Original Article

Integrated Bioinformatics Tools For The Evaluation Of Psychiatric Disorders

Usha Adiga[1], Shreyas Adiga[2*], Tirthal Rai[3]

[1,3] Professor, Nitte-DU, KS Hegde Medical Academy, Department of Biochemistry, Mangalore, India
[2*] First Year Student, BTech Computer Science Dual Degree Program, IIIT, Hyderabad, India

***Corresponding author:-** Mr Shreyas Adiga

*First Year Student, BTech Computer Science Dual Degree Program, IIIT, Hyderabad, India,
Email ID: shreyas.adiga@research.iiit.ac.in
DOI: 10.47750/pnr.2022.13.S05.392

Abstract

Background: However, it can be challenging to assess the complicated interactions between genetics and disease. With a high global prevalence rate, high morbidity, high resource use, and high disability rates, mental disorders are significant diseases. Major depression, bipolar disorder, and schizophrenia are widespread, complicated disorders with high heritabilities that fall under the category of polygenic genetic diseases. High-throughput sequencing, recombinant DNA technologies, and multidimensional nuclear magnetic resonance have all contributed to the rapid advancement of molecular biology. The sequencing of many biological genomes, including the human genome, has produced a wealth of nucleic acid knowledge. These genes can be categorised or grouped depending on different phenotypes, such as disease types or cell types, and their relationships can then be determined through experimental research.

Methodology: The ncbi geo database was used to gather published microarray data gene expression patterns of mental diseases such depression, bipolar disorder, and schizophrenia. The gene expression patterns were analysed and compared using integrated bioinformatics methods. R was used to process the data, Gene ontology (GO) and the KEGG database were used to analyse the function and pathway enrichment of differentially expressed genes (DEGs), and string database was used to study protein-protein interactions.

Results: 12,626 differentially expressed genes (DEGs) in all were investigated. However, there was no statistically significant difference in any of the genes between the patients and controls. Further research was done on the genes that were enriched by greater than 1.5 times. The genes with differential expression were compared using a volcano, mean difference plot, box plot, histogram, and Venn's diagram. In comparison to controls, none of the genes were significantly enriched, downregulated, or elevated in any of the three mental diseases.

Conclusion: The study demonstrates the efficacy of bioinformatics analytic approaches in identifying probable pathogenic genes for mental diseases like schizophrenia, bipolar disorder, and depression as well as their underlying mechanisms.

Key words: mental illnesses, gene, bioinformatics

BACKGROUND:

One of the most common mental diseases, major depressive disorder (MDD), is characterised by sporadic suicidal thoughts and actions in addition to persistent depression [1]. The fatal brain disease known as bipolar disorder affects more than a million individuals globally. Epidemiological research indicates that both genetic and environmental factors have a role in the development of this condition. A mental illness known as schizophrenia is characterised by recurrent or ongoing psychotic episodes. Hallucinations (which generally involve hearing voices), delusions, and disordered thinking are among the main symptoms. Schizophrenia often has a long history, despite the chaotic patterns and cognitive dysfunction it exhibits.

The defining characteristic of this condition is psychosis, which is characterised by positive symptoms like hallucinations and delusions and frequently accompanied by negative (deficit) symptoms like waning emotions, speech, and interest as well as disorganisation symptoms like disordered syntax and behaviour. Severe mood disorders, such as manic and profound depressive episodes, are usually present.

In recent years, there has been a growing body of research that suggests depression is more complex genetic disorder with multiple etiologies than a single brain region or single gene issue. Numerous brain areas associated to the "emotional network" were shown to have abnormal anatomy, function, and coordinated activity in patients with MDD.

Thus, another name for MDD is "disconnection syndrome." Brain activity abnormalities and poor mood regulation are regarded to be the main neuropathology underlying depression [2]. MDD is known to have abnormalities in the prefrontal cortex (PFC) and anterior cingulate cortex (ACC) in addition to the hippocampus [3]. In the ACC, pain, negative affect, and cognitive control are all regulated [4]. The PFC controls the hypothalamo-pituitary-adrenal (HPA) axis in both the

stress response and depression. Gene expression analysis in MDD has revealed many genes as well as details regarding the illness. However, due to heterogeneity and various sources of noise, aetiology is still difficult to identify[5-7].

Tissue-to-tissue network analysis, which is based on genome-wide association studies, provides a mechanism for the identification of disease-specific genes in response to tissue anomalies [8,9]. Converging evidence suggested that gene co-expression studies in the context of transcriptome investigations provide further information on gene changes [10].

Shared regulatory pathways are just one of the basic mechanisms by which co-expression genes can arise and play similar roles [5,8,10]. Signaling pathway disruption is thought to be a contributing factor in a number of illnesses [11]. Notably, by including multiple connections among a large number of genes, the analysis of gene co-expression networks provides a technique to deal with the complexity of biological alterations in polygenic disease [5,12].

