Genetic Identification of Bipolar Disorder

Ziv Cohen <326178266>
Noam Barash Biram <327923595>

Introduction to Bioinformatics – Final Project

Abstract (Up to 250 words)

Brief introduction on the disease:

Bipolar disorder is a psychiatric illness which is characterized by extreme mood swings between euphoria (mania) and depression. It is mainly genetic although the biological mechanisms which are underlying it are mostly unknown.

The knowledge gaps:

The main goal of the analysis:

In this study, we aimed to identify how people who suffer from bipolar disorder differ from healthy people and schizophrenic people on the microenvironmental level – gene expression levels and enriched pathways – in order to better understand, diagnose and treat the bipolar disorder.

Brief overview of your analysis:

We performed differential gene expression to identify potential genes that could act as biomarkers for bipolar disorder and then took a closer look at the genes that seemed to be the best candidates. In addition, we have also conducted GSEA to find pathways which are associated with bipolar disorder. We even tried to classify the bipolar disorder patients into subtypes based on their biological measures.

Your key result(s):

We have identified two candidate genes (LINC02340 and MT1X) to be biomarkers of bipolar disorder and other psychiatric illnesses and we have also found genes which could act as biomarkers in specific regions of the brain: CENPL in area 9 and LACRT and KRT18P11 in area 46. In addition, we found two pathways which are associated with bipolar disorder – the coagulation system and the metabolism of xenobiotics. Furthermore, we succeeded/failed to divide the bipolar patients into subtypes.

Main conclusion(s):

In conclusion, we gathered enough evidence to support our claim that bipolar disorder can be discovered based on biological methods but nonetheless, there is still much research that must be made in order to come to better and more usable findings.

Introduction

Bipolar disorder (BD) is a multicomponent genetic illness that involves severe mood disturbance, neuropsychological deficits, and physiological changes and it is one of the leading causes of disability globally (Rowland and Marwaha, 2018). Patients often experience extreme mood swings from manias to depressions and vice versa. In fact, the name "bipolar disorder" was adopted by the DSM (Diagnostic and Statistical Manual for Mental Disorders) in 1980 to replace the term "manic depression" (Phillips and Kupfer, 2013). The mood swings are different in different individuals and ranges from mild hypomania or depression to severe manias or depressions, sometimes accompanied by psychosis (Miklowitz, 2008; Müller-Oerlinghausen et al., 2002). We tend to classify BD into 3 subtypes: BD I which includes manic episodes, BD II which includes only hypomanic episodes and major depressive episodes and Cyclothymia which is consistent of hypomanias and minor depressions (Cerimele et al., 2014).

BD affects both young and adult people: recently, there have been some evidence that indicates an increase in the prevalence of BD in young people (Moreno et al., 2007). In addition, in the United States, BD patients make up 10% to 25% of all the geriatric patients with mood disorders (Aziz et al., 2006). When it comes to biological sex, men are affected slightly more than women in a ratio of 1.1:1 (Miller and Black, 2020). It is unclear what is the lifetime prevalence of people who are on the bipolar spectrum (suffer from one of the 3 BD subtypes mentioned before) because different studies have come to very different results. In any case, all the studies have found that the patients' lifetime prevalence decreases significantly (Cerimele et al., 2014).

The mortality rate of people with BD is quite high – around 10% to 20% of individuals with this illness has committed suicide and more than a third have attempted suicide at least once (Müller-Oerlinghausen et al., 2002).

As we have established before, the BD portrays a threat on a variety of people in different ages, hence, it is of great importance for us to develop new ways of identifying patients before they experience an outbreak.

In the research literature, it is apparent that diagnosing BD is quite challenging because the diagnosis is made exclusively based on clinical information which is not objective: BD I is diagnosed based on one manic episode, BD II is diagnosed based on depressive and

hypomanic episodes and Cyclothymia is diagnosed based on hypomanic and depressive symptoms that do not count as depressive episodes. In addition, some other psychiatric illnesses resemble the BD's symptoms, especially recurring unipolar depressive disorder (a disorder which is characterized by recurrent depressive episodes). The misdiagnosis between unipolar disorder and BD is made the most when differentiating unipolar disorder and BD II, that's because patients who suffer from BD II do not experience manic episodes. However, it is difficult to differentiate BD patients in general because manic episodes are rarer than the depressive ones (Phillips and Kupfer, 2013). Furthermore, it is also extremely challenging to come to proper findings because of the insufficient sample sizes of the current studies (Medeiros and Goes, 2022).

There are a lot of things which are still unknown about the BD's diagnosis, nature and treatment: first of all, it is unknown how to diagnose patients with BD based on biological methods besides tracking down their family history in order to identify potential risks of having BD which is a tedious and inaccurate method. In addition, there are no known specific biomarkers (biological measures that could indicate about the presence or the severity of the illness) for BD (Frey et al., 2013; Salagre and Vieta, 2022). It is neither known how to differentiate BD patients from people who suffer from similar psychiatric illnesses such as recurring unipolar depressive disorder and schizophrenia (Salagre and Vieta, 2022).

