

Comparative Transcriptomics Data Profiling Reveals E2F Targets as an Important Biological Pathway Overexpressed in Intellectual Disability Disorder

Bioinformatics and Biology Insights
Volume 18: 1–8
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322231224665



Prekshi Garg¹, Farrukh Jamal² and Prachi Srivastava¹

¹Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, India. ²Department of Biochemistry, Dr Rammanohar Lohia Avadh University, Ayodhya, India.

ABSTRACT: Intellectual disability (ID) is an early childhood neurodevelopmental disorder that is characterized by impaired intellectual functioning and adaptive behavior. It is one of the major concerns in the field of neurodevelopmental disorders across the globe. Diversified approaches have been put forward to overcome this problem. Among all these approaches, high throughput transcriptomic analysis has taken an important dimension. The identification of genes causing ID rapidly increased over the past 3 to 5 years owing to the use of sophisticated high throughput sequencing platforms. Early monitoring and preventions are much important for such disorder as their progression occurs during fetal development. This study is an attempt to identify differentially expressed genes (DEGs) and upregulated biological processes involved in development of ID patients through comparative analysis of available transcriptomics data. A total of 7 transcriptomic studies were retrieved from National Center for Biotechnology Information (NCBI) and were subjected to quality check and trimming prior to alignment. The normalization and differential expression analysis were carried out using DESeq2 and EdgeR packages of Rstudio to identify DEGs in ID. In progression of the study, functional enrichment analysis of the results obtained from both DESeq2 and EdgeR was done using gene set enrichment analysis (GSEA) tool to identify major upregulated biological processes involved in ID. Our findings concluded that monitoring the level of E2F targets, estrogen, and genes related to oxidative phosphorylation, DNA repair, and glycolysis during the developmental stage of an individual can help in the early detection of ID disorder.

KEYWORDS: Intellectual disability, comparative transcriptome analysis, neurodevelopmental delay, differential gene expression analysis, functional enrichment

RECEIVED: April 4, 2023. **ACCEPTED:** December 15, 2023.

TYPE: Original Research Article

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Prachi Srivastava, Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh 226028, India. Email: psrivastava@amity.edu

Introduction

Intellectual disability (ID) is an early childhood neurodevelopmental disorder that is characterized by impaired intellectual functioning and adaptive behavior.¹ Its prevalence is estimated to be around 1% in the general population and is gradually increasing in its frequency.^{2,3} Basic symptoms of ID include greater difficulty in learning new things, understanding concepts, solving problems, concentrating, and remembering. Intellectual functioning in general refers to intelligence that includes mental activities such as problem-solving, logical reasoning, learning, and verbal skills. Intellectual functioning is measured in intelligence quotient (IQ). An individual with IQ <70 is known to suffer from ID. Based on IQ, ID can be divided into mild (IQ: 55-70), moderate (IQ: 40-55), severe (IQ: 25-40), and profound (IQ <25).⁴ Adaptive behavior refers to social, conceptual, and practical skills. Social skills include self-esteem, social responsibility, and interpersonal skills. Conceptual skills include an understanding of time, finance, and language. Practical skills include the ability to perform daily activities and interact with other people.⁵

Intellectual disability is grouped into 3 categories based on their inheritance pattern, namely autosomal recessive intellectual disability (ARID), autosomal dominant intellectual disability (ADID), and X-linked intellectual disability (XID).⁶ Autosomal recessive intellectual disability is a heterogeneous disorder that can be either syndromic or nonsyndromic. The

syndromic form of ID is characterized by intellectual problems occurring with a group of other phenotypic features whereas nonsyndromic ID is characterized by a lack of associated pathology.^{7,8} Autosomal dominant intellectual disability generally occurs when an individual carries 1 copy of a mutant allele and 1 normal allele on a gene. It is generally caused due to heterozygous mutations in reported genes and copy number variations (CNVs). X-linked intellectual disability is caused due to genes present on the X chromosome. About 10% to 12% of X chromosome genes are known to be linked with ID.⁶

Diagnosis of ID is usually possible through phenotypic symptoms such as delayed speech, hypotonia, and seizures. Intellectual disability can be caused due to various reasons including genetic, nongenetic, and environmental.⁹ The heterogeneity of ID makes genetic and clinical diagnosis more challenging.^{3,10-12} Although different studies are defining the causes and symptoms of this social trauma, there is a great need to decode the etiopathogenesis of ID as a disease. With the advent of next-generation sequencing (NGS) and transcriptome profiling and analysis, it is possible to investigate such complex human diseases at the molecular level. The use of RNA sequencing (RNA-Seq) analysis can help us identify essential outputs about ID. RNA-Seq is an advanced sequencing technology that has completely revolutionized transcriptome analysis. This technology allows the quantification of gene expression levels



