

GENE CO-EXPRESSION AND TRANSCRIPTIONAL REGULATORY NETWORKS IN AUTISM SPECTRUM DISORDER AND ITS IMPLICATIONS IN GENETIC AND ENVIRONMENTAL PREDISPOSITION TO THE DISORDER

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Science in Bioinformatics and Data Science

By

Harsiddhi. A (41738018)



**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 I Category-1 University by UGC I Approved by AICTE
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119

APRIL – 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A++" by NAAC I Category-1 University by UGC I 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 **HARSIDDHI. A (41738018)**, who carried out the final year project entitled "**GENE CO-EXPRESSION AND TRANSCRIPTIONAL REGULATORY NETWORKS IN AUTISM SPECTRUM DISORDER AND ITS IMPLICATIONS IN GENETIC AND ENVIRONMENTAL PREDISPOSITION TO THE DISORDER**" under our supervision from January 2024 to April 2024.

Dr. H. JEMMY CHRISTY

Internal Guide

Dr. H. JEMMY CHRISTY

Head of the Department

Submitted for Viva Voce Examination held on 27/04/24

Internal Examiner

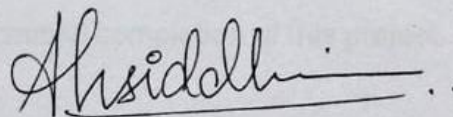
External Examiner

DECLARATION

I, **HARSIDDHI. A**, hereby declare that the Project Report entitled **GENE CO-EXPRESSION AND TRANSCRIPTIONAL REGULATORY NETWORKS IN AUTISM SPECTRUM DISORDER AND ITS IMPLICATIONS IN GENETIC AND ENVIRONMENTAL PREDISPOSITION TO THE DISORDER** done by me under the guidance of **Dr. H. JEMMY CHRISTY** 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 **Bachelor of Science in Bioinformatics and Data Science**.

DATE: 27/04/2024

PLACE: Chennai



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENTS

I am delighted to express my heartfelt acknowledgement to the **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement and support throughout the duration of my project. Their guidance has been invaluable in ensuring the successful completion of this project, and I am sincerely grateful for their contribution.

I convey my thanks to **Dr. Rekha Chakravarthy**, Dean, School of Science and Humanities for providing me necessary support and details at the right time during the progressive reviews.

I sincerely thank **Dr. H. Jemmy Christy**, Head of the Department, Dept. of Bioinformatics, for her unwavering support and timely provision of necessary details during the progressive reviews. Her support has been instrumental in helping me to overcome obstacles and achieve success in this project.

I am deeply indebted to my Project Guide, **Dr. H. Jemmy Christy**, for her invaluable guidance, suggestions, and constant encouragement throughout the project. Her expertise and insights have been instrumental in shaping my work and ensuring its successful completion.

I also wish to extend my gratitude to my family for their support and assistance in various ways, which have contributed to the successful completion of this project.

Harsiddhi. A



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A++" by NAAC ICategory-1 University by UGC I Approved by AICTE
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119
www.sathyabama.ac.in

CERTIFICATE

This is to certify that the dissertation entitled "GENE CO-EXPRESSION AND TRANSCRIPTIONAL REGULATORY NETWORKS IN AUTISM SPECTRUM DISORDER AND ITS IMPLICATIONS IN GENETIC AND ENVIRONMENTAL PREDISPOSITION TO THE DISORDER" was carried out by **Ms. Harsiddhi. A (Reg. No. 41738018)**, Department of Bioinformatics and Data Science, Sathyabama Institute of Science and Technology, Chennai in partial fulfillment of the requirements of the award of the **Degree of Bachelor of Science in Bioinformatics and Data Science 2021-2024**, is a bonafide record of the project work done under my guidance and supervision.

This thesis or part, therefore, has not been submitted elsewhere for any other degree or diploma.

Date: 27/04/24

Place: Chennai

Dr. H. JEMMY CHRISTY

ABSTRACT

Autism spectrum disorder (ASD) presents as a non-mendelian heterogeneous hereditary neuropsychiatric condition with a complex etiology involving a wide array of genes expressed in multiple converging pathways. Inconsistencies and discrepancies in gene expression data from diverse samples of ASD patients are observed. This study utilizes expression-network-analysis to explore pivotal genes and pathways in ASD, aiming to elucidate gene co-expression and transcriptional regulation mechanisms and its implications in genetic and environmental factors' predisposition to the disease. Autism Spectrum Disorder datasets were scrutinized to identify Differentially Expressed Genes (DEGs) sourced from the Gene Expression Omnibus (GEO) Series, including GSE143155, GSE18123, and GSE102741. 4587 common genes studied via Network construction and analysis using Cytoscape, and CytoHubba. PPI Network DEGs were subjected to Functional annotation and enrichment analysis. Our findings highlight GRIN2A, GRIK3, TERT, VCAM1, WNT7A, WNT7B, SRC, ERBB4, and LRP2 genes as critical markers in ASD risk. Genes such as GRIN2A and GRIK3 contribute to neurotransmitter receptor activity in ASD associated with NMDA receptors, while TERT alters excitatory/inhibitory (E/I) potential via telomerase activity. VCAM 1 facilitates leukocyte adhesion and trans endothelial migration and is involved in primary amine oxidase activity, and WNT7A and WNT7B are crucial for neural stem cell renewal, synapse formation, axon guidance, and neuronal progenitor cell progression and lens fibre cell development. Additionally, SRC participates in BMP receptor binding involved in the regulation of adhesion and migration in autistic lymphoblasts, ERBB4 is associated with synaptic communication deficits in hippocampal synaptic plasticity and contextual fear memory, and LRP2 contributes to neuronal development and regeneration through its effects on sonic hedgehog (Shh) signaling and ventricular compact myocardium morphogenesis. This diverse array of associated genes highlights the multifaceted nature of gene expression and its relevance to autism spectrum disorder (ASD) pathology. Differentially expressed genes identified in this study lay the groundwork for constructing and studying gene co-expression and transcriptional regulatory networks in ASD, facilitating potential intervention strategies upon validation.

KEYWORDS: *Autism Spectrum Disorder, Gene Expression-Network Analysis, Functional Annotation, Systems Biology, Gene Ontology.*

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO
	ABSTRACT	v
	LIST OF ABBREVIATIONS	ix
	LIST OF TABLES	x
	LIST OF FIGURES	xi
1	INTRODUCTION	1
	1.1 AUTISM SPECTRUM DISORDER	1
	1.2 CAUSATIVE FACTORS	2
	1.2.1 GENETIC FACTORS	
	1.2.2 ENVIRONMENTAL FACTORS	
	1.3 EPIDEMIOLOGY	5
	1.4 IMPORTANCE OF SYSTEMS BIOLOGY AND MICROARRAY DATA ANALYSIS IN ASD	6
	1.5 TOOLS USED FOR DATA ANALYSIS	7
	1.5.1 GENE EXPRESSION OMNIBUS	
	1.6 DIFFERENTIAL EXPRESSION ANALYSIS	7
	1.6.1 PARAMETERS USED FOR DEG SCREENING	
	1.7 STEPS IN DGE ANALYSIS	9
	1.7.1 ABOUT GEO2R	
	1.8 ABOUT CYTOSCAPE AND PPI NETWORK CONSTRUCTION	10
	1.9 FUNCTIONAL ENRICHMENT ANALYSIS BASED ON GO TERMS	11
2	REVIEW OF LITERATURE	12

3	AIM AND SCOPE	16
	3.1 AIM	
	3.2 OBJECTIVE	
	3.3 SCOPE OF THE WORK	
4	MATERIAL AND METHODS	18
	4.1 AUTISM SPECTRUM DISORDER	18
	EXPRESSION DATASETS	
	4.2 IDENTIFICATION OF	19
	DIFFERENTIALLY EXPRESSED GENES (DEG's)	
	4.3 PROTEIN-PROTEIN INTERACTION NETWORK	20
	CONSTRUCTION	
	4.4 NETWORK ASSESSMENT TO SCREEN HUB	21
	GENES	
	4.4.1 LOCAL BASED METHODS	21
	4.4.2 LOCAL BASED METHODS	22
	4.5 FUNCTIONAL ENRICHMENT AND GO ANALYSIS	24
	OF DEG's	
	4.5.1 ENRICHR	24
5	RESULTS AND DISCUSSION	26
	5.1 IDENTIFICATION OF DEG's	26
	5.2 PPI NETWORK CONSTRUCTION AND ANALYSIS	28
	5.3 NETWORK ASSESSMENT TO SCREEN HUB	29
	GENES	
	5.4 GENE ONTOLOGY AND FUNCTIONAL	40
	ENRICHMENT ANALYSIS OF DEGs	

5.5	BIOLOGICAL SIGNIFICANCE OF GENE MARKERS AND THEIR ROLE IN GENETIC AND ENVIRONMENTAL PREDISPOSITION TO ASD	46
5.5.1	GRIN2A (Glutamate Receptor, Ionotropic, N- methyl-d-aspartate, Subunit 2)	46
5.5.2	GRIK3 (Glutamate Ionotropic Receptor Kainate Type Subunit 3)	47
5.5.3	TERT (Telomerase Reverse Transcriptase)	47
5.5.4	VCAM 1 (Vascular Cell Adhesion Molecule 1)	48
5.5.5	WNT7A and WNT7B (Wingless-related Integration Site Subunit 7A & 7B)	49
5.5.6	SRC (Sarcoma)	50
5.5.7	ERBB4 (Erb-B2 Receptor Tyrosine Kinase 4)	51
5.5.8	LRP2 (Low Density Lipoprotein Receptor- related Protein 2)	52
6	SUMMARY AND CONCLUSION	54
	REFERENCES	56

LIST OF ABBREVIATIONS

ASD	Autism Spectrum Disorder
PDDNOS	Pervasive Developmental Disorder-Not Otherwise Specified
TD	Typical Development
CNV	Copy Number Variation
SNP	Single Nucleotide Polymorphism
PPI	Protein-Protein Interactions
GEO	Gene Expression Omnibus
GEP	Gene Expression Profile
DGEA	Differential Gene Expression Analysis
DEG	Differentially Expressed Genes
LogFC	Logarithmic Fold Change
FC	Fold Change
LIMMA	Linear Models for Microarray Analysis
FDR	False Discovery Rate
GO	Gene Ontology
KEGG	Kyoto Encyclopaedia of Genes and Genomes
MNC	Maximum Neighborhood Component
DMNC	Density of Maximum Neighborhood Component
MCC	Maximal Clique Centrality
EPC	Edge Percolated Component
GRIN2A	Glutamate Receptor, Ionotropic, N-methyl-d-aspartate, Subunit 2
GRIK3	Glutamate Ionotropic Receptor Kainate Type Subunit 3
TERT	Telomerase Reverse Transcriptase
VCAM1	Vascular Cell Adhesion Molecule 1
Wnt7A & Wnt7B	(Wingless-related Integration Site Subunit 7A & 7B)
SRC	Sarcoma
ERBB4	Erb-B2 Receptor Tyrosine Kinase 4
LDRP2	Low Density Lipoprotein Receptor-related Protein 2
GEA	Gene Expression Analysis
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
GWAS	Genome Wide Association Studies

LIST OF TABLES

TABLE NO	TITLE	PAGE NO
1.1	Parameters for DEG screening	10
5.1	Datasets considered for TD vs ASD DEG analysis	27
5.2	Criterion of Statistical Significance for Identifying the Significant Upregulated DEGs	27
5.3	Top 20 genes evaluated in the PPI network using Eleven Topological methods analysed using CytoHubba plugin in Cytoscape.	33
5.4	Comparison of gene from 12 topological methods	36
5.5	High Confidence Factors for ASD	39
5.6	Genes Associated With Biological Process	42
5.7	Genes Associated With Molecular Function	45
5.8	Environmental Factors affecting high risk genes	53

LIST OF FIGURES

FIGURE	TITLE	PAGE
NO		NO
1.1	Causative Factors of ASD	2
1.2	Environmental Factors influencing ASD	5
1.3	Maternal factors affecting ASD risk during gestation period	14
4.1	Proposed work plan	18
5.1	Neurons, fibroblasts, prefrontal cortex and peripheral blood cells-Common DEG's	28
5.2	Reactome-FI Based Network With 1,520 nodes and 2,425 gene-gene interactions edges	29
5.3	Network of Topological Methods- Degree	30
5.4	Network of Topological Methods- Bottleneck	30
5.5	Network of Topological Methods- Maximum Neighborhood Component (MNC)	31
5.6	Network of Topological Methods- Closeness	31
5.7	Network of Topological Methods- Betweenness	31
5.8	Network of Topological Methods- Edge Percolated Component (EPC)	31
5.9	Network of Topological Methods- Maximal Clique Centrality (MCC)	32
5.10	Network of Topological Methods- Radiality	32
5.11	Network of Topological Methods- Stress	32
5.12	Network of Topological Methods- EcCentrality	32
5.13	Network of Topological Methods- Density of Maximum Neighborhood Component (DMNC)	33
5.14	Network of Topological Methods- Clustering Coefficient	33
5.15	Functional Enrichment analysis based on Gene ontology – Biological Process	41

5.16	Cluster Dendrogram- Biological Process	41
5.17	Functional Enrichment Network –Biological Process	42
5.18	Functional Enrichment analysis based on Gene ontology – Molecular Function	43
5.19	Cluster Dendrogram- Molecular Function	44
5.20	Functional Enrichment Network –Molecular Function	44

CHAPTER-1

INTRODUCTION

1.1 AUTISM SPECTRUM DISORDER:

Autism Spectrum Disorder (ASD) is a non-mendelian heterogenous hereditary neuropsychiatric disorder with a complex neurological and developmental condition impacting individuals' social interactions, communication abilities, learning, and behaviors. While ASD can manifest at any age, it is primarily recognized as a developmental disorder due to the typical onset of symptoms within the first two years of infancy in humans (*Masini et al., 2020*).

ASD encompasses a spectrum of conditions, including autistic disorder, Asperger's syndrome, childhood disintegrative disorder, Rett's disorder, and pervasive developmental disorder not otherwise specified (PDD-NOS). Children with ASD often exhibit delays in social, language, and intellectual development, characterized by repetitive behaviors, fixated interests, and sensory sensitivities. They commonly struggle with social reciprocity and non-verbal communication (*AIIMS New Delhi, n.d.*)

Diagnosis of ASD according to DSM-5 criteria mandates persistent deficits in social communication and interaction across various contexts, along with restricted and repetitive patterns of behaviors or interests. These symptoms must significantly impair functioning and be present in the early developmental period. However, they may become more apparent as social demands increase or be masked by learned strategies later in life. It's crucial to note that the symptoms should not be solely explained by intellectual disability or global developmental delay. The American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) outlines these criteria, providing a framework for understanding and diagnosing ASD. (*American Psychiatric Association. Autism Spectrum Disorder. In: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, American Psychiatric Association, Arlington, VA 2013. p.50*)

1.2 CAUSATIVE FACTORS:

ASD involves a multi-factorial etiology characterized by a complex interplay of neurologic, immunologic, environmental, and genetic elements, lacks a singular causative factor (Ashwood *et al.*, 2006). Its strong familial occurrence underscores the significant influence of genetic predisposition on its manifestation. Despite this, only a subset of the presumed associated genes has been identified, leaving gaps in our understanding. Additionally, non-genetic factors, such as parental health conditions and prenatal and perinatal factors, are increasingly recognized for their potential contributions to ASD development. More recently, attention has turned to epigenetic modifications, including DNA methylation and microRNA alterations, as potential biomarkers associated with ASD (Masini *et al.*, 2020)

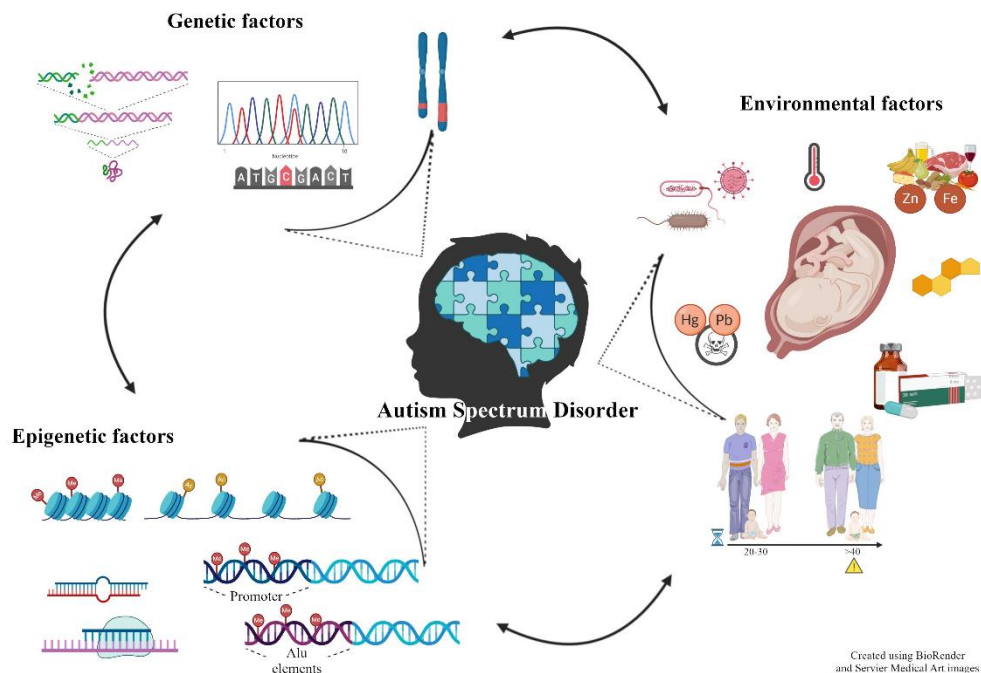


Figure 1.1. Causative factors of ASD (Masini *et al.*, 2020)

1.2.1 GENETIC FACTORS:

In 1977, Folstein and Rutter conducted twin studies, revealing a higher concordance rate in monozygotic twins compared to dizygotic twins, suggesting a genetic influence on autism spectrum disorder (ASD) (*Folstein & Rutter, 1977*). Subsequent karyotype studies identified chromosomal abnormalities, guiding investigations into relevant genomic regions. Genome-wide association studies (GWAS) have identified several genetic risk loci for ASD, although their specific target genes in relevant tissues and cell types remain to be elucidated. Numerous large-scale genetic studies on ASD have led to the identification of hundreds of risk genes, primarily categorized into two classes: those involved in synapse formation and those associated with transcriptional regulation and chromatin remodeling pathways. Additional susceptibility loci affect protein transcription through various mechanisms, including increased de novo mutation load in regulatory elements of ASD risk genes. Somatic mosaicism, resulting from post-zygotic DNA mutations, has emerged as a significant contributor to neurodevelopmental disorders, including ASD, Rett syndrome, and schizophrenia. Pathogenic somatic variations have been linked to ASD susceptibility, highlighting the importance of genetic heterogeneity in ASD etiology. Furthermore, genes with epigenetic-modulating functions play a substantial role in ASD susceptibility, with studies estimating that approximately 19.5% of candidate genes are epigenetic regulators (*Rylaarsdam et al., 2019*).

