Large-scale differential gene expression analysis identifies genes associated with Bipolar Disorder in post-mortem brain

Mohamed N. Omar *[1], Mohamed Youssef [2] and Mohamed Abdellatif[2]

[1,2]Faculty of Medicine, Mansoura University

ABSTRACT

Background and purpose

Bipolar disorder (BD) is a common psychiatric disorder with high morbidity and mortality. Several polymorphisms have been found to be implicated in the pathogenesis of BD, however, these loci have small effect sizes that fail to explain the high heritability of the disease. Here, we provide more insights into the genetic basis of BD by identifying the differentially expressed genes (DEGs) and their associated pathways and biological processes in post-mortem brain tissues of patients with BD.

Methods

Eight datasets were eligible for the differential expression analysis. We used six datasets for the discovery of the gene signature and used the other two for independent validation. We performed the multi-cohort analysis by a random-effect model using R and MetaIntegrator package.

*Corresponding author. Khalid Ibn El-Waleed St, Mansoura 35511, Egypt. Email: mohamed.n.omar1993@gmail.com

Results

The initial analysis resulted in the identification of 126 DEGs (30 up-regulated and 94 down-regulated). We refined this initial signature by a forward search process and resulted in the identification of 22 DEGs (6 up-regulated and 16 down-regulated). We validated the final gene signature in the independent datasets and resulted in an Area Under the ROC Curve (AUC) of 0.756 and 0.76, respectively. We performed gene set enrichment analysis (GSEA) which identified several biological processes and pathways related to BD including Ca transport, inflammation and DNA damage response.

Conclusion

Our findings support the previous findings that link BD pathogenesis to abnormalities in glial inflammation and calcium transport and also identify several other biological processes not previously reported to be associated with the development of the disease. Such findings will improve our understanding of the genetic basis underlying BD and may have future clinical implications.

Keywords— bipolar disorder, differential expression, inflammation, calcium transport, post-mortem brain

---

INTRODUCTION

Bipolar disorder (BD) is a common psychiatric disorder characterized by alternating periods of depression and abnormally elevated mood associated with cognitive impairments and increased impulsivity [1, 2, 3]. The lifetime prevalence of Bipolar disorder is approximately 2.4% with average age of onset of 25 years [4].

Aside from mental and cognitive impairments, BD is also associated with high cardiovascular morbidity and mortality [5, 6, 7]. This association was explained by the co-occurrence of risk factors or due to side effects of psychotropic medications [8].

On the neuro-anatomical level, neuro-imaging studies link BD to abnormalities in various areas of the brain, especially the frontal, prefrontal and parietal cortices with cortical thinning and increased ventricular volume [9, 10].

On the genetic level, twin studies showed that BD is a highly heritable disease with heritability

reaching up to 80% [11, 12]. The risk of getting the disease is 10-fold higher in the first degree relatives of those affected with BD compared to normal population. Although recent studies showed that several susceptibility loci with small effect sizes are associated with BD [13, 14], there is no much evidence supporting the role of specific genes in the pathogenesis of the disease [15].

The molecular mechanisms underlying the etiology of bipolar disorder remain unclear with recent evidence pointing toward a significant role of immune disturbances [16, 17], disrupted Ca dynamics [18, 19] and oxidative DNA damage in the pathogenesis of BD [20].

In this study, we attempt to identify an accurate and consistent gene signature of BD in post-mortem brain rather than detecting polymorphisms or copy number variants (CNVs). We also attempt to identify the biological pathways and processes of this gene signature. For this purpose, we analyze microarray gene expression data to identify the deferentially expressed genes (DEGs) between BD and healthy controls. Detecting those genes and pathways would offer a better understanding of the molecular pathogenesis of this disease and may have future therapeutic implications by targeting these genes or pathways.

METHODS

Data Collection

We performed an extensive search of the Gene Expression Omnibus (GEO) database [21] looking for microarray gene expression datasets from patients with BD and from healthy controls. We used the following search term "Bipolar disorder" and the following filters: entry type (series), study type (Expression profiling by array) and organism (Homo sapiens). This resulted in a total of 39 studies which were then filtered to include only gene expression studies in post-mortem brain tissues only (frontal, prefrontal and parietal cortices).

A total of eight datasets were eligible for the downstream analysis (Table 1). In all datasets, we removed samples from patients with schizophrenia and major depressive disorder and kept only samples from BD patients and controls. This resulted in a total of 509 samples with 248 BD and 261 control samples. We used six datasets (378 samples, 74.3% of the total samples) for discovery of the BD gene signature and kept two datasets (131 samples, 25.7% of the total samples) out for independent validation of the signature performance.

The Meta-Analysis

We used R Programming language [22] and the "MetaIntegrator" package [23] which utilizes a gene expression meta-analysis workflow described by Haynes et al [24]. The package can be used to retrieve the datasets from GEO, then it automatically checks if these datasets are appropriately normalized and log2-scaled. We found that two datasets (GSE12654 and GSE12649) were initially normalized using median centering so we downloaded the raw data and normalized them using Robust Multiarray Average (RMA) method which is the same method used to normalize the other Affymetrix datasets. All datasets were on log2-scale before conducting the meta-analysis.

The MetaIntegrator approach computes a Hedges effect size [25]for each gene in each dataset. These effect sizes are then pooled across all the datasets (Supp. figures1,2) using a random-effect model by assuming that results from each study is drawn from a single distribution and that each inter-study difference is just a random effect. The approach computes the log sum of p-values that each gene is up/down-regulated, then combine p-values using Fisher's method and finally performs Benjamini-Hochberg false discovery rate (FDR) correction across all genes [26].

In our analysis, a gene would be deferentially expressed if the absolute value of its effect size is greater than zero, the FDR is less than 5% across all discovery datasets and it has to be significantly up/down regulated in all of the six training datasets with a heterogeneity P-value cutoff of 0.05 [27]. All discovery steps were conducted on the training datasets only.

In order to optimize the initial gene signature, we performed a Forward Search process by taking the initial gene set, adding one gene at a time and calculating the weighted Area Under the ROC curve (AUC) resulting from the addition of this gene. Weighted AUC is calculated as:

$$W.auc = S.auc \times n$$

where W.auc is the weighted AUC, S.auc is the sum of AUC of each data set and n is the number of samples in the data set .

This process is repeated for each gene until the stopping threshold (0 in our case) is reached and the final set of genes will be those who contributed the most to the weighted AUC.

Validation of the Resulting Gene Signature

We tested the performance and consistency of the resulting BD gene signature in the two independent cohorts (Table 1). We used the Area Under the Receiver Operating Characteristic Curve (AUC), Area Under the Precision Recall Curve (AUPRC) and Matthews correlation coefficient (MCC) as evaluation metrics.

Enrichment Analysis

Functional Enrichment analysis was performed using Enrichr R package [28, 29] and the following databases: "GO-Biological-Process-2018" and "KEGG-2019-Human". The goal was to discover the biological processes and pathways enriched in the discovered gene set.

