

IDENTIFICATION OF RNA-BASED BIOMARKERS ASSOCIATED WITH MANIC EPISODES AND LITHIUM RESPONSE IN BIPOLAR DISORDER

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

By

Noor Niaz

December, 2022

IDENTIFICATION OF RNA-BASED BIOMARKERS ASSOCIATED WITH MANIC EPISODES AND LITHIUM RESPONSE IN BIPOLAR DISORDER

By Noor Niaz

2022

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science:



Ali Osmay Güre (Advisor)

Özlen Konu Karakayalı

Murat İlhan Atagün

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

IDENTIFICATION OF RNA-BASED BIOMARKERS ASSOCIATED WITH MANIC EPISODES AND LITHIUM RESPONSE IN BIPOLAR DISORDER

NOOR NIAZ

M.Sc. in Molecular Biology and Genetics

Advisor: Ali Osmay GÜRE

December 2022

Bipolar disorder (BD) is one of the major mood disorders. A person afflicted with this neuropsychiatric disease undergoes episodes of depression and mania. BD ranks as the highest amongst all the mood disorders for having the most negative affect on a person's life. It is a lifelong illness, which requires constant monitoring and medication. From psychotic behaviours to suicide ideation, this disease is a burden on the afflicted and their families. Although, BD has been shown to have a high heritability factor, none of the studies done so far have been able to identify any biomarkers that could be causative for this disease. It is due to the fact that this disease is multifactorial, in which environment is a major contributing factor. Therefore, we believe that a more 'pattern-seeking' approach would lead to novel findings.

We performed gene-set enrichment analysis (GSEA) with BD datasets that consisted of 3 different biological phenotypes; euthymic BD, manic BD and lithium treated BD samples. Since a long time, lithium has been used as 'the leading' drug to treat BD because it functions as a mood stabilizer. Therefore, we also included BD cohorts that had undergone lithium treatment in our analysis. As a result of GSEA, we were able to discover 2 novel patterns. In pattern 1, specific genes were found to up-regulated in euthymic BD and lithium treated cohorts, while the same genes were down-regulated in manic BD cohorts. In pattern 2, the opposite trend was observed, that is; another list of specific genes was down-regulated in euthymic BD and lithium treated cohorts, while they were found to be up-regulated in manic BD.

The novelty of utilizing GSEA in our analysis was in the fact that we created our own custom gene-sets. The gene-sets were formed after performing differential gene expression (DEG) analysis on all the 3 types of BD datasets. The advantage of using custom gene-sets was that these were genes representing differential expression in different BD phenotypes, therefore, they were biologically relevant to BD. In the publically available curation of gene-sets on various databases, very few gene-sets represent BD. Hence, the custom gene-sets are more relevant and specific to the disease. Then we proceeded to extract core-enriched genes and performed further analysis, such as, plotting fold-change graphs and performing non-parametric tests. In light of these results, we propose putative biomarkers associated with manic episodes and lithium response in bipolar disorder.

We put forth the hypothesis that these patterns can diagnose BD accurately, indicate if a patient

is responding to lithium or not, and predict an oncoming manic episode. Therefore, we propose the genes adhering to the discovery patterns as putative biomarkers.

Simultaneously, we performed biological analyses and literature review with the list of putative biomarkers. We found out inflammation as a potential underlying cause of pathogenicity in BD. We hypothesize that the process of inflammation is disrupted in BD patients, especially between the different mood states and that in order to treat this disease, inflammation as a pathway should also be targeted. We highlight $TNF\alpha$ as one of the main cytokines that 3 of our biomarkers; *ADAMTS9*, *IL-1B* and *STC1* are associated through various pathways, and this disruption of the inflammatory pathway may occur due to alteration in $TNF\alpha$ and biomarkers' levels.

We hope that after further, subsequent research, the inclusion of our biomarkers in the clinics will help with the issue of BD misdiagnosis, help save precious treatment time, help with the choice of medication, and help indicate a patient's mood state. Overall, it will help the clinicians to tailor the treatment to every patient's specific profile, making the it easier to design a more personalized treatment strategy.

Keywords: Bipolar disorder, lithium, mania, inflammation, biomarkers, GSEA

ÖZET

BİPOLAR BOZUKLUKTA MANİK ATAK VE LİTYUM TEPKİYLE İLİŞKİLİ RNA TABANLI BİYOMARKERLERİN TANIMLANMASI

NOOR NIAZ

Yüksek Lisans, Moleküler Biyoloji ve Genetik

Tez Danışmanı: Ali Osmay GÜRE

Aralık 2022

Bipolar bozukluk (BB), başlıca ruhsal bozukluklarından biridir. Bipolar bozukluk olan nöropsikiyatrik hastalığa yakalanmış kişiler, depresyon ve mani dönemleri geçirirler. BB, bireylerin hayatlarını anlamlı oranda olumsuz etkileyen ruhsal bozuklukları arasında en üst sıralarda yer almaktadır. Sürekli takip ve ilaç tedavisi gerektiren, ömür boyu süren bir hastalıktır. Psikotik davranışlardan intihar düşüncesine kadar değişik spektrumlarda görülen bu hastalık, hasta ve aileleri için bir yükür. Bipolar bozukluğun multifaktöriyel kökenli olduđu bilinmesine rağmen, şimdiye kadar yapılan çalışmaların hiçbirisi bu hastalığa neden olabilecek herhangi bir biyobelirteç belirleyememiştir. Bunun nedeni, bu hastalığın multifaktöriyel olmasıdır, hangi çevrede önemli bir katkıda bulunan faktördür. Bu nedenle, daha 'örnek arayan' bir yaklaşımın yeni bulgulara yol açacağına inanıyoruz.

Çalışmamızda 3 farklı biyolojik fenotip olan ; ötimik BB, manik BB ve lityum ile tedavi edilen BB örneklerini içeren BB veri setlerinden oluşan gen seti zenginleştirme analizi (GSZA) uyguladık. Uzun zamandan beri lityum, ruh hali dengeleyici işlevi gördüğü için BB tedavisinde "lider" ilaç olarak kullanılmaktadır. Bu nedenle, yapılan analizlere lityum tedavisi görmüş BB kohortları dahil edilmiştir. GSZA sonucunda 2 yeni model saptanmıştır. Model 1'de, spesifik genlerin ötimik BB ve lityumla tedavi edilen kohortlarda yukarı regüle edildiğı, aynı genlerin manik BB kohortlarında ise aşağı regüle edildiğı gözlenmiştir. Model 2'de ise tam tersi bir eğilim gözlenmiş yani; belirli genlerin başka bir listesi, ötimik BB'de ve lityumla tedavi edilen kohortlarda aşağı regüle edilirken, manik BB'de yukarı regüle edildiğı bulunmuştur.

Çalışmamızın getirdiğı yeniliklerden birisi genetik seti zenginleştirme analizlerinden yararlanılarak kendi özel gen setlerimizi geliştirmemizdir. Gen setleri, tüm 3 tip BB veri seti üzerinde diferansiyel gen ekspresyonu (DGE) analizi yapıldıktan sonra oluşturulmuştur. Özel gen setleri kullanmanın avantajı, bunların farklı BB fenotiplerinde diferansiyel ekspresyonu temsil eden genler olması ve bu nedenle biyolojik olarak BB ile ilgili olmalarıdır. Çeşitli veri tabanlarında, toplumda görülen gen setlerini hakkında çeşitli bilgiler olmakla birlikte, BB hakkında çok az sayıda gen seti bilgisi vardır. Bu nedenle, özel gen setleri hastalık hakkında bilgi edinmede diğer yöntemlere göre daha uygun ve spesifiktir. Çalışmamızda çekirdekle zenginleştirilmiş genleri ayırıp, ileri aşamada kat değişim grafiklerini çizmek ve parametrik

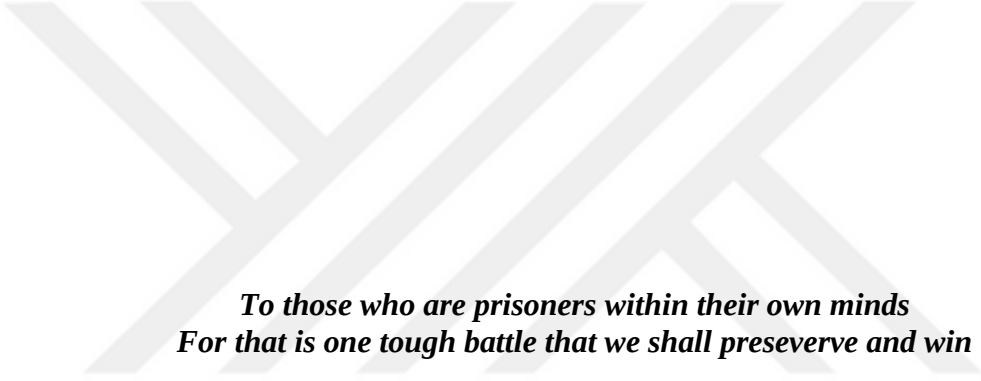
olmayan testler gibi testler yaptık. Bu sonuçların ışığında, bipolar bozuklukta manik epizotlar ve lityum yanıtı ile ilişkili varsayılan biyobelirteçler olduğunu savunuyoruz.

Bu kalıpların, bipolar bozukluğu doğru bir şekilde teşhis edebileceği, bir hastanın lityuma yanıt verip vermediğini gösterebileceği ve yaklaşmakta olan bir manik atağı tahmin edebileceği hipotezini ortaya koyduk. Bu nedenle, tanımlayıcı paternle bağlantılı genlerin biyobelirteçler olarak varsayılabileceğini öneriyoruz.

Çalışmamızda aynı zamanda biyolojik analizler yapıp, varsayılan biyobelirteçler konusunda literatür taraması yaptık. İnflamasyonun, BB'da patojenitenin altında yatan potansiyel bir neden olduğunu bulduk. BB hastalarında, özellikle farklı ruh halleri arasında inflamasyon mekanizmasının bozulduğunu ve bu hastalığı tedavi etmek için inflamasyon sisteminin de hedeflenmesi gerektiğini varsayıyoruz. Ana sitokinlerden biri olan TNF α ; ADAMTS9, IL-1B ve STC1'den oluşan 3 farklı biyobelirteçimizle farklı ara yollarla bağlantılı olduğu ve enflamatuar yollarda herhangi bir bozulmanın, TNF α ve biyobelirteç seviyelerindeki değişikliklerden dolayı olabileceğinin altını çiziyoruz.

Klinik uygulamalarda biyobelirteçleri içeren daha ileri çalışmaların ve detaylı araştırmaların yapılmasının hatalı tedavi sorunu önlemeye, tedavi sürecini kısaltmaya, hastanın ruhsal durumunu tanımlamada ve ilaç seçiminde yardımcı olmasını umuyoruz. Genel olarak, klinisyenlerin tedaviyi her hastanın özel profiline göre uyarlamasına yardımcı olacak ve daha kişiselleştirilmiş bir tedavi stratejisi tasarlamayı kolaylaştıracaktır.

Anahtar Kelimeler: Bipolar Bozukluk, Lityum, Manik Enflamasyon, Biyobelirteç, GSZA



*To those who are prisoners within their own minds
For that is one tough battle that we shall persevere and win*

ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my supervisor, Dr. Ali Osmay Güre, under whose mentorship my passion for scientific research and critical thinking flared up even more. Ali hoca's brilliance and good nature is evident through the admiration people have for him, not just of the AOG lab members but of people from all spheres of life. I am thankful to him for teaching me patience, encouraging me to think outside-the-box, and for teaching me by example; how to show grace in times of crisis.

I would like to thank members of my jury committee, Dr. Özlen Konu and Dr. Murat İlhan Atagün. Dr. Konu, for her meticulous feedback and invaluable suggestions, which helped to further solidify my research. Moreover, I would like to thank her for being an excellent instructor throughout my undergraduate and masters student years. Dr. Atagün, for his encouragement and appraisal of my work. As a young researcher, hearing his words cemented my belief that my work can make a difference in furthering research that will make the world we live in a better place.

I would like to acknowledge Berke Karaahmet as his analysis, supervised by Ali hoca, laid the foundation for this thesis work.

I would like to thank Mehdi Ghasemi for giving me the opportunity to work in the AOG lab in my 4th semester as an undergraduate student. I realized very early that research can be hard and frustrating, and yet wonderfully fulfilling. I would like to thank Ege Dedeoğlu for his mentorship during my senior year internship, and all the rest of the AOG lab members for always being kind and helpful.

As a foreign student, it is sometimes a hard toil to adjust without one's family. I found that to be the case for one semester until I met Ronak Naemae. She is my best friend from undergraduate years, who later became my fellow AOG lab member. I am grateful beyond words for her to-the-point, no-nonsense, fierce-loving presence in my life. She never failed to be there for me in any situation albeit academics, co-curricular, day-to-day life and personal matters. I am and will remain forever grateful to Ronak and her family for being my family away from home. In the same vein, I would like to thank Osama Jawad for his friendship. His belief that I am more than apt and suited for any task that comes my way. Ronak and Osama made my life more colorful and memorable in Turkey. From our study-sessions, presentation preps, proof-readings to food-related adventures, intercity trips, quiz competitions to gift-buying, and to conversations about politics and music and books and movies. I am blessed to have found such amazing friends.

I would like to thank Isli Çela and Shila Azizolli, the dynamic Albanian-duo. Isli, being the voice of reason with her soft love and Shila being the voice of reason with her tough love. I am very grateful for having met them both and to have forged this lasting sisterhood. I thank them for their support in times of crisis from the most mundane to the most insane. I will cherish our yemekhane breaks and our escapes from the lab to get fresh air at our favorite spots around the campus forever.

Marzana, I thank her for her fire-cracker energy and for our long walks at night around the campus, especially during the pandemic. Those were the energy boosters I did not know I needed at the time. Ushna, for her constant reminders from back home of our future plans, that we are making them come true and we can do it all.

My achievements today are the result of the sacrifices made by my parents and the prayers of my family. My mother, Ayesha Ahmad, for her relentless positivity that was my driving force. My brothers, Arsalan, Subhaan and Momin for their belief in me that I can climb mountains higher than Mt. Everest. My father, Alauddin Ahmad, for his silent prayers and wishes, for which long conversations were not needed. My extended family for always having high expectations from me that made me go the extra mile.

Last but not the least, I could not have dealt with the stress and pressure of the thesis writing and presenting process had it not been for the immense support of my significant other, Erdem Savaşcı. I am extremely grateful for his late night stays in winters just to be beside me while I worked, going above and beyond to be there for me in stressful times. His words of encouragement and motivation were the calmness I needed, and his love and belief in me were the drive that got me through. This achievement is ours.

I am grateful to Allah for all that He has blessed me with for I myself could not have known to wish for all the good I have in my life.

TABLE OF CONTENTS

1.	INTRODUCTION	1
1.1.	BIPOLAR DISORDER.....	1
1.1.1.	THE MOOD STATES	1
1.1.2.	TYPES OF BD.....	3
1.2.	DIFFICULTY IN DIAGNOSING BD	4
1.3.	PHARMACEUTICAL TREATMENT OPTIONS FOR BD	5
1.3.1.	MEDICINES IN CURRENT USE	5
1.3.2.	SHORTCOMINGS IN CURRENT MEDICAL TREATMENTS	5
1.3.3.	LITHIUM AT THE CENTRE STAGE FOR TREATING BD	6
1.3.3.1	PHARMACOLOGICAL MECHANISMS OF LITHIUM ..	7
1.3.3.2	LITHIUM'S EFFICACY IN BD	8
1.4.	GENETIC BASIS FOR BD	8
1.5.	RATIONALIZATION BEHIND OUR APPROACH; GSEA	9
1.6.	SUMMARY OF RESULTS	10
2.	AIM	13
3.	METHODS	14
3.1.	PRELIMINARY ANALYSIS: DONE PRIOR TO THIS STUDY	14
3.1.1.	ACQUIRING DATASETS	14
3.1.2.	FIRST APPROACH: DGE ANALYSIS.....	15
3.1.3.	SECOND APPROACH: GSEA	15
3.2.	VALIDATION ANALYSIS: DONE IN THIS STUDY.....	19
3.2.1.	ACQUIRING DATASETS	19
3.2.2.	NORMALIZATION OF DATASETS	21
3.2.3.	DGE ANALYSIS	21
3.2.4.	CREATING CUSTOM GENE-SETS	22
3.2.5.	GSEA PERFORMED WITH CUSTOM GENE-SETS	24
3.2.6.	CORE ENRICHMENT ANALYSIS	24
3.3.	FINAL LIST OF GENES	27
3.3.1.	FC VALUES OF CORE ENRICHED GENES.....	27
3.4.	NETWORK ANALYSIS AND PATHWAY ANALYSIS	28
4.	RESULTS	29
4.1.	PRELIMINARY ANALYSIS	29
4.1.1.	GSEA RESULTS FROM PRELIMINARY ANALYSIS.....	29

4.2.	VALIDATION ANALYSIS	30
4.2.1.	DATASETS USED FOR ANALYSIS.....	31
4.2.2.	GSEA RESULTS.....	32
4.2.3.	EXTRACTION OF CORE ENRICHED GENES.....	36
4.2.4.	VALIDATING THE DISCOVERY PATTERNS WITH CORE ENRICHED GENES.....	42
4.3.	SELECTION OF PUTATIVE BIOMARKERS	47
4.4.	FUNCTIONAL ANALYSIS.....	53
4.4.1.	NETWORK ANALYSIS.....	53
4.4.2.	PATHWAY ENRICHMENT	55
5.	DISCUSSION AND CONCLUSION	58
5.1.	BIPOLAR DISORDER: A HETEROGENEOUS DISEASES.....	58
5.1.1.	HERITABILITY IN BD	58
5.1.2.	INTERPLAY OF GENETICS AND ENVIRONMENT IN BD.....	58
5.2.	FINDING PATTERNS THROUGH GSEA	59
5.2.1.	MULTI-GENE ANALYSIS APPROACH.....	59
5.2.2.	DISCOVERY OF PATTERNS	59
5.3.	RESULTS OF GSEA.....	60
5.3.1.	THE TWO PATTERNS	60
5.3.2.	FINDING BIOMARKERS.....	60
5.4.	BIOLOGICAL ANALYSIS WITH PATTERN 1 GENES.....	61
5.4.1.	NETWORK ANALYSIS.....	61
5.4.2.	PATHWAY ANALYSIS.....	61
5.5.	LITERATURE RESEARCH ON PUTATIVE BIOMARKERS	62
5.6.	INFLAMMATION IN BD	65
5.6.1.	PUTATIVE BIOMARKERS IN INFLAMMATION	65
5.6.2.	TNFA: THE COMMON CYTOKINE	67
5.7.	CONCLUSION & FUTURE PERSPECTIVES.....	68
6.	BIBLIOGRAPHY	71
7.	APPENDIX.....	78
7.1.	APPENDIX A.....	78
7.2.	APPENDIX B	78
7.3.	APPENDIX C	91
7.4.	APPENDIX D.....	103
7.5.	APPENDIX E	104
7.6.	APPENDIX F.....	106

7.7.	APPENDIX G.....	108
7.8.	APPENDIX H.....	111
7.9.	APPENDIX I	112
7.10.	APPENDIX J	120
7.11.	APPENDIX K.....	121
7.12.	APPENDIX L	123
7.13.	APPENDIX M	124
7.14.	APPENDIX N.....	125
7.15.	APPENDIX O.....	125
7.16.	APPENDIX P.....	126
7.17.	APPENDIX Q.....	127
7.18.	APPENDIX R.....	128
7.19.	APPENDIX S.....	130

LIST OF FIGURES

Figure 1.1: The cyclic nature of BD.....	3
Figure 1.2: The original hypothesis	11
Figure 1.3: The new hypothesis generated based on the findings.....	12
Figure 3.1: Preliminary analysis – methodology to obtain custom gene-sets	18
Figure 3.2: GSEA schematic - heatmap and gene-set	25
Figure 3.3: GSEA schematic – Core enriched genes	26
Figure 4.1: Pattern 1 - Enrichment plot of euthymic BD dataset with lithium treated gene-set ...	37
Figure 4.2: Pattern 1 - Enrichment plot of mania dataset with euthymic BD gene-set.....	38
Figure 4.3: Pattern 2 - Enrichment plot of lithium treated dataset with euthymic BD gene-set ...	39
Figure 4.4: Pattern 2 - Enrichment plot of mania dataset with euthymic BD gene-set.....	40
Figure 4.5: Pattern 1 genes correlation plot	46
Figure 4.6: Pattern 2 genes correlation plot	46
Figure 4.7: ENTHD1 and DSC2 log2FCs – direction of expression plots	48
Figure 4.8: COL1A2 and DNAJC7 log2FCs – direction of expression plots.....	49
Figure 4.9: Non-parametric test for P1 genes.....	51
Figure 4.10: Non-parametric test for P2 genes	52
Figure 4.11: Network analysis with pattern 1 genes	54
Figure 4.12: Gene-ontology (GO) pathway analysis	56
Figure 4.13: KEGG and Reactome pathway analysis	57
Figure 5.1.: A simplified illustration of the 3 biomarkers' involvement in inflammation.....	67

LIST OF TABLES

Table 3.1: Preliminary analysis - GSEA advance settings	17
Table 3.2: Datasets used in validation analysis.....	20
Table 3.3.1.: DGE analysis to obtain up-regulated gene-sets.....	22
Table 3.3.2: DGE analysis to obtain down-regulated gene-sets.....	23
Table 3.4: Validation analysis - GSEA advance settings	24
Table 4.1: Preliminary analysis - summarized version of GSEA results with up-regulated gene-sets	29
Table 4.2: Preliminary analysis - summarized version of GSEA results with down-regulated gene-sets.....	30
Table 4.3: Validation analysis' datasets summary.	31
Table 4.4.1: GSEA results with Up-regulated gene-sets.....	33
Table 4.4.2: GSEA results with Down-regulated gene-sets	35
Table 4.5: Extraction of core enriched genes	41
Table 4.6: Pattern 1: Up BD-Li; Down mania list of final genes	43
Table 4.7: Pattern 2: Down BD-Li; Up mania list of final genes.....	45
Table 4.8: Genes that adhered to pattern 1 in more than or equal to 85% of the datasets.....	47
Table 4.9: Genes that adhered to pattern 2 in more than or equal to 75% of the datasets.	47
Table 4.10: Quantitative representation of network analysis with pattern 1 genes.	55
Table 5.1: The putative biomarkers' functions as defined in GeneCards (an online database). ...	62
Table 5.2: Anti-inflammatory drugs for BD treatment	69

CHAPTER 1

1. Introduction

1.1. Bipolar Disorder

Bipolar disorder (BD) is a psychiatric illness characterized by experiencing two types of opposing moods interspersed by a state of “normalcy”, also known as a euthymic state. These opposing moods cycle between two extremes, mania and depression. The spectrum on which these mood states lie is extensive, and these moods and their intensities vary greatly in different patients afflicted with this disease. These are episodic in nature with the durations spanning hours to weeks to months [1, 2, 3]. The extreme highs and lows in the mood that signify mania and depression, respectively, can be very deteriorating to the person’s life. Their social life to work life to personal relationships can be adversely affected [4]. In the very dark moments of suffering from a depressive episode, the patient can even go through suicide ideation, with 15% to 17% of the patients actually committing it [5]. According to a report generated in 2022, 46 million people worldwide have BD, with the prevalence being 1-4% globally [5, 6]. Moreover, the negative impact of BD out of all the other mood disorders is noted to be the highest, about 82.9% [5].

1.1.1. The Mood States

BD patients suffer from both the mood states but to different extents. It has generally been well recorded that for an average BD patient, the state of depression is a more constant or a frequent phase [7]. This is also the reason why BD often gets misdiagnosed with other mood disorders especially those with depressive symptomatology [8]. Mania, however, also exists as mentioned above, on a spectrum. It can present with severe psychotic symptoms or it can be of a milder version. We explain both states in greater detail below.

During a manic episode, especially when it is at its peak, it can lead to a psychotic breakdown that comprises delusions, hallucinations and aggression, this is known to occur in 75% of manic episodes [1]. This then requires the patient to be hospitalized and thus, this too negatively impacts their lives. Even a non-psychotic manic episode can be severe and detrimental to the mental stability of the afflicted individual as it comprises of sleeping disturbances, which is usually lack

of sleep, feeling of grandiosity, misplaced confidence, being a spendthrift, being susceptible to committing sexual transgressions, being more susceptible to substance abuse etc. [1, 2, 9]. All of these traits make a person's life all the more difficult to live and greatly affects their social and work life where such behavioral outputs can be seen as counterproductive and a threat to a safe working environment. Therefore, managing a manic episode becomes all the more important.

In comparison to a manic episode, a somewhat milder version of it has also now been outlined and added into many psychiatric evaluations as a separate mood state, it is called hypomania [6]. This is less severe than a full-blown manic episode. Hypomania causes surges in energy of the person. But the nature of these hypomanic episodes have a negative impact on the productivity of the individual as well because the person is likely to feel more irritable, agitated, and may also have slurred speech [10]. These impaired behaviors make it a problem not just for the afflicted but also for the people around them. All in all, these episodes make BD an extremely debilitating psychiatric disorder.

Depression, on the other hand as expected, is the opposite of the aforementioned behaviors and mind-states. It comprises of low energy state, overwhelming guilt, lack of self-worth, sleeping for prolonged periods of time, loss of appetite, and in severe instances, it can even cause a person to commit suicide [5, 7, 11].

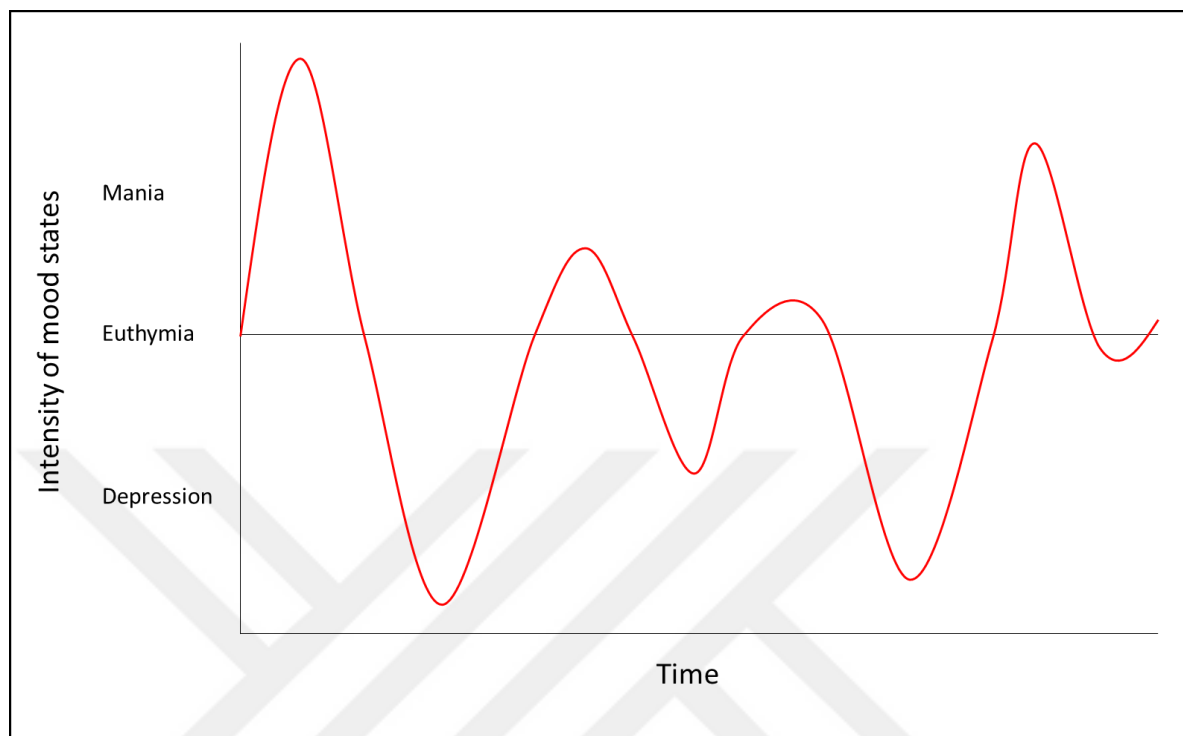


Figure 1.1: The cyclic nature of BD. This is a visualization of the mood spectrum on which BD exists. It ranges from mania to hypomania to depression to euthymic (the ‘normal’ mood-state). The intensity and duration of these moods vary from patient to patient [1].

1.1.2. Types of BD

These different mood states have led to the development of a few classifications of bipolar disorder into BD types I and II, and cyclothymic disorder. According to the *Diagnostic and Statistical Manual of Mental Disorders, 5th edition* (DSM-5), these fall under the category of BD with differences in the extent and intensity of episodes, including slightly different symptomatology [1, 2, 12, 13].

- **BD type I** comprises at least one episode of mania and longer depressive episodes. Mania may or may not be psychotic in nature, but the episodes vary between these two mood states.
- **BD type II** also comprises mania and depression but the severity is less and it does not inhibit the functioning of the patient as much as BD type I, especially their psychosocial interactions. For this reason, some psychiatric classifications define BD type II to include hypomania instead of mania.

- **Cyclothymic disorder** comprises shorter but frequent episodes of hypomania and depression. Again, the intensity of the mood swings is not as disruptive as BD type I.

1.2. Difficulty in Diagnosing BD

The lines drawn between the 3 categories as defined by DSM-5 are blurry and rely heavily on the clinicians' history taking abilities, and patients' narratives and/or their relatives' in cases of adolescents, without much astute biological testing. Moreover, since the depressive episodes for all 3 categories are more prevalent, BD is often misdiagnosed as major depressive disorder (MDD) [14]. Approximately, 70% of BD type 1 patients are misdiagnosed in the start, which then causes a delay in undertaking the correct treatment route for up to 5 to 10 years after the initial report [2]. When the age-of-onset is low, the probability of making the correct diagnosis gets even more difficult. Even though it's been reported that 2% of the world's population has at least been affected by BD types I and II [12]. The overlap of BD symptomatology with other affective disorders such as MDD, borderline personality disorder (BPD), attention deficit hyperactivity disorder (ADHD) and even schizophrenia is great [15, 16]. Therefore, this results in misdiagnosis, late treatment initiations and even adopting wrongful strategies of dealing with the disorder. The fact that clinical screening methods are not very reliable or specific in determining the course of BD adds to the difficulty in treating the disease. These diagnostic methods are heavily based on interviews and psychiatric evaluations such as questionnaires [17, 29]. There are many psychiatric committees and health organizations that have tried to make them more universal but the consensus is still lacking. This causes inconsistencies in diagnosis and screening. They also vary based on the interviewer's skills. Additionally, self-reports and/or relatives' reports have many biases that undermine the validity of the diagnoses. These include recall biases, lack of understanding of the symptoms, confusion while distinguishing between clinical depression versus feeling low etc. [18, 19, 27].

Such obstacles before the accurate conclusion can be made, not only delay treating the BD patient, it increases burden on the health systems, leads to cost-ineffectiveness and it can potentially increase the chances of treatment being ineffective in the long-term due to late intervention. Also, since it's a disorder of the mind, a patient's attitude and positivity as their situations deteriorate causes them to become even more resigned and unmotivated about adopting healthy lifestyles, which is essential in order to be on the road to recovery along with

taking medications [20].

Consequently, it is vital to define more concrete parameters to first identify BD in patients and then to form a meticulous treatment strategy. The ideal situation for clinicians would be to have a gene-signature that can predict or indicate the presence of the disorder but as has been the case for almost all affective disorders, including the extensively researched schizophrenia, identifying biomarkers is a hard task. Numerous genome wide association studies (GWAS) studies have identified some genes from a milieu as being somewhat predictive of BD. However, further research and validation studies failed to establish these putative candidates as clinical biomarkers [21]. Whether this is due to the fact that these genes have small effect sizes or the sampling processes have biases and cannot achieve statistical significance, it is a concern.

1.3. Pharmaceutical Treatment Options for BD

1.3.1. Medicines in Current Use

However, before going down the road of biomarker discovery, we also analyzed the current treatment options for BD patients. Aside from therapy sessions and psychiatric evaluations, we learned that medical treatment relies heavily on mood-stabilizers. For mania, mood-stabilizers such as lithium, valproic acid are commonplace, with lithium being the gold-standard. But some anticonvulsants such as valproate, carbamazepine, and lamotrigine are also used. Along with some antipsychotic medications; olanzapine, risperidone, aripiprazole and quetiapine, which have been shown to have the same efficacy as mood-stabilizers if not more, are also in clinical use [22, 23, 24].

For depression, the treatment is administration of antidepressants, selective serotonin-reuptake inhibitors (SSRIs) being one of the mainstays [11, 22]. Anti-inflammatory drugs such as N-acetylcysteine and polyunsaturated fatty acids have also been shown to have antidepressant properties. Natural compounds; omega-3 fatty acids have been one of the implicated substances that work to reduce BD symptoms without causing many side-effects [1, 25]. Although the studies about natural compounds efficacy in BD cannot be considered extensive as of yet.

1.3.2. Shortcomings in Current Medical Treatments

Even with these available medicinal treatments, there are patients that are non-responsive to them

or display severe side-effects [26]. Another hindrance is the age-of-onset, which if it's too early, the medications have to be prescribed with great care, according to the age and development of the individual as many medications are unsuitable at an early age. This is due to their potential negative impact on the development of the young patient [28, 29]. The nature of the disease itself requires careful monitoring of the dosage and administration of the drugs. The drugs require altering with changes in the mood states. Antidepressants have shown to cause a manic episode, also known as a 'mood-switch' if its dosage and uptake frequency is not monitored with the changes in the depressive phase [29]. Similarly, medications used to reduce the manic state can also trigger a depressive episode. This makes the treatment of BD that much more complicated and difficult. Thus, similar to other diseases, a 'one-fit-all' approach is regressive if not detrimental for the individual who is in the throes of bipolar disorder.

1.3.3. Lithium at the Center Stage for Treating BD

We will now put forth some literature review findings for the case of lithium as a 'treatment front' used in BD for treating both mania and depression. According to Malhi et.al, 1/3rd of BD patients respond positively to lithium [30]. Lithium is generally used as a maintenance therapy for BD, owing to BD being a mostly lifetime prevalent disease. Therefore, understanding lithium's efficacy is an important step. Although a lot of studies have been done in this regard, still little is known about its exact mode of action. The risk reduction capability of lithium has been calculated to be greater than other mood stabilizers, even more than the modern ones. Moreover, the cost of lithium compared to other drugs is 10 times less, also making it more affordable [31].

Before prescribing lithium, doctors have to ensure the blood levels of the patients for particular substances such as thyroid hormones, blood salts, creatinine proteins are within an optimal range. Moreover, clinicians have to monitor the functioning of kidneys especially when prescribing lithium to adolescents. As one of the side-effects of lithium is potentially causing kidneys to have reduced ability to concentrate the urine properly, with as much as 15% reduction in being able to do so [32]. Simultaneously, lithium blood levels should also be monitored constantly. The serum concentrations of lithium that are regarded as treatment effective in clinics range between 0.6-0.8 mmol/L [32]. These tests are undertaken to reduce any risks and susceptibility to side-effects. There are some studies where lithium intake during pregnancy has been linked to cardiac

malfunctioning of the fetus and therefore, it is not recommended for pregnant women [32]. Other side-effects of lithium include nausea, weight gain, cognitive problems and hypothyroidism [32]. Some of the symptoms are said to improve over long-term usage while others worsen. Thus, constant monitoring of lithium dosage and screening of symptoms becomes a must during medication.

1.3.3.1. Pharmacological Mechanisms of Lithium

The pharmacological mechanisms of lithium have been a point of debate and much deliberation for many years since it was first introduced as an agent to treat mania in 1949 [32]. Although still not clearly defined, there has been advancement in elucidating some pathways through which it mediates and seems to function as a mood stabilizer. As expected, more than one pathway has been associated with its function. This is not a surprise because as it has been observed that lithium not only mediates to lessen mania but also regulates depression and for some BD patients, it works as one of the best drugs preventing suicide [32]. The mitigation of depression seems to be potentially through lithium's action of increasing serotonin production. Conversely, it was also found to modify mania by decreasing dopaminergic activity [30, 33]. Another potential target of lithium is glutamate, which is an amino acid neurotransmitter that is excitatory in nature. It affects central nervous system (CNS) processes including modulation of plasticity, memory processing and learning. Another outcome of modulation of glutamate (potentially via GABA receptor pathway) is lithium's function of acting as a neuroprotective agent especially against apoptotic processes [30, 31]. One of the anti-apoptotic proteins, bcl-2, has been implicated to have had high expression as a result of lithium medication [31]. Similarly, other targets of lithium present in the literature include brain-derived neurotrophic factor (BDNF) and glycogen synthase kinase 3 (GSK3). While it increases the levels of BDNF in the brain, it inhibits the enzymatic activity of GSK3 [30, 33]. The former protects neurons against glutamate excitotoxicity while the latter action results in it being anti-manic but how exactly this works is still not known. It might be due to the fact that GSK3 is involved in neuronal survival, circadian clocks, inflammation and regulation of gene expression. GSK3's mode of action is proposed to be affected by lithium as lithium acts upon G-proteins in an inhibiting fashion [30, 34]. A group of studies have reported that lithium inhibits cAMP generation [31]. Another area of target of lithium is the circadian rhythm genes. It has been suggested that lithium tends to increase the expression of CRY1, PER1 and PER2 genes [35]. As these genes affect the sleep/wake cycles

and appetite, it is not surprising that this system should have been found to be affected in a mood disorder such as BD. However, most of these putative roles of lithium have been proposed via experiments done on animal models in the lab, further validation in human subjects is still lacking.

1.3.3.2. Lithium's Efficacy in BD

As for lithium's efficacy in BD patients, there is a lot of heterogeneity within different BD patients. It has been found that not all BD patients respond equally well to it, and these have been segregated into lithium responders versus non-responders. The approximate percentage of patients not successfully being treated by lithium, or they are straight out non-responsive is 30-40% [31]. A pattern observed is that a person with a family history of lithium responders, also responds well to the drug. Thus, there does seem to be a heritable and genetic basis for lithium's action and hence, the genetic basis for BD's etiology. GWAS studies have established some gene linkages and SNPs, which associate with response to lithium again reaffirming the heritability factor of BD [31]. Although many of the results of these studies failed to be reproduced or replicated in lab based settings. This might be owing to the small effect sizes of the genes discovered and/or sample heterogeneity during GWAS, which cannot always be translated to a smaller cohort. Indeed, it will be extremely helpful to have a way of determining a certain BD patient's responsiveness to lithium and thus, not wasting time and money. More importantly, not letting negative effects of lithium affect the patient, when they are non-responsive to it. So, if there is a way of knowing this, this should be assessed and highlighted.

1.4. Genetic Basis for BD

With pedigree studies at least, it is estimated that the heritability factor for BD ranges between 70 to 90% [1]. Especially, if it's the case of 1st degree relatives suffering from BD or other closely associated mood disorders. Similarly, the risk factor increases if there is parental early-onset BD then 49% the offspring become at risk for developing BD [36]. With twin studies, it has been shown that genetics is a strong factor and isolating the environmental aspect as the only way to prevent the occurrence of BD is futile [1, 37].

Along with getting these heritability estimates, another factor that supports molecular hypothesis is the neuroprogression of the disorder. Many MRIs and brain-scans indicate alteration of brain

structure in the nervous system of the afflicted individuals. Compromise of white matter and gray matter integrity being one of the major indicators of neurodegeneration [34]. Thus, neuroimaging can be another method of assessing disease progression. Therefore, it can be said neuroprogression supports the genetic basis for BD. Another factor that supports the progressive nature of BD are the psychological changes. This signifies that there is a need and maybe a way to intervene early to cure and not just manage the disease. Moreover, all of these studies indicate that biomarker discovery, although extremely complicated because of the nature of the disorder, can still be pursued. It was with this aim of biomarker discovery that we started working on this thesis.

1.5. Rationalization behind our Approach; GSEA

Gene-set enrichment analysis (GSEA) is a tool that utilizes a prior knowledge-based approach. While single-gene analysis such as differential gene expression (DGE) analysis is usually done without prior knowledge. For DGE analysis, biological phenotypes or classes are first defined; for example, sample sorting based on tests versus controls, then the rest of the analysis requires statistical tests and significance thresholds to ultimately narrow down the genes to one or a small list from the whole transcriptome [38]. This is then validated further. This approach is vital for discovering genes, especially if their effect sizes are large. Then it makes it easier to spot them out of thousands of genes making up a transcriptome. However, when it comes to heterogeneous diseases, especially those that constitute psychiatric illnesses, finding biomarkers gets very tricky with a single-gene analysis strategy. A multi-gene analysis approach such as GSEA can be a more substantial way to elucidate putative biomarkers or biological pathways for these diseases.

Moreover, while doing a meta-analysis, there is a myriad of data now available that evaluates the transcriptome using several technological methods such as different kinds of microarrays and RNA-Seq, the difficulty then also lies in making the data comparable to each other. This noise reduction while doing DGE can result in finding little to no significant result as multiple hypothesis testing aims to reduce the noise and batch effects, the tradeoff is losing what can be potentially important data. Another shortcoming of a single-gene analysis is that in most biological processes, it's a group of genes that regulate or are involved in these processes, and more than one pathway is causing the pathogenicity in a disease. Thus, this data can also be lost.

GSEA also utilizes DGE analysis as the initial step to distinguish the two phenotypes and sort

the gene expression profile (input file) based on this differential expression. However, unlike in the DGE analysis, it proceeds to determine whether the members (genes) in another list of genes, *the gene-set*, which is defined priory, correlates to the input file. A correlation coefficient called the enrichment score, ES, is used as a parameter to measure the enrichment of the gene-set with the input file. The ES reflects the degree to which members of a gene-set correlate to which phenotype as defined by the user. Then there are additional measures and significance parameters that are also performed to factor in the permutations, gene-set sizes etc. In our analysis, we used the false discovery rate, FDR, a multiple hypothesis correction value, to highlight the significant results [38].

GSEA software incorporates gene-sets curated from various public platforms such as Gene Ontology (GO), Reactome, KEGG etc. Moreover, it also incorporates different categories of gene-sets such as those belonging to biological pathways, molecular functions, oncogenic signatures, immunological pathways, drug responses etc. Therefore, with these thousands of defined gene-sets, analysis can be performed by the user with their own transcriptomic dataset to see whether any pathways, and ultimately genes, come out as significantly enriched.