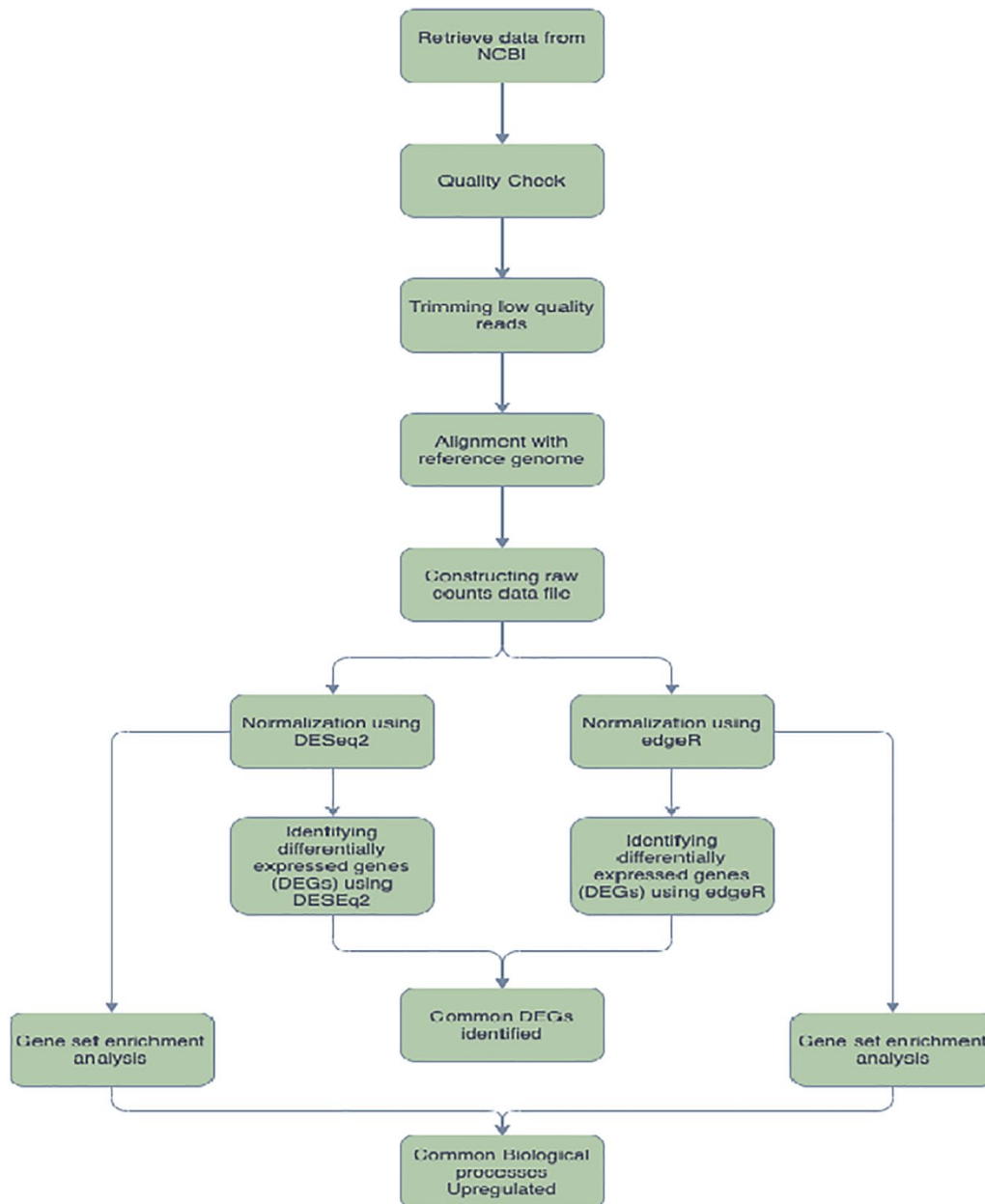


Figure 1. Flowchart illustrating the complete pipeline of DEG analysis. DEG indicates differentially expressed gene; NCBI, National Center for Biotechnology Information.

and allele-specific expressions in a single experiment. RNA-Seq analysis can help us identify novel genes, splice isoforms, and noncoding RNA at an unprecedented level.¹³ Therefore, in this study, we aim to identify differentially expressed genes (DEGs) for ID using available RNA-Seq data and further perform functional enrichment of such genes.

Materials and Methods

An intense literature survey was done to organize a well-defined pipeline for the study. The complete pipeline is illustrated in Figure 1.

Raw data retrieval and preprocessing

The RNA-Seq data from blood and skin cells of ID patients were downloaded from Gene Expression Omnibus (GEO),¹⁴

Sequence Read Archive (SRA),¹⁵ and BioProject¹⁶ (GSE77742, GSE74263, GSE90682, GSE98476, GSE108887, GSE145710, and PRJEB21964) of National Center for Biotechnology Information (NCBI) and were subsequently used for analysis in this study. The obtained data were generated on HiSeq using standard Illumina sequencing workflow. The low-quality bases and adaptor sequences were removed from the raw sequence through FastQC.¹⁷ The human reference genome hg38 (2013 release) was used to align the trimmed sequences of each sample.

