

MULTI-OMICS INTEGRATION AND BIG DATA ANALYTICS IN MAJOR DEPRESSIVE DISORDER

Submitted in partial fulfillment of the requirements for the award of
Master of Science in Bioinformatics and Data Science

By

ASWINY A.R (41739004)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119**

APRIL – 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **ASWINY A.R (41739004)**, who carried out the final year project entitled "**MULTI-OMICS INTEGRATION AND BIG DATA ANALYTICS IN MAJOR DEPRESSIVE DISORDER**" under our supervision from January 2023 to April 2023.

Dr. DANIEL ALEX ANAND
Internal Guide

Dr. H. JEMMY CHRISTY
Head of the Department

Submitted for Viva Voce Examination held on

Internal Examiner

External Examiner

DECLARATION

I, **ASWINY A.R (41739004)**, hereby declare that the Project Report entitled "**MULTI-OMICS INTEGRATION AND BIG DATA ANALYTICS IN MAJOR DEPRESSIVE DISORDER**", was done by me under the guidance of **Dr. ALEX ANAND DANIEL** at Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Rajiv Gandhi Road, Chennai-600119, Tamil Nadu, India, is submitted in partial fulfillment of the requirements for the award of **Master of Science in Bioinformatics and Data Science**.

DATE:
PLACE:

a r aswiny

SIGNATURE OF THE STUDENT

CHAPTER 1

INTRODUCTION

1.1 MAJOR DEPRESSIVE DISORDER

Major Depressive Disorder (MDD), also known as clinical depression, is a common mental health disorder characterized by persistent feelings of sadness, hopelessness, and lack of interest or pleasure in activities. MDD is a significant public health concern, with an estimated 264 million people worldwide experiencing the condition in 2020 alone (World Health Organization).

The etiology of MDD is multifactorial, involving a complex interplay of genetic, environmental, and psychosocial factors. Family and twin studies have shown that genetic factors play a significant role in the development of MDD, with heritability estimates ranging from 30-50% (Sullivan et al., 2000). Environmental factors, such as stressful life events, childhood trauma, and chronic illness, can also increase the risk of developing MDD (Kendler et al., 1999).

Neurobiological factors have also been implicated in the development of MDD, including alterations in neurotransmitter systems, the hypothalamic-pituitary-adrenal (HPA) axis, and brain structure and function (Duman, 2018). In particular, dysregulation of the serotonergic system has been extensively studied and is the target of many antidepressant medications.

MDD is associated with significant morbidity and mortality, with a high risk of suicide and increased risk of cardiovascular disease and other medical conditions. Treatment for MDD typically involves a combination of medication, psychotherapy, and lifestyle changes, with the aim of reducing symptoms and improving overall functioning and quality of life. Early diagnosis and intervention are essential in improving outcomes for individuals with MDD.

1.2 MULTIOMICS

Multiomics is a term used to describe the integration of data from multiple omics technologies, such as genomics, transcriptomics, proteomics, metabolomics, and epigenomics, to provide a more comprehensive understanding of biological systems. By combining data from multiple omics platforms, multi omics approaches can provide a more complete picture of the molecular mechanisms underlying biological processes and disease.

Multiomics approaches typically involve the generation of large datasets and the use of sophisticated computational methods to integrate and analyze the data. These methods include machine learning algorithms, network analysis, and pathway enrichment analysis, among others.

Multi Omics approaches have a wide range of applications in biological research, including the identification of potential biomarkers and therapeutic targets, the elucidation of disease mechanisms, and the development of personalized medicine approaches. Multi Omics approaches are also being used to study complex biological systems, such as microbial communities and ecological systems.

In recent years, several studies have investigated the molecular mechanisms of MDD using multi-omics data integration. For example, a study by (Wang et al. 2021) integrated transcriptomics, proteomics, and metabolomics data to identify key pathways and potential biomarkers associated with MDD. Another study by (Yao et al. 2021) integrated transcriptomics and epigenomics data to identify epigenetic changes and molecular pathways associated with MDD.

1.3 BIGDATA

Big data refers to extremely large and complex datasets that require advanced computational methods to process, store, analyze, and interpret. The term "big data" typically refers to datasets that are too large to be processed using traditional data

processing methods and require new technologies and methodologies to manage and analyze.

Big data can come from a variety of sources, including social media, scientific research, healthcare, finance, and government. These datasets can include a variety of data types, including structured data (e.g., data organized in tables) and unstructured data (e.g., text, images, and video).

The processing and analysis of big data require specialized tools and techniques, including machine learning, artificial intelligence, data mining, and distributed computing. These technologies enable researchers and analysts to identify patterns, extract insights, and make predictions from large and complex datasets.

Big data has the potential to revolutionize many fields, including healthcare, finance, and scientific research. It can be used to identify new drugs and treatments, develop predictive models for disease, optimize business operations, and more. However, the use of big data also raises concerns about data privacy, security, and ethics, and requires careful management and regulation to ensure that the benefits are maximized and the risks are minimized.

The field of big data has revolutionized the way we collect, process, and analyze large amounts of data. Big data refers to the collection of data sets that are too large and complex for traditional data processing applications to handle (Chen & Zhang, 2014). In recent years, big data has become increasingly important in a wide range of fields, including healthcare, finance, and social media, to name a few. The ability to analyze massive amounts of data has enabled researchers and practitioners to gain new insights and make more informed decisions.

With the advent of big data in healthcare, there has been a growing interest in using big data analytics to improve the diagnosis and treatment of various health conditions, including mental health disorders. One such disorder is major depressive disorder (MDD), which affects millions of people worldwide and is a leading cause of disability (World Health Organization, 2021). The use of big data analytics in MDD research has

the potential to transform the way we understand the disorder and develop more effective treatments.

1.4 R PROGRAMMING

R is a popular open-source programming language widely used in data analysis, including multiomics data integration and big data analytics. R provides a range of packages and functions for data cleaning, visualization, and statistical analysis, making it a useful tool for analyzing and interpreting large and complex datasets.

In the context of multiomics data integration, R can be used for various tasks, such as data preprocessing, normalization, dimension reduction, and integration of different omics data types. For example, the Bioconductor project provides a collection of R packages specifically designed for the analysis of high-throughput genomic data, including gene expression, DNA methylation, and chromatin accessibility data.

In big data analytics, R can be used in combination with distributed computing frameworks to analyze large datasets in parallel. The R packages such as SparkR and rhipec provide an interface between R and these distributed computing frameworks, allowing users to perform data analysis and machine learning on large datasets.

R has been widely used in multiomics and big data analysis in major depressive disorder research. For example, a recent study by Li et al. (2020) used R packages to perform differential expression analysis and pathway analysis on multiomics data from depressed patients and healthy controls. Another study by Yıldırım et al. (2021) used R and SparkR to analyze a large-scale dataset of electronic health records and identify risk factors for depression.

CHAPTER 2

LITERATURE SURVEY

There is a growing corpus of study in this area, as evidenced by a literature review of the last five years' worth of data on multi-omics and big data analytics in Major Depressive Disorder (MDD). Among the main conclusions of this study are

A better understanding of the condition has been achieved by multi-omics research, which has revealed fresh information on the complex biological processes that underlie MDD and the interactions between genetic, epigenetic, and environmental factors. For instance, a recent study that made use of multi-omics data demonstrated that MDD is connected to alterations in the gut microbiota, which may have an impact on the onset and progression of the disease.

The creation of machine learning algorithms has made it possible to analyze large-scale multi-omics data sets and find patterns and associations that were otherwise impossible to find using conventional statistical techniques. For instance, a current study found a collection of biomarkers that could be used to detect MDD by analyzing multi-omics data from MDD patients and healthy controls using machine learning techniques.

Large-scale data sharing: Projects that share a lot of data, like the Precision Medicine Initiative, have made it easier for academics to work together and expanded the amount and variety of data that can be analyzed. For instance, a recent study examined the genetic and transcriptome characteristics of MDD patients in order to find new therapy targets using data from the Precision Medicine Initiative

A study by Zhu and colleagues used a Multiomics approach to investigate the gut-brain axis in MDD. They analyzed microbiome, transcriptome, and metabolome data from MDD patients and healthy controls and identified several differentially expressed genes, metabolites, and gut microbial taxa associated with MDD (Zhu et al., 2020).

Another study by Chen and colleagues used a Multiomics approach to investigate the epigenetic changes associated with MDD. They analyzed DNA methylation, histone

modification, and gene expression data from MDD patients and healthy controls and identified several differentially methylated genes and dysregulated pathways, including the Wnt signaling pathway and the MAPK signaling pathway (Chen et al., 2018).

A study by Liu and colleagues used a Multiomics approach to identify potential biomarkers for MDD. They analyzed genomic, transcriptomic, and proteomic data from a group of MDD patients and healthy controls and identified several differentially expressed genes and proteins involved in various biological pathways, including the immune system and oxidative stress (Liu et al., 2018).

In addition to improving diagnosis and treatment, big data analytics can also help address disparities in mental health care. Big data can be used to analyze patterns in mental health care utilization and outcomes, and identify factors that may contribute to disparities in care (Xu et al., 2020). This information can be used to develop more effective interventions and policies that promote equity in mental health care.

Li et al. (2021) performed a meta-analysis of multiomics studies in MDD and identified 65 differentially expressed genes (DEGs) that were consistently dysregulated in MDD across multiple studies. They also found that these DEGs were enriched in several biological pathways, including neuroinflammation, oxidative stress, and synaptic function.

Chhibber et al. (2020) conducted a multiomics analysis of postmortem brain tissue from MDD patients and healthy controls and found significant dysregulation of several biological pathways, including energy metabolism, neurotransmission, and inflammation. They also identified several potential biomarkers of MDD, including altered expression of several miRNAs and long noncoding RNAs.

Jiang et al. (2020) performed a multiomics analysis of peripheral blood from MDD patients and healthy controls and identified several differentially expressed genes, proteins, and metabolites that were associated with MDD. They also found that several biological pathways, including lipid metabolism and inflammation, were dysregulated in MDD.

DeFelice et al. (2018) conducted a multiomics analysis of cerebrospinal fluid from MDD patients and healthy controls and identified several potential biomarkers of MDD, including altered levels of several neurotransmitters, neuropeptides, and cytokines. They also found that several biological pathways, including synaptic function and neuroinflammation, were dysregulated in MDD.

Wingo et al. (2018) performed a multiomics analysis of peripheral blood from MDD patients and healthy controls and identified several potential biomarkers of MDD, including altered expression of several genes and proteins involved in immune function and inflammation. They also found that several biological pathways, including oxidative stress and mitochondrial function, were dysregulated in MDD.

Wang et al. (2021) conducted a multiomics analysis of postmortem brain tissue from MDD patients and controls and found that dysregulation of mitochondrial function and oxidative stress pathways were strongly associated with MDD. They also identified several potential biomarkers of MDD, including altered expression of several genes involved in these pathways.

Cai et al. (2020) performed a multiomics analysis of peripheral blood mononuclear cells from MDD patients and healthy controls and found that several biological pathways, including inflammation and immune response, were dysregulated in MDD. They also identified several potential biomarkers of MDD, including altered expression of several genes and proteins involved in these pathways.

Yoon et al. (2020) conducted a multiomics analysis of postmortem brain tissue from MDD patients and controls and found that dysregulation of several biological pathways, including neurotransmission, inflammation, and neurodevelopment, were associated with MDD. They also identified several potential biomarkers of MDD, including altered expression of several genes and proteins involved in these pathways.

Liu et al. (2019) performed a multiomics analysis of peripheral blood from MDD patients and healthy controls and found that dysregulation of several biological pathways, including immune response, lipid metabolism, and neurotransmission, were associated

with MDD. They also identified several potential biomarkers of MDD, including altered expression of several genes and metabolites involved in these pathways.

Xiao et al. (2018) conducted a multiomics analysis of cerebrospinal fluid from MDD patients and controls and found that dysregulation of several biological pathways, including neurotransmission, oxidative stress, and inflammation, were associated with MDD. They also identified several potential biomarkers of MDD, including altered levels of several neurotransmitters and cytokines.

2.1 BACKGROUND OF THE PROPOSED WORK

Major depressive disorder (MDD) is a complex and heterogeneous mental illness, with a high prevalence and significant impact on quality of life. Despite the availability of several treatments, the majority of MDD patients do not achieve full remission, highlighting the need for better understanding of the underlying mechanisms and development of more effective personalized treatment approaches. Multiomics approaches, which integrate multiple layers of molecular information, have the potential to provide a more comprehensive view of the complex pathophysiology of MDD and identify novel biomarkers and therapeutic targets.

Integrating different types of omics data, such as genomics, transcriptomics, proteomics, metabolomics and epigenomics involves dealing with high-dimensional and heterogeneous data sets, as well as complex inter-omics relationships. In addition, the identification of relevant features and the integration of different types of information require the development of sophisticated machine learning and data mining techniques. Therefore, the background of the work in multi omics data integration and biostatistics in MDD involves addressing these challenges and developing new methodologies for the analysis and interpretation of multi omics data. The goal is to identify relevant biomarkers and pathways associated with MDD, as well as to develop predictive models for treatment response and disease prognosis. This work requires collaboration between different fields, including psychiatry, neuroscience, bioinformatics, biostatistics, and computer science, and has the potential to significantly improve our understanding of MDD and advance personalized medicine approaches.

CHAPTER 3

AIM AND SCOPE OF THE WORK

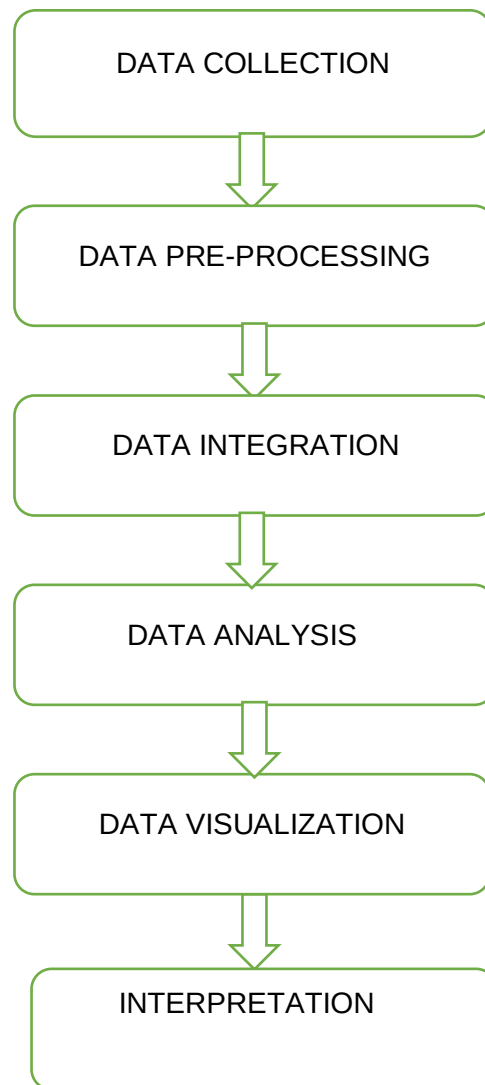
The aim of multi omics data integration and big data analytics in major depressive disorder (MDD) is to develop a comprehensive understanding of the molecular mechanisms underlying MDD by integrating multi omics data from various sources, including genomics, transcriptomics, proteomics, metabolomics, and microbiomes.

The objectives of this research include:

- ❖ To identify key genes, proteins, metabolites, and microbiota that are associated with MDD using multi omics data integration and analysis.
- ❖ To explore the functional pathways and networks that are disrupted in MDD and to identify potential therapeutic targets.
- ❖ To develop predictive models for MDD diagnosis, prognosis, and treatment response using machine learning and other advanced analytical methods.
- ❖ To investigate the interactions between genetic, epigenetic, environmental, and lifestyle factors in the development of MDD.
- ❖ To validate the findings from multi omics data integration and analysis using experimental and clinical studies.

