Genetic Identification of Bipolar Disorder

Ziv Cohen <326178266>
Noam Barash Biram <327923595>

Introduction to Bioinformatics – Final Project

Abstract

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. It is yet unknown how to identify patients with bipolar disorder based on biological measures but rather only based on behavioral patterns which can be detected only after an outbreak of the disorder – too late to prevent or prepare for the condition. 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. We performed differential gene expression and GSEA to identify potential genes and pathways that could act as biomarkers for bipolar, we used xCell to try and find what changes in the cellular composition happen in the presence of bipolar disorder and lastly, we tried to classify the BD samples using unsupervised clustering based on biological measures. We have identified the genes MTND6P4, LINC02340 and MT1X as candidates to be biomarkers of bipolar disorder. We have also found that the genes CHI3L2, MTND6P4 and MT1X could be biomarkers of bipolar disorder specifically in the 46^{th} area of the brain. Moreover, we found a few pathways which are associated with bipolar disorder, the most significant of them were the interferon alpha/gamma proteins and the hypoxia pathway. Unfortunately, we failed both to spot a difference in the cellular composition between bipolar disorder and healthy patients and to divide the bipolar patients into subtypes. 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 mostly genetic illness that involves severe mood disturbances, 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, some new evidence indicating an increase in the prevalence of BD in young people was found (Moreno et al., 2007). In addition for that, 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 real and severe 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 a challenge all of itself since the diagnosis is made exclusively based on non-objective clinical information: 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 most apparent when differentiating unipolar disorder and BD II, that's because patients who suffer from BD II do not experience manic episodes. However, it is also difficult to differentiate BD I patients from unipolar ones because manic episodes are rarer than the depressive ones in both BD I and BD II (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 regarding 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, for example, the genes 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 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) – which, in general, are the summation of all the individual's alleles that 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 for uncovering new ways of identifying and diagnosing 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 shed some light, using the technology and algorithms that are available for us today, on the biological mechanisms underlying BD and identify some significant biological differences between BD patients and healthy individuals; perhaps even between BD patients and people who suffer from similar illnesses such as schizophrenia.

If we succeed to identify any biomarkers for BD, it could enable us to diagnose BD patients earlier – even before they experience some trigger that would cause the outbreak. In addition, if 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. – in order to improve the identification process of BD and possibly even future treatments.

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 (H) individuals. The samples from these two studies combined are shown in table 1.

It is important to mention that the brain areas are numbered according to the Brodmann area system (Brodmann, 1909) which is broadly used in the scientific community.

Brain area	Number of BD samples	Number of SZ samples	Number of H samples	Sum
9	7	6	6	19
11	16	16	12	44
24	7	6	6	19
46	31	0	31	62
Sum	61	28	55	144

Table 1: The combined samples from studies E-GEOD-78936 (Hu et al., 2016) and E-GEOD-53239 (Akula et al., 2014).

Identifying biomarker genes

We aimed to identify genes that are differentially expressed in BD patients relative to the SZ and H control groups. To do so, we started by performing 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 then used PCA (which is a method of visualizing high-dimensional data in a more simplistic and easier to conceive way) to plot 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 the SZ and H patients. We found that there is only one gene which is significantly highly expressed in BD compared to the SZ group while between BD and the H group, 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 (H and SZ) using volcano plots. As the graphs show, in areas 9, 11 and 24, there are no genes which are highly expressed in BD patients relative to neither H nor SZ patients. The only part of the brain that could indicate the presence of BD is area 46 which is, unfortunately, the only area (out of the four areas we are dealing with in this study) that we do not have SZ samples from (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 potential biomarkers we have identified.