Copy number variations (CNVs) are submicroscopic structural variants implicated in ASD susceptibility, with rare, genic CNVs showing a higher prevalence in autistic individuals. These findings underscore the genetic complexity underlying ASD and the potential for diverse disease phenotypes from a limited number of pathogenic variants (*Duffney et al., 2018*). Additionally, risk genes with high penetrance are often involved in gene expression modulation or protein-protein interaction networks crucial for central nervous system developmental patterning.

1.2.2 ENVIRONMENTAL FACTORS:

Autism susceptibility is predominantly attributed to genetic factors, estimated at 40–80%, with environmental influences likely acting via epigenetic mechanisms. Proposed environmental contributors include advanced paternal age, maternal complications during pregnancy, prenatal exposure to various substances, and perinatal factors such as umbilical cord complications, maternal haemorrhage and neonatal trauma. Additionally, fetal exposure to sex steroids, maternal malnutrition, and trace element imbalances are implicated. Notably, altered metal levels can affect synaptic function, with zinc deficiency linked to neuronal defects and ASD-related symptoms. Maternal obesity and stress association with ASD risk shown to prevalently affect oxidative stress response of mitochondria in fetal microenvironment and infancy. Maternal substance consumption during pregnancy, including medications, alcohol, and tobacco may pose risks, as do viral and bacterial infections and exposure to toxic substances like flame retardants and heavy metals. Expecting Mothers with a history of Autoimmune disorder are also more likely to have ASD-affected children. Environmental sanitation during pregnancy is crucial for fetal development, as adverse exposures during early pregnancy can impact organogenesis and fetal growth (*Masini et al., 2020*)

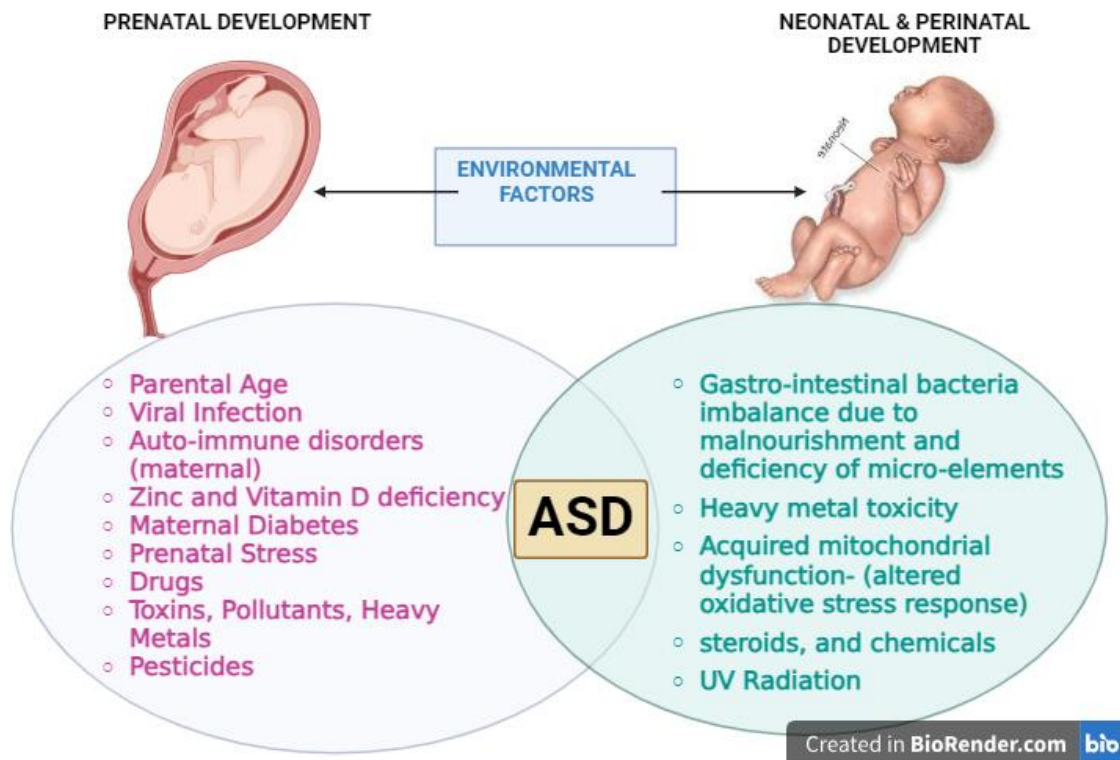


Figure 1.2. Environmental factors influencing ASD

1.3 EPIDEMIOLOGY:

Approximately 1 in 36 children are diagnosed with autism spectrum disorder (ASD) based on estimates from the CDC's Autism and Developmental Disabilities Monitoring (ADDM) Network. ASD is found across all racial, ethnic, and socioeconomic groups, with boys being nearly four times more affected than girls. A study spanning 2009-2017 reported that around 17% of children aged 3–17 years were diagnosed with a developmental disability, including autism, attention-deficit/hyperactivity disorder, blindness, and cerebral palsy (*Autism*, n.d.).

In India, research published in the Indian Journal of Pediatrics in 2021 estimated the prevalence of autism to be approximately 1 in 68 children, with boys being more commonly affected than girls at a ratio of about 3:1. A community-based study conducted among 1–10-year-olds in rural, urban, and tribal areas revealed a prevalence of 0.15% (*Raina S et al., 2017*). Studies worldwide, spanning Asia, Europe, and North America, have identified ASD prevalence ranging from 1% to

2% (*Salari N et al., 2022*).

Recent research indicates a rapid increase in ASD prevalence globally, reaching 1 in 132 children. Moreover, there's notable diversity in ASD prevalence across different regions, with rates ranging from 1.4 per 10,000 children in the Arabian Peninsula to 185 per 10,000 children in Asia. Sweden holds the highest prevalence in Europe at 115 per 10,000 children (*Wang J et al., 2022*).

1.4 IMPORTANCE OF SYSTEMS BIOLOGY AND MICROARRAY DATA ANALYSIS IN ASD:

The analysis of microarray data and the adoption of a systems biology approach play crucial roles in understanding autism spectrum disorder (ASD) and related conditions such as intellectual disability (ID), developmental delay (DD), and multiple congenital anomalies (MCAs). Microarray analysis, recommended as a first-line diagnostic tool by the International Standards for Cytogenomic Arrays (ISCA), has shown a diagnostic yield of 10%-15% in these patient populations. However, many copy number variants (CNVs) identified through microarray analysis are classified as of uncertain clinical significance, highlighting the need for standardized databases and increased data sharing to improve diagnostic accuracy (*Ayaz et al., 2022*).

Despite efforts to understand the genetic basis of ASD, the "missing heritability issue" remains a challenge, with only a small fraction of ASD cases being associated with known genetic variants. Whole-exome sequencing (WES) and microarray technologies have enabled genome-wide studies, revealing a highly heterogeneous genetic architecture characterized by both CNVs and single nucleotide variations (SNVs). While CNVs contribute to about 15% and SNVs to 7% of ASD cases, the majority of ASD cases (>75%) lack a clear genetic explanation. Genome-wide association studies (GWAS) have identified several common genetic variants associated with ASD, although their individual effects are relatively small. WES analysis has proven to be a powerful tool for identifying over 150 new candidate genes for ASD, offering insights into sporadic cases and detecting mutations with incomplete penetrance (*Masini et al., 2020*).

1.5 TOOLS USED FOR DATA ANALYSIS:

The emergence of Bioinformatics revolutionized the management of large-scale biological data generated by genome-wide studies and experiments with high-capacity outputs. This necessitated the establishment of permanent repositories, easily accessible to all. Open-access platforms like GEO play a pivotal role in storing raw and sequence annotated files from microarray and gene sequencing experiments aiding this study.

1.5.1 GENE EXPRESSION OMNIBUS

GEO (<http://www.ncbi.nlm.nih.gov/geo/>) serves as a public repository for high-throughput data, facilitating effective exploration, querying, and visualization of gene expression data. Its architecture allows for versatile design, accommodating heterogeneous data from both gene-centric and study-centric perspectives (*Barrett et al., 2007*). Unlike cancer-specific databases, GEO encompasses profiles from all studies utilizing microarray technologies. Additionally, it includes data obtained from non-array-based techniques such as serial analysis of gene expression (SAGE) (*Barrett & Edgar, 2006*).

1.6 DIFFERENTIAL EXPRESSION ANALYSIS

Differential gene expression (DGE) refers to genes exhibiting varying expression levels between conditions- up-regulated and down-regulated fractions in contrasting conditions. For example, diseased vs. healthy individuals, different tissues, or developmental stages usually classified as test vs control groups. It involves quantifying genes whose expression varies according to phenotype or experimental conditions, sometimes accounting for inter-gene dependencies. This necessitates statistical analysis of large gene sets to detect quantitative expression changes between experimental groups. Microarrays are an integral part of conducting differential gene expression analyses (*McDermaid et al., 2019*).

1.6.1 Parameters used for DEG screening

P-value

The p-value represents the likelihood of rejecting the null hypothesis (H_0) when it is true. In this study, the null hypothesis assumed no differences between control and test samples for a specific gene. Therefore, if the p-value is less than the significance level (<0.05), we reject the null hypothesis, indicating acceptance of a difference between normal and diseased samples for that gene (*About GEO2R - GEO - NCBI*, n.d.). The statistical significance of p-value for DGEs is estimated as Upregulated for a p-value <0.05 and Downregulated for a p-value >0.05 . (Andrade, 2019)

Multiple Testing Problem (False Discovery Rate)

Setting a p-value threshold of 0.05 indicates a 5% chance of obtaining a false positive result. While this level of significance may be acceptable for a single test, conducting multiple tests increases the likelihood of encountering false positives, known as the multiple testing problem. To address this issue, the Benjamini-Hochberg method (Benjamini & Hochberg, 1995) is employed to control the False Discovery Rate (FDR). Accepting a cut-off of 0.05 for the adjusted p-value (q-value) ensures that approximately 5% of differentially expressed cases may be false positives in reality (Jeggari, 2013).

Fold change

Fold change estimates are key indicators of biological significance frequently employed in microarray gene expression analysis. They quantify the change in mRNA/gene expression level between experimental conditions, typically from normal (x) to diseased (y) states, and are expressed as a logarithmic base 2 number, representing the fold increase or decrease in expression. (*About GEO2R - GEO - NCBI*, n.d.). The statistical significance of Fold Change for DGEs is estimated as Upregulated for a FC-value >2 and Downregulated for a FC-value <2 . (Dalman et al., 2012)

1.7 STEPS IN DGE ANALYSIS

Differential expression analysis involves statistically analyzing normalized read count data to detect quantitative changes in expression levels among experimental groups. GEO2R, an interactive web tool, enables users to compare multiple groups of samples within a GEO Series, identifying genes exhibiting differential expression across experimental conditions. In this study, data obtained from the GEO accession viewer was analyzed using the GEO2R tool. Utilizing GEO2R, the obtained GSM files were categorized into distinct groups, facilitating comprehensive analysis.

1.7.1 ABOUT GEO2R:

Bioconductor, an open-source project built on the R programming language, offers tools for high-throughput genomic data analysis. GEO2R, utilizing the GEOquery and limma R packages from Bioconductor, compares the original submitter's processed data tables from the GEO database. GEOquery parses GEO data into R data structures, while limma, a widely used statistical test (*Smyth, 2005*), identifies differentially expressed genes, applying multiple-testing corrections to P-values to mitigate false positives. This user-friendly interface enables R statistical analysis without command line expertise (*Davis & Meltzer, 2007*).

P-value adjustments in GEO2R are essential for correcting false positives. By default, the Benjamini & Hochberg false discovery rate method is applied, offering a balance between identifying significant genes and minimizing false positives (*Benjamini & Yekutieli, 2001*). GEO2R automatically applies log₂ transformation to data values, as Limma expects data in log space. This auto-detect feature ensures compatibility with GEO database standards (*Smyth, 2005*).

Table 1.1 Parameters for DEG screening (*About GEO2R - GEO - NCBI*, n.d.)

Parameter	Definition of the parameter
adj.P.Val	• P-value after adjustment for multiple testing. • Genes with the smallest adj.P.Val are considered more reliable indicators of differential expression.
P.Value	• Raw P-value representing the significance of gene expression changes.
t	• Moderated t-statistic, applicable when comparing two groups of Samples
B	• B-statistic or log-odds indicating the likelihood of differential gene expression, relevant in two-group comparisons.
logFC	• Log2-fold change between two experimental conditions, applicable in two-group comparisons.
F	• Moderated F-statistic, amalgamating t-statistics from pairwise comparisons into an overall test of significance for genes, relevant in analyses involving more than two groups of Samples.

1.8 ABOUT CYTOSCAPE and PPI NETWORK CONSTRUCTION

Cytoscape serves as an open platform, offering a publicly available network visualization and analysis tool complemented by various plugins to enhance both visualization options and network analysis capabilities. Numerous Cytoscape plugins enable scoring and ranking of nodes based on network features. Through Cytoscape, accessing a graphical representation of a network is simple, allowing for the integration of multiple layers of information, including large-scale genome-wide experiments and protein function annotations, onto the interactome (*Shannon et al., 2003*).

Integration of seven publicly available protein-protein interaction (PPI) databases—BioGRID, REACTOME, DIP, HPRD, IntAct, MINT, and SPIKE—offers a robust

approach to construct an integrated PPI network. Leveraging PPI databases facilitates the exploration of disease-related genes/proteins, enhancing our understanding of disease mechanisms. PPI data play a pivotal role in discovering new disease biomarkers, investigating disease-disease relationships, and detecting common biological functions. In the PPI network, proteins serve as nodes or vertices, with interactions between them represented as edges. These interactions delineate the biological connections between nodes and their systemic impact on functional gene and transcriptional regulatory pathways.

Specific statistic and mathematical measurements are employed to assess the significance of node measures and analyse their network features to infer their importance in the network, and estimate centrality of certain proteins between other nodes. These key regulatory genes could be termed as biomarkers.

1.9 FUNCTIONAL ENRICHMENT ANALYSIS BASED ON GO TERMS

Functional enrichment analysis is a computational technique that compares an input gene set to annotated gene sets, aiming to infer biological insights. It utilizes GO terms and pathway enrichment analysis to identify underlying biological processes and mechanisms by leveraging existing knowledge about the organism. GO terms encompass genes associated with specific biological processes, locations, diseases, or pathways (Barry et al., 2005).

Gene Ontology (GO) organizes gene functions and attributes into a directed acyclic graph (DAG), aiding interpretation of biological experiments. GO comprises three domains: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). BP encompasses molecular events, MF defines gene functions, and CC describes gene localizations within cells. Functional Similarity (FS) between genes, inferred from GO terms, is pivotal in detecting genetic interactions, pathways, and prioritizing disease genes. Functional Enrichment Analysis help infer expression of proteins in biological regulatory pathways using network enrichment analysis of differentially expressed genes (*Subramanian et al., 2005*).

CHAPTER-2

REVIEW OF LITERATURE

Autism spectrum disorder (ASD) involves a huge array of genes expressed through multiple converging pathways. Several genetic risk loci have been identified for ASD from genome-wide association studies (GWAS); however, their target genes in relevant tissues and cell types remain to be investigated. Through a literature survey, it was observed that there are inconsistencies and discrepancies in gene expression data from diverse samples of ASD patients. The Research Problem of here is to figure methods to enhance understanding of gene expression patterns and regulatory mechanisms in ASD. This study utilizes expression-network-analysis to explore pivotal genes and pathways in ASD, aiming to elucidate gene co-expression and molecular mechanisms underlying ASD pathogenesis.

Over the past decade there has been a global increase in the occurrence of autism spectrum disorder. At present, Autism has been designated as emergent property resulting from interaction between multigenic and polychromosomal factors including de novo mutations, methylome alterations, Copy Number Variations (CNV), Short Nucleotide Polymorphism (SNP) etc (*Jones et al., 2002*). An effect of this is the unpredictability and lack of specification of risk genes via standard genetic diagnostic techniques for disorders in genome transcriptome and associated variations for Autism in contrast to the affiliate syndromes of Retts, Fragile X, etc. (*Y. Feng et al., 1995*)(*Van D et al., 2001*)

Large scale Sequencing techniques and the development of Genome-Wide Analysis, Whole Exome Sequencing, and Whole Genome Sequencing have been revolutionary in aiding the assessment of multiple variations at once via screening of entire microarrays for variations in gene expression, epigenetic factors, regulatory mechanisms, and transcriptomic profiling between typical developing children and autistic patients (*Schmunk & Gargus, 2013*). Differences in diagnostic results may be attributed to the heterogeneity of the etiology as well as the varying degrees of

phenotypic expression among patients on the spectrum. This primes focus on the measurable qualities as defined in the DSM5 for classification of phenotypic symptoms as part of ASD, as well as quantifiable variations in the neuronal constitution, cortical, myocardial, and nephritic morphogenesis. (AIIMS New Delhi, n.d. rtd.2023).

Attempts to classify patient test subjects to create homogenous subgroups for efficient analysis on the basis of observable characteristics such as repression of speech, intellectual deficits, etc., have been employed in studies such as those conducted by Gregg et al., (Gregg et al., 2008) for the analysis of gene signature expression in blood lymphocytes from patients with early onset autism. Results inferred genes associated with immune-histamine concentrations and natural killer cells retention in circulating blood leading to toxicity. Classification of test subjects with Superior intellectual performance and exceptional traits beyond the average ability of the normal population or general limitations of an individual referred to as savant abilities inferred susceptibility of ASD with chromosomal linkage in 15q11-q13 (Nurmi et al., 2003). While classification based on linguistic skills of patient subsets in Genetic linkage detection revealed gene factors on 7q-13q chromosomes (Bradford et al., 2001). There are also stratification of test groups based on severity of specific phenotypes. However, the reproducibility of these results are ambiguous due to heterogeneity of diverse patients samples on the spectrum.

Hu and Steinberg (Hu V.W. et al., 2009) employed innovative multiple clustering algorithms to categorize phenotypic groups for gene expression profiling. They identified four ASD subgroups based on the severity profiles of their test subjects: individuals with severe language impairment, those with moderate severity and high savant skills, individuals with mild phenotypes and associated syndromes such as Asperger's or PDD-NOS, and individuals with intermediate phenotypes. The Gene expression analyses in lymphoblastoid cell lines (LCL) from different phenotypic clusters revealed differential expression of circadian rhythm genes pertinent as sleep disturbances are prevalent in ASD, especially among individuals with significant language deficits and epilepsy.

