Upregulation of the Ribosomal Pathway as a Potential Blood-Based Genetic Biomarker for Comorbid Major Depressive Disorder (MDD) and PTSD

Karthik Ravi[1] and Inhan Lee[2]

[1] International Academy, Bloomfield, Michigan

[2] miRcore, Ann Arbor, Michigan

Summary

Major Depressive Disorder (MDD), and Post-Traumatic Stress Disorder (PTSD) are two of the fastest growing comorbid diseases in the world. A patient without any genetic predisposition to depression can live a healthy lifestyle but develop depression as a result of their environment, such as the occurrence of a traumatic event. This study addresses the question of whether there are any biomarkers in blood that could alert physicians of a patient's susceptibility to comorbid MDD and PTSD. Since the pathophysiology of comorbid MDD is complex, it has been hypothesized that there could be several gene groups acting as biomarkers. Based on existing literature, the genes *DICER1*, *FK506*, and *MRPS23* are thought to be relevant biomarkers. Using GEO2R, provided by NCBI, we conducted a differential gene expression analysis. We used patient dataset GSE67663, which contained 184 blood samples, including 72 control samples and 112 test samples. The top 253 genes that had p-values from 1.6×10^{-6} to 3.49×10^{-3} were entered into the STRING database in order to identify common protein interactions. In accordance with the KEGG Ribosomal Pathway, two main gene groups, including those involved in mitochondrial ribosomal translation, were further analyzed for common gene ontology and biological processes. In the end, we found that the Ribosomal Pathway (ID: 03010) contained 29 upregulated genes and is a potential blood-based biomarker. This study has applications to the field of personalized medicine, in which tailored treatments based on biomarkers and one's genome are utilized to identify people at risk for disease.

Introduction

Post-traumatic stress disorder (PTSD) is a psychiatric disorder that develops as a result of a traumatic event.

Among victims of severe trauma, nearly 80% experience PTSD in some manner (1). PTSD currently has no definitive cure due to the countless offsets that it has as well as the complexity of the disease. Patients who experience severe trauma often develop other diseases along with PTSD. The number of patients impacted by comorbid Major Depressive Disorder (MDD) and PTSD is on the rise. According to Breslau and colleagues, approximately half of all people with PTSD also suffer from MDD (2). Currently, the most abundant form of treatment for patients with only MDD is antidepressant medication such as selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants (TCAs). However, there exist far fewer treatments for comorbid MDD and PTSD. Treatment for these comorbid diseases is usually approached in three different ways: integrated treatment (treatment of both disorders simultaneously); sequential treatment (treatment of both disorders one at a time); or single diagnosis treatment (treatment of only one disorder) (3). Regardless of the treatment plan, these treatments are often deemed inefficient because they require a trial and error regiment in order to establish which medication works best for each patient and for each disease. In addition, diagnoses for these diseases are commonly given after the patient has been suffering for a long period of time. The current costly and time-consuming process of diagnosing and treating comorbid MDD and PTSD is the leading motivation for this study.

In order to advance the current comorbid MDD diagnosis and treatment process, scientists need a way to diagnose the disease earlier. We propose that blood-based genetic biomarkers can be utilized to identify at-risk patients and to clarify their diagnoses. If gene expression changes predict comorbid MDD and PTSD, then the potential diagnosis may be made before the patient starts suffering from the disease. For patients with comorbid MDD and PTSD, the treatment process can become more focused on the expression of biomarker genes, increasing the effectiveness of the current diagnostic and treatment process. If this goal is accomplished, the pathology of comorbid MDD and PTSD may no longer have such negative and long-term effects on its victims.