RESULTS

The BD Gene Signature

The initial meta-analysis resulted in the identification of 124 DEGs (30 up-regulated and 94 down-regulated). We refined this initial signature by using a forward search process which resulted in the identification of 22 DEGs (6 up-regulated and 16 down-regulated). Figure (1) shows the summary AUC of the discovery datasets with the initial gene signature and the summary AUC with the final signature resulting from forward search.

Of the 22 DEGs, 6 genes were up-regulated (*CTBP2, ELK1, ARHGAP5, RNASE4, MAOA, GOLGA8A*) (Supp. figure1) and 16 were down-regulated (*SST, IGFBP6, CD53, RABIF, DDX18, NXF1, CXCL12, ETV5, HCLS1, NFYC, SULF1, ZFYVE9, CSE1L, DNAJC7, FDFT1, TSN*) (Table 2) (Supp. figure2).

Strong Performance in the Two Validation Datasets

We measured the discriminatory performance of our gene signature by testing it in two independent datasets (N=131) (Fig. 2). In the first validation dataset (GSE35977), the signature achieved an AUC of 0.756 (95%CI: 0.66-0.85) (Fig. 2), AUPRC of 0.70 (95%CI: 0.57-0.83), accuracy of 73% (95% CI: 63%-81%, P=5.3e-05), sensitivity of 76% and MCC of 0.46.

In the second validation dataset (GSE53987), the signature achieved an AUC of 0.762 (95%CI: 0.60-0.92) (Fig. 3), AUPRC of 0.704 (95%CI: 0.49-0.72), accuracy of 72% (95%CI: 55%-86%, P=0.014), sensitivity of 53% and MCC of 0.46.

The signature performance in the validation datasets showed great consistency with the performance in the discovery datasets with minimal overfitting (Summary AUC = 0.81, 95% CI:0.75-0.86).

Enrichment Analysis

Gene set enrichment analysis (GSEA) has revealed several significant ($P < 0.05$) KEGG pathways and GO biological processes enriched in the discovered gene signature. The six up-regulated genes were associated with 26 KEGG pathways (Fig. 3) (Supp.table1) and 18 GO biological processes (Fig. 5) (Supp. table2). The 16 down-regulated genes were associated with 3 KEGG pathways (Fig. 4) (Supp. table3) together with 102 GO biological processes (Fig. 6) (Supp. table4).

DISCUSSION

Bipolar disorder is a highly heterogeneous disease and the genetic basis underlying its pathophysiology is complex [30]. Several genome-wide association studies (GWAS) have identified several copy number variants (CNVs) and single-nucleotide polymorphisms (SNPs) that seem to play a role in the high heritability of the disease but each of these variants has a small effect [31, 32]. Moreover, most of these studies failed to identify specific genes that may have a large effect in the pathogenesis of the disease, most likely due to small sample size.

Here we attempted to overcome the problem of inadequate sample size by integrating multiple microarray gene expression studies in a meta-analysis process to achieve high statistical power. Through this large-scale gene expression analysis of 509 post-mortem brain samples, we managed to identify six up-regulated and sixteen down-regulated genes in the post-mortem brain tissues of patients with BD. The resulting gene signature proved to be accurate and consistent when tested in two independent validation datasets.

GSEA identified important pathways and processes enriched in the discovered BD gene signature. The up-regulated genes were particularly associated with important cell signaling pathways like notch and Wnt signaling pathways which were previously reported to have a significant role

in the pathogenesis of BD together with other psychiatric disorders [33, 34, 35, 36]. Additionally they were also associated with pathways involved in inflammatory cell motility like focal adhesion and leukocyte transendothelial migration pathways. The GO biological processes enriched in the up-regulated genes included dopamine and catecholamine catabolic processes together with processes that regulate signal transduction like Rho protein signal transduction and GTPase-mediated signal transduction supporting the current evidence that links BD to alterations in the cytoskeletal dynamics [37, 38]. Other biological processes included those regulating transcription and mRNA cleavage.

On the other hand, the down-regulated genes were associated with processes that regulate apoptosis, Ca transport, cell migration, granulocytes differentiation, chemotaxis and lymphocytes migration, supporting the notion that glial inflammation/loss, imbalances in the apoptotic process and Ca homeostasis may have a role in the pathogenesis of BD [39, 40, 41, 42, 18].

Most of these biological processes and pathways enriched in the BD gene signature overlap with those responsible for the development of atherosclerosis [43, 44]. This suggests that the association between BD and cardiovascular diseases may have some genetic basis rather than shared environmental risk factors or side effects of psychotropic medications.

One limitation of this study is the lack of experimental validation in another independent cohort. Although our signature showed great performance and consistency in the validation datasets, this assessment was purely computational and it needs experimental validation using real-time PCR to assess its clinical validity.

CONCLUSION

In this study, we provide more insights into the genetic basis of BD through the identification of DEGs in post-mortem brain tissues of BD patients compared with normal controls. Our discovered BD gene signature proved to be accurate and consistent as shown by its performance in the independent validation cohorts. The BD signature genes were associated with a number of biological processes and pathways, most importantly the pathways regulating Ca transport, immune response, oxidative damage and inflammatory processes. These pathways are also involved in the pathogenesis of atherosclerosis, cardiovascular diseases and metabolic syndrome. Such findings, together with

previous research will provide a more clear understanding of the molecular pathophysiology of BD and may have future therapeutic potentials by targeting these specific genes and pathways.

List of Abbreviations

- **BD**: Bipolar Disorder.
- **DEGs**: Differentially Expressed Genes.
- **GEO**: Gene Expression Omnibus.
- **ROC**: Receiver Operating Characteristic.
- **AUC**: Area Under the ROC Curve.
- **AUPRC**: Area Under the Precision Recall Curve.
- **MCC**: Matthews Correlation Coefficient.
- **GO**: Gene Ontology.
- **KEGG**: Kyoto Encyclopedia of Genes and Genomes.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data sets used in this study are publicly available on the Gene Expression Omnibus (GEO) data base under the corresponding accession number. The code for this analysis is available on GitHub and can accessed using the following link: https://github.com/MohamedOmar2020/BD

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was received for this study.

Author's contributions

MO conceived the research question and performed the analysis. MY and MA collected the data sets. All authors contributed equally to the manuscript writing, read and approved the final manuscript.

Acknowledgements

We thank Dr.Mohamed Tarek Badr (University Medical Center Freiburg, Germany) for his valuable suggestions that greatly improved the manuscript.

References

[1] Ian M Anderson, Peter M Haddad, and Jan Scott. Bipolar disorder. *BMJ*, 345, 2012.

[2] Lucy J Robinson, Jill M Thompson, Peter Gallagher, Utpal Goswami, Allan H Young, I Nicol Ferrier, and P Brian Moore. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *J Affect Disord*, 93(1-3):105–115, Jul 2006.

[3] B Sole, C M Bonnin, C Torrent, A Martinez-Aran, D Popovic, R Tabares-Seisdedos, and E Vieta. Neurocognitive impairment across the bipolar spectrum. *CNS Neurosci Ther*, 18(3):194–200, Mar 2012.