The novelty of our analysis was that we utilized GSEA by creating our own *custom* gene-sets. We did not use the publicly defined gene-sets, which can be very general and also, in our case, since we had a neuropsychiatric disease, the availability of gene-sets relevant to such diseases were rare to non-existent. Therefore, as detailed in the methods section, we created our own custom gene-sets from the datasets available online that corresponded to BD cohorts with various additional characteristics. The three types of datasets we downloaded and ran on GSEA were:

1. Euthymic BD vs. healthy controls,
2. lithium-treated vs. non-treated and,
3. manic vs. euthymic BD.

1.6. Summary of Results

The methods and results of this study are divided into two parts and referred to as:

1. Preliminary analysis
2. Validation analysis

The results that laid the foundation for this project were obtained in the preliminary analysis.

This analysis led to the discovery of two surprising patterns; referred to as pattern 1 and pattern 2 throughout the thesis, that led us to a new hypothesis illustrated below. Subsequently, we used these results to validate the data in new datasets; the validation analysis.

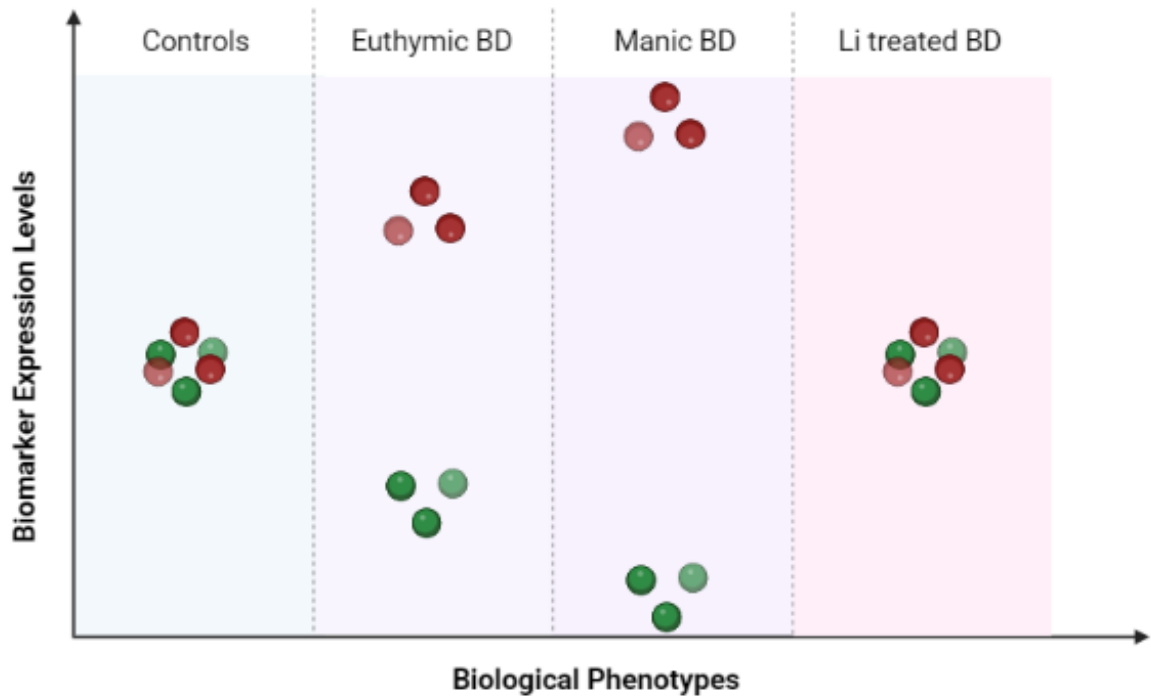


Figure 1.2: The original hypothesis. We expected the biomarkers' expression levels (depicted by red and green dots) to be differentially regulated in euthymic BD compared to other phenotypes. (Abbreviations: BD; Bipolar disorder).

In **Fig.1.2** above, the differential biomarker expression is shown by both increased and decreased levels depicted by red and green dots, respectively. However, if a patient suffered through a manic episode, we expected the discrepancy in biomarkers' expression to be even more pronounced *relative* to euthymic BD patients. Then as a result of administering lithium, we expected the biomarkers to go back to normal levels of expression, i.e., the same levels as in healthy controls. However, due to the findings of the preliminary analysis a new hypothesis was generated. This is illustrated in **Fig.1.3** below.

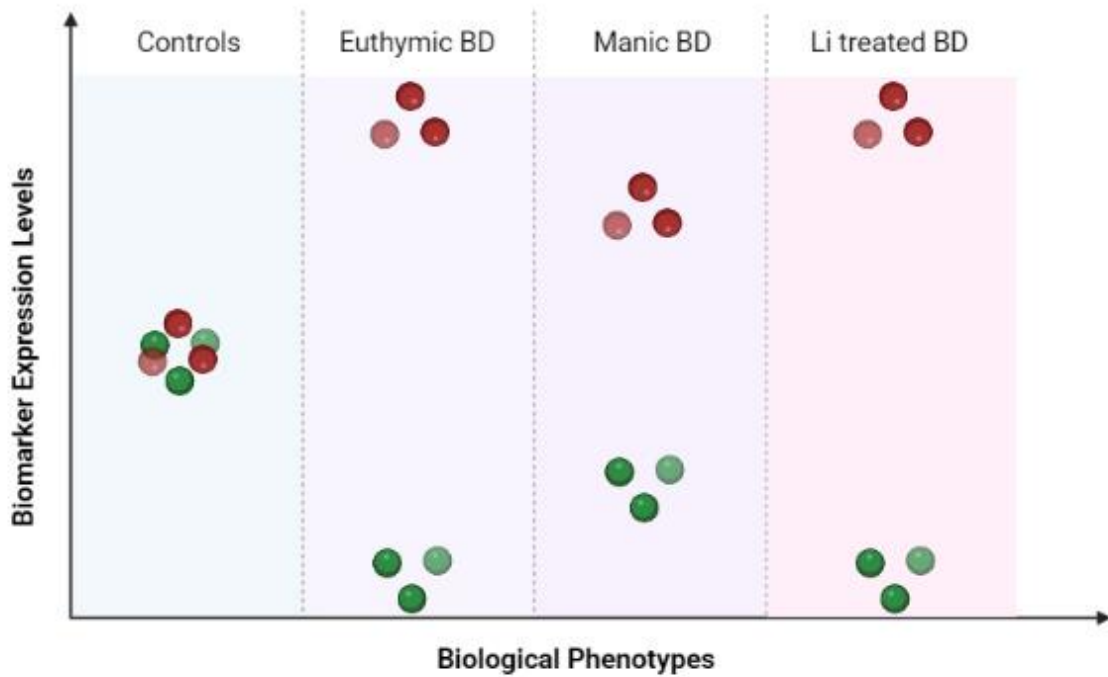


Figure 1.3: The new hypothesis generated based on the findings. This figure illustrates the two patterns of biomarker expressions between the different phenotypes. These are observed after performing GSEA with datasets in preliminary analysis. (Abbreviations: BD; Bipolar disorder).

- **Pattern 1: Up BD-Li; Down Mania:** As expected, certain biomarkers had differential expression between healthy controls and euthymic BD. In this figure, this is shown by elevated expression of the red dots in euthymic BD. In a patient suffering through a manic episode, we observed the *same* biomarkers (red dots) to show a decreased biomarker expression compared to euthymic BD. Contrary to our expectations, however, after administering lithium, the *same* biomarkers' expression was observed to be up-regulated, similar to euthymic BD.
- **Pattern 2: Down BD-Li; Up Mania:** The opposite of the aforementioned expression profile depicted by green dots that represent biomarkers.

The data analysis performed in validation analysis after observing these patterns led to the discovery of putative biomarkers.

CHAPTER 2

2. Aim

The purpose of this study is to identify biomarkers that can be representative of lithium response in bipolar patients, and to find biomarkers that can potentially be indicative of an oncoming manic episode.

- Lithium, as we have extensively discussed above, is the mainstay for BD treatments and there is a genetic factor associated with its efficacy. We chose lithium to see whether we can identify any molecular biomarkers that can be associated with its function.
- As for choosing manic BD patients, we believe that depression has been the most extensively studied mood state. Whereas, the biological understanding of mania, and not just its psychobehavioral attributes, is lacking. Moreover, it is a very debilitating state for a BD patient and we seek to identify any biomarkers associated with mania.
- Lithium is mostly used as an anti-manic drug. Therefore, our aim is to try to gain more insight about the dynamic of this drug with this mood state.

Fulfilling this aim will help diagnose BD in the clinics more astutely. As we have discussed before, misdiagnosis of BD is a huge hurdle when it comes to disease management and treatment. Therefore, determination of molecular biomarkers that can differentiate between the different phenotypes of a bipolar patient are the main goal.

CHAPTER 3

3. Methods

3.1. Preliminary Analysis: *Done prior to this study*

Prior to the work done in this project, the foundation of this research was laid down with results of data analysis done by a former Bilkent University student, Berke Karaahmet.

His analysis included the ‘discovery’ GSEA results that revealed the two surprising patterns of gene expression shown in **Fig.1.3** in the introduction section, which formed the new hypothesis. These results were obtained after performing extensive data analysis and *in-silico* work. The data presented in this thesis and the hypothesis that was formed is a continuation of his work.

Here we summarize the work done by Karaahmet, which we refer to as the preliminary analysis:

3.1.1. Acquiring Datasets

A manual key-word search was conducted on NCBI’s *GEO DataSets* database [40]. The key-words were designed to obtain bipolar disorder patients’ data, studies consisting of lithium-treated samples and studies that could differentiate between euthymic and manic states. A total of 10 datasets were retrieved and downloaded from the online database. These datasets belonged to 3 categories of biological phenotypes:

- BD patients vs healthy controls,
- lithium-treated vs non-treated and,
- manic vs euthymic states.

All the datasets included transcriptomic data obtained after performing mRNA-based microarray analysis. The sources of the mRNA transcripts in each dataset were of various kinds. For example, Dataset GSE35291 included hematopoietic stem cells (HSPCs) extracted from mice treated with lithium and non-treated HSPCs [118], GSE46449 included RNA purified from leukocytes of BD patients, manic patients and healthy controls [106], and GSE29417 included cells obtained from the hippocampus of mice models of mania and euthymia [121]. The details

of all the datasets used, including sample size, sample type and presence of replicates etc. are presented in **Appendix A**.

3.1.2. First Approach: DGE Analysis

The first approach taken with these datasets was to perform differential gene expression (DGE) analysis. The goal of which was to see whether any significant genes could be found that were common genes the different BD datasets or between different lithium-treated and BD datasets. And whether these could be potentially regarded as biomarkers after further *in-vivo* and *in-vitro* validations.

The GEO2R tool from the NCBI website was utilized, this tool uses ‘moderated t-statistic’ as a means of performing DGE analysis and p-values can be selected as per the user’s preference. The tool operates by applying GEOquery or limma R packages for the DGE analysis. For this analysis, multiple-hypothesis testing measure of Benjamini-Hochberg (p-value < 0.05) was used [93].

However, this single-gene analysis approach didn’t yield significant results even without the p-value correction.

3.1.3. Second Approach: GSEA

Gene-set enrichment analysis (GSEA) was performed with all the datasets.

Step 1. Preparation of GSEA input files

First and foremost, all the datasets were downloaded after normalizing them using GEO2R. Secondly, 4 types of input files were created to be uploaded on GSEA:

- i. Text files (.txt); these consist of normalized gene expression values, with genes as rows and samples as columns.
- ii. Chip annotation file (.chip); these consist of all the probesets or identifiers used by the platforms with their matched HGNC gene symbols.
- iii. Phenotype labels file (.cls); this file categorizes the samples into two phenotypic groups. For this study, these were based on the 3 categories of datasets mentioned above.

- iv. Gene-set file (.gmx); this consists of all the list of genes separated based on their different features. These are defined by a title, for example, a list of n number of genes, that are part of the 'Ca²⁺ signaling pathway'. This will constitute one specific gene-set.

In order to create the files with the correct formats, GSEA user guide was referred to [41].

One thing to be noted is that GSEA requires maximum two distinct phenotypes so that it can perform the analysis and find enrichment of the gene-sets with the expression dataset. The input gene expression file is then sorted by the software based on the differential gene expression between the two user-defined phenotypes. For our purposes, these were based on the aforementioned 3 categories, that is, BD vs healthy controls, li-treated vs non-treated and manic vs euthymic states.

Last step is to upload all files to see whether any significant enrichment occurs between any of the gene-sets and the gene expression profiles, which will then distinguish that particular set of genes as having some functional relevance to the disease (or to the test samples of the input dataset).

Step 2-A: GSEA performed with C5 GO gene-sets

The initial analysis was performed for all the datasets with the gene-sets from "c5.all...symbols.gmt" (the full name is missing as an older version of these gene-sets was used relative to the present). This is an internal feature of GSEA, which contains curated gene-sets that are publicly available. This is from the *The Molecular Signatures Database* (MSigDB), which is a library of gene-sets clustered based on specific biological features. From the open access database of Gene Ontology (GO), category C5 was specified, this is sub-categorized into BP (biological pathways), CC (Cellular Component) and MF (Molecular Function) [42]. All of these gene-sets correspond to specific pathways that range from signalling cascades to cancerous processes to extracellular rearrangements etc. More than 15000 gene-sets of various sizes are present in this database. After selecting for this database and changing the additional settings on the software according to the user guide and literature research, (the specifics are shown **Table 3.1**), GSEA was run.

Table 3.1: Preliminary analysis - GSEA advance settings. The parameters and thresholds set before running GSEA (most were the default settings).

Required fields	
Number of permutations	1000
Collapse/Remap to gene symbols	Collapse
*Permutation type	phenotype
Basic fields	
Enrichment statistic	weighted
Metric for ranking genes	Signal2Noise
Gene list sorting mode	real
Gene list ordering mode	descending
Max size: exclude larger sets	500
Min size: exclude smaller sets	10
Advanced fields	
Collapsing mode for probe sets => 1 gene	Max_probe
Normalization mode	meandiv
Randomization mode	no_balance
Omit features with no symbol match	TRUE
Median for class metrics	FALSE
Seed for permutation	timestamp
*If there are fewer than 7 samples in a phenotype/class use 'gene_set'. Otherwise, use 'phenotype'. Creates null distribution [41].	

This step was repeated for all the datasets and the results were extracted. However, although there were several pathways that were enriched, very few had the significant value of false discovery rate, $FDR \leq 0.25$. Moreover, pathways that could be found as common within the same categories of datasets were few to none. Therefore, the idea for making custom gene-sets from the different BD datasets came next.

Step 2-B: GSEA performed with *custom* gene-sets

This step is the novel approach taken in this study. This approach then laid the foundation for further analysis and a new hypothesis was created, after the discovery of the patterns shown in **Fig.1.3**.

Custom gene-sets were created. Contrary to curated gene-sets, available and defined based on biological processes in public databases, *custom* gene-sets were created, using the researched BD datasets in **Appendix A**. For this purpose, DGE analysis was performed again, this time less

stringent cut-offs were used as it was reasoned that during GSEA, more cut-offs will be applied, and significance measures will be used. The scheme is shown in **Fig.3.1**.

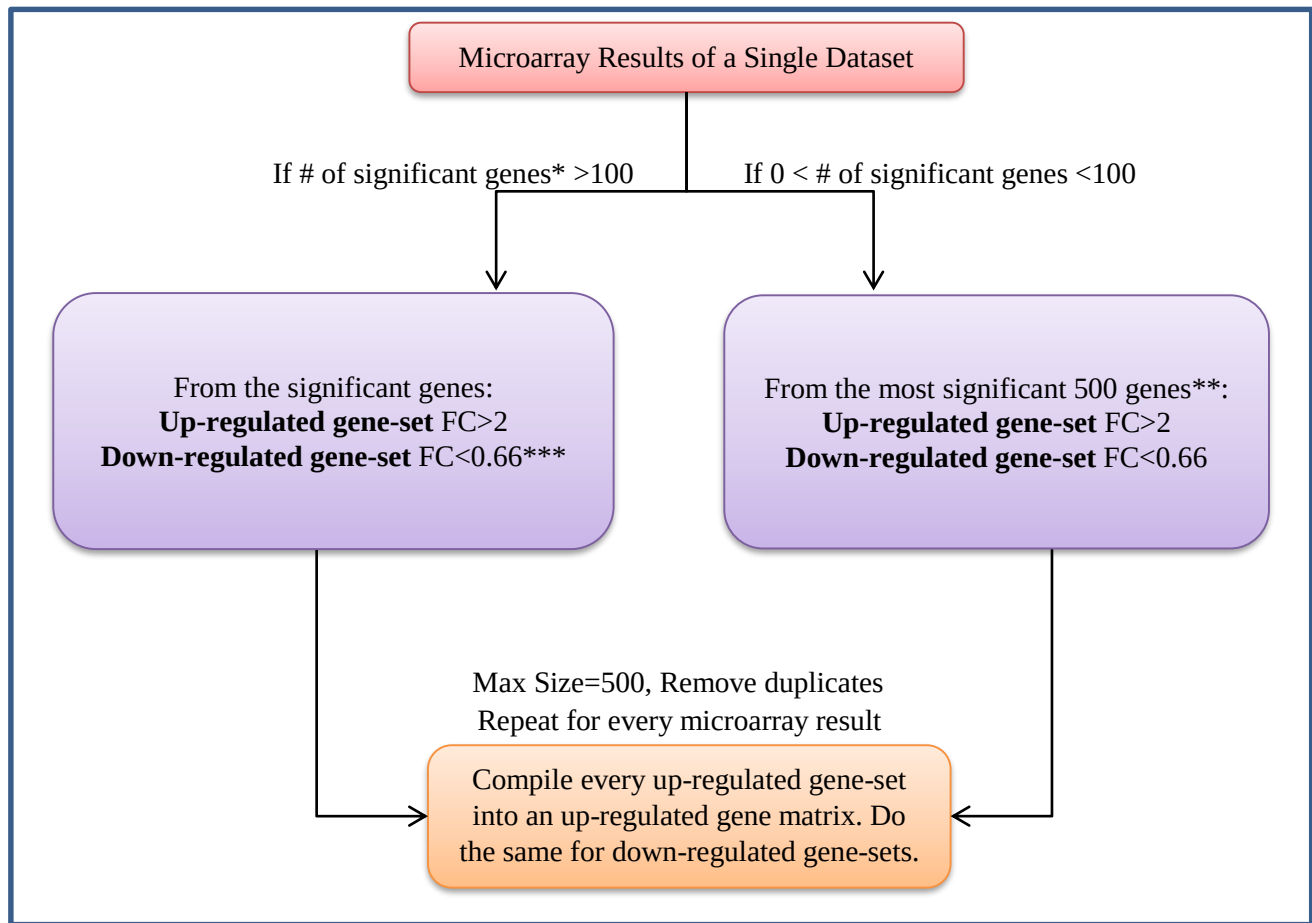


Figure 3.1: Preliminary analysis – methodology to obtain custom gene-sets. For every dataset, this schema was applied to obtain a gene-set. *Significance: $p < 0.05$ after Benjamini-Hochberg correction. **This number was specifically chosen to not lose too much significance. ***If we had used $FC < 0.5$ as criteria, we would have very small gene-sets which are not optimal for GSEA analysis.

Tables in **Appendices B and C** shows the 12 custom gene-sets that were obtained, 6 that were up-regulated in the indicated datasets and 6 that were down-regulated in the same datasets, respectively. Therefore, 2 gene matrix files were created consisting of up-regulated gene-sets and down-regulated gene-sets.

Then GSEA was run again with all the datasets one by one with both the gene matrix files. The same advance settings mentioned in **Table 3.1** were used. The results were obtained. These are

expressed in detail in tables in the **Appendices D and E**, GSEA results with up-regulated and down-regulated custom gene-sets, respectively. A summarized version is depicted in **Tables 4.1 and 4.2**.

3.2. Validation Analysis: *Done in this study*

The methodology was used to perform new analysis with the objective to see if the results of the preliminary analysis can be observed in new datasets with slightly different filters and cut-offs, and with new and more consistent types of datasets and statistical thresholds.

3.2.1. Acquiring Datasets

A key-words search using strings such as “bipolar disorder”, “bipolar disorder mania” and “bipolar disorder lithium” was done on NCBI’s *GEO DataSet* [39]. After evaluating the research summaries and looking at the methodology files, and at the contents of the published papers, we ended up with 24 datasets. Similar to the previous analysis, the categories of datasets consisted of 3 phenotypes:

- BD patients vs healthy controls,
- lithium-treated BD samples vs non-treated BD samples and,
- manic BD patients vs euthymic BD patients.

Two important distinctions of the new datasets from the previous ones were that, number one, these consisted samples from homo-sapiens only, so although the source of genetic material differed, for example, in some datasets fibroblasts were used to extract DNA and in others post-mortem brain cells were used, the species variable was constant. Number two, all the datasets were from bipolar disorder afflicted individuals, that is to say, unlike some datasets in the previous analysis in which HSPCs from mice were treated with lithium, these were not from a bipolar cohort, in this validation study, we made sure that in all 3 categories of datasets, we keep the disease state a constant as well. However, the controls of each category differed with the phenotype itself. The complete table with datasets information is presented in **Table 3.2**.

Table 3.2: Datasets used in validation analysis. These datasets were obtained from NCBI's *GEO DataSets* database. These datasets were used in the validation analysis for biomarker discovery. The transcriptomic data that was collected in these datasets was from microarray analysis and RNA-sequencing.

Phenotypes	Serial no.	GEO DataSets Accession no.	BD status	BD Type Diagnosis	Cases	size	Controls	size	Organism	Description	Year	Gene Profiling Platform	Custom Gene-set Acquired?	References
Euthymic BD datasets	1	GSE112523	YES	not mentioned	BD post-mortem frontal cortex	10	healthy controls	17	Homo sapiens	Gene expression profiling of neurons from post-mortem brains of BD patients	2019	RNA-Seq	Yes (both gene-sets)	95
	2	GSE66196		not mentioned	BD fibroblasts and lymphoblasts	10	control BD fibroblasts and lymphoblasts	11		Global transcriptional response of lympho- and fibroblast cell lines derived from BD patients and unaffected controls to 1 mM LiCl ₂ .	2018	microarray	Yes (only for down-regulated gene-sets)	96
	3	GSE120340		not mentioned	BD patients	4	healthy controls	10		Gene expression profiling of dorsolateral prefrontal cortex of post-mortem brain tissue	2018	microarray	No	97
	4	GSE80655		not mentioned	BD post-mortem brain	71	healthy controls	70		Gene expression profiling was conducted on clinically-annotated human post-mortem brain tissues	2017	RNA-Seq	Yes (only for up-regulated gene-sets)	98
	5	GSE81396		not mentioned	BD post-mortem striatal tissue	8	healthy controls	8		Gene expression profiling of human putamen and caudate nucleus tissues in healthy control and BD individuals	2017	RNA-Seq	Yes (only for down-regulated gene-sets)	99
	6	GSE80336		not mentioned	BD post-mortem dorsal striatum	18	healthy controls	18		Gene expression profiling of the human dorsal striatum of BD patients	2016	RNA-Seq	Yes (only for up-regulated gene-sets)	100
	7	GSE78246		not mentioned	BD post-mortem anterior cingulate cortex	8	healthy controls	8		Gene expression profiling of anterior cingulate post-mortem brain samples	2016	microarray	No	101
	8	GSE74358		BD-1	BD hiPSC-derived neuron from fibroblasts	4	control hiPSC-derived neuron from fibroblasts	4		Gene expression profiling of Induced Pluripotent Stem Cells (iPSCs) from BD patients	2015	microarray	Yes (only for down-regulated gene-sets)	102
	9	GSE42546		not mentioned	BD post-mortem hippocampus granule cells	16	healthy controls	29		Gene expression profiling of post-mortem human hippocampus dentate gyrus granule cells	2014	RNA-Seq	Yes (only for down-regulated gene-sets)	103
	10	GSE53987		not mentioned	BD post-mortem; pre-frontal cortex, striatum and hippocampus	17	healthy controls	19		Gene expression profiling of pre-frontal cortex, striatum and hippocampus obtained from subjects with BD and matched controls	2014	microarray	No	104
	11	GSE53239		not mentioned	BD post-mortem dorsolateral prefrontal cortex	6	healthy controls	6		Gene expression profiling of post-mortem dorsolateral prefrontal cortex	2014	RNA-Seq	No	105
	12	GSE46449		BD-1	BD PBLs	43	control PBLs	39		Gene expression profiling of peripheral blood leukocytes (PBLs) of BD patients, BD patients suffering from 1st manic episode and healthy matched controls	2013	microarray	No	106
	13	GSE35974		not mentioned	BD post-mortem cerebellum, parietal cortex	37	healthy controls	50		Gene expression profiling data from the human cerebellum brain	2012	microarray	No	107
	14	GSE62191		not mentioned	BD post-mortem frontal cortex	29	healthy controls	30		Gene expression profiling of patients with BD and healthy controls	2014	microarray	No	108
	15	GSE87610		not mentioned	BD post-mortem dorsolateral prefrontal cortex	73	healthy controls	72		Gene expression profiling of L3 and L5 pyramidal neurons in the dorsolateral prefrontal cortex (DLPFC) from BD and unaffected subjects	2017	microarray	No	109
	16	GSE92538		not mentioned	BD post-mortem dorsolateral	42	healthy controls	##		Gene expression profiling of BD post-mortem dorsolateral prefrontal cortex	2016	microarray	No	110
	17	GSE116820		not mentioned	BD neurons induced from iPSCs	4	healthy controls	4		Gene expression profiling data of glutamatergic neuron and GABAergic neuron induced from iPSCs	2018	microarray	Yes (both gene-sets)	111

Li-treated datasets	18	GSE66196	YES	not mentioned	Li-treated BD fibroblasts and lymphoblasts	10	untreated BD fibroblasts and lymphoblasts	11	Homosapiens	Global transcriptional response of lympho- and fibroblast cell lines derived from BD patients and unaffected controls to 1 mM LiCl ₂ .	2018	microarray	Yes (both gene-sets)	96
	19	GSE93112		not mentioned	Li+ BD patients	4	Li- BD patients	4		Gene expression profiling data from lymphoblastoid cell lines (LCLs) from BD patients, LCLs were incubated for 4 days with or without LiCl (1mM)	2017	microarray	No	112
	20	GSE58933		not mentioned	Li+ BD iPSC-derived hippocampal dentate gyrus	6	Li- BD iPSC-derived hippocampal dentate gyrus granule neurons	4		Gene expression profiling of total RNAseq from neurons generated from BD patient-specific induced pluripotent stem (iPS) cells	2015	RNA-Seq	No	113
Mania datasets	21	GSE46416		not mentioned	BD manic phase blood samples	11	BD euthymic phase blood samples	11		Gene expression profiling of blood samples from euthymic and manic BD patients	2014	microarray	Yes (only for up-regulated gene-sets)	114
	22	GSE120340		not mentioned	Manic BD patients	6	euthymic BD patients	10		Gene expression profiling of dorsolateral prefrontal cortex of post-mortem brain tissue	2018	microarray	No	97
	23	GSE46449		BD-1	BD-1st episode PBLs	6	BD PBLs	39		Gene expression profiling of peripheral blood leukocytes (PBLs) of BD patients, BD patients suffering from 1st manic episode and healthy matched controls	2013	microarray	No	106

3.2.2. Normalization of Datasets

The first step was to normalize the transcriptomic data. For every datasets that had the raw values, such as ‘.cel’ files from microarray profiling and raw counts from RNA-seq datasets, these were normalized based on a consistent method. Microarray datasets were normalized using the BRB Arraytools feature on excel, which uses RMA normalization method. Whereas the RNA-seq raw counts were normalized on R through the DESeq2 package [94]. It is also to be noted that a few datasets did not contain the raw files, instead these had already processed/normalized values. These were used as it is.

Another thing of importance to be mentioned at this point is that due to the heterogeneity of datasets and values coming from a variety of sources, the gene annotations in the files; probesets, gene-ids etc were very diverse. Therefore, gene annotation was done to make the analysis easy and comparable. For this purpose, HGNC symbols were used throughout the analysis. However, while doing this step, a few of the datasets had to be discarded because the annotation proved to be very difficult owing either to old, outdated microarray analysis or after annotating, too few genes remained to make the dataset usable.

3.2.3. DGE Analysis

After processing the transcriptomic data, differential gene expression (DGE) analysis was performed. The statistical test used to do this analysis was the *student's t-test*. The samples were separated and compared according to the 3 categories mentioned above and indicated in the

datasets (**Table 3.2**). The normalized expression values were converted into a logarithmic scale and Log2 fold-change (log2FC) values were obtained:

$$\text{log2FC of one gene} = \text{average expression value of cases} - \text{average expression value of controls}$$

3.2.4. Creating Custom Gene-sets

As a consequence of DGE analysis done with each of the 23 datasets, the aim was then to create custom gene-sets with these new datasets. Using a standardized cut-off scheme shown in **Tables 3.3.1 and 3.3.2**, up-regulated and down-regulated gene matrices were created, respectively.

Table 3.3.1: DGE analysis to obtain up-regulated gene-sets. For every dataset, the filters in the 3rd column heading were applied to obtain a gene-set. Out of 23 datasets, up-regulated gene matrix was obtained from only 6 datasets. (Abbreviations: DEGs; differentially expressed genes, FC; fold-change).

Up-regulated Gene-sets	Type of Datasets	Total # of DEGs; p-value < 0.05 & FC >=1.5
11. GSE112523	BD vs controls	66
21. GSE116820	BD vs controls	91
2. GSE80655	BD vs controls	15
7. GSE80336	BD vs controls	17
13. GSE66196	Li-treated vs controls	38
10. GSE46416	Manic BD vs controls	19

Total # of gene-sets	6	
----------------------	---	--

Table 3.3.2: DGE analysis to obtain down-regulated gene-sets. For every dataset, the filters in the 3rd column heading were applied to obtain a gene-set. Out of 23 datasets, down-regulated gene matrix was obtained from only 7 datasets. (Abbreviations: DEGs; differentially expressed genes, FC; fold-change).

Down-regulated Gene-sets	Type of Datasets	Total # of DEGs; p-value < 0.05 & FC ≤ -1.5
11. GSE112523	BD vs controls	17
12. GSE66196	BD vs controls	26
21. GSE116820	BD vs controls	31
14. GSE81396	BD vs controls	85
23. GSE74358	BD vs controls	34
9. GSE42546	BD vs controls	22
13. GSE66196	Li-treated vs controls	18
Total # of gene-sets	7	

As result of this, 6 up-regulated gene-sets and 7 down-regulated gene-sets were created. Every dataset as expected did not pass the cut-offs. The gene lists belonging to these custom gene-sets are shown with their p-values and their datasets IDs in tables in **Appendices F and G** of up-regulated and down-regulated custom gene-sets, respectively.

While designing this scheme, a few considerations were kept in mind:

1. Statistical thresholds were not to be made too stringent as it would lead to loss of potentially substantial genes. The aim was to perform GSEA with these gene-sets, which is a multi-gene analysis and it has internal methods of determining significance. Thus, a stringent cut-off scheme would have led to premature limiting of the data.
2. The cut-offs were kept consistent throughout analyzing the different datasets because even though there was much heterogeneity, it was a way to standardize the datasets.

3.2.5. GSEA Performed with Custom Gene-sets

Gene-set enrichment analysis (GSEA) was performed using the *GSEA_4.1.0* version. All the datasets were run with each of the two gene-sets; up-regulated and down-regulated separately. The other specifications of GSEA software settings are mentioned in **Table 3.4**. These were kept constant throughout the analysis of all the datasets. The GSEA results tables are presented in **Tables 4.4.1 and 4.4.2**.

Table 3.4: Validation analysis - GSEA advance settings. The parameters and thresholds set before running GSEA (most were the default settings).

Required fields			
Number of permutations	1000		
Collapse/Remap to gene symbols	Collapse		
*Permutation type	phenotype		
Basic fields			
Enrichment statistic	weighted		
Metric for ranking genes	Signal2Noise		
Gene list sorting mode	real		
Gene list ordering mode	descending		
Max size: exclude larger sets	510 (up-regulated gene-sets)	100 (down-regulated gene-sets)	
Min size: exclude smaller sets	10 (up-regulated gene-sets)	15 (down-regulated gene-sets)	
Advanced fields			
Collapsing mode for probe sets => 1 gene	Max_probe		
Normalization mode	meandiv		
Randomization mode	no_balance		
Omit features with no symbol match	TRUE		
Median for class metrics	FALSE		
Seed for permutation	timestamp		
*If there are fewer than 7 samples in a phenotype/class use ‘gene_set’. Otherwise, use ‘phenotype’. Creates null distribution [41].			

3.2.6. Core Enrichment Analysis

After observing the patterns in the results from GSEA, the objective was to highlight the particular datasets that adhered to these patterns and extract what information we could from them. With this in mind, only the datasets that had significance value, **FDR $\leq$ 0.25**, and adhered

to the aforementioned patterns were selected. This is indicated in the last columns of the two GSEA results' **Tables 4.4.1 and 4.4.2.** The GSEA results of these datasets were extracted.

The GSEA software generates graphs and tables to show the enrichment via the enrichment score (ES). The greater the deviation of the ES from zero, the higher the probability that it will be a significant score. This significance means there is a greater degree of enrichment of a specific gene-set with one of the phenotypes (test vs. control). However, in a gene-set not all genes in it contribute equally to the high ES, it's a certain number of genes that cause this deviation. Therefore, one can be interested only in these genes, which are also referred to as the *leading-edge subset* or *core enriched genes* [38, 41].

Fig.3.2 shows a summarized depiction of GSEA and the 'core enrichment genes' concept that was used to extract the genes, that were causing the most deviation in the ES, from the enriched custom gene-sets.

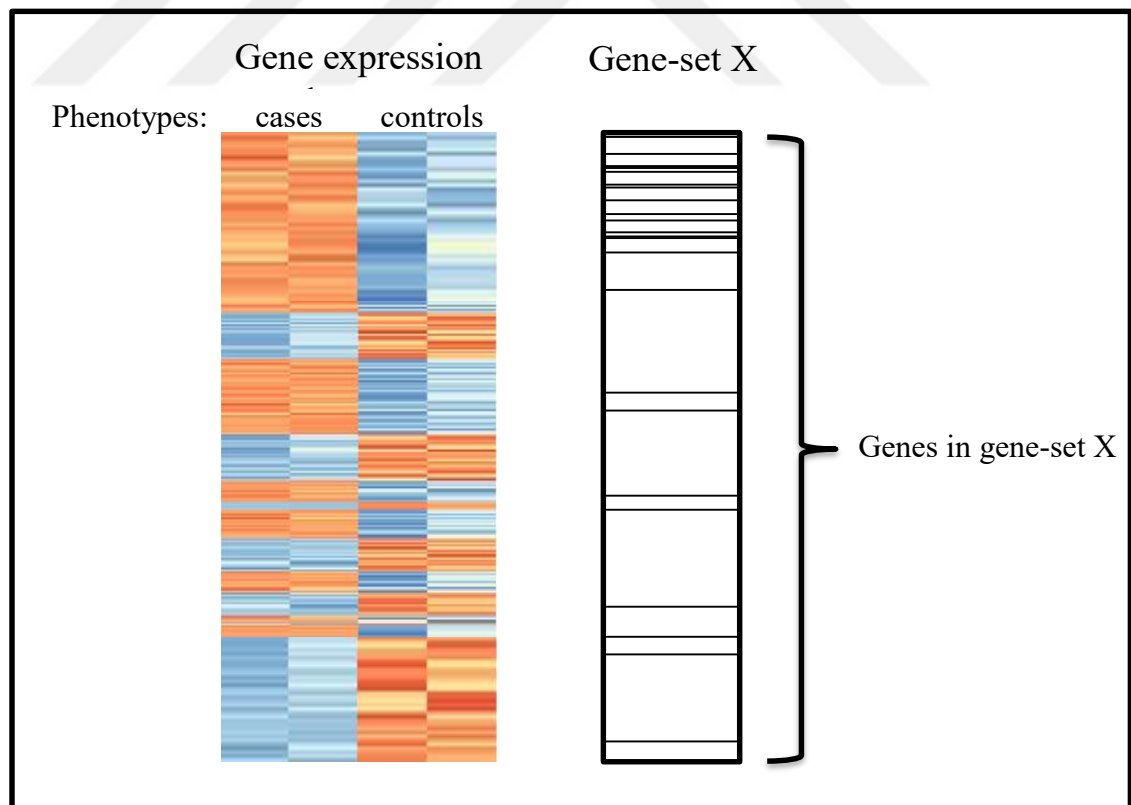


Figure 3.2: GSEA schematic - heatmap and gene-set. The heatmap is the gene expression profile (user input), which is sorted based on two phenotypes'; cases' and controls', differential expression. Thus, the

genes in the profile are ranked from the most differentially expressed genes between the cases and controls, with highest expression in cases vs. lowest expression in controls. The genes get less differentially expressed in the middle of the heatmap and with the opposite expression trend seen at the bottom of the heatmap. A gene-set X, is then compared to the heatmap. If majority of the genes in a gene-set cluster at either the top or the bottom of the heatmap, this will generate a high ES. In this case, it can be seen that the genes cluster at the top, that is, the genes correlate with the cases. *Therefore, gene-set X is enriched in cases.*

It could be the case, that the genes in a particular gene-set are randomly correlating with the heatmap. Then the ES will not be high and will probably have a non-significant FDR value so we can discard such a gene-set. As such a gene-set will not be biologically relevant. If the genes in gene-set X had clustered at the bottom, it would have meant that they are enriched in controls. Thus, this gene-set will still have a high ES and will be considered biologically relevant. This concept is illustrated in **Fig.3.3** below.

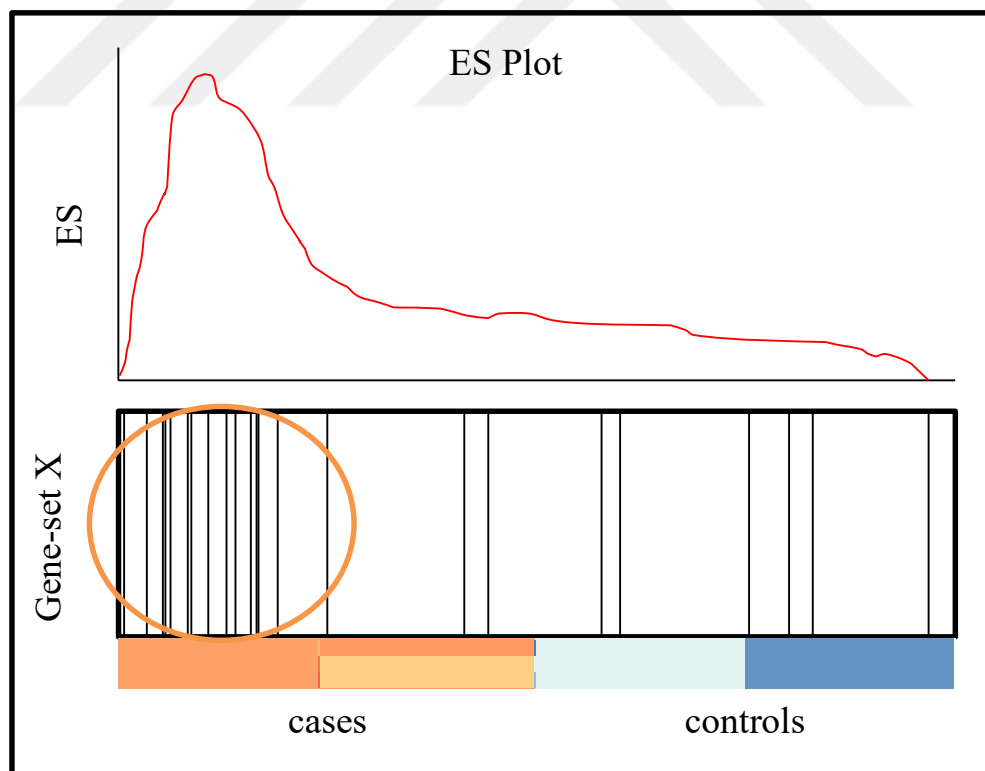


Figure 3.3: GSEA schematic – Core enriched genes. The greater the number of genes in gene-set X (shown by vertical lines) are enriched with cases, the higher the ES. This is shown by the clustering of lines, i.e., genes on the left. The rest of the genes in the gene-set are more spread out along gene-set X,

meaning they are not enriched. Therefore, the gene-of-interests (GOIs) are the genes that have been encircled in green. These genes can be identified from the gene-set X and thus, this is how the core-enriched genes can be obtained from the enriched gene-sets.

We obtained enrichment plots from all the datasets that adhered to patterns 1 and 2. These are depicted in **Fig.4.1 to Fig.4.4** in the results section.

GSEA also creates a table along with the figures that indicate these leading-edge subset genes. This table is shown in **Table 4.5**. The genes that are marked as ‘Yes’ in the core enrichment column are the ones contributing to high ES. These core enriched genes were extracted from each of these datasets.

All these genes were extracted from both the GSEA **Tables 4.4.1 and 4.4.2** and merged. The ones belonging to: Pattern 1: up-regulated in BD and lithium-treated samples while down-regulated in manic samples and, Pattern 2: down-regulated in BD and lithium-treated samples while being up-regulated in manic samples. They are in short form referred to as **pattern 1** and **pattern 2**, respectively. And we refer to this nomenclature throughout the rest of the thesis.

3.3. Final List of Genes

3.3.1. FC Values of Core Enriched Genes

After the lists of core enriched genes that adhered to the two patterns were compiled, a secondary validation step was performed. Since, GSEA results only led to the selection of a few datasets and not all, we still wanted to see the expression profile of the list of these final core enriched genes in the datasets. To observe this, we sought to only see the direction of expression of these genes in *all* the datasets compared to their specific controls. This was done by looking at their log2FC.

Genes were discarded if they were not found in all 3 types of datasets. For example, if one gene was only found in BD vs healthy control datasets and manic vs euthymic datasets but not in one of the lithium-treated datasets, that gene was eliminated. **Tables 4.6 and 4.7** lists the final genes.