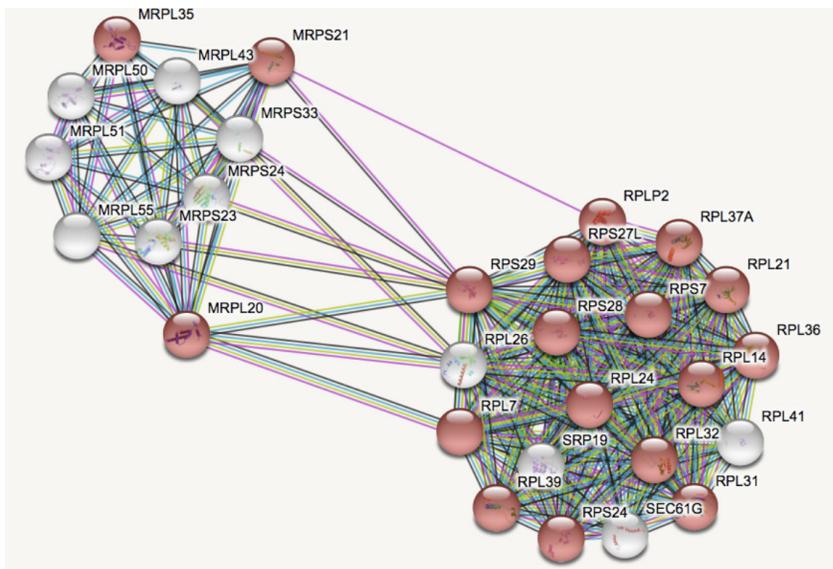The primary motivation for this investigation was to

determine whether there is an accessible biomarker for comorbid MDD and PTSD, and if so, what it is. In particular, we attempted to identify and analyze biomarkers in blood that could be used in the future as alerting precursors to comorbid MDD and PTSD. Prior studies have hypothesized that since the pathophysiology of comorbid MDD and PTSD is complex, there would be several different genes acting together as biomarkers, which would be either significantly upregulated or downregulated. This reasoning is partly based on the many different genes found to be linked with comorbid MDD and PTSD across different studies that analyzed expression profiles. Some of these genes include *DICER1*, *FK506*, *MRPS23*, and *ZXDC* (4,5). Since many different studies reached different conclusions in regards to which genes are linked to the disease, we similarly hypothesized that there would be several genes involved.

Utilizing data from the GSE67663 dataset found in the NCBI database, we performed gene expression analysis, protein interaction analysis, pathway analysis, and gene ontology (GO) analysis in order to identify potential biomarkers. After initial analysis, the *RPL* and *MRP* gene groups were found to be significantly more upregulated compared to others. With further analysis, we found that these genes and their proteins are strongly correlated with each other, forming a majority of the entire ribosomal pathway. This led to the conclusion that the upregulation of the ribosomal pathway is a potential biomarker for comorbid MDD and PTSD.

Results
Differential Gene Expression Analysis
The dataset GSE67663 from the NCBI database

Figure 1: Gene interaction map produced through the STRING database of the 128 downregulated genes. Each circle represents a gene, and connections represent a correlation between genes. Light blue and pink connections represent known interactions established by curated databases, or by experiments, respectively. Light green, red, and dark blue connections represent predicted interactions established by gene neighborhood, gene fusions, and gene co-occurrence. Yellow, black, and purple connections represent interactions established by text mining, co-expression, and protein homology, respectively. The box represents the potential genes of interest.

was used for this study (6). This dataset contains gene expression profiles from blood for patients with comorbid MDD and PTSD as well as control patients. Of the 13,049 genes analyzed in the original study done at Emory University (GSE67663), only the top 253 genes from GEO2R were used for this study because we determined that p-values greater or equal to 3.5×10^{-3} were not statistically significant enough to be identified as relevant biomarkers. Varying from 1.6×10^{-6} to 3.49×10^{-3} the top 253 genes' *p*-values were deemed significant for further analysis. Once the statistical significance of the genes was validated, the upregulation and downregulation of the genes were reported by log fold-change (logFC). The logFC value indicated whether or not the particular gene was upregulated or downregulated and how significantly it was differentially expressed. Of the 253 genes, 125 genes were upregulated with logFC values varying from 0.053 to 0.308. 128 genes were downregulated, with logFC values in the range of -0.313 to -0.0538. Since the logFC values were relatively low, it was understood that one or two genes alone were not significant enough to serve as biomarkers.