Clearly, our limited knowledge about BD and lack of understanding of the biological mechanisms that are underlying it, combined with the insufficiency of proper sized samples, are the main reasons for the challenges we have been facing in regard to BD.

In the last decade, various studies have focused on the genetics of bipolar disorder and the various risk factors that can affect its development (Rowland and Marwaha, 2018). It is found that bipolar disorder has a major genetic component and it seems to be very heritable (Kim et al., 2021). The new findings shows that there are some genes that seem to be associated with bipolar disorder. Those include: SERINC2 (increases the risk of bipolar disorder in Asian population) and SLC6A2 (affects the likelihood of having bipolar disorder I and its severity) (Kim et al., 2021; Yang et al., 2021).

One of the popular approaches in the pursuit to better understand the genetics of bipolar disorder, is to perform GWAS (Genome-Wide Association Study) which helps

identifying significant SNPs (Single Nucleotide Polymorphisms) that are associated with this illness. It is also common to use PRS (polygenic risk scores) – in general, those scores are the summation of all the individual's alleles which are associated with the phenotype (in this context, the phenotype is bipolar disorder) weighted by the size of their effect on it – which provides a way to approximate how well a patient will respond to a clinical treatment. Another useful method is WES (whole-exome sequencing) which helps identifying rare variants in genes and brain-related pathways. Finally, there is WGS (whole-genome sequencing) which is the most extensive yet most expensive and technically challenging method (Oraki Kohshour et al., 2022).

We believe that the brain's microenvironment withholds the potential of uncovering new ways of identifying BD based on biological measures. In this study, we used the data collected in previous studies (Akula et al., 2014; Hu et al., 2016) in order to try and check if the technology and algorithms which are available to us today could shed some light on the biological mechanisms underlying BD and identify some significant biological differences between BD patients and healthy individuals and perhaps even between BD patients and people who suffer from similar illnesses such as schizophrenia.

If we identify some kind of biomarkers for BD, it could enable us to diagnose BD patients earlier – even before they experience an outbreak. In addition, if the said biomarkers would be specific for BD, it could help differentiating it from other psychiatric illnesses and enable many patients to get their appropriate medicine and treatments.

Results

Currently, the identification of bipolar disorder is made based on behavioral factors which could be detected and measured only after an outbreak of the illness. The purpose of our study was to uncover some of the biological mechanisms of this disorder – whether it be identifying genes which are associated specifically with BD, enriched pathways which are affected by BD, new ways to classify BD into subtypes based on biological differences etc.

We used RNA-seq gene expression data from E-GEOD-78936 (Hu et al., 2016) and E-GEOD-53239 (Akula et al., 2014) to compare samples of different brain areas from BD patients, schizophrenia (SZ) patients and healthy individuals.

Identifying biomarker genes

First, we aimed to identify genes that are differentially expressed in BD patients relative to SZ patients and healthy individuals. We performed the differential expression analysis using DESeq2 (Love et al., 2014) by using the raw count data and the corresponding metadata regarding the diagnosis and brain area of each sample.

We have used PCA (which is a method of visualizing high-dimensional data in a more simplistic and easier to conceive way) and plotted three graphs: one is classified based on the diagnosis of each sample, the second is classified based on the brain area which is the source of the samples and the last one is based on both the diagnosis and the source of the samples. The PCA plots have showed complete chaos and nonsense which actually validates that the normalization of the two distinct datasets we have based our study on did not separate them into two clusters but rather succeeded in combining them together.

After validating our data, we have plotted two volcano plots which depict the differentially expressed genes in BD patients relative to SZ patients and healthy patients. We found that there is only one gene which is significantly highly expressed in BD compared to SZ and between BD and healthy people, we have identified 4 significantly highly expressed genes (figure 1). In both those comparisons, we have identified only the gene MTND6P4 as a common significant gene.

In addition, we have examined the highly expressed genes in different areas of the brain in BD patients relative to the control groups (healthy and SZ) using volcano plots. As the graphs show, there is a common gene which is highly expressed in BD patients exclusively in the 9^{th} area of the brain that could indicate the presence of BD. It is also

apparent that neither BD nor SZ affect the 11th area's expression patterns in any significant way and that there are several genes in the 24th and 46th areas that could distinguish BD sample from a healthy sample but there are no genes that could do so between BD samples and SZ samples (figure 2).

Those two findings show us very clearly that biomarkers for BD do exist and that they are not even hard to find. In the next step, we aimed to better understand the biomarkers we have identified.