The top genes that adhered to the patterns were selected as putative biomarkers, **Tables 4.8 and 4.9**. As another statistical validation step and as a way to measure the power of these genes as biomarkers, we performed non-parametric tests; *Wilcoxon* and *Kruskal–Wallis*.

3.4. Network Analysis and Pathway Analysis

Simultaneously, to the biomarker discovery using the final list of genes, we also performed functional analysis to see whether more can be learned about the genes in relation to psychiatry, neurology and specifically to bipolar disorder. For this purpose, the final lists of genes were inputted in *NetworkAnalyst* [45] to perform network analysis using its ‘Gene-disease Associations’ function. The rest of the settings were kept at default. *Gene-Ontology Resource* [47] and *Gene Set Analyzer* [50] were used for pathway analysis, the curated pathways belonged to biological pathways (BP) from the former resource while KEGG and Reactome were databases were used in the latter online tools.

CHAPTER 4

4. Results

Note

In this section, the tables and figures have been color coded as follows:

- orange represents euthymic BD datasets,
- yellow represents lithium-treated datasets, and
- blue represents mania datasets.

4.1. Preliminary Analysis

4.1.1. GSEA Results from Preliminary Analysis

Summary **Tables 4.1 and 4.2** show that 2 patterns emerged in the enrichment of gene-sets with the datasets. These ‘discovery’ patterns then formed the basis of further analysis done in this project. **Fig.1.3** shown in the introduction section represents these newly discovered patterns.

This was the result of GSEA done with custom gene-sets as part of the preliminary analysis. The complete tables of results are presented in the **Appendices D and E as tables**, which correspond to GSEA done with up-regulated and down-regulated custom gene-sets, respectively.

Table 4.1: Preliminary analysis - summarized version of GSEA results with up-regulated gene-sets. In this table only the values; $FDR \leq 0.25$, are shown, the results through which **Pattern 1: Up BD-Li; Down Mania** was observed.

Datasets		Up-regulated Gene-sets									
		Li-treated Gene-sets						Bipolar disorder Gene-sets			
		GSE35291		GSE66277_1		GSE66277_2		GSE46449		GSE46416	
		Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Li-treated datasets (Cases: Li+; Controls: Li-)	GSE35291									0.162	
	GSE66277_1							0.174			
	GSE66277_2							0.063		0.054	
	GSE54294_1									0.034	
	GSE5429_2									0.045	

Bipolar disorder datasets (Cases: BD patients; Controls:	GSE46416	0.072										
	GSE69486			0.085								
Mania datasets (Cases: Manic BD patients; Controls:	GSE46449_1		0.157									0.037
	GSE46449_2								0.008			0.019

Table 4.2: Preliminary analysis - summarized version of GSEA results with down-regulated gene-sets. In this table only the values; $FDR \leq 0.25$, are shown, the results through which **Pattern 2: Down BD-Li; Up Mania** was observed.

Datasets		Down-regulated Gene-sets											
		Li-treated Gene-sets						Bipolar disorder Gene-sets				Mania Gene-sets	
		GSE35291		GSE66277_1		GSE66277_2		GSE46449		GSE46416		GSE29417	
		Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Li-treated datasets (Cases: Li+; Controls: Li-)	GSE35291											0.116	
	GSE66277_2								0.001		0.083		
	GSE66277_3										0.144		
Mania datasets (Cases: Manic BD patients; Controls: Euthymic BD patients)	GSE46449_2			0.129				0					

4.2. Validation Analysis

The foundation of this study was laid down by the results of the GSEA done with custom gene-sets from the preliminary analysis done by Karaahmet, detailed above and in the methods section. The results of this analysis led to the discovery of a novel pattern of gene expression. This was at first counter-intuitive but we proceeded with it to see whether subsequent analysis might highlight a biological aspect as yet missing from BD analysis done so far, and this is what we achieved.

4.2.1. Datasets Used for Analysis

Table 4.3 shows the new datasets obtained from *GEO DataSets* database. Detailed information about these datasets is given in **Table 3.2** in the methods section.

Table 4.3: Validation analysis' datasets summary. This table shows the sources of RNA extraction in each experimental setup from all the datasets and the samples sizes. *Dataset no. is based on the datasets as they are referred to in the GSEA results and other graphs for consistency. **This table has been sorted in descending order based on the number of cases. (Abbreviations: iPSC; induced pluripotent stem cells).

*Dataset no.	GEO Accession no.	Tissue	**No. of Cases	No. of Controls	Year
1	GSE87610	post-mortem dorsolateral prefrontal cortex	73	72	2017
2	GSE80655	post-mortem brain	71	70	2017
3	GSE46449	peripheral blood leukocytes	43	39	2013
4	GSE92538	post-mortem dorsolateral prefrontal cortex	42	119	2016
5	GSE35974	post-mortem cerebellum	37	50	2012
6	GSE62191	post-mortem frontal cortex	29	30	2014
7	GSE80336	post-mortem dorsal striatum	18	18	2016
8	GSE53987	post-mortem brain; pre-frontal cortex, striatum and hippocampus	17	19	2014
9	GSE42546	post-mortem hippocampus granule cells	16	29	2014
10	GSE46416	blood samples	11	10	2014
11	GSE112523	post-mortem brain; frontal cortex	10	17	2019
12	GSE66196	fibroblasts and lymphoblasts	10	11	2018
13	GSE66196	fibroblasts and lymphoblasts	10	11	2018
14	GSE81396	post-mortem putamen and caudate nucleus tissues	8	8	2017
15	GSE78246	post-mortem anterior cingulate cortex	8	8	2016

16	GSE53239	post-mortem dorsolateral prefrontal cortex	6	6	2014
17	GSE58933	iPSC-derived hippocampal dentate gyrus granule neurons	6	4	2015
18	GSE120340	post-mortem brain; dorsolateral prefrontal cortex	6	10	2018
19	GSE46449	peripheral blood leukocytes	6	39	2013
20	GSE120340	post-mortem brain; dorsolateral prefrontal cortex	4	10	2018
21	GSE116820	neurons induced from iPSCs	4	4	2018
22	GSE93112	lymphoblastoid cell-lines	4	4	2017
23	GSE74358	iPSC-derived neuron from fibroblasts	4	4	2015

4.2.2. GSEA Results

Table 4.4.1: GSEA results with Up-regulated gene-sets. The 6 columns are the gene-sets that were obtained after performing DEG analysis. Therefore, these are referred to as custom gene-sets. The rows show all the datasets. For each dataset (row), GSEA was performed using all 6 gene-sets (columns) to observe whether any enrichment of the genes present in these custom gene-sets occurred with the dataset.

A false-discovery rate (FDR) value was obtained for each custom gene-set analysed with each dataset. As **FDR values ≤ 0.25** are considered significant, therefore, only those values have been highlighted.

- All the green cells show significance.
- The red font in green cells shows enrichment of genes that correspond to **Pattern 1: Up BD-Li; Down Mania**.
- The values marked with asterisk are significant values that contradicts the pattern.

The last column shows which dataset was used to obtain core enrichment genes.

		UP-REGULATED CUSTOM GENE-SETS												
		11. GSE112523		21. GSE116820		2. GSE80655		7. GSE80336		13. GSE66196		10. GSE46416		
		BD vs CTRL post-mortem brains; frontal cortex		BD vs CTRL neurons induced from iPSCs		BD vs CTRL post-mortem brain		BD vs CTRL post-mortem dorsal striatum		Li treated lymphoblasts		Manic blood samples		Datasets selected to retrieve core enriched genes
Dataset no.	GEO Accession no.	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	
1	GSE87610	0.767		0.671					0.691	0.565			0.753	
2	GSE80655	0.116			0.916	0.231		0.768			1		1	
3	GSE46449	1		1				0.93			0.922	1		
4	GSE92538		0.603		0.075*			0.207			0.729	0.921		
5	GSE35974		0.985		0.9				0.427		0.461		0.904	
6	GSE62191		0.839		1						1	0.02*		
7	GSE80336	0.063			1			0.449			1		0.909	
8	GSE53987	0.878		0.378					0.735		0.905	0.034*		
9	GSE42546				1					0.761				
11	GSE112523	0.213		0.952				0.25			0.782		0.4	
12	GSE66196	0.002		0.426				0.313			0.681		0.813	
12	GSE66196		0.718	0.618				0.642		0.077			0.776	←
14	GSE81396		1		1				0.868	0.458		0.713		
15	GSE78246		0.916		0.499				0.357	0.798		0.014*		
16	GSE53239	0.083		1					0.411	0.065		0.978		←
18	GSE120340	0.161		0.263				0.3			0.045*		0.146	←
21	GSE116820	0.054		0				0.537		0.044			0.359	←
23	GSE74358		0.036*		0*				0.399		0.459	0.017*		

13	GSE66196		0.151*	0.355				0.289			0.398	0.17*		
13	GSE66196	0.579		0.46					0.287	0		0.129*		
17	GSE58933		0.123*		0.011*				0.075*		0.131*		0.962	
22	GSE93112	0.827		0.651			0.47	0.765			0.771		0.979	
10	GSE46416	0.18*			0.336					0*		0.077		
18	GSE120340		0	0.238*					0.028	0.025*		0.046		←
19	GSE46449		0.653		0.153			0.972			0.007		0.002*	←

Table 4.4.2: GSEA results with Down-regulated gene-sets. The 7 columns are the gene-sets that were obtained after performing differential gene expression analysis. Therefore, these are referred to as custom gene-sets. The rows show all the datasets. For each dataset (row), GSEA was performed using all 7 gene-sets (columns) to observe whether any enrichment of the genes present in these custom gene-sets occurred with the dataset.

A false-discovery rate (FDR) value was obtained for each custom gene-set analyzed with each dataset. As **FDR values ≤ 0.25** are considered significant, therefore, only those values have been highlighted.

- All the green cells show significance.
- The red font in green cells shows enrichment of genes that correspond to **Pattern 2: Down BD-Li; Up Mania**.
- The values marked with asterisk are significant values that contradicts the pattern.

The last column shows which dataset was used to obtained core enrichment genes.

		DOWN-REGULATED CUSTOM GENE-SETS														
		11. GSE112523		12. GSE66196		21. GSE116820		14. GSE81396		23. GSE74358		9. GSE42546		13. GSE66196		
		BD vs CTRL post-mortem brains; frontal cortex		BD vs Ctrl lymphoblasts		BD vs CTRL neurons induced from iPSCs		BD vs CTRL post-mortem putamen and caudate nucleus tissues		BD vs CTRL hiPSC-derived neuron from fibroblasts		BD vs CTRL hiPSC-derived neuron from fibroblasts		Li treated lymphoblasts		Datasets selected to retrieve core enriched genes
Dataset no.	GEO Accession no.	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	
1	GSE87610			0.971			0.55		0.142		0.857	0.762		1		
2	GSE80655	0.807		0.525		0.805			1	0.618			0.908			
3	GSE46449				0.892		0.824		1	0.677			0.99			
4	GSE92538				0.921	0.426			0.145		0.668					
5	GSE35974			0.48		0.331		0.878		0.287		0.887				
6	GSE62191							0.599		0.848						
7	GSE80336			0.272		0.262			0.777		0.665	0.452				
8	GSE53987			0.477		0.812			0.981	0.831		0.545				
9	GSE42546												0.028			
11	GSE112523	0.697		0.798		0.895		0.236*			0.848	1				
12	GSE66196			0.36			0.281	0.072*			0.392		0.778			
12	GSE66196				0	0.244*			0.84		0.194	0.811				
14	GSE81396			0.922		0.685			0	0.702			0.755			
15	GSE78246			1		1			0.254		0.579		0.436			
16	GSE53239			0.033*		0.993			0	0.771		1				
20	GSE120340			0.377			0.796		0	0.293			0.497			
21	GSE116820			0.44			0	0.436		0*			0.451			
23	GSE74358				0.146	0.279			0.541		0		0.662			

13	GSE66196			0.211*		0.235*			0.351	0.247*			0.793			
13	GSE66196				0.14	0.701		0.222*			0.457	0.29				←
17	GSE58933				0.967	0.249*		0*			0.24	0.335				←
22	GSE93112				1		1	0.69		0.776			0.994			
10	GSE46416								0.518		0.483					
18	GSE120340				0.789	0.01		0.289		0.278		0.449				←
19	GSE46449				0.808	0.866		1		0.928			1			

4.2.3. Extraction of Core Enriched Genes

From the GSEA results' **Tables 4.4.1 and 4.4.2**, the whole transcriptome of the indicated datasets, marked with red arrows in the last columns of these tables, were obtained. As these were the datasets that adhered to the discovered patterns. Then further within these transcriptomes, the genes common with the enriched custom gene-sets were extracted. Furthermore, another step was required to identify the leading-edge subset or core enrichment genes. The core enriched genes were the genes in the gene-sets that caused the high, significant ES.

The four ES plots of the highlighted datasets (red arrows) are shown below. **Fig.4.1 and Fig.4.2** adhered to pattern 1, while **Fig.4.3 and Fig.4.4** adhered to pattern 2. The rest of the ES plots are in **Appendix H**.

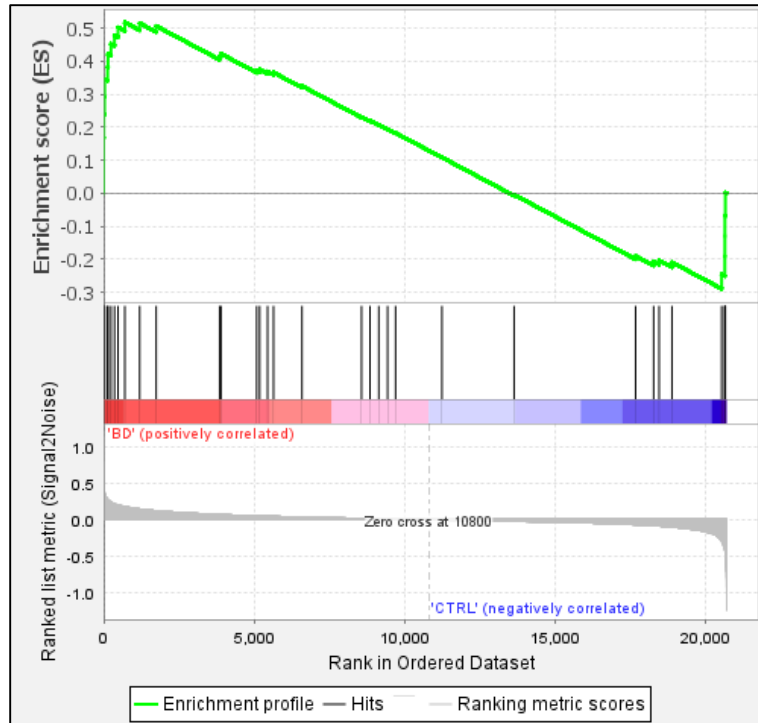


Figure 4.1: Pattern 1 - Enrichment plot of euthymic BD dataset with lithium-treated gene-set. The ES score is 0.5, with a significant **FDR = 0.077** value. The vertical bars, representing the up-regulated genes in this lithium-treated gene-set *no. 13*, cluster at the left side contributing to the high ES value. These genes correlate with the up-regulated genes in the euthymic BD patients of the BD dataset *no. 12*. Therefore, these are our genes of interest, i.e.; core enriched genes.

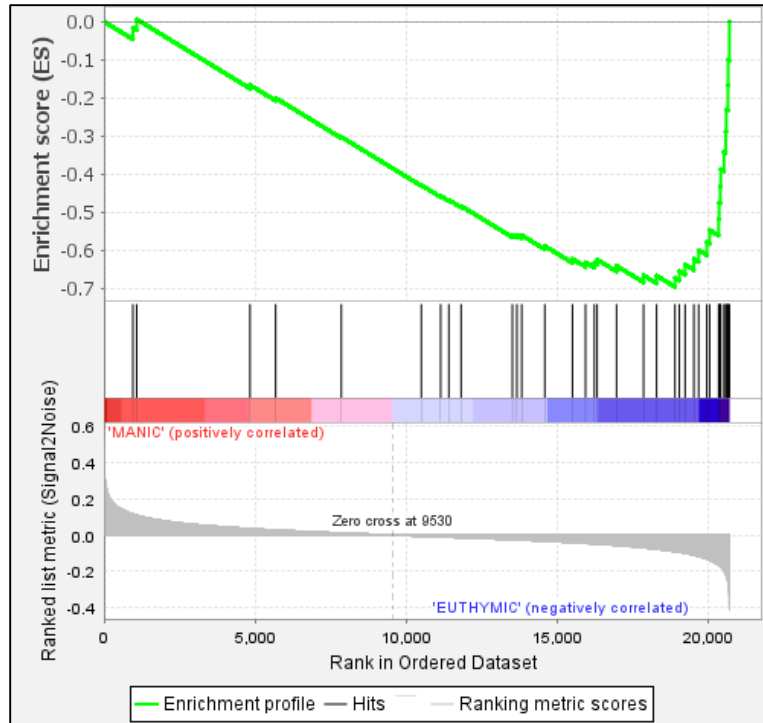


Figure 4.2: Pattern 1 - Enrichment plot of mania dataset with euthymic BD gene-set. The ES score is -0.7, with a significant **FDR = 0.0** value. The vertical bars, representing the up-regulated genes in this euthymic BD gene-set *no. 11*, cluster at the right side contributing to the high, negative ES value. These genes correlate with the up-regulated genes in the euthymic controls of the mania dataset *no. 18*. Hence, these genes are down-regulated in manic BD patients. Therefore, these are our genes of interest i.e.; core enriched genes.

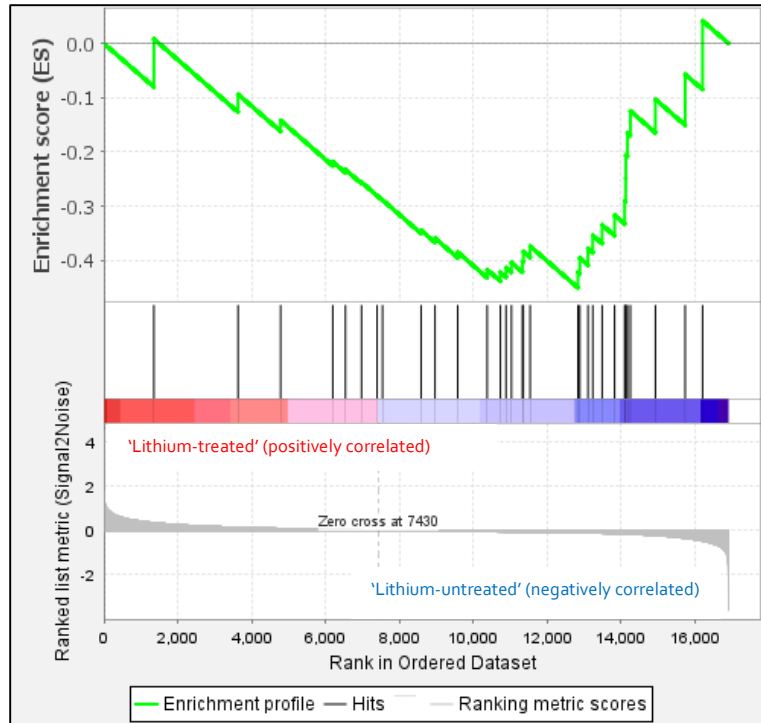


Figure 4.3: Pattern 2 - Enrichment plot of lithium-treated dataset with euthymic BD gene-set. The ES score is -0.4, with a significant **FDR = 0.24** value. The vertical bars, representing the down-regulated genes in this euthymic BD gene-set *no. 23*, cluster at the right side contributing to the high, negative ES value. These genes correlate with the up-regulated genes in the untreated controls of lithium dataset *no. 17*. Hence, these genes are down-regulated in lithium-treated BD patients. Therefore, these are our genes of interest, i.e.; core enriched genes.

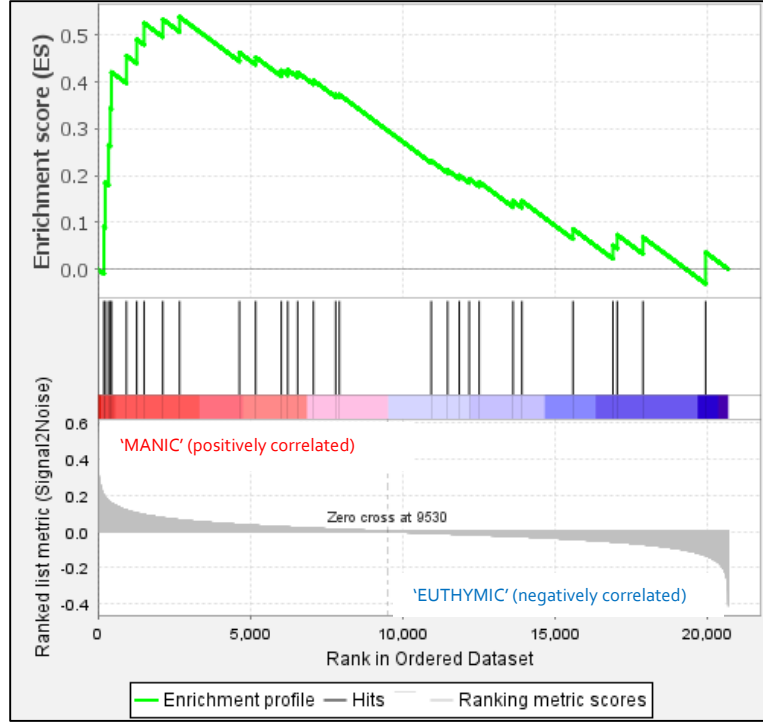


Figure 4.4: Pattern 2 - Enrichment plot of mania dataset with euthymic BD gene-set. The ES score is 0.5, with a significant **FDR = 0.01** value. The vertical bars, representing the down-regulated genes in this euthymic BD gene-set *no. 21*, cluster at the left side contributing to the high ES value. These genes correlate with the up-regulated genes in the manic BD patients of mania dataset *no. 18*. Hence, these genes are down-regulated in euthymic BD patients. Therefore, these are our genes of interest, i.e.; core enriched genes.

The identities of these core enriched genes were obtained from each of the corresponding tables of these ES plots. **Table 4.5** is shown as one example; this table corresponds to **Fig4.1**. The rest are in the **Appendix I**. These genes were extracted from each of the tables and merged as one and then further analysis was done to create the final list of genes.

A point to be noted is that the clustering at the extremes of the horizontal gene-sets are not very dense because we used custom gene-sets, and the number of genes in the list were maximum 510 and minimum 10 for up-regulated gene-sets and maximum 100 and minimum 15 for down-regulated gene-sets. The max and min were defined based on the largest and smallest list of genes present in a gene-set from both the files.

Table 4.5: Extraction of core enriched genes. This table is the corresponding table of the ES plot in **Fig.4.1**. As mentioned in the methods section, the genes that contribute to high ES, are referred to as ‘core enriched genes. These are the vertical lines clustering at the left side in **Fig.4.1** and these are genes indicated as ‘Yes’ in the CORE ENRICHMENT column of this table.

	SYMBOL	TITLE	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	KLRC1	killer cell lectin like receptor C1 [Source:HGNC Symbol;Acc:HGNC:6374]	1	1.13	0.1675	Yes
2	DNM3	dynamin 3 [Source:HGNC Symbol;Acc:HGNC:29125]	8	0.488	0.2396	Yes
3	SV2B	synaptic vesicle glycoprotein 2B [Source:HGNC Symbol;Acc:HGNC:16874]	30	0.366	0.2929	Yes
4	SPINK2	serine peptidase inhibitor Kazal type 2 [Source:HGNC Symbol;Acc:HGNC:11245]	35	0.361	0.3462	Yes
5	RIMBP2	RIMS binding protein 2 [Source:HGNC Symbol;Acc:HGNC:30339]	133	0.275	0.3823	Yes
6	HOPX	HOP homeobox [Source:HGNC Symbol;Acc:HGNC:24961]	147	0.27	0.4217	Yes
7	DSC2	desmocollin 2 [Source:HGNC Symbol;Acc:HGNC:3036]	245	0.241	0.4527	Yes
8	LHX2	LIM homeobox 2 [Source:HGNC Symbol;Acc:HGNC:6594]	362	0.212	0.4786	Yes
9	WNT2	Wnt family member 2 [Source:HGNC Symbol;Acc:HGNC:12780]	480	0.199	0.5024	Yes
10	PRKCH	protein kinase C eta [Source:HGNC Symbol;Acc:HGNC:9403]	705	0.177	0.5178	Yes
11	DSC3	desmocollin 3 [Source:HGNC Symbol;Acc:HGNC:3037]	1198	0.144	0.5154	No
12	CREB5	cAMP responsive element binding protein 5 [Source:HGNC Symbol;Acc:HGNC:16844]	1746	0.12	0.5067	No
13	PIEZO2	piezo type mechanosensitive ion channel component 2 [Source:HGNC Symbol;Acc:HGNC:26270]	3852	0.068	0.4149	No
14	NFAT5	nuclear factor of activated T cells 5 [Source:HGNC Symbol;Acc:HGNC:7774]	3897	0.067	0.4227	No
15	ATXN7L3 B	ataxin 7 like 3B [Source:HGNC Symbol;Acc:HGNC:37931]	5082	0.049	0.3727	No
16	ST3GAL6	ST3 beta-galactoside alpha-2,3-sialyltransferase 6 [Source:HGNC Symbol;Acc:HGNC:18080]	5181	0.048	0.375	No
17	NCR2	natural cytotoxicity triggering receptor 2 [Source:HGNC Symbol;Acc:HGNC:6732]	5443	0.045	0.3691	No
18	CCL20	C-C motif chemokine ligand 20 [Source:HGNC Symbol;Acc:HGNC:10619]	5641	0.043	0.3659	No
19	DNASE1L 3	deoxyribonuclease 1 like 3 [Source:HGNC Symbol;Acc:HGNC:2959]	6586	0.033	0.3251	No
20	BCL2L11	BCL2 like 11 [Source:HGNC Symbol;Acc:HGNC:994]	8562	0.015	0.2317	No
21	PDCD4	programmed cell death 4 [Source:HGNC	8843	0.013	0.22	No

		Symbol;Acc:HGNC:8763]				
22	VPS35	VPS35 retromer complex component [Source:HGNC Symbol;Acc:HGNC:13487]	9135	0.011	0.2076	No
23	GRIA3	glutamate ionotropic receptor AMPA type subunit 3 [Source:HGNC Symbol;Acc:HGNC:4573]	9427	0.009	0.1948	No
24	IL6ST	interleukin 6 signal transducer [Source:HGNC Symbol;Acc:HGNC:6021]	9692	0.007	0.183	No
25	MYB	MYB proto-oncogene, transcription factor [Source:HGNC Symbol;Acc:HGNC:7545]	11232	-0.003	0.1089	No
26	LRP12	LDL receptor related protein 12 [Source:HGNC Symbol;Acc:HGNC:31708]	13626	-0.02	-0.0041	No
27	WSB1	WD repeat and SOCS box containing 1 [Source:HGNC Symbol;Acc:HGNC:19221]	17659	-0.067	-0.1895	No
28	UNC13C	unc-13 homolog C [Source:HGNC Symbol;Acc:HGNC:23149]	18258	-0.08	-0.2066	No
29	PEX13	peroxisomal biogenesis factor 13 [Source:HGNC Symbol;Acc:HGNC:8855]	18441	-0.085	-0.2028	No
30	MUC19	mucin 19, oligomeric [Source:HGNC Symbol;Acc:HGNC:14362]	18867	-0.099	-0.2087	No
31	ENTHD1	ENTH domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26352]	20515	-0.291	-0.2453	No
32	EOMES	eomesodermin [Source:HGNC Symbol;Acc:HGNC:3372]	20622	-0.405	-0.1904	No
33	XCL1	X-C motif chemokine ligand 1 [Source:HGNC Symbol;Acc:HGNC:10645]	20624	-0.413	-0.1291	No
34	FRMD4B	FERM domain containing 4B [Source:HGNC Symbol;Acc:HGNC:24886]	20625	-0.415	-0.0675	No
35	CXCL10	C-X-C motif chemokine ligand 10 [Source:HGNC Symbol;Acc:HGNC:10637]	20639	-0.474	0.0021	No

4.2.4. Validating the Discovery Patterns with Core Enriched Genes

After extracting the core enriched genes, a ‘circulating back’ step to validate the results was performed. With these lists of core enriched genes, we aimed to look at their direction of expression in all the datasets without applying any cutoffs. We wanted to see whether pattern 1 and pattern 2 expressions could be seen in *all* the datasets quantitatively.

Table 4.6: Pattern 1: Up BD-Li; Down mania list of final genes. These are up-regulated in BD and Li-treated samples and down-regulated in mania samples. The 3rd and 4th columns show the % of datasets the genes were found in and subsequently, the % of datasets in which the genes adhered to Pattern 1 (*the genes are sorted in descending order of this value*), respectively.

Pattern 1: Up BD-Li; Down mania			
	Genes	Percentage of datasets in which the gene is found (%)	Percentage of datasets in which the gene adhered to Pattern 1 (%)
1	ENTHD1	74	94
2	DSC2	96	91
3	ADAMTS9	87	90
4	STC1	91	90
5	IL6ST	91	86
6	SFN	91	86
7	THBS1	96	86
8	CHI3L2	87	80
9	CALD1	100	78
10	HILPDA	78	78
11	GRIA3	96	77
12	CYP1B1	91	76
13	IGFBP7	91	76
14	MKI67	91	76
15	ZNF700	91	76
16	CCL20	70	75
17	COL2A1	87	75
18	SPPL2C	70	75
19	XCL1	70	75
20	P2RX2	83	74
21	DGKG	96	73
22	GRM5	96	73
23	MUC19	65	73
24	PTGER3	96	73
25	CLCF1	91	71
26	DSC3	74	71
27	HAMP	91	71
28	KHDC3L	57	69
29	NCR2	70	69
30	BMP5	96	68
31	LHX2	96	68
32	SORCS2	96	68

33	BGN	91	67
34	PDIA2	91	67
35	PIEZO2	78	67
36	IL1B	74	65
37	PRKCH	100	65
38	SV2B	100	65
39	HAMP	87	65
40	MYLK	96	64
41	HOPX	96	64
42	FN1	91	62
43	EOMES	91	62
44	SERPINA1	91	62
45	SORCS2	100	61
46	DHRS3	100	61
47	RGS6	100	61
48	PIRT	100	61
49	S100A9	78	61
50	BDNF	87	60
51	PEX13	74	59
52	SPARC	96	59
53	LRP12	96	59
54	MYB	91	57
55	WNT2	87	55
56	RIMBP2	74	53
57	UNC13C	91	52
58	FRMD4B	87	50
59	SPINK2	70	50
60	WSB1	87	50
61	ACY3	100	48
62	CXCL10	74	47
63	GRM3	96	45
64	DNM3	96	45
65	RNASE3	100	43

Table 4.7: Pattern 2: Down BD-Li; Up mania list of final genes. These are down-regulated in BD and Li- treated samples and up-regulated in mania samples. The 3rd and 4th columns show the % of datasets the genes were found in and subsequently, the % of datasets in which the genes adhered to Pattern 2 (*the genes are sorted in descending order of this value*), respectively.

	Pattern 2: Down BD-Li; Up mania		
	Genes	Percentage of datasets in which the gene is found (%)	Percentage of datasets in which the gene adhered to Pattern 2 (%)
1	COL1A2	91	81
2	DNAJC7	96	77
3	LAMA4	91	76
4	ZNF736	74	76
5	LEO1	87	75
6	LRRTM4	96	73
7	SGCD	96	73
8	GLT8D2	87	70
9	B3GALNT1	96	68
10	PDZD2	96	68
11	SORBS2	96	68
12	ZNF253	91	67
13	OGN	87	65
14	COL5A2	96	64
15	LAMB1	96	64
16	MAP1B	96	64
17	NR1D2	96	64
18	RECQL	96	64
19	KCTD14	35	63
20	HAPLN1	91	62
21	RBPM5	91	62
22	ZNF471	91	62
23	CHRNA5	87	60
24	NAPRT	65	60
25	EDNRA	91	57
26	FAM221A	70	56
27	RGS13	70	56
28	TBX15	78	56
29	ABHD3	96	55
30	ESF1	96	55
31	LRRC34	87	45
32	PRRX1	87	45
33	COL21A1	96	41

To check whether the ‘Percentage of datasets in which the gene adhered to the patterns’ had any correlation with the number of datasets or not, we plotted a correlation graph for each of the above tables. The low R^2 values indicates there is no correlation between the 2 variables. These are shown below in **Fig.4.5** and **Fig.4.6**.

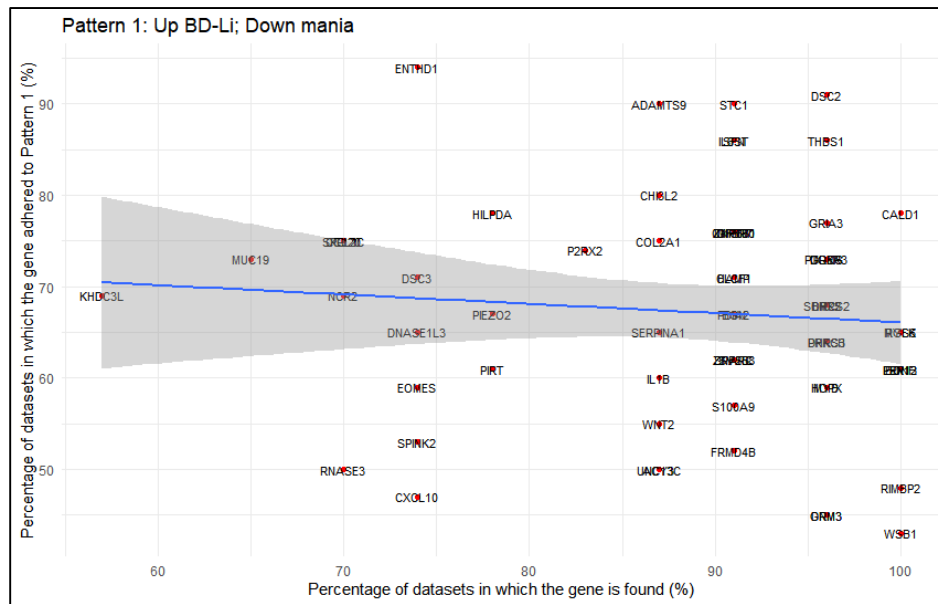


Figure 4.5: Pattern 1 genes correlation plot. A scatterplot of % of datasets in which the genes adhered to Pattern 1 vs. % of datasets in which the gene is found. The blue line is the regression line; $R^2=0.008$.

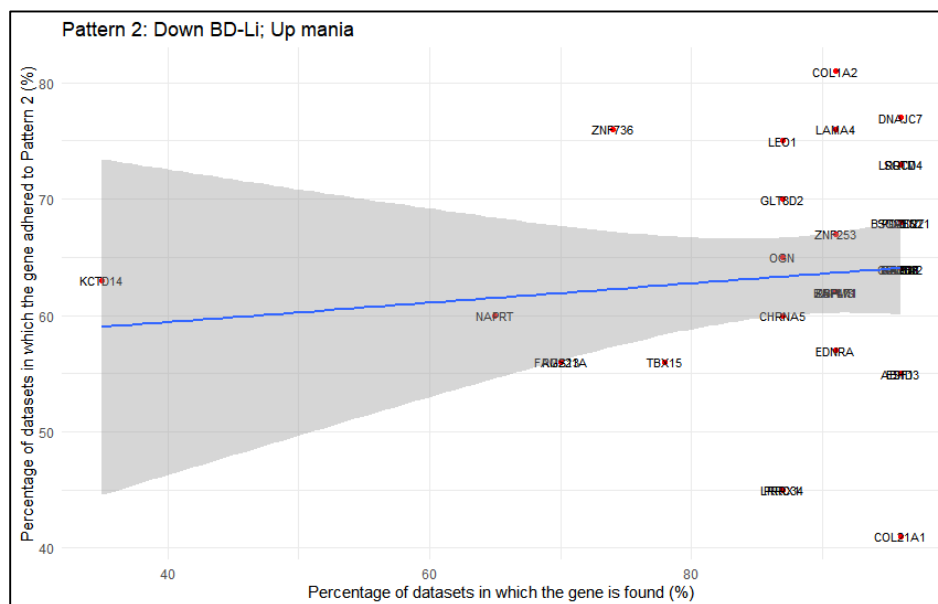


Figure 4.6: Pattern 2 genes correlation plot. A scatterplot of % of datasets in which the genes adhered to Pattern 2 vs. % of datasets in which the gene is found. The blue line is the regression line; $R^2=0.0129$.

4.3. Selection of Putative Biomarkers

From these lists of final genes, that adhered to patterns 1 and 2. We selected the top genes. The cutoff was:

1. **Table 4.8: Genes that adhered to pattern 1 in more than or equal to 85% of the datasets.**

1	ENTHD1
2	DSC2
3	ADAMTS9
4	STC1
5	IL6ST
6	SFN
7	THBS1

2. **Table 4.9: Genes that adhered to pattern 2 in more than or equal to 75% of the datasets.**

1	COL1A2
2	DNAJC7
3	LAMA4
4	ZNF736
5	LEO1

In order to visualize the direction of expression of these genes in all the datasets, we plotted their ‘direction of expression’ graphs. Fold change values, $\log_2FC$, are used as ‘direction of expression’ pattern. The $\log_2FC$ values of each gene were calculated between the cases and controls in all of the datasets.

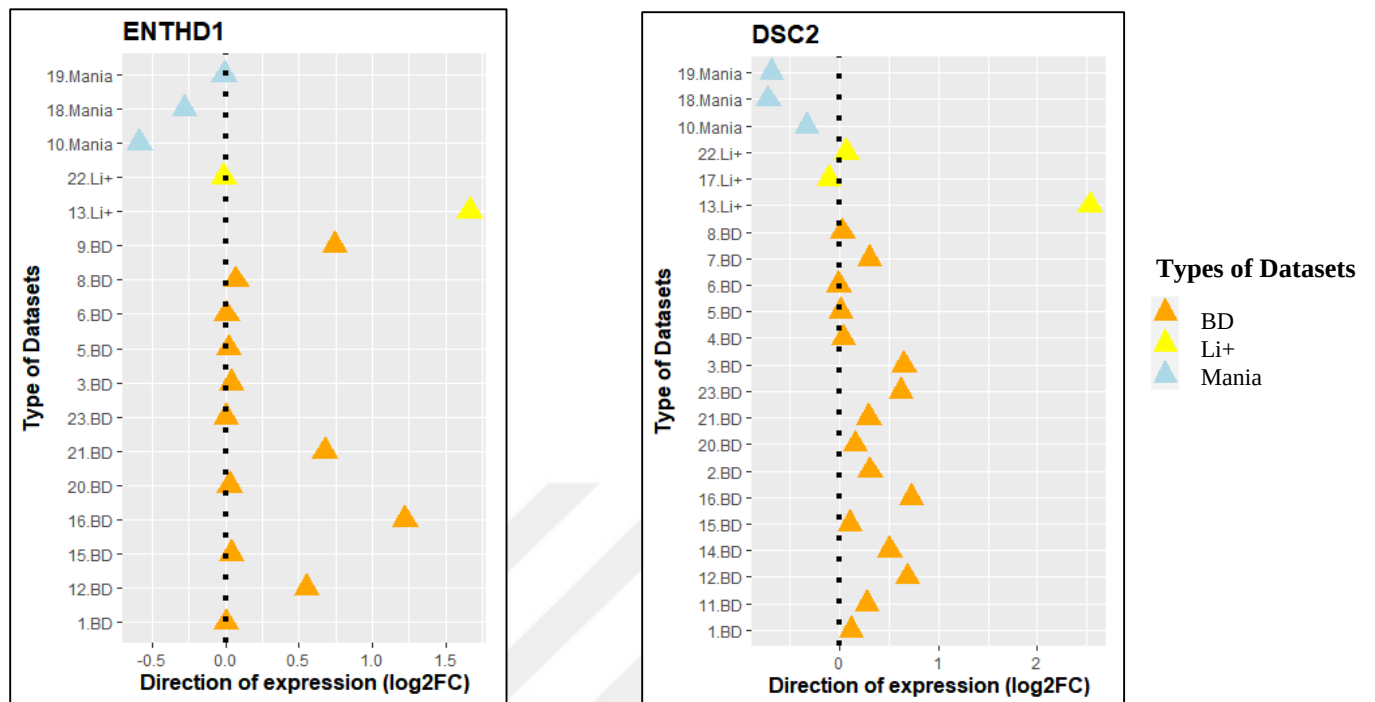


Figure 4.7: ENTHD1 and DSC2 log₂FCs – direction of expression plots. The top 2 genes; *ENTHD1* and *DSC2* are taken from **Table 4.8** to illustrate how these genes adhere to **Pattern 1: Up BD-Li; Down mania**. Direction of expression on x-axis is based on log₂ fold change (FC) values, i.e.; whether the gene is up or down regulated compared to the respective controls, in the datasets labelled on the y-axis. These are the validation datasets; the numbers refer to the datasets as they are numbered in **Table 4.3**. The black dotted line is x-intercept, to demarcate between up and down regulation, i.e.; the positive and negative values, respectively.