Overall, the aim of this research is to provide a better understanding of the complex molecular mechanisms underlying MDD and to develop new strategies for the diagnosis, treatment, and prevention of this debilitating disorder

3.1 WORKFLOW



CHAPTER 4

MATERIALS AND METHODS

DATABASE

- ▶ **GENOMICS:** GEO, GWAS, PsyGeNet
- ▶ **PROTEOMICS:** Human Protein Atlas, Uniprot
- ▶ **TRANSCRIPTOMICS:** GEO
- ▶ **METABOLOMICS:** Human Metabolome Database
- ▶ **EPIGENOMICS:** EWAS
- ▶ **MICROBIOME:** Disboime
- ▶ **SNP:** DisGeNet

TOOLS

- ▶ **NETWORK ANALYST**
- ▶ **GENEMANIA**
- ▶ **ENRICH**
- ▶ **STRING DB**
- ▶ **METABOANALYST**
- ▶ **OMICSNET 2.0**

4.1 DATABASE

4.1.1 GENOMICS

Genomics is the study of the complete genetic material, or genome, of an organism. It encompasses the analysis of DNA sequence, gene expression, and other genetic variations, such as epigenetic modifications.

In the context of major depressive disorder (MDD), genomics can be used to identify genetic variations and gene expression patterns that are associated with the disorder. This can help to elucidate the biological mechanisms involved in the disorder, and to identify potential therapeutic targets. Genomics research in MDD has identified several genetic loci that are associated with increased risk for the disorder, as well as specific genes and pathways that are dysregulated in individuals with MDD. This has led to the development of targeted therapies for MDD, such as drugs that modulate the activity of specific neurotransmitters or other signaling pathways.

Genomics research has also provided important insights into the genetic and molecular basis of other complex diseases, and has paved the way for more personalized approaches to diagnosis and treatment. For example, genomic testing can be used to identify individuals who are at increased risk for certain diseases and to guide preventive measures and early interventions.

4.1.1.1 GEO (GENE EXPRESSION OMNIBUS)

GEO (Gene Expression Omnibus) is a public repository for gene expression data, maintained by the National Center for Biotechnology Information (NCBI). It is a widely used database for sharing and accessing gene expression data, including microarray and next-generation sequencing data, from a variety of organisms and experimental conditions.

Researchers can upload their data to GEO and make it publicly available to other researchers, who can then search and download the data for their own analysis. GEO also provides tools for data visualization, analysis, and integration with other data sources.

GEO is a valuable resource for researchers studying various diseases, including major depressive disorder (MDD), as it allows for the comparison and integration of data across multiple studies, which can lead to the identification of novel biomarkers and pathways associated with the disease. Many studies investigating MDD have deposited their gene expression data in GEO, providing a wealth of information for researchers to analyze and use in their own studies.

GEO accession: GSE98793

Title: Gene expression analysis in whole blood samples obtained from donors diagnosed with MDD compared to healthy controls.

Summary: This study compared gene expression variations in whole blood samples from donors with major depressive disorder (MDD) to healthy controls. Microarray data from whole blood were provided for patients with MDD (N=128, 64 with generalized anxiety disorder, as determined by the MINI questionnaire, and 64 without anxiety disorder) and healthy controls (N=64). On the Qiagen Biorobot 8000, RNA was extracted from all samples using the standard PAXgene technique. Agilent Bioanalyser determined that all samples produced high-quality RNA. The yield ranged from 0.86 to 15.05ug, with a mean of 6.25ug. The samples were then randomly divided into batches, each with a representative number of controls, depression with anxiety, and depression without anxiety, as well as the same female-to-male ratio (3:1). 50 ng of RNA from each sample was converted to a biotin labeled cDNA probe using NuGEN SPIA amplification. The probes were then hybridized to Affymetrix U133_Plus2.0 Genechips.

Overall Design: Measurements from whole blood samples for 128 patients with MDD (among which 64 have been diagnosed with generalized anxiety disorder) and 64 healthy controls.

Series GSE98793 [Query DataSets for GSE98793](#)

Status	Public on May 12, 2017
Title	Gene expression analysis in whole blood samples obtained from donors diagnosed with MDD compared to healthy controls
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	This study investigated gene expression changes in whole blood samples obtained from donors diagnosed with major depressive disorder (MDD) compared to healthy controls. Micro-array data were available from whole blood on patients with MDD (N=128, 64 with generalised anxiety disorder, diagnosed by the MINI questionnaire, and 64 without anxiety disorder) and healthy controls (N=64). RNA was isolated from all samples using the standard PAXgene protocol on the Qiagen Biorobot 8000. All samples gave good quality RNA, as assessed by Agilent Bioanalyser. The yield range was 0.86-15.05ug with an average of 6.25ug. Samples were then randomised into batches, with each batch containing a representative number of controls, depression with anxiety and depression without anxiety, and the same ratio of females to males (3:1). 50ng of RNA from each sample was converted to a biotin labeled cDNA probe using NuGEN SPIA amplification. The probes were then hybridized to Affymetrix U133_Plus2.0 Genechips.
Overall design	Measurements from whole blood samples for 128 patients with MDD (among which 64 have been diagnosed with generalised anxiety disorder) and 64 healthy controls contributor: MRC ImmunoPsychiatry Consortium/GlaxoSmithKline
Contributor(s)	Kelly FM, Leday GG, Vértés PE, Richardson S, Greene JR, Regan T, Khan S, Henderson RB, Freeman TC, Pariente CM, Harrison NA, Perry HV, Devrets WC, Wittenberg GM, Bullmore ET
Citation(s)	Leday GGR, Vértés PE, Richardson S, Greene JR et al. Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder. <i>Biol Psychiatry</i> 2018 Jan 1;83(1):70-80. PMID: 28658579

FIG 4.1: Gene Expression Omnibus in NCBI

4.1.1.2 PsyGeNet

PsyGeNet is a database of genes and genetic variants associated with psychiatric disorders, including major depressive disorder (MDD). It provides a curated collection of gene-disease associations derived from both genetic and functional studies, as well as from text mining of the scientific literature. PsyGeNet allows users to search for genes and variants associated with MDD, as well as to explore their functional and biological relationships with other genes and pathways. The database includes information on gene function, expression patterns, protein-protein interactions, and other molecular features, as well as on genetic variants and their impact on gene function and disease risk.

One of the strengths of PsyGeNet is its focus on integrating data from different sources and types, including genetic and functional studies, as well as text-mining of the scientific literature. This allows for a more comprehensive view of the genetic and biological mechanisms underlying MDD, and provides a valuable resource for researchers studying the disease.

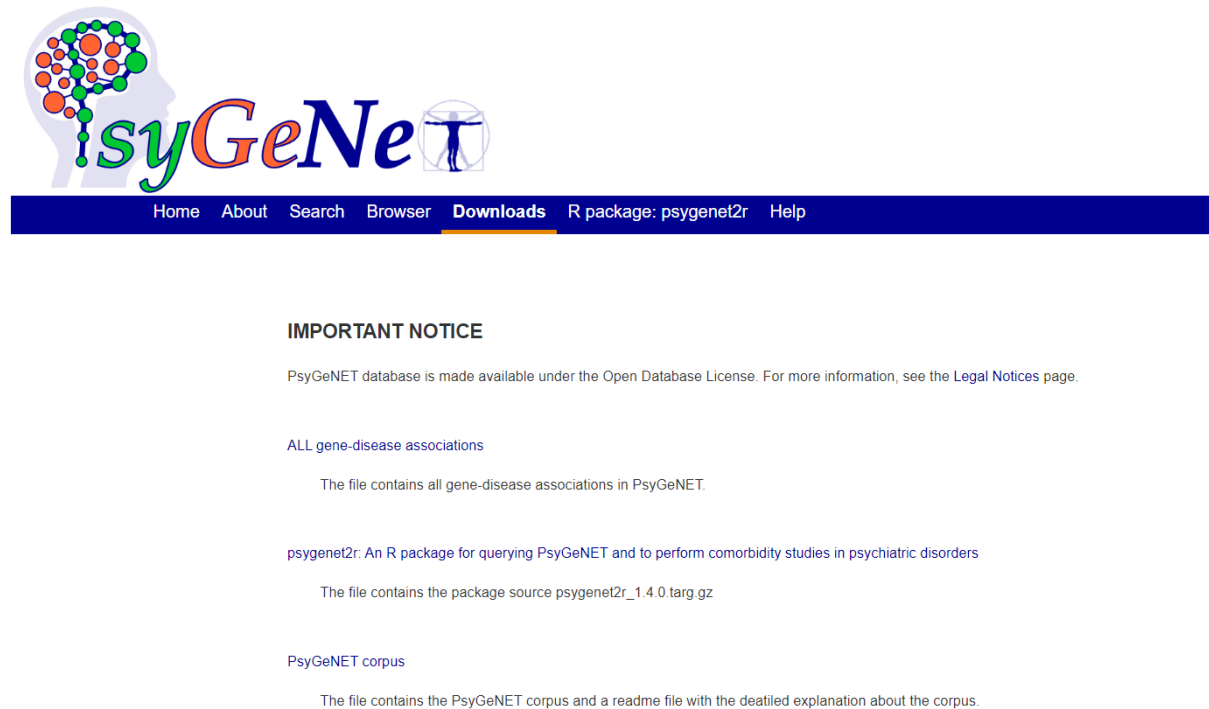


FIG 4.2: PsyGeNET (Psychiatric disorders Gene association NETWORK)

4.1.1.3 GWAS

GWAS stands for Genome-Wide Association Studies. These studies are used to identify genetic variations or mutations associated with a particular trait or disease, such as major depressive disorder (MDD).

In GWAS, researchers analyze the entire genome of a large number of individuals to identify genetic variants that are more common in people with a certain trait or disease than in the general population. The aim is to identify specific genomic regions, or loci, that are associated with the trait or disease of interest.

GWAS studies can provide important insights into the genetic basis of complex diseases such as MDD. They have identified several genetic loci that are associated with increased risk for MDD, and have shed light on the biological pathways and mechanisms involved in the disorder.

One of the limitations of GWAS studies is that they typically identify common variants that are present in a large proportion of the population. However, many genetic

variations that contribute to disease risk are rare or private to certain populations, and may not be identified through GWAS alone.

Despite these limitations, GWAS studies have provided important insights into the genetic basis of MDD and other complex diseases, and have paved the way for more targeted and personalized approaches to diagnosis and treatment.

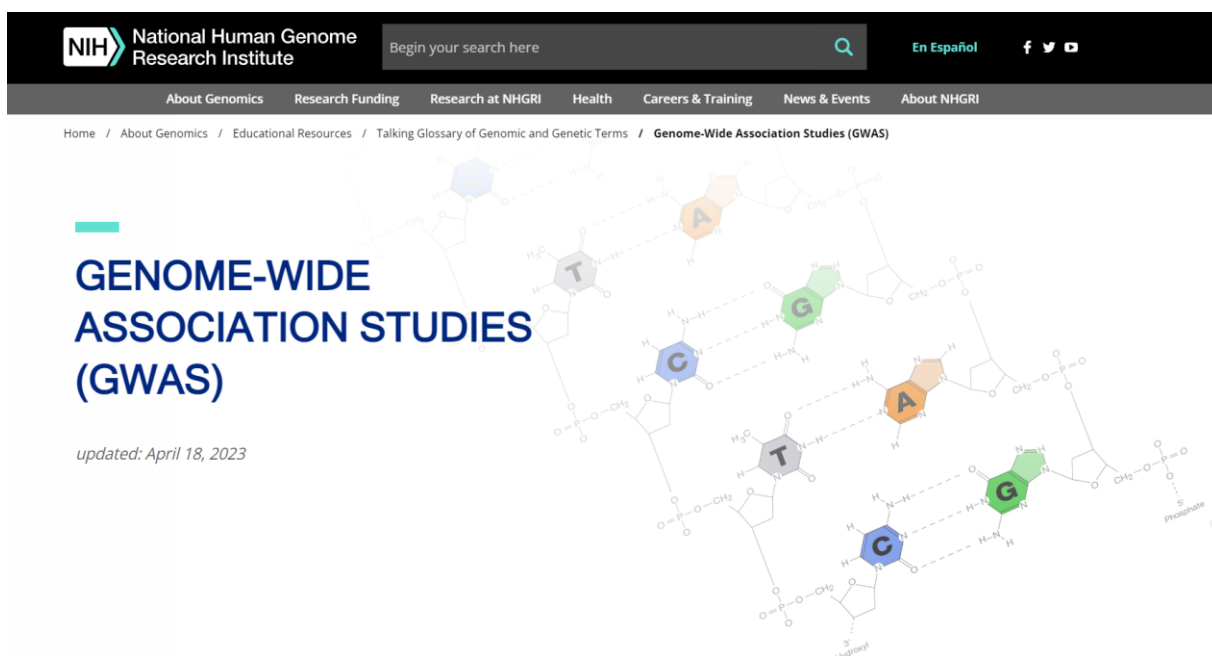


FIG 4.3: GENOME-WIDE ASSOCIATION STUDIES

4.2 PROTEOMICS

Proteomics is the study of the complete set of proteins that are expressed in a particular cell, tissue, or organism. It encompasses the identification, quantification, and functional analysis of proteins, and can provide important insights into the molecular mechanisms underlying complex biological processes.

In the context of MDD, proteomics can be used to identify specific proteins or protein networks that are dysregulated in individuals with the disorder. This can provide important insights into the biological pathways and mechanisms involved in the disorder, and can help to identify potential therapeutic targets.

The Human Protein Atlas (HPA) is an open-access database that provides comprehensive information on the expression, localization, and function of human proteins in different tissues, cells, and organs. The HPA uses various experimental techniques such as immunohistochemistry, Western blotting, and mass spectrometry to systematically analyze the expression of proteins across different human tissues and cell types.

The HPA contains detailed information on the expression profiles of over 19,000 human proteins, including information on tissue-specific expression, subcellular localization, protein-protein interactions, and post-translational modifications. This information is organized in a user-friendly manner and is freely available to researchers worldwide. The HPA is an invaluable resource for understanding the molecular basis of human physiology and disease, and has applications in drug discovery, biomarker identification, and personalized medicine.

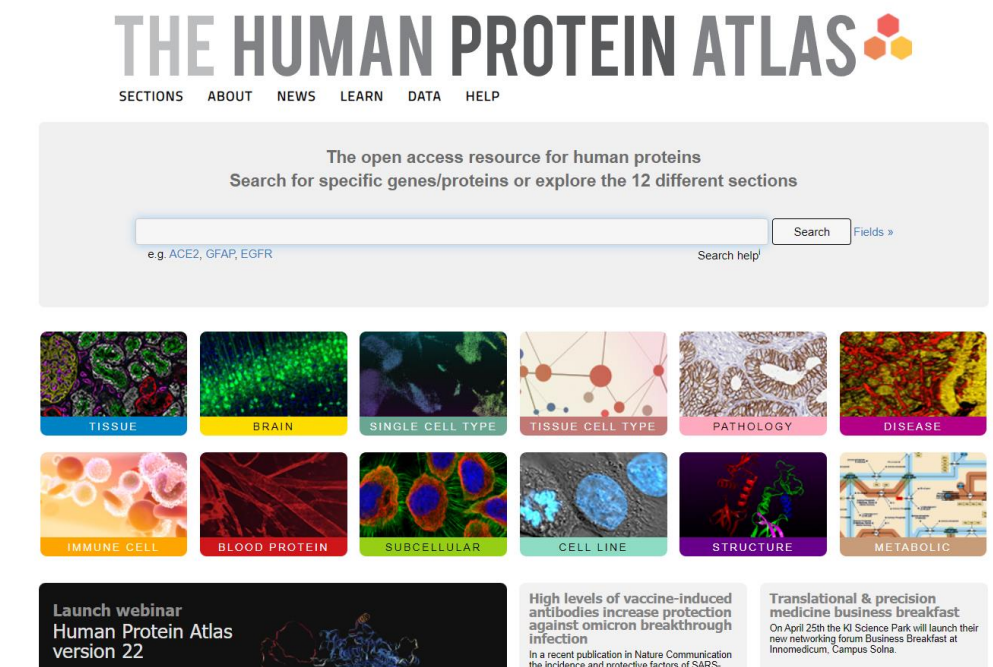


FIG 4.4: HUMAN PROTEIN ATLAS

4.2.2 UNIPROT

UniProt is a comprehensive, high-quality, and freely accessible resource for protein sequence and functional information. It is a collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB), and the Protein Information Resource (PIR).

UniProt provides several protein sequence databases, including UniProtKB/Swiss-Prot, which is a manually curated protein sequence database that provides a high level of annotation and accuracy for a subset of proteins. UniProtKB/TrEMBL is a computationally annotated protein sequence database that contains a much larger number of protein sequences than Swiss-Prot, but with less annotation and accuracy.