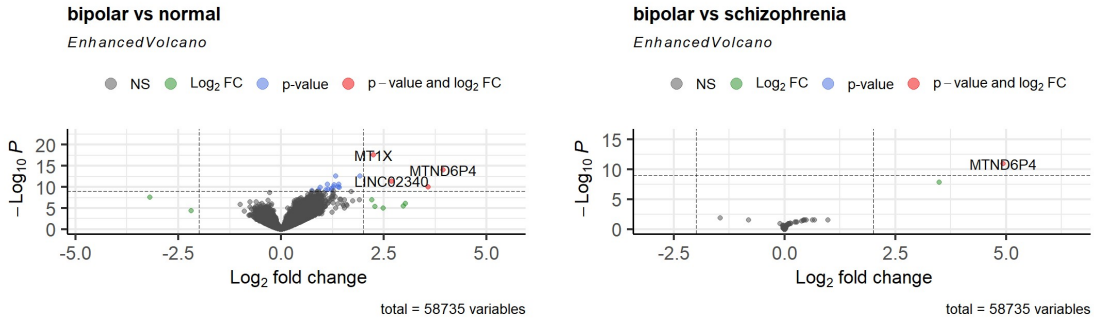
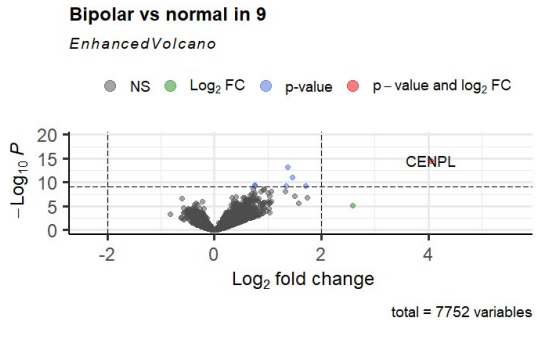
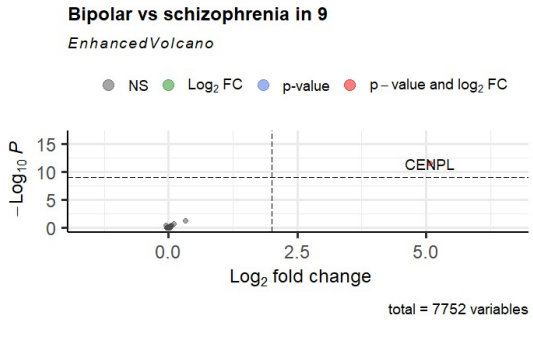


Figure 1: Two volcano plots that display the genes that were most differentially expressed in BD patients (red) relative to healthy individuals (on the left) and SZ patients (on the right).