Gene-to-Gene Interactions

The STRING database, a collection of known and predicted protein interactions, was used to study relationships between the protein products of genes we identified (7). All 128 downregulated genes were analyzed in order to identify possible interactions. When entered into STRING, no significant connections or clusters were present (**Figure 1**). However, near the center around the *Toll Like Receptor* (*TLR*) genes, there are connections branching out from *TLR2*. These are valid and proven connections since a majority of the connections are made with pink lines, which means that

the connections were experimentally determined. Thus, the only downregulated gene group was determined to be the *TLR* group, but even that group did not have many interactions (**Figure 1**). Therefore, none of the downregulated genes were deemed significant enough for further analysis.

When the 125 upregulated genes were put into STRING, there was a larger quantity of results due to the sheer size and number of connections present. Near the outside edges of the STRING database figure, there were not many clusters; however, two main clusters were prominent (**Figure 2**). These were identified as the ribosomal proteins of the *large subunit of the ribosome* (*RPL*) gene group and the *mitochondrial ribosomal protein* (*MRP*) gene group. The RPL gene cluster had 19 genes and the *MRP* gene group had nine genes (**Figure 2**). The other genes outside these clusters were not considered important for this study because the amount of connections was subordinate. In the end, these two gene clusters were chosen as the genes of interest.

Enriched Gene Ontology and Biological Processes

Once the genes of focus were identified, enrichment was the next form of analysis. According to STRING, the top five most significant biological processes, arranged by *p*-value, were translational termination, translational elongation, translational initiation, signal recognition particle (SRP)-dependent cotranslational protein targeting to membrane, and translation (7). These processes contained 26, 26, 26, 18, and 21 genes, respectively (**Figure 3**). From this, it was apparent that the *RPL* and *MRPL* gene groups have some connection to translation, which is evident in ribosomes and mitochondria. This explained why many of the biological processes involved mitochondria and ribosomes, such

Table 1: Enriched Biological Processes table depicting the top 15 enriched biological processes identified by the STRING database. The left column is the pathway ID that represents which GO term the process falls under. The third column represents the number of genes out of the 29 that are known to have the same function or related function. The fourth column is the *p*-value.

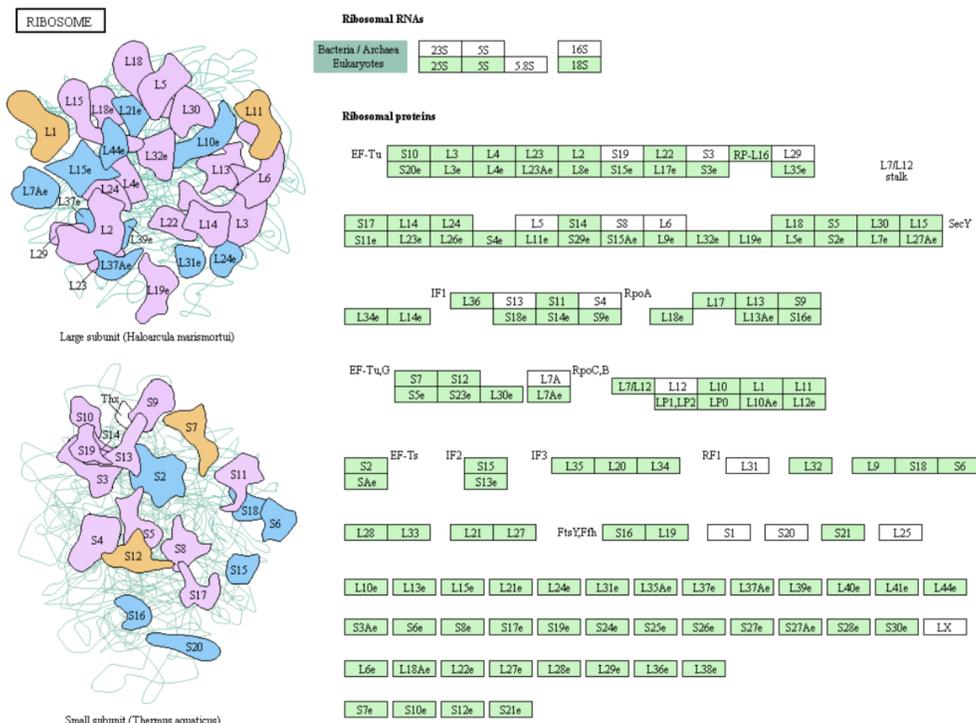| Biological Processes (GO)			
Pathway ID	Pathway Description	Number of genes	P-value
GO:0006415	Translational termination	26	5.43e-48
GO:0006414	Translational elongation	26	3.37e-47
GO:0006413	Translational initiation	26	1.61e-45
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	18	1.08e-31
GO:0006412	Translation	21	3.31e-29
GO:0019083	Viral transcription	16	1.04e-26
GO:0000184	Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	16	1.8e-26
GO:0019080	Viral gene expression	16	3.52e-26
GO:0019058	Viral life cycle	16	1.85e-21
GO:0006605	Protein targeting	17	8.98e-20
GO:0070124	Mitochondrial translational initiation	10	4.69e-15
GO:0070125	Mitochondrial translational elongation	10	4.69e-15
GO:0070126	Mitochondrial translational termination	10	4.69e-15
GO:0042273	Ribosomal large subunit biogenesis	3	0.00034
GO:0042254	Ribosome biogenesis	5	0.000739

