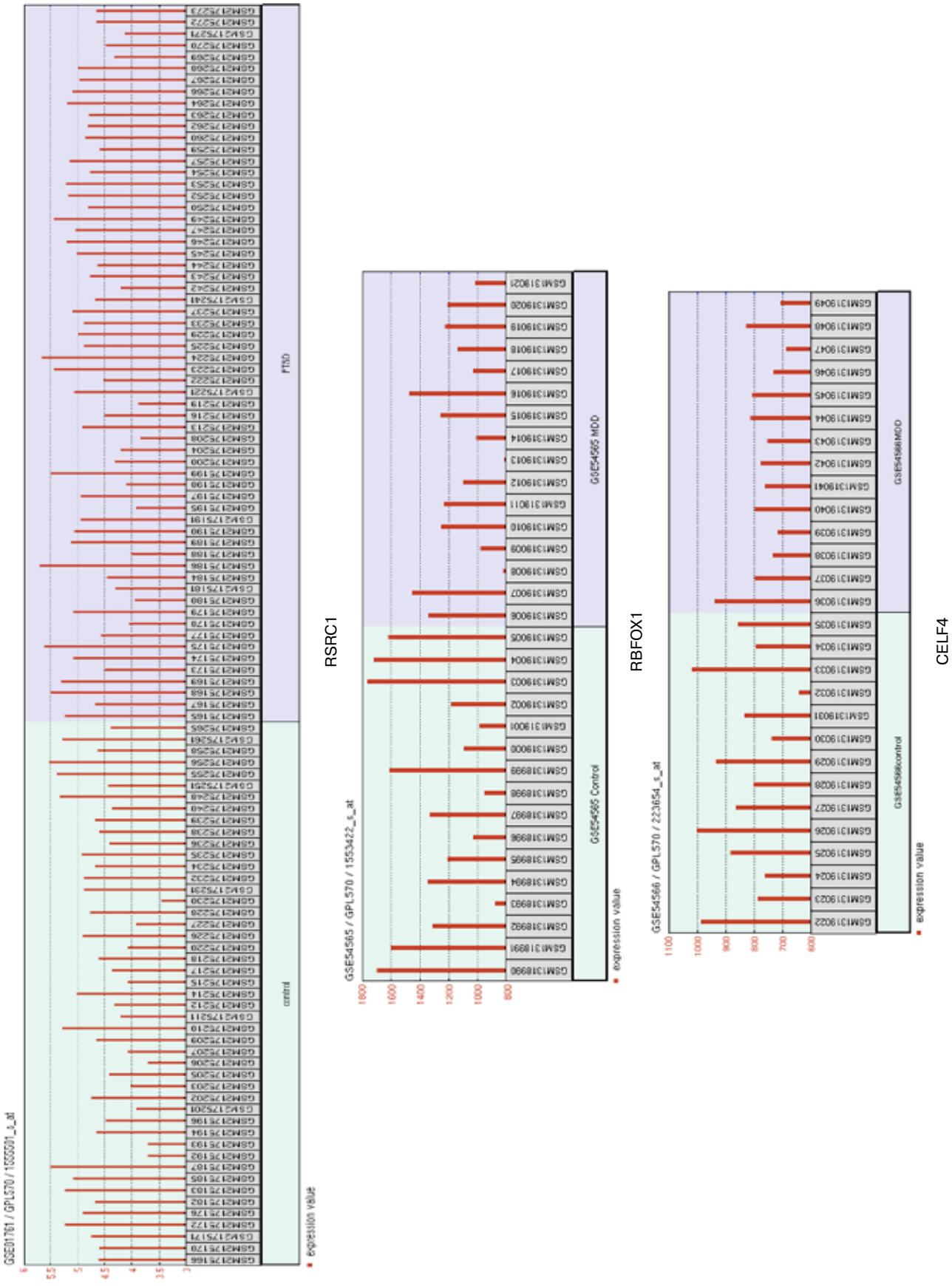

Figure 3. Diagrams of most change in gene expression level for each gene

As the graph has shown, the gene expression level has been increased, causing its original function to be enhanced.

CELF4, RSRC1, and RBFOX1 are Dys-regulated in Gene Expression in MDD

In order to understand how the gene expression level are different in MDD patients and the control, NCBI was used again to search for different studies. There are a total of 8 studies that provide a comparison between the expression level of controlled cases and MDD cases. It's worth noting that not all of these studies focused on the same aspect as this study, for example one of them provided data for blood sample, however they still provided information on whether gene expression level change will contribute to depression. After sorting the data from each study into controlled versus MDD, we searched the IDs for each gene. Each of these three genes can have multiple different Probe Set IDs, and each can have different expression level.