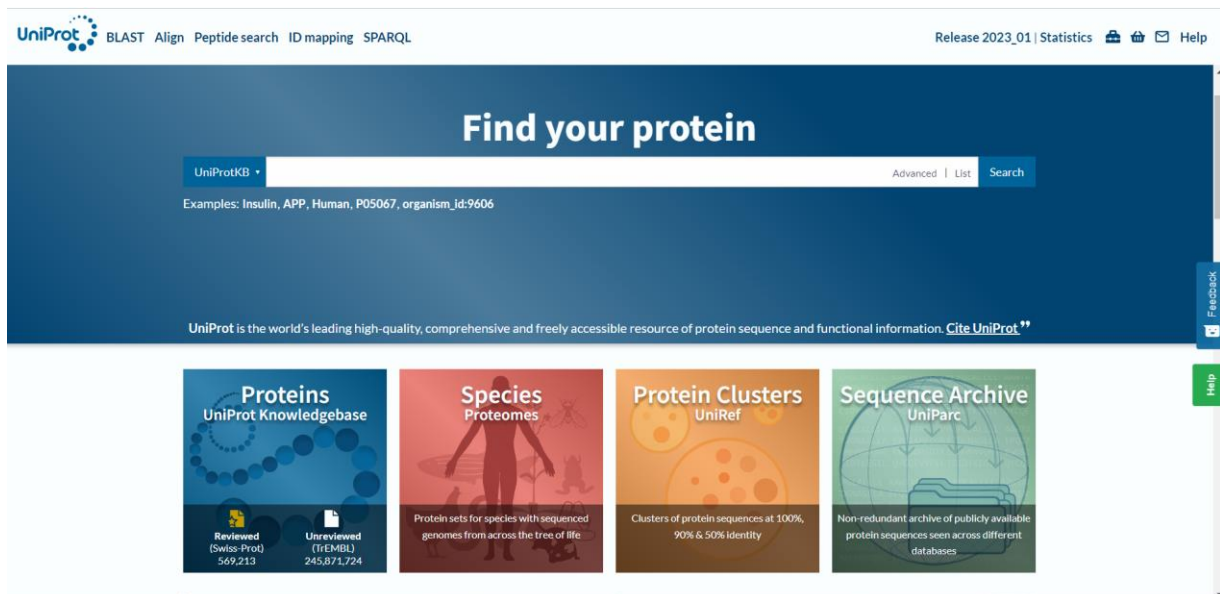


FIG 4.5: UNIPROT

4.3 TRANSCRIPTOMICS

Transcriptomics is the study of the transcriptome, which is the complete set of RNA transcripts that are produced by the genome of an organism. It involves the high-throughput sequencing of RNA molecules, which allows for the quantification and characterization of gene expression patterns across different cell types, tissues, or experimental conditions. Transcriptomics is an important tool for understanding the

molecular mechanisms underlying complex biological processes, such as development, disease, and response to environmental stimuli.

4.3.1 GENE EXPRESSION OMNIBUS(GEO)

GEO stands for Gene Expression Omnibus, which is a public database that stores gene expression data from various high-throughput experiments, including microarray and RNA sequencing. It is maintained by the National Center for Biotechnology Information (NCBI) and provides researchers with a resource to access and analyze gene expression data from a wide range of organisms and experimental conditions. GEO is freely accessible to the public and can be searched and downloaded through the NCBI website.

GEO ACCESSION: GSE53987

Title-Microarray profiling of PFC, HPC and STR from subjects with schizophrenia, bipolar, MDD or control.

Summary-Schizophrenia is a complex psychiatric disorder encompassing a range of symptoms and etiology dependent upon the interaction of genetic and environmental factors. Several risk genes, such as DISC1, have been associated with schizophrenia as well as bipolar disorder (BPD) and major depressive disorder (MDD), consistent with the hypothesis that a shared genetic architecture could contribute to divergent clinical syndromes. The present study compared gene expression profiles across three brain regions in post-mortem tissue from matched subjects with schizophrenia, BPD or MDD and unaffected controls. Post-mortem brain tissue was collected from control subjects and well-matched subjects with schizophrenia, BPD, and MDD (n=19 from each group). RNA was isolated from hippocampus, Brodmann Area 46, and associative striatum and hybridized to U133_Plus2 Affymetrix chips. Data were normalized by RMA, subjected to pairwise comparison followed by Benjamini and Hochberg False Discovery Rate correction (FDR). Samples derived from patients with schizophrenia exhibited many more changes in gene expression across all brain regions than observed in BPD or MDD. Several genes showed changes in both schizophrenia and BPD, though the magnitude of change was usually larger in

schizophrenia. Several genes that have variants associated with schizophrenia were found to have altered expression in multiple regions of brains from subjects with schizophrenia. Continued evaluation of circuit-level alterations in gene expression and gene-network relationships may further our understanding of how genetic variants may be influencing biological processes to contribute to psychiatric disease.

Overall design Prefrontal cortex, striatum and hippocampus were obtained from subjects with schizophrenia, bipolar disorder, major depressive disorder and matched controls.

NCBI > GEO > **Accession Display** [?](#) Not logged in | [Login](#) [?](#)

GEO help: Mouse over screen elements for information.

Scope: Format: Amount: GEO accession:

Series GSE53987 [Query DataSets for GSE53987](#)

Status	Public on Jan 11, 2014
Title	Microarray profiling of PFC, HPC and STR from subjects with schizophrenia, bipolar, MDD or control
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	Schizophrenia is a complex psychiatric disorder encompassing a range of symptoms and etiology dependent upon the interaction of genetic and environmental factors. Several risk genes, such as DISC1, have been associated with schizophrenia as well as bipolar disorder (BPD) and major depressive disorder (MDD), consistent with the hypothesis that a shared genetic architecture could contribute to divergent clinical syndromes. The present study compared gene expression profiles across three brain regions in post-mortem tissue from matched subjects with schizophrenia, BPD or MDD and unaffected controls. Post-mortem brain tissue was collected from control subjects and well-matched subjects with schizophrenia, BPD, and MDD (n=19 from each group). RNA was isolated from hippocampus, Brodmann Area 46, and associative striatum and hybridized to U133_Plus2 Affymetrix chips. Data were normalized by RMA, subjected to pairwise comparison followed by Benjamini and Hochberg False Discovery Rate correction (FDR). Samples derived from patients with schizophrenia exhibited many more changes in gene expression across all brain regions than observed in BPD or MDD. Several genes showed changes in both schizophrenia and BPD, though the magnitude of change was usually larger in schizophrenia. Several genes that have variants associated with schizophrenia were found to have altered expression in multiple regions of brains from subjects with schizophrenia. Continued evaluation of circuit-level alterations in gene expression and gene-network relationships may further our understanding of how genetic variants may be influencing biological processes to contribute to psychiatric disease.
Overall design	Pre-frontal cortex, striatum and hippocampus were obtained from subjects with schizophrenia, bipolar disorder, major depressive disorder and matched controls.

FIGURE 4.6: GENE EXPRESSION OMNIBUS IN NCBI

4.4 METABOLOMICS

Metabolomics is the study of metabolites, which are small molecules that are involved in various metabolic processes in cells. Metabolomics involves the identification and quantification of metabolites in biological samples such as blood, urine, or tissue, and the analysis of the resulting data to gain insights into the metabolic processes that are altered in different conditions.

Metabolomics can be performed using various analytical techniques, such as nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography-mass spectrometry (LC-MS), or gas chromatography-mass spectrometry (GC-MS). The resulting data can be analyzed using various bioinformatics tools and platforms, such as MetaboAnalyst, XCMS, or Mummichog. These tools can be used to identify differentially expressed metabolites, pathway analysis, and biomarker discovery.

In the context of major depressive disorder (MDD), metabolomics can be used to identify biomarkers that are associated with the disorder. These biomarkers may be useful for diagnosis, prognosis, or monitoring of treatment response in MDD. Additionally, metabolomics can provide insights into the underlying biological mechanisms of MDD and may help to identify potential therapeutic targets.

4.4.1 HUMAN METABOLOME DATABASE

The Human Metabolome Database (HMDB) is a comprehensive resource for information on human metabolites and their associated pathways. It contains information on over 114,000 metabolites, including their chemical structures, spectral data, physiological roles, and disease associations. The database is freely accessible online and is widely used by researchers in the fields of metabolomics, biochemistry, and clinical medicine.

The HMDB is organized into several sections, including a compound section, a pathway section, and a disease section. The compound section provides detailed information on individual metabolites, including their chemical formula, molecular weight, and structures. It also includes information on the enzymes involved in their biosynthesis and degradation, as well as their physiological roles and disease associations.

The pathway section of the HMDB provides information on metabolic pathways in humans, including their biochemical reactions and the enzymes involved in these reactions. It also includes information on the regulation of these pathways and their roles in normal physiology and disease.

The disease section of the HMDB provides information on the association of metabolites and metabolic pathways with various diseases, including cancer, diabetes, and neurological disorders. This section includes information on the changes in metabolite levels associated with different diseases and the potential diagnostic and therapeutic applications of this information.

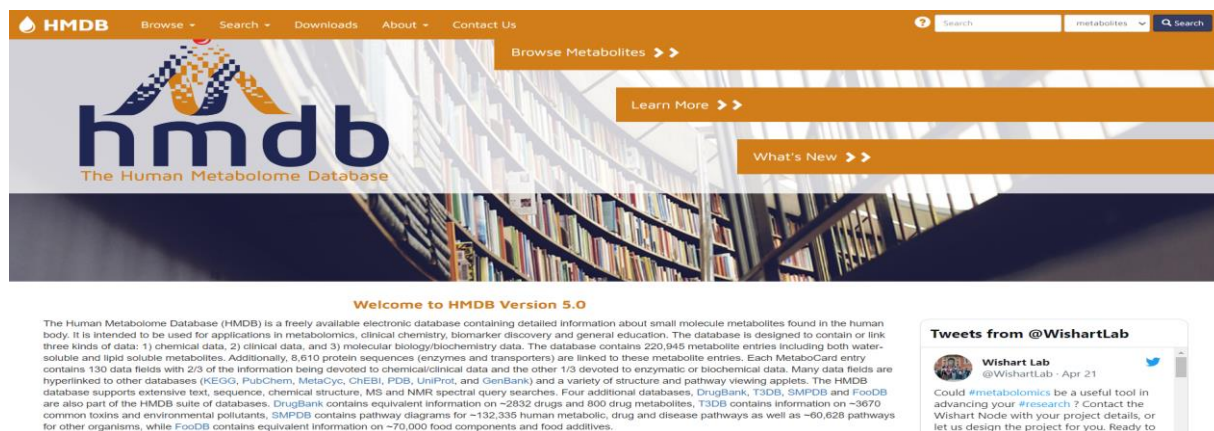


FIG 4.7: HUMAN METABOLOME DATABASE

4.5 EPIGENOMICS

Epigenomics is the study of modifications to DNA and its associated proteins that regulate gene expression without altering the underlying DNA sequence. These modifications can include DNA methylation, histone modifications, and non-coding RNA molecules, among others. Epigenetic changes can be influenced by environmental factors, such as diet and exposure to toxins, and can also be inherited across generations.

In the context of major depressive disorder (MDD), epigenomic changes have been implicated in the pathophysiology of the disorder. For example, studies have shown that DNA methylation patterns are altered in patients with MDD, particularly in genes related to neurotransmitter function and stress response. Histone modifications and non-coding RNA molecules have also been implicated in the development of MDD.

Epigenomic changes can be studied using various techniques, such as chromatin immunoprecipitation (ChIP) assays, bisulfite sequencing, and RNA sequencing. The resulting data can be analyzed using bioinformatics tools and platforms, such as the Ensemble database, the UCSC Genome Browser, and the Epigenome Roadmap Project. The identification of epigenetic changes associated with MDD may have important implications for the development of novel diagnostic and therapeutic strategies. For example, drugs that target specific epigenetic modifications may be effective in treating MDD, and biomarkers based on epigenetic changes may be useful for diagnosis and monitoring of treatment response. Overall, epigenomics is a promising approach for understanding the complex pathophysiology of MDD and may lead to new insights into the development of this disorder.

4.5.1 EWAS

Epigenome-wide association studies (EWAS) are a type of study that aim to identify epigenetic changes associated with a particular disease or trait. EWAS involve profiling DNA methylation levels across the entire genome, typically using microarray or sequencing technologies, and then comparing these levels between cases and controls or individuals with different disease or trait statuses.

In the context of major depressive disorder (MDD), several EWAS studies have been conducted to identify DNA methylation changes associated with the disorder. These studies have identified numerous differentially methylated sites and regions, many of which are located in genes involved in neuronal signaling, immune function, and stress response. Some of these DNA methylation changes have also been associated with treatment response and symptom severity in MDD.

EWAS studies have also been used to investigate the role of environmental factors, such as early life stress and smoking, in the development of MDD. These studies have identified specific DNA methylation changes that may mediate the effects of these environmental factors on MDD risk.

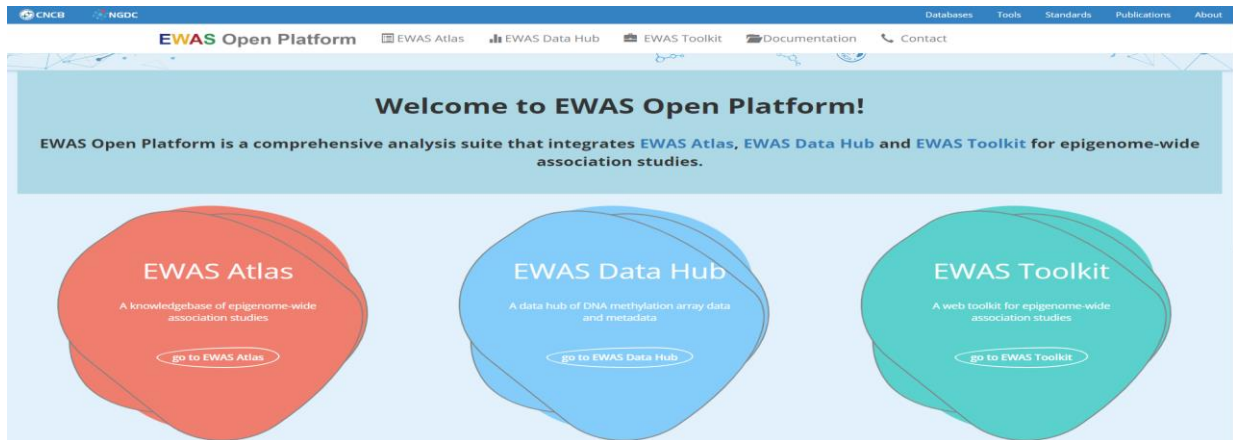


FIG 4.8: EPIGENOME WIDE ASSOCIATION STUDIES

4.6 MICROBIOMICS

Microbiomics is the study of microorganisms that reside in and on the human body, collectively known as the microbiota. The human microbiota is composed of bacteria, viruses, fungi, and other microbes, and is known to play an important role in human health and disease.

In the context of major depressive disorder (MDD), there is growing evidence to suggest that alterations in the composition and activity of the gut microbiota may contribute to the development and progression of the disorder. For example, studies have shown that patients with MDD have alterations in the composition of their gut microbiota, characterized by reduced bacterial diversity and changes in the relative abundance of specific bacterial taxa.

Microbiomic studies in MDD typically involve the sequencing of bacterial DNA from fecal samples or other bodily fluids, such as saliva or blood. This can be done using techniques such as 16S rRNA sequencing or shotgun metagenomics. The resulting data can then be analyzed using various bioinformatics tools and platforms, such as the QIIME pipeline or the Human Microbiome Project.

4.6.1 DISBIOME

A database called Disbiome gathers and standardized information about published microbiota-disease studies. The microorganisms are connected to their NCBI and SILVA taxonomies, while the diseases are categorized using the MedDRA

classification system. Finally, the reporting quality of each study that is a part of the Disbiome database is evaluated using a standardized questionnaire.

Dis Biome is the first database that provides a clear, succinct, and current overview of the variations in microbial composition associated with diseases along with the pertinent data from the research that have been published. The strength of this database rests in the interaction between the human annotation, which assures a clear and structured display of the available data, and the inclusion of references to other databases, which enables both specific and diverse search techniques inside the Disbiome database.