The molecular pathogenesis of mental disease is still poorly understood at the molecular level.

The study's goal was to use microarray data to analyse biological information and find the genes that are differentially expressed (DEG) in distinct brain tissues. In order to comprehend the molecular mechanisms behind mental diseases, the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), pathway enrichment, and protein-protein interaction (PPI) network were examined.

MATERIALS AND METHODS

Microarray data

High throughput gene expression data, chips, and microarrays can be found in the Gene Expression Omnibus, a freely accessible and useful genomics database. Microarray data for schizophrenia, bipolar disorder, and depression were downloaded for this study's analysis from https://www.ncbi.nlm.nih.gov/geo/. For analysis, the GSE12654 [13] microarray data were retrieved. Postmortem prefrontal cortices were used in the study's oligonucleotide microarray investigation of bipolar disorder, major depression, schizophrenia, and control patients. The study investigated the molecular differences between bipolar disorder and other mental diseases by comparing the gene expression profiles of related but distinct mental disorders.

In the study, postmortem prefrontal cortex (Broadmann's Area 10) brain samples from healthy controls, patients with bipolar disorder, depression, and schizophrenia were analyzed. 15 subjects made comprised each group [13]. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [14] had been used to make diagnoses.

Bipolar disorder, major depression, schizophrenia, and control participants' oligonucleotide microarray analysis results using postmortem prefrontal cortices were compared to the gene expression profiles of healthy controls.

Data processing

To analyse the raw data and find differentially expressed genes, Bioconductor (http://bioconductor.org/biocLite.R) and R 4.0.1 statistical tools (https://www.r-project.org/) were used. Using the limma package, the data was batch calibrated and normalised. The main functionality of the Limma programme is to fit a gene linear model to gene expression data in order to assess the potential for differential expression [15]. This package offers particularly potent capabilities for reading, standardising, and analysing such data. The limma software was then used to filter the differentially expressed genes. The screening cutoff was a fold-change of 1.5 and a p-value of 0.05. The DEGs were visualised as a volcano map using the ggplot2 software, and the important DEGs were clustered using the pheatmap package.

Function and pathway enrichment analysis of DEGs

A community-based bioinformatics tool is the Gene Ontology (GO; http://www.geneontology.org). It offers details on the roles of genes and gene products and applies ontology to advance biological understanding [16]. The KEGG (https://www.kegg.jp) is a database for the qualitative interpretation of genetic sequences and other biological data, which includes systematic, genomic, and chemical data as well as a separate category of health data that is specifically related to humans[17]. GO/KEGG enrichment and the cluster Profiler software package were used to analyse the linked biological processes and signal pathways, with a p-value of less than 0.05 being deemed statistically significant.

Protein-protein interaction network construction and identification and validation of hub genes

The function of interacting proteins is mostly predicted by protein-protein interaction (PPI) network analysis. It is a practical tool that can be utilised to comprehend disease mechanisms and cell function [18,19]. By combining numerous known and anticipated protein-protein association data, the STRING database (http://string-db.org) aims to provide a critical evaluation of protein-protein interactions [20,21]. The STRING database was used to create a PPI network that was displayed using the Cytoscape programme. The microarray data available at GEO2R was used to confirm the statistical significance of these genes. Users can compare two or more sets of data in a GEO sequence using the interactive network programme GEO2R to find genes that are differentially expressed [22]. Statistics were deemed significant at $p<0.05$.

RESULTS AND DISCUSSION

12,626 differentially expressed genes (DEGs) in all were investigated. However, there was no statistically significant difference in any of the genes between the patients and controls. Further research was done on the genes that were enriched by greater than 1.5 times.

Following the filtering processes, a scatterplot based on the genes revealed that most of the genes' expression levels, from highly expressed to barely expressed, were identical in all three groups and control groups. The volcano plot created using Lima helps to visualise differentially expressed genes by displaying statistical significance (-log10 P value) vs the size of the change (log2 fold change). At a default adjusted p-value cutoff of 0.05, genes are significantly differentially expressed.However none of the genes were significant (fig 1a,1b,1c).