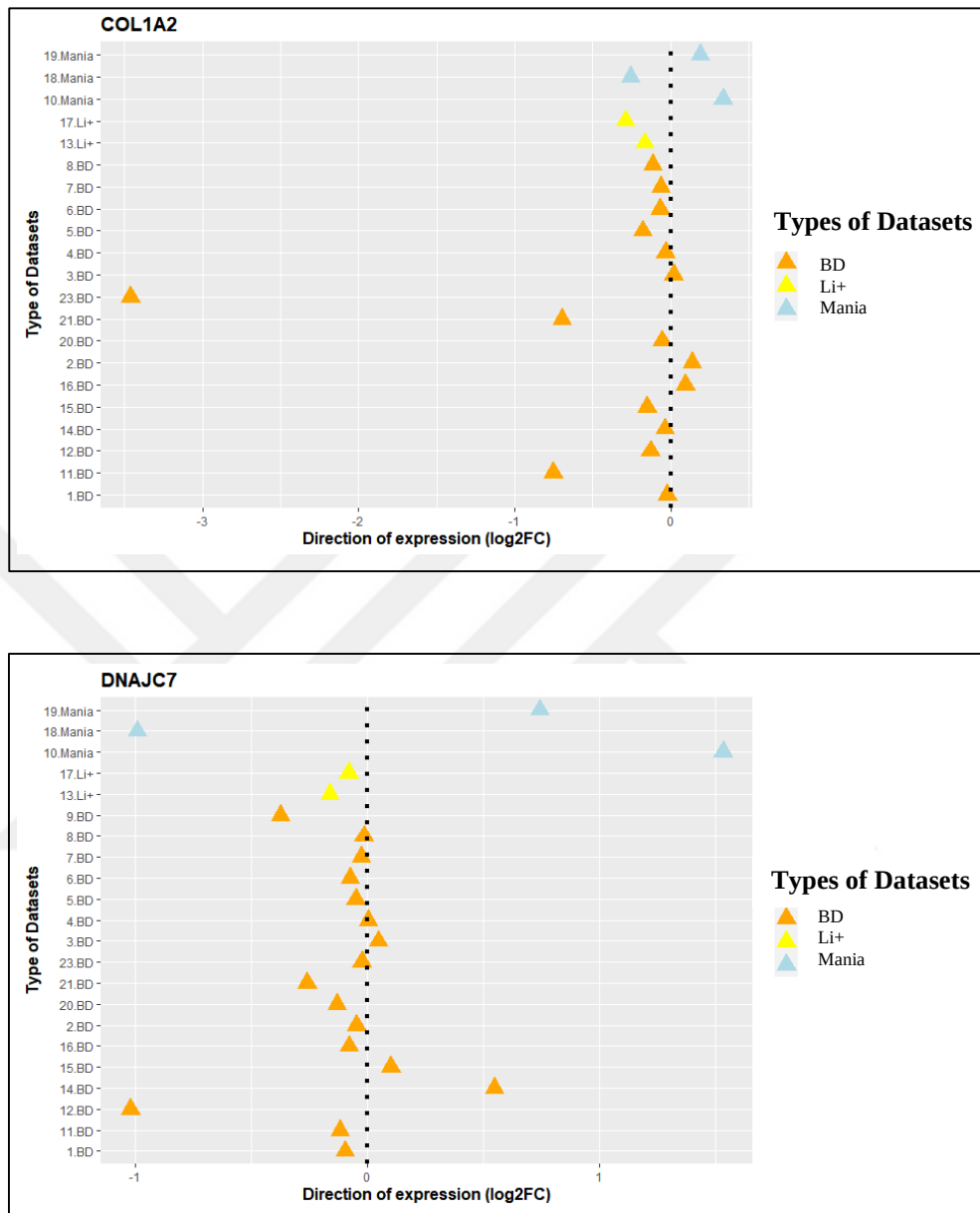


Figure 4.8: COL1A2 and DNAJC7 log2FCs – direction of expression plots. The top gene; *COL1A2* and *DNAJC7* are taken from **Table 4.9** to illustrate how these genes adhered to **Pattern 2: Down BD-Li; Up mania**. Direction of expression on x-axis is based on log2 fold change (FC) values, i.e.; whether the gene is up or down regulated compared to the controls respectively, in the datasets labelled on the y-axis. These are the validation datasets; the numbers refer to the datasets as they are numbered in **Table 4.3**. The black dotted line is x-intercept, to demarcate between up and down regulation, i.e.; the positive and negative values, respectively.

The graphs for the rest of the genes are in the **Appendix J** for pattern 1 genes and **Appendix L** for pattern 2 genes.

The next step was to perform non-parametric statistical testing to see whether significance is observed in the log2FCs of these genes between the 3 different phenotypes, i.e.; the types of datasets.

We observed that between euthymic BD datasets and lithium-treated datasets, the difference was non-significant. Whereas, between euthymic BD and mania, and lithium-treated and mania datasets, the pairwise tests were significant. Therefore, these results further validated the patterns we have observed from the GSEA results as significant.



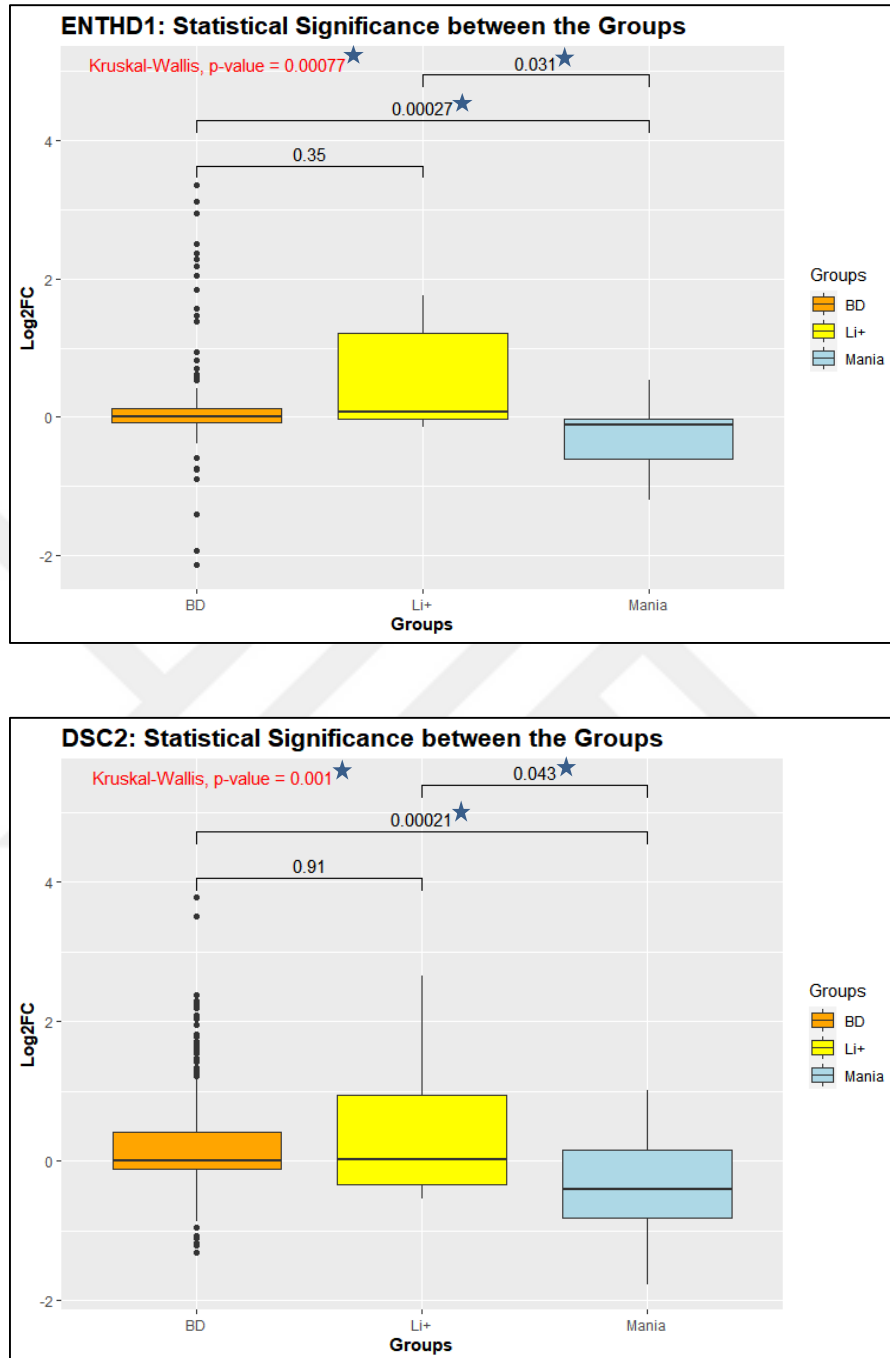


Figure 4.9: Non-parametric test for P1 genes. For the top 2 genes from **Table 4.8**, we performed non-parametric pairwise and one-way variance tests, Wilcoxon and Kruskal–Wallis, respectively. As can be seen, blue star marks significant values, p-value < 0.05. For pairwise tests, the log2FCs from mania samples were significantly different from samples of the other two groups. While, the ANOVA equivalent parameter, Kruskal–Wallis, was also significant for both genes. The values in black font depict Wilcoxon results.

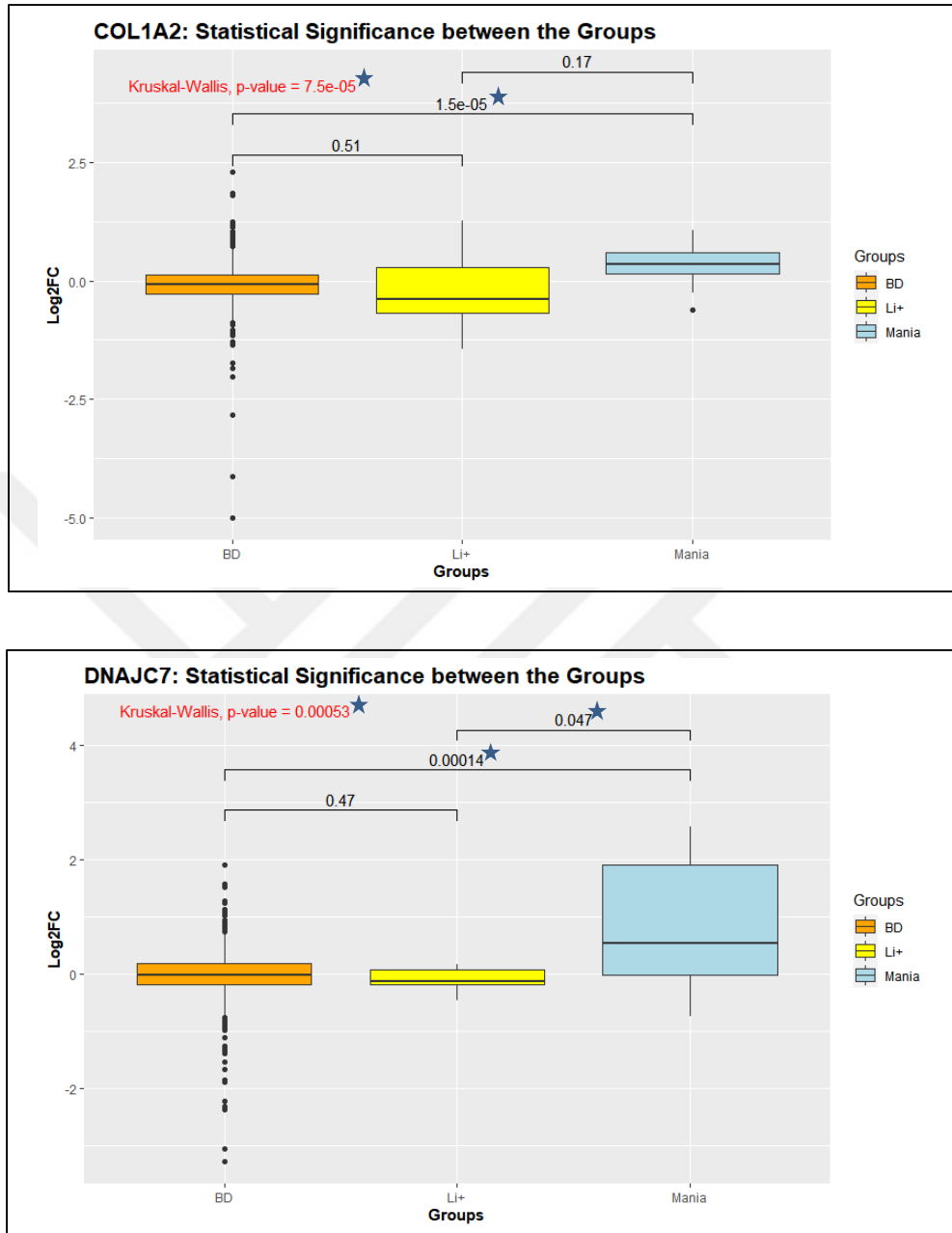


Figure 4.10: Non-parametric test for P2 genes. For the top 2 genes from **Table 4.9**, we performed non-parametric pairwise and one-way variance tests, Wilcoxon and Kruskal–Wallis, respectively. As can be seen, blue star marks significant values, p-value < 0.05. For pairwise tests, the log2FCs from mania samples were significantly different from samples of the other two groups. While, the ANOVA equivalent parameter, Kruskal–Wallis, was also significant for both genes. The values in black font depict Wilcoxon results.

The graphs for the rest of the genes are in the **Appendix K** for pattern 1 genes and **Appendix M** for pattern 2 genes.

4.4. Functional Analysis

Simultaneous to the analysis done in the previous step, with the list of genes presented in **Tables 4.6 and 4.7**, we performed functional analyses to elucidate biological functioning of these genes in concert with BD.

We performed these analyses separately for each of the gene lists adhering to patterns 1 and 2.

4.4.1. Network Analysis

For the pattern 1 genes, network analysis was done to show association of genes with diseases. As result of doing this analysis without any selection of a criteria of diseases, the output showed high degree with BD and other neuropsychiatric disorders such as schizophrenia. The subnetwork showing the identity of genes (circular nodes) directly associating with BD are shown in **Appendix N**. Pattern 2 genes didn't show any relevant disease association. The network for pattern 2 genes is shown in **Appendix O**.

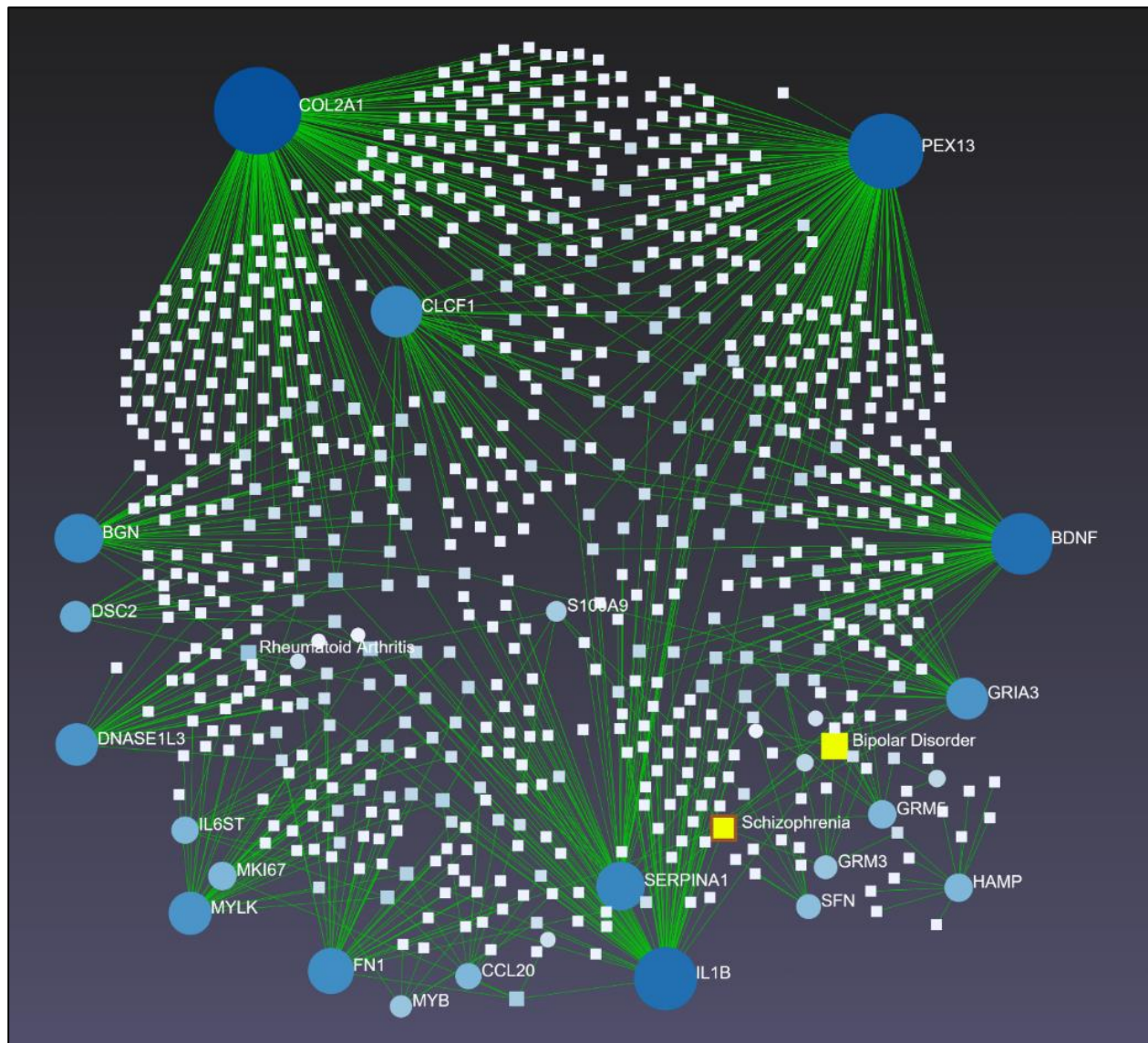


Figure 4.11: Network analysis with pattern 1 genes. This network was obtained from *NetworkAnalyst* using its ‘Gene-disease Associations’ function. It was performed using the final list of genes from **Pattern 1: Up BD-Li; Down mania**. The squares represent diseases while the circular nodes represent genes. As can be seen Bipolar disorder, highlighted in yellow (along with schizophrenia) appears in the network.

Table 4.10: Quantitative representation of network analysis with pattern 1 genes. These tables are a quantitative representation of the whole network shown in **Fig.4.11**. These consist of the Degree and Betweenness values of the all the genes from the network and the top diseases based on their degree (≥ 3) on the left and right, respectively. BD is the 2nd highest disease with degree 7.

Label	Degree	Betweenness
COL2A1	242	152305.7
PEX13	152	89852.91
IL1B	89	71258.47
BDNF	82	60642.16
CLCF1	50	26159.61
SERPINA1	43	35204.62
BGN	43	22774.46
FN1	37	20871.51
MYLK	31	19870.4
DNASE1L3	30	20727.04
GRIA3	29	14578.47
DSC2	14	8123.45
HAMP	11	7605
GRM5	11	5985.1
MKI67	11	4098.89
IL6ST	10	1642.48
CCL20	9	906.68
SFN	8	4687.19
GRM3	7	10900.36
MYB	6	3095.64
S100A9	5	2581.96
PRKCH	3	1529
WNT2	3	801.94
DSC3	2	765
CALD1	2	765
EOMES	2	765
SORCS2	1	0
DHRS3	1	0
LHX2	1	0

Label	Degree	Betweenness
Schizophrenia	8	15298.39
Bipolar Disorder	7	11681.78
Rheumatoid Arthritis	6	8409.64
Autosomal recessive predisposition	5	24442.28
Liver Cirrhosis, Experimental	5	6329.25
Mammary Neoplasms	5	3227.31
Hypersensitivity	4	1558.36
Pneumonia	4	1558.36
Colonic Neoplasms	4	940.03
Seizures	3	14489.07
Lens Opacities	3	10915.44
Hypoplastic mandible condyle	3	10915.44
Short stature	3	4630.55
Anteverted nostril	3	4191.8
Pulmonary Fibrosis	3	3659.37
Mood Disorders	3	3638.14
Lung Neoplasms	3	3429.62
Feeding difficulties	3	2740.78
Stomach Neoplasms	3	2596.44
Hyperkyphosis	3	2289.48
Limited elbow extension	3	2289.48
Hypertensive disease	3	2284.74
Neoplasm Invasiveness	3	2284.74
Dyspnea	3	2093.3
Mental Depression	3	2017.95
Depressive disorder	3	2017.95
Unipolar Depression	3	1956.7
Major Depressive Disorder	3	1956.7

4.4.2. Pathway Enrichment

Pathway enrichment yielded results with pattern 1 genes. There was a significant enrichment of inflammatory-related pathways for both the analyses done with different databases. This inflammation phenomenon formed our hypothesis as elaborated in the discussion part. The identities of genes from pattern 1 gene-list participating in specific pathways shown in **Fig.4.12** and **Fig.4.13** are displayed in **Appendices Q and R**, respectively. For pathway enrichment, the pattern 2 gene list was too small to give any significant results.

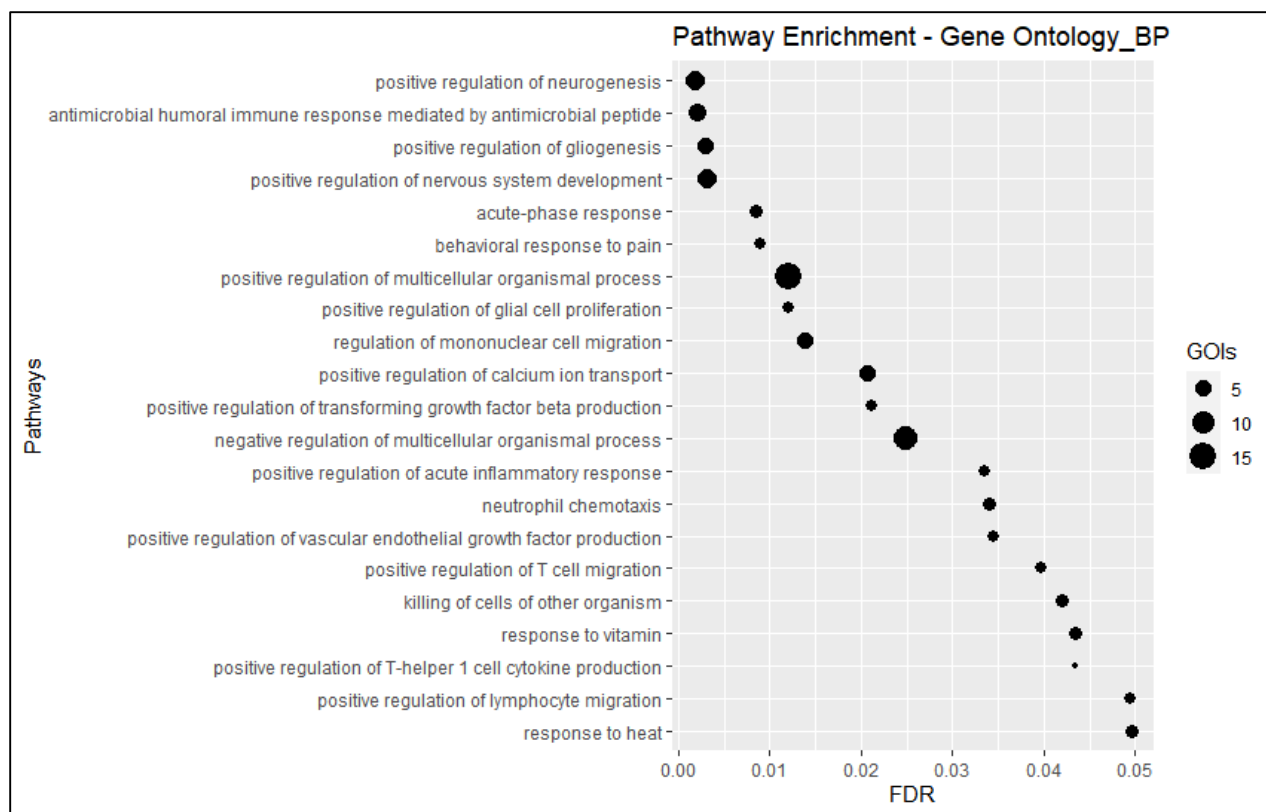


Figure 4.12: Gene-ontology (GO) pathway analysis. This was performed using the final list of genes from **Pattern 1: Up BD-Li; Down mania**. All of the pathways shown in the table were significantly enriched (FDR < 0.05). GOIs are genes-of-interest, i.e.; genes from the final list that participate in these pathways. The size of the circles represents the number of GOIs in a particular pathway.

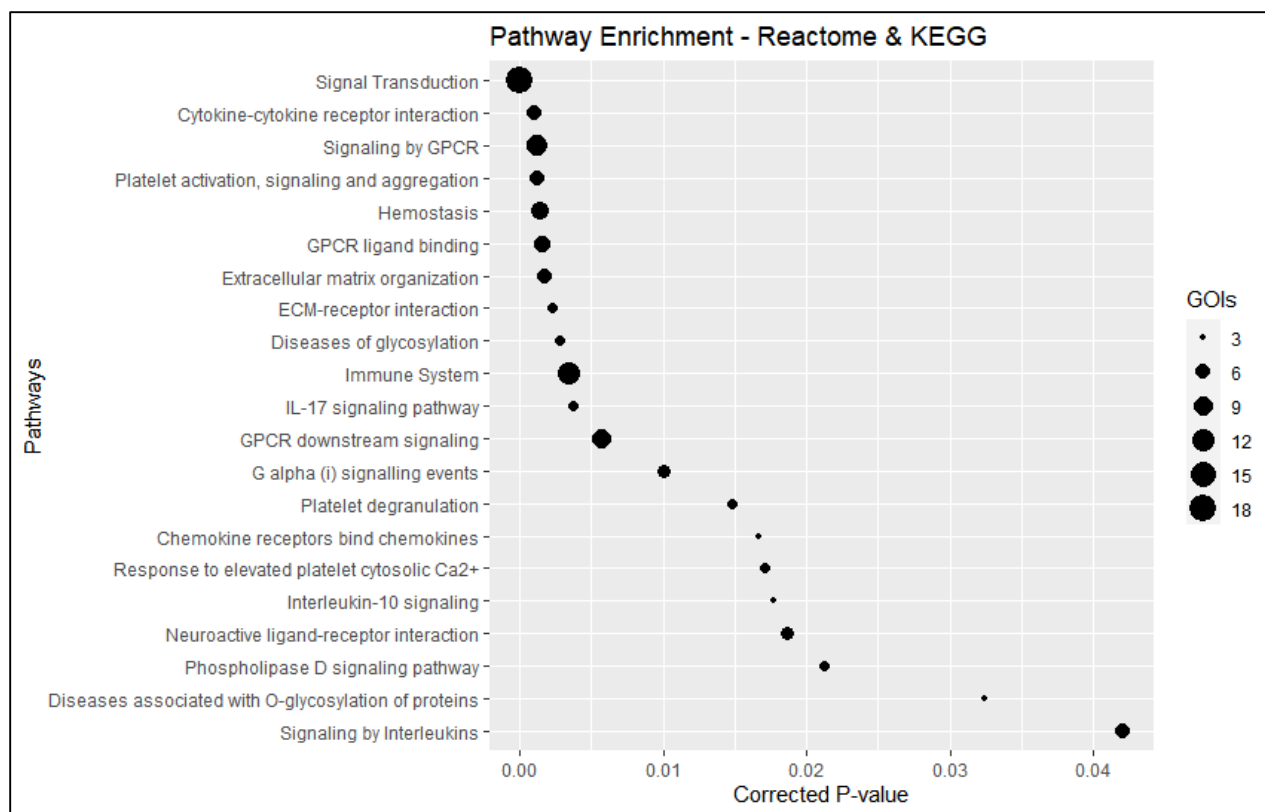


Figure 4.13: KEGG and Reactome pathway analysis. This was performed using the final list of genes from **Pattern 1: Up BD-Li; Down mania**. All of the pathways shown in the table were significantly enriched (corrected P-value < 0.05). GOs are genes-of-interest, i.e.; genes from the final list that participate in these pathways. The size of the circles represents the number of GOs in a particular pathway.

CHAPTER 5

5. Discussion and Conclusion

5.1. Bipolar Disorder: A Heterogeneous Disease

5.1.1. Heritability in BD

Bipolar disorder is a heterogeneous disease. A single-gene analysis approach so far has been unsuccessful in finding the biological basis for this disease. One would expect given the wide range of symptoms observed, that a pleiotropic gene might be causative but so far none has been found that could account for the variation in BD patients [52]. Additionally, a number of GWAS studies have been done with large cohorts of diseased individuals in order to find significant SNPs, however, the results could not be validated in labs and clinics [21]. Moreover, in heritability studies done with familial genotyping and twin studies, a strong heritability factor was found to be associated with BD. This led to the conclusion that BD is passed onto the offspring. Furthermore, the age-of-onset of the disease is early if there is a greater incidence of this disease in the parents and first-grade relatives [36, 37]. Additionally, there is a genetic association with lithium treatment as well because not all BD patients respond to the drug equally, and often these patients are segregated based on lithium responders and nonresponders [31].

Then why has there not been any gene(s), mutated or differentially expressed, discovered as solely causing BD?

5.1.2. Interplay of Genetics and Environment in BD

A few models relating to mental illnesses and the influence of the environment, bipolar disorder in this case, have been postulated in the scientific community. These include:

1. A two-hit model, in which two individuals with the same genetic makeup might suffer from two different variations of BD. One suffers from a mild form of BD, and the second, confers a worse form of BD. This is owing to the fact that one of them has been exposed to another disease such as a viral disease or prenatal infection etc. This causes the second

individual BD to be more adversely affected. While the first individual suffers from a milder form of BD.

2. A risk-resilience model, in which two individuals with the same genetic makeup can be differentially at risk from suffering from the disorder. One might be more resilient against it depending on their environmental conditions, which affect their development and the interplay of genes with their surroundings.

These models again highlight the complicated nature of diseases such as BD, where heritability is a strong factor but there have been cases where the offspring is more resilient against conferring this mental illness, this is attributed to them having a more favorable environment [53].

Therefore, the strong heritability factor for BD that predisposes individuals to the disease suggests the existence of genetic markers. However, the role of environment and its influence on the onset of disease and progression, suggest that these heritability factors are hard to isolate individually, instead a group of them might be more reflective of the etiology of the disease. Therefore, we reasoned that focusing on finding patterns between different BD phenotypes might lead to new discoveries.

5.2. Finding patterns through GSEA

5.2.1. Multi-Gene Analysis Approach

GSEA is a multi-gene analysis approach that is intended to capture a more biologically apt picture for different biological processes [38]. GSEA was created and utilized on the understanding that in biology, it is very rare that a state, especially a common disease, is associated with rare variants. Rather such multifactorial disorders, in which environment and genetics have an intimate role to play potentially have lists of genes (gene-sets) that are differentially expressed in a diseased person compared to a healthy individual. And although not fully proven, bipolar disorder seems to fit this criterion of a common psychiatric disease, associated with genes with small effect sizes [54].

5.2.2. Discovery of Patterns

BD is a polygenic disease. This fact directed us to take the GSEA route, with which we

discovered 2 patterns. Due to the variation in the disease and its cyclic nature, it was highly improbable that one single gene or even a small list of genes could be isolated as the cause of the disease. We reason that our multi-gene analysis is a much more robust way in identifying biomarkers since genetics is a strong factor in BD's etiology. Therefore, we propose the genes in **Table 5.1** as putative biomarkers of BD and its phenotypes.

5.3. Results of GSEA

5.3.1. The Two Patterns

As a result of GSEA, we obtained two patterns of gene expression, shown in GSEA results **Tables 4.4.1 and 4.4.2**:

- **Pattern 1: Up BD-Li; Down Mania**

Certain genes that were up-regulated in euthymic BD as compared to the controls were also up-regulated in lithium treated samples. This same set of genes were found to be down-regulated in manic individuals as compared to the BD and lithium treated groups.

- **Pattern 2: Down BD-Li; Up Mania**

Certain genes that were down-regulated in euthymic BD as compared to the controls were also down-regulated in lithium treated samples. This same set of genes were found to be up-regulated in manic individuals as compared to the BD and lithium treated groups (vice versa of pattern 1).

5.3.2. Finding Biomarkers

After discovering these patterns in both preliminary and validation analysis, we identified the datasets adhering to these patterns from the validation analysis' results, with a significance score of $FDR \leq 0.25$. Then the *core-enriched genes* were extracted.

In order to further highlight these genes as putative biomarkers, we did a 'doubling back' step in validation analysis. This included observing the direction of expression of these genes in *all* the datasets and not just the ones selected from GSEA results. For this purpose, graphs of the genes' fold-change values for each of the phenotypes; euthymic BD, lithium treated and mania, were generated, the fold-change ($\log_2FC$) graphs in **Fig.4.7 and Fig.4.8**.

In order to establish these genes as putative biomarkers, we performed non-parametric tests on

log2FC values; **Fig.4.9 and Fig.4.10**. We observed a significant difference in direction of expression between euthymic BD and manic, and lithium treated and manic phenotypes, and non-significant difference between euthymic BD and lithium treated. This again supports the theory that euthymic BD patients have these genes at a similar level of expression to patients who are treated with lithium, and that this expression is altered when a patient undergoes a manic episode.

After finalizing the list of core-enriched genes that adhered to the discovered patterns, we wanted to quantitate their degree of robustness. In order to do this, we utilized another parameter, we calculated the percentages of datasets in which these core-enriched genes adhered to the discovery patterns 1 and 2, **Tables 4.6 and 4.7**, respectively. Then we proceeded to select the top genes that adhered to pattern 1 in 85% or more datasets, and the genes that adhered to pattern 2 in 75% or more datasets, we obtained 7 and 5 genes from patterns 1 and 2 (**Tables 4.8 and 4.9**), respectively. Furthermore, the number of datasets did not affect the percent of datasets in which the genes adhered to the patterns, **Fig.4.5 and Fig.4.6** show this for pattern 1 and pattern 2, respectively.

5.4. Biological Analysis with Pattern 1 Genes

Biological analysis was also done with Pattern 2 genes. However, due to the heterogeneity in the genes and the smaller size of the genelist, the results were not significant. Therefore, only pattern 1 genes are being proposed as putative biomarkers for BD (**Table 4.8**).

5.4.1. Network Analysis

Simultaneous to the statistical analysis done with core-enriched genes, network analysis was performed. We observed with pattern 1 genes' network analysis, that bipolar disorder showed up as a disease with the second highest degree of betweenness, **Fig.4.11 and Table 4.10**. This result corroborated that pattern 1 genes have known associations with BD in research done previously. The highest degree of betweenness in the network analysis was with schizophrenia (SCZ), which is also a supportive finding since BD and SCZ share similar symptomatologies.

5.4.2. Pathway Analysis

Pathway analysis, done with publicly curated gene-sets belonging to three different databases;

GO, KEGG and Reactome, showed inflammatory pathways as most significant. From more general inflammatory pathways; ‘acute-phase response’ to more specific pathways; ‘positive regulation of T-helper cell cytokine production’, an inflammation associated role of these genes was observed, depicted in **Fig.4.12 and Fig.4.13**. This is supported by the literature research done on the genes This analysis led us to propose inflammation as an underlying phenomenon in BD that can be targeted for treating the disease more comprehensively.

5.5. Literature Research on Putative Biomarkers

Literature research was performed for each of the 7 genes listed in **Table 5.1** to better understand the nature of these genes in cohort with BD and their potential roles in the disease etiology. Additionally, their role in inflammation as a consequence of the findings of pathway analyses was also researched. We will go over these genes one-by-one:

Table 5.1: The putative biomarkers’ functions as defined in GeneCards (an online database).

No.	Gene Name	Type of Protein: GeneCards db.
1	ENTHD1	ENTH Domain Containing 1: This gene is predicted to be involved in the process of endocytosis and clathrin vesicle binding [55].
2	DSC2	Desmocollin 2: These are transmembrane glycoproteins that are present in cell-to-cell junctions; desmosomes [56].
3	ADAMTS9	ADAM Metalloproteinase With Thrombospondin Type 1 Motif 9: A metalloproteinase; predicted to cleave proteoglycans and involved in controlling developmental stage organ formation and inhibiting the process of angiogenesis [57].
4	STC1	Stanniocalcin 1: A glycoprotein that is implicated to have autocrine and/or paracrine function. It is phosphorylated by protein kinase C (PKC) [58].
5	IL6ST	Interleukin 6 Signal Transducer (gp130): A common cytokine-signal transducer such as interleukin 6 (IL6) and ciliary neurotrophic factor (CTNF). Involved in JAK-STAT3 and MAPK pathways [59].
6	SFN	Stratifin (14-3-3 Sigma): A cell cycle protein, specifically a checkpoint protein. It regulates mitotic translation and prevents DNA errors during the process of mitosis. It is implicated to be involved in the Akt/mTOR pathway [60].
7	THBS1	Thrombospondin-1: It is an adhesive glycoprotein. It regulates cell-to-cell and cell-to-extracellular matrix binding [61].

As a result of extensive literature research, we mention in detail whether these genes have known BD associations. Furthermore, their roles in the pathogenesis of other neurological diseases are

mentioned, that is; the ones with which BD is often misdiagnosed as they share overlapping symptomatology.

1. ENTHD1

From one of the meta-analysis studies about neuropsychiatric disorders, that included BD, a SNP; rs11570190 located in this gene was found to be associated significantly with BD and schizophrenia (SCZ). Moreover, further functional analysis indicated this gene's involvement in cognitive processes in the nervous system [64].

2. DSC2

It has been implicated to be associated with different stages of SCZ [65]. Another study done with patients suffering from autism and intellectual disabilities, copy number variations in DSC2 were found to be associated with cognitive decline. One of the translocations was found to be in DSC2, leading to its monoallelic expression. This decline was explained to be caused by a disruption in cell adhesion especially at the synapses, which causes communication issues/speech problems [66]. The point of interest from this observation in concert with BD is that manic patients often exhibit disrupted speech patterns such as slurred speech or during intense manic episodes, they display excited speech with psychotic behaviors [1, 9]. Another study analyzed the post-mortem brain samples of individuals who died by suicide vs. those who died by accidents or natural causes, DSC2 gene was found to be over-expressed in suicide samples compared to controls [67]. Suicide is the worst outcome of a BD patient in depressive state. Therefore, the association of DSC2 with this action is also of interest.

3. ADAMTS9

It also plays an important role in the CNS as a proteolytic molecule. The mechanisms have yet to be discovered, which would show whether the function is of a degenerative or of a repairing nature in the brain [68]. Structural changes in the brain in BD have been implicated owing to its progressive nature via several brain-imaging techniques, therefore, ADAMTS9, being abundantly present in the brain, is a potential strong candidate.

BD is often misdiagnosed as major depressive disorder (MDD). It might be owing to the fact that depression is a more prevalent BD phenotype than mania, and that depressive behaviors are easier to observe. ADAMTS9 was found to be up-regulated in MDD in a study done with post-mortem brain samples. In the same study, the patients had been sub-categorized into suicide vs controls, and ADAMTS9 was found to be differentially expressed. ADAMTS9, is a chemokine-related transcript, it was overexpressed in the dentate gyrus. This highlighted its association to neuro-immune alterations in the brains of the diseased individuals. In addition to its inhibition of angiogenesis, this also showcases its involvement in neuronal plasticity [69]. In another study done on babies *in-utero* about synthetic hormones, it was found that they grew up to develop psychotic features. ADAMTS9 was found to be differentially expressed in these affected individuals. It was implicated that it altered the CNS during the developmental stages. Although the psychotic individuals were all clustered together, BD was one of the disorders reported amongst them [70].

4. STC1

In a downstream analysis of transcriptomic data obtained from a cohort of BD and SCZ patients, STC1 was found to be enriched as a result of pathway analysis. The function was related to negative regulation of Ca²⁺ transport [71]. Moreover, in Alzheimer's disease, STC1 has been proposed as a potential biomarker. It has shown to exhibit reduction in oxidative stress by suppressing the ERK1/2 pathway [72].

5. IL6ST

In a monozygotic twin study, where 5 sets of twins diagnosed with different severities of autism were analyzed. *IL6ST* was one of the candidate genes to be differentially expressed. It was in concert with the neuro-inflammatory pathway to be significantly enriched in these twins. Twins with a more severe form of autism had higher fold expression of this gene [73]. Moreover, its pro-inflammatory role has been proposed in BD and SCZ patients in their CNS [74].

6. SFN

In MDD patients, SFN was found to be up-regulated compared to controls after the post-

mortem brain samples were analyzed. On the other hand, its reduced expression has been reported in SCZ patients [75, 76]. A study was done on SCZ individuals at risk for psychosis, their peripheral blood transcriptome was analyzed. SFN was one of the genes found to be differentially expressed. The YWHA family of proteins is involved in CNS as it regulates synaptic and neuronal plasticity and development. SFN, also known as YWHAS, was up-regulated in these patients, who were at risk of psychosis [77].

7. THBS1

It functions in synaptogenesis and synaptic functions [78]. This happens via astrocyte-derived THBS1. Thus, these have also been shown to repair brain damage through neuronal formation. Two THBS1's SNPs were found to be associated with SCZ in a Korean population. The implication of this is with the neuro-degenerative nature of the disease [79]. Additionally, this gene has also been linked to autism at both mRNA and protein levels, as part of the inflammatory reaction in autistic individuals [80]. In another study done to compare and classify MDD and subsyndromal symptomatic depression (SSD), THBS1 came up as a potential biomarker candidate, mostly associated with MDD vs. controls [81].

5.6. Inflammation in BD

Inflammation has been postulated as a mechanism disrupted in BD patients and a wide array of mood disorders. Recent and growing research has now found a bidirectional relationship between mood disorders and inflammation. For many diseases, inflammation is an underlying constant state. Similarly, in BD patients, increased levels of pro-inflammatory cytokines have been found. These have shown to affect the CNS structurally and functionally. Inflammation is an important part of homeostasis that is stimulated by stress, and during infection to combat the negative effects in the body. But a deregulated inflammatory process can be harmful.

5.6.1. Putative Biomarkers in Inflammation

From the putative biomarkers proposed, we further highlight genes; *ADAMTS9*, *IL6ST* and *STC1* as potential candidate genes involved in inflammatory responses in BD.

In a BD and SCZ study done to see whether sex differences could be attributed to genetics, *IL6ST*

was found to account for some of the neuroinflammatory alterations between the genders. However, irrespective of the genders, it was found that in both diseases, neuroinflammation is disrupted due to which cytokine expression is mostly elevated. Due to the fact that *IL6ST* is a common signal-transducer, its expression is correlated to these cytokines, thus, it is up-regulated in BD and SCZ compared to controls [74].

STC1 has shown to demonstrate a protective role, whereby it suppresses superoxide production. It protects the CNS by reducing the neuroinflammation in microglial cells by inhibiting cytokine production. These cytokines, which are pro-inflammatory, include $\text{TNF}\alpha$, IL-1B, IL-6 and INF- γ [82]. One of the mechanisms through which it exerts its anti-inflammatory role is by decreasing the mobility of the macrophages as being a glycoprotein, it maintains endothelial tight junctions [83].

IL-1B has been shown to be the inducer of *ADAMTS9* via MAPK and NF-KB pathway. The exact role of this gene in inflammation is yet to be elucidated but it is implicated to be one of the pro-inflammatory promoters that are stimulated in these signaling pathways [84]. Cytokines such as IL-1B and $\text{TNF}\alpha$ up-regulated this gene [85]. Anti-inflammatory molecules have been shown to down-regulate *ADAMTS9*. This supports its pro-inflammatory role [68].