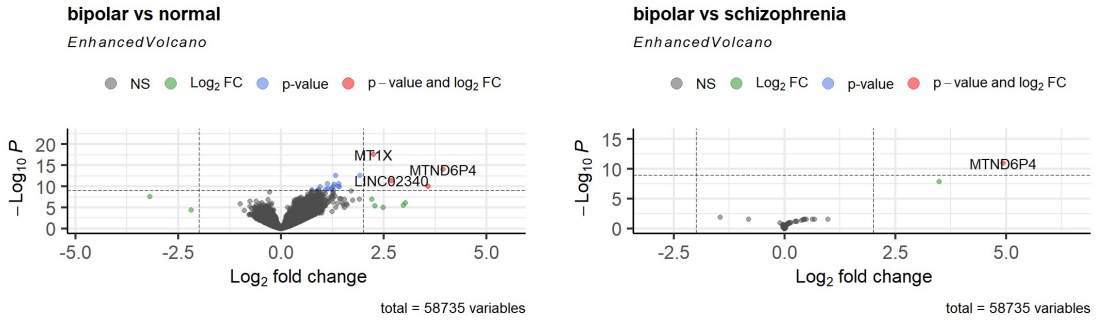


Figure 1: Two volcano plots that display the genes that were most differentially expressed in BD patients (red) relative to H patients (on the left) and SZ patients (on the right).

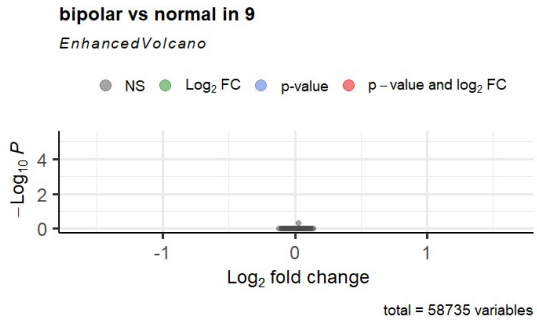
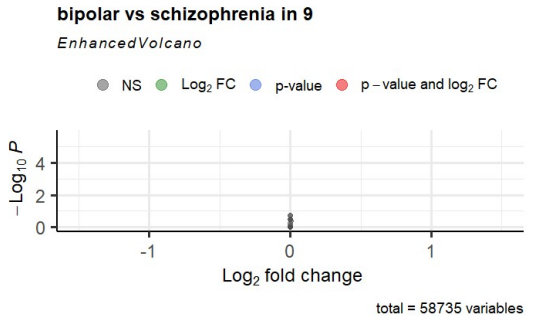
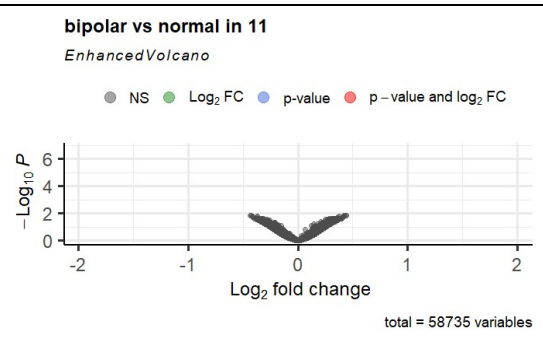
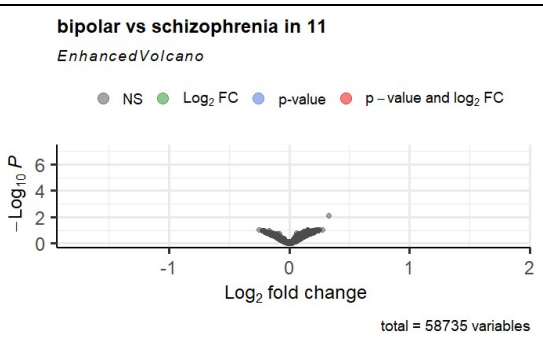
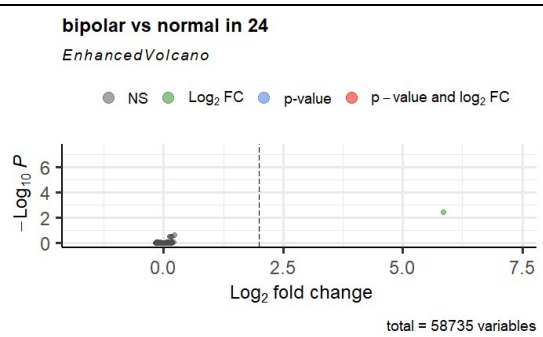
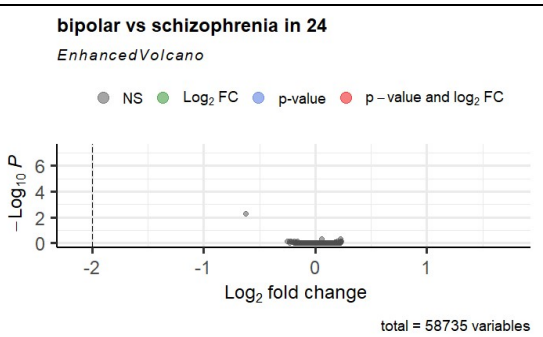
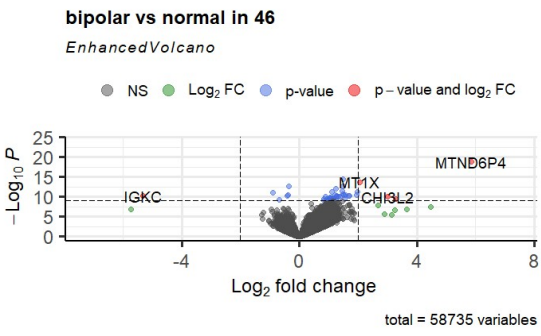
Brain Area	BD relative to H	BD relative to SZ
9		
11		
24		
46		X