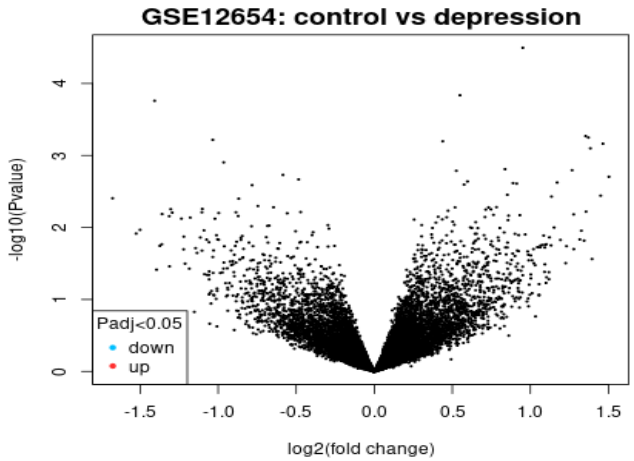

Fig 1a: Volcano plots of Depression

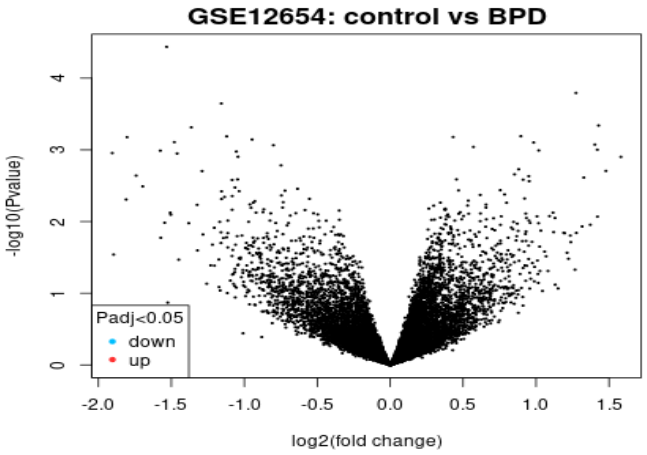

Fig 1b: Volcano plots of bipolar disorder

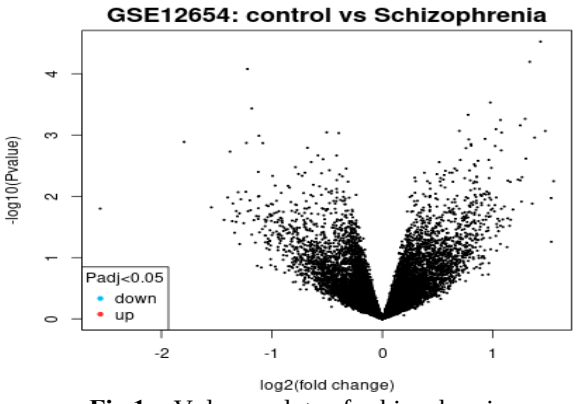

Fig 1c: Volcano plots of schizophrenia

To visualise differentially expressed genes, the mean difference (MD) plot created with limma compares log2 fold change to average log2 expression levels.

The plot shows the test results for a single comparison between a group of controls and a group with depression, bipolar

disorder, or schizophrenia. (fig 2a, 2b,2c).None of the genes were significantly differently expressed in mental illness compared to controls.