Other GES experiments involving LCL or primary lymphocytes, have highlighted the significance of natural killer (NK) activity in autism (*Baron et al., 2006*) and identified SEMA5A, a gene involved in axon guidance (*Melin et al., 2006*). A research on idiopathic autism has revealed differential expression of genes related to neuronal morphogenesis, steroidal hormone synthesis, and immune-modulatory response regulation in test-control sibling pairs (Hu et al., 2009), inferring dysregulation of hormone biosynthesis may contribute to systemic issues associated with autism, such as disruption of sleep, gastro-intestinal dysregulation and immune related disorders.

Significant studies have been reviewed to understand Maternal and prenatal conditions (maternal age, obesity, dietary nutrients, etc.) influence ASD expression through epigenetic mechanisms such as DNA methylation or histone acetylation. The impact of these factors on foetal development significantly increases risk of ASD (*Balachandar et al., 2022*).

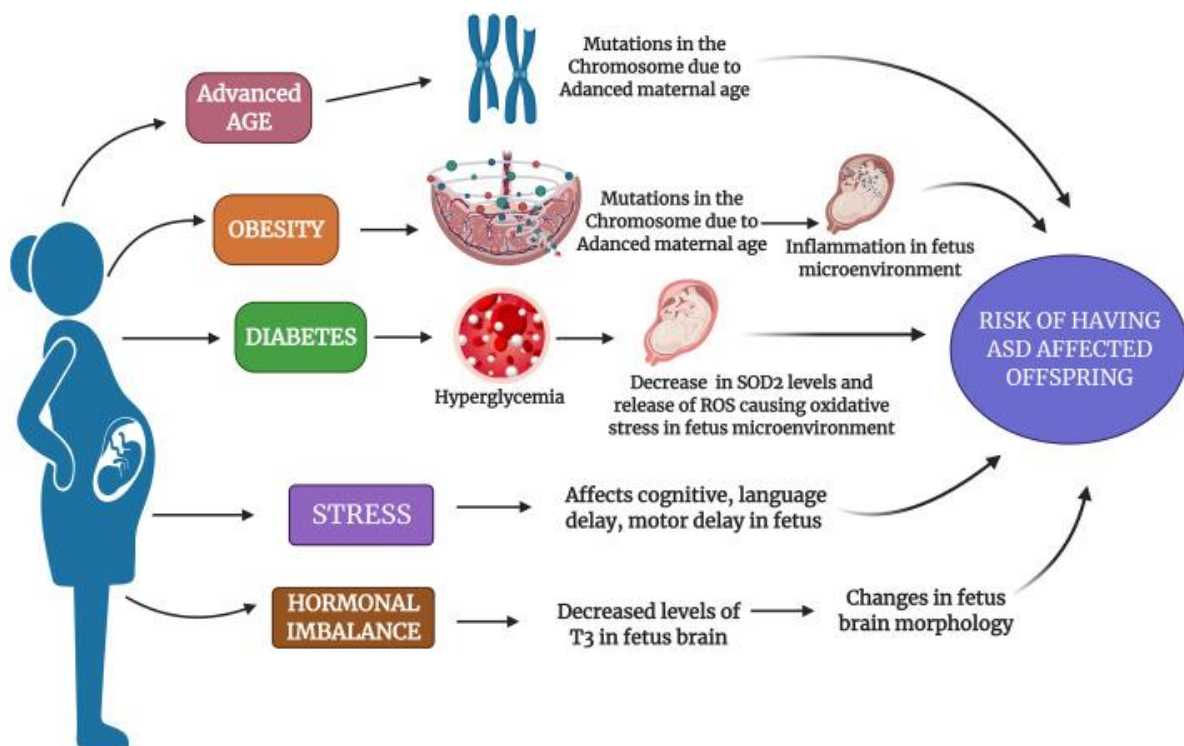


Figure 1.3- Maternal factors affecting ASD risk during gestation period
(*Balachandar et al., 2022*)

The least investigated aspect of autism involves environmental triggers, encompassing external factors like drugs, toxins, or pesticides, as well as intrinsic biological factors such as hormones or inflammatory mediators that a fetus may

encounter in-utero foetal microenvironment. These factors can interact with susceptibility genes, potentially influencing gene expression and phenotype through genetic or epigenetic mechanisms, or both.

Widespread variations in genetic markers detected from the previous experiments underscore the lack of reproducibility of these results as an effective applicant for diagnostics and treatment strategies of autism. In the prevalent evidence from twin and family studies demonstrating high heritability, however specific genes markers are unspecifiable due to discrepancies and high variability both in disease expression as well as the heterogeneity of experimental conditions and systemics of controls and test groups of healthy and diagnosed individuals as evidenced by the variation in co-expression results of microarray studies employed in different tissue samples of autism positive test subjects against controls (*Kong et al., 2012*)(*Wright et al., 2017*)(*Bai et al., 2022*)(*GEO Accession Viewer, n.d.*). This prompts a need to promote comprehensive understanding of autism at a systems biology level requires integrating information on genetic, environmental, and biological factors contributing to ASD. The ultimate aim is to predict how perturbations in these factors may lead to specific ASD phenotypes and how therapeutic interventions could potentially reverse or mitigate identified deficits.

CHAPTER-3

AIM AND SCOPE

This study aims to implement Network Biology and Systems Biology to gain an avid understanding of common regulatory pathways of diverse molecular mechanisms involved in the expression of Autism Spectrum Disorder in affected individuals.

3.1 OBJECTIVE

- Literature survey on Autism Spectrum Disorder (ASD), its genetic and environmental risk factors and its connection to phenotypic expression of symptoms associated with the disorder.
- Retrieval of diverse tissue samples' microarray experiments with test vs control groups from datasets in GEO and to construct gene expression network based on publicly available mRNA datasets from GEO.
- Examine Differentially Expressed Genes (DEGs) expression in ASD and typically developing patients via assessment of P-value and logFC value. Identification of statistically significant upregulated gene markers.
- Construction of Protein-Protein Interaction (PPI) network based on default metrics of Graph theory and distance measures.
- Hub Gene assessment based on eleven node ranking methods (Degree, Clustering Coefficient, Stress, Betweenness, Radiality, Closeness, Eccentricity, Bottleneck, MNC, DMNC, EPC and MCC) to evaluate the significance of gene nodes within the framework of the biological network.
- Functional Enrichment Analysis and Gene Annotation- Using GO terms, KEGG PATHWAYS and Enrichr to identify overlapping candidate genes filtered from hub genes list based on frequency of occurrence.
- To understand how these genetic variants lead to the ASD phenotype, focused on identifying points of convergence in the targets and functions of the ASD risk genes.

3.2 SCOPE OF THE WORK

Investigating the molecular mechanisms underlying ASD pathogenesis at the genetic and transcriptional levels- The differentially expressed genes (DEGs) and hub genes highlighted from the gene-expression network in this study offer valuable insights into the mechanisms involved in the onset and progression of ASD and its associated symptoms. Our research findings underscore the significance of GRIN2A, GRIK3, TERT, VCAM1, WNT7A, WNT7B, SRC, ERBB4, and LRP2 genes as indispensable indicators in ASD susceptibility. These genes hold promise as potential targets for diagnostic modalities and therapeutic interventions.

CHAPTER-4

MATERIAL AND METHODS

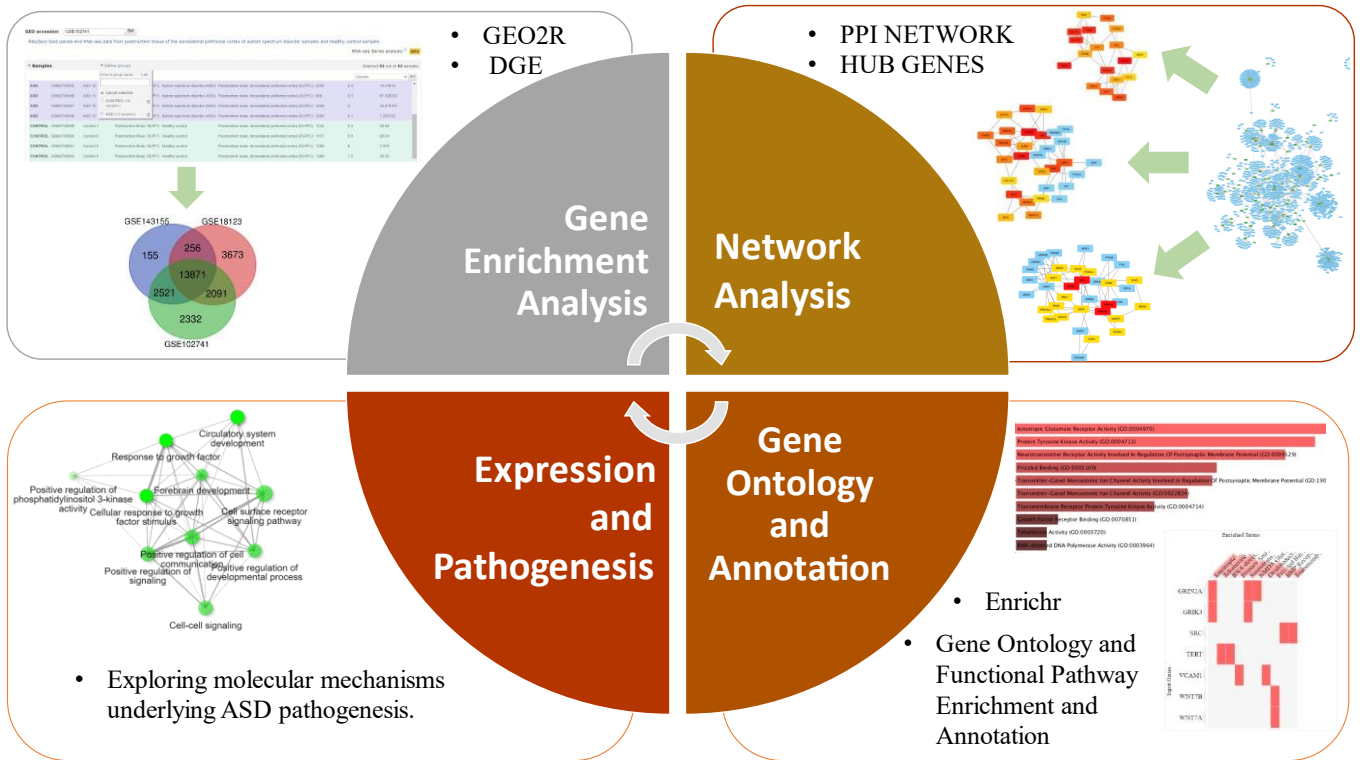


Fig 4.1 Proposed work plan

4.1 AUTISM SPECTRUM DISORDER EXPRESSION DATASETS

Datasets from Gene Expression Omnibus (GEO) for control and test samples, including experimental groups related to ASD. Autism Spectrum Disorder datasets were scrutinized to identify Differentially Expressed Genes (DEGs) sourced from the Gene Expression Omnibus (GEO) Series, including GSE143155, GSE18123, and GSE102741. GSE143155 contains microarray data on “Poor Mitochondrial Responses to Oxidative Stress in a Subset of Autism Spectral Disorder Patients” involving tissue samples of [neurons, fibroblasts] submitted to GEO on January 07, 2020 by *Baldwin WS et al.*, (GEO Accession Viewer, n.d.). The GEP Analysis was performed on GPL21697 NextSeq 550 (Homo sapiens) platform with a sample size of 30- 10 controls for ASD negative and 20 test (ASD positive). GSE140702 involves gene data sequenced to analyze “RiboZero Gold paired-end RNA-seq data from postmortem tissue of the dorsolateral prefrontal cortex of autism spectrum disorder

samples and healthy control samples” involving tissues extracted from [prefrontal cortex, neurons] submitted on August 16, 2017 by Wright C et al., (*Wright C et al., 2017*). The GEP Analysis was performed on GPL11154 Illumina HiSeq 2000 (Homo sapiens) platform with a total sample size of 52- 39 healthy controls and 13 ASD affected. GSE18123 contains microarray data on “Blood gene expression signatures distinguish autism spectrum disorders from controls” assessing samples of [peripheral blood cells], made public on GEO by Kong S et al., (*Kong S et al., 2012*) on August 01, 2012. GPE performed on platforms GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array and GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] for a sample size of 285 of which 33 healthy controls (typical development) and 31 samples for ASD affected individuals were screened for this study (Table 5.1).

4.2. IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES (DEG's)

Differentially Gene Expression Analysis between healthy control and test samples of the various tissues was performed using GEO2R. GEO2R facilitates comparing sample groups within a GEO Series to identify differentially expressed genes. Utilizing R packages from Bioconductor, it presents results in tables ranked by P-value and graphical plots, aiding in data interpretation. The package employed depends on type of data employed for analysis. GEO2R employs DESeq2 for RNA-seq data analysis. DESeq2, leveraging negative binomial models, ensures consistent performance across diverse data types, suitable for both small and large studies. For microarray data, GEO2R employs GEOquery for parsing GEO data into R structures and limma for differential expression analysis. Limma, a statistical test, accommodates various experimental designs and data types, applying P-value corrections to mitigate false positives.

One of the advantages of GEO2R is in its quantitative algorithm design for statistical pointers and mathematical analysis. GEO2R accessing Series Matrix files directly, it bypasses curated DataSets, broadening its applicability but necessitating awareness of limitations. This tool empowers researchers with versatile capabilities for gene expression exploration across experimental conditions and data types.

A comprehensive analysis was conducted utilizing t-test and Fold Change (FC) to detect Differentially Expressed Genes (DEGs) between Typical Development (TD) vs

ASD samples. The Benjamini and Hochberg method was applied to compute adjusted P-values (False Discovery Rate [FDR]), and the criteria of $\log FC > 2$ and adjusted P-value < 0.05 were considered statistically significant for identifying upregulated DEGs for each of the GEO Series.

Subsequently, the extracted DEGs were subject to pre-processing removing errors and gas in data along with anomalous, unidentified Gene IDs, and array markers along with preferential sorting of gene lists based on Prominence of LogFC value. The resultant lists were comparatively analyzed through venn analysis for common genes (Figure 5.1)

4.3. PROTEIN-PROTEIN INTERACTION NETWORK CONSTRUCTION

Protein-protein interactions (PPIs) play a crucial role in coordinating cellular processes, including signal transduction pathways and transcriptional regulation. These interactions form the foundation of various cellular events and can be represented as a protein-protein interaction network, aiding in disease analysis and drug discovery. In this network, proteins serve as nodes, while interactions between them are depicted as edges. (*Raman K, 2010*) The Proteomics Standard Initiative Common QUery InterfaCe (PSICQUIC) facilitates computational access to molecular interaction data from multiple sources worldwide. With the PSICQUIC interface, users can simultaneously search numerous molecular interaction databases, such as BioGRID, REACTOME, DIP, HPRD, IntAct, MINT, SPIKE, and iRefIndex, offering access to over 150 million binary interactions using a single query in the Molecular Interactions Query Language (MIQL). (*Del-Toro, N., et al., 2013*)

The Reactome FI network implemented to construct PPI-network model integrates curated pathways from Reactome, KEGG PATHWAY, PANTHER pathways, and gene-gene interactions from diverse sources via machine learning. The 2018 version, entails enhanced the Naive Bayesian Classifier training, leading to a higher recall rate providing increased SwissProt coverage expanded from 12,441 to 13,469 proteins (8.3% increase), and functional interactions among genes from 241,338 to 262,321 (8.7% increase) Via boosting predicted FIs per query. Covering 60% of human protein-coding genes, the Reactome Functional Interaction (FI) network combines manually curated pathways with predicted interactions via machine learning. The Cytoscape app, "ReactomeFIViz," leverages this reliable FI network for network-based data

visualization and analysis. (*Intnl:OpenAW- FIViz, 2018*). The Protein-Protein Interaction (PPI) network elucidates the mechanism of action resulting from disruptions in proteins and associated targets. Essential targets and associated network nodes analyzed for significance.

4.4 NETWORK ASSESSMENT TO SCREEN HUB GENES

The centrality-lethality rule posits that deletion of highly connected protein nodes (hubs) in protein-protein interaction (PPI) networks is more likely lethal, indicating their crucial role in network organization and biological significance. Hub nodes, with high degree, are pivotal in molecular mechanisms like protein interactions, gene regulation by transcription factors, or multi-regulation by elements like microRNAs. In human cells, hub genes often denote essential rather than disease genes, suggesting their potential as drug targets for diseases like cancer. (*He X, Zhang J, et al., 2006*)

To assess hub genes, the CytoHubba Cytoscape plugin offers a user-friendly interface for exploring biological networks and ranking nodes by network features. It provides 11 topological analysis methods, categorized into Local and Global Based Methods. Local methods [Degree, Maximum Neighborhood Component (MNC), Density of Maximum Neighborhood Component (DMNC), Maximal Clique Centrality (MCC)] evaluate a node's relationship with its direct neighbors, whereas global methods [Bottleneck, Eccentricity, Closeness, Radiality, Betweenness, Stress, Clustering Coefficient, and Edge Percolated Component (EPC)] consider its relationship with the entire network when calculating node scores. This tool aids in searching and exploring networks, facilitating the extraction of user-defined subnetworks for further analysis.

(*Chia-Hao Chin et al., 2014*).

4.4.1 LOCAL BASED METHODS

Degree (Deg)

Characterized by the number of edges connected with the node. It is denoted as 'Degree' of node. Higher the Degree, increase in likelihood of a node being a potential Hub Gene. (*Chia-Hao Chin et al., 2014*).

Maximal Clique Centrality (MCC)

Increased specificity in detection of highly-ranked feature nodes via the integration of maximal cliques thereby filtering essential hubs. (*Chia-Hao Chin et al., 2014*).

Maximum Neighborhood Component (MNC)

Hubs with maximum number of connected nodes filtered from the degree subgraph of the main PPI network. Aid in recognition of unrecognized hubs. (*Chia-Hao Chin et al., 2014*).

Density of Maximum Neighborhood Component (DMNC)

Based on MNC, measures density of connections among nodes by estimating the ratio of edge sets with respect to node collections. (*Chia-Hao Chin et al., 2014*).

4.4.2 GLOBAL-BASED METHODS

CytoHubba involves six methods where nodes are ranked based on shortest paths and two methods based percolated connectivity.

Closeness (Clo)

Closeness measures the inverse of the distance- the shortest path between two nodes to estimate centrality of a node. (*Chia-Hao Chin et al., 2014*). Comparing values to the average closeness value of a network enhances their significance. A high closeness value for a node suggests proximity to all other nodes, while a low value indicates greater distance from other nodes (*Freeman et al., 1997*).

EcCentricity (EC)

EcCentrality defined by the inverse of the maximum distance between the component containing node *v* and another node that is the farthest by distance from *v*. (*Chia-Hao Chin et al., 2014*).

Radiality (Rad)

The radiality is measured by the ratio of maximum distance between any two vertices of a component with the maximum distance between two nodes farthest from each other in the overall network. (*Chia-Hao Chin et al., 2014*).