as mitochondrial translational initiation and ribosome biogenesis.

After the GO analysis, we decided that these original two distinct clusters of genes were similar in that the genes in both clusters had connections to ribosomal translation, whether it be mitochondrial ribosomes or ribosomes dispersed in the cytoplasm. Hence, we began to view these clusters as a single unit of genes with similar functions. In order to analyze the signaling pathways between these genes, we used Kyoto Encyclopedia of Genes and Genomes (KEGG). According to the STRING database, the enriched KEGG pathway for these 29 genes in the *RPL* and *MRP* groups was the Ribosome Pathway (8). Analysis of this pathway shows how the genes affect the ribosomal proteins in the large and small subunits of the ribosome. This pathway portrayed the genes in charge of coding the proteins that carry out many of the functions of the ribosome. According to STRING, 18 of these 29 genes were confirmed to be a part of the pathway (**Figure 4**). However, STRING only counted genes that have been experimentally determined to be a part of the pathway. In fact, under the elongation factors in the pathway, all of the genes that were not labeled under the ribosome pathway in STRING were identified. This means that they are also a part of the ribosomal pathway, and that these genes also code for proteins found in accordance with the subunits of the ribosome. Therefore, 29 of the 48 genes in the ribosomal pathway are upregulated in the comorbid MDD and PTSD patients in this study. That is 60.412% of the entire ribosomal pathway, implying that these genes studied constitute a major component.

Discussion

The significance of the analysis done shows that the majority of the ribosomal pathway was upregulated in the

Genes in final biomarker: MRPL35, MRPL50, MRPL43, MRPS21, MRPL51, MRPS33, MRPS24, MRPL55, MRPS23, MRPL20, RPLP2, RPL37A, RPS27L, RPL21, RPS29, RPS7, RPS28, RPL26, RPL24, RPL14, RPL36, RPL7, RPL39, RPS24, RPL31, RPL32, RPL41, SRP19, SEC61G

Figure 4: Depiction of Ribosomal Pathway in KEGG Database. The left images represent the large and small subunits of the ribosome. The particular proteins that are coded for are drawn. To the right of the figure are the proteins themselves with their according elongation factors. The specific genes that code for these proteins are identified by the ending name of the protein. For example, the S10 protein is coded by the MRPS10 gene and the L3 protein is coded by the RPL3 gene. The table shows all included genes in the final biomarker.