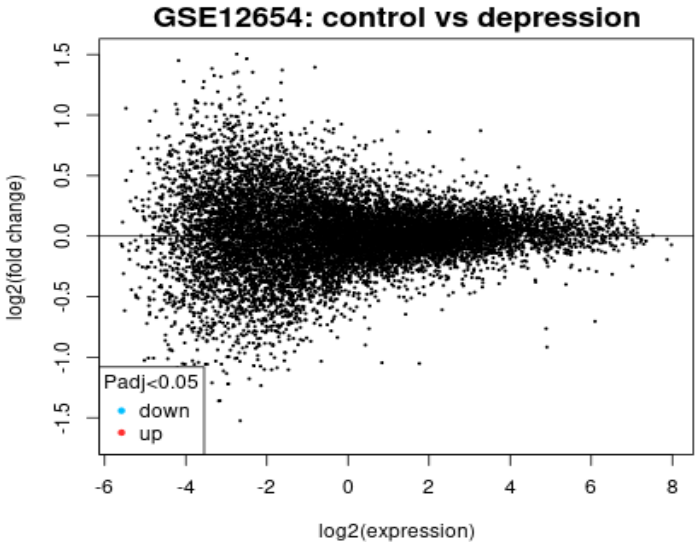

Fig 2a: Mean difference plots of depression

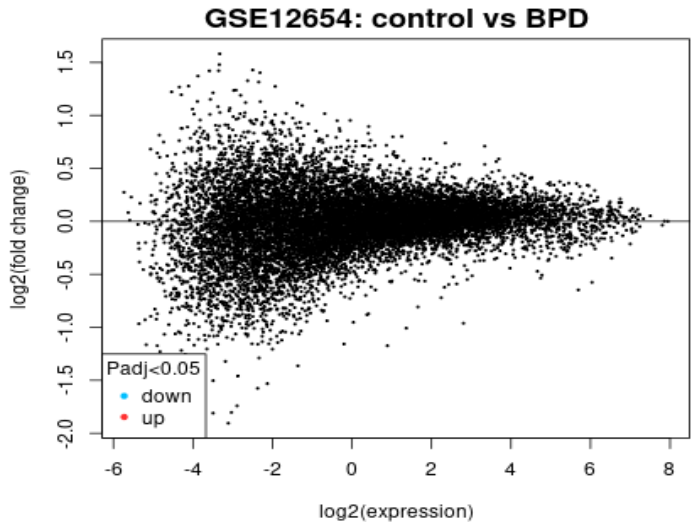

Fig 2b: Mean difference plots of BPD

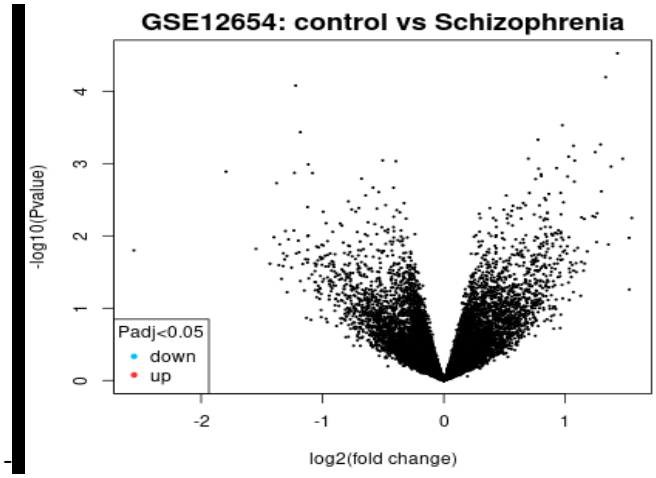

Fig 2c: Mean difference plots of schizophrenia

To see how the values of the chosen samples were distributed, a box plot created from boxplot R was employed. Samples are coloured based on groupings. Determining if particular samples are appropriate for differential expression analysis can be done by looking at the distribution. Typically, the plot displays the data after normalisation and log transformation. The results showed that the data were normalised and cross-comparable since they were median-centered. (fig 3).

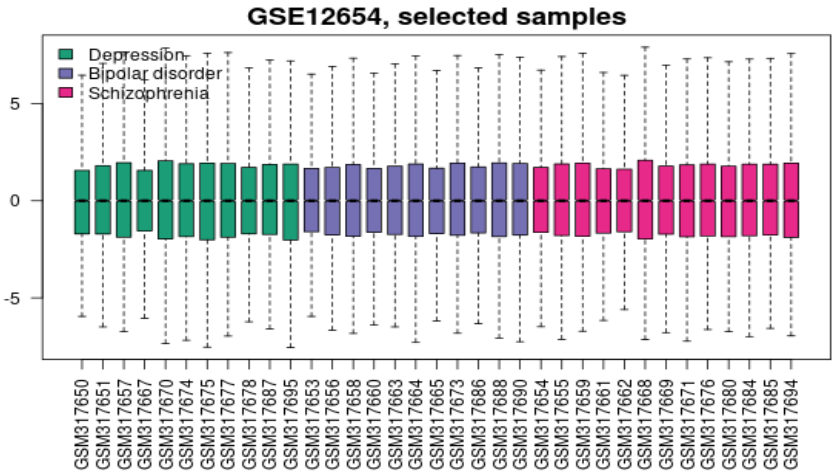

Fig 3: Boxplot for three psychiatric conditions

It was possible to see how the values of the chosen samples were distributed using an expression density plot created using R limma. Samples were coloured based on groupings. When testing for data normalisation prior to differential expression analysis, this graphic complements the boxplot. In comparison to controls, there was no appreciable difference in the expression densities of the various conditions (fig 4a,4b, 4c).

.