The radiality score of a node signifies the shortest path between the node v and all other nodes in the network. A node with high eccentricity + high closeness+ high radiality is an indicator of a high central position in the network (*Valente TW, Foreman RK, 1998*).

Betweenness (BC)

Betweenness is measured by the integral ratio of the shortest paths connecting two with degree. It is defined as the extent to which a node participates in the shortest paths connecting other nodes (*Freeman L, 1977*).

BottleNeck (BN)

Nodes possessing high betweenness are commonly referred to as 'bottlenecks', strategically positioned at crucial intersections between hubs, serving as bridges facilitating intergroup communication among nodes. Their elimination could result in the complete loss of connectivity between node clusters, profoundly impacting the network's overall topology and consequently, information dissemination. This can be translated into the molecular pathways where proteins exemplify bottleneck nodes, whose functional loss can deactivate specific processes (*Freeman L, 1977*).

Stress (Str)

Stress of a node measure of the total number of shortest paths between two connecting nodes that pass through an interlinking node v . (*Chia-Hao Chin et al., 2014*).

Higher the stress, higher the relevance of the node connecting other nodes is higher. In biological terms, the stress of a node indicates its relevance in holding together communicating nodes. (*Hage P, Harary F, 1995*).

Edge Percolated Component (EPC)

The Edge Percolated Component relies on a user-defined threshold ($0 \leq \text{threshold} \leq 1$). It generates 1000 reduced networks by assigning a random number within the range of 0 to 1 to each edge on the basis of which, edges are eliminated if their corresponding random numbers fall below the specified threshold (*Chia-Hao Chin et al., 2014*).

Clustering Co-efficient

The clustering coefficient is a measure of likelihood of clustering among two nodes nodes in a network. It is estimated by dividing the number of edges connecting a vertex's neighbors by the total number of possible edges between the vertex's neighbors. (*Liang J et al., 2022*)

4.5 FUNCTIONAL ENRICHMENT AND GO ANALYSIS OF DEG's

To determine the biological significance of overlapping DEG's, Functional Enrichment analysis was performed based on Gene Ontology (GO) in Enrichr.

4.5.1 Enrichr

Enrichr serves as an extensive web server for gene set enrichment analysis, associating with functional or pathway terms. It includes features with ability to submit fuzzy sets, upload BED files, and improved application programming interface (API), along with result visualization via cluster-grams. Enrichr functions as both a search engine and a comprehensive repository for curated gene sets, synthesizing biological knowledge to facilitate further discoveries. Leveraging Enrichr's KEGG gene set libraries, utilizes various enrichment analysis methods, including p-value, rank (Z-score), and combined score, to explore the signaling pathways of identified DEGs (*Maxim V. Kuleshov et al., 2016*). It also is linked with several repositories that enable it to provide information on enrichment terms- Transcription, Pathways, Ontologies, Drugs, Diseases, and Cell types for a given input gene sets.

p- value

The p-value is determined through standard statistical methods such as Fisher's exact test or the hypergeometric test. This test assumes binomial distribution and independence in gene set probabilities.

q-value

The q-value is an adjusted p-value, calculated using the Benjamini-Hochberg method to correct for multiple hypotheses testing. (*Yoav Benjamini and Yosef Hochberg, 1995*).

Odds ratio

The odds ratio is computed using this formula:

$$\text{oddsRatio} = (1.0 * a * d) / \text{Math.max}(1.0 * b * c, 1)$$

here: a = overlapping genes, b = genes in the annotated set - overlapping genes, c = genes in the input set - overlapping genes, and d= total genes in the background - genes in the annotated set - genes in the input set + overlapping genes.

Combined score

The combined score is a combination of the p-value and odds ratio calculated by multiplying the two scores as: **$c = -\log(p) * \text{oddsRatio}$**

Here c = combined score, p = p-value computed using Fisher's exact test, and oddsRatio = odds ratio. The combined score provides a compromise between both methods and in several benchmarks, it reports the best rankings when compared with the other scoring schemes.

CHAPTER-5

RESULTS AND DISCUSSION

The network construction based on DEG analysis involved selecting appropriate correlation measures and applying thresholding to identify significant gene-expression relationships, thereby shortlisting statistically significant upregulated genes. A total of 13,871 common genes from DEGs of which a shortlisted 200+ were studied through network construction and analysis using Cytoscape, and CytoHubba. Subsequently, the PPI network of DEGs underwent functional annotation and enrichment analysis using tool Enrichr. The results elaborate on how Bioinformatics tools were employed to predict regulatory relationships and identify key regulatory genes associated with Autism Spectrum Disorder (ASD). The role of identified candidate genes and associated functional pathways in ASD expression and regulation was predicted and verified. Additionally, the predictive value of biomarkers was analyzed within this framework.

5.1 IDENTIFICATION OF DEGs

Using criteria of $\log FC > 2$ and adjusted $P\text{-value} < 0.05$, significant upregulated DEGs were identified (Table 5.2). In GSE143155 (fibroblasts), 2511 upregulated and 14,294 downregulated genes were found, while GSE102471 (prefrontal cortex) showed 2527 upregulated and 18,290 downregulated genes. Additionally, GSE18123 (peripheral blood cells) exhibited 410 upregulated and 42,276 downregulated genes. By comparing DEG lists between these three tissues, 13871 common genes were identified through Venn diagram intersection for further analysis. Of which based on prominence of Log FC value > 2 for statistical significance, 226 genes were filtered for PPI network construction.

Table 5.1 Datasets considered for TD vs ASD DEG analysis

GEO Accession Number	Tissue Type	Sample size	Healthy Controls	Test (ASD +ve)	Platform for GEP Analysis
GSE143155	neurons, fibroblasts	30	10	20	GPL21697 NextSeq 550 (Homo sapiens)
GSE102741	Prefrontal cortex, neurons	52	39	13	GPL11154 Illumina HiSeq 2000 (Homo sapiens)
GSE18123	peripheral blood cells	285	33	31	GPL570 [HG- U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array

Table 5.2 Criterion of Statistical Significance for Identifying the Significant Upregulated DEGs.

GSE143155	GSE102741	GSE18123
16803(DEG'S)	20815(DEG'S)	42684 (DEG'S)
p value <0.05	p value<0.05	p value<0.05
FC>2	FC>2	FC>2
SAMPLE SIZE:30	SAMPLE SIZE:52	SAMPLE SIZE:285



***Fig 5.1 Neurons, fibroblasts, prefrontal cortex and peripheral blood cells-
Common DEG's***

5.2 PPI NETWORK CONSTRUCTION AND ANALYSIS

Utilizing predicted interactions of common DEGs, we constructed a Protein-Protein Interaction (PPI) network employing the Reactome-FI Database to unveil pivotal proteins potentially crucial in ASD development. The network encompasses 1,520 genes (nodes) and 2,425 gene-gene interactions (edges). Each node signifies a

protein, and edges denote interactions between proteins, as depicted in Fig 5.2. Nodes were assigned degrees reflecting the number of neighboring nodes and changes in proteins/genes. We employed the novel Cytoscape plugin, CytoHubba, offering a user-friendly interface for hub gene assessment. CytoHubba ranks nodes by network features, facilitating network exploration and extraction of user-selected subnetworks. It encompasses 11 topological analysis methods to unravel the network's intricate features.

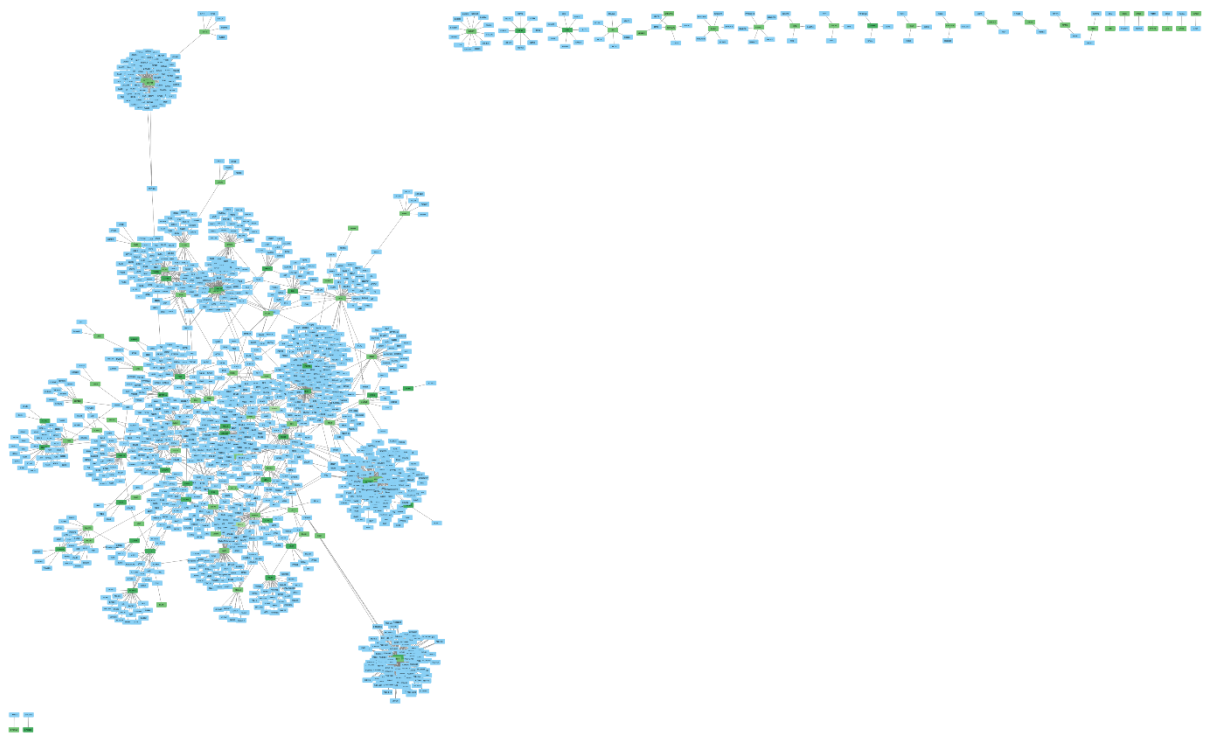


Fig 5.2 Reactome-FI Based Network With 1,520 nodes and 2,425 gene-gene interactions edges

5.3 NETWORK ASSESSMENT TO SCREEN HUB GENES

The hub genes were identified from the Protein-Protein Interaction (PPI) network utilizing 12 topological methods available in the CytoHubba plugin. These methods encompassed both local and global approaches, such as Degree, Maximum Neighborhood Component (MNC), Density of Maximum Neighborhood Component

(DMNC), and Maximal Clique Centrality (MCC), among others. Each gene was assessed based on its degree, representing the number of neighboring nodes in the network and changes in the proteins/genes. Notably, the top 20 hub nodes with the highest degrees in ASD included ITGB8, ERBB4, GRIN2A, FGFR3, WNT7A, WNT7B, CDC25A, VCAM1, COL17A1, COL22A1, HAP1, LRP2, SRC, GJD2, MYBL2, CDC2, CENPM, PCDH11Y, SPC24, GRIK3, and TERT. These genes were identified as high confidence factors and potent risk genes for ASD occurrence, based on their rankings across all topological methods. The high degree of these hub genes suggests their crucial roles within the protein interaction network. (See Table 5.3 and Table 5.4 for details).

Figures: Network of Topological Methods-Cytohubba

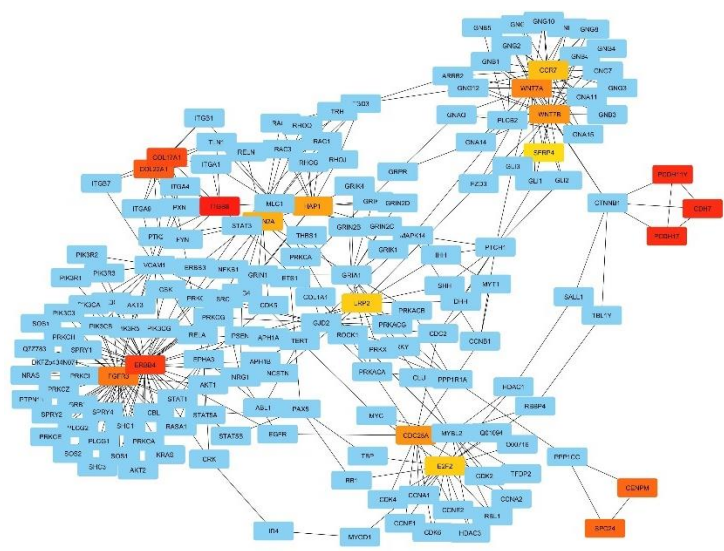


Fig 5.3 Degree

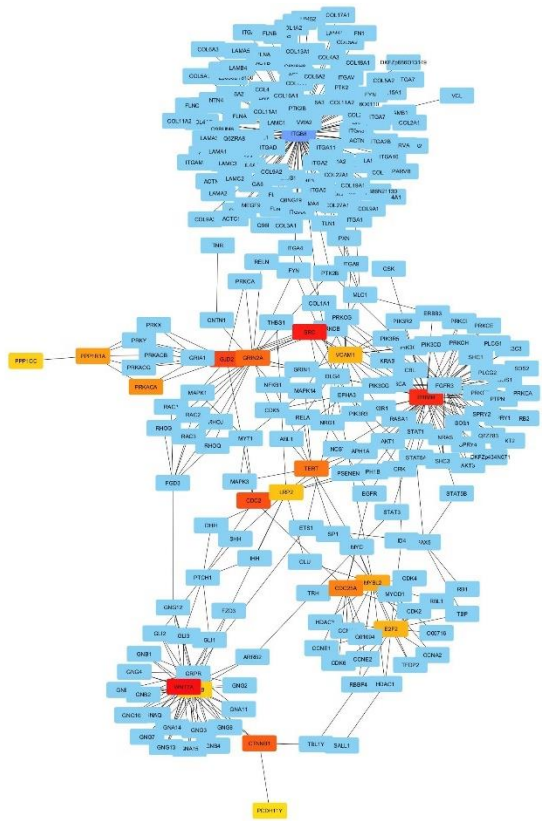


Fig 5.4 Bottleneck

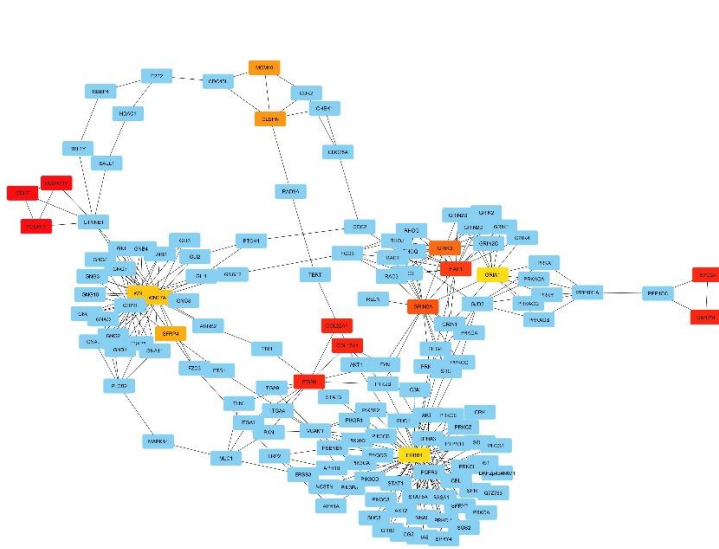


Fig 5.5 Maximum Neighborhood Component (MNC)

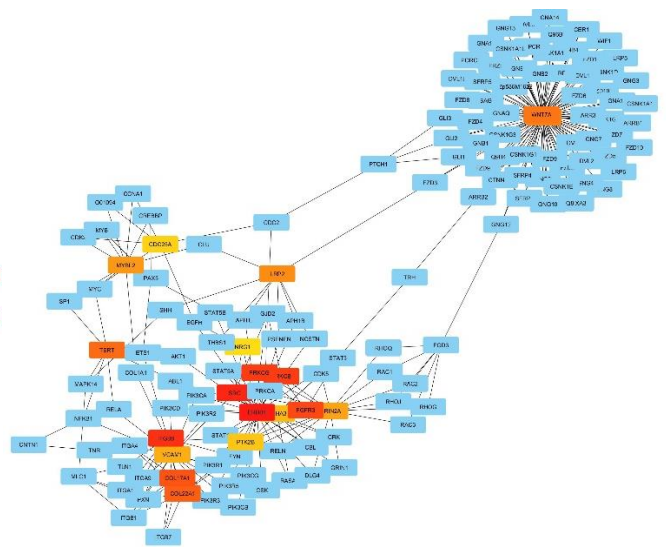


Fig 5.6 Closeness

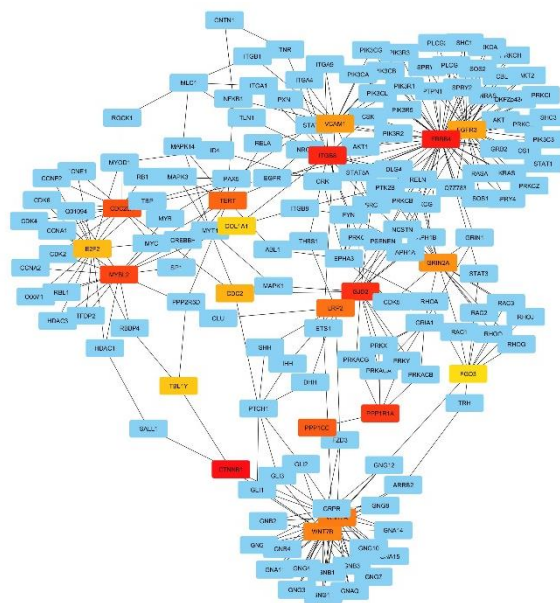


Fig 5.7 Betweenness

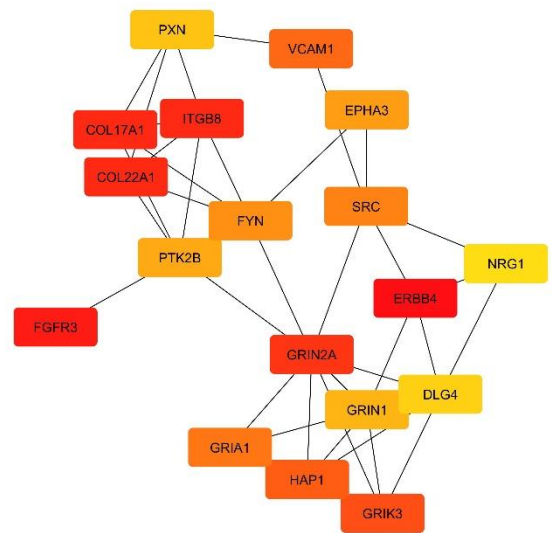


Fig 5.8 Edge Percolated Component (EPC)