Normalization and statistical analysis

The aligned sequences of each sample were normalized using the Trimmed Mean of M values (TMM) method of EdgeR,¹⁸ and the relative log₂ expression of the DESeq2 package¹⁹ in

Table 1. Upregulated differentially expressed genes in patients of ID obtained from DESeq2 and EdgeR when compared with control healthy samples.

STUDY ID	NUMBER OF SAMPLES	DESEQ2	EDGER	INTERSECTING UPREGULATED GENES
GSE77742	2 control, 1 patient	3933	802	799
GSE74263	2 control, 2 patient	1087	170	167
PRJEB21964	3 control, 1 patient	129	16	15
GSE90682	3 control, 7 patient	19	11	1
GSE98476 (mutated)	10 control, 6 patient	13	1	1
GSE98476 (hemizygous)	10 control, 2 patient	59	6	6
GSE108887 (CCNT2 mutation)	2 control, 2 patient	34	4	4
GSE108887 (CDK9 mutation)	2 control, 2 patient	51	5	2
GSE108887 (TAF2 mutation)	2 control, 2 patient	558	67	66
GSE145710	3 control, 6 patient	3690	687	648

Abbreviation: ID, intellectual disability.

R.²⁰ Factoextra package of R was used to construct Principal Component Analysis (PCA) Biplots of each data set to ensure minimum batch effect in this study. Both the normalized gene count files were subjected to the Benjamini-Hochberg algorithm for the DEG analysis at the threshold of 5% false-discovery rate (FDR). The recognized DEGs from DESeq2 and EdgeR were compared for the identification of common DEGs.

Statistical analysis using gene set enrichment analysis

The normalized count file of each study was further used for the identification of upregulated biological processes in the patient sample as compared with the normal sample using gene set enrichment analysis (GSEA).²¹ Gene set enrichment analysis is a computational methodology that tests the statistical significance of a priori-defined set of genes between 2 biological states. The normalized counts of each data were analyzed using the hallmark data set²² of the molecular signature database (MSigDB)²³ of GSEA. The GSEA method calculates an enrichment score (ES) based on weighted Kolmogorov-Smirnov statistics. The statistical significance of the calculated ES (nominal *P* value) is estimated based on an empirical phenotype-based permutation test. In this study, default statistical parameters were used, that is, the number of permutations was set to 1000, the signal-to-noise metric was used for ranking genes, and the minimum and maximum size of gene sets was 15 and 500.

Results

Data retrieved and preprocessing

The NCBI GEO and BioProject were the main sources of NGS data, from where the transcriptomic sequencing data of ID patients and control (GSE77742,²⁴ GSE74263, GSE90682,²⁵ GSE98476,²⁶ GSE108887, GSE145710,²⁷ and PRJEB21964)

were downloaded and subsequently used in this study. These included 26 control and 25 diseased samples. The low-quality sequences and adapter sequences were removed using trimmomatic and aligned with reference genome (hg38) for further analysis. Each sample of all the studies was independently run through FastQC to check the quality of the reads. The poor-quality reads and duplicated sequences were deleted. The alignment of all the sequences was done through the human genome (hg38, 2013 release). The aligned sequences were further used to count the expression of genes in each sample.

Normalization and statistical analysis

The results obtained from DESeq2 and EdgeR were compared for each study, and genes that were regarded as significantly expressed by both packages were further taken into consideration. Table 1 summarizes the results of DEG obtained from DESeq2 and EdgeR. The studies with multiple conditions were further divided into groups to avoid the normalization of data that could alter the results. Supplementary Figure 1 shows the results obtained from DESeq2 in the form of a heatmap, volcano plot, and MA (log ratio vs mean average) plot. Supplementary Figure 2 depicts the results obtained from EdgeR in the form of a heatmap, smear plot, and BCV (Biological coefficient of variant) plot. Supplementary Figure 5 shows the independence of patient and control samples in the form of PCA-Biplots. Supplemental materials 8 and 9 depict the results including *P* value, log2FoldChange, and FDR obtained through DESeq2 and EdgeR analysis.

Statistical analysis using gene set enrichment analysis

The normalized gene counts from both DESeq2 and edgeR were further used for GSEA. Each study into consideration

Table 2. Gene set enrichment analysis indicating the number of biological pathways upregulated in patients of ID.

STUDY ID	NUMBER OF BIOLOGICAL PROCESSES UPREGULATED USING DESEQ2 NORMALIZED FILE	NUMBER OF BIOLOGICAL PROCESSES UPREGULATED USING EDGER NORMALIZED FILE	NUMBER OF BIOLOGICAL PROCESSES FOUND COMMON IN BOTH THE STUDIES
GSE77742	19	8	5
GSE74263	14	2	2
PRJEB21964	31	27	11
GSE98476_Hemi	22	26	16
GSE98476_Mut	15	4	2
GSE108887_CCNT2	10	9	5
GSE108887_CDK9	30	33	13
GSE108887_TAF2	26	33	7
GSE90682	5	9	2
GSE145710	8	8	4

Abbreviation: ID, intellectual disability.

