Studying the effects of bipolar disorder on the microenvironment of different areas in the brain

Ziv Cohen <326178266>
Noam Barash Biram <327923595>

Introduction to Bioinformatics – Final Project

Abstract

*Up to 250 words.

Brief introduction on the disease:

The knowledge gaps:

The main goal of the analysis: We aim to identify how people who suffer from bipolar disorder differ from healthy people and schizophrenic people on the microenvironment level – gene expression levels, enriched pathways and tissue composition – in order to better understand, diagnose and treat the bipolar disorder.

Brief overview of your analysis:

Your key result(s):

Main conclusion(s):

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 came 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 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), 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 order 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 researches (Akula et al., 2014; Hu et al., 2016) in order to try and check if the machine learning algorithms 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.

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

As we have previously established, there is a lack of knowledge and understanding of BD which results in an insufficient diagnosis and treatment. We are hoping that we could uncover some of the mysteries of this disorder using machine learning algorithms – 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.

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

We have performed PCA which is a method of visualizing high-dimensional data in a more simplistic and easier to conceive way. We have plotted three PCAs: 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 area of the samples. The PCA plots have showed complete chaos which actually validates that the normalization of the two distinct datasets we have based our study on did not separate them into two clusters.

After validating our data, we have plotted volcano plots which depict the differentially expressed genes. We found that there is only one gene which is significantly highly expressed in BD compared to SZ, between BD and healthy people, we have found 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 difference between the highly expressed genes in different areas of the brain in BD patients relative to healthy individuals. As the plots show, there are different genes which are highly expressed in different areas of the brain and can indicate the presence of BD (figure 2).

Those two findings show us very clearly that biomarkers for BD do exist and that they are not that rare or even hard to find. Unfortunately, the research literature does not support our claims, nor it disagree with it but rather it lacks any references regarding those genes in the context of BD. It is possible that this missing validation is caused due to insufficient research of the topic as we have mentioned before.

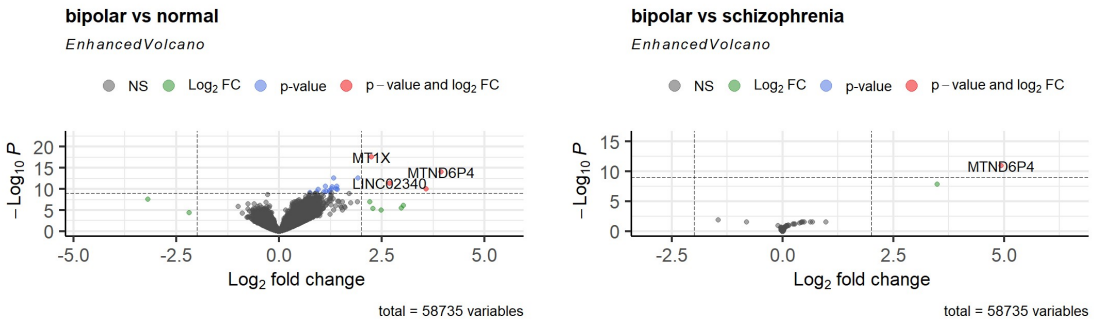
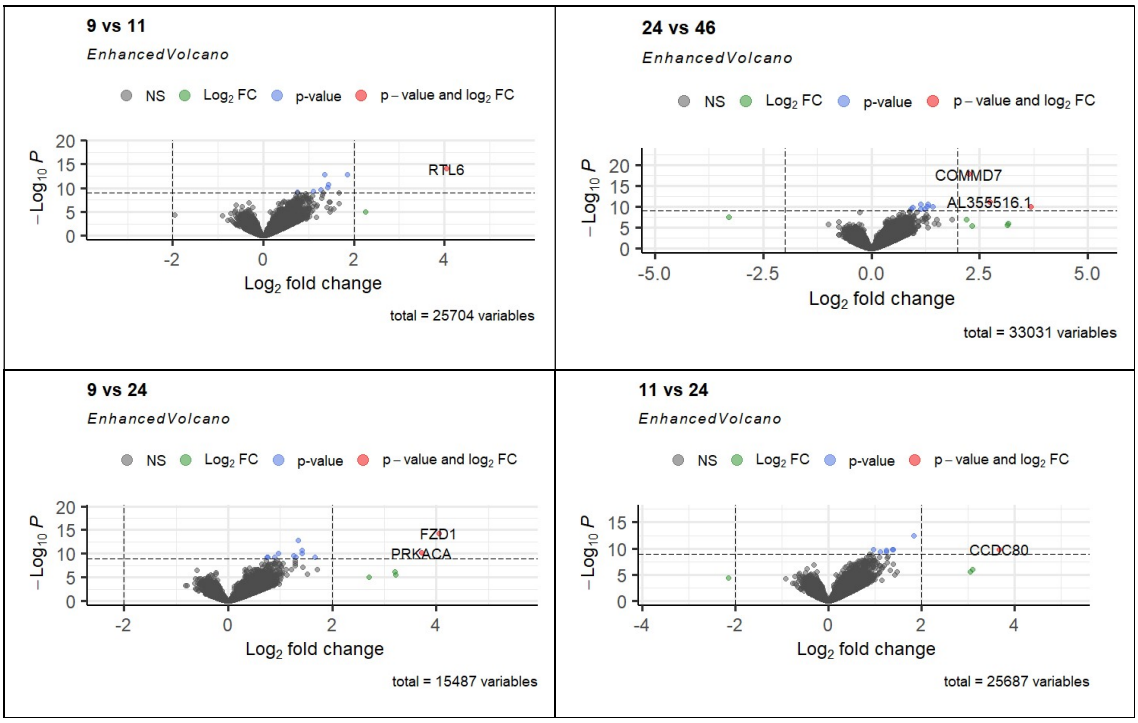