test group. However, the reason as to why this pathway may be associated with MDD and PTSD is unclear in the literature. The function of the ribosome pathway is mainly to facilitate the assembly of the ribosome and the translation of mRNA. Though the ribosomal pathway and ribosome biogenesis have been connected with cancers by activating the tumor suppressor *p53* pathway in response to ribosomal stress, there is significantly less literature on why they may be connected to the brain and comorbid MDD (12). We speculate that they are connected because it has previously been established that ribosomal dysfunction affects serotonergic activity in the brain (6). Dysfunction in serotonin activity is known to be connected with many psychiatric disorders, including comorbid MDD and PTSD (7). This finding leads us to postulate that the upregulation of genes in the ribosome pathway is connected to comorbid MDD because of its relation to serotonergic activity in the brain. It seems

plausible that the upregulation of the ribosome pathway can serve as a biomarker for comorbid MDD since the overexpression of ribosomal genes leads to dysfunction in serotonergic activity (7).

This specific set of biomarkers has also come up in other studies (10). Using mice in a depressed state, researchers have measured gene expression in both the hypothalamus and hippocampus (10). After differential gene expression analysis, they found ribosomal genes, including the *RPS*, *RPL*, *MPRS*, and *MRP* genes, to be upregulated in the test group (10). This supports the notion that ribosomal dysfunction may indeed have a correlation and effect on the pathology of comorbid MDD. Although the biological and social nature of mice stand in stark contrast to that of humans, mice can serve as informative models about human disease. However, it is important to note that this study was only based on induced stress or depression. PTSD was not a part of this
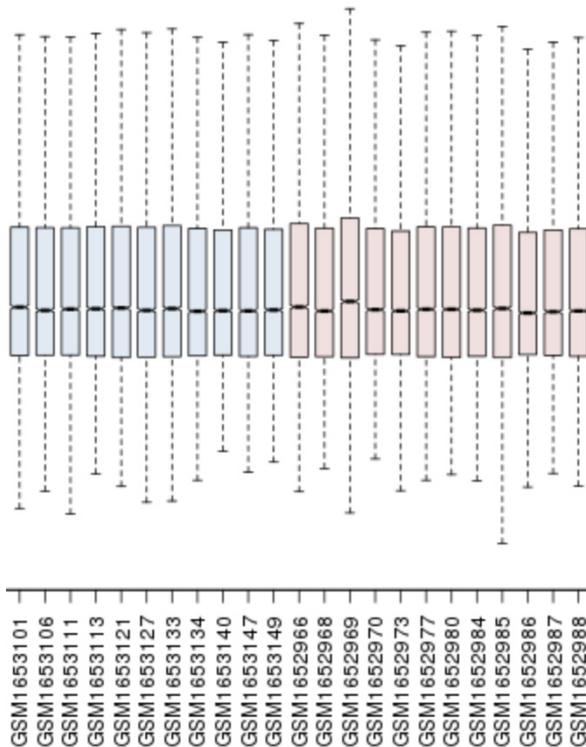
samples and RNA sequencing are much more useful because the gene expression can be measured from a standard blood test. However, there are also limitations attached to using blood samples for measuring gene expression. The main limitation is that blood expression profiles do not show what is really happening in various parts of the brain. For example, this pathway can be significantly upregulated in the hippocampus and significantly downregulated in the hypothalamus, and the blood sample would only show that it is upregulated. Hence, using brain tissue provides more context as to where this over-expression is occuring. In the future, this study should be repeated using human brain tissue samples for expression profiles in order to know which parts of the brain are undergoing this upregulation of the ribosomal pathway.

Another potential limitation that exists in this study is the amount of data, as this study focuses on only one dataset. Usually, multiple datasets are used for scientific investigation, but not many publicly available datasets of comorbid MDD and PTSD gene expression profiles exist. The analysis of only one dataset limits the scope of analysis and the conclusions that can be made. If this study is cross-referenced with other datasets as well, the results will become more reliable. Another experiment could be held without the comorbidity of MDD and PTSD in the experimental group. The results may not be totally definitive for depression alone and the presence of PTSD in experimental patients could be an alternative explanation for some of the observations that were used to draw the conclusions. Currently, there is limited literature on gene expression profiles of patients with only PTSD, so cross-referencing may be more difficult. Thus, the results in this study are only applicable to both diseases concurrently and cannot be applied to either MDD or PTSD separately. Through a study with an experimental group made of patients with only depression, the possibility of another disease influencing the results could be eliminated. In addition, when interpreting the results of this investigation, it is important to understand that gene expression itself is not the same as protein expression.