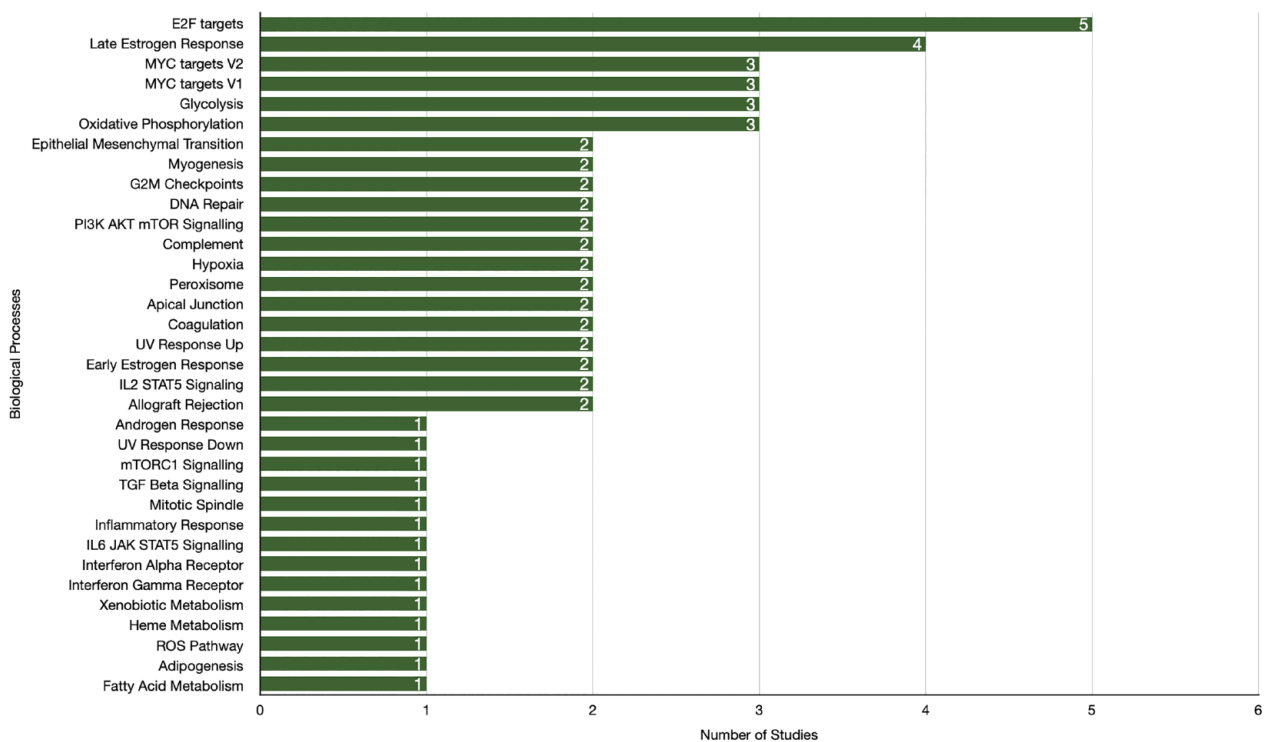


Figure 2. The major biological processes found to be upregulated in the patient sample. (AKT: Ak strain transforming; DNA: Deoxyribonucleic acid; G2M: Growth phase II and M phase checkpoint; IL6: Inflammatory cytokine interleukin-6; JAK: Janus kinase; mTOR: Mammalian target of rapamycin; MYC: Myelocytomatosis oncogene; PI3K: Phosphatidylinositol-3-kinase; ROS: Reactive oxygen species; STAT5: Signal transducer and activator of transcription 5; TGF: Transforming growth factor; UV: Ultraviolet).

was analyzed independently of the other. The role of biological processes which were found to be upregulated in the patient sample under both DESeq2 and EdgeR analysis was further analyzed. The number of permutations for the analysis was set to 1000. Weighted enrichment static and signal-to-noise metric were used for ranking genes. Table 2 summarizes the results

of the GSEA. Figure 2 illustrates the major biological processes and the number of studies in which they are found to be upregulated in the patient sample. Biological processes that were found to be upregulated in the patient sample in maximum studies, that is, E2F targets and late estrogen response are depicted through graphs in Supplementary Figures 3 and 4.

Table 3. Details of data used for the analysis of differentially expressed genes in epilepsy, bipolar disorder, dementia, and fragile X syndrome.

DISEASE	GEO (GENE EXPRESSION OMNIBUS) ID	NUMBER OF SAMPLES	UPREGULATED GENES OBTAINED THROUGH DESEQ2	UPREGULATED GENES OBTAINED THROUGH EDGER	COMMON UPREGULATED GENES
Bipolar disorder	GSE78936 ²⁸	16P, 12C	381	160	128
Dementia	GSE92340 ²⁹	3P, 4C	1679	377	0
Epilepsy	GSE190451	3P, 3C	2604	237	228
FXS	GSE76490 ³⁰	4P, 4C	56	4	2
ASD	GSE64018 ³¹	12P, 12C	62	10	0

Abbreviations: ASD: autism spectrum disorder; FXS: fragile X syndrome; ID, intellectual disability.