Figure 1: Two volcano plots that display the genes that were most differentially expressed in BD patients relative to healthy individuals (on the left) and SZ patients (on the right).



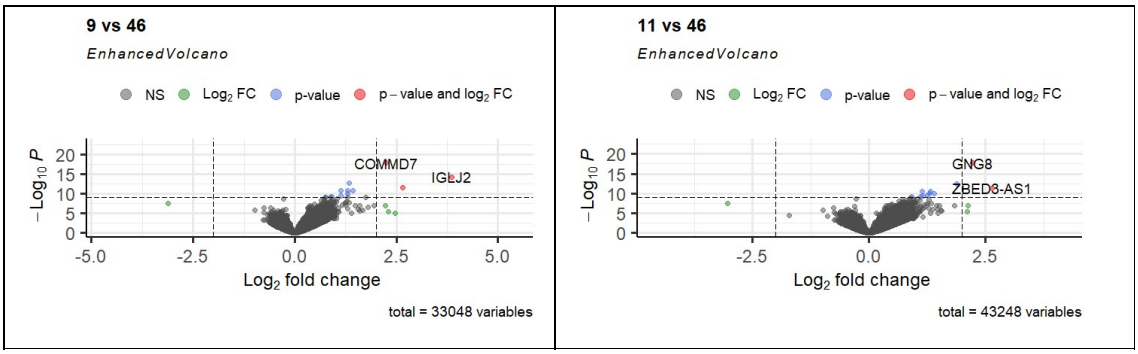
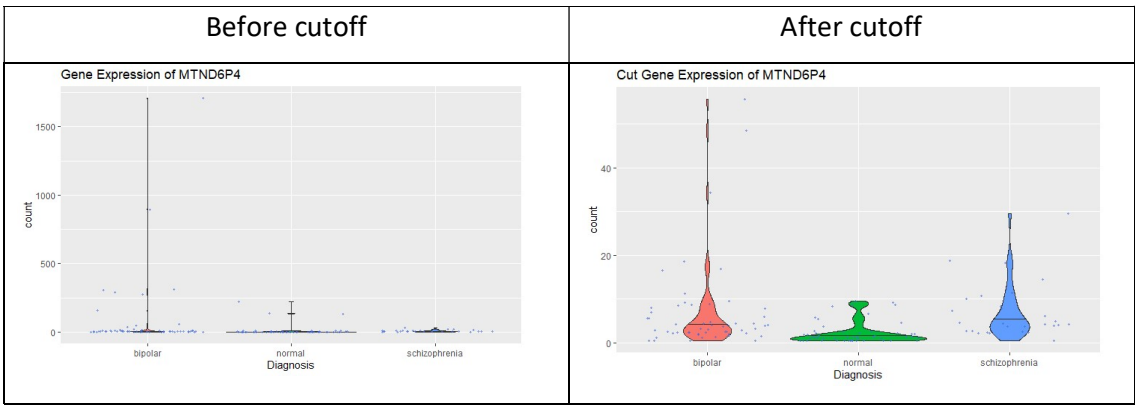