In regards to application of the biomarkers we identified, they may serve a diagnostic role. More research, such as a longitudinal dataset on patients who eventually get the disease, is needed to determine whether these biomarkers may also be predictors of the disease. In addition, it must be proven that the upregulation of the pathway does not occur as a result of the presence of the disease. Another future exploration can be whether the upregulation of the ribosomal pathway is a specific marker for comorbid MDD and PTSD or if it is a marker for psychiatric disorders in general. In order to study this problem, essentially the same procedure

Figure 5: Value Distribution Table. Representative box plots of all RNA expression values of selected patients and controls. Red and blue boxes are patients and control groups, respectively.

study, so its application to this investigation's results is limited. What this complementary finding does suggest is that ribosomal dysfunction likely is more related to MDD than to PTSD since both our study and other studies have identified the same biomarker (10). All things considered, the new biomarkers identified in this study are supported by both studies since they identified a similar pathway and came to a similar conclusion. Therefore, we propose that the upregulation of the ribosomal pathway, and its genes within, is a potential blood-based biomarker for comorbid MDD and PTSD.

Another conclusion that can be made from this study is that blood samples can be accurate in finding biomarkers for diseases of the brain. Usually, when searching for biomarkers of brain diseases, only brain tissue samples are used. However, with our study, blood samples from the dataset were used to find that the *RPL* and *MRP* gene groups were upregulated. This conclusion is supported by similar findings that use brain tissue from mice (10). Pragmatically speaking, using blood samples for future biomarker studies is better because the data collection is easier and less invasive. In addition, for future treatment processes in which diagnosis for susceptibility to comorbid MDD may be done before the patient has the disease, blood

used in this investigation should be done with various other psychiatric diseases to see if the upregulation of the ribosome pathway comes up in them, as well. According to current literature, dysfunction in ribosomal gene expression is known to have an influence on ribosomopathies, such as Diamond Blackfan Anemia, 5q-syndrome, and Shwachman Diamond Syndrome, and other diseases such as obesity, metabolic syndrome, and liver disease (9). However, these diseases are not related to neurological diseases, and many of these ribosomopathies are caused by mutations in a few genes, which stands in contrast to the genetic complexity of comorbid MDD and PTSD. In order to determine whether this pathway is a specific marker for comorbid MDD and PTSD, further research on its prevalence in other neurological diseases is needed.

Materials and Methods
Accessing Data

In order to identify biomarkers for comorbid MDD and PTSD, data needs to be obtained and analyzed. Since we lacked the ability to collect blood or brain samples from real patients, publicly available Dataset GSE67663 from NCBI (National Center for Biotechnology Information) was utilized (6). This was a dataset published in December 01, 2015 from Emory University/Atlanta VAMC. In this dataset, blood samples were extracted from 184 patients for analysis. Of the 184 patients, 72 of them served as the control group with a comorbidity status of 0, which means that the patients had neither MDD or PTSD. The other 112 patients had a comorbidity status of 1, which means that they suffered from both MDD and PTSD. Using microarray expression analysis, this study was able to perform a genome-wide differential gene expression survey of the control and experimental group. The expression data was arranged and then analyzed using GEO by sex, age, comorbid PTSD and Depression status, Rin, and Hybridization batch (11). Statistical readouts such as logFC and p-value were used to show the differentially expressed genes. This was useful for preliminary analysis for outliers and for defining groups for further analysis. It is important to note that the p-values determined were based on the one null hypothesis, so multiple hypothesis testing does not affect the true p-value. Also, the value distribution graph was used to view the median centered values, which implies that there were no significant outliers in the data (**Figure 5**).