The dot plots of each data set representing the biological processes, the gene ratio, and FDR are given in Supplementary Figures 6 and 7.

Comparison of the results with other disorders

The similar studies were carried out on different disorders to identify whether the identified gene is specific for ID or not. Different disorders such as epilepsy, bipolar disorder, dementia, and fragile X syndrome (FXS) were taken into consideration for the comparison. The DEGs in autism spectrum disorder (ASD) were also analyzed following similar methodology because of its close link with ID disorder. The data used for different studies and the number of DEGs obtained using DESeq2 and EdgeR have been listed in Table 3 below. On comparing the results of all the diseases, there were no genes identified as commonly expressed in all 5 diseases. Only 5 genes, that is, BTBD1, DNAJA1, HSPA8, PLEKHB2, and G3BP2 were identified as the upregulated genes between bipolar disorder and epilepsy. The 2 genes, that is, NNAT and ZNF208 found to be upregulated in FXS were explicitly enriched in patients with FXS only. Figure 3 depicts the commonly expressed genes in different disorders in the form of a Venn diagram.

Discussion

Intellectual disability is one of the major issues of concern in the field of neurodevelopmental disorders across the globe. Since such disorders are caused during the developmental stages of a child, it is very important to monitor them at an early stage itself. High throughput sequencing put reveals various facts of genomic science underlying the disease. This gives enormous scope to the researcher to decode the mystery of the genome. Previous studies were based on gene findings and expression analysis at a limited level but the transcriptome data have given a new vision in the form of DEG analysis coupled with functional enrichment. This study signifies the identification of DEGs through comparative studies in the results of DESeq2 and EdgeR analysis. The GSEA was also carried out on both results to identify upregulated biological processes in the case of patients with ID.

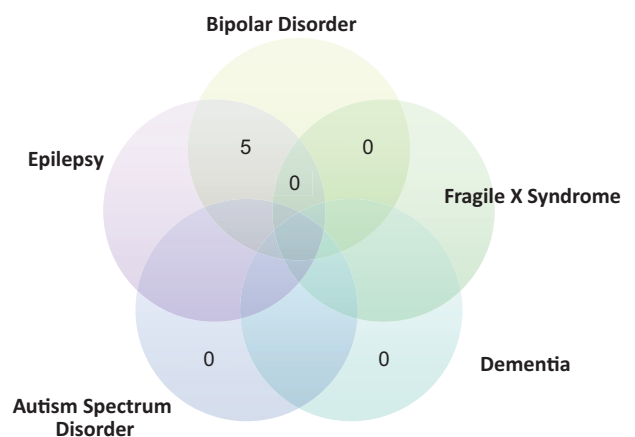


Figure 3. The upregulated genes expressed commonly in different disorders depicted in the form of a Venn diagram.

Our findings revealed a very significant output about the identification of related biological processes and pathways found to be involved in the progression of ID. It is well evident that E2F targets are strongly related to neurodevelopmental delay and other pathways such as glycolysis, oxidative phosphorylation, late estrogen response, DNA repair, and MYC (Myelocytomatosis oncogene) targets V1 and V2 also contribute to the onset of ID. E2F target genes are important genes that regulate the cell cycle.³² The cell cycle machinery is a pivotal regulator of brain development and function by influencing key cell fate decisions, typically via E2F transcription factor activity.³³⁻³⁶ pRb/E2f pathway is a cell cycle regulatory pathway that acts as an important effector of fate decisions in various cell types, including the brain.³⁷ Cell cycle proteins play an essential part in various cell fate decisions and, therefore, have pivotal roles in neurogenesis and brain development. Cell cycle dynamics strongly influence neural precursor cell (NPC) maintenance and neurogenesis, and gain- or loss-of-function studies have demonstrated key roles for cell cycle proteins, including the E2f family, in NPC fate decisions.³⁸⁻⁴⁰ Therefore, upregulation of E2F target genes can disrupt cell cycle machinery and lead to impaired development of the brain which causes ID. Similarly, MYC targets V1 and V2 also play an important role in the regulation of the cell cycle. The upregulation of MYC