[4] Lakshmi N Yatham, Sidney H Kennedy, Sagar V Parikh, Ayal Schaffer, David J Bond, Benicio N Frey, Verinder Sharma, Benjamin I Goldstein, Soham Rej, Serge Beaulieu, Martin Alda, Glenda MacQueen, Roumen V Milev, Arun Ravindran, Claire O'Donovan, Diane McIntosh, Raymond W Lam, Gustavo Vazquez, Flavio Kapczinski, Roger S McIntyre, Jan Kozicky, Shigenobu Kanba, Beny Lafer, Trisha Suppes, Joseph R Calabrese, Eduard Vieta, Gin Malhi, Robert M Post, and Michael Berk. Canadian network for mood and anxiety treatments (canmat) and international society for bipolar disorders (isbd) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord*, 20(2):97–170, Mar 2018.

[5] F Angst, H H Stassen, P J Clayton, and J Angst. Mortality of patients with mood disorders: follow-up over 34-38 years. *J Affect Disord*, 68(2-3):167–181, Apr 2002.

[6] Giuseppe Maina, Virginio Salvi, Alberto Vitalucci, Virginia D'Ambrosio, and Filippo Bogetto. Prevalence and correlates of overweight in drug-naive patients with bipolar disorder. *J Affect Disord*, 110(1-2):149–155, Sep 2008.

[7] M De Hert, J M Dekker, D Wood, K G Kahl, R I G Holt, and H-J Moller. Cardiovascular disease and diabetes in people with severe mental illness position statement from the european psychiatric association (epa), supported by the european association for the study of diabetes (easd) and the european society of cardiology (esc). *Eur Psychiatry*, 24(6):412–424, Sep 2009.

[8] Christoph U Correll, Julie B Penzner, Umesh H Parikh, Tahir Mughal, Tariq Javed, Maren Carbon, and Anil K Malhotra. Recognizing and monitoring adverse events of second-generation antipsychotics in children and adolescents. *Child Adolesc Psychiatr Clin N Am*, 15(1):177–206, Jan 2006.

[9] Scott C Fears, Remmelt Schür, Rachel Sjouwerman, Susan K Service, Carmen Araya, Xinia Araya, Julio Bejarano, Emma Knowles, Juliana Gomez-Makhinson, Maria C Lopez, Ileana Aldana, Terri M Teshiba, Zvart Abaryan, Noor B Al-Sharif, Linda Navarro, Todd A Tishler, Lori Altshuler, George Bartzokis, Javier I Escobar, David C Glahn, Paul M Thompson, Carlos Lopez-Jaramillo, Gabriel Macaya, Julio Molina, Victor I Reus, Chiara Sabatti, Rita M Cantor, Nelson B Freimer, and Carrie E Bearden. Brain structure-function associations in multi-generational families genetically enriched for bipolar disorder. *Brain*, 138(Pt 7):2087–2102, Jul 2015.

[10] Morgan Haldane, Giles Cunningham, Chris Androutsos, and Sophia Frangou. Structural brain correlates of response inhibition in bipolar disorder i. *J Psychopharmacol*, 22(2):138–143, Mar 2008.

[11] Jack Edvardsen, Svenn Torgersen, Espen Roysamb, Sissel Lygren, Ingunn Skre, Sidsel Onstad, and Per Anders Oien. Heritability of bipolar spectrum disorders. unity or heterogeneity? *J Affect Disord*, 106(3):229–240, Mar 2008.

[12] Nick Craddock and Pamela Sklar. Genetics of bipolar disorder. *Lancet*, 381(9878):1654–1662, May 2013.

[13] Seth A Ament, Szabolcs Szelinger, Gustavo Glusman, Justin Ashworth, Liping Hou, Nirmala Akula, Tatyana Shekhtman, Judith A Badner, Mary E Brunkow, Denise E Mauldin, Anna-Barbara Stittrich, Katherine Rouleau, Sevilla D Detera-Wadleigh, John I Jr Nurnberger, Howard J Edenberg, Elliot S Gershon, Nicholas Schork, Nathan D Price, Richard Gelinas, Leroy Hood, David Craig, Francis J McMahon, John R Kelsoe, and Jared C Roach. Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci U S A*, 112(11):3576–3581, Mar 2015.

[14] M Kataoka, N Matoba, T Sawada, A-A Kazuno, M Ishiwata, K Fujii, K Matsuo, A Takata, and T Kato. Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. *Mol Psychiatry*, 21(7):885–893, Jul 2016.

[15] Michael C O'Donovan and Michael J Owen. The implications of the shared genetics of psychiatric disorders. *Nat Med*, 22(11):1214–1219, Nov 2016.

[16] Sinead M O'Brien, Paul Scully, Lucinda V Scott, and Timothy G Dinan. Cytokine profiles in bipolar affective disorder: focus on acutely ill patients. *J Affect Disord*, 90(2-3):263–267, Feb 2006.

[17] Elisa Brietzke, Laura Stertz, Brisa Simoes Fernandes, Marcia Kauer-Sant'anna, Marcello Mascarenhas, Andreia Escosteguy Vargas, Jose Artur Chies, and Flavio Kapczinski. Comparison of cytokine levels in depressed, manic and euthymic patients with bipolar disorder. *J Affect Disord*, 116(3):214–217, Aug 2009.

[18] Michael J Berridge. Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia. *Cell Tissue Res*, 357(2):477–492, Aug 2014.

[19] Xiaoyun Guo, Dengtang Liu, Tong Wang, and Xingguang Luo. Aetiology of bipolar disorder: contribution of the l-type voltage-gated calcium channels. *Gen Psychiatr*, 32(2):e100009, 2019.

[20] Klaus Munkholm, Henrik Enghusen Poulsen, Lars Vedel Kessing, and Maj Vinberg. Elevated levels of urinary markers of oxidatively generated dna and rna damage in bipolar disorder. *Bipolar Disord*, 17(3):257–268, May 2015.

[21] Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, and Alexandra Soboleva. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1):D991–D995, 11 2012.

[22] Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. 5(3):299–314, 1996.

[23] Purvesh Khatri, Silke Roedder, Naoyuki Kimura, Katrien De Vusser, Alexander A. Morgan, Yongquan Gong, Michael P. Fischbein, Robert C. Robbins, Maarten Naesens, Atul J. Butte, and Minnie M. Sarwal. A common rejection module (crm) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *Journal of Experimental Medicine*, 210(11):2205–2221, 2013.

[24] Winston A Haynes, Francesco Vallania, Charles Liu, Erika Bongen, Aurelie Tomczak, Marta Andres-Terrè, Shane Lofgren, Andrew Tam, Cole A Deisseroth, Matthew D Li, Timothy E Sweeney, and Purvesh Khatri. Empowering multi-cohort gene expression analysis to increase reproducibility. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 22:144–153, 2017.

[25] Larry V. Hedges. Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2):107–128, 1981.

[26] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 2019/07/24 1995.

[27] David C. Hoaglin. Misunderstandings about q and 'cochran's q test' in meta-analysis. *Statistics in Medicine*, 35(4):485–495, 2016.

[28] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, Michael G McDermott, Caroline D Monteiro, Gregory W Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, 44(W1):W90–7, Jul 2016.

[29] Edward Y Chen, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R Clark, and Avi Ma'ayan. Enrichr: interactive and collaborative html5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14:128, Apr 2013.