The GEO2R analysis function in NCBI provided a diagram that compares the expression level of both MDD cases and controlled cases. Out of all the studies, there were 13 cases where the expression level is visibly different. Then took the sample data and used the AVERAGE and TTEST function to find the average and p-value for each set of comparison. There are a total of 4 cases where the MDD cases' expression level is slightly higher, while there are 9 cases where controlled have a higher expression level. In all of the 13 cases, 8 cases have a low enough p-value to be considered significant. In those 8 cases, 7 cases showed that the controlled case has a higher expression level. In all of the diagrams, the left (green) represents the expression level from controlled cases, while the right (purple) represents the expression level from MDD cases.

Table 4. – Gene expression level comparison from different studies. The GEO accession is the code for each study that can allow for easy navigation in NCBI. The Expression Level Change shows whether the gene expression level increased or decreased from controlled to MDD

GENE	GEO accession	Probe Set ID	Expression Level Change	p-value
RSRC1	GSE81761	1555501_s_at	increase	0.036166143
RBFOX1	GSE81761	1553422_s_at	decrease	0.033248424
RBFOX1	GSE81761	235070_at	decrease	0.056357117
RBFOX1	GSE54565	1553422_s_at	decrease	0.046421725
CELF4	GSE81761	231220_at	decrease	0.051945508
CELF4	GSE81761	232719_at	decrease	0.026451222
CELF4	GSE54568	223653_x_at	decrease	0.061491506
CELF4	GSE54566	223654_s_at	decrease	0.034136362

As shown in the table above, most studies showed a decrease in gene expression for the MDD cases while still keeping the p-value relatively low. It's worth noting that although the trend shows a decrease of expression level, the genes expression level for RSRC1 (first row in table 4) increased. This stayed consistent with the data in (figure 2), where both SNPs showed an increase in gene expression in the brain and nerves.

In addition to the table, below are three graphs, one for each gene, that represent the most change in gene expression.

In the above figures, the letter to the left of each diagram correspond to the label in (table 7). For each figure, each red bar represents the expression level for one case. The controlled cases are always to the left in the green section while the MDD cases are always to the right in the purple section.

In conclusion to the data gather, these genes can cause systematic abnormality in the alternative splicing process. This can lead to an altar of expression level in many places and organs. As (table 5) states, the gene RSRC1 increased the expression level in the brain and nerves, enhancing the function of this gene. In other organs, such as blood or different part of the brain, the expression level was decreased, as stated in (table 10).

Network Analysis Reveals Interaction Among MDD-associated Genes

In order to further understand the relationship between all of the genes involved, STRING as used to produce a network of all the genes identified at the beginning of the study. Many of the genes that might contribute to MDD have some level of connection between them, while there are also a decent number of genes isolated from the overall network.