Brain Area	BD relative to Healthy	BD relative to SZ
9		

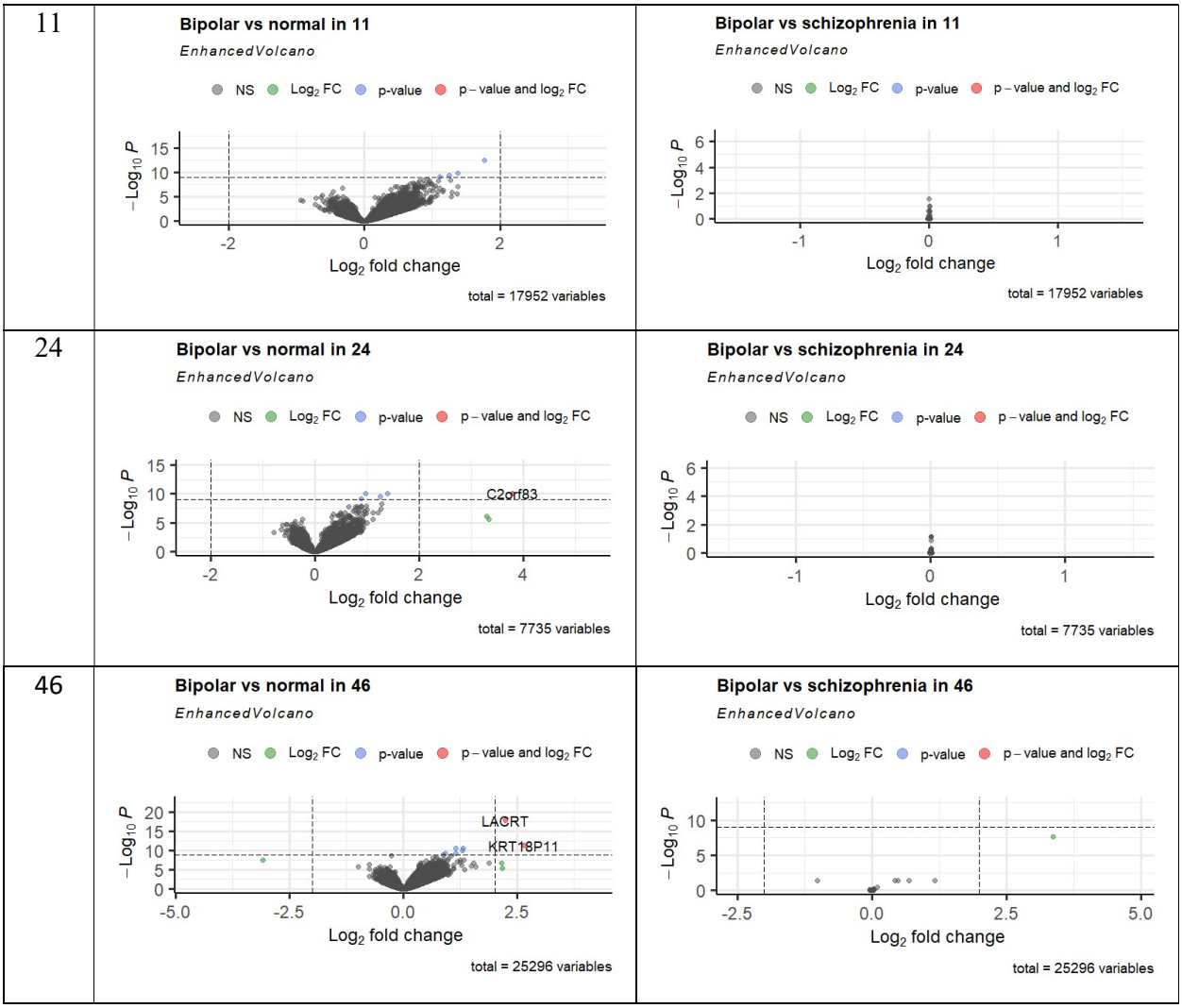


Figure 2: Eight volcano plots that display the genes that were most differentially expressed in different areas of the brain in BD patients (red) relative to healthy patients (left column) and SZ patients (right column).

After we have completed the differential expression analysis, we aimed to further understand the four highly expressed genes that we have found in figure 1 (MTND6P4, LINC02340, IL1RL1, MT1X). Using violin plots, we compared the expression levels of the genes in the three populations we are dealing with (BD patients, SZ patients and healthy patients) but to our inconvenience, the plots looked a bit odd because of a small number of samples which had extremely high gene counts. To solve this problem, we have used a logarithmic scale which enabled us to see them properly. As expected, the plots of BD and SZ were very similar to each other while mostly distinct from those of the healthy

individuals (figure 3). It seems that the MTND6P4 which seemed promising as a specific biomarker for BD, is not actually differentially expressed between the BD and SZ patients but rather that the difference perhaps has occurred as a result of noisy data. Nevertheless, MTND6P4 does function as a great biomarker to distinguish between healthy samples and either BD or SZ samples and so does the genes LINC02340 and MT1X who shows great resemblance between the counts of BD and SZ while maintaining very distinct measures in comparison to the healthy control group. It is important to mention that the IL1RL1's graph looks insignificant and it is plausible that the volcano plot from figure 1 identified it as a significant gene because of noisy data which got removed as part of the algorithm that calculates the significancy of the differences between the violins. In any case, it is better not to use this gene as a biomarker because of its BD gene counts debatable resemblance to both the SZ gene counts and the healthy gene counts.

In addition, we have also used violin plots to further understand the genes CENPL (which is allegedly a specific biomarker for BD in the 9th area of the brain), C2orf83 (which is allegedly a biomarker to distinguish BD from healthy patients in the 24th area of the brain), LACRT and KRT18P11 (which are allegedly biomarkers to distinguish BD from healthy patients in the 46th area of the brain) which we have identified in figure 2. Again, the plots of BD and SZ seemed very similar to each other while mostly distinct from those of the healthy individuals (figure 4). It seems that like MTND6P4, CENPL which first seemed like a potential specific biomarker for BD in area 9, was not really significantly differentially expressed between BS and SZ. If that's so, then unfortunately we have failed to find any gene that acts as a biomarker to distinguish between BD patients and SZ patients. However, CENPL does work as a great biomarker to distinguish between healthy samples and either BD or SZ samples in area 9 and so does the genes LACRT and KRT18P11 in area 46. It is important to mention that the C2orf83's graph in area 24 looks insignificant and it is plausible that it was suggested as a significant biomarker due to the same mistake we have witnessed before regarding IL1RL1.