[30] Nick Craddock and Pamela Sklar. Genetics of bipolar disorder. *Lancet*, 381(9878):1654–1662, May 2013.

[31] Liping Hou, Sarah E Bergen, Nirmala Akula, Jie Song, Christina M Hultman, Mikael Landén, Mazda Adli, Martin Alda, Raffaella Ardau, Bárbara Arias, Jean-Michel Aubry, Lena Backlund, Judith A Badner, Thomas B Barrett, Michael Bauer, Bernhard T Baune, Frank Bellivier, Antonio Benabarre, Susanne Bengesser, Wade H Berrettini, Abesh Kumar Bhattacharjee, Joanna M Biernacka, Armin Birner, Cinnamon S Bloss, Clara Brichant-Petitjean, Elise T Bui,

William Byerley, Pablo Cervantes, Caterina Chillotti, Sven Cichon, Francesc Colom, William Coryell, David W Craig, Cristiana Cruceanu, Piotr M Czerski, Tony Davis, Alexandre Dayer, Franziska Degenhardt, Maria Del Zompo, J Raymond DePaulo, Howard J Edenberg, Bruno Étain, Peter Falkai, Tatiana Foroud, Andreas J Forstner, Louise Frisén, Mark A Frye, Janice M Fullerton, Sébastien Gard, Julie S Garnham, Elliot S Gershon, Fernando S Goes, Tiffany A Greenwood, Maria Grigoroiu-Serbanescu, Joanna Hauser, Urs Heilbronner, Stefanie Heilmann-Heimbach, Stefan Herms, Maria Hipolito, Shashi Hitturlingappa, Per Hoffmann, Andrea Hofmann, Stephane Jamain, Esther Jiménez, Jean-Pierre Kahn, Layla Kassem, John R Kelsoe, Sarah Kittel-Schneider, Sebastian Kliwicki, Daniel L Koller, Barbara König, Nina Lackner, Gonzalo Laje, Maren Lang, Catharina Lavebratt, William B Lawson, Marion Leboyer, Susan G Leckband, Chunyu Liu, Anna Maaser, Pamela B Mahon, Wolfgang Maier, Mario Maj, Mirko Manchia, Lina Martinsson, Michael J McCarthy, Susan L McElroy, Melvin G McInnis, Rebecca McKinney, Philip B Mitchell, Marina Mitjans, Francis M Mondimore, Palmiero Monteleone, Thomas W Mühleisen, Caroline M Nievergelt, Markus M Nöthen, Tomas Novák, John I Jr Nurnberger, Evaristus A Nwulia, Urban Ösby, Andrea Pfennig, James B Potash, Peter Propping, Andreas Reif, Eva Reininghaus, John Rice, Marcella Rietschel, Guy A Rouleau, Janusz K Rybakowski, Martin Schalling, William A Scheftner, Peter R Schofield, Nicholas J Schork, Thomas G Schulze, Johannes Schumacher, Barbara W Schweizer, Giovanni Severino, Tatyana Shekhtman, Paul D Shilling, Christian Simhandl, Claire M Slaney, Erin N Smith, Alessio Squassina, Thomas Stamm, Pavla Stopkova, Fabian Streit, Jana Strohmaier, Szabolcs Szelinger, Sarah K Tighe, Alfonso Tortorella, Gustavo Turecki, Eduard Vieta, Julia Volkert, Stephanie H Witt, Adam Wright, Peter P Zandi, Peng Zhang, Sebastian Zollner, and Francis J McMahon. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum Mol Genet*, 25(15):3383–3394, Aug 2016.

[32] Consortium Cross-Disorder Group of the Psychiatric Genomics. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, 381(9875):1371–1379, Apr 2013.

[33] Eva Z Hoseth, Florian Krull, Ingrid Dieset, Ragni H Mørch, Sigrun Hope, Erlend S Gardsjord, Nils Eiel Steen, Ingrid Melle, Hans-Richard Brattbakk, Vidar M Steen, Pål Aukrust,

Srdjan Djurovic, Ole A Andreassen, and Thor Ueland. Exploring the wnt signaling pathway in schizophrenia and bipolar disorder. *Transl Psychiatry*, 8(1):55, Mar 2018.

[34] Kimberly A Mulligan and Benjamin N R Cheyette. Neurodevelopmental perspectives on wnt signaling in psychiatry. *Mol Neuropsychiatry*, 2(4):219–246, Feb 2017.

[35] Eva Z Hoseth, Florian Krull, Ingrid Dieset, Ragni H Mørch, Sigrun Hope, Erlend S Gardsjord, Nils Eiel Steen, Ingrid Melle, Hans-Richard Brattbakk, Vidar M Steen, Pål Aukrust, Srdjan Djurovic, Ole A Andreassen, and Thor Ueland. Attenuated notch signaling in schizophrenia and bipolar disorder. *Sci Rep*, 8(1):5349, Mar 2018.

[36] Inti Pedroso, Anbarasu Lourdusamy, Marcella Rietschel, Markus M Nothen, Sven Cichon, Peter McGuffin, Ammar Al-Chalabi, Michael R Barnes, and Gerome Breen. Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. *Biol Psychiatry*, 72(4):311–317, Aug 2012.

[37] Yarema Bezchlibnyk and L Trevor Young. The neurobiology of bipolar disorder: Focus on signal transduction pathways and the regulation of gene expression. *The Canadian Journal of Psychiatry*, 47(2):135–148, 2019/10/27 2002.

[38] Gregory S. Berns and Charles B. Nemeroff. The neurobiology of bipolar disorder. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 123C(1):76–84, 2003.

[39] Jairo Vinicius Pinto, Ives Cavalcante Passos, Diego Librenza-Garcia, Grasiela Marcon, Maiko Abel Schneider, Joao Henrique Conte, Joao Pedro Abreu da Silva, Luiza Pereira Lima, Andre Quincozes-Santos, Marcia Kauer-Sant Anna, and Flavio Kapczinski. Neuron-glia interaction as a possible pathophysiological mechanism of bipolar disorder. *Curr Neuropharmacol*, 16(5):519–532, 2018.

[40] G Z Reus, G R Fries, L Stertz, M Badawy, I C Passos, T Barichello, F Kapczinski, and J Quevedo. The role of inflammation and microglial activation in the pathophysiology of psychiatric disorders. *Neuroscience*, 300:141–154, Aug 2015.

[41] Tobias Rowland, Benjamin I Perry, Rachel Upthegrove, Nicholas Barnes, Jayanta Chatterjee, Daniel Gallacher, and Steven Marwaha. Neurotrophins, cytokines, oxidative stress mediators

and mood state in bipolar disorder: systematic review and meta-analyses. *Br J Psychiatry*, 213(3):514–525, Sep 2018.

[42] Jerry J. Warsh, Stavroula Andreopoulos, and Peter P. Li. Role of intracellular calcium signaling in the pathophysiology and pharmacotherapy of bipolar disorder: current status. *Clinical Neuroscience Research*, 4(3):201 – 213, 2004. Lithium and Mood Stabilizers: Mechanisms of Action.