A visual representation of the effect on inflammatory processes, which includes the putative biomarkers; *ADAMTS9*, *IL6ST* and *STC1* in relation to $\text{TNF}\alpha$ is summarized in **Fig.5.1**.

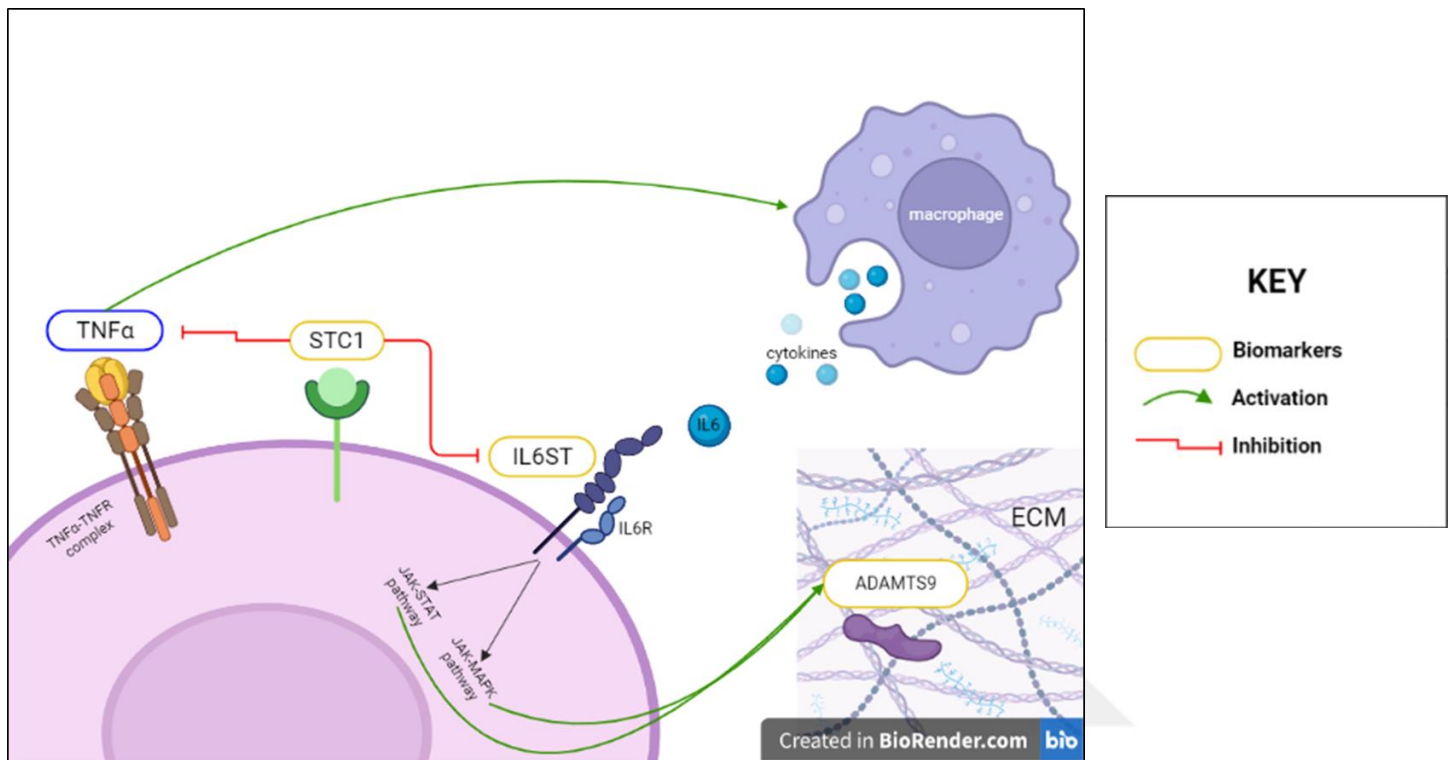


Figure 5.1: A simplified illustration of the 3 biomarkers' involvement in inflammation. This diagram depicts the functions and regulation of these molecules based on the findings of the literature research. *This diagram is just for illustrative purposes; it is not drawn to accuracy [86].*

5.6.2. $\text{TNF}\alpha$: The Common Cytokine

As can be understood from the roles of the aforementioned genes in inflammation, $\text{TNF}\alpha$, a cytokine secreted by macrophages, is involved in a plethora of processes, inflammation being one. NF-KB is the major transcription factor stimulated by $\text{TNF}\alpha$, which plays an important role in stress-induced inflammation [87]. $\text{TNF}\alpha$ comes up as a common candidate regulated by or regulating these 3 genes.

Cytokines amongst other immune agents are now known to be able to cross the blood-brain-barrier [88]. Thereby, the CNS is greatly susceptible to immune system-related changes. There are various ways in which an inflammatory response occurs, release of cytokines by macrophages, glial cells, interleukins are one of them. These cytokines, which include $\text{TNF}\alpha$ act

as chemoattractants for other immune cells such as lymphocytes and leukocytes. It can activate the production of antibodies, affect cytotoxicity, and stimulate antiviral response. After binding to its specific receptors, it ultimately regulates the processes of cellular division, differentiation and apoptosis [89].

TNF α levels have been reported as elevated in both manic patients and depressed BD patients [89, 90]. Subsequently, there have been reports of TNF α being involved in neuronal functions, and it interacts with both serotonergic and dopaminergic hormonal pathways. Moreover, it has been implicated that high levels of TNF α are predictive of a poor lithium response [91]. Cognitive decline is one of the unfortunate consequences of BD, and TNF α has been implicated in it as well [89].

Moreover, there have been numerous studies showing impairment of the hypothalamic-pituitary-adrenal (HPA) axis, which is instrumental in the immune system homeostasis. The HPA axis activity is increased, which results in increased cortisol levels. High cortisol levels have been implicated in decreasing serotonin synthesis. The deregulation of this axis through the aforementioned processes have been implicated in mood disruptions. Additionally, adverse effects on the structure of the brain and neuronal plasticity have been observed as a result of HPA axis impairment [90, 92].

The alteration in the cytokine levels as potential causation of mood swings is one of the main hypotheses for mood disorders. Therefore, a way in which to assess whether inflammation is disrupted in any way in a mood disorder can be by measuring the levels of these cytokines in the blood. Hence, as a result of our analysis, we also hypothesize that these can act as potential biomarkers.

5.7. Conclusion & Future Perspectives

The process of inflammation in association to these proposed biomarkers is also something that should not be overlooked. Especially with *ADAMTS9*, *IL-1B* and *STC1*, as we found that these genes have a common cytokine, TNF α . Therefore, more research should be done to discover whether treating mood disorders with anti-inflammatory agents is beneficial. Additionally, it has been suggested that lithium, has anti-inflammatory properties along with some other mood-stabilizers, such as valproate, which might also be the reason for their good efficacy in treating

BD [92]. As a future perspective we mention 4 anti-inflammatory agents that can be used in BD treatment in **Table 5.2** below.

Table 5.2: Anti-inflammatory drugs for BD treatment. These drugs act by reducing inflammation via different pathways. These have been shown to have positive outcomes not just in BD but also other mood disorders such as MDD and SCZ [25, 92].

	Anti-Inflammatory Drugs	Type of Drug	Direct/Indirect Molecular Targets	Effect	Results	Have been used in BD Treatment Trials?	Ref
1	Acetyl-salicylic acid (ASA)	NSAID	COX-1, COX-2, TNF α , IL-6	Decrease	Lowering cytokine levels	Yes	[26, 92]
2	Curcumin				Lowering cytokine levels, Improvement of HPA axis activity, ROS balance	Yes	[83]
3	Omega-3 polyunsaturated fatty acids (O-3 PUFA)		PGE2	Decrease	Acts as an anti-depressant	Yes	[26, 92]
4	Celecoxib	NSAID	COX-2, IL-6, IL-1B	Decrease	Lowering cytokine levels	Yes	[92]

These drugs require more research and experimental validations to be used in clinics but we propose these as good candidates for the treatment strategizing of BD patients.

We propose genes listed in **Table 5.1** as putative biomarkers for BD. These genes have been found to be significantly differentially expressed between the three BD phenotypes' euthymic BD, lithium treated BD and manic BD patients in *in-silico* analysis. We propose the patterns in which these genes are expressed in the different BD phenotypes as an important discovery as biomarkers. As with further validation of these biomarkers in patients, these can potentially:

1. Be indicative of BD patients vs. healthy control. Thus, they will help increase the diagnostic power of the disease.
2. Be indicative of lithium response. Thus, patients who are not responsive to lithium can be shifted onto other medications sooner, before they start suffering from lithium's side-effects and their treatment gets delayed.

3. Be predictive of manic episodes. These can be prevented before turning into full-blown psychotic episodes and the medicines can be modulated accordingly.

Therefore, the patterns discovered can form gene signatures that can potentially be used in clinics. Another important factor is that we propose these as serum biomarkers. Blood samples are easy to obtain and an important non-invasive method, especially for neurological disorders. In **Appendix S**, the table indicates presence of these proteins in the blood. We acknowledge the limitations of our research and our claims as these are a result of *in-silico* work only. Therefore, this necessitates *in-vitro* validation before these biomarkers can be used in the clinics.

Another hypothesis we put forth as a result of the findings depicted in **Fig.5.1** is that ratios of biomarkers' expressions can be important predictors of BD. For the 3 biomarkers in relation to inflammation, it can be observed that *STC1* is anti-inflammatory while *IL6ST* and *ADAMTS9* are both pro-inflammatory, these opposing roles can mean that inflammation is not just decreased or increased, that is; regulated uni-directionally, instead it requires a balanced regulation. Therefore, after further *in-vitro* validation, we propose that by monitoring serums levels of these biomarkers, ratios can be established, which can indicate BD in euthymic phase versus manic phase, and similarly indicate a BD patient's response to lithium.

We do recognize and believe that this biomarker approach should also be in concert with assessing other types of biomarkers. These include, neuroimaging tests, which reflect structural alterations as result of BD progression, neurocognitive tests and psychological markers in addition to the molecular biomarkers we propose in this thesis. There have only been a few studies such as one done by Le-Niculescu *et al.*, which determined a biomarker profile to track suicidality and mood states including both depression and mania [22]. We need to do much more comprehensive research to integrate biomarkers and clinical assessments. Together these can help establish a guideline and criteria to not only diagnose BD, but also give an indication of an oncoming episode and/or the responsiveness to lithium. Eventually it will help establish a more personalized mode of treatment for the afflicted.

CHAPTER 6

6. BIBLIOGRAPHY

1. Carvalho, A. F., Firth, J., & Vieta, E. (2020). Bipolar disorder. *New England Journal of Medicine*, 383(1), 58-66.
2. Phillips, M. L., & Kupfer, D. J. (2013). Bipolar disorder diagnosis: challenges and future directions. *The Lancet*, 381(9878), 1663-1671.
3. Samalin, L., Murru, A., & Vieta, E. (2016). Management of inter-episodic periods in patients with bipolar disorder. *Expert Review of Neurotherapeutics*, 16(6), 659-670.
4. IsHak, W. W., Brown, K., Aye, S. S., Kahloon, M., Mobaraki, S., & Hanna, R. (2012). Health-related quality of life in bipolar disorder. *Bipolar disorders*, 14(1), 6-18.
5. SingleCare Team. (2022, February 15). Bipolar disorder statistics 2022. *SingleCare*. <https://www.singlecare.com/blog/news/bipolar-disorder-statistics/>
6. Angst, J., Gamma, A., Benazzi, F., Ajdacic, V., Eich, D., & Rössler, W. (2003). Diagnostic issues in bipolar disorder. *European Neuropsychopharmacology*, 13, 43-50.
7. Frye, M. A. (2011). Bipolar disorder—a focus on depression. *New England Journal of Medicine*, 364(1), 51-59.
8. Awad, A. G., Rajagopalan, K., Bolge, S. C., & McDonnell, D. D. (2007). Quality of life among bipolar disorder patients misdiagnosed with major depressive disorder. *Primary care companion to the Journal of clinical psychiatry*, 9(3), 195.
9. Keck Jr, P. E., McElroy, S. L., & Arnold, L. M. (2001). Bipolar disorder. *Medical Clinics of North America*, 85(3), 645-661.
10. Angst, J. (1998). The emerging epidemiology of hypomania and bipolar II disorder. *Journal of affective disorders*, 50(2-3), 143-151.
11. Benazzi, F. (2007). Bipolar disorder—focus on bipolar II disorder and mixed depression. *The Lancet*, 369(9565), 935-945.
12. Geddes, J. R., & Miklowitz, D. J. (2013). Treatment of bipolar disorder. *The Lancet*, 381(9878), 1672-1682.
13. Bega, S., Schaffer, A., Goldstein, B., & Levitt, A. (2012). Differentiating between bipolar disorder types I and II: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Journal of affective disorders*, 138(1-2), 46-53.
14. Hirschfeld, R. M. (2014). Differential diagnosis of bipolar disorder and major depressive disorder. *Journal of affective disorders*, 169, S12-S16.
15. Schulze, T. G., Akula, N., Breuer, R., Steele, J., Nalls, M. A., Singleton, A. B., ... & McMahon, F. J. (2014). Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *The World Journal of Biological Psychiatry*, 15(3), 200-208.
16. Brus, M. J., Solanto, M. V., & Goldberg, J. F. (2014). Adult ADHD vs. bipolar disorder in the DSM-5 era: a challenging differentiation for clinicians. *Journal of Psychiatric Practice®*, 20(6), 428-437.
17. Shen, H., Zhang, L., Xu, C., Zhu, J., Chen, M., & Fang, Y. (2018). Analysis of misdiagnosis of bipolar disorder in an outpatient setting. *Shanghai archives of psychiatry*, 30(2), 93.
18. Youngstrom, E. A., Findling, R. L., Calabrese, J. R., Gracious, B. L., Demeter, C., Bedoya, D. D., & Price, M. (2004). Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(7), 847-858.
19. Simpson, S. G., McMahon, F. J., McInnis, M. G., MacKinnon, D. F., Edwin, D., Folstein, S. E., & DePaulo, J. R. (2002). Diagnostic reliability of bipolar II disorder. *Archives of General Psychiatry*, 59(8), 736-740.

20. Bauer, I. E., Galvez, J. F., Hamilton, J. E., Balanza-Martinez, V., Zunta-Soares, G. B., Soares, J. C., & Meyer, T. D. (2016). Lifestyle interventions targeting dietary habits and exercise in bipolar disorder: a systematic review. *Journal of psychiatric research*, 74, 1-7.
21. Ikeda, M., Saito, T., Kondo, K., & Iwata, N. (2018). Genome-wide association studies of bipolar disorder: A systematic review of recent findings and their clinical implications. *Psychiatry and clinical neurosciences*, 72(2), 52-63.
22. Le-Niculescu, H., Roseberry, K., Gill, S. S., Levey, D. F., Phalen, P. L., Mullen, J., ... & Niculescu, A. B. (2021). Precision medicine for mood disorders: objective assessment, risk prediction, pharmacogenomics, and repurposed drugs. *Molecular psychiatry*, 26(7), 2776-2804.
23. Sachs, G. S., Printz, D. J., Kahn, D. A., Carpenter, D., & Docherty, J. P. (2000). The expert consensus guideline series: medication treatment of bipolar disorder. *Postgrad Med*, 1, 1-104.
24. Keck Jr, P. E., McElroy, S. L., & Strakowski, S. M. (1998). Anticonvulsants and antipsychotics in the treatment of bipolar disorder. *Journal of Clinical Psychiatry*, 59(6), 74.
25. Rosenblatt, J. D., Kakar, R., Berk, M., Kessing, L. V., Vinberg, M., Baune, B. T., ... & McIntyre, R. S. (2016). Anti-inflammatory agents in the treatment of bipolar depression: a systematic review and meta-analysis. *Bipolar disorders*, 18(2), 89-101.
26. Kemp, D. E. (2014). Managing the side effects associated with commonly used treatments for bipolar depression. *Journal of affective disorders*, 169, S34-S44.
27. Carlson, G. A. (2005). Early onset bipolar disorder: clinical and research considerations. *Journal of Clinical Child and Adolescent Psychology*, 34(2), 333-343.
28. Shulman, K. I., Tohen, M., & Kutcher, S. P. (Eds.). (1996). *Mood disorders across the life span*. Wiley-Liss.
29. Vieta, E., Berk, M., Schulze, T. G., Carvalho, A. F., Suppes, T., Calabrese, J. R., ... & Grande, I. (2018). Bipolar disorders. *Nature reviews Disease primers*, 4(1), 1-16.
30. Malhi, G. S., & Outhred, T. (2016). Therapeutic mechanisms of lithium in bipolar disorder: recent advances and current understanding. *CNS drugs*, 30(10), 931-949.
31. Oedegaard, K. J., Alda, M., Anand, A., Andreassen, O. A., Balaraman, Y., Berrettini, W. H., ... & Kelsoe, J. R. (2016). The Pharmacogenomics of Bipolar Disorder study (PGBD): identification of genes for lithium response in a prospective sample. *BMC psychiatry*, 16(1), 1-15.
32. Volkmann, C., Bschor, T., & Köhler, S. (2020). Lithium treatment over the lifespan in bipolar disorders. *Frontiers in Psychiatry*, 11, 377.
33. Alda, M. (2015). Lithium in the treatment of bipolar disorder: pharmacology and pharmacogenetics. *Molecular psychiatry*, 20(6), 661-670.
34. Ochoa, E. L. (2022). Lithium as a neuroprotective agent for bipolar disorder: an overview. *Cellular and Molecular Neurobiology*, 42(1), 85-97.
35. Moreira, J., & Geoffroy, P. A. (2016). Lithium and bipolar disorder: impacts from molecular to behavioural circadian rhythms. *Chronobiology international*, 33(4), 351-373.
36. Vieta, E., Salagre, E., Grande, I., Carvalho, A. F., Fernandes, B. S., Berk, M., ... & Suppes, T. (2018). Early intervention in bipolar disorder. *American Journal of Psychiatry*, 175(5), 411-426.
37. Craddock, N., & Sklar, P. (2013). Genetics of bipolar disorder. *The Lancet*, 381(9878), 1654-1662.
38. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545-15550.
39. National Center for Biotechnology Information (NCBI). (2022). *Gene Expression Omnibus (GEO) DataSets* [Data set]. National Library of Medicine (NLM). <https://www.ncbi.nlm.nih.gov/gds>
40. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*, 30(1), 207-210.

41. The GSEA Team. (2019). *Gene Set Enrichment Analysis (GSEA) User Guide*. Broad Institute. <https://www.gsea-msigdb.org/gsea/doc/GSEAUserGuideFrame.html>
42. Broad Institute, UC San Diego. (2022). *The Molecular Signatures Database (MSigDB)* (Version 2022.1.Hs) [Data set]. Broad Institute. <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>
43. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545-15550.
44. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6), 417-425.
45. Xia Research Group. (2019). *NetworkAnalyst* (Version 3) [Computer software]. <https://www.networkanalyst.ca/>
46. Zhou, G., Soufan, O., & Ewald, J. (2019). Robert EW Hancock, Niladri Basu, and Jianguo Xia. Network-analyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*, 47(W1), W234-W241.
47. The Gene Ontology. (2022). Gene Ontology (GO) Resource. In *Gene Ontology Knowledgebase*. Retrieved December 19, 2022, from <http://geneontology.org/>
48. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29.
49. Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic acids research*, 49(D1), D325–D334.
50. North Carolina State University. (2022). *Set Analyzer* (Revision 16952M) [Computer software]. <https://ctdbase.org/tools/analyzer.go?q>
51. MDI Biological Laboratory, NC State University. (2022). *Comparative Toxicogenomics Database* (Version 2022, November 29) [Data set]. NC State University. <https://ctdbase.org/>
52. Lobo, I. (2008). Pleiotropy: one gene can affect multiple traits. *Nature Education* 1(1), 10.
53. Gordovez, F. J. A., & McMahon, F. J. (2020). The genetics of bipolar disorder. *Molecular psychiatry*, 25(3), 544-559.
54. Ioannidis, J. P., Trikalinos, T. A., & Khoury, M. J. (2006). Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *American journal of epidemiology*, 164(7), 609-614.
55. Weizmann Institute of Science. (2022). ENTHD1 Gene - GeneCards | ENTHD1 Protein | ENTHD1 Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ENTHD1>
56. Weizmann Institute of Science. (2022). DSC2 Gene - GeneCards | DSC2 Protein | DSC2 Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=DSC2&keywords=DSC2>
57. Weizmann Institute of Science. (2022). ADAMTS9 Gene - GeneCards | AT9S Protein | AT9S Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAMTS9&keywords=adamts9>
58. Weizmann Institute of Science. (2022). STC1 Gene - GeneCards | STC1 Protein | STC1 Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=STC1&keywords=STC1>
59. Weizmann Institute of Science. (2022). IL6ST Gene - GeneCards | IL6RB Protein | IL6RB Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=IL6ST&keywords=il6st>
60. Weizmann Institute of Science. (2022). SFN Gene - GeneCards | 1433S Protein | 1433S Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SFN&keywords=sfn>
61. Weizmann Institute of Science. (2022). THBS1 Gene - GeneCards | TSP1 Protein | TSP1 Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=THBS1&keywords=thbs1>

62. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., ... & Lancet, D. (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics*, 54(1), 1-30.
63. Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T. I., Dahary, D., ... & Lancet, D. (2021). The genecards suite. In *Practical guide to life science databases* (pp. 27-56). Springer, Singapore.
64. Lee, P. H., Anttila, V., Won, H., Feng, Y. C. A., Rosenthal, J., Zhu, Z., ... & Burmeister, M. (2019). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*, 179(7), 1469-1482.
65. Narayan, S., Tang, B., Head, S. R., Gilmartin, T. J., Sutcliffe, J. G., Dean, B., & Thomas, E. A. (2008). Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain research*, 1239, 235-248.
66. Wang, P., Carrion, P., Qiao, Y., Tyson, C., Hrynychak, M., Calli, K., ... & Lewis, M. S. (2013). Genotype–phenotype analysis of 18q12. 1-q12. 2 copy number variation in autism. *European journal of medical genetics*, 56(8), 420-425.
67. Galfalvy, H., Zalsman, G., Huang, Y. Y., Murphy, L., Rosoklija, G., Dwork, A. J., ... & Mann, J. J. (2013). A pilot genome wide association and gene expression array study of suicide with and without major depression. *The world journal of biological psychiatry*, 14(8), 574-582.
68. Reid, M. J., Cross, A. K., Haddock, G., Allan, S. M., Stock, C. J., Woodroffe, M. N., ... & Bunning, R. A. (2009). ADAMTS-9 expression is up-regulated following transient middle cerebral artery occlusion (tMCAo) in the rat. *Neuroscience letters*, 452(3), 252-257.
69. Mahajan, G. J., Vallender, E. J., Garrett, M. R., Challagundla, L., Overholser, J. C., Jurjus, G., ... & Stockmeier, C. A. (2018). Altered neuro-inflammatory gene expression in hippocampus in major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 82, 177-186.
70. Soyer-Gobillard, M. O., Gaspari, L., Courtet, P., Puillandre, M., Paris, F., & Sultan, C. (2019). Neurodevelopmental disorders in children exposed in utero to synthetic progestins: analysis from the national cohort of the Hhorages Association. *Gynecological Endocrinology*, 35(3), 247-250.
71. Sahu, A., Chowdhury, H. A., Gaikwad, M., Chongtham, C., Talukdar, U., Phukan, J. K., ... & Barah, P. (2020). Integrative network analysis identifies differential regulation of neuroimmune system in Schizophrenia and Bipolar disorder. *Brain, Behavior, & Immunity-Health*, 2, 100023.
72. Wang, P., Li, X. L., & Cao, Z. H. (2021). STC1 ameliorates cognitive impairment and neuroinflammation of Alzheimer's disease mice via inhibition of ERK1/2 pathway. *Immunobiology*, 226(3), 152092.
73. Hu, V. W., Frank, B. C., Heine, S., Lee, N. H., & Quackenbush, J. (2006). Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. *BMC genomics*, 7(1), 1-18.
74. Lobentanzer, S., Hanin, G., Klein, J., & Soreq, H. (2019). Integrative transcriptomics reveals sexually dimorphic microRNA control of the cholinergic/neurokinin interface in schizophrenia and bipolar disorder. *bioRxiv*, 600932.
75. Wang, Q., & Dwivedi, Y. (2017). Transcriptional profiling of mitochondria associated genes in prefrontal cortex of subjects with major depressive disorder. *The World Journal of Biological Psychiatry*, 18(8), 592-603.
76. Petschner, P., Gal, Z., & Gonda, X. (2021). Impaired mitochondrial bioenergetics in psychiatric disorders. In *Clinical Bioenergetics* (pp. 195-221). Academic Press.
77. Demars, F., Kebir, O., Marzo, A., Iftimovici, A., Schramm, C., Krebs, M. O., & Chaumette, B. (2020). Dysregulation of peripheral expression of the YWHA genes during conversion to psychosis. *Scientific reports*, 10(1), 1-9.
78. Grasby, K. L., Jahanshad, N., Painter, J. N., Colodro-Conde, L., Bralten, J., Hibar, D. P., ... & Van Rooij, D. (2020). The genetic architecture of the human cerebral cortex. *Science*, 367(6484), eaay6690.

79. Park, H. J., Kim, S. K., Kim, J. W., Kang, W. S., & Chung, J. H. (2012). Association of thrombospondin 1 gene with schizophrenia in Korean population. *Molecular biology reports*, 39(6), 6875-6880.
80. Shen, L., Zhang, K., Feng, C., Chen, Y., Li, S., Iqbal, J., ... & Zhai, J. (2018). iTRAQ-based proteomic analysis reveals protein profile in plasma from children with autism. *PROTEOMICS—Clinical Applications*, 12(3), 1700085.
81. Yi, Z., Li, Z., Yu, S., Yuan, C., Hong, W., Wang, Z., ... & Fang, Y. (2012). Blood-based gene expression profiles models for classification of subsyndromal symptomatic depression and major depressive disorder. *PloS one*, 7(2), e31283.
82. Bonfante, S., Joaquim, L., Fileti, M. E., Giustina, A. D., de Souza Goldim, M. P., Danielski, L. G., ... & Petronilho, F. (2021). Stanniocalcin 1 inhibits the inflammatory response in microglia and protects against sepsis-associated encephalopathy. *Neurotoxicity Research*, 39(2), 119-132.
83. Huang, L., Garcia, G., Lou, Y., Zhou, Q., Truong, L. D., DiMattia, G., ... & Sheikh-Hamad, D. (2009). Anti-inflammatory and renal protective actions of stanniocalcin-1 in a model of anti-glomerular basement membrane glomerulonephritis. *The American journal of pathology*, 174(4), 1368-1378.
84. Uysal, S., Ünal, Z. N., Erdoğan, S., Akyol, S., Ramazan Yiğitoğlu, M., Hirohata, S., ... & Demircan, K. (2013). Augmentation of ADAMTS9 gene expression by IL-1 β is reversed by NF κ B and MAPK inhibitors, but not PI3 kinase inhibitors. *Cell biochemistry and function*, 31(7), 539-544.
85. Ohtsuki, T., Shinaoka, A., Kumagishi-Shinaoka, K., Asano, K., Hatipoglu, O. F., Inagaki, J., ... & Hirohata, S. (2019). Mechanical strain attenuates cytokine-induced ADAMTS9 expression via transient receptor potential vanilloid type 1. *Experimental Cell Research*, 383(2), 111556.
86. BioRender. (2022) *BioRender App* (Version 2022, September 6) [Computer software]. <https://app.biorender.com/>
87. Weizmann Institute of Science. (2022). TNF Gene - GeneCards | TNFA Protein | TNFA Antibody. In *Genecards, The Human Gene Database*. Retrieved December 19, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=TNF&keywords=TNFA>
88. Leboyer, M., Soreca, I., Scott, J., Frye, M., Henry, C., Tamouza, R., & Kupfer, D. J. (2012). Can bipolar disorder be viewed as a multi-system inflammatory disease?. *Journal of affective disorders*, 141(1), 1-10.
89. Brietzke, E., & Kapczinski, F. (2008). TNF- α as a molecular target in bipolar disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 32(6), 1355-1361.
90. Ortiz-Domínguez, A., Hernández, M. E., Berlanga, C., Gutiérrez-Mora, D., Moreno, J., Heinze, G., & Pavón, L. (2007). Immune variations in bipolar disorder: phasic differences. *Bipolar disorders*, 9(6), 596-602.
91. Bai, Y. M., Su, T. P., Tsai, S. J., Wen-Fei, C., Li, C. T., Pei-Chi, T., & Mu-Hong, C. (2014). Comparison of inflammatory cytokine levels among type I/type II and manic/hypomanic/euthymic/depressive states of bipolar disorder. *Journal of affective disorders*, 166, 187-192.
92. Rosenblat, J. D., Cha, D. S., Mansur, R. B., & McIntyre, R. S. (2014). Inflamed moods: a review of the interactions between inflammation and mood disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 53, 23-34.
93. GEO2R [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2022 Dec 19]. Available from: <https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html>
94. Love, M. I., Anders, S., & Huber, W. (2017). Analyzing RNA-seq data with DESeq2. *Bioconductor*, 2, 1-63.
95. Pai, S., Li, P., Killinger, B., Marshall, L., Jia, P., Liao, J., ... & Labrie, V. (2019). Differential methylation of enhancer at IGF2 is associated with abnormal dopamine synthesis in major psychosis. *Nature communications*, 10(1), 1-12 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE112523).

96. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE66196 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66196>).
97. Abdolmaleky, H. M., Gower, A. C., Wong, C. K., Cox, J. W., Zhang, X., Thiagalingam, A., ... & Thiagalingam, S. (2019). Aberrant transcriptomes and DNA methylomes define pathways that drive pathogenesis and loss of brain laterality/asymmetry in schizophrenia and bipolar disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 180(2), 138-149 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE120340).
98. Ramaker, R. C., Bowling, K. M., Lasseigne, B. N., Hagenauer, M. H., Hardigan, A. A., Davis, N. S., ... & Myers, R. M. (2017). Post-mortem molecular profiling of three psychiatric disorders. *Genome medicine*, 9(1), 1-12 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE80655).
99. MacMullen, C. M., Fallahi, M., & Davis, R. L. (2017). Novel PDE10A transcript diversity in the human striatum: insights into gene complexity, conservation and regulation. *Gene*, 606, 17-24 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE81396).
100. Pacifico, R., & Davis, R. L. (2017). Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. *Molecular psychiatry*, 22(3), 441-449 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE80336).
101. Morgan, L. Z., Rollins, B., Sequeira, A., Byerley, W., DeLisi, L. E., Schatzberg, A. F., ... & Vawter, M. P. (2016). Quantitative trait locus and brain expression of HLA-DPA1 offers evidence of shared immune alterations in psychiatric disorders. *Microarrays*, 5(1), 6 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE78246).
102. Kim, K. H., Liu, J., Sells Galvin, R. J., Dage, J. L., Egeland, J. A., Smith, R. C., ... & Paul, S. M. (2015). Transcriptomic analysis of induced pluripotent stem cells derived from patients with bipolar disorder from an old order Amish pedigree. *PloS one*, 10(11), e0142693 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE74358).
103. Kohen, R., Dobra, A., Tracy, J. H., & Haugen, E. (2014). Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. *Translational psychiatry*, 4(3), e366-e366 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE42546).
104. Lanz, T. A., Reinhart, V., Sheehan, M. J., Rizzo, S. J. S., Bove, S. E., James, L. C., ... & Kleiman, R. J. (2019). Postmortem transcriptional profiling reveals widespread increase in inflammation in schizophrenia: a comparison of prefrontal cortex, striatum, and hippocampus among matched tetrads of controls with subjects diagnosed with schizophrenia, bipolar or major depressive disorder. *Translational psychiatry*, 9(1), 1-13 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE53987).
105. Akula, N., Barb, J., Jiang, X., Wendland, J. R., Choi, K. H., Sen, S. K., ... & McMahon, F. J. (2014). RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Molecular psychiatry*, 19(11), 1179-1185 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE53239).
106. Clelland, C. L., Read, L. L., Panek, L. J., Nadrach, R. H., Bancroft, C., & Clelland, J. D. (2013). Utilization of never-medicated bipolar disorder patients towards development and validation of a peripheral biomarker profile. *PLoS One*, 8(6), e69082 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE46449).
107. Chen, C., Cheng, L., Grennan, K., Pibiri, F., Zhang, C., Badner, J. A., ... & Liu, C. (2013). Two gene co-expression modules differentiate psychotics and controls. *Molecular psychiatry*, 18(12), 1308-1314 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE35974).
108. De Baumont, A., Maschietto, M., Lima, L., Carraro, D. M., Olivieri, E. H., Fiorini, A., ... & Brentani, H. (2015). Innate immune response is differentially dysregulated between bipolar disease and schizophrenia. *Schizophrenia research*, 161(2-3), 215-221 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE62191).

109. Arion, D., Huo, Z., Enwright, J. F., Corradi, J. P., Tseng, G., & Lewis, D. A. (2017). Transcriptome alterations in prefrontal pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders. *Biological psychiatry*, 82(8), 594-600 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE87610).
110. Hagenauer, M. H., Schulmann, A., Li, J. Z., Vawter, M. P., Walsh, D. M., Thompson, R. C., ... & Akil, H. (2018). Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis. *PloS one*, 13(7), e0200003 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE92538).
111. Ishii, T., Ishikawa, M., Fujimori, K., Maeda, T., Kushima, I., Arioka, Y., ... & Okano, H. (2019). In vitro modeling of the bipolar disorder and schizophrenia using patient-derived induced pluripotent stem cells with copy number variations of PCDH15 and RELN. *Eneuro*, 6(5) (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE116820).
112. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE93112 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93112>).
113. Mertens, J., Wang, Q. W., Kim, Y., Yu, D. X., Pham, S., Yang, B., ... & Yao, J. (2015). Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature*, 527(7576), 95-99 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE58933).
114. Witt, S. H., Juraeva, D., Sticht, C., Strohmaier, J., Meier, S., Treutlein, J., ... & Rietschel, M. (2014). Investigation of manic and euthymic episodes identifies state-and trait-specific gene expression and STAB1 as a new candidate gene for bipolar disorder. *Translational psychiatry*, 4(8), e426-e426 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE46416).
115. Iwamoto, K., Kakiuchi, C., Bundo, M., Ikeda, K., & Kato, T. (2004). Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Molecular psychiatry*, 9(4), 406-416 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE12654).
116. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE69486 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69486>).
117. Satija, N. K., Sharma, D., Afrin, F., Tripathi, R. P., & Gangenahalli, G. (2013). High throughput transcriptome profiling of lithium stimulated human mesenchymal stem cells reveals priming towards osteoblastic lineage. *PloS one*, 8(1), e55769 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE34747).
118. Walasek, M. A., Bystrykh, L., van den Boom, V., Olthof, S., Ausema, A., Ritsema, M., ... & van Os, R. (2012). The combination of valproic acid and lithium delays hematopoietic stem/progenitor cell differentiation. *Blood, The Journal of the American Society of Hematology*, 119(13), 3050-3059 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE35291).
119. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE66277 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66277>).
120. Thorfve, A., Bergstrand, A., Ekström, K., Lindahl, A., Thomsen, P., Larsson, A., & Tengvall, P. (2014). Gene expression profiling of peri-implant healing of PLGA-Li+ implants suggests an activated Wnt signaling pathway in vivo. *PLoS One*, 9(7), e102597 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE54294).
121. Saul, M. C., Gessay, G. M., & Gammie, S. C. (2012). A new mouse model for mania shares genetic correlates with human bipolar disorder. *PloS one*, 7(6), e38128 (data accessible at NCBI GEO database (Edgar *et al.*, 2002), accession GSE29417).

CHAPTER 7

7. Appendix

7.1. Appendix A

Detailed table of datasets used in preliminary analysis.

GEO DataSets Accession no.	BD status	BD Type Diagnosis	Cases	size	Controls	size	Organism	Description	Year	Gene Profiling Platform	Custom Gene-set Acquired?	References
			BD replicates 1 PBLs	25	healthy replicates 1	25	human	Gene expression profiling of peripheral blood leukocytes (PBLs) of BD patients, BD patients suffering from 1st manic episode and healthy matched controls	2013	Microarray	Yes	106
GSE46449	Yes	BD 1	BD replicates 2 PBLs	18	healthy replicates 2	14						
GSE12654	Yes	not mentioned	BD post-mortem prefrontal cortex	11	healthy controls	15	human	Gene expression profiling of human prefrontal cortex, BD patients' post-mortem samples	2008		No	115
GSE46416	Yes	not mentioned	BD euthymic phase blood samples	11	healthy controls	10	human	Gene expression profiling of blood samples from euthymic and manic BD patients	2014		Yes	114
GSE69486	Yes	not mentioned	BD skin fibroblasts	10	healthy controls	2	human	Gene expression profiling of skin fibroblast cells from BD patients	2015		No	116
GSE34747	No	-	Li+ hMSCs	3	Li- hMSCs	3	human	Gene expression profiling of lithium stimulated mesenchymal stem cells	2013		No	117
GSE35291	No	-	Li+ HSPCs	3	Li- HSPCs	3	mouse	Gene expression profiling of 7-day cultured hematopoietic stem cells in the presence or absence of lithium	2012		Yes	118
GSE66277	No	-	Li+ Hipp	5	Li-Hipp	5	rat	Gene expression profiling of rat brain following chronic treatment with antipsychotic drugs or mood stabilizers	2015		Yes	119
			Li+ Cortex	5	Li-Cortex	5					No	
			Li+ Striatum	5	Li-Striatum	5						
GSE54294	No	-	Li+ PLGA day 7	6	Li- PLGA day 7	6	rat	Gene expression profile during peri-implant healing of poly(lactic-co-glycolic acid) (PLGA) implants with incorporated Li+, while PLGA without Li+ was used as control	2014		No	120
			Li+ PLGA day 28	6	Li- PLGA day 28	6						
GSE29417	No	-	manic mouse strain hippocampus	6	control strain of mice	6	mouse	Gene expression profiling of hippocampus from Madison (MSN) mice strain displaying generationally-stable mania	2012		Yes	121
GSE46449	Yes	BD 1	1st episode replicates 1 PBLs	3	BD replicates 1 PBLs	25	human	Gene expression profiling of peripheral blood leukocytes (PBLs) of BD patients, BD patients suffering from 1st manic episode and healthy matched controls	2013		Yes	106
			1st episode replicates 2 PBLs	3	BD replicates 2 PBLs	18						
GSE46416	Yes	not mentioned	BD manic phase blood samples	11	BD euthymic phase blood samples	11	human	Gene expression profiling of blood samples from euthymic and manic BD patients	2014		Yes	114

7.2. Appendix B

Table of up-regulated gene-sets used in preliminary analysis.