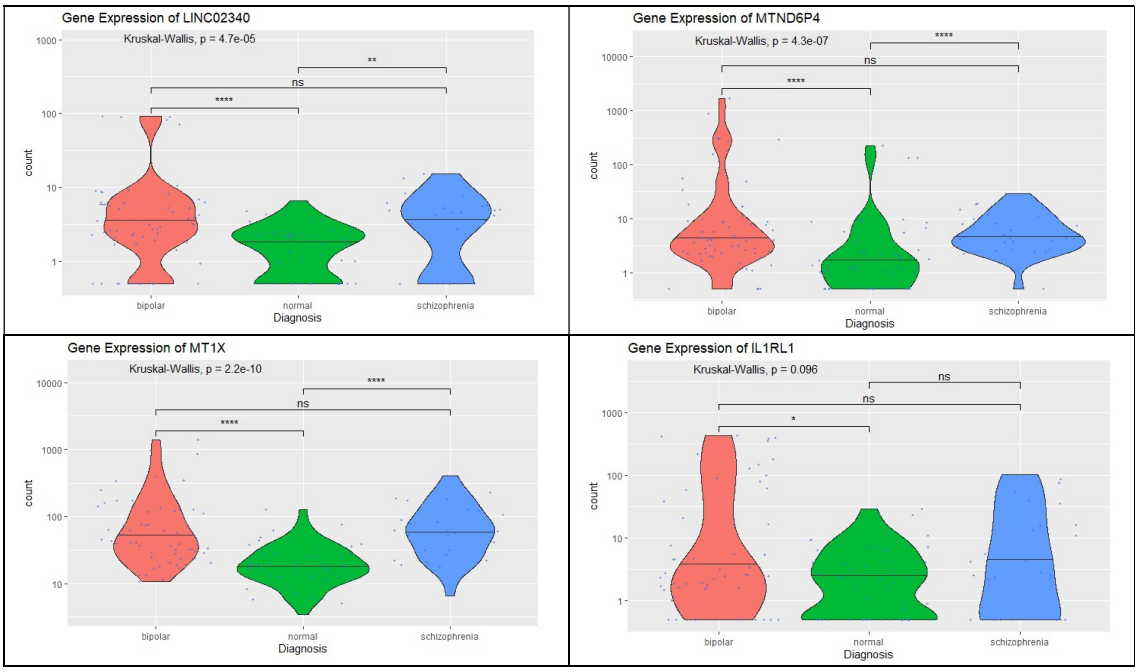


Figure 3: Four violin plots that correspond to the gene expression of the four significant genes (top-right: MTND6P4, top-left: LINC02340, bottom-left: IL1RL1, bottom-right: MT1X) in BD (red), SZ (blue) and healthy people (green).

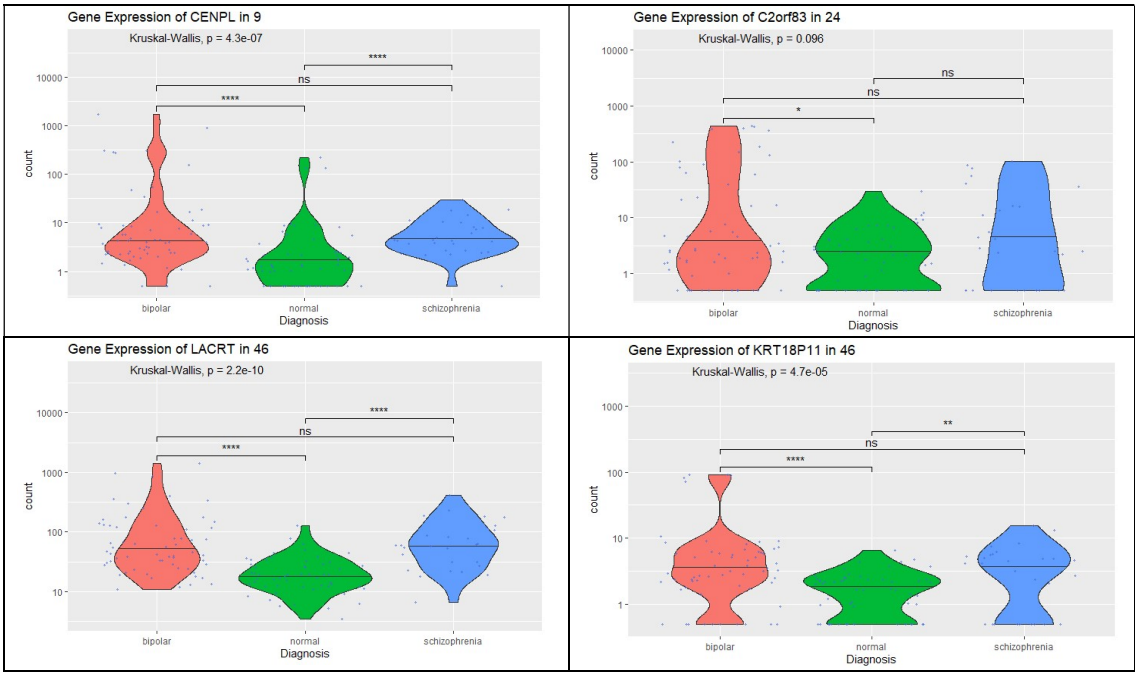


Figure 4: Four violin plots that correspond to the gene expression of the four area-specific significant genes (top-right: C2orf83 in area 24, top-left: CENPL in area 9, bottom-left: LACRT

in area 46, bottom-right: KRT18P11 in area 46) in BD (red), SZ (blue) and healthy people (green).