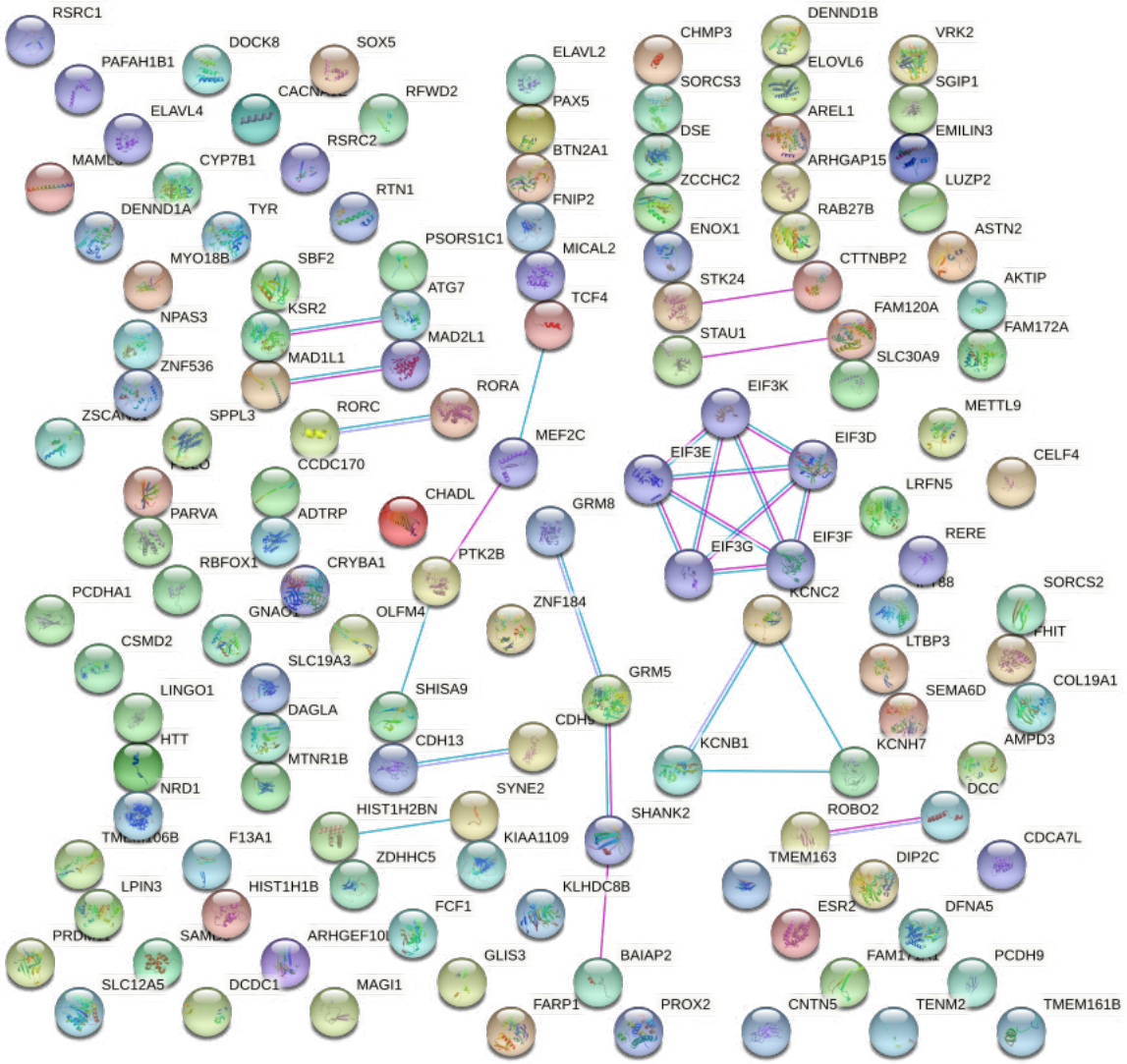

Figure 4. MDD related Gene Connection Network

In the above diagram, blue lines connecting two genes represent known interactions from curated databases and pink lines between two genes represent experimentally determined interactions.

A large portion of these genes don't have known interactions with each other, including the three genes this study primarily focused on. Most of the

genes are not in one network, rather, effecting a wide variety of genes and organs.

There are a few clusters of genes with known interaction with each other. The most notable interactions happened between the EIFs family. These genes include EIF3E, EIF3K, EIF3D, EIF3F, and EIF3G. All of these 5 genes have both known interactions from databases and experiments. Some other examples are the connection between GRM8, ZNF184, GRM5, SHANK2, and BAIAP2, and the connection between KCNC2, KCNB1, and KCH7.

Discussion

Alternative splicing happens with a unique pattern in the human brain and controls multiple aspects of early neuronal development. The diversity of these splicing patterns dictate important regulatory decisions in many stages of neuronal development. In mature neurons, synaptic remodeling and strengthening are all regulated by alternative splicing and the relative expression levels of Amplicon sequence variant. The splicing events that control neuronal activity are regulated by incoming stimuli, for example chronic excitatory depolarization [14].

The precursor-mRNA splicing reaction is essential in the regulation of gene expressions. Most genes produce multiple mRNA isoforms through such process to produce proteins with different structures and functions. In addition, the nervous system makes use of splicing regulation to produce specialized protein isoforms that will importantly affect many aspects of neuronal development [15].

The alternative splicing process is regulated by specialized pre-mRNA binding proteins that alter spliceosome assembly. Some of these regulators show tissue-specific expression, whereas others show more universal expression level, however they all regulate large overlapping programs of neuronal alternative splicing events. This support the why the gene expression level increased in the brain and nerves, as shown in figure 5, however showed a decreasing trend in other studies that are surveyed later [15].