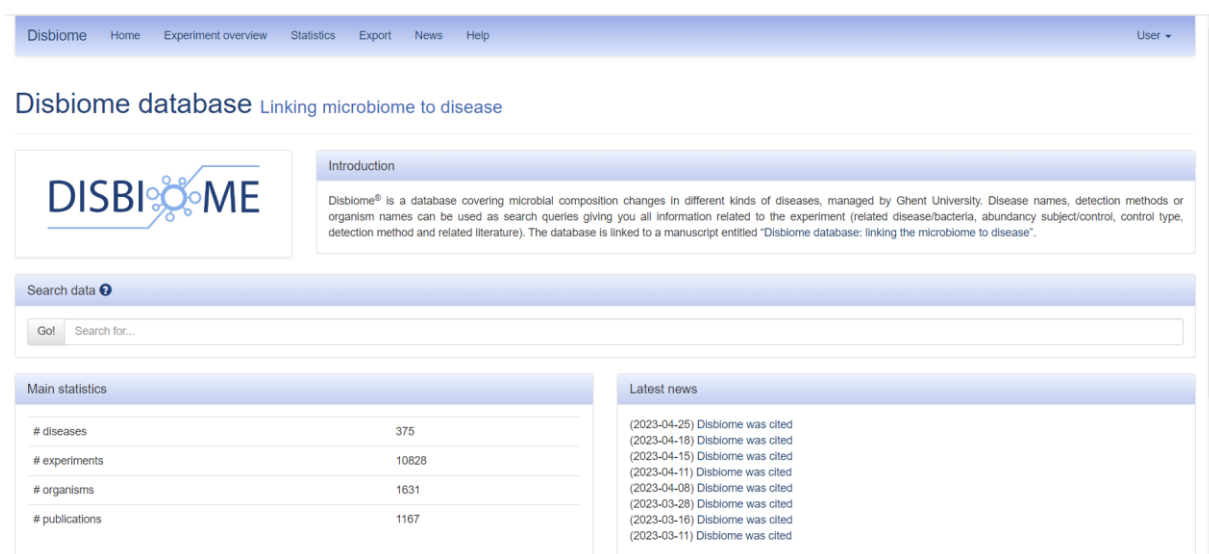


FIGURE 4.9: DISBIOME DATABASE

4.7 SNP

4.7.1 DISGENET

DisGeNET is a publicly available database that integrates information on genes, variants, diseases, and their relationships from various sources, including GWAS, genome-wide association meta-analyses, and other curated databases. DisGeNET includes information on genetic variations, including SNPs, that are associated with various diseases, including major depressive disorder (MDD).

SNP-related data in DisGeNET can be accessed through its user-friendly interface or via programmatic access using its application programming interface (API). The SNP

data in DisGeNET includes information on the location of the SNP, the effect allele and non-effect allele, the minor allele frequency, and the p-value of the association with MDD or other diseases. In addition, DisGeNET also provides information on the genes that are associated with the SNPs, as well as their functional annotations, pathways, and gene-disease associations.

Researchers can use DisGeNET to identify SNPs that are associated with MDD and to explore the potential functional consequences of these SNPs on gene expression, protein function, and other biological processes. This can help to identify potential therapeutic targets and to gain a better understanding of the genetic basis of MDD.

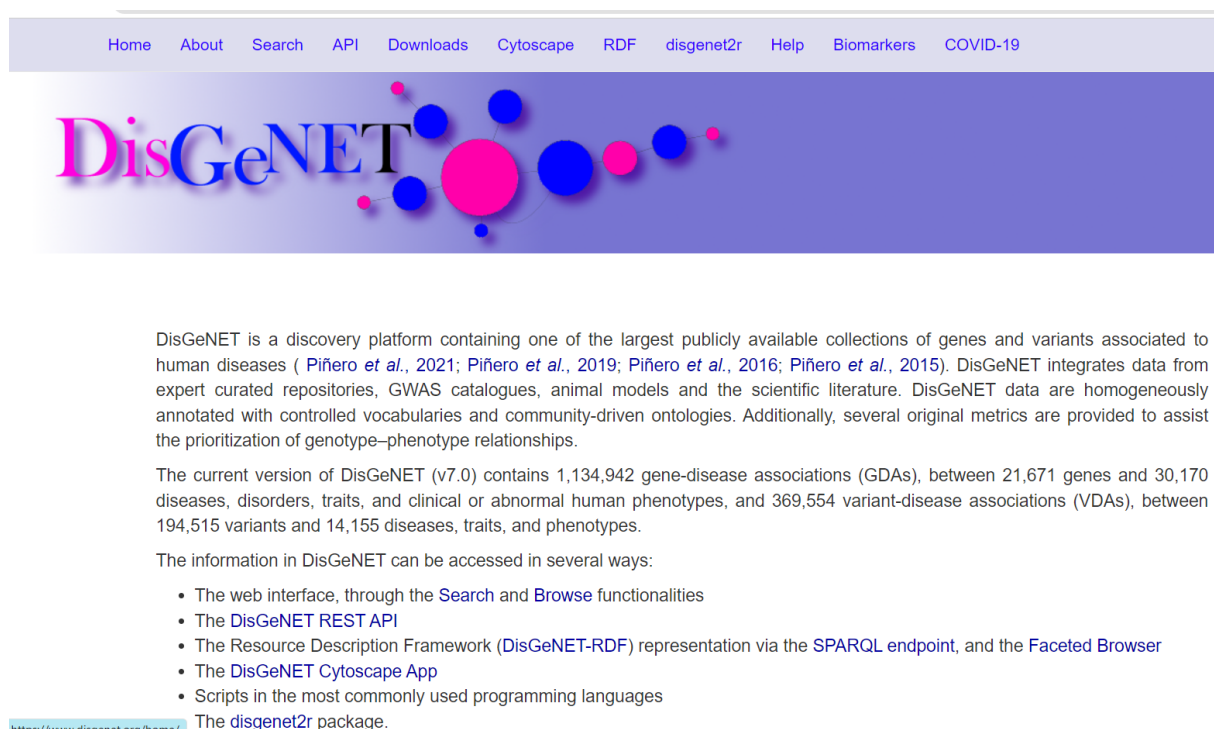


FIG 4.10: DisGeNET database

4.2 TOOLS

4.2.1 NETWORK ANALYST

NetworkAnalyst is a web-based platform for biological network analysis and visualization. It provides a suite of tools for the analysis of molecular interaction networks, including protein-protein interaction (PPI) networks, gene co-expression networks, and metabolic networks. NetworkAnalyst allows users to upload their own network data or to access and analyze publically available data from a variety of sources, including databases such as STRING, KEGG, and BioGRID.

One of the main features of NetworkAnalyst is the ability to perform network enrichment analysis, which identifies functional modules or pathways that are overrepresented in a given network. This allows researchers to gain insight into the biological processes that are involved in a particular disease or condition, such as major depressive disorder (MDD). NetworkAnalyst also provides tools for visualizing and exploring network data, including interactive network diagrams and heatmaps.

NetworkAnalyst is a powerful tool for researchers studying complex biological systems, as it allows for the integration of multiple types of data and the identification of novel biomarkers and therapeutic targets. In the context of MDD, NetworkAnalyst can be used to identify key genes, pathways, and biological processes that are dysregulated in the disease, as well as to explore the relationships between different molecular players and potential treatment targets.

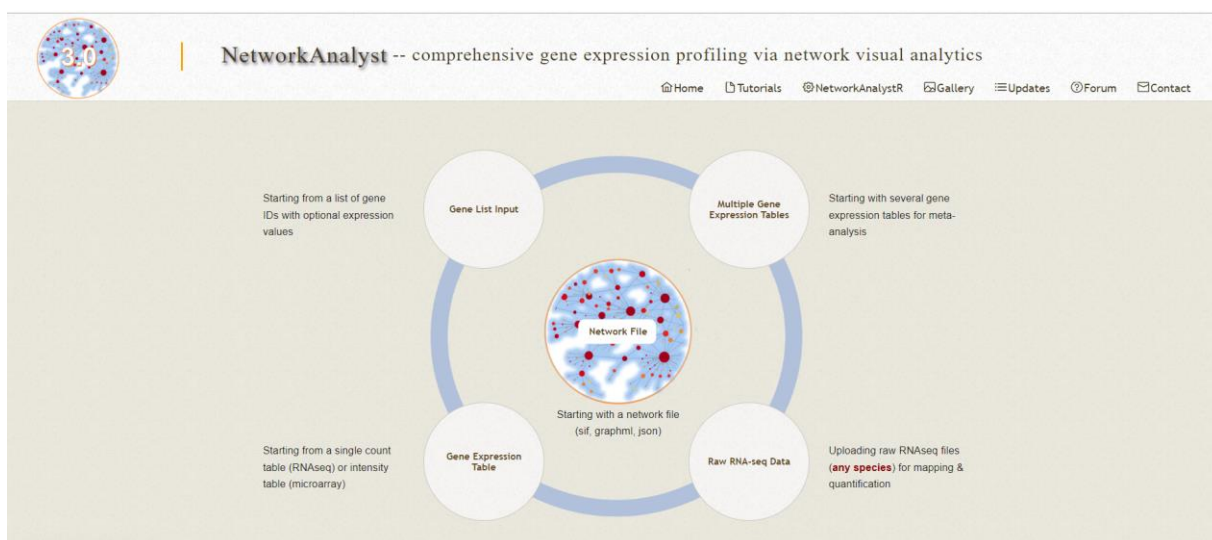


FIGURE 4.11: NETWORK ANALYST TOOL

In this study, Network analysis can be used to identify key biomolecules that are associated with MDD. By analyzing the complex interactions between these molecules, network analysis can reveal which ones play a critical role in the disease.

4.2.2 GENEMANIA

GENEMANIA is a web-based tool for gene function prediction and prioritization based on network analysis. It integrates a variety of functional genomics data, including protein-protein interactions, gene co-expression, pathways, and protein domains, to identify functionally related genes.

GENEMANIA allows users to enter a gene list or a single gene of interest and generate an interactive network of genes and their functional relationships. The tool enables users to visualize the network, explore the functional annotation of genes in the network, and prioritize genes based on their functional relevance to the input gene(s).

GENEMANIA is widely used in biological research, particularly in the identification of novel disease genes, drug targets, and biomarkers.

GENEMANIA

[Cytoscape app](#) | [Help](#) | [Contact](#) | [Cite](#) | [© UToronto](#)

GeneMANIA helps you predict the function of your favourite genes and gene sets.

Indexing 2,830 association networks containing 660,554,667 interactions mapped to 166,691 genes from 9 organisms.

FIGURE 4.12: GENEMANIA (NETWORK ANALYSIS TOOL)

In this study, GeneMANIA can prioritize candidate genes based on their functional relationships with other genes in the network. This can help to identify potential biomarkers and therapeutic targets for MDD.

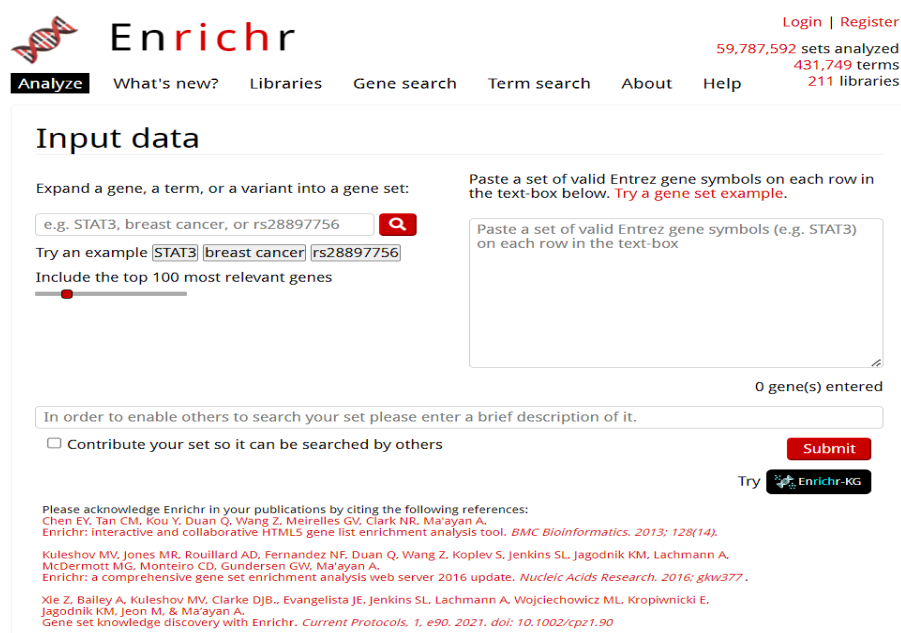
4.2.3 ENRICHR

ENRICHR (ENrichment analysis and Visualization of Gene Ontology and biological pathways) is a web-based tool that provides a comprehensive gene set enrichment analysis for a variety of functional categories. It allows users to input a list of genes and identify enriched biological pathways, gene ontology terms, transcription factor targets, and other functional annotations associated with the gene list.

ENRICHR integrates data from several public databases, including the Gene Ontology (GO), KEGG Pathways, Reactome Pathways, and ChEA transcription factor targets, to identify functional annotations that are enriched in the input gene list.

ENRICHR also provides an intuitive and interactive visualization of the results, including heat maps, bar charts, and scatter plots, to aid in the interpretation of the enriched functional categories. ENRICHR is widely used in biological research to

understand the biological significance of a set of genes, identify potential gene targets, and prioritize genes for further study.



The screenshot shows the Enrichr website interface. At the top, there is a navigation bar with links: Analyze, What's new?, Libraries, Gene search, Term search, About, and Help. On the right, it displays statistics: 59,787,592 sets analyzed, 431,749 terms, and 211 libraries. The main section is titled "Input data" and contains two input areas. The left area is for "Expand a gene, a term, or a variant into a gene set:" and includes a search bar with a magnifying glass icon. Below the search bar, there are example inputs: "e.g. STAT3, breast cancer, or rs28897756" and "Try an example [STAT3] [breast cancer] [rs28897756]". A progress bar indicates "Include the top 100 most relevant genes". The right area is for "Paste a set of valid Entrez gene symbols on each row in the text-box below. Try a gene set example." and includes a large text box for pasting gene symbols. Below the text box, it says "0 gene(s) entered". At the bottom, there is a checkbox labeled "Contribute your set so it can be searched by others" and a "Submit" button. A footer section contains a "Try Enrichr-KG" button and a list of references.

Enrichr

Login | Register

59,787,592 sets analyzed
431,749 terms
211 libraries

Analyze What's new? Libraries Gene search Term search About Help

Input data

Expand a gene, a term, or a variant into a gene set:

e.g. STAT3, breast cancer, or rs28897756

Try an example [STAT3] [breast cancer] [rs28897756]

Include the top 100 most relevant genes

Paste a set of valid Entrez gene symbols on each row in the text-box below. Try a gene set example.

Paste a set of valid Entrez gene symbols (e.g. STAT3) on each row in the text-box

0 gene(s) entered

In order to enable others to search your set please enter a brief description of it.

☐ Contribute your set so it can be searched by others

Submit

Try Enrichr-KG

Please acknowledge Enrichr in your publications by citing the following references:
Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A.
Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013; 12(14).
Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A.
Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016; gkw377.
Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowski ML, Kropiwnicki E, Jagodnik KM, Jeon M, & Ma'ayan A.
Gene set knowledge discovery with Enrichr. *Current Protocols*. 1, e90. 2021. doi: 10.1002/cpz1.90

FIGURE 4.13: ENrichment Analysis Tool

In this study, EnrichR is used for identifying enriched pathways and biological functions associated with MDD-related genes across multiple datasets.

4.2.4 STRING DB

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a database and web-based tool that provides a comprehensive resource for the prediction and analysis of protein-protein interactions (PPIs). It integrates information from various sources, including experimental data, computational predictions, and publicly available databases, to generate a network of predicted functional interactions among proteins. STRING allows users to input a list of proteins or genes and generate an interactive network of predicted interactions, including direct and indirect interactions, functional associations, and pathways. The tool also provides information on the strength of the interactions, the type of evidence supporting the interactions, and the confidence score for each interaction.

STRING also provides additional features, such as gene neighborhood analysis, co-expression analysis, and pathway enrichment analysis, to aid in the interpretation of

the results. STRING is widely used in biological research to identify potential protein-protein interactions, understand the functional relationships among proteins, and prioritize genes for further study. It is a valuable resource for researchers working in systems biology, genomics, and proteomics.