targets in patient samples further ensures the role of cell cycle progression in neurodevelopmental delay and ID. Estrogen has the capability to impose neuroprotective actions.^{41,42} Estrogen signaling plays an important role in normal nervous system development. Estrogen signaling has also been seen to be associated with neurodegenerative disorders in females and neurodevelopmental disorders in males. Estrogen signaling ensures healthy brain development. Therefore, a late estrogen response in children can make them susceptible to neurodevelopmental delay that leads to ID in such individuals. Mutations in the genes involved in glycolysis can further upregulate glycolysis. Glycerol kinase is an important enzyme that links glycolysis, glycogenesis, and gluconeogenesis together.⁴³ In the case of glycerol kinase deficiency (GKD), glycerol accumulates in the bloodstream and can lead to an infantile or juvenile form of GKD. The infantile form is associated with severe developmental delay and juvenile patients are often symptomatic. Glycerol kinase deficiency in infants often leads to ID in such patients because of the close association between genes of GKD and ID.⁴⁴ Oxidative phosphorylation or mitochondrial biogenesis plays a key role in neurogenesis and neuroplasticity.⁴⁵ Dysfunctioning of respiratory chain complexes and ATP (Adenosine triphosphate) synthase in mitochondria leads to reduced ATP production and increased ROS (Reactive oxygen species) production leading to oxidative stress. This defect in mitochondrial energy production machinery plays an important role in the pathogenesis of numerous neurological disorders.^{46,47} This results in phenotypes such as developmental delay, muscle weakness, hypotonia,⁴⁸ and more that affects highly oxidative tissues such as the brain, heart, muscle, and kidney.^{49,50} Therefore, oxidative phosphorylation alters brain tissues leading to neurodevelopmental delay and ID.

Furthermore, the study identified IFI44L (interferon-induced protein 44 like) and IFIT3 (interferon-induced protein with tetratricopeptide repeats 3) as the significantly upregulated gene in most of the patients of ID. Interferon-induced proteins have been majorly studied for their role in immune response. However, the recent researches done in neuroscience indicate an active role of interferon in brain functions.⁵¹ The normal development of cerebral cortex depends on the migration of neural crest cell (NCC). The prolonged exposure or accumulation of interferon-induced proteins in brain leads to abnormal migration of NCC to brain leading to developmental delay and ID.⁵²

Therefore, our findings conclude that monitoring the level of E2F targets, estrogen, and genes related to oxidative phosphorylation, DNA repair, and glycolysis during the developmental stage of an individual can help in the early detection of ID as disease and other related neurodevelopmental disorders as well. Prevention is always better than cure so these early detections at the genetic or genomic level can further aid in proper and early diagnosis so that the right decision can be taken at the right time.

A similar approach was used by Trivedi et al,⁵³ in her recent study done on patients with ASD. Autism spectrum disorder is a neurodevelopmental disorder that is very closely linked to ID. Therefore, it is crucial to compare the results with ASD to analyze the explicit expression of the pathways in ID patients. The study identified CYC1 gene (cycle gene) as the common DEG in patients with ASD. CYC1 gene is an important regulatory gene in pathways such as ATP synthesis pathway, FAS signaling pathway, and Huntington's disease pathway. The gene is also known to play an important role in mitochondrial dysfunction and generation of allergic or inflammatory responses in ASD patients. The same methodology was run on different disorders as well to determine that the gene sets are enriched explicitly in patients with ID. It was a clear revelation that the gene sets obtained in different disorders were different and rarely matched with each other. Only 5 genes were seen to be expressed in both epilepsy and bipolar disorder patients, that is, BTBD1, DNAJA1, HSPA8, PLEKHB2, and G3BP2. These genes are known to be involved in exercise-induced circadian regulation,⁵⁴ methylation, and C3HC4-type RING finger domain binding. NNAT and ZNF208 were found to be upregulated in patients with FXS only. NNAT gene is known to be involved in brain development, protein lipoylation,⁵⁵ and positive regulation of insulin secretion.⁵⁶ ZNF208 is more actively involved in facilitating DNA binding,⁵⁶ zinc ion binding,⁵⁷ and metal ion binding. The gene set explicitly expressed in bipolar disorder is known to be involved in biological processes such as synaptic signaling, trans-synaptic signaling, chemical synaptic transmission, cell-cell signaling, and nervous system development. The gene set explicitly expressed in epilepsy patients is known to be involved in biological processes such as synapse pruning, integrin activation, cellular response to testosterone stimulus, regulation of L-glutamate import across plasma membrane, positive regulation of protein localization to early endosome, fatty acid degradation, and metabolism.

Thus, the transcriptomic data analysis reveals the overexpression of different biological pathways in different disorders.

Although the results are very much convincing and give much impact on findings of ID, yet the only limitation of the study lies that the available experimental data of ID from GEO are critically analyzed and well interpreted rather than own sequenced data. It is well established that *in silico* interpretations majorly give significant findings as this is the importance and cost effectiveness of computational experiments; hence, this limitation may also be overruled when results are in linearity with the study hypothesis.

Acknowledgements

The authors would like to acknowledge the support provided by Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Lucknow Campus for the completion of the study. They would also acknowledge all individuals who supported

their work directly or indirectly. They gratefully acknowledge the support provided by Bioinformatics tools, software and databases in the work.

Author Contributions

Prekshi Garg conceived, designed, and performed the experiments, and prepared figures, tables, and the original draft. Farrukh Jamal reviewed the draft and finalized the manuscript. Prachi Srivastava conceived and designed the experiment, and reviewed and finalized the final draft.

Availability of Data and Materials

All the supporting data relevant to the article are available at the official website of NCBI (www.ncbi.nlm.nih.gov).

Ethics Approval and Consent to Participate

Not applicable.