[43] Yuri V Bobryshev, Ekaterina A Ivanova, Dimitry A Chistiakov, Nikita G Nikiforov, and Alexander N Orekhov. Macrophages and their role in atherosclerosis: Pathophysiology and transcriptome analysis. *Biomed Res Int*, 2016:9582430, 2016.

[44] Bin Li, Weihong Li, Xiaoli Li, and Hong Zhou. Inflammation: A novel therapeutic target/direction in atherosclerosis. *Curr Pharm Des*, 23(8):1216–1227, 2017.

[45] Hamid M Abdolmaleky, Adam C Gower, Chen-Khuan Wong, Jiayi W Cox, Xiaoling Zhang, Arunthathi Thiagalingam, Rahim Shafa, Vadivelu Sivaraman, Jin-Rong Zhou, and Sam Thiagalingam. Aberrant transcriptomes and dna methylomes define pathways that drive pathogenesis and loss of brain laterality/asymmetry in schizophrenia and bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*, 180(2):138–149, Mar 2019.

[46] Dominique Arion, Zhiguang Huo, John F Enwright, John P Corradi, George Tseng, and David A Lewis. Transcriptome alterations in prefrontal pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders. *Biol Psychiatry*, 82(8):594–600, Oct 2017.

[47] K Iwamoto, C Kakiuchi, M Bundo, K Ikeda, and T Kato. Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Mol Psychiatry*, 9(4):406–416, Apr 2004.

[48] Angelica de Baumont, Mariana Maschietto, Leandro Lima, Dirce Maria Carraro, Eloisa Helena Olivieri, Alex Fiorini, Luiz Andre Nardin Barreta, Joana Almeida Palha, Paulo Belmonte-de Abreu, Carlos Alberto Moreira Filho, and Helena Brentani. Innate immune response is differentially dysregulated between bipolar disease and schizophrenia. *Schizophr Res*, 161(2-3):215–221, Feb 2015.

[49] M M Ryan, H E Lockstone, S J Huffaker, M T Wayland, M J Webster, and S Bahn. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry*, 11(10):965–978, Oct 2006.

[50] C Chen, L Cheng, K Grennan, F Pibiri, C Zhang, J A Badner, E S Gershon, and C Liu. Two gene co-expression modules differentiate psychotics and controls. *Mol Psychiatry*, 18(12):1308–1314, Dec 2013.

[51] Chao Chen, Qingtuan Meng, Yan Xia, Chaodong Ding, Le Wang, Rujia Dai, Lijun Cheng, Preethi Gunaratne, Richard A Gibbs, Shishi Min, Cristian Coarfa, Jeffrey G Reid, Chunling Zhang, Chuan Jiao, Yi Jiang, Gina Giase, Amber Thomas, Dominic Fitzgerald, Tonya Brunetti, Annie Shieh, Cuihua Xia, Yongjun Wang, Yunpeng Wang, Judith A Badner, Elliot S Gershon, Kevin P White, and Chunyu Liu. The transcription factor pou3f2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci Transl Med*, 10(472), Dec 2018.

[52] Thomas A Lanz, Veronica Reinhart, Mark J Sheehan, Stacey J Sukoff Rizzo, Susan E Bove, Larry C James, Dmitri Volfson, David A Lewis, and Robin J Kleiman. 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. *Transl Psychiatry*, 9(1):151, May 2019.

Figures

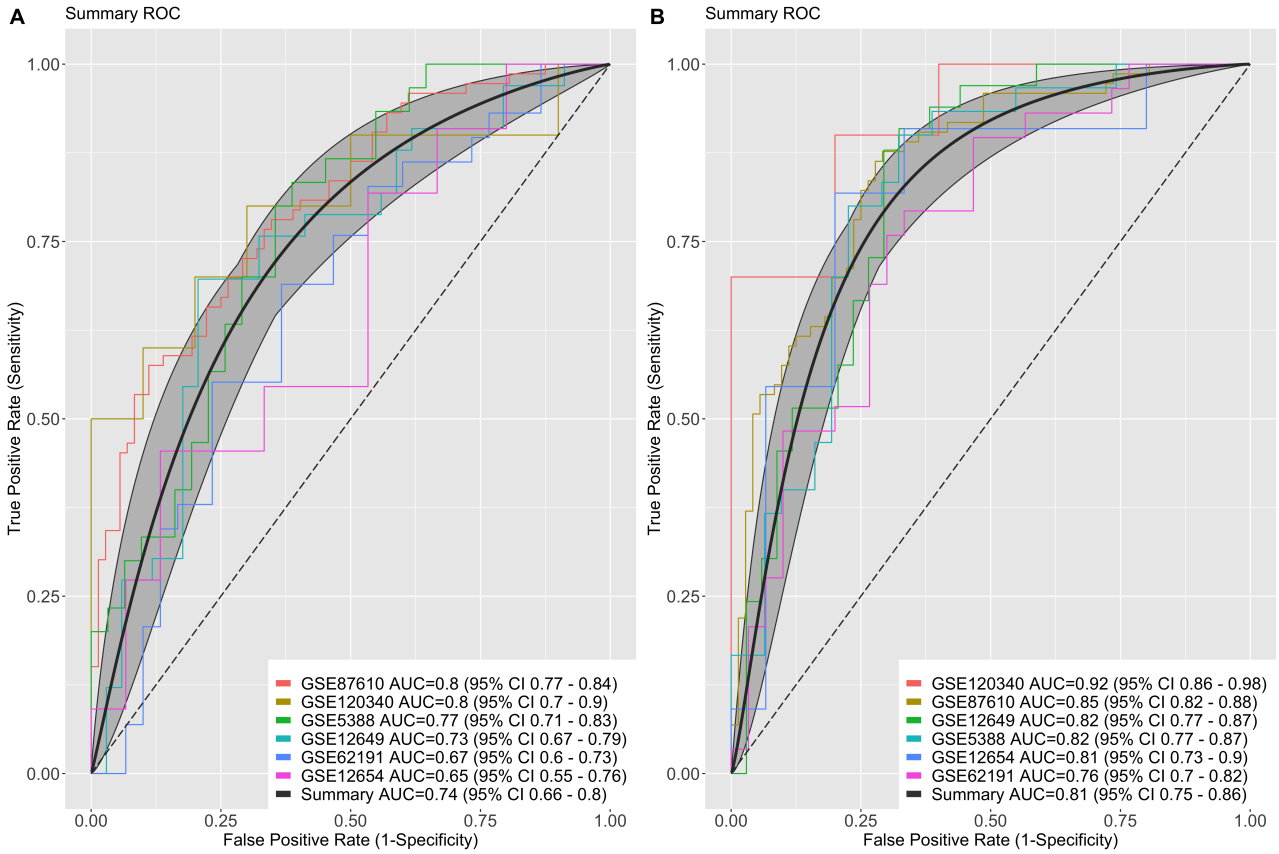

Figure 1: **Summary AUC in the training datasets.** with the initial gene signature (A) and the final gene signature resulting from forward search (B)