Identifying enriched pathways

In addition, we wanted to search for enriched pathways in BD patients relative to the healthy control group and SZ patients. We used the GSEA algorithm (Aravind et al., 2005; Mootha et al., 2003) to find the enriched pathways and found that sadly, there are no enriched pathways between BD and SZ. Fortunately, we were able to find some enriched pathways between BD and the control group (figure 5). It is not surprising to find out that such a complex disorder has great influence on so many pathways.

The least significant pathways we have found were "HALLMARK COAGULATION" and "HALLMARK XENOBIOTIC METABOLISM". Looking at the GSEA website, we can see that the first pathway is made up of genes which encode for components in the blood coagulation system and the second pathway is made up of genes which encode for proteins that are involved in the processing of drugs and xenobiotics.

The most significant pathways we have found were "HALLMARK INTERFERON ALPHA RESPONSE", "HALLMARK INTERFERON GAMMA RESPONSE " and "HALLMARK IL6 JAK STAT3 SIGNALING". Looking at the GSEA website, we can see that the first pathway is made up of genes which are up-regulated by alpha interferon proteins, the second pathway is made up of genes which are up-regulated by gamma interferon proteins and the third one is made up of genes which are up-regulated by IL6 via STAT3. The interferon proteins are proteins which are usually get secreted by a cell as part of an immunological response to threats like viruses, germs, parasites or in the case of the cell becoming cancerous.