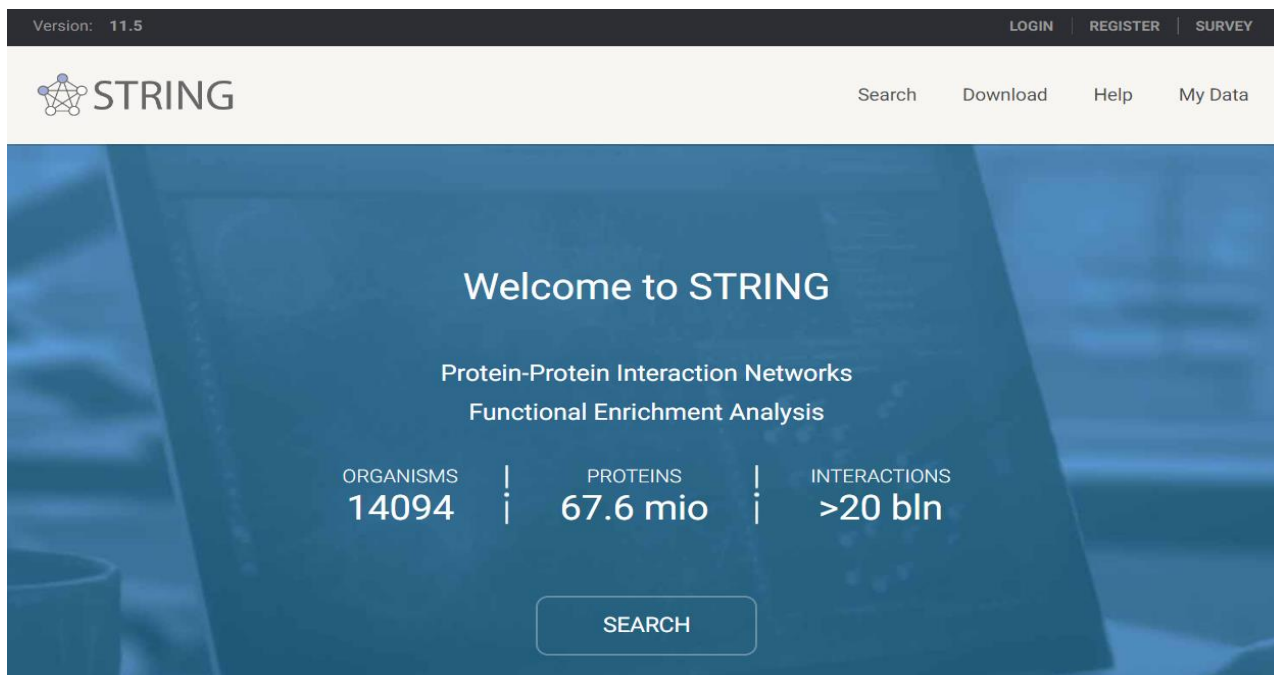


FIGURE 4.14: STRING DATABASE

In this study, String DB is used to construct a protein-protein interaction (PPI) network that includes proteins of interest from multiple omics data types. This can help identify key pathways and biological processes involved in MDD pathogenesis.

4.2.5 METABOANALYST

MetaboAnalyst is a web-based platform for metabolomics data analysis and interpretation. It provides a comprehensive suite of tools for processing, analyzing, and visualizing metabolomics data from a variety of analytical platforms, including mass spectrometry, nuclear magnetic resonance (NMR), and chromatography.

MetaboAnalyst offers a range of statistical and machine learning methods for data processing and normalization, as well as for univariate and multivariate data analysis. It also provides functional analysis tools, such as pathway analysis and metabolite set enrichment analysis, to help identify the biological pathways and processes that are

most relevant to the metabolomic data. MetaboAnalyst provides a user-friendly interface, with interactive visualization tools that allow users to explore their data and interpret their results. It also offers a range of tutorials, online help, and example datasets to assist users with their analysis.

MetaboAnalyst is widely used in metabolomics research to analyze and interpret metabolomic data, identify potential biomarkers, and understand the metabolic pathways involved in biological processes and disease. It is a valuable resource for researchers working in metabolomics, systems biology, and bioinformatics.

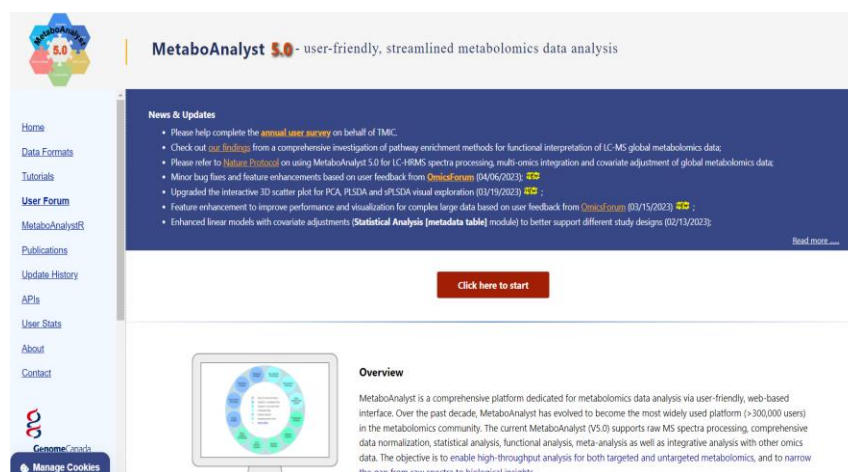


FIGURE 4.15: METABOANALYST

In this study, MetaboAnalyst is used to integrate metabolomics and transcriptomics data to identify metabolites and pathways that are differentially regulated in MDD.

4.2.6 OMICSNET 2.0

OmicsNet is an R package that provides a network-based approach for the analysis and interpretation of high-throughput omics data, such as transcriptomics, proteomics, and metabolomics data. It allows users to construct a network of genes, proteins, or metabolites based on their co-expression or co-regulation, and then analyze the network to identify important functional modules and pathways.

OmicsNet offers a range of network analysis tools, such as centrality analysis, clustering, and module detection, to help identify the most important nodes and subnetworks in the network. It also provides functional enrichment analysis tools, such as Gene Ontology (GO) and pathway enrichment analysis, to help interpret the biological significance of the network modules. OmicsNet is highly flexible, allowing users to integrate data from multiple omics platforms and customize the network construction and analysis parameters to suit their specific needs. It also provides an

interactive user interface that allows users to visualize and explore the network and its modules.

OmicsNet is widely used in systems biology and omics research to identify important biological pathways and processes, understand the relationships between genes, proteins, and metabolites, and identify potential biomarkers and therapeutic targets. It is a valuable resource for researchers working in genomics, proteomics, metabolomics, and bioinformatics.

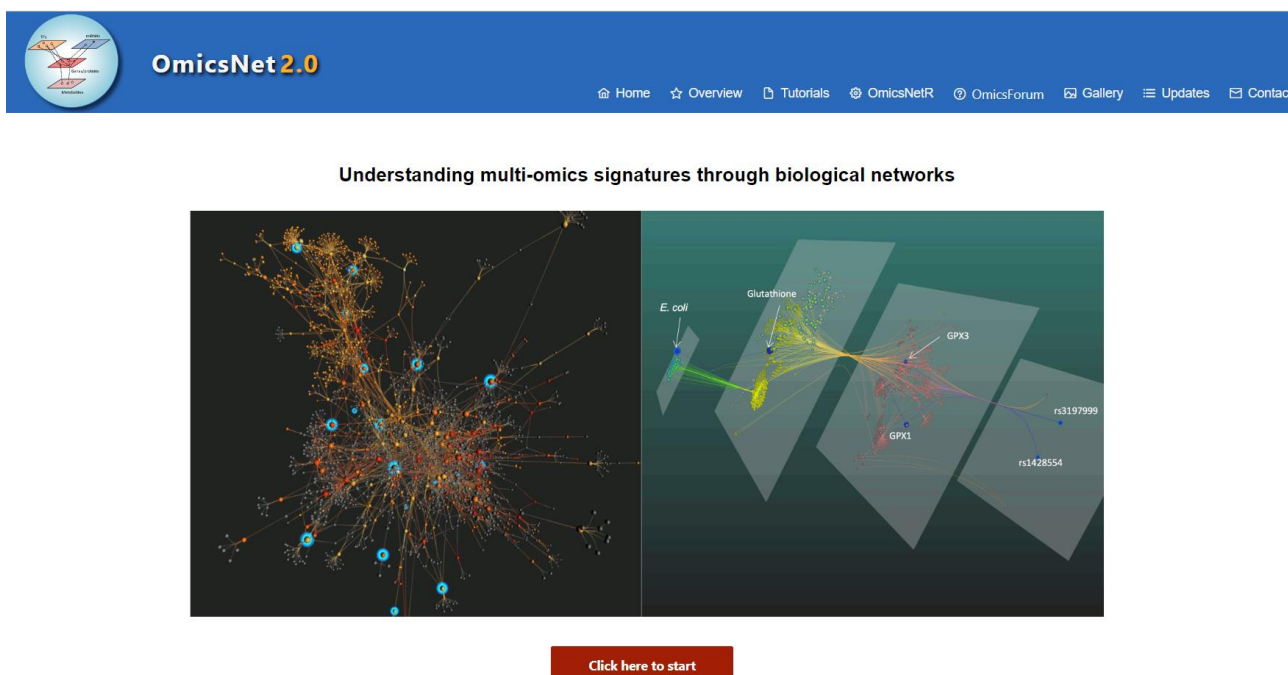


FIGURE 4.16: OMICSNET 2.0

In this study, OmicsNet2.0 is used for constructing and visualizing network models to help understand the complex relationships between different omics data.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 GEO AND NETWORK ANALYST

GEO2R tool used to analyze the GSE98793 dataset, which includes gene expression data from 128 MDD patients and 64 healthy controls. "Case" vs. "Control" comparison groups were established, and the quantile normalization approach was used to preprocess the data. To find DEGs, I utilized a significance level of $p < 0.05$ and a fold change limit of 1.5.

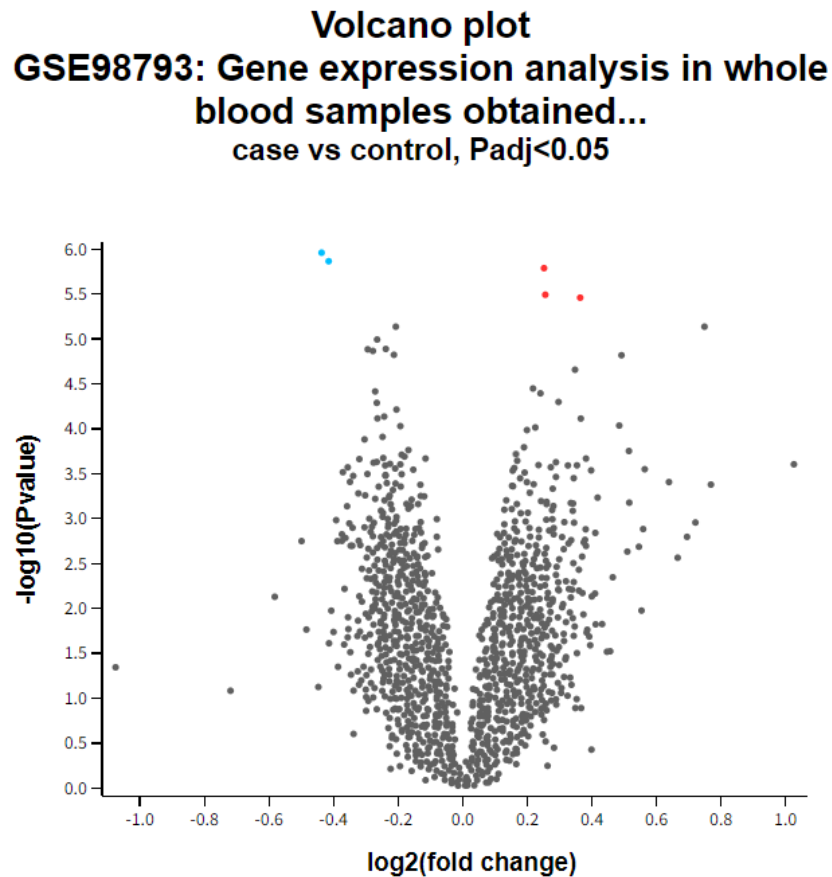


FIG:5.1(Volcano diagram, black indicates genes with no differential expression, blue indicates down-regulated genes, and red indicates up-regulated genes.)

Venn Diagram
GSE98793: limma, Padj<0.05

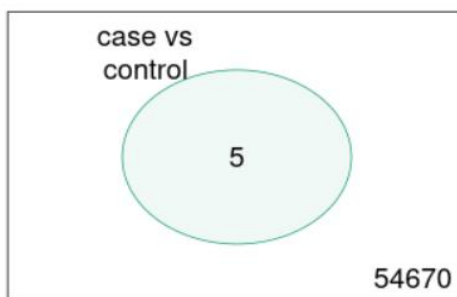


FIGURE 5.2: VENN DIAGRAM OF CASE AND CONTROL FROM GEO2R

The top 10 upregulated genes include OLFM4,CEACAM8,MMP8,LCN2,LTF,DEFA4,ARG1,while the top 10 downregulated genes included KCNA3, TREML4,PPP1R17.

TABLE 5.1: UPREGULATED GENES FROM GEO2R ANALYSIS

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symk	Gene.title	
212768_s_	0.2654	0.000249	3.732128	0.06389	1.03	OLFM4	olfactomedin 4	
206676_at	0.2874	0.000419	3.589853	-0.33375	0.769	CEACAM8	carcinoembryonic ant	
231688_at	0.0572	7.32E-06	4.608674	2.7897	0.749	MMP8	matrix metallopeptida	
212531_at	0.4234	0.001108	3.311084	-1.07433	0.721	LCN2	lipocalin 2	
202018_s_	0.4273	0.0016	3.200823	-1.35284	0.695	LTF	lactotransferrin	
207269_at	0.4541	0.002727	3.035789	-1.75411	0.666	DEFA4	defensin alpha 4	
206177_s_	0.2871	0.000392	3.607979	-0.28382	0.639	ARG1	arginase 1	
211657_at	0.2654	0.000282	3.698315	-0.0318	0.564	CEACAM6	carcinoembryonic ant	
205557_at	0.4273	0.001309	3.261413	-1.20082	0.559	LOC14968	uncharacterized LOC1	
206371_at	0.5809	0.010581	2.581189	-2.76006	0.554	FOLR3	folate receptor 3	

TABLE 5.2: DOWNREGULATED GENES FROM GEO2R ANALYSIS

207509_s_	0.5809	0.010617	-2.57998	-2.76254	-0.407	LAIR2	leukocyte associated
209686_at	0.6741	0.024508	-2.2667	-3.36792	-0.414	S100B	S100 calcium binding
206666_at	0.0297	1.36E-06	-4.98521	4.09623	-0.415	GZMK	granzyme K
236203_at	0.7974	0.074988	-1.79012	-4.14667	-0.447	LOC10050	HLA class II histocomp
212999_x_	0.636	0.017213	-2.40271	-3.11404	-0.484	LOC10106	HLA class II histocomp
235104_at	0.4349	0.001781	-3.16829	-1.43344	-0.499	ERAP2	endoplasmic reticular
227194_at	0.5468	0.007435	-2.70502	-2.50074	-0.582	FAM3B	family with sequence
209480_at	0.808	0.082575	-1.74493	-4.21142	-0.719	HLA-DQB1	major histocompatibi
213831_at	0.7381	0.04527	-2.01509	-3.80076	-1.08	LOC10050	HLA class II histocomp