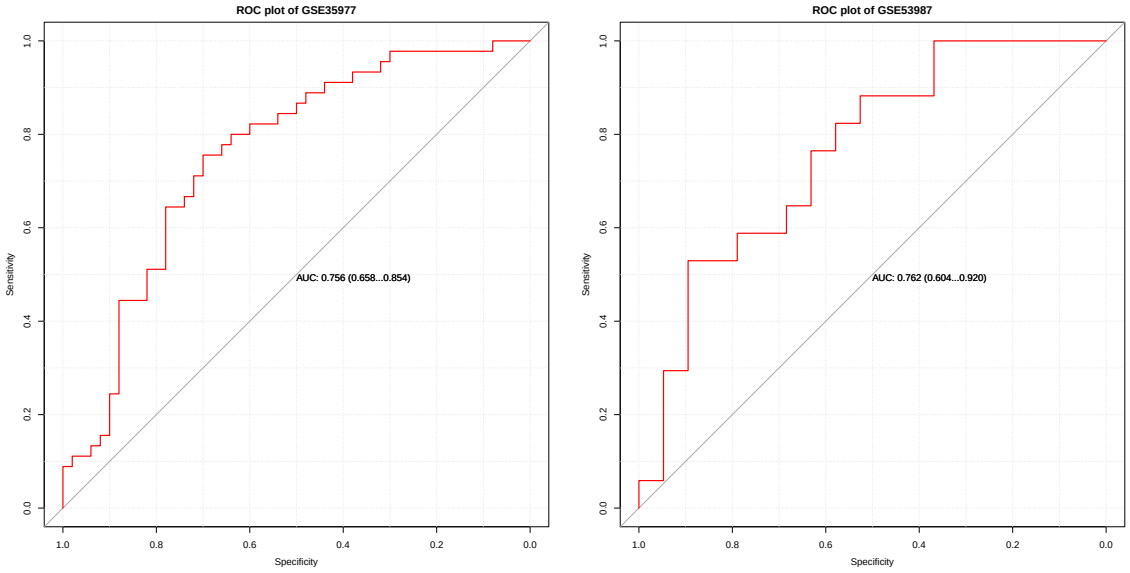

Figure 2: **ROC curves of the two validation data sets.** Strong classification performance (AUC = 0.756 and 0.76 respectively).

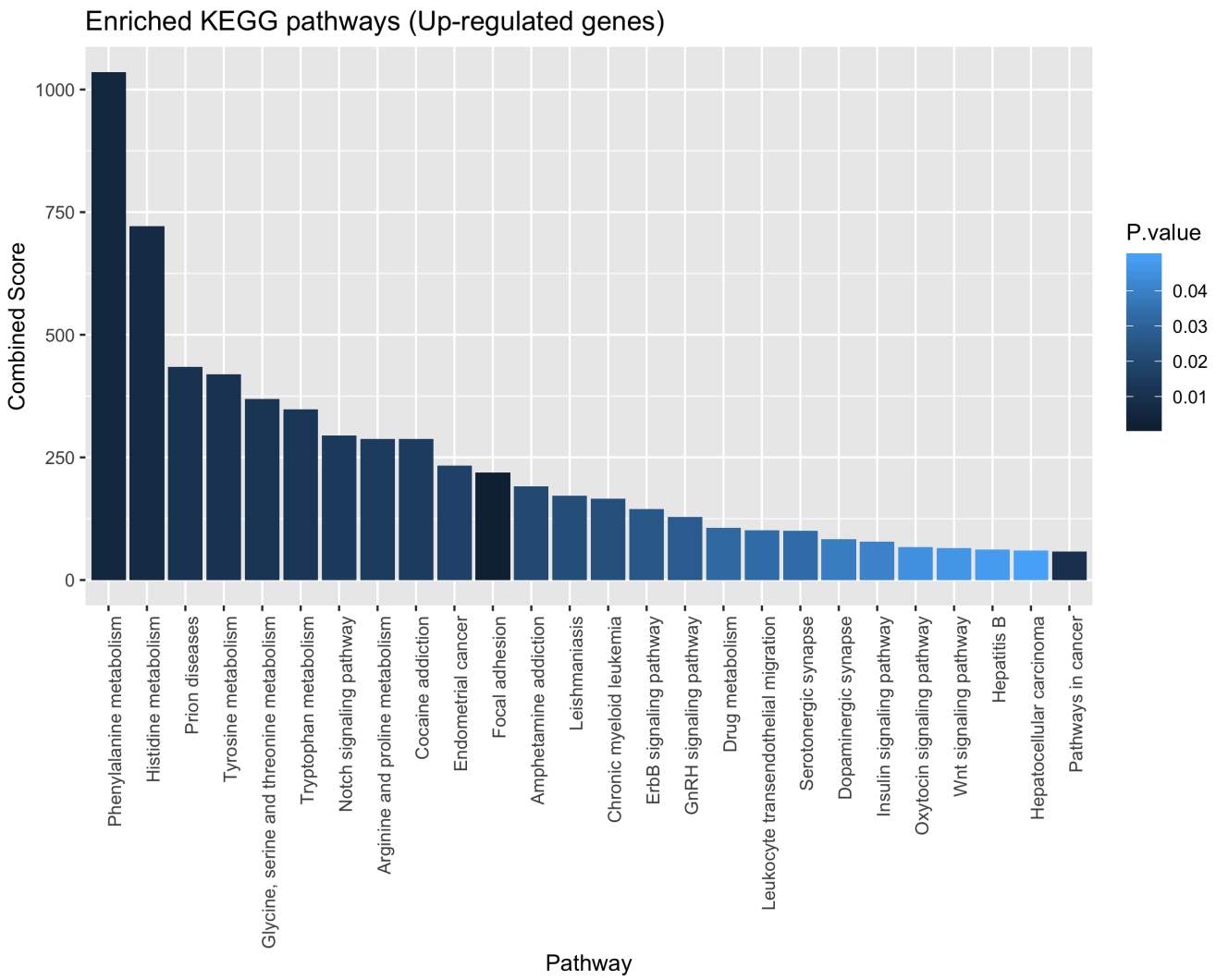

Figure 3: KEGG pathways of the 6 up-regulated genes.