The serotonin-1A (5-HT1A) receptor is crucially in regulating serotonergic activity and is implicated to have emotional effect. When the human HTR1A RNA is alternately spliced, this splicing removes a microRNA site increase HTR1A expression. This splicing can varies in different brain regions but is generally reduced in MDD. Un-spliced HTR1A was also shows stronger expression in the hippocampus and midbrain in comparison to the prefrontal cortex. This could explain why the systematic abnormality in alternative splicing can decrease the expression level in many of the studies that we surveyed [6].

References:

1. Douglas F. Levinson M. D., Walter E. Nichols M. D. Major Depression and Genetics.n.d.
2. Felix Torres M. D. What Is Depression. October, 2020.
3. Alyssa. Depressed Brain vs. Normal Brain. 24 Mar. 2021.
4. Nancy Schimelpfening. How Genetics Can Play a Role in Depression. December 03, 2020.
5. Eric J. Nestler. The Genetics of Depression: What is Known, What is Next. November 2017. Video available from: URL: https://www.hopefordepression.org/research-funding-and-impact/video-genetics-depression
6. Le François B., Zhang L., Mahajan G. J., Stockmeier C. A., Friedman E., Albert P. R. A novel alternative splicing mechanism that enhances human 5-HT1A receptor RNA stability is altered in major depression. J Neurosci. 2018; 38: 8200–10. 10.1523/JNEUROSCI.0902-18.2018
7. Chang L. C., Jamain S., Lin C. W., Rujescu D. et al. A conserved BDNF, glutamate- and GABA enriched gene module related to human depression identified by coexpression meta-analysis and DNA variant genome-wide association studies. PLoS One 2014; 9(3): e90980. PMID: 24608543

8. Rusch H. L., Robinson J., Yun S., Osier N. D. et al. Gene expression differences in PTSD are uniquely related to the intrusion symptom cluster: A transcriptome-wide analysis in military service members. Brain Behav Immun 2019. Aug; 80: 904–908. PMID: 31039430
9. Leday G. G.R., Vértes P. E., Richardson S., Greene J. R. et al. Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder. Biol Psychiatry, 2018. Jan 1; 83(1): 70–80. PMID: 28688579
10. Sibille E. Molecular Evidence for a Dimensional Basis of Depression. Jan 01, 2016.
11. Chang L. C., Jamain S., Lin C. W., Rujescu D. et al. A conserved BDNF, glutamate- and GABA-enriched gene module related to human depression identified by coexpression meta-analysis and DNA variant genome-wide association studies. PLoS One 2014; 9(3): e90980. PMID: 24608543
12. Lanz T. A., Reinhart V., Sheehan M. J., Rizzo S. J. S. et al. Postmortem transcriptional profiling reveals widespread increase in inflammation in schizophrenia: a comparison of prefrontal cortex, striatum, and hippocampus among matched tetrads of controls with subjects diagnosed with schizophrenia, bipolar or major depressive disorder. Transl Psychiatry, 2019. May 23; 9(1): 151. PMID: 31123247
13. Kong S. W., Collins C. D., Shimizu-Motohashi Y., Holm I. A. et al. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One 2012; 7(12): e49475. PMID: 23227143
14. Glatt S. J., Cohen O. S., Faraone S. V., Tsuang M. T. Dysfunctional gene splicing as a potential contributor to neuropsychiatric disorders. Am. J. Med. Genet. Part B: Neuropsychiatric Genet. 2011; 156:382–392.
15. Vuong C. K., Black D. L., Zheng S. The neurogenetics of alternative splicing. Nat. Rev. Neurosci. 2016; 17: 265–281. Doi: 10.1038/nrn.2016.27.
16. Pandya M., Altinay M., Malone D. A., Anand A. Where in the Brain Is Depression? Curr. Psychiatry Rep. 2012; 14: 634–642. Doi: 10.1007/s11920–012–0322–7.
17. [Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2021, Aug, 31]. Available from: URL: https://www.ncbi.nlm.nih.gov/gene/56853
18. Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2021, Aug, 31]. Available from: URL: https://www.ncbi.nlm.nih.gov/gene/51319
19. Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2021, Aug, 31]. Available from: URL: https://www.ncbi.nlm.nih.gov/gene/54715
20. Stephanie Faris. Is Depression Genetic? April 5, 2021.
21. The American College of Obstetricians and Gynecologists. Depression. June 2005. Available from: URL: https://dmh.lacounty.gov/our-services/employment-education/education/depression