Figure 2: Seven volcano plots that display the genes that were most differentially expressed in different areas of the brain in BD patients (red) relative to H 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 H 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 H group (figure 3). It seems that the MTND6P4 which was once 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 (in the violin plots, we have used the Wilcoxon test to determine significancy in the differences between the plots while the DESeq algorithm is using Wald test with a BH correction to point out genes which are highly differentially expressed). Nevertheless, MTND6P4 does function as a great biomarker to distinguish between H 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 H 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 H gene counts.

We have also used violin plots to further understand the genes CHI3L2 and IGKC which allegedly could function as biomarkers in area 46. Moreover, we have plotted the counts of the genes MTND6P4 and MT1X, that we have already encountered in figure 3, this time only based on samples from area 46. Except for the graph of the gene IGKC, all of the graphs seemed significant enough for us to consider the three remaining genes as potential biomarkers to distinguish BD from H patients based on the 46^{th} area of the brain (figure 4).

The discovery of the genes MTND6P4 and MT1X as biomarkers twice (once when searching for a general biomarker and once when searching for biomarkers in the 46^{th} area of the brain) is very promising and reassuring.

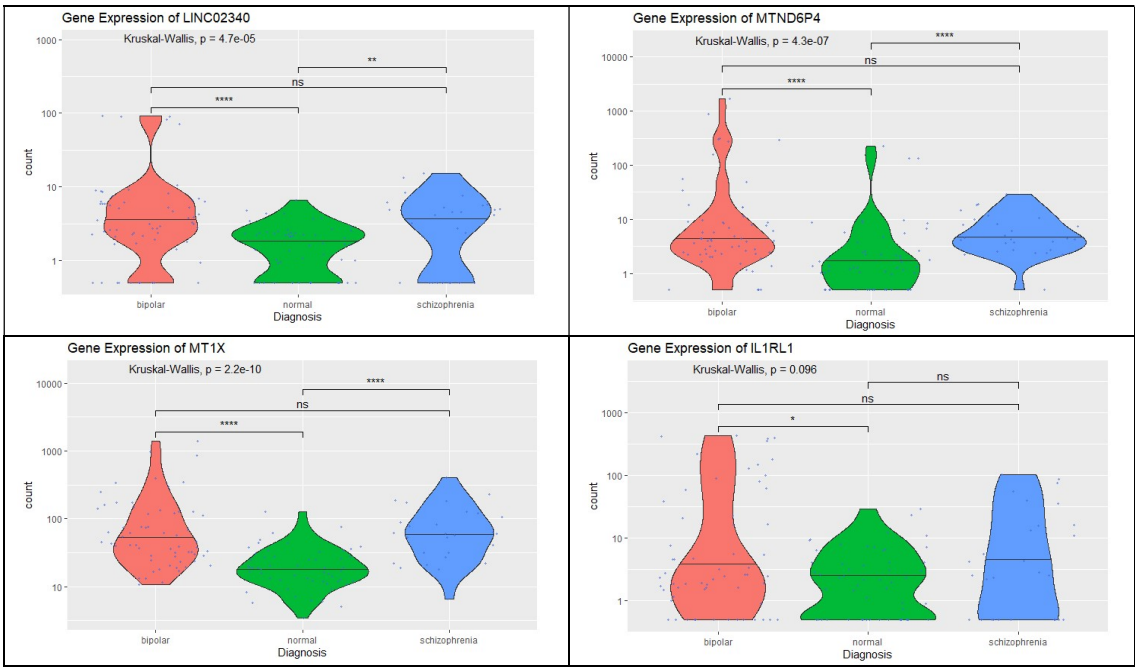


Figure 3: Four violin plots that correspond to the gene expression of the four significant genes (top-right: MTND6P4, top-left: LINC02340, bottom-left: IL1RL1, bottom-right: MT1X) in the BD (red), SZ (blue) and H (green) groups.

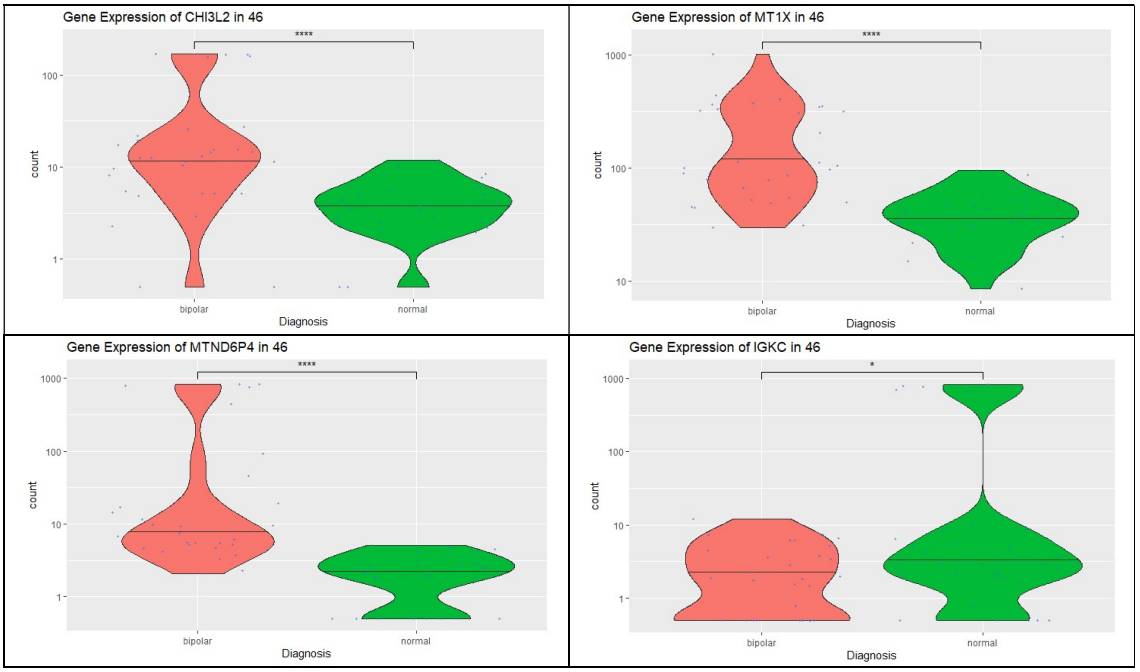


Figure 4: Four violin plots that correspond to the gene expression of the four significant genes in area 46 (top-right: MT1X, top-left: CHI3L2, bottom-left: MTND6P4, bottom-right: IGKC) in BD (red) and H (green) groups.