Li-treated datasets		Euthymic BD datasets		Mania dataset	
GSE35291	GSE66277	GSE66277	GSE46449	GSE46416	GSE29417
MPO	BAIAP2	BTG2	SPTLC2	KCNH3	HPS3
SPNS3	RGD1561413 /// RGD1565622	ELN	PHYKPL	GRINA	A530054K11RIK
PDLIM1	ARC	KCNJ6	GPR153	BTG3	INVS
CTSG	GAS5	CXCR5	GJC1	FKBP9	FBXL12OS
GPR65	ITSN2	LOC102554740 ///	ARVCF	SLPI	QTRT1

		LOC103693750			
E130012A19RIK	P4HA2	TEAD2	CYP2C9	CNPY4	KLHDC7A
TMEM176B	RIMKLA	RT1-CE12	CHRD	FIZ1	HPGD
SAA3	EIF4E3	NR4A3	ADD2	EZH2	GM6034
HP	ITSN2	ZFP423	RAB27A	SLPI	RNF187
CD97	PAPOLA	LOC100910319	CADM3	PLOD2	CAT
MMP14	NLGN2	TRIM21	OLFM1	C8ORF48	HSPE1
IGFBP4	STEAP4	AIF1	KCNMB2	PTGFRN	TAF1D
MS4A3	MPHOSPH8	H19	PDE8A	SLPI	GIT1
MYL10	FGL2	PLBD1	MIOX	MAMDC4	HFM1
MRPS28		H6PD	MAP1LC3C	MCM10	EPOR
GFI1		MLPH	BOK	UHRF1	IFFO2
PRPS1		BIRC5	WDR72	DLGAP5	PLOD3
RPS24		HOXD11	ANO7	CYP1B1	9930111J21RIK1/ //GM5431///993 0111J21RIK2
HBA-A1		MAU2	LSAMP	MITF	EMCN
EXOSC2		PDCD4	OTUD7B	PVALB	SMARCA1
EBNA1BP2		LGALS3	EVPL	CD109	LRRC25
THY1		CSF1	FBLN1	CASC5	RGL1
PRMT5		PTGFR	OR1A1	DLG4	GM15772///RPL2 6
TIMM8A1		ADAMTS7	PIP5K1A	MBOAT2	HSD17B7
CCL7		TRPV6	LINC00112	PFN2	9930111J21RIK1/ //GM12185///99 30111J21RIK2
OLFML2B		PADI2	RNF146	RAD21L1	PDGFRA
GNPTAB		FAM219B	LMX1A	HIST1H3G	SPINK10
GPR183		MCM3AP	IFITM10	DSC2	GREM1
PPA1		NFATC4	GPR124	TOP2A	TERC
SLC11A1		PNLDC1	CYBA	SLC34A2	CMKLR1
MMP2		UOX	DBF4B	CYP1B1	PSAT1
TPPP3		AQP12A	NUTM2A-AS1	DLGAP5	TIMMDC1
ABCB1B		LOC100365450	PKD1P1///NPIPA1	GLDC	CD59B///CD59A
CST7		LOC103691267	YY1AP1	PID1	TRAPPC2L
HSPD1		RAC2	DKFZP434A062	AMZ2	
SLPI		OLR1867	LAT2	DCUN1D4	
BHLHB9		CDC20	CMTM3	TET1	
HSPH1		LOC100911090 /// LOC102555319	KRT78	KCTD3	
SHMT2		ANPEP	RASAL2	BAIAP3	
PRF1		BRD1	MTF1	TP63	
HIRIP3		ATP2A1	DKFZP564H213	STARD8	
CRIP2		LDB3	DPP6	CLSPN	
NCL		GEM	MYH11	MND1	
BCAT1		CRYGS	CYP1B1	CDKN3	

PRTN3		BCL2L14	CA7	PLP2	
PDLIM4		NRADD	GREM2	CYP1B1	
SMPDL3B		HIC1	TMEM72	CDC42BPA	
PER2		CALML5	NADSYN1	COX6C	
GJA1		LOC103695179	FLJ32742	ASPM	
PHGDH		TMEM229A	CYP3A7	PLP2	
GSPT2		NMUR1	LOC283177	PRC1	
PPAT		RT1-BA	DUOXA1	CRYBB1	
PA2G4		LOC100911950	CLSTN3	MELK	
BZW2		LOC103694460	CARHSP1	TOM1L1	
TIMELESS		CYR61	FBXL19-AS1	NID1	
PF4		SLF2	NAV1	CYP1B1	
RFX2		GATA1 /// NEWGENE_2663	DRAXIN	TDRD9	
CFB		LOC302999	NEUROD2	RGL1	
SSSCA1		C1QTNF4	THY1	AP4M1	
WDR74		IFI44L	OR3A3	PLP2	
ENKUR		NCK2	LOC100286895	RAB23	
TCOF1		CH25H	CLASRP	UBAP2	
MYBL2		KATNBL1	PITX3	CYP1B1	
NKG7		ZIC2	TP53AIP1	CUL7	
MSANTD3		ITPR2	NCKAP1L	MMP19	
GPR84		RCC1	STC1	PTGFR	
HSPE1		RSU1	BSN	KIF14	
PLAC8		RANGAP1	LINC00672	AXL	
GPR97		DCTD	HHIPL2	CCDC71L	
NOP2		DYSF	PCDHGA1	KCTD11	
RBM38		BEST2	SAMD1	CRLS1	
GART		ITGA6	SLC5A9	ARHGAP6	
2310033P09RIK		MYO1G	SCRT1	FBN2	
1110017D15RIK		SPRED3	LINC00960	PI15	
CD69		CIITA	SYT7	SCAMP5	
POLR3H		FAAP100	CLDN3	MDM1	
LYAR		MLXIP	PPP1R14D	TPX2	
TONSL		TMEM245	SGSM1	RANBP17	
TIMM10		LOC100912870 /// RUP2	PCDH1	MBOAT2	
CPA3		LIPF	FAM166A	CFTR	
SRM		DPEP1	CD47	HFE	
ERICH5		RGS9	PROX1	NUDT4	
SERPINF1		GPR1	TBX10	S100A8	
ABCF2		CFDP1	RNF207	ADAMTSL4	
GSTT2		GFI1B	NCOR2	UBAP2	
SOD2		ITGA1	SULT4A1	ZNF614	
CCL2		OPRM1	CLN6	MGST1	

RCC1		PAX8	HLA-DOB	RAD51C	
ELANE		ACRBP	STAU2-AS1	NEIL3	
BOP1		CLRN3	LOC100505650	LRCH3	
NDC1		TDRP	NBPF20///NBPF1 0///NBPF1	LRP1	
MUC13		HCLS1	CHI3L1	HADHB	
		SLC9A2	POM121L12	LRP1	
		TPBPA	ADAMTSL1	GPBP1L1	
		PSME1	ADAM30	DLGAP5	
		LOC102555574	FNDC5	C14ORF2	
		ZFP385C	AKAP5	TYMS	
		NGB	FTCD	C2ORF88	
		FAM83A	SPG20	STAB1	
		FAM46A	TRUB1	FBN2	
		SULT5A1	LCN6	POMT1	
		PMCH	MAT1A	RANBP17	
		HOXB8	CEP89	CYP1B1	
		TXNIP	ABCC11	PRELID2	
		NPFF	FBLIM1	ERI2	
		HAND1 /// LOC102557319 /// PSG19	PAX8	HP	
		SRCIN1	LOC100133920/// LOC286297	RAB3IP	
		EDNRA	IQCH	CD109	
		IFNAR1	U2AF2	DLGAP5	
		ABCC3	ZFYVE9	HMMR	
		GYLTL1B	VAT1L	CASC5	
		CYP8B1	CDH24	RPL31	
		IL9R	OR7C2	ALS2CR12	
		LTBP2	GRK5	FBN2	
		BPIFA2F	LENEP	TMEM40	
		ESRP1	TOP1	PTPRA	
		SPATA1	CDC42BPA	LTBP1	
		LOC103691310	C1ORF61	EZH2	
		LOC500300	NRSN2	PER3	
		C2 /// CFB	ALPK3	GPBP1L1	
		LOC56764	FAM161A	KIAA1841	
		CXCL3	MALAT1	CASC5	
		GTSE1	KRT31	TPX2	
		DNAAF1	C2ORF82	SLC7A11	
		DDX19A /// DDX19B	C13ORF41	LRTOMT	
		KMT2D /// PRKAG1	DGCR5	ZNF236	
		EGLN1	HNRNPM	BRCA2	
		TBXA2R	SCAF11	MXI1	

		PRLR	PVRL4	LRP1	
		RETNLA	TMEM132A	ANLN	
		EVC	EVPLL	LEPR	
		MILR1	CCDC80	KIAA1841	
		ASGR2	POU5F1B	ESPN	
		AMMECR1L	LINC01477	NUP62CL	
		HAL	LINC00881	EYA3	
		LOC102546843	ADRBK1	NDUFA4	
		ARHGAP28	GNN	SLC25A48	
		C5AR1	PTPRC	FBN2	
		PDE6G	AP2B1	HIST1H3H	
		GUCA1B	MAGIX	RAB34	
		RELB	GALE	LRP1	
		SCARF1	SLIT1-AS1	KIAA0101	
		CREB3L3	CADM4	RNF11	
		LOC102549724 /// LOC102549828 /// LOC103691626 /// LOC103695130	TANK	PLOD2	
		FNDC1	ELFN2	CYP1B1	
		IL18BP	RCE1	MYH10	
		AKR1B7	LOC100506667	CASC5	
		PTPRV	SEZ6L2	CREG1	
		ZHX3	PDILT	KIF14	
		CHCHD7	C17ORF104	GLE1	
		GRM3	APCDD1L-AS1	ALS2CR12	
		GHRHR	LOC100507675	MYL6	
		FUT4	LINC01016	GDPD3	
		ADGRD1	TTLL10	TEKT2	
		LRCH4	C9ORF85	DMC1	
		KRT77	GPRIN2	DCLRE1B	
		EHD2		SOX6	
		CCDC71		CLSPN	
		CHRNA		ACTC1	
		PELI3		ASPM	
		PRL7D1		LTC4S	
		TFF3		PI3	
		HBQ1B		GPR152	
		SLC27A2		PLXND1	
		UCP3		TOP2A	
		PRL4A1		CASC5	
		CYP4A8		EEF1B2	
		PFKM		MYH4	

	GGT1	MED12L	
	SMIM24	CD99L2	
	LAMA1	TSPAN7	
	TRPM1	MBOAT2	
	RELA	DSC2	
	IGJ	AFF2	
	LENG8	STAB1	
	CNGA3	NCAPG	
	NGFRAP1	ANLN	
	RANBP1	KIAA1841	
	ZFP704	RGS6	
	LENG9	LRCH3	
	SGSM2	ANLN	
	UNC5CL	RBP7	
	ARID1B	KIF15	
	RGD1565166	NCAPH	
	SERPINF2	WDHD1	
	NR5A2	EPB41L4A	
	SLC6A13	RSAD2	
	CAMK2G /// NEWGENE_62180 2	PLOD2	
	LOC683313	FBN2	
	FSD2	TBCA	
	ZG16	CYP1B1	
	CES2C /// LOC100910040	FAM105A	
	OLR1493	MBOAT2	
	GUCY2C	TEX9	
	HOXD4	CASC5	
	ANXA3	EEF1B2	
	SEC14L3 /// SEC14L4	HIST1H3C	
	TMEM243	MBNL3	
	CRYGD /// CRYGE /// CRYGF	PDCD11	
	LOC102551046 /// LOC102557219 /// PHF11 /// PHF11B /// SETDB2	ARL4A	
	VOM1R90	LRP1	
	QTRT1	COL24A1	
	HJURP	NCAPG	
	FKBP2	DMC1	
	BNIP1	DIAPH3	
	DUXBL1	BRIP1	

	LOC102547242	SOCS3
	CASZ1	GRIN2B
	LOC303566	GAPT
	FCRLB	SOX6
	ASPDH	PLOD2
	LOC102548848	C1QTNF6
	NPPB	TMEM56
	AGTR2	LRP1
	GSAP	BUB1
	KDM4B	CENPF
	TAF12	DROSHA
	DHRS11	LRP1
	CDC45	FAM13A
	BMPR2	FBN2
	AQP1	LEPR
	PTPRCAP	SKA3
	LOC100911669	RSAD2
	TPM3	AFF2
	KCNE3	NUSAP1
	ITGA9	TOP2A
	LOC100911030	FOXO3
	DHX37	PTPRA
	SLC38A7	RPS27L
	WNT6	RBBP8
	RCSD1	CDC42BPA
	FBR5	BEND2
	ZMYND10	XK
	RELT	PRR11
	LOC100910809	GYPE
	CDC14B	PFDN5
	SEMA6C	NCAPG
	GUCY2D	ZWINT
	HAAO	MBNL3
	SOX7	PLA2G7
	WISP2	GLRX5
	PFKFB3	NID1
	TEAD4	CYP1B1
	FXD3	RBM23
	GULP1	SLC18A2
	DUSP13	KIF14
	STAT5A	UBE2E3
	LOC100909928	LONP1
	FOXO2	NID1
	CYP1B1	FBN2

	TMEM101	DLGAP5
	ARHGEF12	CCNB1
	CD99 /// LOC103694460	FAM13B
	TNNI1	PVALB
	IAPP	TOP2A
	RGD1310262	KIF15
	FAM83H	COPS9
	P2RY4	RANBP17
	KLRC2	ATXN2L
	CHRNA6	ANXA3
	RHOD	ZNF236
	KCNQ1	CASC5
	SCYL2	MELK
	PROB1	DBI
	LOC360231	CISD2
	SP100	MELK
	FLT4	WASF3
	PRIMPOL	PGM2L1
	LHB	STIL
	SRPX	ASPM
	KRT36	LARP1B
	SLC6A4	NCAPG
	SLC01A5	TOP2A
	SCNN1B	CENPE
	LRRC75B	MBNL3
	SLC30A2	TOP2A
	MLLT1	ANLN
	ADAM22	RPL30
	CSF2RB	ECT2
	SLC25A41	NFIA
	MS4A8	RPL35A
	ATG5	SGO2
	COX11	ZNF569
	SLC27A4	GNG11
	CASP12	GBA
	TFAP2E	MXD3
	PRPH2	CYP1B1
	RAB43	DMC1
	SLC6A18	JCHAIN
	RPS6KA4	RAB23
	NADK	DNMT1
	HARBI1	SIRT5
	SERPINB6B	BUB1
	LOC102549803	PTPRA

		/// LOC102556148			
		CLEC4F		LRP1	
		LRBA		HIST1H3H	
		PLXNB1		SKA3	
		TAOK3		PLOD2	
		SYTL1		SEC62	
		RT1-DB1		EFCAB13	
		SLFN13		HIST1H2AG	
		EPCAM		EGF	
		PKLR		RBP7	
		CYTH1		RABGAP1L	
		MARVELD1		CASC5	
		NR0B2		CASC5	
		GAS2L3		KIAA0101	
		FCGR1A		ANAPC15	
		RBM10		ENKUR	
		ZFP36L2		FBN2	
		CCDC24		FBXW7	
		LOC102552895		KIF14	
		ABCD1		DMC1	
		RFFL		KIF4A	
		LOC102553732		MITD1	
		WWC1		HEBP1	
		TBC1D10C		AURKA	
		SLC25A45		CTNNAL1	
		ZBTB42		FBN2	
		AFM		GPR15	
		SH3TC2		USP12	
		KCNK2		POTEE	
		PDLIM1		CTNNAL1	
		APOB		KIF20B	
		PTRF		DTL	
		SGTB		SGO1	
		TRAF3		DSC2	
		MMS22L		STAB1	
		SENPA8		RIDA	
		SBF1		DLGAP5	
		SMAD6		WDHD1	
		HGFAC		HP	
		PTTG1		FCGR1A	
		CYP2C23		LRP1	
		SHOX2		F5	
		S100A11		MELK	
		BIN2		CLIC2	

	KRT78	PUS3	
	REXO1	NUP62CL	
	TPBGL	GPATCH1	
	ZFP455	LRP1	
	ITGAL	RPL31	
	TMEM51	RPL35A	
	MAPK8	ATP23	
	CIDEB	NCAPH	
	PITX2	CDK8	
	SSPN	EIF3J	
	RASIP1	RCHY1	
	TBC1D8	SIAE	
	ZFP316	FBN2	
	ZFP36	DNAJC15	
	LOC100365008	SPTA1	
	CYP2A1 /// CYP2A2 /// CYP2A3	MINPP1	
	PRKAB2	TBCA	
	HOXC10	ISCA1	
	JAM2	LRP1	
	MYL2	HMGN3	
	CRB3	KIF14	
	APOC4	BCL2L14	
	ATL3	LRP1	
	IFIH1	PDCD11	
	ARRDC1	CENPI	
	TACC3	HMGN5	
	GALNT2	EEF1B2	
	CDKN2C	CEP70	
	LHCGR	FANCI	
	SAMD4B	KIF11	
	YAP1	UBAP2	
	FANCC	RBP7	
	NEDD1	SVBP	
	MYOC	FAM210B	
	DENR	ALS2CR12	
	OPRK1	CDC42BPA	
	SPHK1	SERPING1	
	NKX2-5	PFDN5	
	LOC100912451 /// OLR857	TOP2A	
	GNRH1	CNTLN	
	CNGB1	FBXW7	
	DHH	TRIB3	

	SIRPA	ASH2L	
	TK1	DNAJC6	
	HSF4	BPI	
	LOC102548822	MELK	
	PAPOLB	CCNB1	
	LOC100911486 /// MEGF6	DPP3	
	LIF	SLC26A8	
	HIST1H2AA	FBN2	
	FOXD4	LRP1	
	ACER1	RGS6	
	ZFP623	MBNL3	
	HNF1A	MCM10	
	LOC290595	FBN2	
	NCOR2	FBN2	
	SCGB2A2	CMAS	
	ITIH1	TOP2A	
	ZFP53	PPA1	
	SPRR3	COL6A3	
	ACTG1 /// LOC100361457 /// PTEF	STIL	
	GUCY2E	GPR15	
	INSRR	DMC1	
	HPS1	RBP7	
	THBS2	MCM10	
	IFIT1	CALD1	
	ANKLE2	CSTA	
	LOC102551341	SERPINB10	
	SDCBP2	LRP1	
	SLBP	LRP1	
	IGF2	LRP1	
	LIPE	APLF	
	FGF6	S100A8	
	NCKAP5L	BPI	
	PRCP	MGLL	
	TMED6	PLXND1	
	ZFP438	S100A8	
	KCTD5	CEACAM6	
	FKBP8	FBN2	
	FAM134B	KIF4A	
	KCNT1	GYPA	
	LOC100910487 /// ZNRD1A51	SLIRP	
	IVL	LTBP1	
	MUC2	ANAPC15	

	TIA1	RRM2	
	LIMS2	LRP1	
	RGD1566029	PCNT	
	RGD1563354	LRP1	
	CYP4A3	CASC5	
	NFAM1	KIAA1841	
	ZFP202	RPIA	
	ZXDC	TOP2A	
	SECTM1B	FCGR1A	
	TBX2	RIDA	
	SPIN4	LRP1	
	FBXL8	JCHAIN	
	FCRLA	LEPR	
	SLC28A3	CASC5	
	DHFR	FBXO10	
	CELA2A	PGM2L1	
	XKR5	GCSAML	
	MRAP	SLIRP	
	CRYGD /// CRYGE	KIAA0101	
	GAS5	ATAD2	
	KCNK3	KIF23	
	NPFFR2	ITGA1	
	CCDC13	CASC5	
	FMO4	RPS24	
	AHR	ISCA1	
	UNCX	PPA1	
	UMOD	XK	
	TRIM26	ALS2CR12	
	PTGER3	MBNL3	
	LOC688415	CASC5	
	EIF4G1 /// LOC100912571 /// LOC682402	TYMS	
	LOC100365542	KIF14	
	LOC685152	KIF14	
	PHF7	LRP1	
	EPS8	TFDP1	
	CXCL9	CXCL5	
	LOC103694545 /// LOC103694550 /// LOC103694552 /// LOC103694554 /// LOC103694555	LRP1	

		/// SRY			
		KIF1B		NUP62CL	
		GRM8		SGO1	
		CDSN		SERPINB10	
		HIST1H1T		LRP1	
		SLAMF9		LPAR1	
		GZMK		RGS6	
		SMARCAD1		TOP2A	
		RGS12		NEK1	
		CPM		WDR78	
		TMEM132E		LRP1	
		RMDN2		PFDN5	
		ZFP36L1		ARHGAP6	
		LOC102552530		FBN2	
		SELL		LETMD1	
		DSC1		DBI	
		SMCP		RBBP8	
		LOC103691629		LTBP1	
		RBM24		MINPP1	
		CCDC74A		CEP70	
		CYBB		LRP1	
		RARA		MAOA	
		KREMEN1		PIWIL4	
		SCARF2		LRP1	
		KLHL30		MKI67	
		RGD1308775		MAD2L2	
		BCL2L10		SERPINI1	
		HOXD8		FECH	
		DRD4		AIF1	
		CACNB1		NUP54	
		NSUN5		BARD1	
		RGD1560289		EFCAB13	
		WDR95		TEX9	
		CYP4F17 /// CYP4F5 /// LOC100912471 /// LOC691320		LRP1	
		SLC52A3		DMC1	
		STOM		LRP1	
		LOC102548586		SKA3	
		NAV2		SKA3	
		PAPD7		RPL35A	
		PROK1		FBN2	
		SSFA2		LEPR	

7.3. Appendix C

Table of down-regulated gene-sets used in preliminary analysis.

Li-treated datasets		Euthymic BD datasets		Mania dataset	
GSE35291	GSE66277	GSE66277	GSE46449	GSE46416	GSE29417
RETNLG	PKIB	DCAF6	ZNF585B	4632415L05RIK	RAD51B
NT5E	CLEC4F	LINGO2	TNFRSF17	H2-T9///H2-T22///H2-T10	FAM53B
P2RY6	SPHK1	PPP1R9A	C1ORF86	CP	ZNF528
SLFN1	PKIB	ZFP189	C1ORF228	TCP11L1	MAX
EAR4	MOCOS	PPP1R2	ABCC4	MUT	CSE1L
SELL	RGD1566368	PTGS2	ARGLU1	MYOC	ZSCAN25
KLK1B27	RNF207	CDH12	ZNF329	ZFP459	C18ORF25
CCR5	TOR2A	TPR	CEP41	E330013P04RIK	COA1
KLF9	TMEM79	LOC103693948	PLD4	LTN1	ARL17A
ARG1	PKIB	NLK	SF3A2	A430033K04RIK	RASGEF1B
EAR2	TRAK2	LCAT	FCER1A	HLTF	ARL17A
MMP9	SLBP	FAM19A2 /// LOC103692867	CD8B	TAC1	CAMK2D
TRF	SNCG	RSPH1	TMEM204	PNPO	ZNF546
CYP4A12A	ADCY8	CNTNAP2	VAR5	GLRP1	CSGALNACT1
GMPR	TRAK2	CDKN1A	MZB1	CCT8	CNOT6L
RIN2	SLBP	ASS1	NT5E	NUP133	KANSL1
MGST3	SNCG	NTNG1	INADL	LSM6	SNX29
KLK1B11	ADCY8	ACVR1C	FLT3LG	AP3B1	GSTZ1
AMICA1		EGR2	MAP7	FHIT	PYHIN1
TMEM154		HOMER1	NOG	GM3696///GM2897///4930555G01RIK	CFAP43
STK17B		BCAT1	LSM5	FBXL12	ANGEL2
CLEC4A3		BRD4	CHRM3-AS2	TMEM181C-PS///TMEM181B-PS///TMEM181A	PSPC1
FCGRT		PWP2	IGHV4-31///IGHM///IGHG2///IGHG1	MXRA8	MIOS
THBS1		GYPC	GABPB1	PIK3IP1	SCAPER
EGR1		HARS	PCIF1	MIR421	HAUS2
DUSP6		LRRC4C	TMEM97	SMARCA4	ZFYVE9
SDC3		APPBP2	GID8	CDON	LRRC37A3
TCF7L2		NMT1	SLC35D1	DYNLT1A///DYNLT1C///DYNLT1F///DYNLT1B	ENSA
KLK1B4		NLGN1	CYB5D1	GM5089	RORA
GCNT1		KCNJ6	PAPOLG	GM10404	MESDC2
CXCR2		UBE2G1	CHD2	MIR374B	BDP1
SCEL		FKRP	VPS72	CAV2	THADA
DCXR		ASNS	RBM4B	CCL28	KRT7

PLEKHG2		TANC2	STT3A	SLC16A14	ZBTB44
CCND1		RELN	PEX13	PIGZ	RAPGEF6
TOB1		USF2	HIBADH	PIK3CD	COL19A1
IFITM1		BLVRB	GMEB1	PLEKHG4	SIN3A
EGR2		TRIP12	RMND5A	SLC38A6	ZNF337
RCAN3		VPS52	ALDH6A1	ENO3	SH3YL1
ST8SIA4		LOC100912478 /// LOC103689964 /// LOC103694079 /// VMA21	DNPH1	XYLB	PPP2R5C
CDC42EP2		DPCD	ALKBH7	TMA16	SLFN12
HIST1H1C		PCSK2	C12ORF10	LMF2	RASGRP1
CH25H		ZFP874B	OPA1	PTTG1	HIST4H4
AXL		SPATS2L	MGAT4A	LAIR1	RDM1
B3GNT8		MLIP	DNMT3A	IDUA	C16ORF89
LY6A		EXOC7 /// LOC100911734	ZNF706	RIOK1	ZNF507
IL4I1		SCYL3	CDC37L1	ZFP704	NBEA
SERPINB6A		EPC2	ZCCHC10	MPC1	EPG5
SBK1		DAB1	FAM110A	YIPF2///CARM1	NPHP4
H2-T23		PAPOLA	PGLS	VMN2R121///VM N2R- PS159///VMN2R1 23	CDKL3
IER3		KCTD1	DNAJC16	OLFR165	SNX29
NREP		BCL11A	ZNF764	H2AFY2	FCRL2
FAM129A		FAM19A1	PDSS2	CDKL4	FAM20A
LTB4R1		ICA1L	LINC00094	MIR680-2	YLPM1
TRIM29		CABYR	C9ORF16	2310061104RIK	KLHL20
TMEM106A		EPHA7	USP37	4930466F19RIK	BNIP3L
CD274		DGKB	PIGA	OLFR1289	PDK1
ADRA2A		NDEL1	CCNB1	GPR137C	RBM19
SIRPB1A		TMOD1	ATXN3		TAS2R4
DPEP2		EPHX4	SUSD4		SP140L
AP5S1		YAF2	USP42		PDZD9
SLC52A3		NPY1R	UBL7		TBC1D29
SLAMF9		RNF6	GOLPH3L		RAPGEF1
HGFAC		STX6	BPTF		KRT8
RAB19		PLCB4	MRPL57		RPL41
RUSC2		HNRNPD	NMRAL1		RFXAP
F13A1		MAP4K5	TFAM		TMEM135
LIMS2		NIN	COQ4		UBE3B
MGLL		HNRNPA3	ALKBH2		TGIF2
KANK3		PLXNA2	FXR1		PPM1D
TBC1D2		HS6ST1	ZNF766		TBC1D4
VWF		PRRT2	LAMTOR3		ROS1

IFI27		TET3	RPL13		EMSY
GADD45G		ING2	TAF4B		CUBN
CAMK1		PRPS2	DSEL		N4BP2L2
ITGAM		NELL2	SEC24A		ZDHHC17
KLK1B1		PEX7	MUT		DDI2
H2-EB1		SPHKAP	LEF1-AS1		FUBP3
TTC39A		EIF3C	USE1		PYHIN1
AQP9		GIGYF2	DEPDC7		YME1L1
RORA		RUNDC3B	PWWP2B		SCML1
CD9		AEBP2	NTMT1		TRMO
F11R		SLIT2	KIAA0141		MFSD4B
TREM1		ARL1	SLC35C2		MFSD13A
GPR146		PTK2B	DHPS		EHMT1
CASP6		NAP1L1	MRPS10		CTDP1
CAR4		NXPH1	PKIB		KCTD18
GOLM1		OBFC1	ZGPAT		SEMA6B
GRK5		CACNA1A	HINFP		MSANTD2
H2-DMB1		GRIN2A	ARFGAP2		RBM12B
CTSK		MLLT1	SPG7		NCOA5
SERTAD3		LOC102548415 /// NEWGENE_1594 149 /// TCEAL3 /// TCEAL5 /// TCEAL6	TP53RK		CUL1
FLOT1		BET1L	ILF3-AS1		ARL17A
THEMIS2		SLC30A5	BSCL2		PATJ
PRR13		ACKR2	SLC25A45		SRSF5
TMEM25		AZI2	RASA4B///RASA 4CP///RASA4		AFF3
MYCBP2		ZFP292	PHF20		RUNX1
CELA1		BHLHE40	MIEF2		ABCB4
ITGB7		MSH3	FAM98C		CBLB
KLRA2		MBD5	CSTF1		CNST
LILRA6		RBM15	PPIB		RBL2
P2RY13		BEX1 /// LOC100912195	FABP5		SYNM
TNFRSF9		EXOSC7	TNFAIP8L1		ZNF133
USP22		PSMD12	NDUFB1		SCAPER
NAGK		EID2	MAN1C1		ARHGEF16
CYP4F18		FAM219A	LYRM4		PRDM2
SIDT2		RPAP2	CCDC124		ITPKB
FPR2		ANAPC10	PSTK		PYHIN1
ALDH3B1		LOC103690102 /// ZFP655	MKRN2		PROM2
KCP		SGSM3	PMM2		ATP8A1
HIST2H2AA1		AREL1	CCDC94		LRBA
TDRD7		DNAL4 ///	UMPS		SLC47A1

		RGD1559781		
H2-DMA		RALYL	COBL1	UTRN
NECAP1		PCDHB12	YRDC	PYHIN1
RABAC1		IFT57	CCS	SCAPER
ADD3		PTPN3	TNRC6C-AS1	NR2C1
AMDHD2		DYNC1LI1	TERF2	GPATCH2L
ADSSL1		KCND2	RCAN3	R3HDM1
ADAM8		RNF111	PUSL1	CUBN
PLEC		GDPD1	DDX28	ZBTB1
CTNNA1		PLPPR4	DGCR14	MYO9A
ZFP36L1		RASGEF1B	PTPMT1///KBTB D4	PSPC1
DYRK1B		NUDT10 /// NUDT11	KIF22	FCRL3
HIST2H2BB		AFF4	GPANK1	C9ORF3
CD300LF		RAB14	YIPF3	ATP5S
FBXO6		IFRD1	PDE7A	RBM19
GPR171		EIF4E	MRPS16	PTPN11
RAB3D		RORA	PPP1R2	TAS2R13
P2RY14		FHL2	CCDC61	COASY
PDZK1IP1		PPP4R2	SSNA1	NCOA5
SPSB2		COMMD10	ZNF320	COL19A1
DENND2A		ITPKA	PLEKHG4	RTTN
CD300LG		RAD23B	TSSC4	PARP15
SMPD1		SYT17	ZBTB3	NDUFAF6
TSPAN17		GUCY1B3	ZNF252P	PDE7A
GATSL3		TMEM237	MRPL21	BORA
CAPN2		ZKSCAN1	BCAS4	COL19A1
AB124611		CSNK1E	TTC28-AS1	WDPCP
NCKAP1L		BRIX1	THAP6	U2AF1
TMEM51		SCFD1	MPHOSPH8	HIBCH
3110082I17RIK		TMED9	AGBL3	COL19A1
CARD11		KIFAP3	POP7	NEMF
MARCKS		CADM3	CKMT2-AS1	WDR33
DAB2		HPS6	CAPRIN1	KCNQ5
FAM69A		ACOT9	C15ORF40	YLPM1
GDF15		KALRN	ODF2	PPM1D
MMP8		FOXP1	RANBP2	ZNF587
HSD3B7		DSEL	DCXR	RFXAP
GIPC1		METTL5	HIGD1A	PBX3
ING4		TENM3	TYSND1	ANKRD44
CYTIP		COPS5	TOX2	TSC22D2
COPZ2		IPP	TSN	PRKCQ
ZFHX3		AFAP1L1	PPP3CB	SCAPER
SMAD4		MPLKIP	ZNF507	CYP4B1

DHRS1		DPP10	ARL15		LATS1
CYP4F13		PPP1R21	IPO7		UTRN
CYTH3		KATNA1	RMND5B		KIAA1468
CBR3		EPB41L3	LGMN		SPTBN5
SLC41A3		STXBP1	LOXL1-AS1		MKL2
RHOF		ENC1	SLC10A3		CLEC2D
PDDC1		CUX2	TSTD1		TARDBP
PIRA11		NAA15	TARBP2		CLNK
GALE		USP13	TSTA3		STAMBPL1
PRNP		LOC103689968 /// LOC103690028 /// ZFP622	EID2B		KLC1
NFIL3		TTL11	SIVA1		ADAMTS17
LITAF		TPM1	CSTF3		RBM14
ERRFI1		RABGGTA	FAM161A		PRDM2
YPEL3		ACKR1	ZNF569		FOCAD
EIF4E3		FSCN1	ARIH2		ANKRD10
TMEM198B		ZMYM2	ZBTB17		RCBTB1
DRAM2		RTN4RL1	LZTFL1		C9ORF3
F2RL2		UCHL5	EXOSC2		CEP85
SMAP2		PRMT8	CD248		KIF21A
CPNE8		RANBP2	CD27		KRIT1
OSM		MTRF1	SLC25A16		RUNX1
DBNDD2		CHRM4	MRPL11		HP1BP3
GUCD1		PPP1R12A	CHCHD5		RUNX1
TLR13		DCUN1D2 /// NEWGENE_1582 994	ANKRD46		ZCCHC7
AKAP7		FHL1	CNNM2		N4BP2L2
RNF103		AJAP1	SSSCA1		MALT1
OBFC1		SLITRK1	RUNX2		ANKRD10
AKT3		USP11	ZKSCAN7		FNDC3A
PHLDA3		RBM34	TBRG4		NBEA
CMC1		FAM207A	RAD1		ZNF75A
HCAR2		LRRC17	ZNF747		UTY
CIB2		TRPC6	FAM73A		ITPR1
PID1		ICAM5	RPL37		TSC22D2
RHOC		ZRSR1	NIPAL3		KLC1
HLX		GHR	PRKCQ-AS1		EPB41
IFT81		KRT10	ZNF736		SCAF11
TAZ		PIP4K2C	HSF5		ZMYM6
TAP1		RGD1562218	MRPS2		WIPF3
TMEM65		TSLP	CBY1		IL16
MPST		ZFP322A	DDX51		ZFAT
SHISA5		RIT1	VPS51		LRRC37A3

BAG3		TPP2	CPSF3L		ERAP2
H2-AB1		DCUN1D1	IDUA		KLRC1
ANKLE2		TNFRSF21	FAM195A		ATM
FAM126A		GNL2	PAICS		ZC3H14
DUSP22		C1QTNF4	PPAPDC2		SLC25A45
S100A1		MAT2A	MRPL43		TCERG1L
GRINA		ACOT2	ZNF682		ZNF706
IL13RA1		ZFP131	CCDC137		GEMIN2
FAM214A		PPP3R1	PQLC2		DNPEP
FGF13		PNO1	VSIG1		ANKRD10
H2-DMB2		ATG101	SCARB1		ULK4
HEBP1		TM2D1	PDHA1		STAG3
PIRA4		MRPL47	NDC1		SRSF5
CD84		ZBTB6	HIC2		THADA
CMTM4		FAM179B	COG1		MDM4
PIWIL2		WDR12	TRMT1		UTRN
RNASE2A		XK	FOXP1		ZSCAN25
LPL		RGD1562747	DIS3L2		STAMBPL1
B3GNT3		CEBPB	SRRM1		LRRC37A3
PRKRA		IFT81	ZNF33B		RTTN
HPS5		SLC2A13	TCAIM		PSPC1
INSL6		EPHA4	PSMF1		CNOT6L
AIM1		RC3H1	CNPY3		CBLB
WIP1		ASNA1	ZC3H3		KLF12
GAMT		MRPS30	YIF1B		ATM
KCTD12		CDADC1	PER2		ST3GAL1
MCF2L		CMSS1	WDR43		FOXP1
KLK1B8		GRPEL1	MED19		ATF7IP
APLP2		TAF1	MRT04		ZNF586
EVA1B		RABEP1	CCT5		ANKRD36B
TGFBR2		LOC100909747 /// SRP72	TMEM57		PALM2-AKAP2
SLC15A3		NIFK	PATZ1		UTRN
H2-OB		LOC100909512 /// ZFP605	RWDD2A		COX11
ATRAID		TM9SF4	BAD		PTPN2
2510009E07RIK		KCTD17	THEMIS		B4GALT4
IDH1		PAK3	USP31		TUBGCP3
CASP7		MNAT1	NUDT4		CCSER2
PRUNE		CDK11B	SEH1L		CDK13
FAM132A		PRP2L1	GNAS		SH2D3A
FAHD2A		USP32	PLAG1		ACAT1
MARCH2		SYT4	HDLBP		PMM2
BCL2L11		HEXIM2	LOC100130856		ZNF107
HINT3		COG7	CHML		C9ORF3

SLC2A3		STIM1	NBEA		ZNF280D
FOXJ2		ZYG11B	LY6G5C		ZNF586
CHMP2A		FAM81A	PFAS		ZNF154
ZCCHC6		WDR26	WDR3		CDK13
MOCOS		RTN4R	CENPV		RTTN
ARHGAP25		KCNB2	C20ORF196		SP2
GAL		TSSC1	SERBP1		MRI1
ZER1		EIF3E	DPH5		PLEKHA1
CAMK2D		VPS33A	TC2N		GALK2
GNB1		GPR161	C7ORF55		KRT18
TRP53INP1		MBNL2	PIAS2		LRRC37A3
CERCAM		YTHDF3	TTC12		UTRN
SUN1		LOC103690803	GADD45GIP1		IREB2
SORT1		RGD1564804	CNOT10		TTC12
SNX21		FAM122A	LAGE3		RUNX1
NCOA1		JAZF1	RHBDD1		ITPR1
TICAM2		STOML2	PIGO		DDX31
RB1		SMARCA4	PPM1A		ATAD2B
TMEM134		KCNH1 /// RCOR3	JRKL		SLC4A9
ZFP473		PAFAH1B3	COPS6		ITPR1
AATK		MED29	RBM12		MKL2
LHFPL2		NIPSNAP1	BRMS1		SYNE3
CCPG1		BCKDK	DALRD3		ADD1
GALM		ZFP68	EMG1		RTTN
PQLC3		LOC102551114	ADPRHL2		ZBTB44
PGPEP1		ARHGAP32	MKKS		KCNH8
STFA3		SLC45A1	FOXRED1		RTTN
KIF1B		PIK3C3	BRWD1		YLPM1
H2-M3		CPNE5	SNHG10		CUBN
LEPROT		ICA1	RNMTL1		SEC31A
FCHO2		VSIG10L	C11ORF1		ZNF510
IKBKG		SLK	RNASEH2B		FCRL1
ACOX3		LOC103692984	LOC645513		ATP8A1
IL10RA		L1CAM	RPRD2		HECTD2
KLC1		EXPH5	MPG		FAM53B
BLVRB		GCC2	TAF4		RBM4
GYG		TTI2	LRRC47		UTY
GMCL1		VPS37B	FBXO3		NOTCH2
AP2A2		CKAP4	LIME1		TMEM68
PYCARD		RPUSD1	C1ORF50		YLPM1
TLR4		RGD1307100	BUD13		MGEA5
NFATC1		KCNIP4	DDT		TSC22D2
OGT		SNUPN	TMEM41B		GALNT1
HAL		ANKRD33B	KAT8		TRMT44

IRF4		CASD1	DTD2		CAMK2D
RFTN1		TMED1	RAB28		KLRG1
CDKN1A		GTF3A	LUC7L2		GJB7
ACOT1		NUP205	GRSF1		FAM120A
CNNM2		RNF126	ZNF529		PTPRK
AACS		COQ2	PRPSAP2		SERINC3
TIMP2		ERP44	C19ORF52		NFATC3
GAB1		FEM1C	FBXO11		NR1D2
UPP1		ADGRL2	NAA60		MON2
IRF2		FBXO33	RYR1		CAPRIN2
ADRB2		MALL	ZNF7		RALGAPA1
EBPL		LOC685619	TP53		NRF1
NOTCH4		MRPL20	CIRBP		RTTN
SLC27A1		NRN1	RRP15		FYN
PECAM1		PCNXL4	WDR74		ZNF407
CISD1		ZFP574	SNRNP25		ZXDC
APPL2		CCT5	LOC285074		THADA
DOK2		CALB1	TMEM160		AKAP13
NRBP2		AAED1	ZBTB10		KRT40
PLD3		RGD1307704	COG7		FCRL1
POLD4		CAMSAP2	MCPH1		ZNF44
ITFG3		LOC102546678 /// PRRG3	SPECC1L///ADO RA2A		ZNF33A
CDH1		PCDH20	NUDT15		RAB3IP
AGFG2		PIK3CA	DEAF1		PDPR
INPPL1		AK6 /// TAF9	INO80C		MAGOHB
NDRG1		MTDH	IL18BP		NFX1
LENG9		APOO	NOSIP		ANKRD10
BTG1		PELI3	CNOT2		CEP120
EPHX1		BRINP1	LOC100507448		NFAT5
SFT2D2		CAPRIN1	THAP7		LATS1
SOCS3		PPP1R7	DMRTC1B///DM RTC1		JADE1
IL17RA		FKBP3	WDR48		CCDC84
ITGAE		FBXO11	RFWD3		GNL3L
NTPCR		EIF4ENIF1	ITM2C		AAK1
SLC7A2		CAMKK2	ZNF738		KYNU
IL18R1		TTC33	C19ORF60		ANKRD36B
TGM2		CCNH	DDX23		C9ORF3
F7		DDX46	ERAL1		ZNF581
BDH1		VPS4A	MRPS15		FOXJ3
ABCG1		IREB2	LOC148413		SIN3A
OLFM1		GADD45B	MAFK		MORN1
MAVS		EXOSC8	RPL23		CBLB
COX7A1		UBP1	ZNF264		ELMO1

AGRN		TFAM	COX15		RBM33
GM2A		MEDAG	CFL2		ZBED5
SH2D1B1		CHCHD3	TMEM256		SCN3A
PRSS34		TIAM1	WDR54		TRAPPC10
ZFP36		VPS53	TRIP6		FCRL1
KLF13		ARL8B	ZBTB48		MTR
CD33		ARHGAP35	ACYP2		KCNQ2
CRTAC1		GPN3	ARID5A		PSMD13
HAAO		FAM155A	PPP3R1		BNIP3L
IDS		CHRM1	NDUFA11		AURKA
KLHDC4		LGI1	SREK1		VPS13B
KDM4B		GNA13	NOP14		SPECC1L
CD276		GEMIN4	TPPP3		ARL17A
S100A5		CHRM3	SFXN1		ATM
CSRNP1		ATRNL1	MTMR14		RFTN1
ELMO2		MRPS28	MOB2		ARID1B
NACC2		PBDC1	C9ORF114		ABCG1
AA467197		ALDH3B2	SGOL2		ELK4
TLR6		MYT1L	TRABD2A		NDUFAF6
CLEC4A1		ALDH1B1	SKP1		FOXP1
KLF4		AKAP10	ATP5D		RPRD1A
CIITA		SESTD1	CTNS		FAM188A
ENPP1		UBL3	POLD4		UTY
LRG1		ZFP777	CORO1C		PMM2
GBP2		MICU2	FBXO46		MTMR10
PTTG1IP		NUP62	RPIA		GPR135
TRAPPC6A		CRHR1	EEA1		KCP
CTSC		FAM69B	CYB5A		EWSR1
CAR5B		PHOSPHO2	C11ORF30		C19ORF33
CD244		PHF24	MRPL37		AAK1
TNFRSF21		PLCL2	LINC00957		TBC1D19
PEX11G		ZFP91	MAN1B1		SYNRG
GDI1		SYT3	URB1		SMG1
SIRPA		JAK2	MAVS		DIS3L2
PDLIM2		GPATCH11	ZNF880		KLF12
N4BP2L1		SLC7A6OS	THNSL1		BANP
EHD1		PRICKLE2	LINC00263///PP7080		ANKS6
ARHGEF3		FBXW7	ABCF1		RBM19
IL15		PI4KA	BZW2		OCLM
TGFBR3		CYHR1	HADH		PTPRK
SIGIRR		PTPRT	UBXN1		TMEM87A
TNNI2		MEF2C	CIAO1		SCAI
VAMP4		UBE2H	UBL7-AS1		ZNF559-ZNF177
SREBF2		SNAPC5	ZDHHC4		NBEA

HMGCS1		AAGAB	ZNHIT6		ZNF83
SELENBP1		R3HDM2	CENPT		SHPRH
RNF11		BFAR	HAUS6		ITGB1
H2-OA		NUP133	SSBP2		LRRRC37A3
MAPK9		PELO	PAIP1		PCDH8
ZC3H6		PRAGMIN	ECSIT		CSGALNACT1
CTSH		RBSN	SHARPIN		ABCG1
CD74		AHCTF1	RASA4		MINA
MAGEE1		RGS4	COMM5		FCRL2
SLC6A8		FAM168B	ZNF653		UTY
LSP1		SUMF1	TMTC3		NFATC3
KAZALD1		GNPAT	HSPA14		CCDC84
MED23		PROSER1	SATB1		ATP5S
1600012H06RIK		PRMT10	RAD51B		IL16
HEMK1		MAGOH	APEH		MACF1
6330416G13RIK		SLC41A2	CLUAP1		RBM14
RAP2A		MAGI1	SH2D3A		DOCK8
FXVD5		VAMP4	SNORA72		ZNF587
ANO10		LEO1	HSPA9		FCRL1
CD300LD		GPBP1	RPP30		RTTN
PTGIR		VSTM2L	KIF9		ZNF343
5430435G22RIK		CPA2	TMCO6		ZNF37A
STX7		STAT3	C6ORF136		CDKL1
MARCH1		GBF1	WDR13		PDPR
CSNK1D		SEC24B	ZNF688		NAF1
WRN		GABARAPL1	HAUS2		ANKRD46
SMOX		KCNJ4	ARIH2OS		PRDM2
ALDH3A1		CCDC9	HNRNPC		RBM39
GANC		USP46	ATRX		FBXW7
FAM198B		CUL5	TECPR1		RALGAP1
DCTN4		LOC100362572 /// LOC103693261	ATP6V0A2		CCM2
ANG		ATP6V0A1	ZBED4		FRAT1
NR1H3		FLRT2	ZNF561		ARL6
TAP2		EMC7	TNRC6C		ARHGEF6
APOBEC1		GRAMD1B	CYC1		THADA
HSPB6		ASAP1	PPP2R1B		ARFGEF1
CCDC88A		CCT2	EIF4ENIF1		ADAM28
CRYL1		ZBTB45	COX7B		UTRN
LBH		UBE2Q2	EID3		RTTN
PCBD2		TXNDC9	EXOC3		ZNF546
ST6GAL1		DLGAP1	ZNF234		ITPR1
PAFAH1B3		ANK3	STYX		L3MBTL4
ZC3H12D		PHRF1	INTS5		RC3H2

KLF3		DDX10	TOMM22		TRIM52
NICN1		NELL1	TPGS1		RBBP6
GNAQ		CIRBP	PCBD2		C14ORF159
CYB561		RBM11	ILVBL		ZNF75D
HAGHL		VPS26B	NHP2		CD84
AU040320		GTDC1	DHX30		CUBN
2310022B05RIK		SF3A3	TIMM9		CAMK2D
PLEKHB2		RALGAPA1	MZF1		PATJ
WFDC17		AGTPBP1	ZNF174		WDR33
SLC12A6		PSMA2	ZNF260		ZNF44
CLEC4B1		RTF1	RNF123		C9ORF3
AIFM2		GATAD1	ITPKB		DCAF16
PAK4		GPR180	CREBZF		CCNH
APOA1BP		LOC102554302	SLC35E1		OGT
TRAF6		DENND4A	ATG5		TRMO
CYHR1		SF3A1	THADA		KLF12
PSKH1		HIRIP3	SMPD1		CELF2
SERTAD2		PDS5B	SMARCC2		ATM
PRMT2		SDR39U1	FEN1		DHRS3
OSTF1		UBXN4	RPP40		ZNF585B
CCSER2		CACNG3	FAM193A		SLC26A1
SEMA4B		HPCA	SLC52A2		PBX4
CD47		VLDLR	GTF2F1		ZNF189
BOK		PAFAH1B2	TBC1D15		NBEA
ZFP658		TTL	GPATCH1		UTRN
PLGRKT		TLN1	MTMR12		BACH2
HTATIP2		ZFP157	SUB1		PHTF2
JUNB		POU3F2	CLN5		STXBP5
IGF2BP2		RPL28	NAA15		E2F5
EPS8		TIMM10B	PTPMT1		SP2
NCK2		ACADSB	CETN2		KIAA0226L
NDFIP1		CLNS1A	C2ORF40		ZNF43
LGALS8		DIRC2	LIN37		ATM
SGK1		CACNA2D1	RAB30		LPIN1
LY96		USF3	RLN2		DGKH
CCL6		RGD1564053	PKHD1L1		BTBD11
CDC25B		LOC102556259 /// LOC301748	GEMIN2		COA1
EMILIN2		AMIGO1	TMEM219		CEP72
COL5A1		BABAM1	LEF1		RTTN
OSBPL9		NCKIPSD	GEMIN5		CLTA
RND3		PTPRO	GRPEL2		CCNH
CKB		DDIT3	TFPT		ZNF546
TMEM71		CIPC	PNPO		ADGRF5
CRACR2B		TRIM2	TDRKH		HKR1

PREX1		PRKAG2	TMEM134		HTRA4
DDHD1		FAM45A	TMEM86B		EXOGE
RNF19A		BIRC2	LOC100287628		MACF1
TIAM1		CORO1A	FAM173A		TTLL5
TMEM205		AURKAIP1 /// CCNL2	DENND6B		ZNF215
SLC15A4		COQ5	INTS6-AS1		PYHIN1
FCGR4		MCHR1	MDP1		SLC25A13
KCTD21		MAGI3	ERGIC3		TARDBP
GPR108		TAB3	AARSD1		FERMT1
EAR12		LOC100360120	TMEM101		KLC1
		KRI1	DBP		CRACR2A
		TLE1	TMEM189		CD22
		CERK	SUGP1		RFC2
		RSRC1	RANBP1		LRRRC37A3
		NICN1	RSPRY1		ULK4
		DIRAS2	ITGB8		MAF
		RND1			HIP1R
		CTTNBP2NL			FCRL1
		LOC100910308 /// PAICS			TTC17
		FAM84A			MON2
		AGFG1			ZNF587
		LOC102549115 /// ZFP260			ZNF587
		TRHDE			PHYKPL
		LOC689499			KIF6
		MCTS2			TDP1
		DLG3			BOD1
		CDK8			ECHDC2
		UBR3			UTRN
		ESRRG			CYP4V2
		FAM174A			SCAND1
		HEATR5B			N4BP2L1
		UGGT2			COL19A1
		NFYC			SCAPER
		SON			SLC7A5
		NUDCD1			VPS13D
		GLTSCR1L			ZNF44
		ZWILCH			SYNE2
		VSNL1			CBLB
		DCLK3			NR3C2
		LOC100362176			SAMD12
		POLR2B			CHD2
		LOC103690405 /// MXRA7			SLC4A7

7.4. Appendix D

GSEA results with up-regulated gene-sets from preliminary analysis.