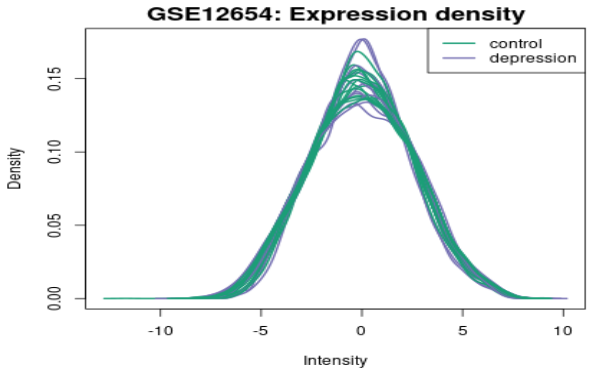

Fig 4a: Expression density plots of Depression

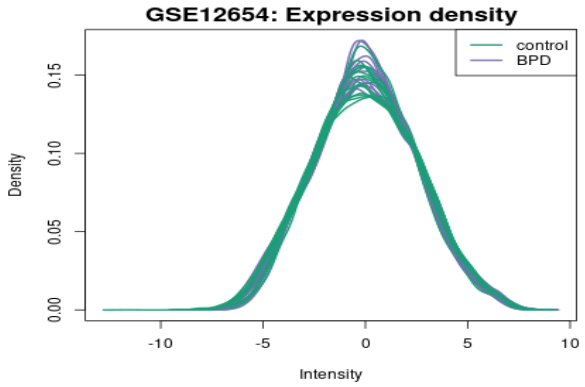

Fig 4b: Expression density plots of BPD

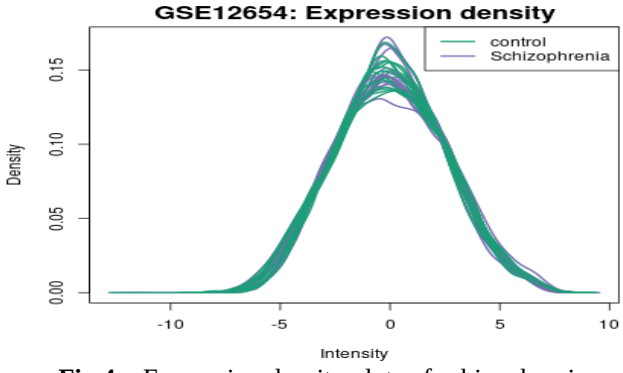

Fig 4c: Expression density plots of schizophrenia

Histogram constructed by hist is for viewing the distribution of the P-values in the analysis results. The P-value here is the same as in the top differentially expressed genes table and computed using all selected contrasts. P values were not significant for any of the conditions (fig 5)

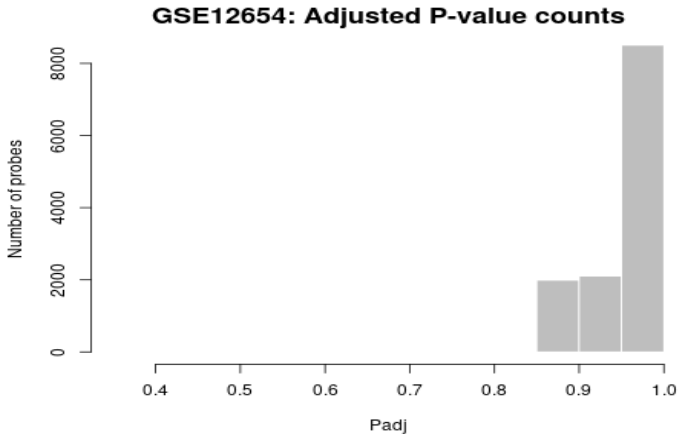

Fig 5: Histogram comparing depression Vs bipolar disorder Vs schizophrenia

Moderated t-statistic quantile quantile plot obtained by limma (qqt). Plots the quantiles of a data sample against the theoretical quantiles of a Student's t distribution. This plot helps to assess the quality of the limma test results (fig 6).

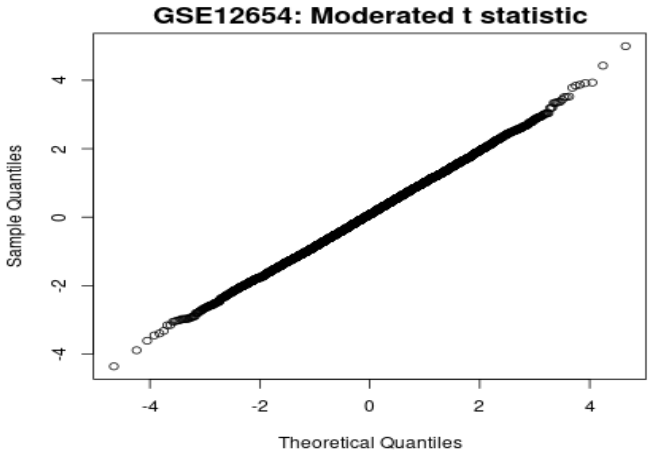

Fig 6: q-q plot

After fitting a linear model, a mean variance plot created using R limma (plotSA, vooma) is used to examine the mean-variance connection of the expression data. It can assist in demonstrating whether the data are very variable. This graphic can be used to determine whether it is advised to use the precision weights option to take mean-variance trend into consideration. When a strong mean-variance trend is present, precision weights increase the accuracy of test results.

A gene is represented by each point. When analysing differential gene expression, the red line is an approximation of the mean-variance trend. The blue line approximates constant variance. (fig 7).