Identifying enriched pathways

In addition, we wanted to search for enriched pathways in BD patients relative to the H and SZ control groups. 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 H control group but because the GSEA algorithm is partly random, we had to perform the process in an iterative way (we found 1 million iterations to be sufficient) until we got an absolute result (figure 5). It is not surprising to find out that such a complex disorder has great influence on so many pathways.

The most significant pathways we have found were "HALLMARK INTERFERON ALPHA RESPONSE", "HALLMARK INTERFERON GAMMA RESPONSE " and "HALLMARK HYPOXIA". 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 in a response to low oxygen. The interferon proteins are proteins that usually get up-regulated by a cell as part of an immunological response to threats such as viral infections (Andrea et al., 2002).

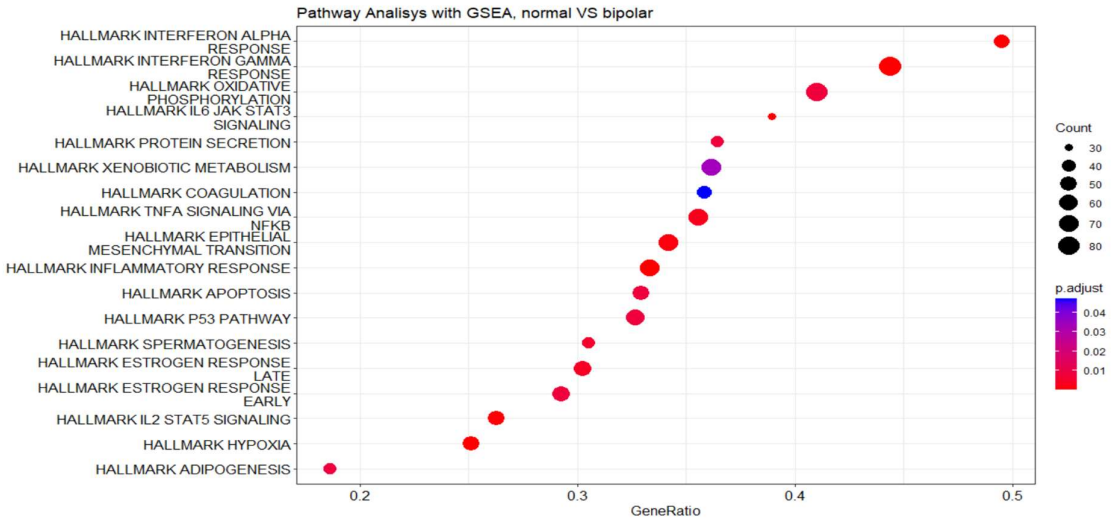


Figure 5: Enriched pathways in BD relative to the H control group.

Examining the changes in the cellular composition

We wanted to try and figure out how does the cellular composition of the different areas of the brain changes due to the presence of BD. We used xCell (Aran et al., 2017) to analyze the samples' cellular composition even though we did not expect it to work out because xCell is trained by a reference which is made mostly by immune cells. To our great surprise we did get some results out of the xCell analysis (figure 6) although they were of no use for us – we could not manage to identify cell types which separated the sample population into distinct clusters of BD and H patients.

We have come up with some plausible explanations for the odd division: it could be that the cells are dividing the population according to an unknown factor (e.g. sex, age, ethnicity, etc.), that the division is successful but based on different parameters than we have expected hance seems wrong or perhaps that an improved version of xCell, containing the gene signatures of more cell types, could identify a cell type which is more suitable for the division.