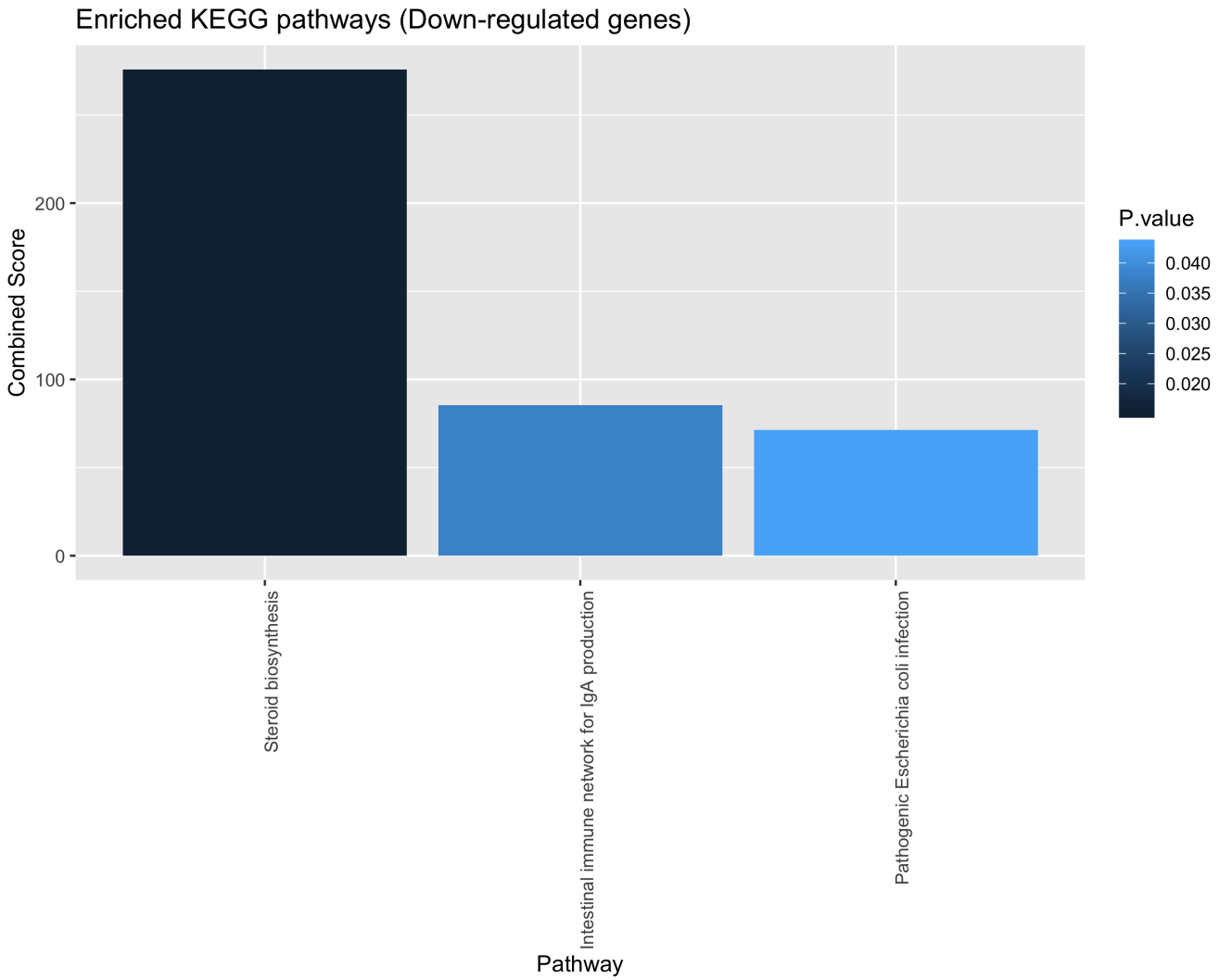

Figure 4: KEGG pathways of the 16 down-regulated genes.

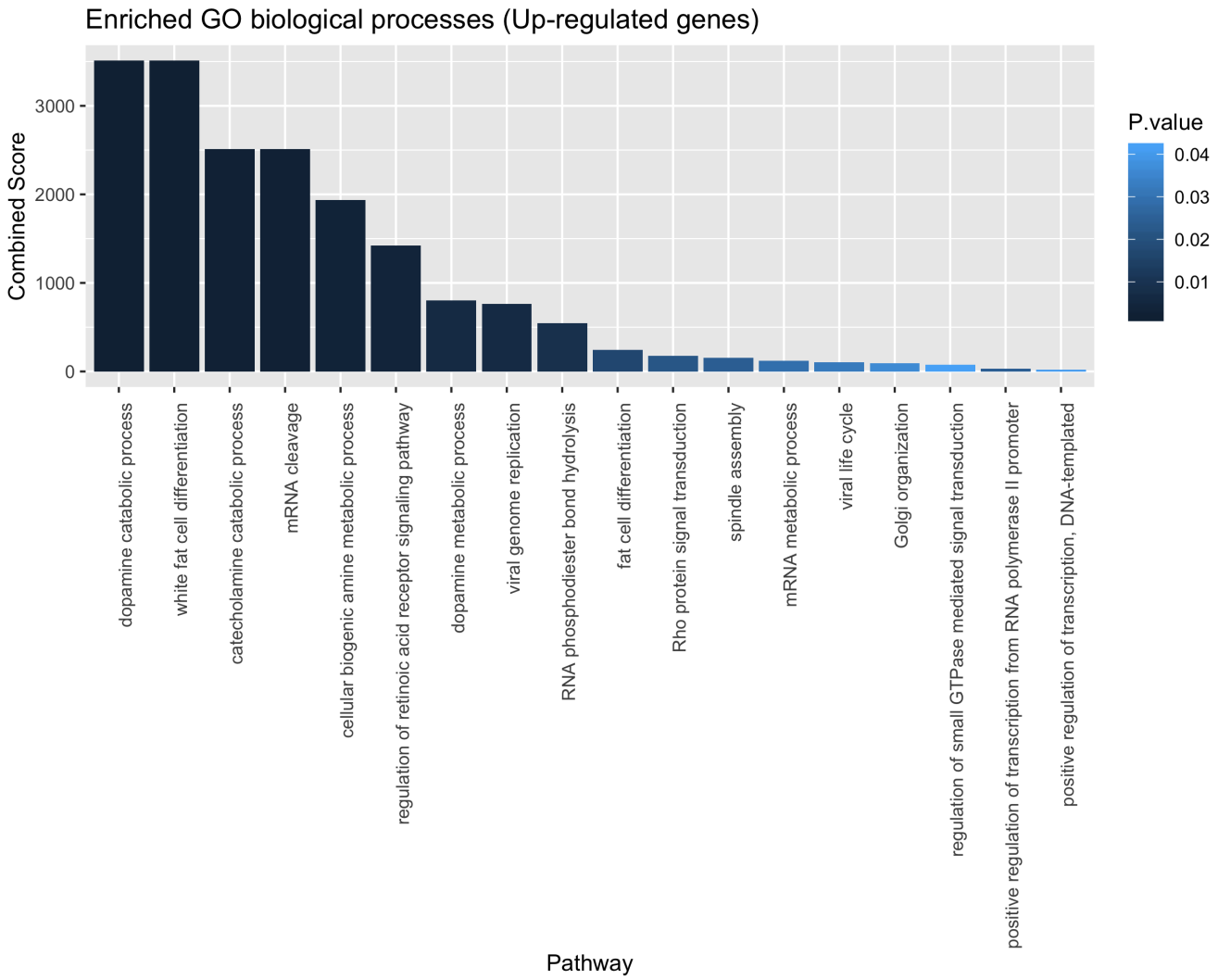

Figure 5: GO biological processes of the 6 up-regulated genes.