Patient Consent for Publication

Not applicable.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- The American Association on intellectual and developmental disabilities. *Intellect Dev Disabil*. 2010;48:307-309. doi:10.1352/1934-9556-48.4.307
- Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Res Dev Disabil*. 2011;32:419-436.
- Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet*. 2016;17:9-18.
- Wieczorek D. Autosomal dominant intellectual disability. *Med Genet*. 2018;30:318-322. doi:10.1007/s11825-018-0206-2
- Lee K, Cascella M, Marwaha R. *Intellectual Disability*. StatPearls Publishing; 2020.
- Ilyas M, Mir A, Efthymiou S, Houlden H. The genetics of intellectual disability: advancing technology and gene editing. *F1000Res*. 2020;9:F1000 Faculty Rev-22. doi:10.12688/f1000research.16315.1
- Marangi G, Leuzzi V, Manti F, et al. TRAPPC9-related autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. *Eur J Hum Genet*. 2013;21:229-232.
- You J, Sobreira NL, Gable DL, et al. A syndromic intellectual disability disorder caused by variants in TEO2, a gene encoding a component of the TTT complex. *Am J Hum Genet*. 2016;98:909-918.
- Karam SM, Barros AJ, Matijasevich A, et al. Intellectual disability in a birth cohort: prevalence, etiology, and determinants at the age of 4 years. *Public Health Genomics*. 2016;19:290-297.
- Grozeva D, Carss K, Spasic-Boskovic O, et al. Targeted next-generation sequencing analysis of 1,000 individuals with intellectual disability. *Hum Mutat*. 2015;36:1197-1204.
- Heslop P, Blair PS, Fleming P, Hoghton M, Marriott A, Russ L. The confidential inquiry into premature deaths of people with intellectual disabilities in the UK: a population-based study. *Lancet*. 2014;383:889-895.
- Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. *N Engl J Med*. 2012;366:733-743.
- Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome profiling in human diseases: new advances and perspectives. *Int J Mol Sci*. 2017;18:1652. doi:10.3390/ijms18081652
- Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol*. 2016;1418:93-110. doi:10.1007/978-1-4939-3578-9_5
- Leinonen R, Sugawara H, Shumway M; International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res*. 2011;39:D19-D21. doi:10.1093/nar/gkq1019
- Barrett T, Clark K, Gevorgyan R, et al. BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res*. 2012;40:D57-D63. doi:10.1093/nar/gkr1163
- Andrews S. FastQC a quality control tool for high throughput sequence data, n.d. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40:4288-4297. doi:10.1093/nar/gks042
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi:10.1186/s13059-014-0550-8
- RStudio Team. *RStudio: Integrated Development for R*. RStudio, PBC; 2020. <http://www.rstudio.com/>
- Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139-140. doi:10.1093/bioinformatics/btp616
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739-1740. doi:10.1093/bioinformatics/btr260
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545-15550. doi:10.1073/pnas.0506580102
- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417-425. doi:10.1016/j.cels.2015.12.004
- Maduro V, Pusey BN, Cherukuri PF, et al. Complex translocation disrupting TCF4 and altering TCF4 isoform expression segregates as mild autosomal dominant intellectual disability. *Orphanet J Rare Dis*. 2016;11:62.
- Harms FL, Girisha KM, Hardigan AA, et al. Mutations in EBF3 disturb transcriptional profiles and cause intellectual disability, ataxia, and facial dysmorphism. *Am J Hum Genet*. 2017;100:117-127.
- Gabriele M, Vulto-van Silfhout AT, Germain PL, et al. YY1 haploinsufficiency causes an intellectual disability syndrome featuring transcriptional and chromatin dysfunction. *Am J Hum Genet*. 2017;100:907-925.
- Hu J, Xu J, Pang L, et al. Systematically characterizing dysfunctional long intergenic non-coding RNAs in multiple brain regions of major psychosis. *Oncotarget*. 2016;7:71087-71098.
- Zhang Y, Schmid B, Nikolaisen NK, et al. Patient iPSC-derived neurons for disease modeling of frontotemporal dementia with mutation in CHMP2B. *Stem Cell Reports*. 2017;8:648-658.
- Lu P, Chen X, Feng Y, et al. Integrated transcriptome analysis of human iPSC cells derived from a fragile X syndrome patient during neuronal differentiation. *Sci China Life Sci*. 2016;59:1093-1105.
- Irimia M, Weatheritt RJ, Ellis JD, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*. 2014;159:1511-1523.
- Lee YR, Khan K, Armfield-Uhas K, et al. Mutations in FAM50A suggest that Armfield XLID syndrome is a spliceosomopathy. *Nat Commun*. 2020;11:3698. doi:10.1038/s41467-020-17452-6. PMID
- Subtil-Rodríguez A, Vázquez-Chávez E, Ceballos-Chávez M, et al. The chromatin remodeller CHD8 is required for E2F-dependent transcription activation of S-phase genes. *Nucleic Acids Res*. 2014;42:2185-2196.
- Julian LM, Vandenbosch R, Pakenham CA, et al. Opposing regulation of Sox2 by cell-cycle effectors E2f3a and E2f3b in neural stem cells. *Stem Cell*. 2013;12:440-452.
- Andrusiak MG, McClellan KA, Dugal-Tessier D, et al. Rb/E2F regulates expression of neogenin during neuronal migration. *Mol Cell Biol*. 2011;31:238-247.
- Herrup K. Post-mitotic role of the cell cycle machinery. *Curr Opin Cell Biol*. 2013;25:711-716.
- Chen D, Chen Y, Forrest D, Bremner R. E2f2 induces cone photoreceptor apoptosis independent of E2f1 and E2f3. *Cell Death Differ*. 2013;20:931-940.
- Julian LM, Liu Y, Pakenham CA, et al. Tissue-specific targeting of cell fate regulatory genes by E2f factors. *Cell Death Differ*. 2016;23:565-575. doi:10.1038/cdd.2015.36
- Arai Y, Pulvers JN, Haffner C, et al. Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat Commun*. 2011;2:154.
- Lim S, Kaldis P. Loss of Cdk2 and Cdk4 induces a switch from proliferation to differentiation in neural. *Stem Cells*. 2012;30:1509-1520.
- Svoboda DS, Paquin A, Park DS, Slack RS. Pocket proteins pRb and p107 are required for cortical lamination independent of apoptosis. *Dev Biol*. 2013;384:101-113.
- Gillies GE, McArthur S. Estrogen actions in the brain and the basis for differential action in men and women: a case for sex-specific medicines. *Pharmacol Rev*. 2010;62:155-198.