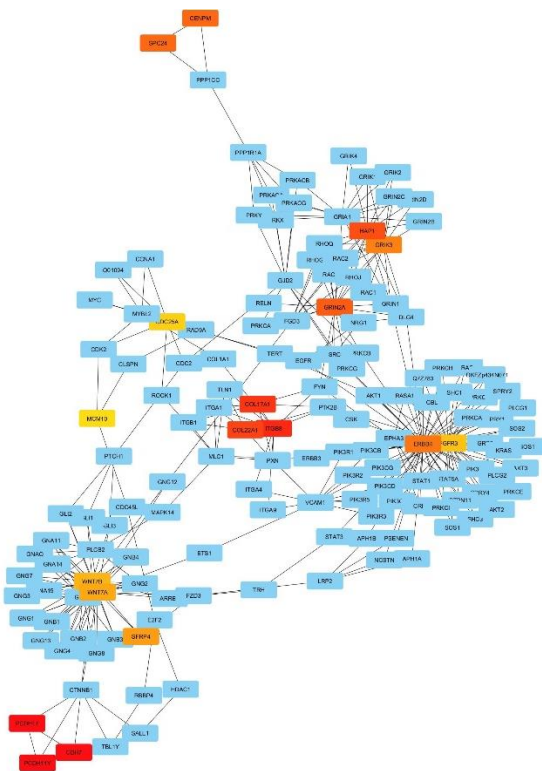


Fig 5.9 Maximal Clique Centrality(MCC)

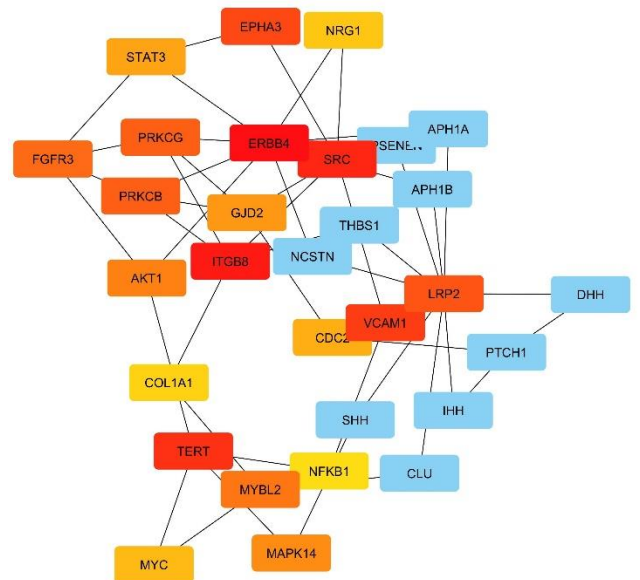


Fig 5.10 Radiality

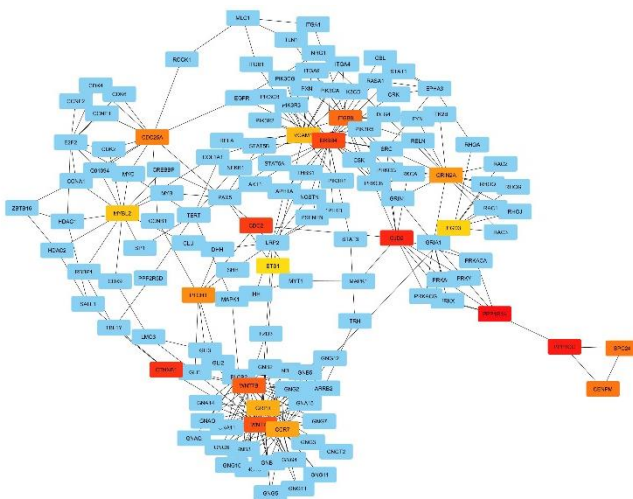


Fig 5.11 Stress

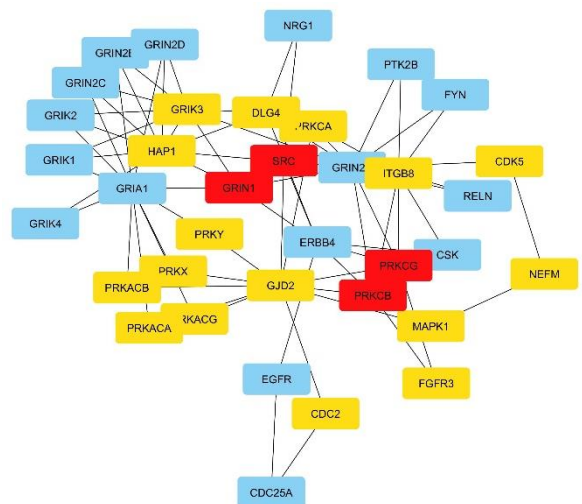


Fig 5.12 EcCentricity

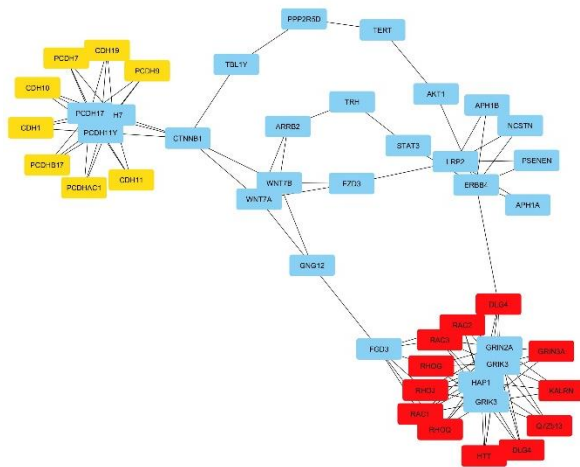


Fig 5.13 Density of Maximum Neighborhood Component (DMNC)

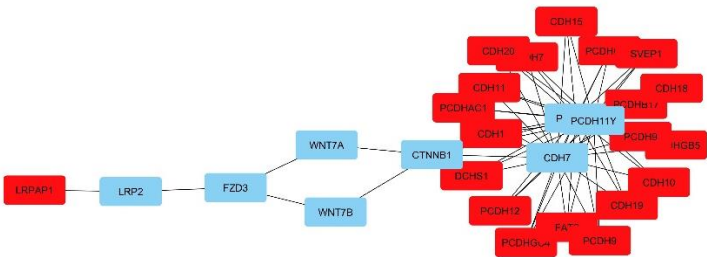


Fig 5.14 Clustering Coefficient

Table 5.3 Top 20 genes evaluated in the PPI network using Eleven Topological methods analysed using CytoHubba plugin in Cytoscape

Gene	Degree	Gene	Stress	Gene	Betweenness	Gene	Radiality
ITGB8	154	PPP1R1A	20000000.00	CTNNB1	352373.2	ERBB4	10.94801
ITGB8	122	PPP1CC	18700000.00	ERBB4	345367.3	ITGB8	10.80289
PCDH17	117	GJD2	18700000.00	ITGB8	339576.6	SRC	10.76612
CDH7	117	CTNNB1	15200000.00	GJD2	267923.8	TERT	10.76481
PCDH11Y	117	CDC2	10900000.00	PPP1R1A	250500.7	VCAM1	10.67025
ERBB4	101	ERBB4	10400000.00	MYBL2	233779.1	MYBL2	10.66763
COL17A1	89	WNT7A	9722452.00	CDC25A	221574.4	LRP2	10.66434
COL22A1	87	WNT7B	9696918.00	PPP1CC	221444	PRKCG	10.66237
CENPM	77	ITGB8	9368752.00	TERT	215889.2	PRKCB	10.66237
SPC24	77	CENPM	9254532.00	LRP2	214658.1	FGFR3	10.61904
FGFR3	76	SPC24	9254532.00	WNT7A	209378	MYBL2	10.60919
WNT7A	65	CDC25A	8920786.00	WNT7B	206546	AKT1	10.60787

WNT7B	64	PTCH1	8866208.00	GRIN2A	181110.2	MAPK14	10.57307
CDC25A	64	GRIN2A	8333432.00	VCAM1	145940.7	GJD2	10.5606
HAP1	58	CCR7	7900264.00	CCR7	140445	CCR7	10.54549
GRIN2A	48	GRPR	6769030.00	GRPR	139290.3	GRPR	10.53105
CCR7	46	CCR7	6626736.00	CCR7	134169.5	CCR7	10.5258
LRP2	42	MYBL2	6354028.00	MYBL2	128132	MYBL2	10.52317
Q14209	42	FGD3	6220532.00	COL1A1	119600.4	COL1A1	10.51857
SFRP4	40	ETS1	5858884.00	FGD3	116877.2	NFKB1	10.47852

Gene	Closeness	Gene	Eccentricity	Gene	Bottleneck	Gene	MNC
ITGB8	446.8381	PRKCG	0.116306	WNT7A	305	PCDH17	117
ERBB4	429.4996	PRKCB	0.116306	SRC	223	CDH7	117
SRC	399.4738	SRC	0.116306	ERBB4	208	PCDH11Y	117
ITGB8	397.0901	GRIN1	0.116306	ITGB8	186	COL17A1	77
PRKCG	391.2762	PRKCA	0.103383	GJD2	173	ITGB8	77
MYBL2	391.2762	MYBL2	0.103383	MYBL2	167	MYBL2	77
FGFR3	385.4246	DLG4	0.103383	CTNNB1	149	ITGB8	77
COL17A1	384.1813	PRKACG	0.103383	GRIN2A	120	CENPM	77
COL22A1	382.848	PRKY	0.103383	ITGB8	113	SPC24	77
TERT	379.2653	PRKACB	0.103383	TERT	113	HAP1	29
MYBL2	373.101	MYBL2	0.103383	MYBL2	100	MYBL2	28
WNT7B	372.4343	PRKACA	0.103383	PRKACA	96	GRIK3	27
LRP2	372.2688	CDC2	0.103383	PPP1R1A	95	GRIK3	24
MYBL2	366.874	MYBL2	0.103383	MYBL2	94	MYBL2	21
CCR7	366.3528	CCR7	0.103383	CCR7	92	CCR7	21
GRPR	365.1508	GRPR	0.103383	GRPR	92	GRPR	18
CCR7	363.6806	CCR7	0.103383	CCR7	91	CCR7	17
MYBL2	363.5226	MYBL2	0.103383	MYBL2	84	MYBL2	17
CDC25A	362.7631	HAP1	0.103383	WNT7B	84	ERBB4	13
NRG1	357.8746	GRIK3	0.103383	PCDH11Y	72	GRIA1	13

Gene	DMNC	Gene	EPC	Gene	MCC	Gene	Clustering Coefficient
RAC3	0.473661	ERBB4	420.228	PCDH17	690	LRPAP1	1
RAC1	0.473661	FGFR3	419.451	CDH7	690	CDH10	1
RHOG	0.473661	COL17A1	418.284	PCDH11Y	690	PCDHB17	1
RHOJ	0.473661	ITGB8	418.284	ITGB8	377	CDH19	1
RHOQ	0.473661	COL22A1	418.284	ITGB8	345	CDH11	1
MYBL2	0.473661	MYBL2	418.284	MYBL2	312	MYBL2	1
DLG4	0.473661	GRIN2A	411.951	COL22A1	310	PCDHAC1	1
DLG4	0.473661	GRIK3	411.676	HAP1	265	CDH1	1
Q7Z513	0.473661	GRIK3	411.379	GRIN2A	254	PCDH9	1
HTT	0.473661	HAP1	410.851	CENPM	152	SVEP1	1
MYBL2	0.473661	MYBL2	403.135	MYBL2	152	MYBL2	1
GRIN3A	0.473661	GRIA1	402.608	ERBB4	128	DCHS1	1
CDH10	0.463463	SRC	384.152	GRIK3	126	PCDHGA5	1
MYBL2	0.463463	MYBL2	382.379	MYBL2	108	MYBL2	1
CCR7	0.463463	CCR7	380.19	CCR7	86	CCR7	1
GRPR	0.463463	GRPR	374.528	GRPR	80	GRPR	1
CCR7	0.463463	CCR7	371.816	CCR7	79	CCR7	1
MYBL2	0.463463	MYBL2	370.628	MYBL2	76	MYBL2	1
CDH1	0.463463	DLG4	369.033	CDC25A	64	PCDH12	1
PCDH9	0.463463	NRG1	368.54	MCM10	51	PCDHGB5	1

Table 5.4 Comparison of gene from 12 topological methods

GENES	CCR7	CDC25	CDH7	CENP	COL17	COL22	E2F2	ERB	FGFR3	GRIN2	HAP1	ITGB8	LRP2	PCDH1	PCD	SFRP4	SPC	WNT7	WNT7
	A	A	A1	A1	M	A1	A1	B4	A	A	1Y	H17	24	A	B				
DEGREE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CLUSTERING COEFFICIENT																			
STRESS	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BETWEENNESS	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RADIALITY								✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CLOSENESS	✓		✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ECCENTRICITY									✓			✓							
BOTTLENECK	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MNC			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DMNC																			
EPC			✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MCC	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

GENES	MCM 10	SFR P4	GRIK3	SRC	GRIA1	NRG1	EPH A3	PTK 2B	VCAM1	DLG4	PRK ACA	PPP1 R1A	TERT	GJD2	PPP1 CC	MYBL2	CTNN B1
DEGREE		✓															
CLUSTERING COEFFICIENT																	
STRESS								✓				✓		✓		✓	✓
BETWEENNESS								✓				✓			✓		✓
RADIALITY				✓		✓	✓	✓	✓				✓	✓			✓
CLOSENESS				✓		✓	✓	✓	✓				✓	✓			✓
ECCENTRICITY			✓	✓						✓		✓		✓			
BOTTLENECK				✓				✓				✓	✓	✓		✓	✓
MNC	✓	✓	✓			✓											
DMNC										✓							
EPC			✓	✓	✓	✓	✓	✓	✓	✓							
MCC	✓	✓	✓														

GENES	CDC2	PRKCB	PRKCG	PCDH9	CDH11	CDH19	PCD HAC1	PCDH7	CDH10	PCDHB1 7	PCDH 17	CDH11	CDH19
DEGREE											✓		
CLUSTERING COEFFICIENT				✓	✓	✓	✓	✓	✓	✓		✓	✓
STRESS	✓												
BETWEENNESS	✓												
RADIALITY	✓	✓	✓										
CLOSENESS		✓	✓										
ECCENTRICITY	✓	✓	✓										
BOTTLENECK	✓												
MNC											✓		
DMNC				✓	✓	✓	✓	✓	✓	✓		✓	✓
EPC													
MCC												✓	

Table 5.5 High Confidence Factors for ASD

GENES	FREQUENCY
ITGB8	10
ERBB4	9
GRIN2A	8
FGFR3	7
WNT7A	7
WNT7B	7
CDC25A	6
VCAM1	6
COL17A1	5
COL22A1	5
HAP1	5
LRP2	5
SRC	5
GJD2	5
MYBL2	5
CDC2	5
CENPM	4
PCDH11Y	4
SPC24	4
GRIK3	4
TERT	4

5.4 GENE ONTOLOGY AND FUNCTIONAL ENRICHMENT ANALYSIS OF DEG's

GO and functional Enrichment enrichment analysis of the critical risk genes was performed in Enrichr to understand the biological relevance. On analysis of Enrichment in Biological processes the following GO were observed to be predominant: Lens Fiber Cell Development (GO:0070307), Lens Fiber Cell Differentiation (GO:0070306), Mammary Gland Epithelium Development (GO:0061180), Establishment Of Protein Localization To Telomere (GO:0070200), Regulation Of Excitatory Synapse Assembly (GO:1904889), Ventricular Compact Myocardium Morphogenesis (GO:0003223), Positive Regulation Of Protein Catabolic Process In The Vacuole (GO:1904352), Intracellular Oxygen Homeostasis (GO:0032364), Kidney Morphogenesis (GO:0060993), Metanephric Collecting Duct Development (GO:0072205), and Embryonic Hindlimb Morphogenesis (GO:0035116). (Figure 5.15, 5.16, 5.17 and Table 5.6).

Similar enrichment analysis for Molecular Function revealed that the critical risk genes were primarily involved in the Ionotropic Glutamate Receptor Activity (GO:0004970), Telomerase Activity (GO:0003720), RNA-directed DNA Polymerase Activity (GO:0003964), Primary Amine Oxidase Activity (GO:0008131), Neurotransmitter Receptor Activity Involved In Regulation Of Postsynaptic Membrane Potential (GO:0099529), Oxidoreductase Activity, Acting On The CH-NH₂ Group Of Donors, Oxygen As Acceptor (GO:0016641), NMDA Glutamate Receptor Activity (GO:0004972), Frizzled Binding (GO:0005109), BMP Receptor Binding (GO:0070700), Transmembrane Receptor Protein Serine/Threonine Kinase Binding (GO:0070696), and Phospholipase Activator Activity (GO:0016004). (Figure 5.18, 5.19, 5.20, and Table 5.7)

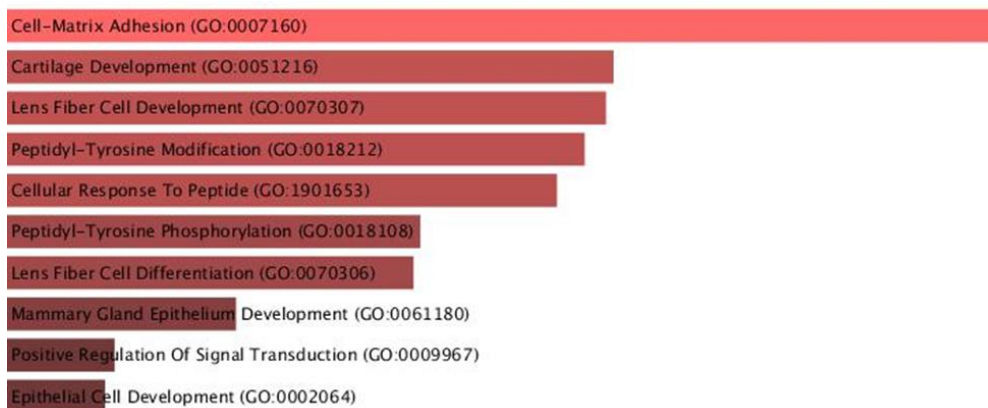


Fig 5.15 Functional Enrichment analysis based on Gene ontology –Biological Process