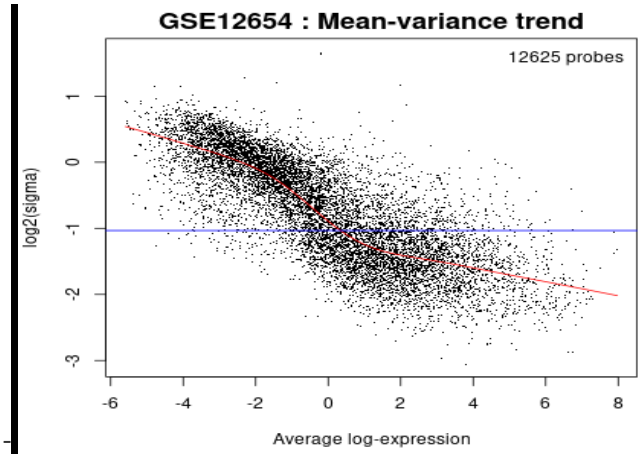

Fig 7: Mean variance trend plot

Uniform Manifold Approximation and Projection (UMAP) derived by using umap, is a dimension reduction technique useful for visualizing how samples are related to each other. The number of nearest neighbors used in the calculation is indicated in the plot(fig 8a,8b,8c).

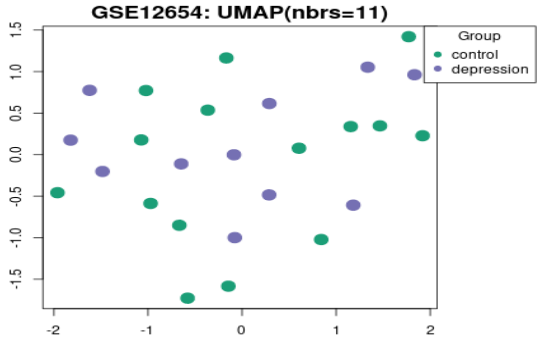

Fig 8a: UMAP plots of depression

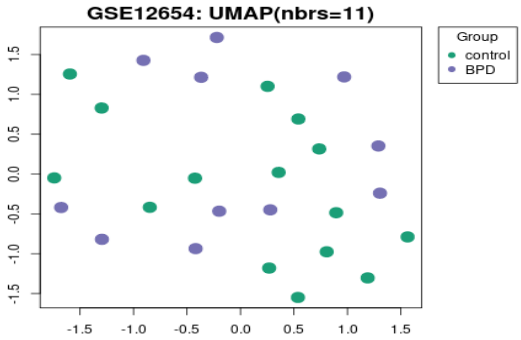

Fig 8b: UMAP plots of BPD

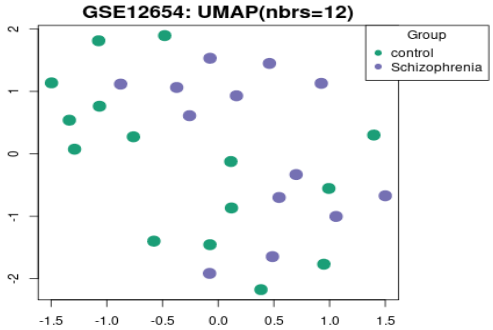

Fig 8c: UMAP plots of schizophrenia

Venn diagram generated using limma (vennDiagram) is used to explore and download the overlap in significant genes between different groups. There was no gene that was common for all the three conditions (depression, bipolar disorder and schizophrenia) suggesting that genes involved in the pathogenesis of the three mental illnesses are different (fig 9).

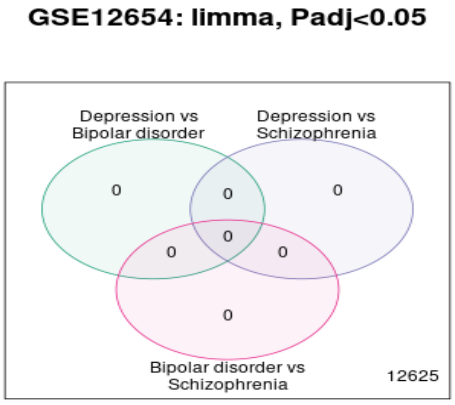

Fig 9: Venn diagram comparing the genes for three conditions

ENRICHMENT PATHWAY ANALYSIS

Differentially expressed genes were defined based on the fold change (1.5- fold or greater) and the P-value (P<0.05) of the mean expression values. The number of genes that met this criteria and the pathways involving these genes is as in fig 10.