Figure 2: Six volcano plots that display the genes that were most differentially expressed in different areas of the brain relative to each other.

After we have completed the differential expression analysis, we aimed to further understand the four highly expressed genes that we have found in the first section of the analysis (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). The plots looked a bit odd because of a small number of samples which had extremely high counts, so we had to cut off the plots' tops in order to see them properly. As expected, the plots of BD and SZ were very similar to each other while very distinct from the healthy individuals (figure 3). It seems that the MTND6P4 is not actually differentially expressed between the BD and SZ patients but rather that the difference has occurred as a result of the rogue samples we have removed. 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.



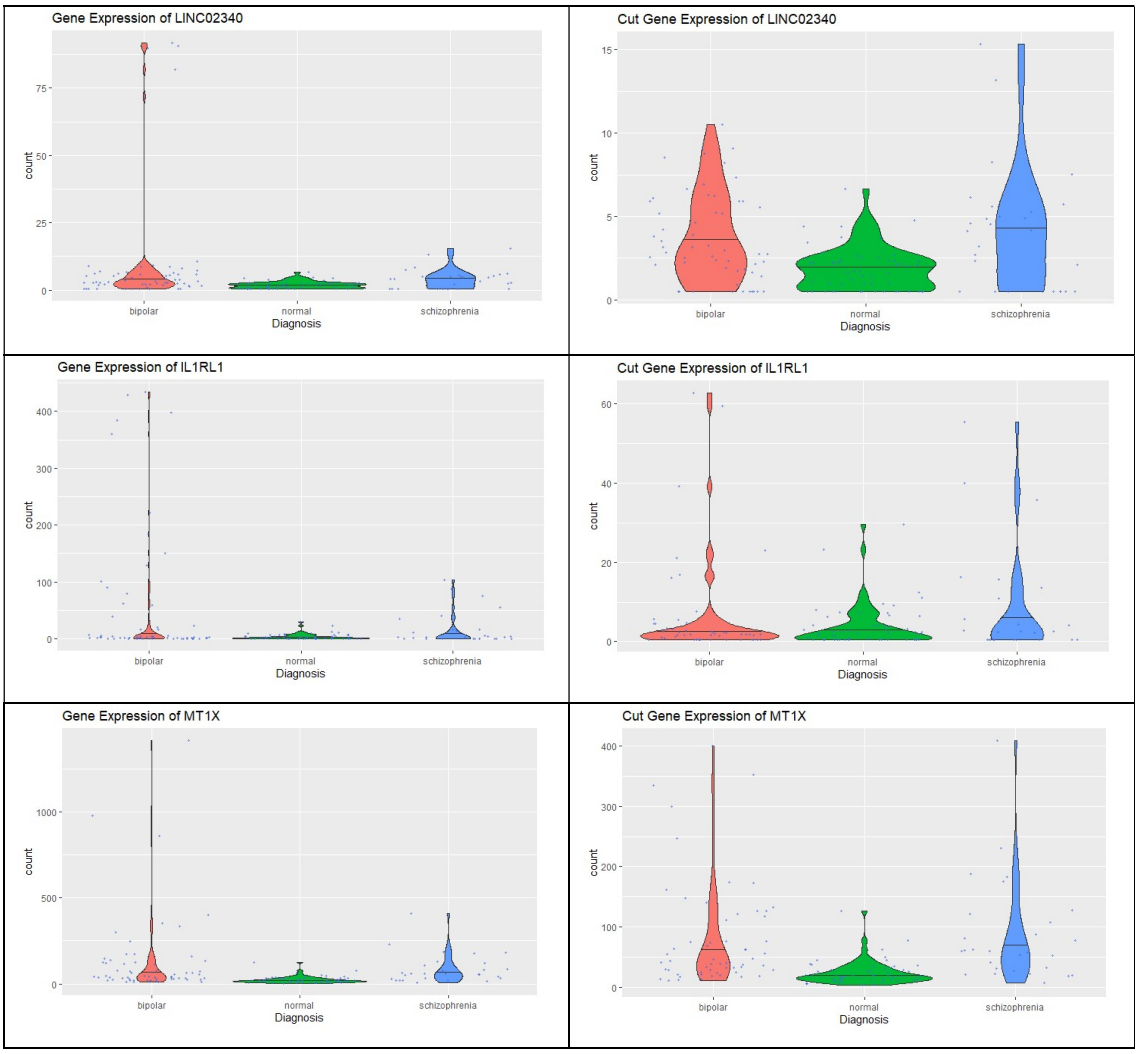


Figure 3: In each column there are four violin plots that correspond to the gene expression of four significant genes (top: MTND6P4, middle-top: LINC02340, middle-bottom: IL1RL1, bottom: MT1X).