		Controls	Cases	UP-REGULATED CUSTOM GENE-SETS											
				Li-treated HSPC (GSE35291)		Li-treated RAT_Hippocampus (GSE66277)		Li-treated RAT_Cortex (GSE66277)		BD PBLs (GSE46449)		Mouse mania model (GSE29417)		BD blood samples (GSE46416)	
				Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
1	GSE34747	mesenchymal stem cells (MSCs)	MSCs with lithium (5mM) for 7 days		0.002		0.981	1		0.592		0.909			1
2	GSE35291	hematopoietic stem/progenitor cells (HSPC)	HSPC with lithium (5mM) for 7 days		0		0.519	0.081*		0.295			0.506		0.162
3	GSE66277 (hippocampus)	rats treated with vehicle	rats treated for 21 days with Lithium chloride (0.2% in chow)		0.788			0*			0.174		0.205*	0.885	
	GSE66277 (cortex)				0.955				0		0.063		0.345		0.054
	GSE66277 (striatum)			0.658				0.003*		0.35			0.51		0.651
4	GSE54294 (day7 samples)	peri-implant bone not exposed to Li after 7 and 28 days	peri-implant bone exposed to Li released from bone, for 7 and 28 days	0.87				0.454			0.54	0.775			0.034
	GSE54294 (day28 samples)			0.977				0.316		0.407			0.979		0.045
5	GSE46449 (rep1s)	peripheral blood lymphocytes (PBLs) from matched healthy controls	PBL gene expression in BD patients	0.736			0.268		0.454		0.005	1		1	
	GSE46449 (rep2s)			1			0.207		0.148		0	1		0.922	
6	GSE12654	post-mortem healthy brain	post-mortem BD brain	1				0.596		1		0.94			0.051
7	GSE46416	male, age-matched healthy controls (mean age=47.6±11.98 years)	BD euthymic states (mean difference: 20.09±7.3)		0.072		0.998		0.615		0.566		0.052*		0.036
8	GSE69486	RNA obtained from skin fibroblast cells of healthy control samples	RNA obtained from skin fibroblast cells of bipolar disorder patients		0.991		0.085		0.969		1		1	0.137*	

9	GSE29417	hippocampal RNA in both groups. ancestral outbred hsd:ICR (ICR) strain of MSN	mouse strain, termed Madison (MSN), which naturally displays a manic phenotype		0.497		0.109*	1		0.695		0	0.754	
7	GSE46416	Euthymic phase blood samples	Manic phase blood samples		0.631		0.403	0.527		0.316		0.968	0.995	
5	GSE46449 (rep1s)	PBL from matched BD	First-episode BD (bipolar 1) patients - no medication	0.157			0.575		0.485	0.946		0.719	0.037	
	GSE46449 (rep2s)	PBL from matched BD	First-episode BD (bipolar 1) patients - no medication		0.785		1		1	0.008		0.933	0.019	

7.5. Appendix E

GSEA results with down-regulated gene-sets from preliminary analysis.

		Controls	Cases	DOWN-REGULATED CUSTOM GENE-SETS											
				Li-treated HSPC (GSE35291)		Li-treated RAT_Hippocampus (GSE66277)		Li-treated RAT_Cortex (GSE66277)		BD PBLs (GSE46449)		Mouse mania model (GSE29417)		BD blood samples (GSE46416)	
				Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
1	GSE34747	mesenchymal stem cells (MSCs)	MSCs with lithium (5mM) for 7 days	0.896			0.78		0.636		0.097*	1		0.968	
2	GSE35291	hematopoietic stem/progenitor cells (HSPC)	HSPC with lithium (5mM) for 7 days	0		0.899			0.157*		0.016*		0.116	1	
3	GSE66277 (hippocampus)	rats treated with vehicle	rats treated for 21 days with Lithium chloride (0.2% in chow)		0.227*				0.351		0.936	1		0.911	
	GSE66277 (cortex)				0.348			0		0.001			0.611	0.083	
	GSE66277 (striatum)				0.005*			0.01			0.162*	0.034*		0.144	
4	GSE54294 (day7)	peri-implant bone not exposed to Li after 7 and 28 days	peri-implant bone exposed to Li released	0.596			0.471		0.046*	0.481		0.679			0.016*
	GSE54294 (day28)				0.491		0.601		0.485		0.789		0.614		0.671

			from bone, for 7 and 28 days												
5	GSE46449 (rep1s)	peripheral blood lymphocytes (PBL) from matched healthy controls	PBL gene expression BD patients	0.581		0.552		0.366		0		0.064*		0.395	
	GSE46449 (rep2s)			0.733		0.239		0.274		0		0.12*		0.233	
6	GSE12654	post-mortem healthy brain	post- mortem BD brain		0.228*			1		1		0.902		1	
7	GSE46416	male, age- matched healthy controls (mean age=47.6±11.98 years)	BD euthymic states (mean difference: 20.09±7.3)		1		0.471		0.586		0.575		0.729	0.005	
8	GSE69486	RNA obtained from skin fibroblast cells of healthy control samples	RNA obtained from skin fibroblast cells of bipolar disorder patients	1		1			1	0.999			1		0.954
9	GSE29417	hippocampal RNA in both groups. ancestral outbred hsd:ICR (ICR) strain of MSN	mouse strain, Madison (MSN), which naturally displays a manic phenotype	1		1		1		1		0		0.983	
7	GSE46416	euthymic phase blood samples	Manic phase blood samples		0.832	0.118*			0.514		0.306		0.501		0.251
5	GSE46449 (rep1s)	PBL from matched BD	First- episode BD (bipolar 1) patients - no medication	0.391			1	0.432			0.411		0.936	0*	
	GSE46449 (rep2s)	PBL from matched BD	First- episode BD (bipolar 1) patients - no medication	0.642			0.129		0.903		0		0.443	0.152*	

7.6. Appendix F

Table of up-regulated gene-sets and their p-values from validation analysis.

11. GSE112523		21. GSE116820		2. GSE80655		7. GSE80336		13. GSE66196		10. GSE46416	
BD	P-value	BD	P-value	BD	P-value	BD	P-value	Li treated	P-value	Manic	P-value
SFN	2.95E-05	IGFBP7	0.008427	HSPA6	3.43E-11	PTGER3	3.19E-05	HOPX	0.013704	PMPCB	0.002759
GBP1	6.54E-06	BGN	0.012649	CHI3L1	2.54E-13	CHI3L2	0.006471	KLRC1	0.013605	LEO1	0.003119
CHI3L2	3.55E-05	FN1	0.01859	IL1RL1	7.52E-06	SLC12A3	0.000815	DSC2	0.006629	PSMD12	0.00123
AC241377.3	3.1E-07	ACTA2	0.001364	AC068491.3	0.005125	APOL4	8.67E-06	DSC3	0.016437	TMF1	0.008219
S100A9	9.23E-05	SPARC	0.024953	ADAMTS9	2.02E-09	SLC17A7	2.51E-05	UNC13C	0.002827	ZNF800	0.003547
S100A8	2.17E-05	MYL9	0.01348	C5orf17	1.16E-07	HAMP	0.000826	VPS35	0.019906	HNRNPH3	0.001533
SLAMF8	6.55E-05	PLPP4	0.040741	XIST	4.54E-12	TTR	0.000873	PIEZO2	0.040686	KLRB1	0.015361
FCGR3A	3.61E-06	RBP4	0.005115	BEND7P1	6.93E-09	KCNS1	0.003098	XCL1	0.026296	CCNG1	0.002981
SLC11A1	8.15E-05	THBS1	0.030826	MTCO1P53	4.37E-07	ACY3	0.000281	CCL20	0.009565	TARP	0.025325
LINC00694	7.89E-05	SLITRK6	0.041807	MIR3151	6.46E-08	CHI3L1	0.016918	SPINK2	0.001718	NR1D2	0.006776
ADAMTS9	4.39E-05	CYP4V2	0.00186	RNU6ATAC39P	4.73E-06	MKI67	0.002729	MUC19	0.006262	RECQL	0.006615
LINC01471	0.000463	NPAS4	0.01619	MT1JP	1.65E-07	PART1	4.9E-06	SV2B	0.021706	KPNA5	0.001532
SPP1	6.32E-06	TAGLN	0.024006	MT1A	2.18E-11	C1QL3	6.24E-05	CREB5	0.000209	DNAJC7	0.016292
AC107208.1	0.000287	TGFB1	0.04552	SOC3	1.87E-09	NWD2	0.008918	XCL1	0.022046	ESF1	0.00163
AC139495.3	1.2E-05	NPAS2	0.006582	AC022144.1	1.19E-10	SERPINA5	0.000317	NCR2	0.021516	SAP18	0.010002
AL391422.3	8.35E-06	ZNF558	0.01288			AC099552.3	0.000706	NFAT5	0.007869	ACTR10	0.00445
C4B	6.74E-05	LINC01561	0.003294			LILRA4	0.000366	LHX2	0.006007	ZNF253	0.002791
PGC	0.009173	MEIS2	0.028475					FRMD4B	0.047541	IPO5	0.02256
BMP5	0.001662	ACTG2	0.021679					ATXN7L3B	0.000853	LASS3	0.003705
KHDC3L	6.55E-05	GRM5	0.000129					ENTHD1	0.009809		
AC093627.7	3.44E-05	PLP1	0.021378					CXCL10	0.027541		
MYO1G	0.001529	LAMA3	0.025721					GRIA3	0.000113		
HILPDA	3.89E-05	GRM3	0.026782					ST3GAL6	0.013064		
SLC13A4	2.01E-05	CTGF	0.007296					MYB	0.006154		
AC004941.1	0.000681	CYP1B1	0.023556					BCL2L11	0.022458		
AC073314.1	4.51E-05	FAT1	0.034503					IQCJ-SCHIP1	0.023511		
PRSS55	4.07E-05	MYLK	0.016794					WSB1	0.035805		
SLC18A1	0.001704	UNC13C	0.014684					IL6ST	0.044499		
STC1	1.37E-05	GPR26	0.001695					DNASE1L3	0.012787		
NUGGC	0.00614	NTS	0.043738					PDCD4	0.026205		
AC108002.3	0.00017	CRYAB	0.007409					LOC101927287	0.002963		
SNX18P5	1.76E-05	EDNRA	0.040097					EOMES	0.040963		
SPATA31B1P	0.003244	SCRG1	0.000432					PRKCH	0.012898		
AL136981.1	3.69E-05	GUCY1A3	0.03104					RIMBP2	0.039488		

AL807752.3	0.0019 81	COL3A1	0.0245 42					LRP12	0.0078 94		
PARD3- AS1	0.0013 19	ENC1	0.0050 91					DNM3	0.0144 3		
IFITM2	9.27E- 05	PPARGC1 A	0.0446 98					PEX13	0.0422 12		
IFITM3	0.0003 15	HLA-A	0.0443 37					WNT2	0.0194 39		
CD44	2.59E- 07	OPRM1	0.0225 01								
SLC22A8	1.61E- 05	COL4A6	0.0240 94								
OVOL1	1.77E- 05	TNS3	0.0116 71								
CLCF1	0.0001 2	RGS6	0.0348 39								
LINC02470	3.09E- 05	SOSTDC1	0.0408 31								
COL2A1	0.0097 34	MYOF	0.0030 84								
RF00019	0.7683 82	SDR42E1	0.0010 93								
GPR84	0.0012 78	SHISA6	0.0213 25								
P2RX2	0.0001 99	DOCK10	0.0440 25								
CASC4P1	0.0097 66	RORB	0.0467 66								
SERPINA1	3.49E- 05	BDNF	0.0127 9								
LINC00524	0.0010 31	GPR3	0.0399 95								
AC021979. 2	0.0103 34	TINAGL1	0.0031 06								
ULK4P3	0.0002 19	ELMOD1	0.0257 97								
DUOX2	0.0052 96	SLC6A1	0.0254 38								
AC104758. 1	0.0001 7	COL1A2	0.0148 21								
41153	0.0050 67	CHST15	0.0451 13								
TEKT5	0.0010 94	DHRS3	0.0474 37								
AC133065. 2	0.0010 29	RNASE3	0.0024 07								
AC020765. 2	0.0010 37	CAV1	0.0021 77								
MT1A	0.0046 93	MYO3A	0.0290 64								
MT1X	9.63E- 05	PNMA5	0.0006 78								
PIRT	0.0053 42	ARHGEF35	0.0195 72								
CCL2	0.0009 27	CIT	0.0238 81								
SPPL2C	0.0001 42	RBFOX3	0.0209 14								
NSFP1	0.0013 16	DGKG	0.0018 84								
ITGB3	0.0011 66	FAM198B	0.0087 29								
AC027801. 4	0.0018 02	IL1B	0.0168 94								
		ALDH1A3	0.0011 79								
		GPRC5A	0.0192 53								
		CTCF	0.0175 76								
		ANO3	0.0208 15								
		CALD1	0.0465 03								
		NGEF	0.0096 89								
		SORCS2	0.0193 27								
		CD99	0.0001 46								

		CPA4	0.0243 34								
		TMEM132 C	0.0181 76								
		KCNK12	0.0266 87								
		TLE4	0.0100 95								
		COLEC12	0.0066 55								
		ADARB2	0.0194 85								
		NDNF	0.0040 14								
		FAM182B	0.0039 09								
		ZNF700	0.0398 6								
		IFITM3	0.0370 9								
		GPRIN3	0.0178 63								
		NWD2	0.0187 51								
		OR5AR1	3.39E- 05								
		CSPG5	0.0432 94								
		PDIA2	0.0085 36								
		GAS8-AS1	1.22E- 05								
		PLPP2	0.0367 83								

7.7. Appendix G

Table of down-regulated gene-sets and their p-values from validation analysis.

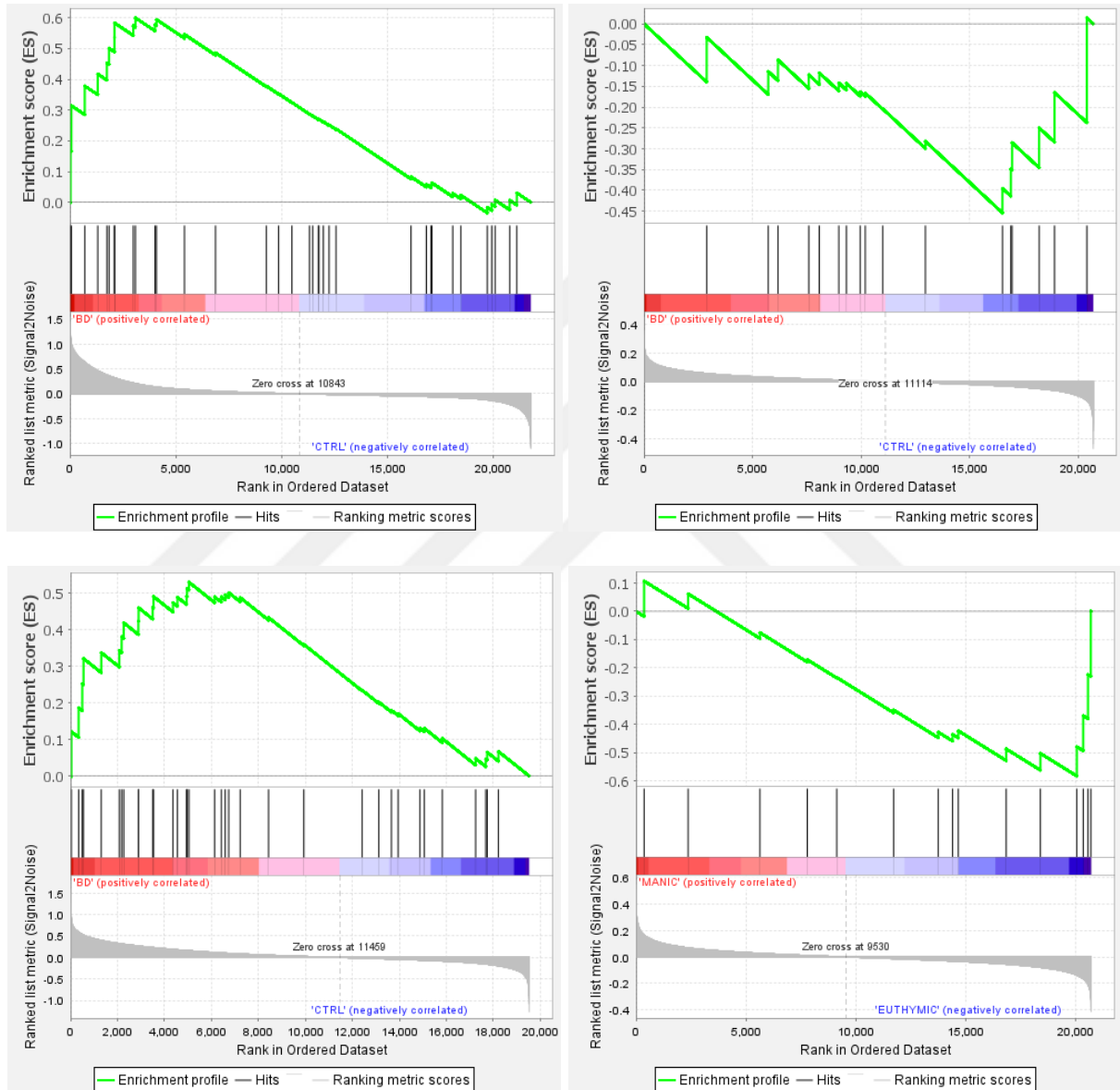
11. GSE1125 23		12. GSE6619 6		21. GSE1168 20		14. GSE813 96		23. GSE743 58		9. GSE42 546		13. GSE661 96	
BD	P- value	BD	P- value	BD	P- value	BD	P- value	BD	P- value	BD	P- value	Li treated	P- value
AC02780 1.3	6.15E -05	KCNJ15	0.032 785	PQBP1	0.019 619	HAPLN 2	0.048 814	COL5A2	0.028 255	GAB1	0.000 547	NFATC2	0.006 061
DLX4	1.31E -06	C14orf10 5	0.042 626	RRP7A	0.003 987	TMEM1 19	0.007 293	COL5A1	0.011 417	GRK7	0.000 199	IGLV1- 44	0.008 938
DNAH17 -AS1	0.003 266	MAP1B	0.000 349	ABHD3	0.003 959	PAQR6	0.047 012	ZNF736	0.022 445	HS3ST3 A1	0.000 179	KLHL14	0.003 686
ANGPTL 4	0.004 381	ZNF503	0.048 796	AMOT	0.046 93	SFTPC	0.013 99	SGCD	0.001 642	IMPG1	0.000 436	IGLJ3	0.002 271
HAMP	0.017 722	CTSZ	0.021 602	EFCAB1 3	0.005 979	SEMA4 C	0.006 467	COL6A2	0.005 729	INO80C	0.017 864	IGKC	0.011 242
IL4I1	0.011 581	MGA	0.021 131	TIMM17 B	0.025 256	AEBP1	0.026 461	ARHGA P29	0.038 216	LATS2	0.000 402	APELA	0.032 876
SIGLEC9	0.004 91	XCL1	0.019 337	CEBPZO S	0.021 742	CNP	0.017 39	B3GAL NT1	0.005 321	LOC284 100	0.002 09	IGLC1	0.004 185
SIGLEC1 4	0.018 725	PDZD2	0.032 486	FAM228 A	1.45E -05	NEURL 2	0.001 974	BGN	0.043 198	LRFN1	0.000 858	CKAP2	0.012 611
FPR1	0.032 823	HS3ST3B 1	0.003 877	KCND1	0.029 187	BGLAP	0.005 713	LOX	0.037 384	NHEDC 2	0.000 371	AQP3	0.014 355
HELZ2	0.005 603	STEAP1	0.009 368	ORAI2	0.019 736	SMPD1	0.038 272	RHOJ	0.030 166	NHLRC 2	0.007 009	CT45A1	0.005 367
TMPRSS 3	0.000 767	RBPMS	0.003 957	CCT6B	0.000 606	SPNS1	0.020 528	ITM2A	0.045 18	NPTX2	0.000 216	SLCO4C 1	0.044 387
SIK1	0.012 635	CXCL10	0.001 619	FAM221 A	2.67E -05	INCA1	0.017 573	ZC3HA V1L	0.013 832	NRN1	0.004 043	C1orf18 6	0.006 812
TSPEAR- AS1	0.001 809	RGS13	0.005 174	GCNT2	0.018 897	NECTIN 1	0.013 499	DCN	0.030 595	PAQR6	5.91E -05	KLF2	0.017 523
APOL4	0.000 701	LOC1002 93211	0.023 098	C1orf52	0.000 711	LDHD	0.021 213	IGF2	0.046 825	PLIN3	1.09E -07	PLEKH A2	0.001 103
IL17REL	0.005 16	PTPN14	0.023 749	LOC3398 62	0.025 772	PRAF2	0.007 902	COL14A 1	0.014 681	PRKG2	2.47E -05	IL12B	0.035 843
TYMP	0.004 982	TBX15	0.002 992	TMEM26 0	0.001 188	GSTM5	0.013 454	GLT8D2	0.003 013	PRTFD C1	4.1E- 07	FAM46 C	0.026 273

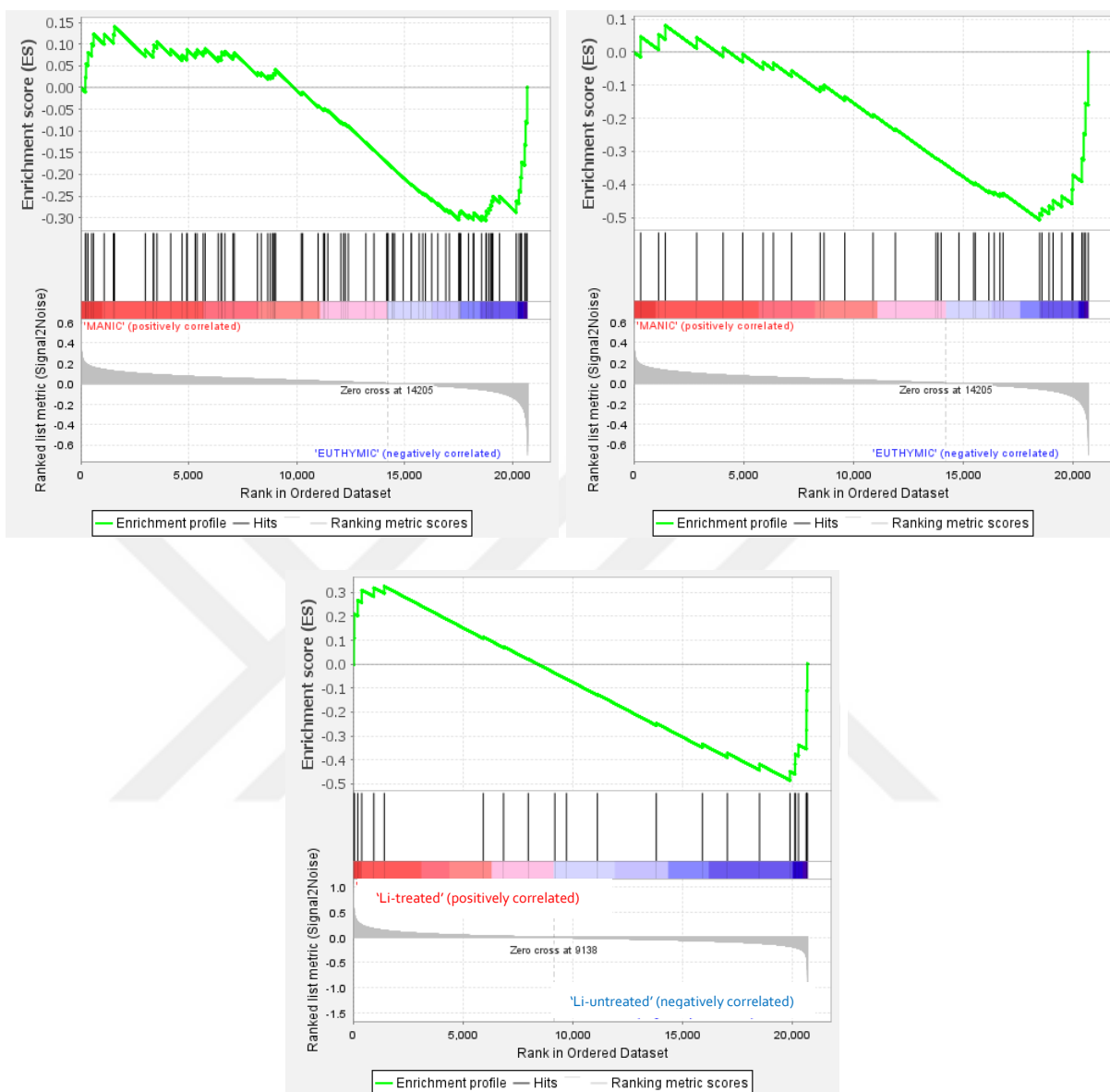
Z82205.1	0.009 033	GPC4	0.010 537	WDR43	0.001 908	SSTR1	0.002 556	LAMB1	0.045 119	PTPRA	6.97E -05	CXCL13	0.000 355
		ARHGD1 A	0.024 905	EPS15	0.039 413	LGALS3 BP	0.013 136	CYP1B1	0.037 172	SLC36 A1	2.83E -06	IGLV4- 60	0.018 231
		CCND2	0.004 568	PIN4	0.000 12	ADAMT S4	0.029 025	NID1	0.016 928	SRPR	7.69E -07		
		VPS35	0.022 56	EFCAB1 1	0.006 174	LLGL1	0.013 022	LGALS1	0.021 533	SYNGA P1	3.8E- 05		
		SPEN	0.027 956	PIM2	0.001 348	TGM2	0.036 118	LAMA4	0.027 708	TTYH3	2.75E -07		
		MAB21L 3	0.021 884	NAPRT	9.53E -05	SOX10	0.011 045	PRRX1	0.021 604	WDR76	5.97E -12		
		KCTD14	0.047 705	CPXM1	0.006 589	HSPA2	0.004 305	COL3A1	0.023 482				
		IGHV3- 73	0.010 448	LRRC34	0.001 024	MT1M	0.005 81	INHBA	0.044 535				
		SYTL2	0.001 649	ZNF471	0.002 044	SLC44A 2	0.003 892	EDNRA	0.042 826				
		SORBS2	0.002 053	ZDHHC1 1	0.029 277	TSC22D 4	0.014 804	PDGFR A	0.042 265				
				PRPH2	0.003 426	PFKL	0.008 481	EGFL6	0.018 459				
				CHRNA5	0.044 133	TPPP3	0.002 864	COL6A3	0.014 793				
				FABP6	0.009 074	EMILIN 3	0.017 889	COL1A1	0.021 027				
				GNAT1	0.001 045	AGTRA P	0.011 578	OGN	0.041 03				
				LRRTM4	0.002 767	ELFN1	0.021 217	COL1A2	0.006 56				
						CDC42E P1	0.017 728	ASPN	0.041 616				
						ZNF414	0.007 822	HAPLN 1	0.039 848				
						ELOVL1	0.003 356	COL21A 1	0.024 037				
						HSPB6	0.029 319						
						ADAM1 5	0.014 265						
						NOC4L	0.027 678						
						ENTPD2	0.024 138						
						BSG	0.008 953						
						B3GNT7	0.006 423						
						CSPG4	0.026 657						
						HSPG2	0.047 762						
						SH2D3C	0.049 788						
						ARIH2O S	0.005 797						
						IZUMO4	0.005 668						
						MARCK SL1	0.008 722						
						KLK6	0.006 536						
						PEBP4	0.022 189						
						ITGB4	0.026 762						
						GPR62	0.007 191						
						C1QL1	0.020 916						
						PRR18	0.009 007						
						TELO2	0.015 547						
						LIPE	0.005 004						
						APLP1	0.001 515						
						PGAP3	0.021 603						

					IGSF21	0.046 712						
					C4B	0.023 823						
					OLIG2	0.000 795						
					PTGDS	0.003 683						
					EML2- AS1	0.004 32						
					PLEKHF 1	0.003 849						
					PAQR4	0.005 499						
					CD99	9.3E- 05						
					LIMS2	0.003 741						
					RHOG	0.038 142						
					GJB1	0.006 127						
					SUSD2	0.015 677						
					GPR37L 1	0.009 645						
					ARL17B	0.004 042						
					GPR17	0.009 327						
					CX3CR1	0.038 404						
					ECEL1	0.014 84						
					JUP	0.028 493						
					PPDPF	0.015 109						
					CARNS 1	0.010 107						
					CMTM5	0.001 165						
					NGFR	0.028 091						
					GGT5	0.010 533						
					TMEM1 25	0.002 492						
					MAG	0.004 972						
					CCDC17	0.000 613						
					KIF19	0.030 943						
					MAL	0.000 183						
					CHI3L2	0.003 745						

7.8. Appendix H

ES plots of the datasets selected from the GSEA result **Tables 4.4.1 and 4.4.2** are shown here (these are indicated with red arrows in the aforementioned tables).





7.9. Appendix I

The core enriched genes from each of the ES plots, shown above, are listed in the tables below. The tables of the plots mentioned in the results section are placed at the end. *The column with genes' description has been deleted to make the results more clear.*

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	DNASE1L3	32	1.153	0.1672	Yes
2	DSC3	68	1.015	0.3141	Yes
3	ENTHD1	703	0.637	0.3781	Yes
4	XCL1	1318	0.459	0.417	Yes
5	CCL20	1748	0.38	0.4529	Yes
6	NCR2	1851	0.36	0.5008	Yes
7	EOMES	2108	0.322	0.5361	Yes
8	KLRC1	2110	0.322	0.5831	Yes
9	CXCL10	2994	0.216	0.5741	Yes
10	DSC2	3097	0.206	0.5996	Yes

11	SPINK2	4020	0.139	0.5775	No
12	MYB	4091	0.136	0.5941	No
13	LHX2	5413	0.085	0.5458	No
14	HOPX	6880	0.052	0.486	No
15	IL6ST	9284	0.016	0.3779	No
16	UNC13C	9862	0.01	0.3528	No
17	GRIA3	10489	0.003	0.3245	No
18	ST3GAL6	11315	-0.005	0.2872	No
19	BCL2L11	11468	-0.006	0.2811	No
20	LRP12	11736	-0.008	0.2701	No
21	NFAT5	11764	-0.009	0.2701	No
22	WSB1	11975	-0.011	0.262	No
23	PDCD4	12244	-0.013	0.2515	No
24	VPS35	12571	-0.016	0.2388	No
25	ATXN7L3B	16116	-0.048	0.0828	No
26	RIMBP2	16839	-0.056	0.0578	No
27	PEX13	17063	-0.059	0.0561	No
28	DNM3	17099	-0.059	0.0632	No
29	WNT2	18087	-0.073	0.0285	No
30	PRKCH	18467	-0.081	0.0229	No
31	SV2B	19717	-0.121	-0.0168	No
32	CREB5	19930	-0.134	-0.0069	No
33	FRMD4B	20089	-0.145	0.007	No
34	IQCJ-SCHIP1	20777	-0.212	0.0064	No
35	PIEZO2	21114	-0.27	0.0304	No

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	KPNA5	2888	0.056	-0.0327	No
2	KLRB1	5722	0.029	-0.1146	No
3	TMF1	6169	0.026	-0.0868	No
4	PSMD12	7588	0.017	-0.1224	No
5	PMPCB	8073	0.015	-0.1179	No
6	HNRNPH3	8970	0.01	-0.1421	No
7	ZNF800	9320	0.008	-0.1429	No
8	CCNG1	9953	0.005	-0.1633	No
9	ACTR10	10188	0.004	-0.1663	No
10	SAP18	10996	0	-0.2044	No
11	IPO5	12956	-0.009	-0.2824	No
12	ZNF253	16504	-0.031	-0.3953	Yes
13	DNAJC7	16890	-0.034	-0.3487	Yes
14	NR1D2	16947	-0.035	-0.2854	Yes
15	RECQL	18192	-0.05	-0.2498	Yes
16	LEO1	18900	-0.062	-0.1651	Yes
17	ESF1	20391	-0.132	0.0141	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	UNC13C	38	0.9	0.1213	Yes
2	NCR2	348	0.592	0.1865	Yes
3	DSC3	501	0.54	0.2526	Yes
4	ENTHD1	555	0.522	0.3215	Yes
5	PIEZO2	1310	0.4	0.3376	Yes
6	MYB	2079	0.326	0.3428	Yes
7	CCL20	2178	0.317	0.3812	Yes
8	DNASE1L3	2267	0.31	0.4191	Yes
9	HOPX	2890	0.268	0.424	Yes
10	MUC19	2899	0.268	0.4603	Yes
11	FRMD4B	3499	0.232	0.4613	Yes
12	SV2B	3534	0.23	0.4911	Yes
13	PRKCH	4363	0.188	0.4743	Yes
14	GRIA3	4552	0.179	0.4892	Yes

15	IL6ST	4942	0.163	0.4915	Yes
16	WNT2	4961	0.162	0.5128	Yes
17	RIMBP2	5043	0.159	0.5304	Yes
18	BCL2L11	6134	0.117	0.4904	No
19	XCL1	6420	0.107	0.4905	No
20	LHX2	6586	0.102	0.4959	No
21	DSC2	6737	0.098	0.5016	No
22	KLRC1	7224	0.084	0.4881	No
23	SPINK2	8432	0.054	0.4336	No
24	LRP12	9933	0.025	0.36	No
25	NFAT5	12411	-0.015	0.235	No
26	WSB1	13122	-0.027	0.2023	No
27	DNM3	13661	-0.036	0.1797	No
28	PEX13	13947	-0.041	0.1707	No
29	ATXN7L3B	14876	-0.059	0.1312	No
30	CXCL10	15062	-0.063	0.1304	No
31	VPS35	15828	-0.083	0.1025	No
32	CREB5	17246	-0.131	0.0476	No
33	ST3GAL6	17676	-0.154	0.0467	No
34	PDCD4	17738	-0.157	0.0651	No
35	EOMES	18217	-0.193	0.067	No

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	TTR	372	0.167	0.107	No
2	SLC12A3	2371	0.069	0.0619	No
3	CHI3L1	5634	0.027	-0.0754	No
4	SERPINA5	7786	0.011	-0.1713	No
5	PART1	9124	0.002	-0.2341	No
6	APOL4	11709	-0.013	-0.3493	No
7	KCNS1	13737	-0.026	-0.4276	No
8	LILRA4	14386	-0.031	-0.4355	No
9	NWD2	14647	-0.033	-0.423	No
10	SLC17A7	16823	-0.055	-0.487	No
11	MKI67	18377	-0.079	-0.5028	Yes
12	ACY3	20033	-0.138	-0.4794	Yes
13	CHI3L2	20324	-0.165	-0.3696	Yes
14	PTGER3	20540	-0.208	-0.224	Yes
15	HAMP	20669	-0.308	0.0007	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	RBP4	214	0.197	0.0235	No
2	ALDH1A3	243	0.19	0.0549	No
3	CSPG5	338	0.176	0.0806	No
4	ACTA2	505	0.163	0.1005	No