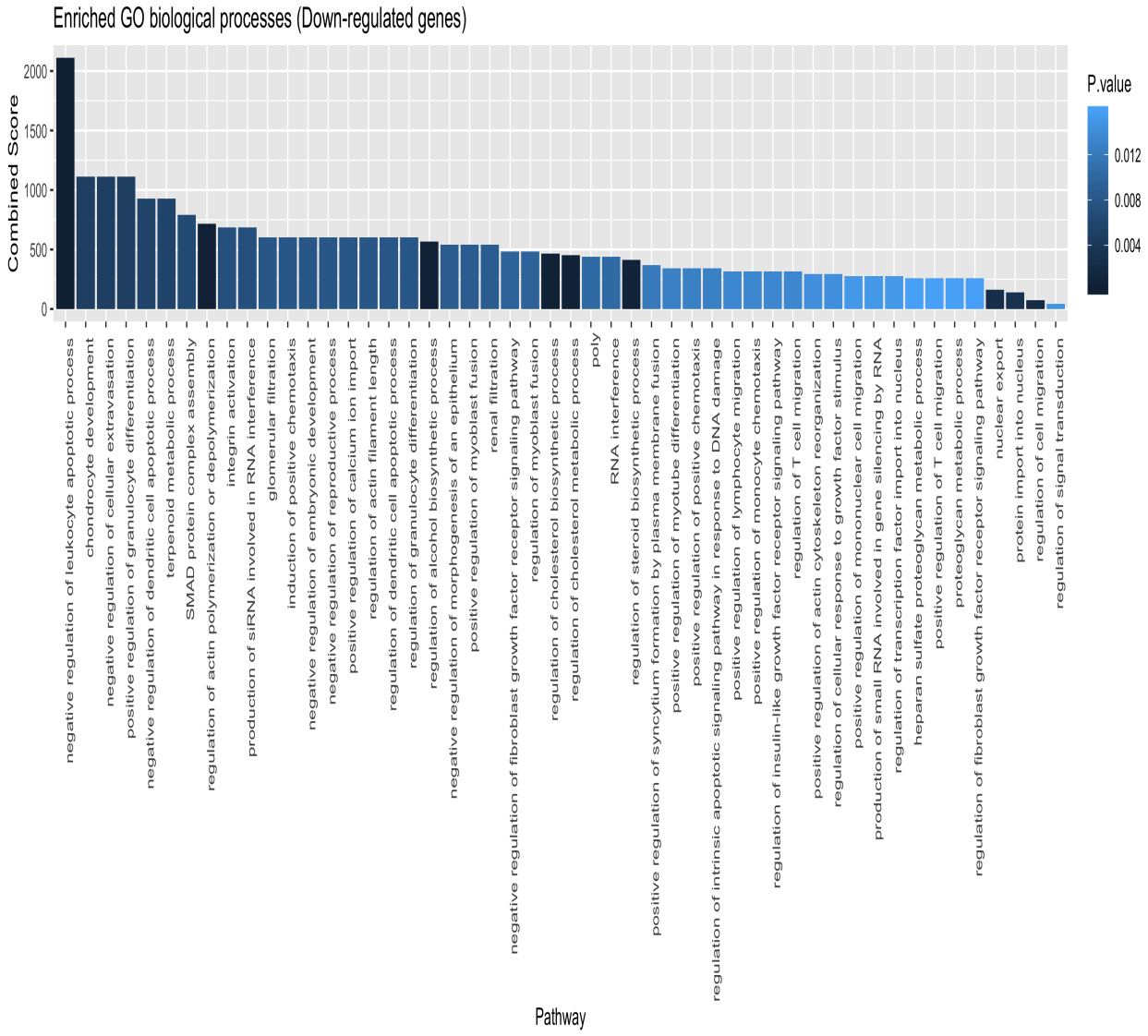

Figure 6: GO biological processes of the 16 down-regulated genes.

Tables

Table 1: Collected data sets showing the number of samples with the number of BD cases.

GEO accession	Type of experiment	Platform	Number of samples (BD cases)	Tissue source
		Discovery		
GSE120340 [45]	Expression profiling by array	GPL16311	20 (10)	Dorsolateral prefrontal cortex
GSE87610 [46]	Expression profiling by array	GPL13667	145 (73)	Dorsolateral prefrontal cortex
GSE12654 [47]	Expression profiling by array	GPL8300	26 (11)	Anterior prefrontal cortex
GSE12649 [47]	Expression profiling by array	GPL96	67 (33)	Dorsolateral prefrontal cortex
GSE62191 [48]	Expression profiling by array	GPL4133	59 (29)	Frontal cortex
GSE5388 [49]	Expression profiling by array	GPL96	61 (30)	Dorsolateral prefrontal cortex
		Validation		
GSE35977 [50, 51]	Expression profiling by array	GPL6244	95 (45)	Parietal cortex
GSE53987 [52]	Expression profiling by array	GPL570	36 (17)	Dorsolateral prefrontal cortex

Table 2: The BD signature genes. 13 up-regulated and 9 down-regulated genes

Gene Symbol	Effect size	Effect size SE	Effect size p-value	Effect size FDR
Up-regulated genes				
CTBP2	0.295	0.046	1.08e-10	7.56e-07
ARHGAP5	0.308	0.056	5.13e-08	9.82e-05
GOLGA8A	0.192	0.044	1.58e-05	0.004
RNASE4	0.321	0.084	0.0001	0.012
MAOA	0.339	0.102	0.0009	0.037
ELK1	0.327	0.099	0.0009	0.037
Down-regulated genes				
NFYC	-0.341	0.069	6.65e-07	0.0005
NXF1	-0.385	0.089	1.37e-05	0.0043
SST	-0.652	0.151	1.54e-05	0.0043
DDX18	-0.258	0.06	1.53e-05	0.0043
ETV5	-0.430	0.105	4.22e-05	0.0071
FDFT1	-0.234	0.057	4.42e-05	0.0073
ZFYVE9	-0.279	0.069	5.11e-05	0.0076
TSN	-0.266	0.066	4.99e-05	0.0076
IGFBP6	-0.419	0.104	5.9e-05	0.0082
CSE1L	-0.247	0.062	7.25e-05	0.0087
SULF1	-0.382	0.098	9.69e-05	0.0104
CXCL12	-0.271	0.070	0.0001	0.011
CD53	-0.442	0.115	0.0001	0.0117
RABIF	-0.441	0.117	0.0002	0.014
HCLS1	-0.267	0.078	0.0006	0.03
DNAJC7	-0.313	0.098	0.0014	0.047

Table 2: The BD signature genes. 13 up-regulated and 9 down-regulated genes

Gene Symbol	Effect size	Effect size SE	Effect size p-value	Effect size FDR

Additional Files

Supp. figure1— EffectSizes-UpRegulated

PDF file showing forest plots of the effect sizes of each of the 6 up-regulated genes.

Supp. figure2— EffectSizes-DownRegulated

PDF file showing forest plots of the effect sizes of each of the 16 down-regulated genes.

Supp. table1 — KEGG-UpRegulated

CSV table showing the significant KEGG pathways of the 6 up-regulated genes.

Supp. table2— GO-UpRegulated

CSV table showing the significant GO biological processes of the 6 up-regulated genes.

Supp. table3 — KEGG-DownRegulated

CSVtable showing the significant KEGG pathways of the 16 down-regulated genes.

Supp. table4 — GO-DownRegulated

CSVtable showing the significant GO biological processes of the 16 down-regulated genes.