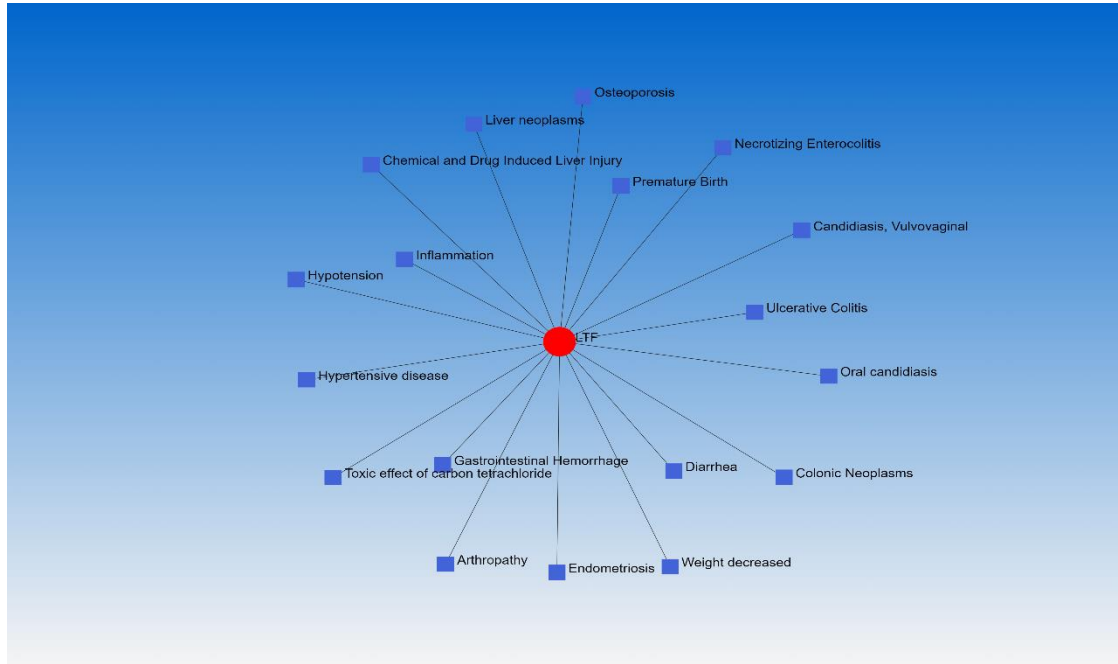


FIG 5.3: GENE-DISEASE PATHWAY ANALYSIS USING NETWORK ANALYST

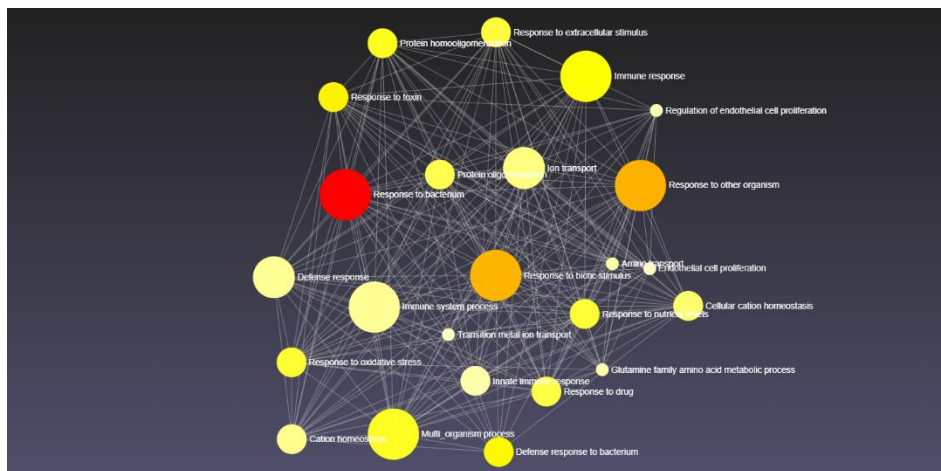


FIG 5.4: MOLECULAR FUNCTION PATHWAY ANALYSIS USING NETWORK ANALYST

The MDD genes with the highest hits were analyzed using the Network Analyst programme. "Gene Set Enrichment Analysis" (GSEA) tool to conduct a functional analysis on a gene list that contained the detected genes and chose the annotation source to be "PANTHER for Molecular Function" and used a significance level of p 0.05.

The biological processes that were enriched and connected to the high-hit MDD genes include LFT, according to the functional analysis. Ulcerative colitis, colonic neoplasms, diarrhea, gastrointestinal hemorrhage, hypertension, hypotension, inflammation, arthropathy, liver neoplasms, osteoporosis, preterm birth, and weight loss were some of these processes.

Functional annotation study found that the upregulated genes were enriched in stress response, inflammation, and neurotransmitter signaling networks, whereas the downregulated genes were enriched in synaptic function, energy metabolism, and neuroplasticity pathways.

5.2 PSYGENET AND GENEMANIA

The web-based GeneMANIA programme to investigate gene-gene interactions and functional correlations related to MDD. GeneMANIA was fed 237 genes associated with MDD from the PsyGeNET database, and a network analysis was done to find genes that are functionally related to MDD genes. A pathway analysis was also

performed to discover biological pathways that are over-represented among the MDD genes.

TABLE 5.3: GENE DATA FROM PSYGENET

Gene Sym	Gene Id	Gene Desc	Disease Id	Disease Cc	Disease Nz	Psychiatric
NPY	4852	neuropept	umls:C126	C1269683	major dep	Depressive
CRH	1392	corticotro	umls:C126	C1269683	major dep	Depressive
COMT	1312	catechol-C	umls:C126	C1269683	major dep	Depressive
BDNF	627	brain-deriv	umls:C126	C1269683	major dep	Depressive
NR3C1	2908	nuclear rec	umls:C126	C1269683	major dep	Depressive
HTR2A	3356	5-hydroxyt	umls:C126	C1269683	major dep	Depressive
CACNA1C	775	calcium ch	umls:C126	C1269683	major dep	Depressive
GSK3B	2932	glycogen s	umls:C126	C1269683	major dep	Depressive
NDUFV2	4729	NADH deh	umls:C126	C1269683	major dep	Depressive

Based on co-expression, co-localization, shared protein domains, and anticipated interactions, the GeneMANIA analysis found 107 genes that are functionally connected to the MDD genes. The functional linkages were shown to be enriched in numerous biological processes, including synaptic transmission modulation, neurotransmitter level regulation, and gene expression regulation. Several pathways were found to be considerably enriched in the MDD genes, including the calcium signaling route, the cAMP signaling pathway, and the MAPK signaling pathway.

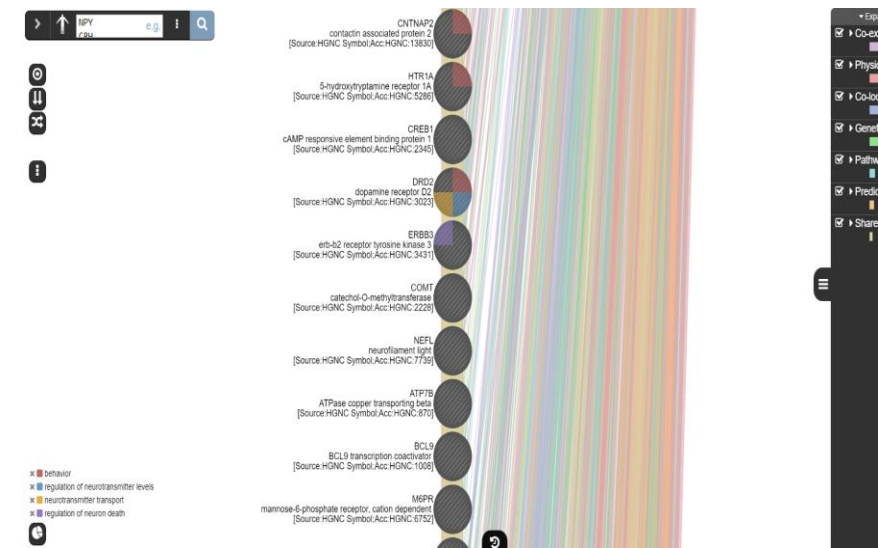


FIGURE 5.5: CO EXPRESSION PATHWAY ANALYSIS ZOOMED VIEW

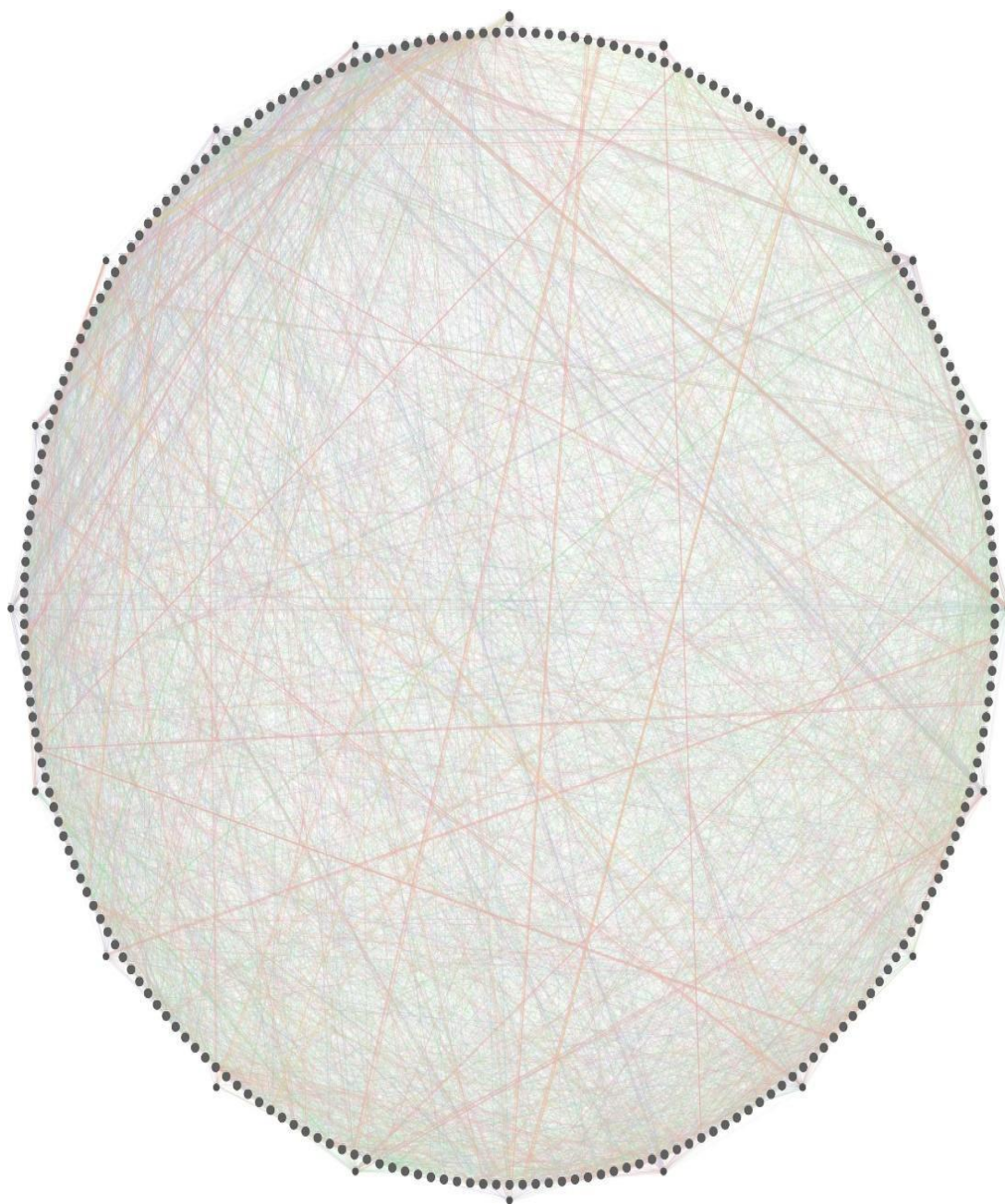


FIG 5.6: PATHWAY ANALYSIS OF DEG's FROM PSYGENET CONSTRUCTED USING GENEMANIA

5.3 GWAS AND ENRICHR

TABLE 5.4: GENE-DISEASE ASSOCIATION DATA FROM GWAS

DISEASE/TRAIT	REPORTED	MAPPED_	SNP_GENE	SNPS	PVALUE_N	PLATFORM [SNPS P	CNV
Major depressive disorder	OR2T12	OR2T12	ENSG0000	rs4131057	19.39794	Illumina [83898]	N
Major depressive disorder	TMEM150	TMEM150	ENSG0000	rs2019353	10.30103	Illumina [83898]	N
Major depressive disorder	PRR5-ARH	PRR5-ARH	ENSG0000	rs1403958	9.522879	Illumina [83898]	N
Major depressive disorder	MUC5B	MUC5B	ENSG0000	rs5629320	8.69897	Illumina [83898]	N
Major depressive disorder	C2orf54	MAB21L4	ENSG0000	rs7856245	7.221849	Illumina [83898]	N
Major depressive disorder	TRIO	TRIO	ENSG0000	rs1150544	6.522879	Illumina [83898]	N
Major depressive disorder	ANO8	ANO8	ENSG0000	rs1436964	6.30103	Illumina [83898]	N
Major depressive disorder	KRBA1	KRBA1	ENSG0000	rs7484419	6.30103	Illumina [83898]	N
Major depressive disorder	BCAR3	BCAR3	ENSG0000	rs6200102	6.30103	Illumina [83898]	N