5	KCNK12	594	0.157	0.1232	No
6	ADARB2	1085	0.134	0.1224	No
7	NDNF	1511	0.12	0.1225	No
8	SLITRK6	1561	0.119	0.1406	No
9	CRYAB	2988	0.092	0.0871	No
10	COL1A2	3352	0.087	0.0844	No
11	CAV1	3378	0.086	0.098	No
12	NPAS4	3529	0.084	0.1051	No
13	CTCFL	4163	0.077	0.0876	No
14	NPAS2	4691	0.071	0.0743	No
15	COL3A1	4916	0.069	0.0753	No
16	TINAGL1	4919	0.069	0.087	No
17	PLPP2	5298	0.065	0.0798	No
18	SOSTDC1	5383	0.064	0.0868	No
19	FAM182B	5670	0.062	0.0835	No
20	EDNRA	5760	0.061	0.0897	No
21	PLPP4	6363	0.056	0.0701	No
22	RORB	6509	0.055	0.0725	No
23	ACTG2	6516	0.055	0.0816	No
24	RBFOX3	6668	0.054	0.0835	No
25	OPRM1	7049	0.051	0.0738	No
26	ELMOD1	7111	0.05	0.0795	No
27	DOCK10	8183	0.043	0.0349	No
28	SDR42E1	8357	0.041	0.0336	No
29	UNC13C	8656	0.039	0.0258	No
30	CHST15	8775	0.038	0.0267	No
31	TLE4	8885	0.038	0.0278	No
32	NGEF	8921	0.037	0.0326	No
33	SLC6A1	9004	0.037	0.0349	No
34	TMEM132C	9007	0.037	0.0412	No
35	COL4A6	10207	0.028	-0.0122	No
36	NTS	10276	0.028	-0.0107	No
37	ENC1	10990	0.023	-0.0414	No
38	GPR26	11234	0.021	-0.0496	No
39	MEIS2	11290	0.021	-0.0488	No
40	PLP1	11438	0.02	-0.0526	No
41	IFITM3	12043	0.016	-0.0792	No
42	ANO3	12173	0.015	-0.083	No
43	LAMA3	12249	0.014	-0.0842	No
44	FAT1	12390	0.013	-0.0888	No
45	TAGLN	13199	0.008	-0.1267	No
46	LINC01561	13575	0.005	-0.144	No
47	HLA-A	14157	0	-0.1722	No
48	TGFBI	14205	0	-0.1745	No
49	COLEC12	14229	0	-0.1755	No
50	GPRC5A	14433	-0.002	-0.1851	No
51	TNS3	14444	-0.002	-0.1852	No
52	CIT	14536	-0.003	-0.1891	No
53	GAS8-AS1	14925	-0.006	-0.2069	No
54	CYP4V2	15294	-0.009	-0.2232	No
55	GPRIN3	15307	-0.009	-0.2222	No
56	SCRG1	15669	-0.013	-0.2376	No
57	NWD2	15834	-0.014	-0.2431	No
58	GPR3	15957	-0.016	-0.2464	No
59	CD99	16246	-0.018	-0.2572	No
60	PPARGC1A	16532	-0.022	-0.2673	No
61	MYO3A	16903	-0.027	-0.2807	No
62	SHISA6	17063	-0.029	-0.2834	No
63	ZNF700	17514	-0.036	-0.2991	Yes

64	PDIA2	17518	-0.036	-0.2931	Yes
65	MYOF	17533	-0.036	-0.2875	Yes
66	IL1B	17597	-0.037	-0.2841	Yes
67	FN1	17955	-0.044	-0.2939	Yes
68	BDNF	18175	-0.048	-0.2962	Yes
69	ZNF558	18198	-0.049	-0.2889	Yes
70	GRM3	18555	-0.058	-0.2963	Yes
71	BGN	18764	-0.064	-0.2955	Yes
72	GRM5	18768	-0.064	-0.2847	Yes
73	SORCS2	18883	-0.068	-0.2786	Yes
74	DGKG	18964	-0.07	-0.2704	Yes
75	RGS6	19024	-0.072	-0.2608	Yes
76	CYP1B1	19082	-0.075	-0.2508	Yes
77	DHRS3	19386	-0.087	-0.2505	Yes
78	RNASE3	20161	-0.154	-0.2616	Yes
79	MYLK	20286	-0.178	-0.237	Yes
80	IGFBP7	20368	-0.2	-0.2066	Yes
81	SPARC	20407	-0.212	-0.172	Yes
82	MYL9	20555	-0.279	-0.1312	Yes
83	THBS1	20597	-0.318	-0.0785	Yes
84	CALD1	20669	-0.481	0.0007	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	CREB5	317	0.18	0.0469	No
2	DNASE1L3	1136	0.133	0.0531	No
3	ST3GAL6	1440	0.122	0.0807	No
4	SPINK2	2868	0.094	0.0441	No
5	LHX2	4067	0.078	0.0131	No
6	CCL20	4960	0.069	-0.0064	No
7	MUC19	5887	0.06	-0.0305	No
8	PIEZO2	6350	0.056	-0.0334	No
9	SV2B	7184	0.05	-0.0565	No
10	DSC3	8480	0.04	-0.1052	No
11	UNC13C	8656	0.039	-0.1002	No
12	WNT2	9600	0.033	-0.1346	No
13	RIMBP2	10883	0.023	-0.1886	No
14	IL6ST	11897	0.017	-0.2319	No
15	ENTHD1	13747	0.004	-0.3202	No
16	PDCD4	13841	0.003	-0.3237	No
17	ATXN7L3B	13984	0.002	-0.3299	No
18	DSC2	14793	-0.005	-0.3674	No
19	GRIA3	15462	-0.011	-0.3961	No
20	BCL2L11	15532	-0.011	-0.3955	No
21	PRKCH	16151	-0.018	-0.4194	No
22	NFAT5	16390	-0.02	-0.4239	No
23	NCR2	16649	-0.023	-0.4283	No
24	VPS35	16795	-0.025	-0.4267	No
25	PEX13	18455	-0.055	-0.4879	Yes
26	CXCL10	18562	-0.058	-0.4731	Yes
27	HOPX	18889	-0.068	-0.4654	Yes
28	FRMD4B	19071	-0.074	-0.4484	Yes
29	WSB1	19462	-0.092	-0.4355	Yes
30	MYB	19935	-0.126	-0.4147	Yes
31	EOMES	19968	-0.129	-0.3716	Yes
32	XCL1	20379	-0.202	-0.3216	Yes
33	LRP12	20464	-0.232	-0.2454	Yes
34	KLRC1	20544	-0.273	-0.1547	Yes
35	DNM3	20668	-0.466	0.0007	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	XCL1	44	0.521	0.1087	No
2	CXCL10	63	0.475	0.2089	No

3	ZNF503	218	0.305	0.2662	No
4	CTSZ	400	0.244	0.3093	No
5	IGHV3-73	949	0.17	0.3189	No
6	STEAP1	1426	0.138	0.3252	No
7	VPS35	5929	0.031	0.1138	No
8	SYTL2	6848	0.021	0.0737	No
9	HS3ST3B1	7973	0.01	0.0214	No
10	GPC4	9181	0	-0.037	No
11	CCND2	9712	-0.005	-0.0616	No
12	MGA	11115	-0.016	-0.126	No
13	ARHGDIA	13794	-0.041	-0.2469	No
14	PTPN14	15888	-0.065	-0.3344	No
15	SPEN	17026	-0.082	-0.3719	No
16	KCNJ15	18489	-0.114	-0.4185	No
17	RBPMS	19872	-0.175	-0.4482	Yes
18	KCTD14	20102	-0.196	-0.4176	Yes
19	SORBS2	20112	-0.197	-0.3762	Yes
20	TBX15	20258	-0.217	-0.337	Yes
21	RGS13	20619	-0.369	-0.276	Yes
22	MAB21L3	20624	-0.386	-0.1942	Yes
23	PDZD2	20632	-0.396	-0.1104	Yes
24	MAP1B	20663	-0.53	0.001	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	PARD3-AS1	957	0.115	-0.0163	No
2	SLAMF8	1084	0.108	0.0059	No
3	MT1X	4830	0.035	-0.1664	No
4	FCGR3A	5677	0.027	-0.2003	No
5	SLC18A1	7848	0.011	-0.3027	No
6	TEKT5	10493	-0.006	-0.4292	No
7	CD44	11132	-0.01	-0.4576	No
8	DUOX2	11403	-0.011	-0.4677	No
9	MYO1G	11808	-0.014	-0.4836	No
10	IFITM3	13505	-0.025	-0.5593	No
11	SLC13A4	13653	-0.026	-0.5597	No
12	SLC11A1	13813	-0.027	-0.5604	No
13	IFITM2	14576	-0.033	-0.5888	No
14	GBP1	15490	-0.041	-0.6222	No
15	ITGB3	15922	-0.045	-0.6313	No
16	SPP1	16214	-0.048	-0.6328	No
17	PGC	16306	-0.049	-0.6244	No
18	OVOL1	16948	-0.056	-0.6407	No
19	S100A8	17841	-0.07	-0.6657	No
20	PRSS55	18272	-0.077	-0.6663	No
21	SFN	18880	-0.091	-0.672	Yes
22	GPR84	19024	-0.094	-0.6543	Yes
23	KHDC3L	19214	-0.1	-0.6374	Yes
24	SLC22A8	19503	-0.11	-0.6226	Yes
25	CLCF1	19663	-0.117	-0.5996	Yes
26	S100A9	19932	-0.131	-0.5782	Yes
27	BMP5	20025	-0.137	-0.5469	Yes
28	CHI3L2	20324	-0.165	-0.5182	Yes
29	COL2A1	20338	-0.167	-0.4753	Yes
30	SPPL2C	20376	-0.172	-0.4321	Yes
31	P2RX2	20391	-0.174	-0.3873	Yes
32	ADAMTS9	20500	-0.196	-0.3414	Yes
33	SERPINA1	20555	-0.213	-0.2882	Yes
34	STC1	20576	-0.223	-0.231	Yes
35	PIRT	20631	-0.254	-0.1673	Yes
36	HILPDA	20638	-0.259	-0.1	Yes
37	CCL2	20680	-0.391	0.0001	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	LGALS1	1355	0.334	0.0083	No
2	ASPN	3632	0.124	-0.0938	No
3	COL5A1	4785	0.075	-0.1422	No
4	ZC3HAV1L	6196	0.029	-0.218	No
5	COL6A2	6539	0.02	-0.2329	No
6	NID1	6974	0.01	-0.2559	No
7	ARHGAP29	7396	0.001	-0.2807	No
8	COL1A1	7543	-0.002	-0.2888	No
9	CYP1B1	8586	-0.022	-0.3446	No
10	DCN	8963	-0.029	-0.3592	No
11	LOX	9578	-0.04	-0.3849	No
12	RHOJ	10375	-0.056	-0.4174	No
13	IGF2	10733	-0.063	-0.4219	No
14	ITM2A	10884	-0.065	-0.4135	No
15	COL3A1	11032	-0.068	-0.4042	No
16	COL6A3	11330	-0.074	-0.4022	No
17	PDGFRA	11353	-0.075	-0.3838	No
18	COL14A1	11537	-0.078	-0.374	No
19	HAPLN1	12831	-0.108	-0.422	Yes
20	COL1A2	12883	-0.11	-0.3959	Yes
21	ZNF736	13103	-0.116	-0.3781	Yes
22	LAMA4	13233	-0.12	-0.354	Yes
23	LAMB1	13486	-0.129	-0.3348	Yes
24	INHBA	13824	-0.144	-0.3166	Yes
25	COL5A2	14096	-0.161	-0.2901	Yes
26	COL21A1	14123	-0.162	-0.2488	Yes
27	B3GALNT1	14126	-0.162	-0.206	Yes
28	PRRX1	14173	-0.165	-0.165	Yes
29	OGN	14252	-0.17	-0.1247	Yes
30	EDNRA	14928	-0.233	-0.1031	Yes
31	GLT8D2	15734	-0.354	-0.057	Yes
32	SGCD	16205	-0.478	0.0417	Yes

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	ABHD3	203	0.198	0.091	Yes
2	LRRC34	245	0.189	0.185	Yes
3	NAPRT	355	0.168	0.2654	Yes
4	LRRTM4	411	0.16	0.3442	Yes
5	CHRNA5	449	0.155	0.4213	Yes
6	FAM221A	939	0.116	0.4565	Yes
7	ZDHHC11	1280	0.101	0.4913	Yes
8	ZNF471	1520	0.091	0.5261	Yes
9	CEBPZOS	2134	0.074	0.5341	Yes
10	TMEM260	2683	0.063	0.5394	Yes
11	CCT6B	4648	0.037	0.463	No
12	AMOT	5179	0.031	0.4532	No
13	FAM228A	6021	0.024	0.4248	No
14	CPXM1	6232	0.022	0.426	No
15	C1orf52	6557	0.02	0.4205	No
16	PIN4	7075	0.016	0.4036	No
17	EPS15	7804	0.011	0.3739	No
18	EFCAB13	7925	0.01	0.3732	No
19	KCND1	10939	-0.009	0.2316	No
20	GCNT2	11475	-0.012	0.2117	No
21	ORAI2	11858	-0.014	0.2005	No
22	RRP7A	12185	-0.016	0.1929	No

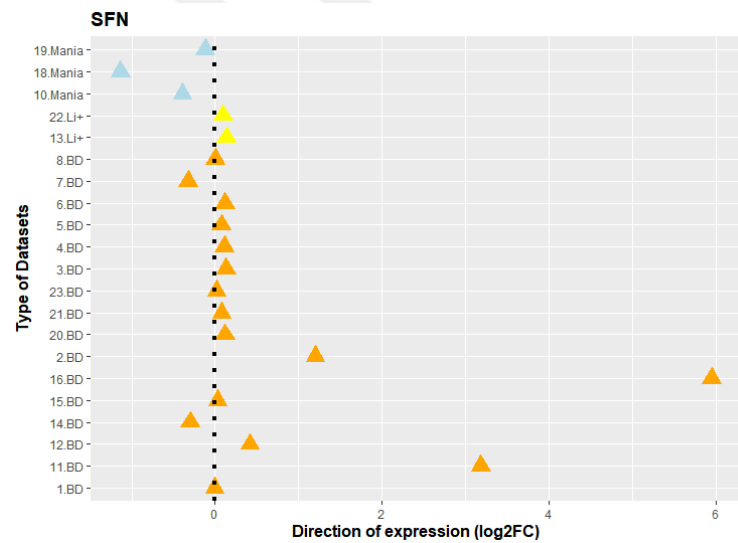
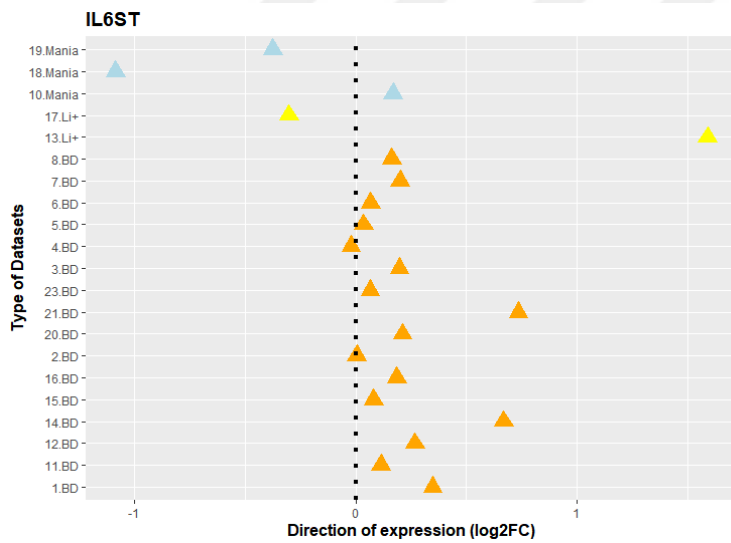
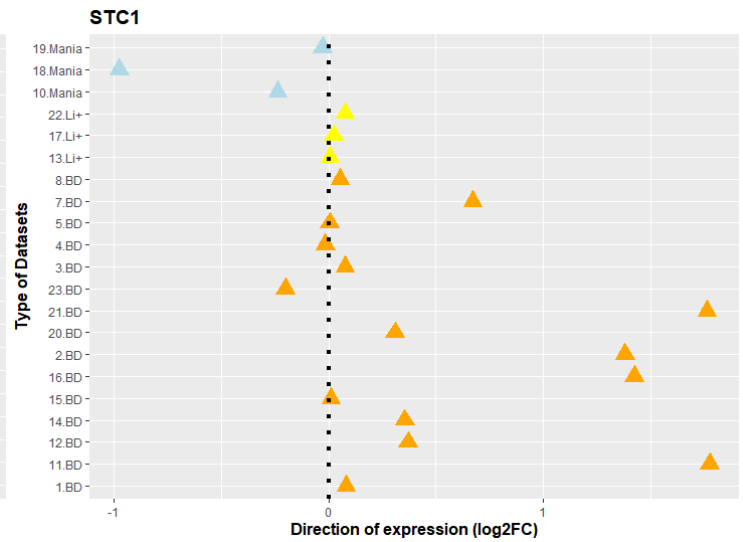
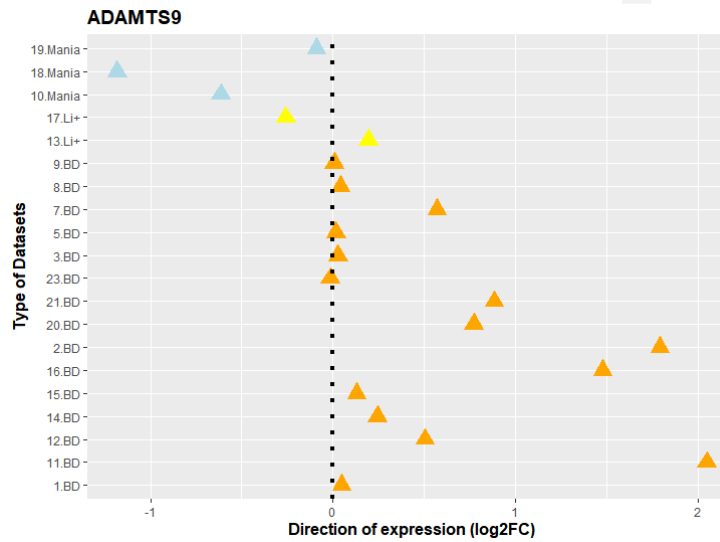
23	WDR43	12512	-0.018	0.1865	No
24	EFCAB11	13624	-0.025	0.1457	No
25	PQBP1	13914	-0.028	0.1457	No
26	TIMM17B	15589	-0.042	0.086	No
27	GNAT1	16899	-0.056	0.051	No
28	FABP6	17039	-0.057	0.0735	No
29	PIM2	17884	-0.07	0.0684	No
30	PRPH2	19940	-0.132	0.036	No

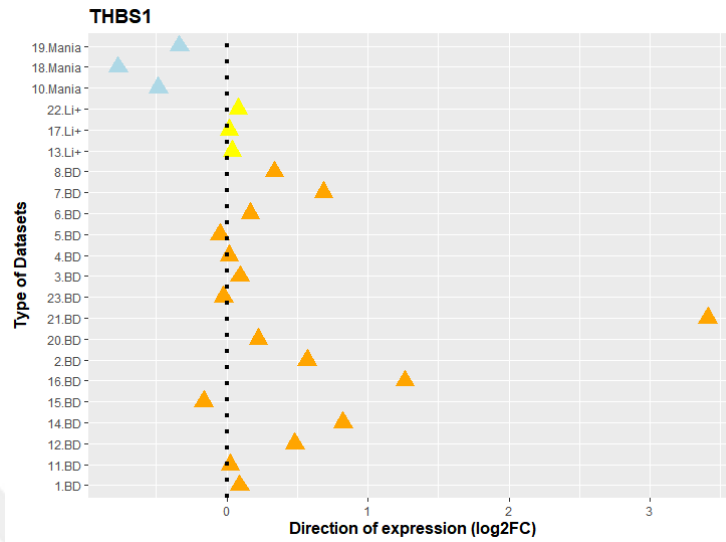


7.10. Appendix J

Log2FC plots of pattern 1 genes.

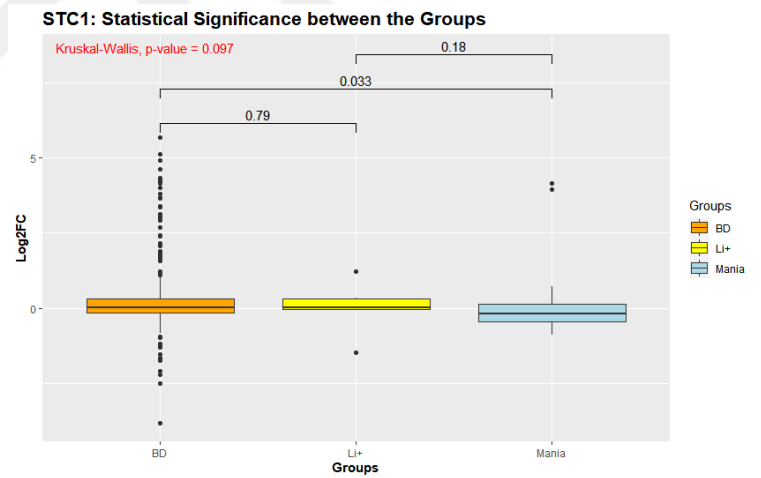
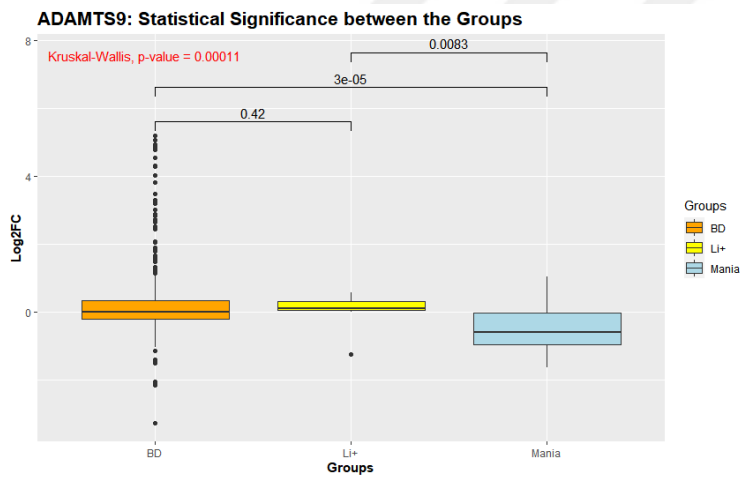
Types of Datasets

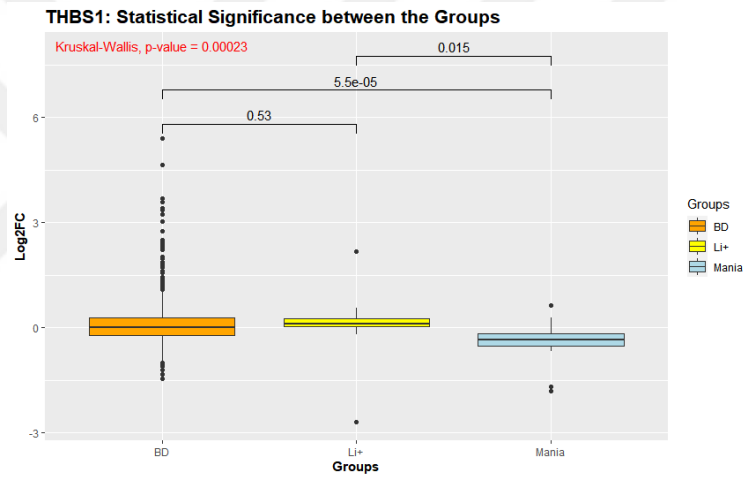
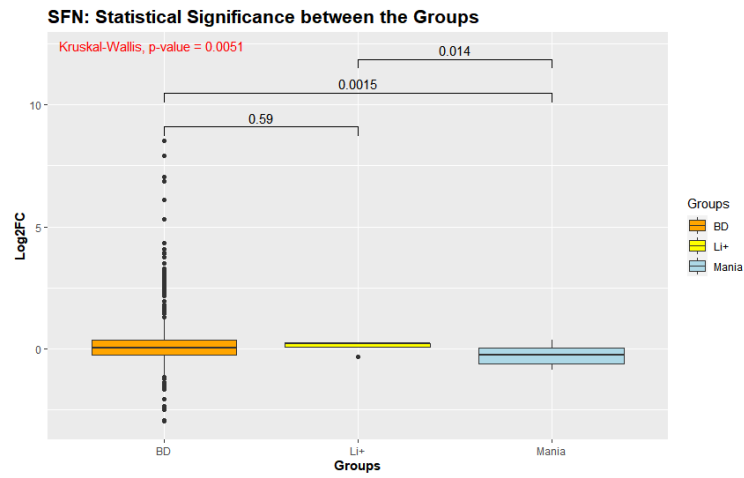
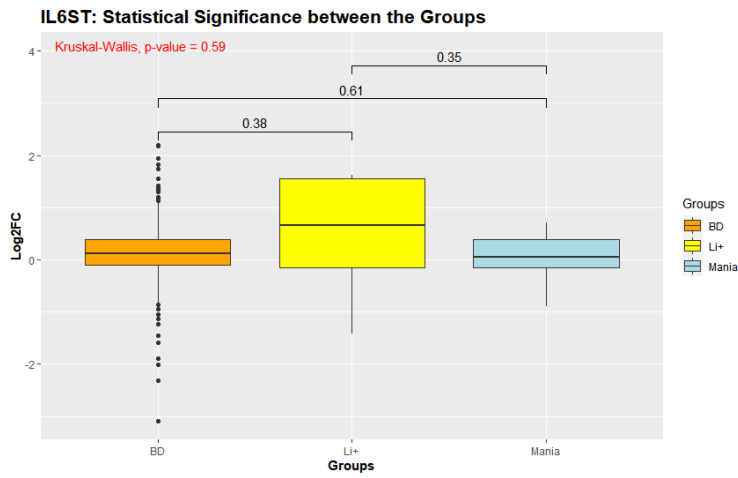




7.11. Appendix K

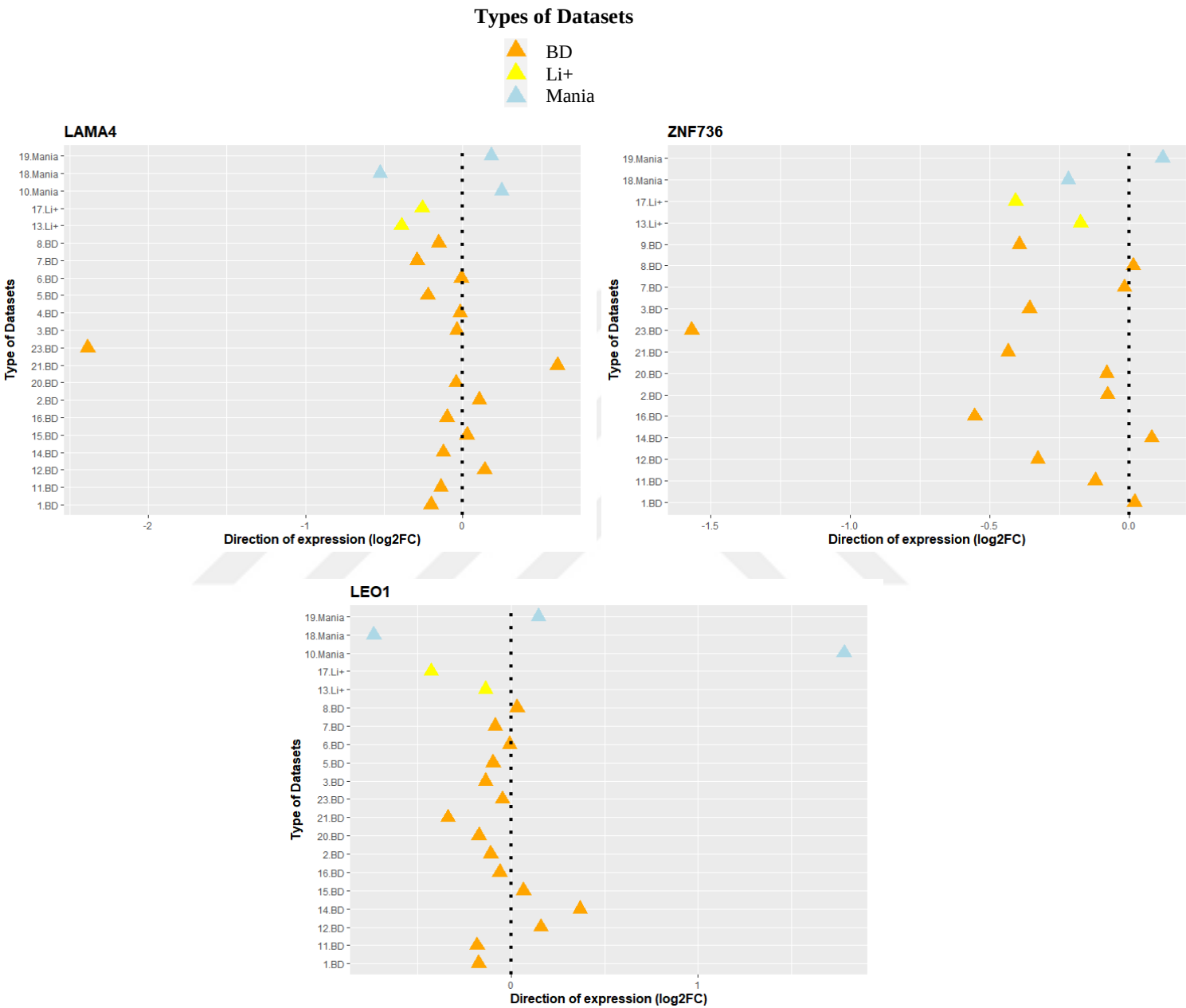
Non-parametric tests for pattern 1 genes.





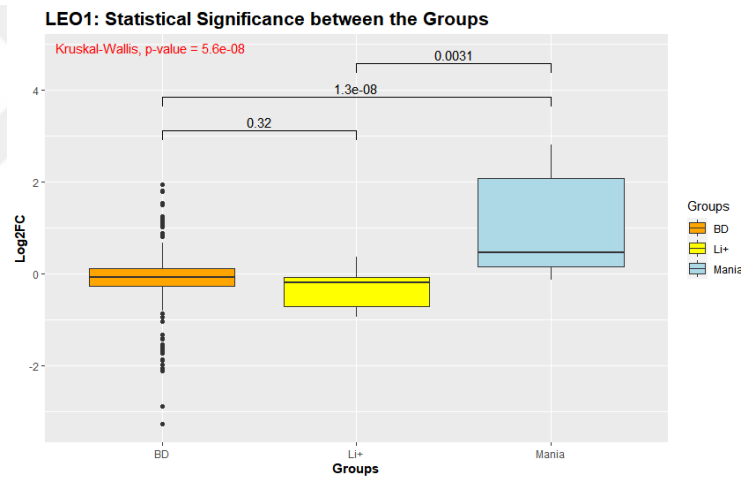
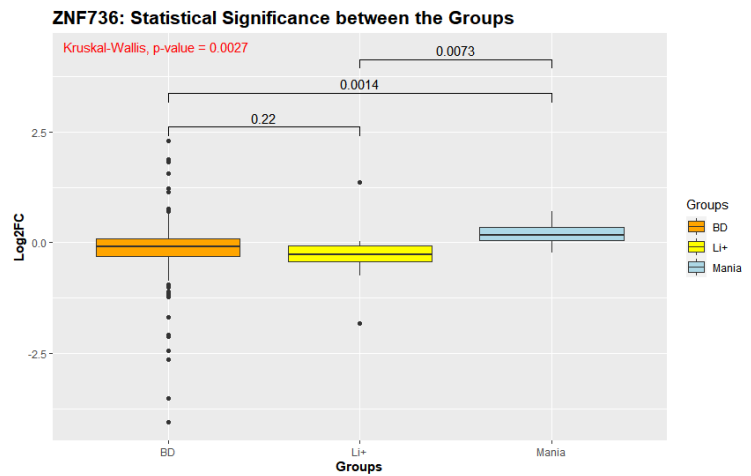
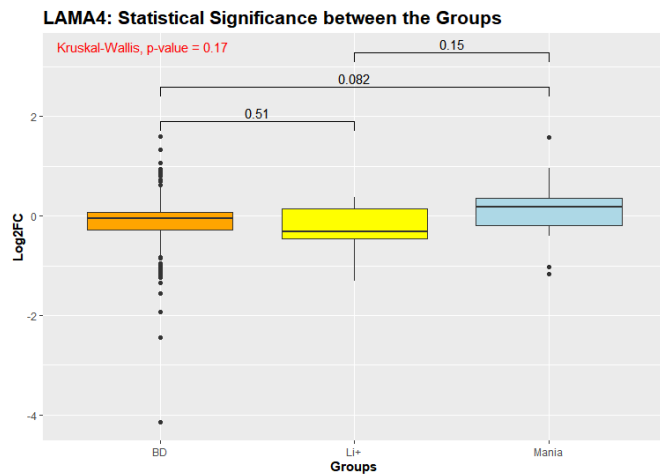
7.12. Appendix L

Log2FC plots of pattern 2 genes.



7.13. Appendix M

Non-parametric tests for pattern 2 genes.



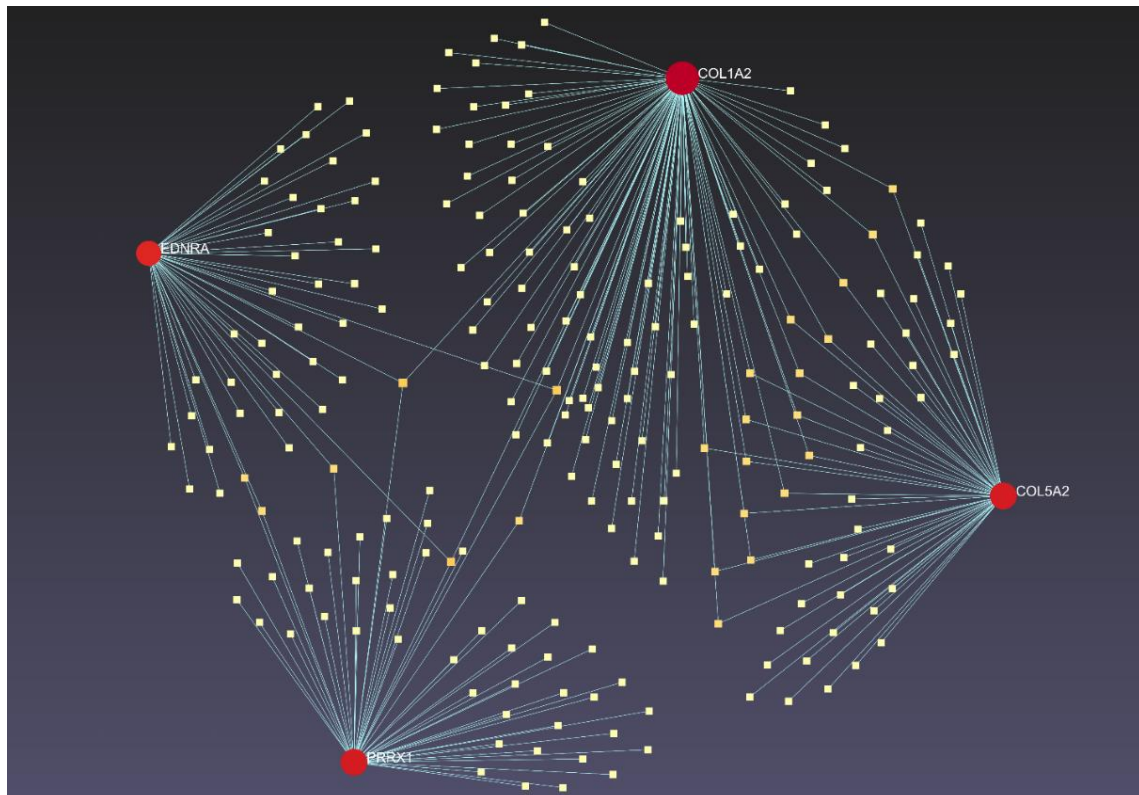
7.14. Appendix N

Subnetwork from *NetworkAnalyst* of pattern 1 genes showing direct association with BD.



7.15. Appendix O

Network analysis with pattern 2 genes.



7.16. Appendix P

Quantitative representation of network analysis with pattern 2 genes.

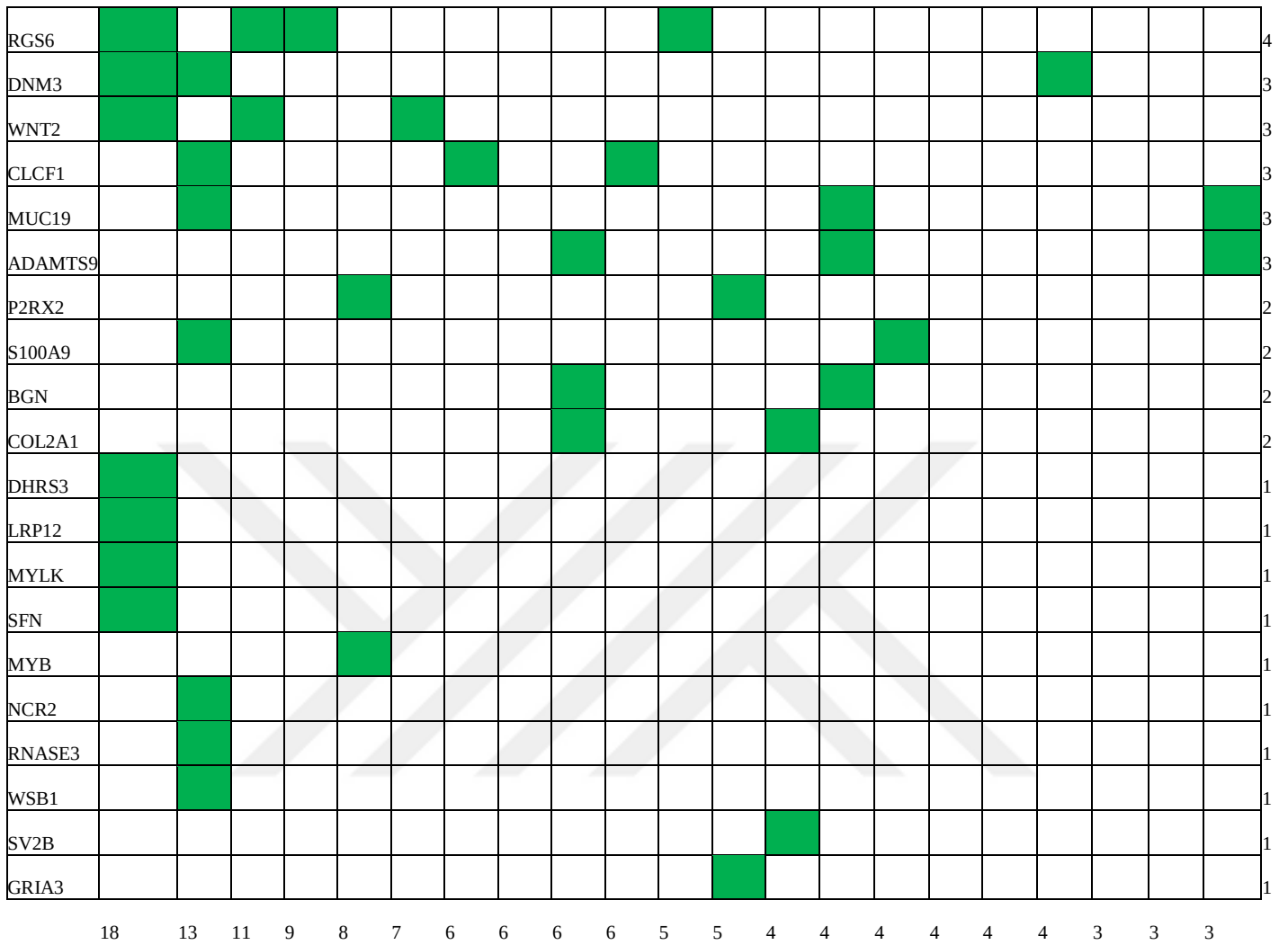
Label	Degree	Betweenness
COL1A2	99	18842.17
PRRX1	51	8951.4
COL5A2	51	6953
EDNRA	43	7612.43

Label	Degree	Betweenness
Micrognathism	3	3538.81
Mandibular hypoplasia	3	3538.81
Hypoplastic mandible condyle	3	3538.81
Autosomal recessive predisposition	2	1529.89
Mitral Valve Prolapse Syndrome	2	347.94
Degenerative polyarthritis	2	347.94
Curvature of spine	2	347.94
Atrophic scar	2	347.94
Skin hyperelastic	2	347.94
Velvety skin	2	347.94
Acquired flat foot	2	347.94
Short stature	2	347.94
Thin skin	2	347.94
Increased tendency to bruise	2	347.94
Blue sclera	2	347.94
Genu recurvatum	2	347.94
Acquired scoliosis	2	347.94
Soft skin	2	347.94
Poor wound healing	2	347.94
Pectus excavatum	2	347.94
Joint hyperflexibility	2	347.94
Cleft Palate	2	289.22
Conductive hearing loss	2	289.22
Uranostaphyloschisis	2	289.22

7.17. Appendix Q

Gene-ontology (GO) pathway analysis with pattern 1 genes.

Pathways	positive regulation of multicellular organismal process	negative regulation of multicellular organismal process	positive regulation of neurogenesis	positive regulation of nervous system development	antimicrobial humoral immune response mediated by antimicrobial response	positive regulation of gliogenesis	regulation of mononuclear cell migration	positive regulation of calcium ion transport	acute-phase response	neutrophil chemotaxis	killing of cells of other organism	response to vitamin	response to heat	behavioral response to pain	positive regulation of glial cell proliferation	positive regulation of transforming growth factor beta production	positive regulation of acute inflammatory response	positive regulation of vascular endothelial growth factor production	positive regulation of T cell migration	positive regulation of lymphocyte migration	positive regulation of T-helper 1 cell cytokine production
XCL1																					11
CXCL10																					10
IL1B																					10
CCL20																					6
HAMP																					6
IL6ST																					6
MYB																					6
THBS1																					6
FN1																					5
PRKCH																					5
CLCF1																					4
BDNF																					3
KHDC3L																					3
PTGER3																					3
STC1																					3
WNT2																					3
BMP5																					2
CYP1B1																					2
P2RX2																					2
PIRT																					2



7.19. Appendix S

Pattern 1 genes found in the blood. Peptide Atlas (PA) and Human Protein Atlas (HPA) were used to find which genes are present in the blood.

	UP BD-Li; Down Mania	Plasma Proteins PA	Secreted in blood HPA	Plasma Proteins HPA
1	IL6ST	.	.	.
2	SPARC	.	.	.
3	HAMP	.	.	.
4	SERPINA1	.	.	.
5	FN1	.	.	.
6	S100A9	.	.	.
7	IGFBP7	.	.	.
8	BGN	.	.	.
9	SFN	.	.	.
10	DSC2	.	.	.
11	STC1	.	.	.
12	DSC3	.	.	.
13	THBS1	.	.	.
14	XCL1	.	.	.
15	BDNF	.	.	.
16	CXCL10	.	.	.
17	CCL20	.	.	.
18	RNASE3	.	.	.
19	IL1B	.	.	.
20	CHI3L2	.	.	.
21	SPINK2	.	.	.
22	CLCF1	.	.	.
23	DNASE1L3	.	.	.
24	MYLK	.	.	.
25	FRMD4B	.	.	.
26	CALD1	.	.	.
27	SORCS2	.	.	.