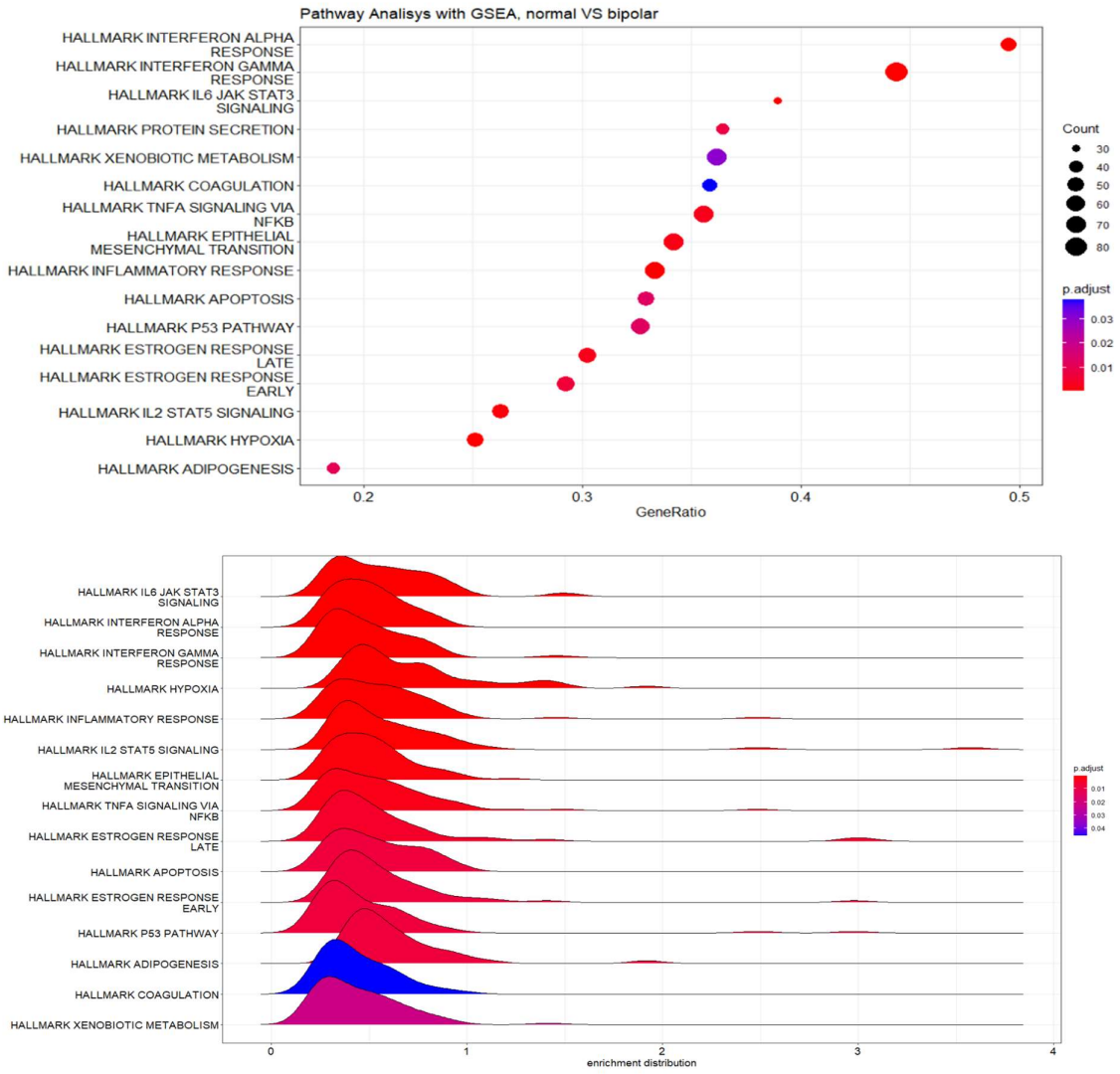


Figure 5: Enriched pathways in BD relative to the healthy control group.

Classifying BD patients into subtypes

D

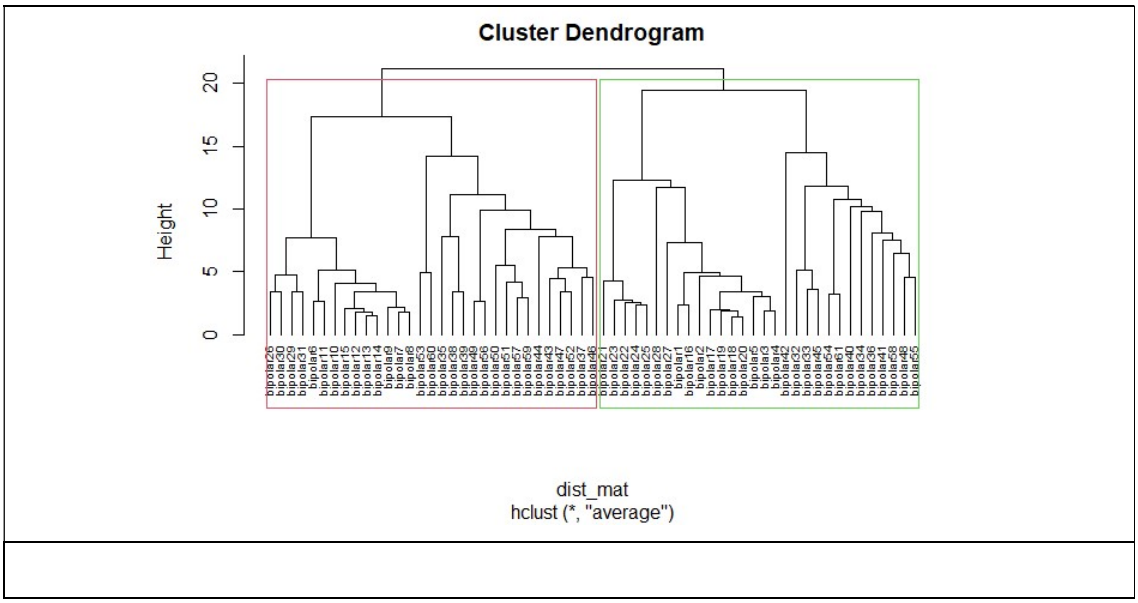


Figure 6: Unsupervised classification of the BD patients (top) and the enriched pathways that distinguish them from one another (bottom).

Discussion

In this current study, we have identified the genes MTND6P4, LINC02340 and MT1X as potential biomarkers of bipolar disorder and possibly more psychiatric illnesses. Furthermore, we have discovered that the gene CENPL could act as a biomarker for bipolar disorder in the 9th area of the brain and that the genes LACRT and KRT18P11 could do so in the 46th area of the brain. In addition, we have found that there is a strong connection between bipolar disorder and the coagulation system and a slightly lesser connection between bipolar disorder and the metabolism of xenobiotics.

Unfortunately, the research literature lacks any references regarding the biomarker genes we have identified in the context of bipolar disorder. It is possible that this missing validation is caused due to the insufficient research of the topic as we have mentioned before. However, the research literature does support the connection we have found between the bipolar disorder and the coagulation system (Hoirisch-Clapauch et al., 2014) and for the association between the metabolism of xenobiotics using the enzyme Cytochrome P450 and bipolar disorder (Altaf-Ul-Amin et al., 2021). Those are very important findings that proves that as we have thought, bipolar disorder's effects on the biological microenvironment can be measured and identified. Furthermore, being able to address a pathway which is known to be affected by bipolar disorder could help future clinical developments in the field.

Mention the limitations of your analysis.

The main reason for our analysis' limitations is the inability to validate most of our results using the research literature. In addition,

What would you do next? Are there any ways to overcome those limitations? What future experiment can you suggest answering your biological question that will address what is still unknown?