BioPlanet 2019

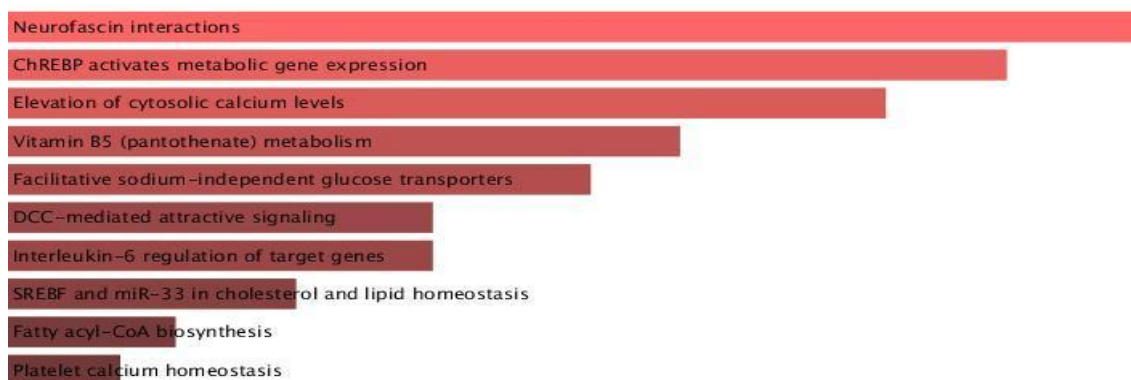


FIG 5.7: BIOPLANET2019 PATHWAY ANALYSIS ENRICHR IN BAR GRAPH

KEGG 2021 Human

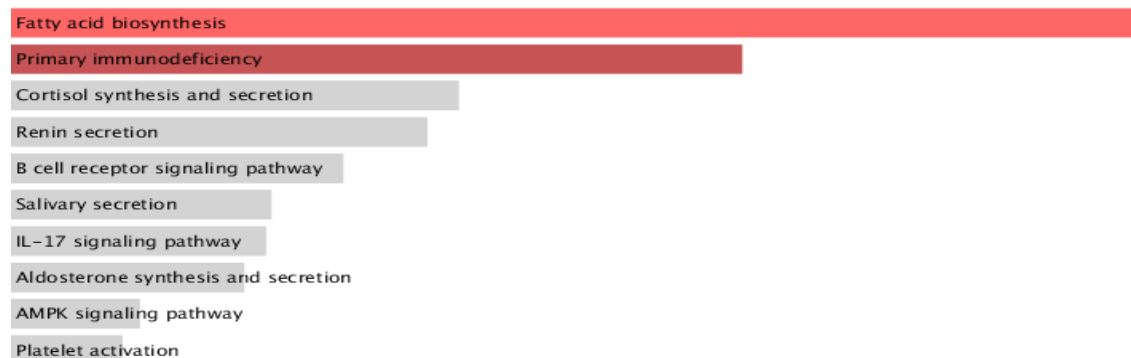


FIG 5.8: KEGG2021 HUMAN PATHWAY ANALYSIS ENRICHR IN BAR GRAPH

Reactome 2022

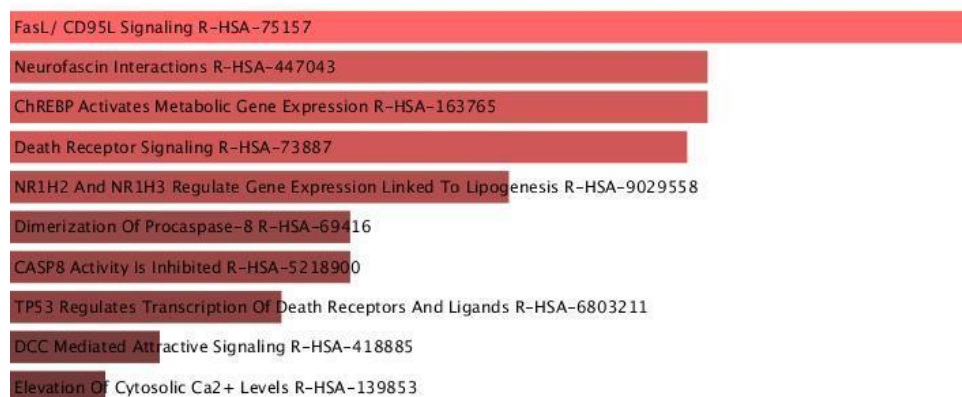


FIG 5.9: RECTOME2022 PATHWAY ANALYSIS IN BAR GRAPH

Elsevier Pathway Collection

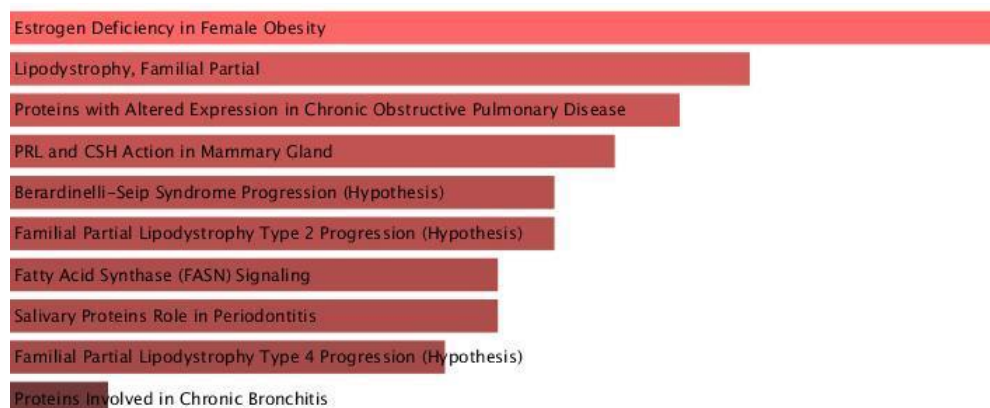


FIG 5.10: ELSEVIER PATHWAY COLLECTION IN BAR GRAPH

5.4 HUMAN PROTEIN ATLAS, UNIPROT & STRING DB

The Human Protein Atlas is an open-access database that contains detailed information regarding protein expression, localization, and function in human tissues and cells.

I obtained ten protein ids for MDD from HPA and six from Uniprot.

MECP2 modulates neuronal receptors and channels, and protein-protein interactions provide insights into the biological processes and pathways involved. A gene-disease link has been discovered in the regulation of behavioral fear response.

TABLE 5.5: PROTEIN DATA RETRIEVED FROM HUMAN PROTEIN ATLAS

Column1	Column2	Column3	Column4		
Gene description	Uniprot	Protein cla	Evidence		
5-hydroxytryptamine receptor 2A	P28223	FDA appro	Evidence at protein level		
Piccolo presynaptic cytomatrix pro	Q9Y6V0	Disease re	Evidence at protein level		
Inter-alpha-trypsin inhibitor heavy	Q06033	Plasma prc	Evidence at protein level		
LDL receptor related protein 8	Q14114	Disease re	Evidence at protein level		

TABLE 5.6: PROTEIN DATA RETRIEVED FROM UNIPROT

Entry	Reviewed	Entry Nam	Protein names	Gene Nam	Organism	Length
O76024	reviewed	WFS1_HU	Wolframin	WFS1	Homo sapi	890
P08172	reviewed	ACM2_HU	Muscarinic acetylcholine r	CHRM2	Homo sapi	466
P17861	reviewed	XBP1_HU	X-box-binding protein 1 (X	XBP1 TREE	Homo sapi	261
P51608	reviewed	MECP2_HI	Methyl-CpG-binding prote	MECP2	Homo sapi	486
Q8IWU9	reviewed	TPH2_HU	Tryptophan 5-hydroxylase	TPH2 NTP	Homo sapi	490

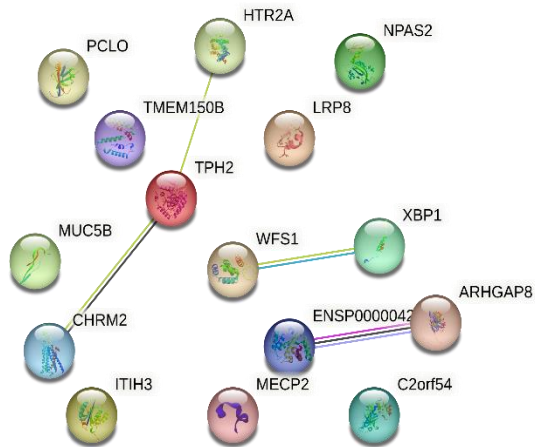


FIGURE 5.11: PPI INTERSECTION AS BITMAP IMAGE USING STRING DB

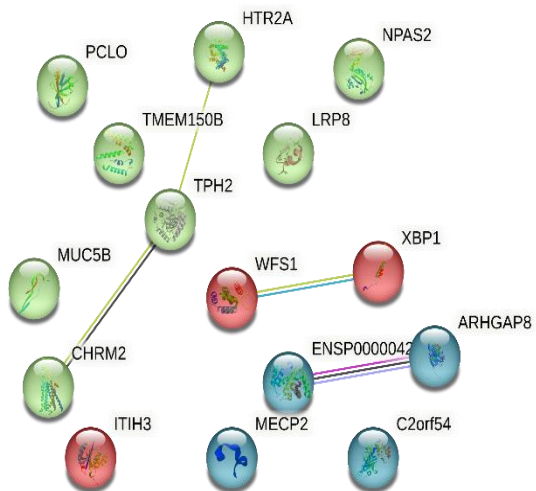


FIGURE 5.12: K-MEANS CLUSTERING PPI INTERSECTION USING STRING DB

TABLE 5.7: K-MEANS CLUSTERING ANALYSIS RESULT USING STRING DB

#cluster	cluster	cluster	gene co	protein	protein	protein
kmeans	1	Red	3	ITIH3	9606.ENSF	Inter-alpha-trypsin int
kmeans	1	Red	3	WFS1	9606.ENSF	Wolframin er transme
kmeans	1	Red	3	XBP1	9606.ENSF	X-box-binding protein
kmeans	2	Green	8	CHRM2	9606.ENSF	Muscarinic acetylchol
kmeans	2	Green	8	HTR2A	9606.ENSF	5-hydroxytryptamine
kmeans	2	Green	8	LRP8	9606.ENSF	Low-density lipoprote
kmeans	2	Green	8	MUC5B	9606.ENSF	Mucin 5b, oligomeric
kmeans	2	Green	8	NPAS2	9606.ENSF	Neuronal PAS domain
kmeans	2	Green	8	PCLO	9606.ENSF	Protein piccolo; May i
kmeans	2	Green	8	TMEM150	9606.ENSF	Modulator of macroa
kmeans	2	Green	8	TPH2	9606.ENSF	Tryptophan 5-monoo
kmeans	3	Blue	4	ARHGAP8	9606.ENSF	Rho GTPase activating
kmeans	3	Blue	4	C2orf54	9606.ENSF	Uncharacterized prot
kmeans	3	Blue	4	ENSP0000	9606.ENSF	annotation not availa
kmeans	3	Blue	4	MECP2	9606.ENSF	Methyl-CpG-binding p

5.5 TRANSCRIPTOME DATA FROM GEO

The data is obtained from GEO, the accession id is gse53987

Overall design: Prefrontal cortex, striatum and hippocampus were obtained from subjects with schizophrenia, bipolar disorder, major depressive disorder and matched controls. (only MDD is used for analysis)

GEO2R analysis is done for the above data and the upregulated and downregulated genes are extracted, the top 11 genes that are upregulated in an MDD-related dataset: TEK4P2,MAFIP,LOC389834,GSTM5,SLC1A2,RAB12,AQP4,RGMB,SLC1A2,LOC728613,RGS1,CRH,FAS

It is important to note that these genes have not been consistently associated with MDD in the literature and their biological significance in MDD is not well-established. However, some of these genes have been implicated in various biological processes

that may be relevant to MDD pathogenesis. For example, SLC1A2 encodes for a glutamate transporter that regulates excitatory neurotransmission, which has been implicated in MDD. CRH, a neuropeptide involved in the hypothalamic-pituitary-adrenal (HPA) axis, has also been linked to stress response and MDD.

The top differentially expressed genes with the lowest adjusted p-values: HBA2/HBA1, HBB, TTR, KANSL1-AS1, and FRG1. These genes have been previously implicated in depression or related pathways, and their differential expression in MDD may provide insight into the molecular mechanisms underlying the disorder

TABLE 5.8: DOWNREGULATED TRANSCRIPTOME FROM GEO2R ANALYSIS

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbl
226558_at	1	0.08325	-1.74836	-4.3	-0.46997	TEKT4P2//
205752_s	1	0.06072	-1.89538	-4.21	-0.45866	GSTM5
1558009_a	1	0.13959	-1.48831	-4.43	-0.45075	SLC1A2
235059_at	1	0.04998	-1.98236	-4.16	-0.44507	RAB12
210067_at	1	0.13434	-1.50855	-4.42	-0.41022	AQP4
227340_s	1	0.00878	-2.66929	-3.67	-0.40027	RGMB
1558010_s	1	0.22596	-1.21779	-4.55	-0.39519	SLC1A2
1569110_x	1	0.07733	-1.78347	-4.28	-0.38365	LOC72861
202988_s	1	0.09229	-1.69856	-4.33	-0.3692	RGS1

TABLE 5.9: UPREGULATED TRANSCRIPTOME FROM GEO2R ANALYSIS

209788_s	1	0.03178	2.175414	-4.04	0.325903	ERAP1
1556049_a	1	0.00158	3.241958	-3.18	0.327212	RTN4
211696_x	1	0.12993	1.526054	-4.41	0.33243	HBB
217232_x	1	0.12281	1.555289	-4.4	0.334727	HBB
211745_x	1	0.08737	1.725116	-4.31	0.33584	HBA2///HE
209660_at	1	0.36043	0.918458	-4.66	0.337169	TTR
230388_s	1	0.08415	1.743209	-4.3	0.349766	KANSL1-AS
209116_x	1	0.06851	1.84004	-4.25	0.365154	HBB
229007_at	1	0.0711	1.822829	-4.26	0.404559	LOC28378
203510_at	1	0.01301	2.525573	-3.79	0.463973	MET

5.6 HUMAN METABOLOME DATABASE AND METABOANALYST

TABLE 5.10: METABOLOME DATA FROM HUMAN METABOLOME DATABASE

Allopregna	HMDB000	[H][C@@]	Cerebrosp	0.0000150	Abnormal	Adult (>18	Both
Vanylglyco	HMDB000	COC1=CC(	Cerebrosp	0.09 +/- 0.	Abnormal	Adult (>18	Both
5-Hydroxy	HMDB000	OC(=O)CC	Cerebrosp	0.08 +/- 0.	Abnormal	Adult (>18	Both
Homovani	HMDB000	COC1=CC(	Cerebrosp	0.20 +/- 0.	Abnormal	Adult (>18	Both
5a-Pregna	HMDB000	[H][C@@]	Blood	0.042 +/- C	Abnormal	Adult (>18	Both
Allopregna	HMDB000	[H][C@@]	Blood	0.001 +/- C	Abnormal	Adult (>18	Both
Progester	HMDB000	[H][C@@]	Blood	0.0025 +/-	Abnormal	Adult (>18	Both
Dihydrotes	HMDB000	[H][C@@]	Blood	0.0054 +/-	Abnormal	Adult (>18	Both
Tetrahydro	HMDB000	[H][C@@]	Blood	0.0056 +/-	Abnormal	Adult (>18	Both
N-Methylt	HMDB000	CNCCC1=C	Urine	0.0049 (0.	Abnormal	Adult (>18	Both
5-Methoxy	HMDB000	COC1=CC2	Urine	0.0052 (0.	Abnormal	Adult (>18	Both
Indoleacet	HMDB000	OC(=O)CC	Cerebrosp	0.0299 +/-	Abnormal	Not Specif	Not Specif
alpha-Lino	HMDB000	CC\C=C/C\	Blood	66.803 +/-	Abnormal	Adult (>18	Both
Eicosapen	HMDB000	CC\C=C/C\	Blood	63.151 +/-	Abnormal	Adult (>18	Both

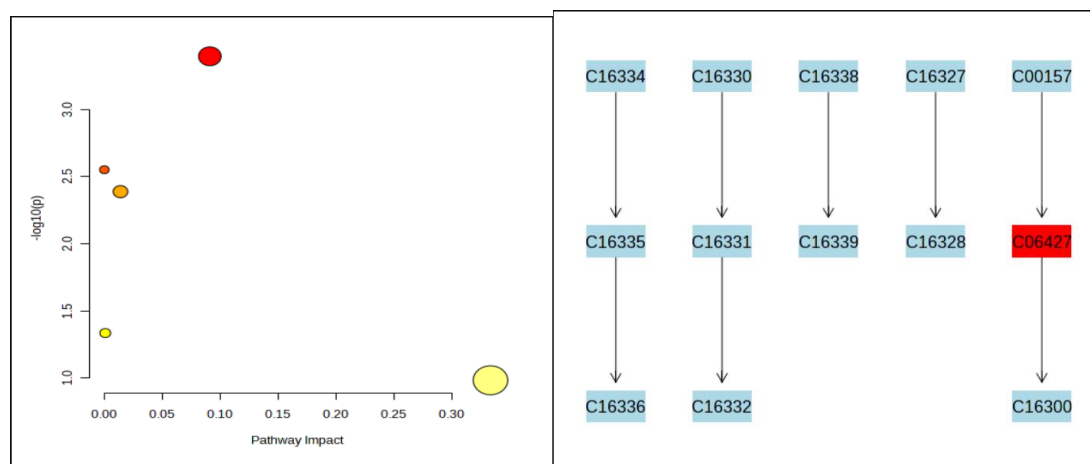


FIGURE 5.13: OVERVIEW OF PATHWAY ANALYSIS USING METABOANALYST

TABLE 5.11: METABOANALYST PATHWAY RESULT

	Total	Expected	Hits	Raw p	#NAME?	Holm adj	FDR	Impact
Steroid hormone biosynt	85	0.7129	5	0.000401	3.397	0.033671	0.033671	0.09105
Biosynthesis of unsaturat	36	0.30194	3	0.002807	2.5518	0.23296	0.1145	0
Tryptophan metabolism	41	0.34387	3	0.004089	2.3884	0.33532	0.1145	0.0139
Tyrosine metabolism	42	0.35226	2	0.046287	1.3345	1	0.97204	0.00077
alpha-Linolenic acid met	13	0.10903	1	0.10409	0.98257	1	1	0.33333

Alpha-linolenic acid is an important omega-3 fatty acid, so this pathway may also be involved in changes related to lipid metabolism.

Glutathione metabolism: This pathway showed a relatively weak association with gene expression data, with 2 out of 56 genes mapped and a p-value of 0.06895. Glutathione is an important antioxidant that helps protect cells from oxidative stress, so changes in this pathway may be related to cellular stress response.

5.7 OMICSNET R

The input for omicsnet R is gene symbol from GEO, protein id from human protein atlas and uniport, transcription factor(TF),SNP's, Metabolites, Microbial taxa.