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 4). The most significant pathways we have found were "HALLMARK COAGULATION" and "HALLMARK XENOBIOTIC METABOLISM". Looking at the research literature, 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.

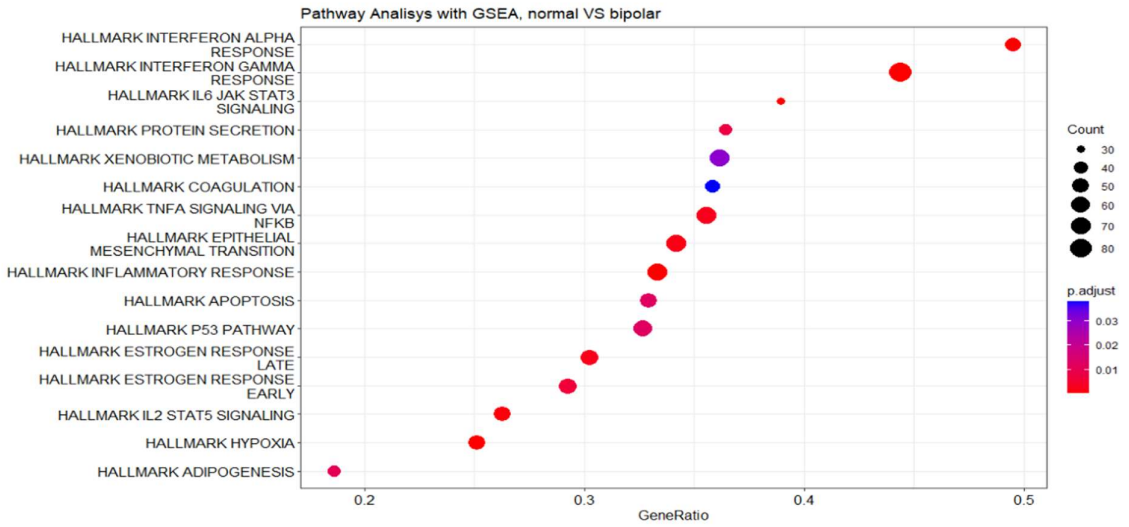


Figure 4: Enriched pathways in BD relative to the healthy control group.

Ignore for now:

The brain's Dorsolateral prefrontal cortex is equivalent the 9th and 46th areas of the brain according to the Brodmann areas system (Horn and Leigh, 2011).

Discussion

What conclusions have you drawn from the analysis? Do they provide any insight into the biological question?

Mention the limitations of your analysis.

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?

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.

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.

Horn, A.K.E., and Leigh, R.J. (2011). Chapter 2 - The anatomy and physiology of the ocular motor system. In Handbook of Clinical Neurology, C. Kennard, and R.J. Leigh, eds. (Elsevier), pp. 21–69.

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.