43. Pinares-García P, Stratikopoulos M, Zagato A, Loke H, Lee J. Sex: a significant risk factor for neurodevelopmental and neurodegenerative disorders. *Brain Sci.* 2018;8:154. doi:10.3390/brainsci8080154
44. MacLennan NK, Rahib L, Shin C, et al. Targeted disruption of glycerol kinase gene in mice: expression analysis in liver shows alterations in network partners related to glycerol kinase activity. *Hum Mol Genet.* 2006;15:405-415. doi:10.1093/hmg/ddi457
45. Wamelink MMC, Valayannopoulos V, Garavaglia B. Disorders of glycolysis and the pentose phosphate pathway. In: Saudubray JM, Baumgartner M, Walter J, eds *Inborn Metabolic Diseases.* Springer; 2016:149-160. doi:10.1007/978-3-662-49771-5_7
46. Valenti D, de Bari L, De Filippis B, Henrion-Caude A, Vacca RA. Mitochondrial dysfunction as a central actor in intellectual disability-related diseases: an overview of Down syndrome, autism, Fragile X and Rett syndrome. *Neurosci Biobehav Rev.* 2014;46:202-217. doi:10.1016/j.neubiorev.2014.01.012
47. Sai Y, Zou Z, Peng K, Dong Z. The Parkinson's disease-related genes act in mitochondrial homeostasis. *Neurosci Biobehav Rev.* 2012;36:2034-2043.
48. Breuer ME, Koopman WJ, Koene S, et al. The role of mitochondrial OXPHOS dysfunction in the development of neurologic diseases. *Neurobiol Dis.* 2013;51:27-34.
49. Gibson GE, Starkov A, Blass JP, Ratan RR, Beal MF. Cause and consequence: mitochondrial dysfunction initiates and propagates neuronal dysfunction, neuronal death and behavioral abnormalities in age-associated neurodegenerative diseases. *Biochim Biophys Acta.* 2010;1802:122-134.
50. Wallace DC, Fan W. Energetics, epigenetics, mitochondrial genetics. *Mitochondrion.* 2010;10:12-31.
51. Bhat H, Lang KS, Hardt C, Lang J. Interferon in the CNS. *Neurosignals.* 2019;27:44-53.
52. Aigner L, Uyanik G, Couillard-Despres S, et al. Somatic mosaicism and variable penetrance in doublecortin-associated migration disorders. *Neurology.* 2003;60:329-332.
53. Trivedi P, Pandey M, Kumar Rai P, Singh P, Srivastava P. A meta-analysis of differentially expressed and regulatory genes with their functional enrichment analysis for brain transcriptome data in autism spectrum disorder. *J Biomol Struct Dyn.* 2023;41:9382-9388.
54. Wolfen M, Klatt D, Salybekov AA, et al. Hematopoietic stem-cell senescence and myocardial repair—coronary artery disease genotype/phenotype analysis of post-MI myocardial regeneration response induced by CABG/CD133+ bone marrow hematopoietic stem cell treatment in RCT PERFECT Phase 3. *EBioMedicine.* 2020;57:102862.
55. Dou D, Joseph R. Cloning of human neuronatin gene and its localization to chromosome-20q 11.2-12: the deduced protein is a novel "proteolipid." *Brain Res.* 1996;723:8-22.
56. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform.* 2011;12:449-462.
57. Eichler EE, Hoffman SM, Adamson AA, et al. Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res.* 1998;8:791-808.