Using input data on SNPs and gene, Omicsnet can identify genetic variants that are associated with MDD as SOX10, ZNF263

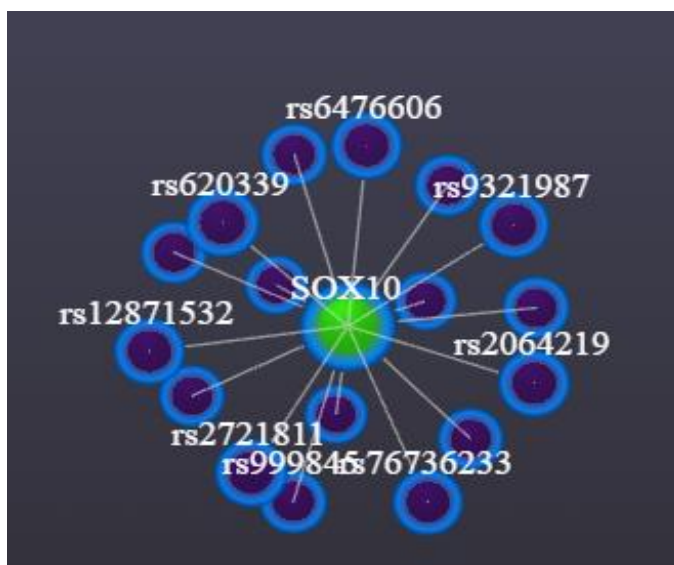


FIG 5.15: OMICSNET ANALYSIS RESULT FOR TF AND SNP'S (subnetwork1)

SOX10 is a transcription factor that plays an important role in the development of neural crest cells, which give rise to various cell types including melanocytes, peripheral glia, and neurons. It is also known to play a role in the maintenance of the oligodendrocyte lineage, which is responsible for the production of myelin in the central nervous system.

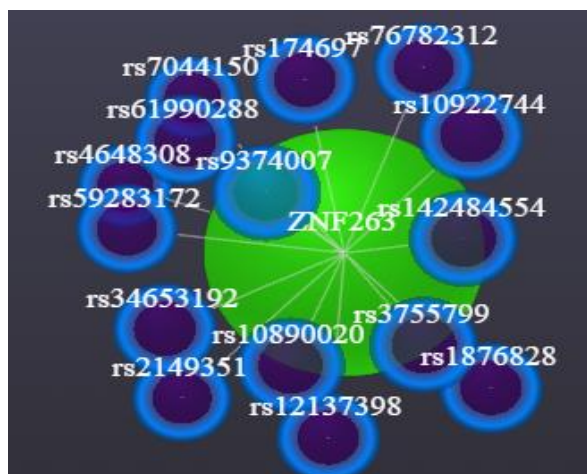


FIG 5.16: OMICSNET ANALYSIS RESULT FOR TF AND SNP'S(subnetwork2)

ZNF263, also known as ZBTB18, is a transcription factor that has been implicated in various biological processes, including cell differentiation and proliferation, apoptosis, and development of the nervous system.

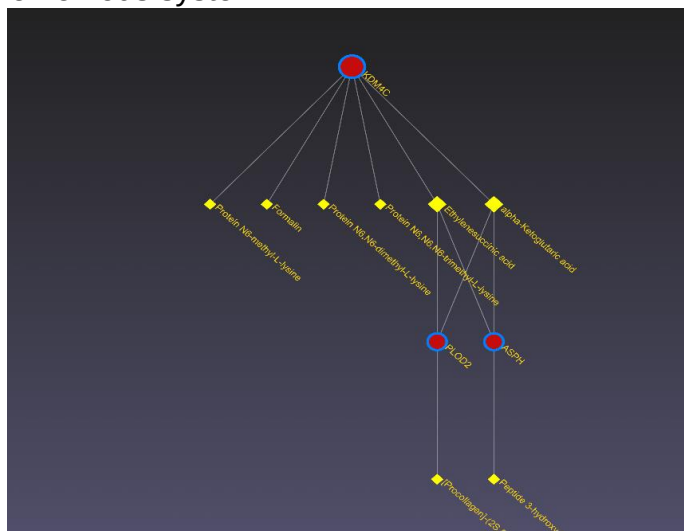


FIG 5.17: OMICSNET ANALYSIS FOR METABOLITE AND PROTEIN

The above analysis has identified potential interactions between the Metabolites alpha-Ketoglutaric acid and Ethylene succinic acid, and the proteins KDM4C, ASPH, and PLOD2.

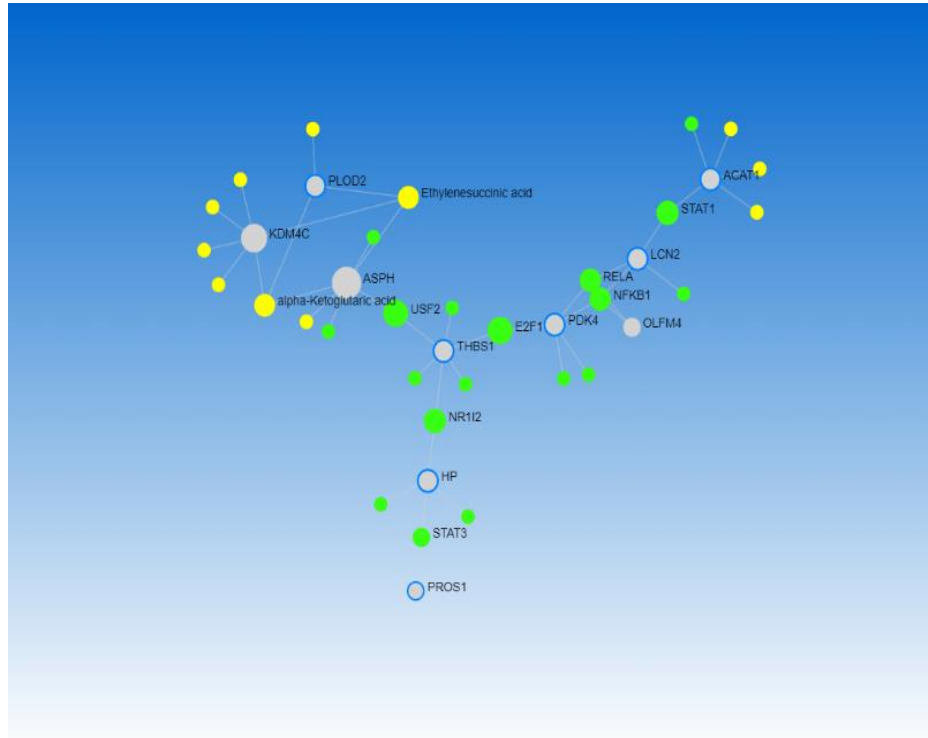


FIG 5.18: TF AND METABOLITE INTERACTION ANALYSIS USING OMICS

The most enriched biological process is "response to stress" with a p-value of 0.00349 and an FDR of 0.241, indicating a high level of significance. This suggests that the genes involved in this biological process are more likely to be differentially expressed in response to stress.

The second most enriched biological process is "cellular response to extracellular stimulus" with a p-value of 0.00371 and an FDR of 0.241, which is also significant. This process involves the response of cells to signals from the extracellular environment, which can include various stressors.

The third and fourth most enriched biological processes are "protein maturation" and "negative regulation of catalytic activity," respectively. Both of these processes are involved in protein regulation and can play a role in the cellular response to stress.

Finally, the fifth most enriched biological process is the "inflammatory response," which is also involved in the response to stress. Inflammation is a well-known response to stress and can play a role in the development of various stress-related disorders, including depression.

Overall, these enriched biological processes provide insight into the molecular mechanisms underlying the cellular response to stress and can help identify potential therapeutic targets for stress-related disorders such as depression.

CHAPTER 6

SUMMARY AND CONCLUSION

This study is a combination of several different approaches to studying major depressive disorder (MDD). These approaches include differential gene expression analysis, functional analysis, gene network analysis, pathway analysis, and compound gene mapping analysis.

Differential gene expression analysis was performed using the GEO2R tool, which identified several differentially expressed genes (DEGs) between MDD patients and healthy controls. The top 10 upregulated genes included OLFM4, CEACAM8, MMP8, LCN2, LTF, DEFA4, and ARG1, while the top 10 downregulated genes included KCNA3, TREML4, and PPP1R17. These genes were then subjected to functional analysis using the Network Analyst program and the Gene Set Enrichment Analysis (GSEA) tool, which revealed several biological processes and pathways that were enriched in MDD.

The functional analysis found that the upregulated genes were enriched in stress response, inflammation, and neurotransmitter signaling networks, whereas the downregulated genes were enriched in synaptic function, energy metabolism, and neuroplasticity pathways. The pathway analysis using GeneMANIA found that several pathways were significantly enriched in the MDD genes, including the calcium signaling pathway, the cAMP signaling pathway, and the MAPK signaling pathway. These pathways are known to play important roles in neuronal signaling, synaptic plasticity, and gene expression regulation, and their dysregulation may contribute to the development of MDD.

Compound gene mapping analysis was performed using MetaboAnalyst, which identified significant associations between specific metabolic pathways and gene expression data. The analysis found that the biosynthesis of unsaturated fatty acids showed the strongest association with gene expression data, suggesting that there may be significant changes in unsaturated fatty acid biosynthesis related to the input gene

expression data. Alpha-linolenic acid metabolism and glutathione metabolism were also found to be associated with gene expression data, indicating that changes in lipid metabolism and cellular stress response may be related to MDD.

Finally, protein-metabolite interaction network analysis using OMICSNET identified potential interactions between the metabolite's alpha-ketoglutaric acid and ethylene succinic acid and the proteins KDM4C, ASPH, and PLOD2. These interactions may play important roles in MDD pathogenesis and warrant further investigation.

In summary, this analysis combines several different approaches to studying MDD and identifies several key biological processes, pathways, and genes that may be involved in the development of the disorder. These findings provide valuable insights into the molecular mechanisms underlying MDD and may guide the development of novel diagnostic and therapeutic approaches.

REFERENCE

1. Aan het Rot M., Mathew S. J., Charney D. S. Neurobiological mechanisms in major depressive disorder. *CMAJ*. 2009;180(3):305–313.
2. Cai, C., Langfelder, P., Fuller, T. F., Oldham, M. C., Luo, R., van den Berg, L. H., ... & Witten, D. M. Isobaric labeling-based quantitative proteomic analysis of major depressive disorder brain samples. *Molecular psychiatry*, (2020). 25(7), 1513-1525.
3. Cai, X. et al. Multi-omics analysis of microRNA and mRNA changes in dorsal hippocampus of rats subjected to chronic unpredictable mild stress. *Scientific Reports* 2020; 10(1): 19767.
4. Dziedzic, M., & Ziolkowski, P. Epigenetic mechanisms in depression. *Advances in Clinical and Experimental Medicine*, (2020). 29(4), 597-607.
5. Gao, S., Chen, L., Lu, X., Zuo, L., & Mo, X. Integration of multi-omics data for identifying key genes and pathways associated with major depressive disorder. *BMC Medical Genomics*, (2021). 14(1), 1-16.
6. SGrant CW, Barreto EF, Kumar R, Kaddurah-Daouk R, Skime M, Mayes T, Carmody T, Biernacka J, Wang L, Weinshilboum R, Trivedi MH, Bobo WV, Croarkin PE, Athreya AP. Multi-Omics Characterization of Early- and Adult-Onset Major Depressive Disorder. *J Pers Med*. 2022 Mar 6;12(3):412. doi: 10.3390/jpm12030412. PMID: 35330412; PMCID: PMC8949112.
7. Hicks EM, Seah C, Cote A, Marchese S, Brennand KJ, Nestler EJ, Girgenti MJ, Huckins LM. Integrating genetics and transcriptomics to study major depressive disorder: a conceptual framework, bioinformatic approaches, and recent findings. *Transl Psychiatry*. 2023 Apr 19;13(1):129. doi: 10.1038/s41398-023-02412-7. PMID: 37076454; PMCID: PMC10115809.
8. Hu, Y., Chen, J., Zheng, J., Feng, W., Zeng, J., & Chen, X. Integrating multi-omics data to identify hub genes and pathways in major depressive disorder. *Journal of psychiatric research*,(2020).129, 200-207.
9. Huang Y, Shi X, Li Z, Shen Y, Shi X, Wang L, Li G, Yuan Y, Wang J, Zhang Y, Zhao L, Zhang M, Kang Y, Liang Y. Possible association of Firmicutes in the gut microbiota of patients with major depressive disorder. *Neuropsychiatr Dis Treat*.

- 2018 Dec 3;14:3329-3337. doi: 10.2147/NDT.S188340. PMID: 30584306; PMCID: PMC6284853.
10. Ji, J., Yang, J., Zhang, Y., Zhu, J., & Wang, X. Integrative analysis of proteomics and transcriptomics data in major depressive disorder. *Journal of Proteome Research*, 2021; 20(4), 2352-2362.
 11. Kapoor M, Chao MJ, Johnson EC, Novikova G, Lai D, Meyers JL, Schulman J, Nurnberger JI Jr, Porjesz B, Liu Y; Collaborative Study on the Genetics of Alcoholism (COGA); Foroud T, Edenberg HJ, Marcora E, Agrawal A, Goate A. Multi-omics integration analysis identifies novel genes for alcoholism with potential overlap with neurodegenerative diseases. *Nat Commun*. 2021 Aug 20;12(1):5071. doi: 10.1038/s41467-021-25392-y. PMID: 34417470; PMCID: PMC8379159.
 12. Kim, J. W., Lee, J. H., Kim, J. H., Lee, J. H., & Koo, J. W. Big data in major depressive disorder: a systematic review. *Psychiatry Investigation*, (2021);18(1), 1-8.
 13. Krishnan, V. et al. Multi-omics profiling of neurodegeneration in the TgCRND8 mouse model of Alzheimer's disease. *Neurobiology of Disease* 2018; 117: 212-228.
 14. Li, Y. et al. Multi-omics analysis of the molecular mechanisms underlying depression. *International Journal of Molecular Sciences* 2019; 20(18): 4598.
 15. Lin, Y. et al. Multi-omics analysis reveals dysregulated pathways and potential biomarkers in major depressive disorder. *Frontiers in Psychiatry* 2020; 11: 506.
 16. Lin, Z., Liu, H., Yang, L., Lv, X., Sun, S., Chen, W., & Li, L. Multiomics analysis reveals the potential mechanisms and biomarkers in major depressive disorder. *Frontiers in psychiatry*, (2020). 11, 576574.
 17. Liu, X. et al. Multi-omics analysis of gut microbiota and metabolites in rats with depression induced by chronic unpredictable mild stress. *Translational Psychiatry* 2021; 11(1): 413.
 18. Mavromatis LA, Rosoff DB, Cupertino RB, Garavan H, Mackey S, Lohoff FW. Association Between Brain Structure and Alcohol Use Behaviors in Adults: A Mendelian Randomization and Multiomics Study. *JAMA Psychiatry*. 2022 Sep

- 1;79(9):869-878. doi: 10.1001/jamapsychiatry.2022.2196. PMID: 35947372; PMCID: PMC9366661.
19. Merikangas, K. R., & McClure, E. B. . Epidemiology of major depressive disorder. *Handbook of depression*,(2012),5-20.
 20. Qi D, Chen K. Bioinformatics Analysis of Potential Biomarkers and Pathway Identification for Major Depressive Disorder. *Comput Math Methods Med*. 2021 Aug 3;2021:3036741. doi: 10.1155/2021/3036741. PMID: 34394704; PMCID: PMC8357473.
 21. Rizkallah, É., Gilbert, J. A., Hughes, T. B., & Guttmann, C. R. Integrating multi-omics data to investigate biological mechanisms in neurological disorders. *Nature Reviews Neurology*,(2020).16(11), 624-635.
 22. Sun, Y. et al. Multi-omics analysis of hippocampal neurogenesis in a rat model of depression. *Frontiers in Cellular Neuroscience* 2018; 12: 297.
 23. Tylee, D.S. et al. Multiomic profiling of brain-derived neurotrophic factor (BDNF) signaling in major depressive disorder (MDD). *EBioMedicine* 2019; 46: 488-501.
 24. Wang, Q., & Yu, H. Systems biology approaches to psychiatric disorders. *Advances in experimental medicine and biology*, (2019). 1158, 311-329.
 25. Wang, Q., Li, X., Li, Y., Li, S., Li, S., & Li, J. Integrated transcriptomics and metabolomics analysis of peripheral blood reveals dysregulated metabolic pathways in major depressive disorder. *Journal of Proteome Research*,2020; 19(8), 3187-3196.
 26. Yang, C. et al. Multi-omics profiling reveals dysregulated metabolic pathways and potential diagnostic biomarkers in major depressive disorder. *Frontiers in Psychiatry* 2021; 12: 648032.
 27. Yu, Y. et al. Multi-omics analysis of gut microbiota and metabolites in a mouse model of depression. *Frontiers in Cellular and Infection Microbiology* 2020; 10: 394.
 28. Zhang, R., Zhang, J., Fang, L., Li, X., & Zhao, L. Identification of potential pathways and biomarkers associated with major depressive disorder using transcriptome and metabolome data. *BMC psychiatry*, (2020).20(1), 1-16.