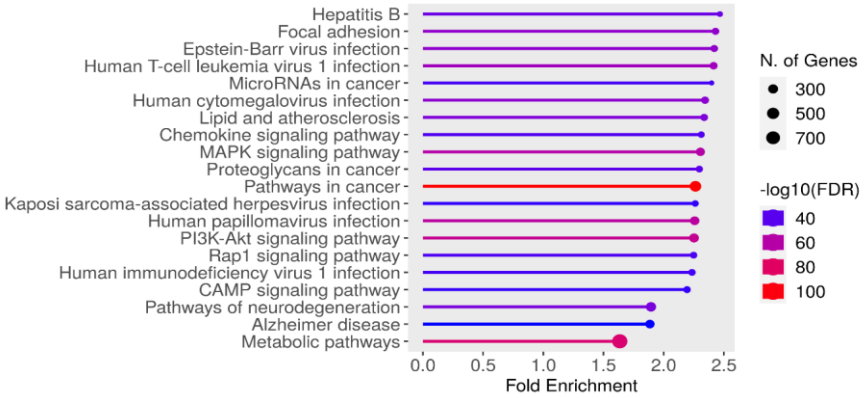

Fig 10: Enrichment pathways

Despite the fact that bipolar disorder and major depression have hereditary roots and similar symptoms, their gene expression patterns were very different. The majority of the genes exhibiting altered expression in two or more illnesses turned out to code for the transcription- or translation-related proteins. This data would suggest that the many mental diseases share a common pathophysiological basis to some extent.

STRING ANALYSIS

On string analysis, 34 nodes, 8 edges were found. Average node degree was 0.471, average local cluster coefficient was 0.191 and p value for PPI enrichment was 0.326(fig 11).

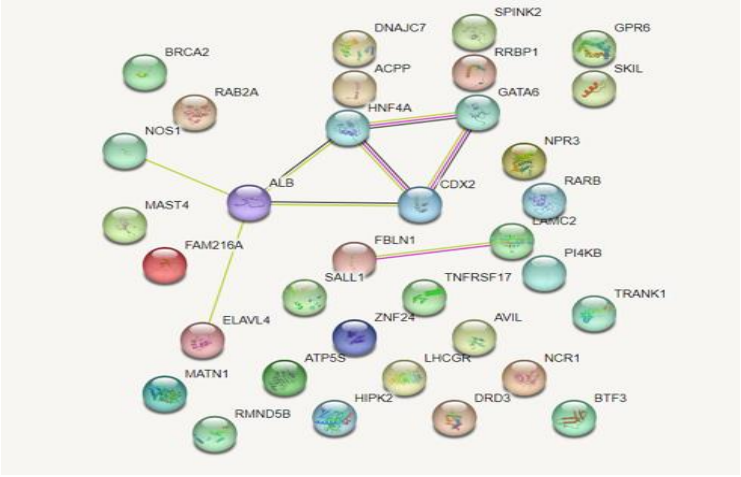

Fig 11: Protein-protein interaction by String analysis

Strength of gene association was 0.91 on the basis of GO enrichment, with a false discovery rate (FDR) of 0.0373. The MAPK signalling pathway, PI3K-Akt route, RAP 1 signalling pathway, CAMP signalling pathway, and several metabolic pathways were among the pathways in which the genes were enriched. Biological processes included the intrinsic apoptotic signalling pathway in response to DNA damage, signal transduction by p53, and FDR of 0.018-0.0259 with a strength of 1.58-2.0. Transcription regulator complex was revealed by cellular component enrichment with a strength of 0.91 and FDR of 0.0373.

One high-throughput technology that makes it possible to examine the genome's overall expression levels and spot changes in gene expression using a method without making any assumptions is the use of microarrays. Over the past ten years, these technologies have been used in several studies to identify abnormalities in gene expression linked to mental diseases. In addition to the hypothesis-driven approach, which is primarily based on the analysis of candidate genes expression levels, transcriptomics studies can enable the identification of new biomarkers associated with various mental illnesses that can aid in the development of novel intervention strategies and the introduction of personalised medicine.
A study by Ryan et al compared the gene expression patterns of bipolar disorder patients' postmortem brain tissue to those of controls and found no statistically significant differences in gene expression in the dorsolateral prefrontal cortex [23]. According to Clifton et al., the expression of schizophrenia in the prefrontal cortex is positively connected with early midfetal development and early infancy, and negatively correlated with expression in late childhood and adolescence [24]. According to a study by Kumar et al., the expression of forty genes had dramatically changed throughout the brain's

regions. Genes involved in biological processes and events were uncovered through analyses [25].

There aren't many research that have examined the microarray data deposited by diverse studies using bioinformatics techniques. This is an attempt to investigate the variations in gene expression patterns between normal controls and mental diseases like depression, bipolar disorder, and schizophrenia.

CONCLUSION

The work demonstrates the efficiency of bioinformatics analysis methods in identifying probable mental illness-related pathogenic genes and their underlying mechanisms. Additionally, it may be possible to successfully anticipate potential biomarkers of mental disorders and the pathways in which those genes function, both of which could offer attractive targets for the treatment of mental illnesses to some extent.