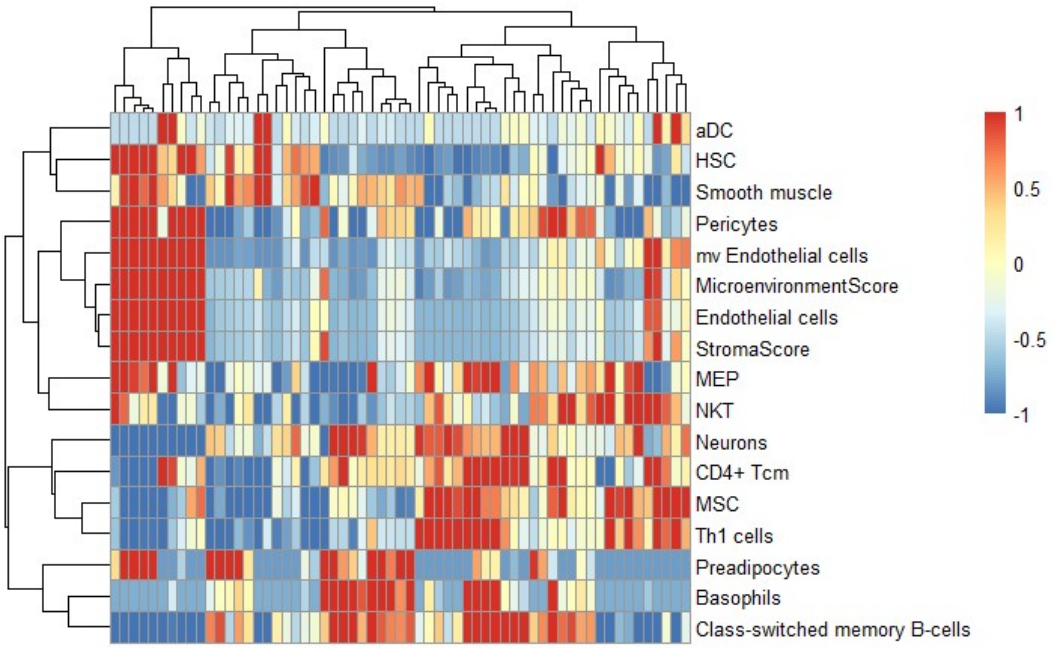


Figure 6: Heatmap of the most diverse cell types between BD and H patients.

Classifying BD patients into subtypes

To finish up, we wanted to search for a way to classify BD patients into subtypes which can be differentiated by biological measures. We used unsupervised clustering in the form of a

dendrogram to classify the samples (figure 7) but unfortunately, the classification process has provided us a poor classification (one cluster included 2 samples while the second and third contained around 20 samples each).

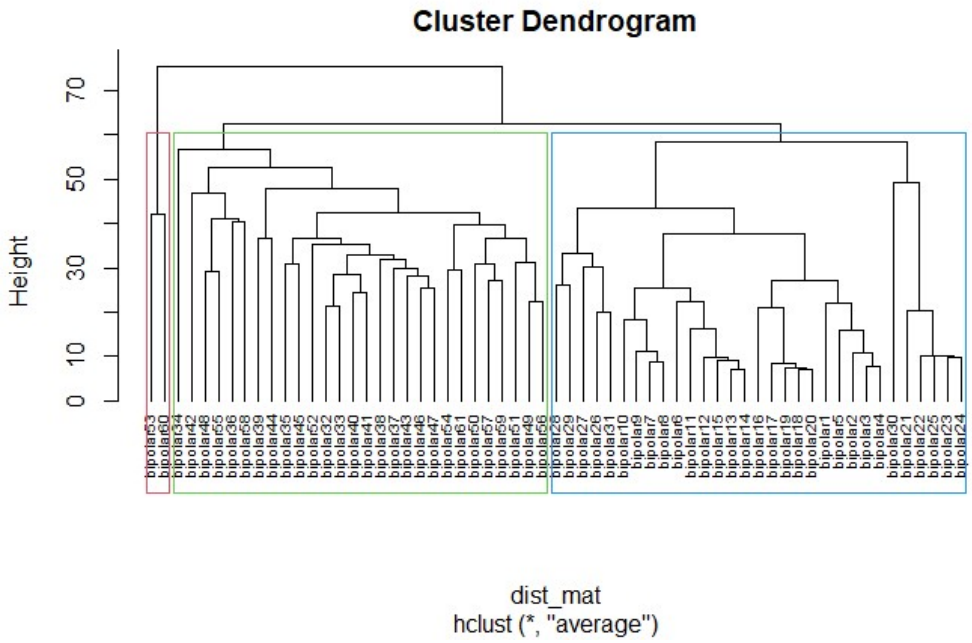


Figure 7: Unsupervised classification of the BD patients.