In order to surpass the current limitations, extensive research on the biological mechanisms which are underlying the bipolar disorder must be conducted. After accumulating sufficient data, differential gene expression should be made in order to try and identify specific biomarkers for bipolar disorder as we have failed to do so.

References

Akula, N., Barb, J., Jiang, X., Wendland, J.R., Choi, K.H., Sen, S.K., Hou, L., Chen, D.T.W., Laje, G., Johnson, K., et al. (2014). RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. Molecular Psychiatry *19*, 1179–1185.

Altaf-Ul-Amin, Md., Hirose, K., Nani, J. v, Porta, L.C., Tasic, L., Hossain, S.F., Huang, M., Ono, N., Hayashi, M.A.F., and Kanaya, S. (2021). A system biology approach based on metabolic biomarkers and protein–protein interactions for identifying pathways underlying schizophrenia and bipolar disorder. Scientific Reports *11*, 14450.

Aravind, S., Pablo, T., K, M.V., Sayan, M., L, E.B., A, G.M., Amanda, P., L, P.S., R, G.T., S, L.E., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences *102*, 15545–15550.

Aziz, R., Lorberg, B., and Tampi, R.R. (2006). Treatments for late-life bipolar disorder. The American Journal of Geriatric Pharmacotherapy *4*, 347–364.

Cerimele, J.M., Chwastiak, L.A., Dodson, S., and Katon, W.J. (2014). The prevalence of bipolar disorder in general primary care samples: a systematic review. General Hospital Psychiatry *36*, 19–25.

Frey, B.N., Andreazza, A.C., Houenou, J., Jamain, S., Goldstein, B.I., Frye, M.A., Leboyer, M., Berk, M., Malhi, G.S., Lopez-Jaramillo, C., et al. (2013). Biomarkers in bipolar disorder: A positional paper from the International Society for Bipolar Disorders Biomarkers Task Force. Australian and New Zealand Journal of Psychiatry *47*, 321–332.

Hoirisch-Clapauch, S., Nardi, A.E., Gris, J.-C., and Brenner, B. (2014). Coagulation and mental disorders. Rambam Maimonides Medical Journal *5*, e0036–e0036.

Hu, J., Xu, J., Pang, L., Zhao, H., Li, F., Deng, Y., Liu, L., Lan, Y., Zhang, X., Zhao, T., et al. (2016). Systematically characterizing dysfunctional long intergenic non-coding RNAs in multiple brain regions of major psychosis. Oncotarget *7*, 71087–71098.

Kim, S.Y., Kim, H.N., Jeon, S.W., Lim, W.J., Kim, S.I., Lee, Y.J., Kim, S.Y., and Kim, Y.K. (2021). Association between genetic variants of the norepinephrine transporter gene (SLC6A2) and bipolar I disorder. Progress in Neuro-Psychopharmacology and Biological Psychiatry *107*.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

Medeiros, G.C., and Goes, F.S. (2022). Genome-wide association study biomarkers in bipolar disorder. In Biomarkers in Bipolar Disorders, (Elsevier), pp. 125–139.

Miklowitz, D.J. (2008). Bipolar disorder : a family-focused treatment approach (Guilford Press).

Miller, J., and Black, D. (2020). Bipolar Disorder and Suicide: a Review. Current Psychiatry Reports *22*.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics *34*, 267–273.

Moreno, C., Laje, G., Blanco, C., Jiang, H., Schmidt, A.B., and Olfson, M. (2007). National Trends in the Outpatient Diagnosis and Treatment of Bipolar Disorder in Youth. Archives of General Psychiatry *64*, 1032–1039.

Müller-Oerlinghausen, B., Berghöfer, A., and Bauer, M. (2002). Bipolar disorder. The Lancet *359*, 241–247.

Oraki Kohshour, M., Papiol, S., Ching, C.R.K., and Schulze, T.G. (2022). Genomic and neuroimaging approaches to bipolar disorder. BJPsych Open *8*, e36.

Phillips, M.L., and Kupfer, D.J. (2013). Bipolar disorder diagnosis: challenges and future directions. The Lancet *381*, 1663–1671.

Rowland, T.A., and Marwaha, S. (2018). Epidemiology and risk factors for bipolar disorder. Therapeutic Advances in Psychopharmacology *8*, 251–269.

Salagre, E., and Vieta, E. (2022). Biomarkers in bipolar disorder: an overview. In Biomarkers in Bipolar Disorders, (Elsevier), pp. 1–18.

Yang, D., Chen, J., Cheng, X., Cao, B., Chang, H., Li, X., Yang, C., Wu, Q., Sun, J., Manry, D., et al. (2021). SERINC2 increases the risk of bipolar disorder in the Chinese population. Depression and Anxiety *38*, 985–995.