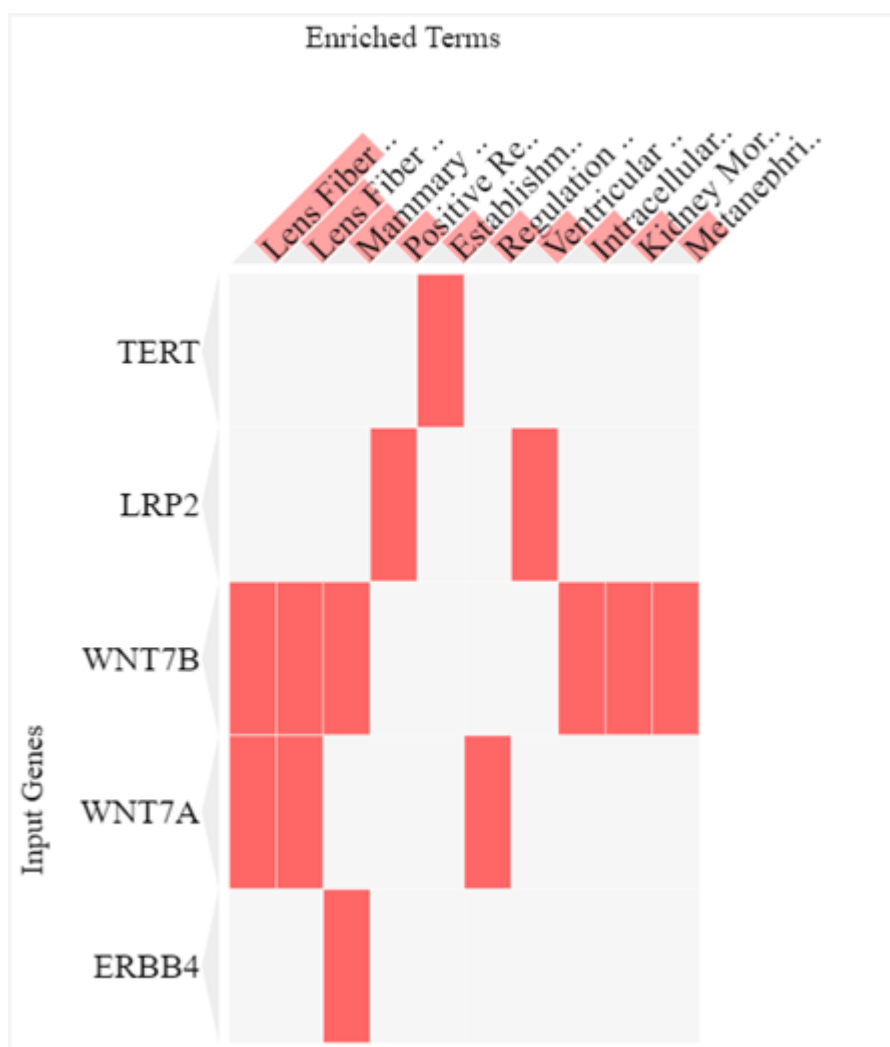


Fig 5.16 Cluster Dendrogram- Biological Process

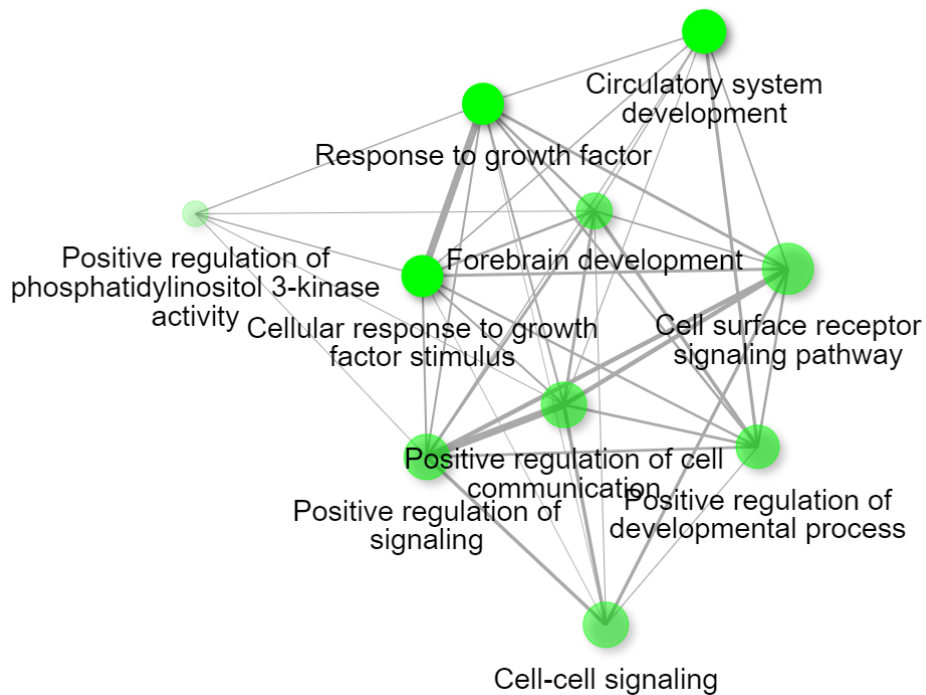


Fig 5.17 Functional Enrichment Network –Biological Process

Table 5.6 Genes Associated With Biological Process

GO TERMS	ASSOCIATED GENES
Lens Fiber Cell Development (GO:0070307)	WNT7B;WNT7A
Lens Fiber Cell Differentiation (GO:0070306)	WNT7B;WNT7A
Mammary Gland Epithelium Development (GO:0061180)	ERBB4;WNT7B
Establishment Of Protein Localization To Telomere (GO:0070200)	TERT
Regulation Of Excitatory Synapse Assembly (GO:1904889)	WNT7A
Ventricular Compact Myocardium Morphogenesis (GO:0003223)	LRP2

Positive Regulation Of Protein Catabolic Process In The Vacuole (GO:1904352)	LRP2
Intracellular Oxygen Homeostasis (GO:0032364)	WNT7B
Kidney Morphogenesis (GO:0060993)	WNT7B
Metanephric Collecting Duct Development (GO:0072205)	WNT7B
Embryonic Hindlimb Morphogenesis (GO:0035116)	WNT7A

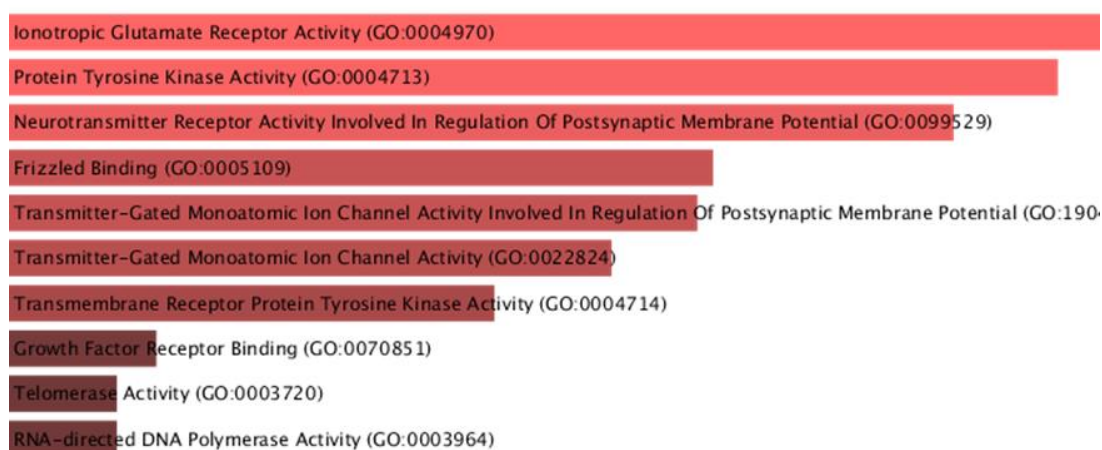


Fig 5.18 Functional Enrichment analysis based on Gene ontology –Molecular Function

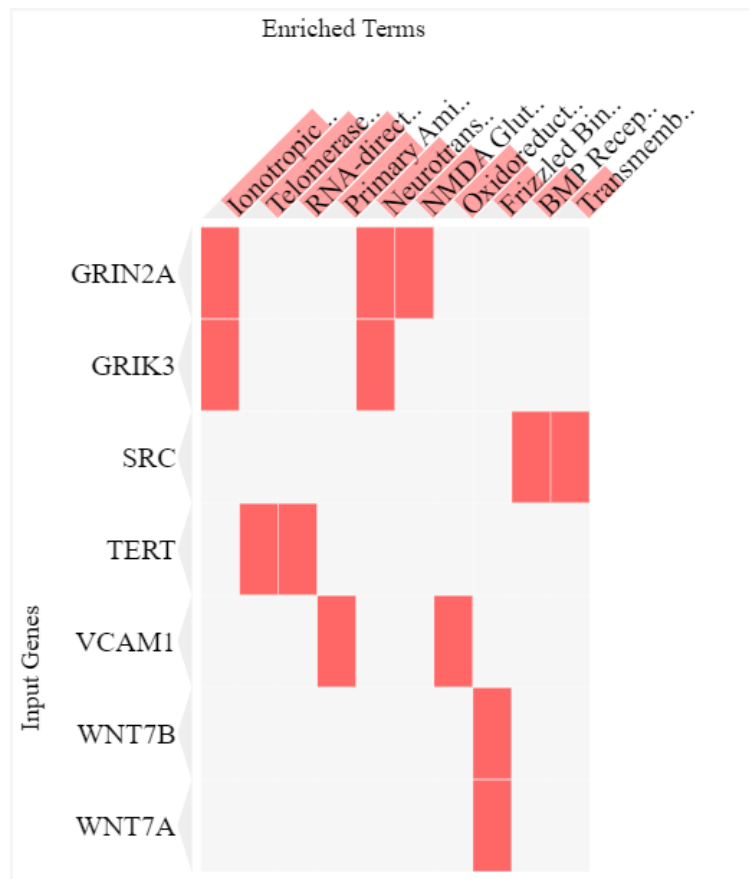


Fig 5.19 Cluster Dendrogram- Molecular Function

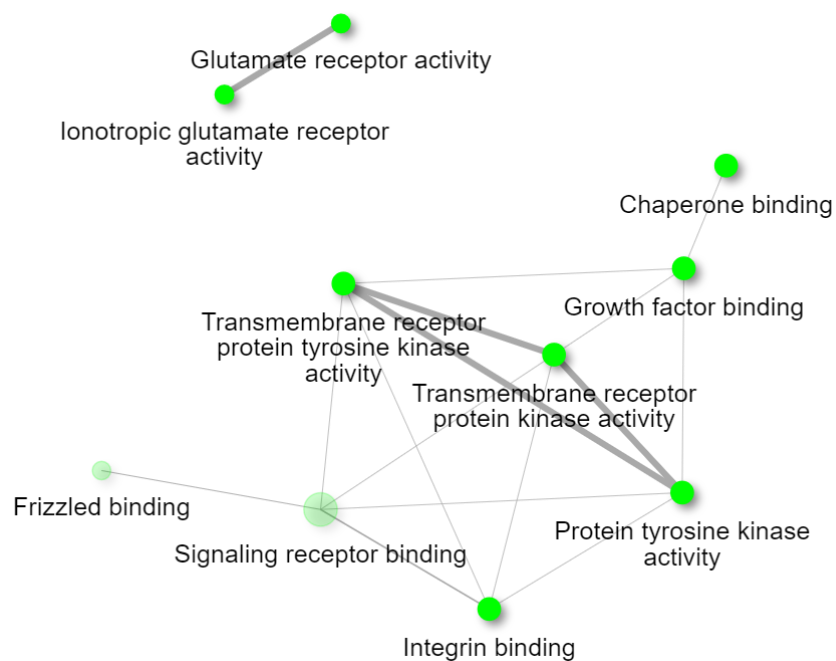


Fig 5.20 Functional Enrichment Network –Molecular Function

Table 5.7 Genes Associated With Molecular Function

GO TERMS	ASSOCIATED GENES
Ionotropic Glutamate Receptor Activity (GO:0004970)	GRIN2A;GRIK3
Telomerase Activity (GO:0003720)	TERT
RNA-directed DNA Polymerase Activity (GO:0003964)	TERT
Primary Amine Oxidase Activity (GO:0008131)	VCAM1
Neurotransmitter Receptor Activity Involved In Regulation Of Postsynaptic Membrane Potential (GO:0099529)	GRIN2A;GRIK3
Oxidoreductase Activity, Acting On The CH- NH2 Group Of Donors, Oxygen As Acceptor (GO:0016641)	VCAM1
NMDA Glutamate Receptor Activity (GO:0004972)	GRIN2A
Frizzled Binding (GO:0005109)	WNT7B;WNT7A
BMP Receptor Binding (GO:0070700)	SRC
Transmembrane Receptor Protein Serine/Threonine Kinase Binding (GO:0070696)	SRC
Phospholipase Activator Activity (GO:0016004)	SRC

5.5 BIOLOGICAL SIGNIFICANCE OF GENE MARKERS AND THEIR ROLE IN GENETIC AND ENVIRONMENTAL PREDISPOSITION TO ASD

GO and functional Enrichment enrichment analysis of the critical risk genes was revealed the impact of individual genes on diverse converging molecular pathways, which may hold the key to understand the significant impact these genes and their expression entail in the regulation of gene co-expression and transcriptionary mechanisms that may affect the expression of phenotypical and physiological symptoms associated with ASD. The influence of Environmental Factors in their characteristic expression and functioning also provide insights into the depth of the impact of aggregating environmental changes both evolutionary and due to external factors alter normal development in both pre-natal and neo-natal development stages, thereby explaining the increasing prevalence of Autism affected infants and adults in newer generations. Our findings highlight GRIN2A, GRIK3, TERT, VCAM1, WNT7A, WNT7B, SRC, ERBB4, and LRP2 genes as critical markers in ASD risk.

5.5.1 GRIN2A (Glutamate Receptor, Ionotropic, N-methyl-d-aspartate, Subunit 2)

GRIN2A (NR2A), a gene implicated in glutamatergic pathways in schizophrenia, showcases functional effects of disease-linked variants, highlighting the critical role of the S1-M1 linker in NMDA receptor channel opening. Early-Onset Psychosis in children correlates with an increased prevalence of rare GRIN2A variants (*Xie L et al., 2023*). In ASD GRIN2A encodes a subunit of the N-methyl-D-aspartate (NMDA) receptor, contributing to the formation of an NMDA-receptor channel. NMDA receptors play a crucial role in long-term potentiation (LTP), a process linked to learning and memory. Disruption of the mouse counterpart of GRIN2A, GluR ϵ , has been shown to decrease LTP and result in various behavioral deficits- mainly speech deficit, mood disorders (Bipolar) and epilepsy. (*Barnby G et al., 2005*).

The association of GRIN1 and GRIN2A-D with schizophrenia and their genetic interaction with maternal herpes simplex virus-2 (HSV-2) infection influencing disease susceptibility have been explored. Candidate gene investigations have indicated that interactions between maternal HSV-2 infection and GRIN2B elevate the risk of perturbation in glutamate receptor activity and induce a wide array of psychosis and epileptic syndromes. (*Demontis et al., 2011*). Exposure to Particulate

Matter less than 2.5 micrometre at the placental stage and increased oxidative stress response are also shown to affect GRIN2A activity. (*Ahadullah et al., 2021*)

5.5.2 GRIK3 (Glutamate Ionotropic Receptor Kainate Type Subunit 3)

GRIK3, also known as Glutamate Ionotropic Receptor Kainate Type Subunit 3, is a Protein Coding gene. It is implicated in diseases such as Schizophrenia and Depersonalization Disorder. Associated pathways include Presynaptic function of Kainate receptors and Transmission across Chemical Synapses. Gene Ontology (GO) annotations linked to this gene encompass glutamate-gated receptor activity and glutamate receptor activity. Notably, GRIK3 shares a significant relationship with GRIK2, an important paralog. Elevated levels of GRIK in the forebrain of mice result in social decline, heightened anxiety, and depressive symptoms, coupled with changes in synaptic communication. The replication of GRIK leads to behavioral traits akin to those seen in individuals with ASD and its expression in the amygdala perpetuates an enduring disruption in the balance between excitatory and inhibitory signals, disturbing the neural circuits governing essential amygdala functions. This imbalance disrupts the normal outputs of the amygdala (*Negrete-Díaz et al., 2022*).

Persistent exposure to PM 2.5 pollutant microparticles (*Sordillo et al., 2021*) and endocrine disrupting compounds from pesticides DDT, nad a wide array of seafood neurotoxicants (*Maggio A et al., 2022*) have been shown to disrupt methylome in regions associated with ASD in respiratory pathways as well as parent sperm and serum constitution.

Overall genes associated with functioning of glutamate receptor affect social deterioration, increased anxiety and depressive states, accompanied by altered synaptic transmission.

5.5.3 TERT (Telomerase Reverse Transcriptase)

TERT is a catalytic subunit that plays a key role in synthesizing telomeric DNA repeats from an RNA template. It influences the DNA repair response, modulating gene transcription, and impacting ubiquitin-proteasomal function (*Dratwa M et al., 2020*). TERT overexpression also disrupts the excitatory/inhibitory (E/I) balance in the medial prefrontal cortex, resulting in heightened hippocampal excitation

alongside unaltered inhibition, leading to significant impairment of long-term synaptic plasticity (Rhee et al., 2018). This alteration in the E/I ratio via telomerase activity corresponds to autism-like behaviors and synaptic abnormalities in the medial prefrontal cortex (Kim K.C et al., 2016). Notably, in mice studies the behavioral effects of TERT overexpression are observed predominantly in males, suggesting that perturbations in this gene and its downstream pathways preferentially impact the functional development of the male brain, aligning with the male preponderance observed in ASD (Kim K.C et al., 2016).

Overexpression of telomerase reverse transcriptase (TERT) in animals leads to diminished social interaction, reduced preference for novel social interactions, and impaired nest-building behaviors, reminiscent of symptoms observed in human autism spectrum disorders (ASD) (Rhee et al., 2018). Environmental stressors such as cortisol induced by stress, social status, smoking and obesity are known to have a significant impact on telomere activity mediated by TERT (Parsa N, 2012).