AUTHOR CONTRIBUTIONS

Usha Adiga: Conceptualization, manuscript preparation
Shreyas Adiga : Bioinformatics softwares, R software
Tirthal Rai: Review and proof reading of the manuscript

CONFLICTS OF INTEREST:

None

FUNDING:

Nil

REFERENCES

1. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M. Major depressive disorder. Nat Rev Dis Primers 2016;2:16065.
2. Song Z, Zhang M, Huang P. Aberrant emotion networks in early major depressive disorder patients: an eigenvector centrality mapping study. Transl Psychiatry 2016;6:e819.
3. Rodriguez-Cano E, Sarro S, Monte GC, Maristany T, Salvador R, McKenna PJ, et al. Evidence for structural and functional abnormality in the subgenual anterior cingulate cortex in major depressive disorder. Psychol Med 2014;44(15):3263–73.
4. Tolomeo S, Christmas D, Jentzsch I, Johnston B, Sprengelmeyer R, Matthews K, et al. A causal role for the anterior mid-cingulate cortex in negative affect and cognitive control. Brain 2016;139(Pt 6):1844–54.
5. Gaiteri C, Guilloux JP, Lewis DA, Sibille E. Altered gene synchrony suggests a combined hormone-mediated dysregulated state in major depression. PLoS ONE 2010;5(4):e9970.
6. Li JZ, Bunney BG, Meng F, Hagenauer MH, Walsh DM, Vawter MP, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc Natl Acad Sci USA 2013;110(24):9950–5.
7. Wang Q, Dwivedi Y. Transcriptional profiling of mitochondria associated genes in prefrontal cortex of subjects with major depressive disorder. World J Biol Psychiatry 2016:1–29.
8. Dobrin R, Zhu J, Molony C, Argman C, Parrish ML, Carlson S, et al. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. Genome Biol 2009;10(5):1–13.
9. Geng R, Li Z, Yu S, Yuan C, Hong W, Wang Z, Wang Q, Yi Z, Fang Y. Weighted gene co-expression network analysis identifies specific modules and hub genes related to subsyndromal symptomatic depression. World J Biol Psychiatry 2020; 21(2):102-110.
10. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. Genome Res 2004;14(6):1085–94.
11. Hughey JJ, Lee TK, Covert MW. Computational modeling of mammalian signaling networks. Wiley Interdiscip Rev Syst Biol Med 2010;2(2):194–209.
12. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond Modules & Hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. Genes Brain Behav 2014;13(1):13–24.
13. Iwamoto, K., Kakiuchi, C., Bundo, M. et al. Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. Mol Psychiatry 2004; **9**: 406–416.
14. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th edn. American Psychiatric Association: Washington, DC, 1994.
15. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature 2013;496(7446):504–7.
16. Guzman MG, Harris E. Dengue. The Lancet 2015; 385(9966):453–65.
17. Shahen M, Guo Z, Shar AH, Ebaid R, Tao Q, Zhang W, et al. Dengue virus causes changes of MicroRNA-genes regulatory network revealing potential targets for antiviral drugs. BMC Syst Biol 2018;12(1):2.
18. Tsai CY, Lee IK, Lee CH, Yang KD, Liu JW. Comparisons of dengue illness classified based on the 1997 and 2009 World Health Organization dengue classification schemes. J Microbiol Immunol Infect 2013;46(4):271–81.
19. Castro MC, Wilson ME, Bloom DE. Disease and economic burdens of dengue. Lancet Infect Dis 2017;17(3):e70–e8.
20. Pei H, Zuo L, Ma J, Cui L, Yu F, Lin Y. Transcriptome profiling reveals differential expression of interferon family induced by dengue virus 2 in human endothelial cells on tissue culture plastic and polyacrylamide hydrogel. J Med Virol 2016;88(7):1137–51.
21. Biswal S, Reynales H, Saez-Llorens X, Lopez P, Borja-Tabora C, Kosalaraksa P, et al. Efficacy of a Tetravalent Dengue Vaccine in Healthy Children and Adolescents. N Engl J Med 2019;381(21):2009–19.
22. Moodie Z, Juraska M, Huang Y, Zhuang Y, Fong Y, Carpp LN, et al. Neutralizing Antibody Correlates Analysis of Tetravalent Dengue Vaccine Efficacy Trials in Asia and Latin America. J Infect Dis 2018;217(5):742–53.
23. Ryan, M., Lockstone, H., Huffaker, S. et al. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. Mol Psychiatry 2006;11: 965–978.
24. Clifton NE, Hannon E, Harwood JC, Di Florio A, Thomas KL, Holmans PA, Walters JTR, O'Donovan MC, Owen MJ, Pocklington AJ, Hall J. Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. Transl Psychiatry. 2019 ;9(1):74.
25. Kumar A, Pareek V, Singh HN, Faiq MA, Narayan RK, Raza K, Kumar P. Altered Expression of a Unique Set of Genes Reveals Complex Etiology of Schizophrenia. Front Psychiatry 2019 ;10:906.