GEO, STRING, and KEGG

After this preliminary analysis, Gene Expression Omnibus' (GEO) top 250 genes of the dataset were used in order to view specific genes and to begin the differential gene expression analysis (8). Arranged by

p-value, the top 250 genes were listed with statistical values, of which we used p-value and logFC. Only the top 250 genes were used because we decided that p-values after a certain point significantly reduce the reliability of the data (>0.004). For most complex diseases and disorders such as MDD and PTSD that have environmental, traumatic, and genetic onsets, one gene alone is not significant enough to serve as a biomarker. Enriched gene networks and pathways are more significant and valuable. In order to find gene networks and relationships of these 250 genes, we used the STRING Consortium Database of Protein to Protein Interaction Networks (7). This database was useful because it not only showed a visual view of the networks, but also classified the connections into known connections and predicted connections.

Once common gene groups were identified, we began to identify common gene ontology and biological processes. According to STRING analysis, some of the top biological processes were translational termination, translational elongation, and SRP-dependent cotranslational targeting to membrane. The same tools and process were used to identify other GO (gene ontology) terms, including molecular function and cellular component. More results and implications on the study of the GO were explained in the results section. Once GO and networks were established and further analyzed, the Kyoto Encyclopedia of Genes (KEGG) was used, which provides information on complex biological systems and pathways on a genetic level (8). For this particular study, its purpose was to provide further insight into the ribosomal pathway, which was the enriched pathway of the 250 genes on STRING, and how specific genes interact within the pathway. Using the pathway ID 03010, the ribosomal pathway was broken down and analyzed. With the implementation of several bioinformatic tools, we were able to finally propose that the upregulation of the ribosome pathway is a potential biomarker for comorbid MDD and PTSD.

References
1. "PTSD Statistics." Heal My PTSD, 20 Dec. 2017, www.healmyptsd.com/education/post-traumatic-stress-disorder-statistics.
2. Breslau, Naomi, et al. "Psychiatric Sequelae of Posttraumatic Stress Disorder in Women." Archives of General Psychiatry, vol. 54, no. 1, Jan. 1997, pp. 81-87, doi:10.1001/archpsyc.1997.01830130087016.
3. Flory, Janine D., and Rachel Yehuda. "Comorbidity between Post-Traumatic Stress Disorder and Major Depressive Disorder: Alternative Explanations and Treatment Considerations." Dialogues in Clinical Neuroscience 17.2 (2015): 141–150. Print.
4. Wingo, Aliza P. et al. "DICER1 and microRNA

Regulation in Post-Traumatic Stress Disorder with Comorbid Depression." Nature Communications 6 (2015): 10106. PMC. Web.

5. Darby, M M, R H Yolken, and S Sabunciyan. "Consistently Altered Expression of Gene Sets in Postmortem Brains of Individuals with Major Psychiatric Disorders." Translational Psychiatry 6.9 (2016): e890–. PMC. Web.

6. Wingo, Aliza P. "Gene Expression Profiles in Comorbid PTSD and Depression."National Center for Biotechnology Information, U.S. National Library of Medicine.

7. Szklarczyk, Damian et al. "STRING v10: Protein–protein Interaction Networks, Integrated over the Tree of Life." Nucleic Acids Research 43.Database issue (2015): D447–D452. PMC. Web.

8. Kanehisa, Minoru, and Susumu Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes." Nucleic Acids Research 28.1 (2000): 27–30. Print.

9. Narla, Anupama, and Benjamin L. Ebert. "Ribosomopathies: Human Disorders of Ribosome Dysfunction." Blood 115.16 (2010): 3196–3205. PMC. Web. 2 May 2018.

10. Smagin, Dmitry A. et al. "Dysfunction in Ribosomal Gene Expression in the Hypothalamus and Hippocampus Following Chronic Social Defeat Stress in Male Mice as Revealed by RNA-Seq." Neural Plasticity 2016 (2016): 3289187. PMC. Web. .

11. Barrett, Tanya et al. "NCBI GEO: Archive for Functional Genomics Data Sets—update." Nucleic Acids Research 41.Database issue (2013): D991–D995. PMC. Web.

12. Zhou, Xiang et al. "Ribosomal Proteins: Functions beyond the Ribosome." Journal of Molecular Cell Biology 7.2 (2015): 92–104. PMC. Web.