5.5.4 VCAM 1 (Vascular Cell Adhesion Molecule 1)

Immune cell entry into the central nervous system (CNS) necessitates traversing the blood-brain barrier (BBB), a barrier between circulating blood and brain tissue. Adhesion molecules play a pivotal role in facilitating leukocyte migration, thus influencing BBB permeability to immune cells. Vascular adhesion molecule-1 (VCAM-1) aids in leukocyte adhesion and transendothelial migration and is implicated in primary amine oxidase activity. Notably, VCAM-1 levels are diminished in Autism Spectrum Disorder (ASD) resulting in cytokine retention within circulating blood plasma. Evidence suggests ongoing neuroinflammation in ASD, characterized by heightened astroglial and microglial activation, and elevated levels of inflammatory cytokines and chemokines in the cerebral spinal fluid (CSF). Moreover, children with ASD exhibit peripheral immune anomalies, including irregular cytokine levels and atypical plasma immunoglobulin levels. Cellular immune dysregulation is also prevalent in ASD, manifesting as abnormal T cell function, aberrant NK cell activation, and anomalous monocyte responses. These findings underscore the involvement of cellular dysfunction and neuroinflammation in ASD, indicating potential aberrant neuroimmune interactions within the CNS. (Onore, C.E et al., 2012)

Studies On effects of Air pollutants on VCAM regulation observe that Both human brain capillaries and the brain parenchyma harbor air-borne particulate matter, although the precise transport mechanisms remain elusive. Moreover, heightened expression of intercellular adhesion molecule (ICAM) and vascular cell adhesion molecule (VCAM) in cerebral vasculature indicates endothelial activation. An ex vivo investigation showcased that Diesel Exhaust Particles (DEPs) induce oxidative stress, proinflammatory signaling, and up-regulation of P-glycoprotein in rat brain capillaries. These findings underscore the significance of the blood-brain barrier (BBB) as a crucial target for air pollutants. Therapeutic approaches aimed at modulating BBB permeability could potentially mitigate the neurotoxic impact of air pollutants on the central nervous system (CNS). (*Genc. S. et al., 2012*)

5.5.5 WNT7A and WNT7B (Wingless-related Integration Site Subunit 7A & 7B)

Wnt7a plays a crucial role in regulating axon guidance through the activation of Wnt/ β -catenin and divergent Wnt pathways across various neuronal subtypes. Experimental evidence suggests that Wnt7a expression induces axonal spreading, branching, and increased growth cone size and complexity in cerebellar granule cells. Moreover, it is essential for hippocampal dendrite development, with data indicating signaling through the Wnt/ Ca^{2+} pathway modulator CaMKII (*Turner et al., 2016*). This pathway activation influences both presynaptic and postsynaptic assembly and promotes excitatory synapse formation. Furthermore, Wnt7a is implicated in various neural processes, including neural stem cell renewal, synapse formation, axon guidance, neuronal progenitor cell progression, and gastrointestinal dysfunction (seen in associated disease with ASD). (*Mulligan & Cheyette, 2017*)

Whereas in Wnt7b decreased expression of Wnt7b is associated with deficits in dendritic spine development, axonal thalamic arborization, and social interaction. Ectopic activation of Wnt signaling with Wnt7b or GSK3 β inhibitors rescues synaptic defects, thereby impacting social behaviors. Deficiencies in chromatin remodeling proteins and transcription factors associated with ASD mirror phenotypes observed in experiments altering β -catenin levels, underscoring the importance of Wnt/ β -catenin dependent transcription in ASD and related neurodevelopmental disorders. Additionally, the Wnt/ Ca^{2+} pathway is implicated in dendrite development in the

hippocampus, similar to Wnt7a (*Caracca M, et al., 2021*).

Both WNT7A and WNT7B play critical roles in cell growth, differentiation, and Wnt signaling pathways. While WNT7A primarily influences axon guidance, synapse formation, and dendrite development, WNT7B is associated with dendritic spine development and social interaction. A recent study reviewed the exposure to environmental heavy metals can disrupt normal cellular function by either inhibiting or overexpressing various components of Wnt signaling pathways. These components include Wnt proteins, transcription factors, receptors, ligands, or transducers. Such interference adversely affects Wnt signaling pathways, leading to impaired cellular function (*Khalid M et al., 2021*).

5.5.6 SRC (Sarcoma)

Src, a non-receptor protein tyrosine kinase, governs essential cellular functions such as cell growth, differentiation, morphogenesis, and motility. Upon activation, Src facilitates survival, angiogenesis, proliferation, and invasion pathways. SRC protein expression levels, along with integrin $\beta 1$ and FAK, are markedly reduced in lymphoblasts from individuals with autism. Additionally, Src protein expression and phosphorylation at the active site (Y416) are significantly diminished. Autistic lymphoblasts exhibit decreased migration, heightened adhesion, and impaired IgG production. The introduction of FAK overexpression in autistic lymphoblasts counteracts adhesion and migration defects. Furthermore, our findings suggest that FAK operates through the activation of Src, phosphatidylinositol 3-kinase–Akt, and mitogen-activated protein kinase signaling pathways, with paxillin potentially involved in adhesion and migration regulation (*Wei H et al., 2011*). Reactivation of SRC restores synaptic integrity-related protein expression in primary neurons from TSPAN7 knockout brains. This indicates that TSPAN7 knockout induces ASD-like and ID-like behaviors in rats and disrupts neuronal synapses, likely via the down-regulation of the integrin $\beta 1$ /FAK/SRC signaling pathway (*Pang. S et al., 2022*). These insights shed light on SRC's role in psychiatric disorders and identify integrin $\beta 1$ /FAK/SRC as a potential therapeutic target for ASD.

Heavy metal toxins found in underground water in industrial areas affect the Src/Ras pathway and known to trigger Cr(VI)-initiated tumor prone actions. (*Ganapathy. S et al., 2022*). Also exposure to UV radiation and Particulate Matter in air pollutants

induce abnormal activation of SRC protein kinases and receptors (*Paik. S.J et al., 2023*).

5.5.7 ERBB4 (Erb-B2 Receptor Tyrosine Kinase 4)

ERBB4, a single-pass type I transmembrane protein, interacts with and gets activated by neuregulins, heparin-binding EGF-like growth factor, and betacellulin. This binding triggers diverse cellular responses, such as mitogenesis and differentiation. ERBB4 is notably involved in distinct functions during normal tissue development, particularly in the heart, nervous system, and mammary gland system. (*Sergers. V.F.M. et al., 2020*). Synaptic communication deficits emerge as a common thread in cognitive disorders, evidenced by a direct association between schizophrenia and Angelman syndrome (AS), an associated sub-expression autism spectrum disorder, at the level of synaptic plasticity and learning. Kaphzan et al., demonstrated aberrant neuregulin-ErbB4 (NRG1-ErbB4) signaling in a Ube3a-deficient mouse model of Angelman syndrome, leading to hippocampal synaptic plasticity deficits and impaired contextual fear memory (*Kaphzan et al., 2012*). ERBB4 encodes the tyrosine kinase receptor HER4, pivotal in normal cellular function and neurodevelopment in the brain. The HER4-NRG1 signaling pathway plays a critical role in neural crest cell migration and neuronal differentiation. Pharmacological inhibition of HER4 accelerates epileptogenesis progression in rodent models, while heterozygous ERBB4 null mice exhibit cognitive deficits and delayed motor development. Chromosome 2q34 deletions, resulting in heterozygous intragenic loss of multiple ERBB4 exons, lead to haploinsufficiency of ERBB4, manifesting as non-syndromic intellectual disability or epilepsy (*Heck. D.H. et al., 2012*).

Environmental air pollutants, known for their association with cancer and respiratory ailments, possess the capability to activate EGFR. Specifically, the environmental electrophile 1,2-naphthoquinone (1,2-NQ), found in diesel exhaust particles and particulate matter, has previously demonstrated an impact on EGFR signaling (*Nakahara. K et al., 2021*).

Di-(2-ethylhexyl) phthalate (DEHP), a synthetic chemical used to impart flexibility to plastics, has been implicated in altering gene expression profiles. Transcriptome sequencing studies have revealed that DEHP decreases NRG1 expression and

downregulates target genes associated with the ErbB2/ErbB4-PI3K/AKT signaling pathway. Prenatal exposure to DEHP has been shown to reduce NRG1 expression in fetal heart tissue, potentially influencing myocardial development and inducing myocardial cytotoxicity through NRG1-dependent ErbB pathway modulation (Yu *et al.*, 2022).

5.5.8 LRP2 (Low Density Lipoprotein Receptor-related Protein 2)

LRP2, a member of the low density lipoprotein receptor (LDLR) gene family, exhibits expression in the brain, hinting at its involvement in neuronal development and regeneration. Notably, LRP2's interaction with certain ligands underscores its role in these processes. Recently, a *de novo* deleterious mutation in LRP2 was identified in an ASD case through whole-exome sequencing. Additionally, LRP2 influences brain development, particularly forebrain development, by modulating sonic hedgehog (Shh) signaling. This mechanism involves LRP2 binding to BMP4, promoting its internalization and degradation. Diminished LRP2 levels correlate with increased BMP4 levels, disrupting normal brain development (Ionita-Laza *et al.*, 2012). Mouse mutants lacking or having underdeveloped corpus callosum, defective nervous systems, and abnormal behavioral traits further underscore LRP2's significance in neuronal and brain development (Li *et al.*, 2014).

A cohort study in 2021 revealed the association of Smoking and cardiovascular risk factors and its influence on the expression of the LRP2 gene via “Linkage Disequilibrium” (Chow *et al.*, 2021). Other factors including low sunlight exposure, Vitamin D deficiency, Viral infections and neonatal- trauma have also been deduced to affect the regulation of lipoprotein receptor and downstream processes (Waubant *et al.*, 2019).

Table 5.8 Environmental Factors affecting high risk genes

GENES	ENVIRONMENTAL FACTORS	EFFECTIVE STAGES	MECHANISM	REFERENCES
GRIN2A - Glutamate Receptor, Ionotropic, N-methyl-d-aspartate, Subunit 2	HSV-2 infection, PM2.5µm	Maternal, Placental	Glutamate receptor signalling, and oxidative stress response	(Demontis et al., 2011), (Ahadullah et
GRIK3 - Glutamate Ionotropic Receptor Kainate Type Subunit 3	Seafood neurotoxins, pesticides DDT, PM2.5µm	Paternal, Neonatal	Sperm methylation and paternal serum alterations, effect on respiratory pathways of Asd patients	(Sordillo et al., 2021), (Maggio A et al., 2022)
TERT -Telomerase Reverse Transcriptase	Cortisol induced by stress, social status, smoking, obesity	Parental/ Foetal	Telomere shortening, induced mutations	(Parsa N, 2012)
VCAM1 - Vascular Cell Adhesion Molecule 1	Diesel Exhaust Particles (DEPs), PM2.5µm	Neonatal, Perinatal	Endothelial activation, oxidative stress, proinflammatory signaling	(Genc. S. et al., 2012)
WNT7A and WNT7B - (Wingless-related Integration Site Subunit 7A & 7B)	Environmental heavy metals, PM2.5µm	Perinatal, (contd.) adulthood	inhibition/ overexpression of Wnt proteins, ligands, receptors	(Khalid M et al., 2021).
SRC - Sarcoma	Heavy metal toxins, UV radiation, PM2.5µm	Maternal/ Neonatal	abnormal activation of SRC protein kinases, receptors	(Ganapathy. S et al., 2022), (Paik. S.J et al.,
ERBB4 - Erb-B2 Receptor Tyrosine Kinase 4	Di-(2-ethylhexyl) phthalate (DEHP), PM2.5µm	Maternal/ Neonatal	downregulates target genes associated with the ErbB2/ErbB4-PI3K/AKT	(Nakahara. K et al., 2021), (Yu et al., 2022).
LRP2 - Low Density Lipoprotein Receptor-related Protein 2	Smoking, low sunlight, Vitamin D deficiency	Maternal/ Neonatal	linkage disequilibrium affect regulation of lipoprotein receptor	(Chow et al., 2021), (Waubant et al.,

CHAPTER-6

SUMMARY AND CONCLUSION

This study provides compelling evidence of convergent molecular abnormalities in ASD, shedding light on the underlying mechanisms of neuronal dysfunction in this disorder. Utilizing gene expression data, it supports emerging evidence of abnormalities in peripheral blood leukocytes and cortical neurons in autism, potentially indicating genetic predispositions during perinatal, pre-natal and neonatal development stages. The findings are analyzed within the context of genetic variations' transmission to clinical characteristics, offering insights into underlying molecular mechanisms and potential subtype-specific biomarkers. This comprehensive understanding enhances our grasp of ASD heterogeneity and can steer future research and personalized interventions for affected individuals.

Our findings highlight GRIN2A, GRIK3, TERT, VCAM1, WNT7A, WNT7B, SRC, ERBB4, and LRP2 genes as critical markers in ASD risk. Genes such as GRIN2A and GRIK3 contribute to neurotransmitter receptor activity in ASD associated with NMDA receptors, while TERT alters excitatory/inhibitory (E/I) potential via telomerase activity. VCAM 1 facilitate leukocyte adhesion and trans endothelial migration and is involved in primary amine oxidase activity, and WNT7A and WNT7B are crucial for neural stem cell renewal, synapse formation, axon guidance, and neuronal progenitor cell progression and lens fibre cell development. Additionally, SRC participates in BMP receptor binding involved in the regulation of adhesion and migration in autistic lymphoblasts, ERBB4 is associated with synaptic communication deficits in hippocampal synaptic plasticity and contextual fear memory, and LRP2 contributes to neuronal development and regeneration through its effects on sonic hedgehog (Shh) signaling and ventricular compact myocardium morphogenesis.

The identified candidate genes predominantly influence various aspects of neural development, synaptic signaling, and morphogenesis processes such as cortical fiber cell development and myocardial morphogenesis.

Additionally, they are involved in telomerase and primary amine oxidase activities, hinting at potential associations with oxidative stress, abnormal neuron excitation/inhibition, and delayed or incomplete physiological development in ASD pathogenesis. These insights highlight the regulation of disease phenotypes across the ASD spectrum.

Furthermore, numerous studies underscore the impact of environmental pollutants resulting from industrialization, including PM_{2.5}, pesticide microcompounds, persistent organic pollutants, heavy metal pollutants, diesel residues, microplastics, and modern-day stressors like obesity and diabetes and exposure to it affecting sperm methylome, placental transmission of altered epigenetic factors, de-novo variations and general susceptibility of nascent stem-cells in developing foetuses to these pollutants underscoring the environmental predisposition to heightened occurrence of Autism Spectrum Disorder in modern times.

Further, these findings may serve as pivotal targets for therapeutic interventions, offering potential benefits for a broad spectrum of individuals affected by ASD.

REFERENCES

- 1) About GEO2R - GEO - NCBI. (n.d.). Retrieved April 12, 2024, from <https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html#rnaseq>
- 2) Ahadullah, Yau, S., Lu, H., Lee, T. M. C., Guo, H., & Chan, C. C. H. (2021). PM2.5 as a potential risk factor for autism spectrum disorder: Its possible link to neuroinflammation, oxidative stress and changes in gene expression. *Neuroscience & Biobehavioral Reviews*, 128, 534–548. doi:10.1016/j.neubiorev.2021.06.043
- 3) AIIMS New Delhi. (n.d.). Retrieved April 12, 2024, from <http://pedneuroaiims.org/autism-spectrum-disorder-diagnosis.html>
- 4) American Psychiatric Association. Autism spectrum disorder. In: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, American Psychiatric Association, Arlington, VA 2013. p.50
- 5) Andrade C. (2019). The *P* Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian journal of psychological medicine*, 41(3), 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19
- 6) Ashwood, P., Wills, S., & Van de Water, J. (2006). The immune response in autism: a new frontier for autism research. *Journal of leukocyte biology*, 80(1), 1–15. <https://doi.org/10.1189/jlb.1205707>
- 7) Autism. (n.d.). Retrieved April 12, 2024, from <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>
- 8) Ayaz, A., Gezdirici, A., Yilmaz Gulec, E., Ozalp, O., Koseoglu, A. H., Dogru, Z., & Yalcintepe, S. (2022). Diagnostic Value of Microarray Method in Autism Spectrum Disorder, Intellectual Disability, and Multiple Congenital Anomalies and Some Candidate Genes for Autism: Experience of Two Centers. *Medeniyet medical journal*, 37(2), 180–193. <https://doi.org/10.4274/MMJ.galenos.2022.70962>
- 9) Bai, Jun & Tan, Renbo & An, Zheng & Xu, Ying. (2022). Quantitative estimation of intracellular oxidative stress in human tissues. *Briefings in bioinformatics*. 23. 10.1093/bib/bbac206.
- 10) Balachandar, V., Mahalaxmi, I., Neethu, R., Arul, N., & Abhilash, V. G. (2022). New insights into epigenetics as an influencer: An associative study between maternal prenatal factors in Autism Spectrum Disorder (ASD). *Neurology Perspectives*, 2(2), 78–86. <https://doi.org/10.1016/j.neurop.2022.01.002>
- 11) Baldwin WS, Heintz MM, Noorai RE, Clemson University (2020). (GEO Accession

- Viewer,GSE143155- Poor Mitochondrial Responses to Oxidative Stress in a Subset of Autism Spectral Disorder Patients),
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143155>
- 12) Barnby, G., Abbott, A., Sykes, N., Morris, A., Weeks, D. E., Mott, R., Lamb, J., Bailey, A. J., Monaco, A. P., & International Molecular Genetics Study of Autism Consortium (2005). Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. *American journal of human genetics*, 76(6), 950–966.
<https://doi.org/10.1086/430454>
 - 13) Baron, C. A., Liu, S. Y., Hicks, C., & Gregg, J. P. (2006). Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. *Journal of autism and developmental disorders*, 36(8), 973–982. <https://doi.org/10.1007/s10803-006-0134-x>
 - 14) Barrett T, Edgar R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO), 2006.
 - 15) Barrett T, Troup DB, Wilhite SE, et al., NCBI GEO: mining tens of millions of expression profiles--database and tools update, 2007.
 - 16) Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29, 1165-1188.
 - 17) Bradford, Y., Haines, J., Hutcheson, H., Gardiner, M., Braun, T., Sheffield, V., Cassavant, T., Huang, W., Wang, K., Vieland, V., Folstein, S., Santangelo, S., & Piven, J. (2001). Incorporating language phenotypes strengthens evidence of linkage to autism. *American journal of medical genetics*, 105(6), 539–547.
 - 18) Caracci, M. O., Avila, M. E., Espinoza-Cavieres, F. A., López, H. R., Ugarte, G. D., & De Ferrari, G. V. (2021). Wnt/ β -Catenin-Dependent Transcription in Autism Spectrum Disorders. *Frontiers in molecular neuroscience*, 14, 764756.
<https://doi.org/10.3389/fnmol.2021.764756>
 - 19) Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A.
 - 20) Chia-Hao Chin, Shu-Hwa Chen, and Hsin-Hung Wu et al., cytoHubba: identifying hub objects and sub-networks from complex interactome, 2014.
 - 21) Chia-Hao Chin, Shu-Hwa Chen, Hsin-Hung Wu et al., Hubba: hub objects analyzer—a framework of interactome hubs identification for network biology, 2008.
 - 22) Chow, H. H., Talbot, J., Marstrand, L., Lundell, H., Roman Siebner, H., Bach Sørensgaard, H., & Sellebjerg, F. (2021). Smoking, cardiovascular risk factors and LRP2 gene variation: Associations with disease severity, cognitive function and brain

- structure in primary progressive multiple sclerosis. *Multiple Sclerosis and Related Disorders*, 56, 103296. <https://doi.org/10.1016/j.msard.2021.103296>
- 23) Dalman, M. R., Deeter, A., Nimishakavi, G., & Duan, Z. H. (2012). Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC bioinformatics*, 13 Suppl 2(Suppl 2), S11. <https://doi.org/10.1186/1471-2105-13-S2-S11>
 - 24) Del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R. C., Galeota, E., Launay, G., Hermjakob, H. (2013). A new reference implementation of the PSICQUIC web service. *Nucleic Acids Research*, 41(W1), W601–W606. <https://doi.org/10.1093/nar/gkt392>.
 - 25) Demontis, D., Nyegaard, M., Buttenschøn, H. N., Hedemand, A., Pedersen, C. B., Grove, J., Flint, T. J., Nordentoft, M., Werge, T., Hougaard, D. M., Sørensen, K. M., Yolken, R. H., Mors, O., Børglum, A. D., & Mortensen, P. B. (2011). Association of GRIN1 and GRIN2A-D with schizophrenia and genetic interaction with maternal herpes simplex virus-2 infection affecting disease risk. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 156B(8), 913–922. <https://doi.org/10.1002/ajmg.b.31234>
 - 26) Dratwa, M., Wysoczańska, B., Łacina, P., Kubik, T., & Bogunia-Kubik, K. (2020). TERT-Regulation and Roles in Cancer Formation. *Frontiers in immunology*, 11, 589929. <https://doi.org/10.3389/fimmu.2020.589929>
 - 27) Duffney, L. J., Valdez, P., Tremblay, M. W., Cao, X., Montgomery, S., McConkie-Rosell, A., & Jiang, Y. H. (2018). Epigenetics and autism spectrum disorder: A report of an autism case with mutation in H1 linker histone HIST1H1E and literature review. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 177(4), 426–433. <https://doi.org/10.1002/ajmg.b.32631>
 - 28) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013; 128(14).
 - 29) Feng, Y., Zhang, F., Lokey, L. K., Chastain, J. L., Lakkis, L., Eberhart, D., & Warren, S. T. (1995). Translational suppression by trinucleotide repeat expansion at FMR1. *Science (New York, N.Y.)*, 268(5211), 731–734. <https://doi.org/10.1126/science.7732383>
 - 30) Folstein, S., & Rutter, M. (1977). Genetic influences and infantile autism. *Nature*, 265(5596), 726–728. <https://doi.org/10.1038/265726a0>
 - 31) Freeman L. A Set of Measures of Centrality Based on Betweenness. *Sociometry*.