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 CHI3L2 could act as a biomarker for bipolar disorder in the 46th area of the brain as well as the forementioned genes MTND6P4 and MT1X. In addition, we have found that there is a strong connection between bipolar disorder and the expression of interferon proteins and the hypoxia pathway.

One of the studies we have based our study on, has confirmed our finding that MTND6P4 and MT1X are potential biomarkers for BD (Akula et al., 2014) but unfortunately, the research literature lacks any references regarding the rest of the biomarker genes we have identified in the context of bipolar disorder (LINC02340 for all of the samples and CHI3L2 for samples from area 46 alone). 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 interferon gamma proteins (Yoon and Kim, 2012) and for the association between hypoxia and bipolar disorder (Haukvik et al., 2014). It is also suspected that interferon alpha treatment could result in a bipolar disorder (Iancu et al., 1997) this might not support the association between the interferon alpha pathway and bipolar disorder directly but it does gives us an indication that there is a connection between the two.

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.

One of the main reasons for our analysis' limitations is the inability to validate some of our results using the research literature. In addition, we have used a limited number of samples and perhaps using a larger and wider range of samples we could have gotten better and more concise results. Furthermore, bipolar disorder has a very complex biological mechanisms so trying to map it entirely is almost impossible, what we have found in this study is merely the very peak of the iceberg which is the bipolar disorder. We have also been limited by our lack of information about both the environmental conditions the specimens underwent during their lives (and especially the trigger that led to the outbreak)

and about the rest of the specimens' phenotypes beside their diagnosis. Both of these factors could greatly affect the specimens' gene expressions and tendency of having the illness. Another major limitation is that all of the samples have been accumulated from deceased individuals so their cause of death and the death itself, might have affected their gene expressions.

In order to surpass the current data-based 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.

It is also recommended to reperform the clustering we have done in the end of our analysis while using a larger sample size with more known phenotypes. It could be that as we have mentioned before, the division we have come to was successful but based on something we did not expect but could have a clinical value in regard to treatment and prevention.

Moreover, when an improved version of xCell will be available, it's best to try and check if changes in the cellular composition of the brain due to bipolar disorder would become apparent.

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.

Andrea, M. de, Ravera, R., Gioia, D., Gariglio, M., and Landolfo, S. (2002). The interferon system: an overview. European Journal of Paediatric Neurology : EJPN : Official Journal of the European Paediatric Neurology Society *6 Suppl A*, A41-6; discussion A55-8.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biology *18*, 220.

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.

Brodmann, K. (1909). Vergleichende lokalisationslehre der grobhirnrinde. In Vergleichende Lokalisationslehre Der Grobhirnrinde, p. 324.

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.

Haukvik, U.K., McNeil, T., Lange, E.H., Melle, I., Dale, A.M., Andreassen, O.A., and Agartz, I. (2014). Pre- and perinatal hypoxia associated with hippocampus/amygdala volume in bipolar disorder. Psychological Medicine *44*, 975–985.

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.

Iancu, I., Sverdlik, A., Dannon, P.N., and Lepkifker, E. (1997). Bipolar disorder associated with interferon-alpha treatment. Postgraduate Medical Journal *73*, 834–835.

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.

Yoon, H.-K., and Kim, Y.-K. (2012). The T allele of the interferon-gamma +874A/T polymorphism is associated with bipolar disorder. Nordic Journal of Psychiatry *66*, 14–18.