- 1977; 40:35–41.
- 32)G. K. Smyth, J. Michaud, and H. Scott. The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, 21(9):2067– 2075, 2005.(pvalue).
 - 33)Ganapathy, S., Liu, J., Yu, T., Xiong, R., Zhang, Q., Makriyannis, A., & Chen, C. (2022). PKC is an indispensable factor in promoting environmental toxin chromium-mediated transformation and drug resistance. *Aging*, 14(4), 1678–1690. <https://doi.org/10.18632/aging.203917>
 - 34)Genc, S., Zadeoglulari, Z., Fuss, S. H., & Genc, K. (2012). The adverse effects of air pollution on the nervous system. *Journal of toxicology*, 2012, 782462. <https://doi.org/10.1155/2012/782462>
 - 35)GEO Accession viewer. (n.d.). Retrieved April 13, 2024, from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143155>
 - 36)Gregg, J. P., Lit, L., Baron, C. A., Hertz-Picciotto, I., Walker, W., Davis, R. A., Croen, L. A., Ozonoff, S., Hansen, R., Pessah, I. N., & Sharp, F. R. (2008). Gene expression changes in children with autism. *Genomics*, 91(1), 22–29. <https://doi.org/10.1016/j.ygeno.2007.09.003>
 - 37)Hage P, Harary F. Eccentricity and centrality in networks. *Social Networks*. 1995;17:57–63.
 - 38)He X, Zhang J (2006) Why Do Hubs Tend to Be Essential in Protein Networks? *PLoS Genet* 2(6): e88. <https://doi.org/10.1371/journal.pgen.0020088>
 - 39)Heck, D. H., & Lu, L. (2012). The social life of neurons: synaptic communication deficits as a common denominator of autism, schizophrenia, and other cognitive disorders. *Biological psychiatry*, 72(3), 173–174. <https://doi.org/10.1016/j.biopsych.2012.05.013>
 - 40)Hu, V. W., & Steinberg, M. E. (2009). Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism research : official journal of the International Society for Autism Research*, 2(2), 67–77. <https://doi.org/10.1002/aur.72>
 - 41)Hu, V. W., Frank, B. C., Heine, S., Lee, N. H., & Quackenbush, J. (2006). Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. *BMC genomics*, 7, 118. <https://doi.org/10.1186/1471-2164-7-118>
 - 42)Hu, V. W., Nguyen, A., Kim, K. S., Steinberg, M. E., Sarachana, T., Scully, M. A.,

- Soldin, S. J., Luu, T., & Lee, N. H. (2009). Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. *PloS one*, 4(6), e5775. <https://doi.org/10.1371/journal.pone.0005775>
- 43) International Open Access Week: Network Visualization & Analysis Tool (2018), Reactome FIViz-Cytoscape Plugin, 23 October 2018. (<https://reactome.org/about/news/112-international-open-access-week-network-analysis-tools>)
- 44) Ionita-Laza, I., Makarov, V., ARRA Autism Sequencing Consortium, & Buxbaum, J. D. (2012). Scan-statistic approach identifies clusters of rare disease variants in LRP2, a gene linked and associated with autism spectrum disorders, in three datasets. *American journal of human genetics*, 90(6), 1002–1013. <https://doi.org/10.1016/j.ajhg.2012.04.010>
- 45) Ionita-Laza, I., Makarov, V., ARRA Autism Sequencing Consortium, & Buxbaum, J. D. (2012). Scan-statistic approach identifies clusters of rare disease variants in LRP2, a gene linked and associated with autism spectrum disorders, in three datasets. *American journal of human genetics*, 90(6), 1002–1013. <https://doi.org/10.1016/j.ajhg.2012.04.010>
- 46) Jones, M. B., & Szatmari, P. (2002). A risk-factor model of epistatic interaction, focusing on autism. *American journal of medical genetics*, 114(5), 558–565. <https://doi.org/10.1002/ajmg.10513>
- 47) Kaphzan, H., Hernandez, P., Jung, J. I., Cowansage, K. K., Deinhardt, K., Chao, M. V., Abel, T., & Klann, E. (2012). Reversal of impaired hippocampal long-term potentiation and contextual fear memory deficits in Angelman syndrome model mice by ErbB inhibitors. *Biological psychiatry*, 72(3), 182–190. <https://doi.org/10.1016/j.biopsych.2012.01.021>
- 48) Khalid, M., Hodjat, M., & Abdollahi, M. (2021). Environmental Exposure to Heavy Metals Contributes to Diseases Via Deregulated Wnt Signaling Pathways. *Iranian journal of pharmaceutical research : IJPR*, 20(2), 370–382. <https://doi.org/10.22037/ijpr.2021.114897.15089>
- 49) Kim, K. C., Rhee, J., Park, J. E., Lee, D. K., Choi, C. S., Kim, J. W., Lee, H. W., Song, M. R., Yoo, H. J., Chung, C., & Shin, C. Y. (2016). Overexpression of Telomerase Reverse Transcriptase Induces Autism-like Excitatory Phenotypes in Mice. *Molecular neurobiology*, 53(10), 7312–7328. <https://doi.org/10.1007/s12035-015-9630-3>

- 50)Kong, S. W., Collins, C. D., Shimizu-Motohashi, Y., Holm, I. A., Campbell, M. G., Lee, I. H., Brewster, S. J., Hanson, E., Harris, H. K., Lowe, K. R., Saada, A., Mora, A., Madison, K., Hundley, R., Egan, J., McCarthy, J., Eran, A., Galdzicki, M., Rappaport, L., Kunkel, L. M., ... Kohane, I. S. (2012). Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PloS one*, 7(12), e49475. <https://doi.org/10.1371/journal.pone.0049475>
- 51)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.
- 52)Li, J., Shi, M., Ma, Z., Zhao, S., Euskirchen, G., Ziskin, J., Urban, A., Hallmayer, J., & Snyder, M. (2014). Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. *Molecular Systems Biology*, 10(12), 774. <https://doi.org/10.15252/msb.20145487>
- 53)Liang J, Chen X, Wang T. Percolation Distribution in Small-World Quantum Networks. *Applied Sciences*. 2022; 12(2):701. <https://doi.org/10.3390/app12020701>
- 54)Maggio, A. G., Shu, H. T., Laufer, B. I., Bi, C., Lai, Y., LaSalle, J. M., & Hu, V. W. (2022). Elevated exposures to persistent endocrine disrupting compounds impact the sperm methylome in regions associated with autism spectrum disorder. *Frontiers in genetics*, 13, 929471. <https://doi.org/10.3389/fgene.2022.929471>
- 55)Masini, E., Loi, E., Vega-Benedetti, A. F., Carta, M., Doneddu, G., Fadda, R., & Zavattari, P. (2020). An Overview of the Main Genetic, Epigenetic and Environmental Factors Involved in Autism Spectrum Disorder Focusing on Synaptic Activity. *International Journal of Molecular Sciences*, 21(21), Article 21. <https://doi.org/10.3390/ijms21218290>
- 56)McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: review and integration. *Briefings in bioinformatics*, 20(6), 2044–2054. <https://doi.org/10.1093/bib/bby067>
- 57)Melin, M., Carlsson, B., Anckarsater, H., Rastam, M., Betancur, C., Isaksson, A., Gillberg, C., & Dahl, N. (2006). Constitutional downregulation of SEMA5A expression in autism. *Neuropsychobiology*, 54(1), 64–69. <https://doi.org/10.1159/000096040>
- 58)Mulligan, K. A., & Cheyette, B. N. R. (2017). Neurodevelopmental Perspectives on Wnt Signaling in Psychiatry. *Molecular Neuropsychiatry*, 2(4), 219–246. <https://doi.org/10.1159/000453266>
- 59)Nakahara, K., Hamada, K., Tsuchida, T., Takasugi, N., Abiko, Y., Shien, K., Toyooka,

- S., Kumagai, Y., & Uehara, T. (2021). Covalent N-arylation by the pollutant 1,2-naphthoquinone activates the EGF receptor. *The Journal of biological chemistry*, 296, 100524. <https://doi.org/10.1016/j.jbc.2021.100524>
- 60) Negrete-Díaz, J. V., Falcón-Moya, R., & Rodríguez-Moreno, A. (2022). Kainate receptors: From synaptic activity to disease. *The FEBS Journal*, 289(17), 5074–5088. <https://doi.org/10.1111/febs.16081>
- 61) Nurmi, E. L., Dowd, M., Tadevosyan-Leyfer, O., Haines, J. L., Folstein, S. E., & Sutcliffe, J. S. (2003). Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(7), 856–863. <https://doi.org/10.1097/01.CHI.0000046868.56865.0F>
- 62) Onore, C. E., Nordahl, C. W., Young, G. S., Van de Water, J. A., Rogers, S. J., & Ashwood, P. (2012). Levels of soluble platelet endothelial cell adhesion molecule-1 and P-selectin are decreased in children with autism spectrum disorder. *Biological psychiatry*, 72(12), 1020–1025. <https://doi.org/10.1016/j.biopsych.2012.05.004>
- 63) Paik, S. J., Kim, D. J., & Jung, S. K. (2023). Preventive Effect of Pharmaceutical Phytochemicals Targeting the Src Family of Protein Tyrosine Kinases and Aryl Hydrocarbon Receptor on Environmental Stress-Induced Skin Disease. *International journal of molecular sciences*, 24(6), 5953. <https://doi.org/10.3390/ijms24065953>
- 64) Pang, S., Luo, Z., Dong, W., Gao, S., Chen, W., Liu, N., Zhang, X., Gao, X., Li, J., Gao, K., Shi, X., Guan, F., Zhang, L., & Zhang, L. (2022). Integrin β 1/FAK/SRC signal pathway is involved in autism spectrum disorder in Tspan7 knockout rats. *Life science alliance*, 6(3), e202201616. <https://doi.org/10.26508/lsa.202201616>
- 65) Parsa N. (2012). Environmental factors inducing human cancers. *Iranian journal of public health*, 41(11), 1–9.
- 66) Purcell, A. E., Jeon, O. H., Zimmerman, A. W., Blue, M. E., & Pevsner, J. (2001). Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. *Neurology*, 57(9), 1618–1628. <https://doi.org/10.1212/wnl.57.9.1618>
- 67) Raina, S. K., Chander, V., Bhardwaj, A. K., Kumar, D., Sharma, S., Kashyap, V., Singh, M., & Bhardwaj, A. (2017). Prevalence of Autism Spectrum Disorder among Rural, Urban, and Tribal Children (1-10 Years of Age). *Journal of neurosciences in rural practice*, 8(3), 368–374. https://doi.org/10.4103/jnrp.jnrp_329_16
- 68) Raman K. (2010). Construction and analysis of protein-protein interaction networks. *Automated experimentation*, 2(1), 2. <https://doi.org/10.1186/1759-4499-2-2>

- 69)Rhee, J., Park, K., Kim, K. C., Shin, C. Y., & Chung, C. (2018). Impaired Hippocampal Synaptic Plasticity and Enhanced Excitatory Transmission in a Novel Animal Model of Autism Spectrum Disorders with Telomerase Reverse Transcriptase Overexpression. *Molecules and Cells*, 41(5), 486–494. <https://doi.org/10.14348/molcells.2018.0145>
- 70)Rylaarsdam, L., & Guemez-Gamboa, A. (2019). Genetic Causes and Modifiers of Autism Spectrum Disorder. *Frontiers in cellular neuroscience*, 13, 385. <https://doi.org/10.3389/fncel.2019.00385>
- 71)Salari, N., Rasoulpoor, S., Rasoulpoor, S. et al. The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis. *Ital J Pediatr* 48, 112 (2022). <https://doi.org/10.1186/s13052-022-01310-w>
- 72)Schmunk, G., & Gargus, J. J. (2013). Channelopathy pathogenesis in autism spectrum disorders. *Frontiers in Genetics*, 4. <https://doi.org/10.3389/fgene.2013.00222>
- 73)Sean Davis and Paul S. Meltzer (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23(14): 1846-1847
- 74)Segers, V. F. M., Dugaucquier, L., Feyen, E., Shakeri, H., & De Keulenaer, G. W. (2020). The role of ErbB4 in cancer. *Cellular oncology (Dordrecht)*, 43(3), 335–352. <https://doi.org/10.1007/s13402-020-00499-4>
- 75)Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498– 2504.
- 76)Sordillo, J. E., Cardenas, A., Qi, C., Rifas-Shiman, S. L., Coull, B., Luttmann-Gibson, H., Schwartz, J., Kloog, I., Hivert, M. F., DeMeo, D. L., Baccarelli, A. A., Xu, C. J., Gehring, U., Vonk, J. M., Koppelman, G., Oken, E., & Gold, D. R. (2021). Residential PM2.5 exposure and the nasal methylome in children. *Environment international*, 153, 106505. <https://doi.org/10.1016/j.envint.2021.106505>
- 77)Subramanian, Aravind; Tamayo, Pablo; Mootha, Vamsi K.; Mukherjee, Sayan; Ebert, Benjamin L.; Gillette, Michael A.; Paulovich, Amanda; et al., (2005). "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression".
- 78)Turner, T. N., Hormozdiari, F., Duyzend, M. H., McClymont, S. A., Hook, P. W., Iossifov, I., Raja, A., Baker, C., Hoekzema, K., Stessman, H. A., Zody, M. C., Nelson, B. J., Huddleston, J., Sandstrom, R., Smith, J. D., Hanna, D., Swanson, J. M., Faustman, E. M., Bamshad, M. J., ... Eichler, E. E. (2016). Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA.

- The American Journal of Human Genetics, 98(1), 58–74.
<https://doi.org/10.1016/j.ajhg.2015.11.023>
- 79)Valente TW, Foreman RK. Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*. 1998;20:89–105. doi: 10.1016/S0378-8733(97)00007-5.
 - 80)Van den Veyver, I. B., & Zoghbi, H. Y. (2001). Mutations in the gene encoding methyl-CpG-binding protein 2 cause Rett syndrome. *Brain & development*, 23 Suppl 1, S147–S151. [https://doi.org/10.1016/s0387-7604\(01\)00376-x](https://doi.org/10.1016/s0387-7604(01)00376-x)
 - 81)Wang, J., Ma, B., Wang, J., Zhang, Z., & Chen, O. (2022). Global prevalence of autism spectrum disorder and its gastrointestinal symptoms: A systematic review and meta-analysis. *Frontiers in psychiatry*, 13, 963102. <https://doi.org/10.3389/fpsy.2022.963102>
 - 82) Waubant, E., Lucas, R., Mowry, E., Graves, J., Olsson, T., Alfredsson, L., & Langer-Gould, A. (2019). Environmental and genetic risk factors for MS: An integrated review. *Annals of Clinical and Translational Neurology*, 6(9), 1905–1922. <https://doi.org/10.1002/acn3.50862>
 - 83)Wei, H., Malik, M., Sheikh, A. M., Merz, G., Ted Brown, W., & Li, X. (2011). Abnormal cell properties and down-regulated FAK-Src complex signaling in B lymphoblasts of autistic subjects. *The American journal of pathology*, 179(1), 66–74. <https://doi.org/10.1016/j.ajpath.2011.03.034>
 - 84)William T. Barry, Andrew B. Nobel, and Fred Wright. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics*, 21(9):1943–1949, May 2005
 - 85)Wright, C., Shin, J. H., Rajpurohit, A., Deep-Soboslay, A., Collado-Torres, L., Brandon, N. J., Hyde, T. M., Kleinman, J. E., Jaffe, A. E., Cross, A. J., & Weinberger, D. R. (2017). Altered expression of histamine signaling genes in autism spectrum disorder. *Translational psychiatry*, 7(5), e1126. <https://doi.org/10.1038/tp.2017.87>
 - 86)Xie, L., McDaniel, M. J., Perszyk, R. E., Kim, S., Cappuccio, G., Shapiro, K. A., Muñoz-Cabello, B., Sanchez-Lara, P. A., Grand, K., Zhang, J., Nocilla, K. A., Sheikh, R., Armengol, L., Romano, R., Pierson, T. M., Yuan, H., Myers, S. J., & Traynelis, S. F. (2023). Functional effects of disease-associated variants reveal that the S1-M1 linker of the NMDA receptor critically controls channel opening. *Cellular and molecular life sciences : CMLS*, 80(4), 110. <https://doi.org/10.1007/s00018-023-04705-y>
 - 87)Yoav Benjamini and Yosef Hochberg, Controlling the False Discovery Rate: A

Practical and Powerful Approach to Multiple Testing, Vol. 57, No. 1 (1995), pp. 289-300.

- 88) Yu, D., Zhu, D., Wang, X., Li, B., Li, J., Lu, P., Ji, Y., & Wang, X. (2022). Prenatal di-(2-ethylhexyl) phthalate exposure induced myocardial cytotoxicity via the regulation of the NRG1-dependent ErbB2/ErbB4-PI3K/AKT signaling pathway in fetal mice. *Ecotoxicology and Environmental Safety*, 241, 113771. <https://doi.org/10.1016/j.ecoenv.2022.113771>