scientific reports

OPEN

Identification of cell specific biomarkers for intellectual disability via single cell RNA sequencing and transcriptomic bioinformatics approaches

Kh Mujahidul Islam[1,2,7], Md Habibur Rahman[1,2,7 ✉], Md. Arju Hossain[3], Mahfuj Khan[1], Md. Mizanur Rahman[1], Prosenjit Saha Apu[1], Md Imran Hasan[4], Siddique Akber Ansari[5] & Mahammad Humayoo[6 ✉]

Limitations in cognitive functioning and adaptive behavior are hallmarks of Intellectual Disability (ID), a neurodevelopmental disease. Specific genetic disorders that result in ID can also have immune system anomalies, such as changes in T (CD4+ and CD8+) cell activity. This work aimed to compare single-cell RNA-sequencing (scRNA-seq) and transcriptome data to find biomarkers linked to T cells that could potentially be utilized for the diagnosis and assessment of ID. After integrating genes and performing a comparative analysis 196 genes were identified as differentially expressed genes (DEGs). Furthermore, the DAVID online platform and FunRich software were utilized to detect signal transduction and translation, immune response, MHC (Major Histocompatibility Complex) class II, antigen processing and presentation, allograft rejection and important pathways of type I diabetes mellitus. In this investigation, six ribosomal proteins (RPS27A, RPS21, RPS18, RPS7, RPS5, and RPL9) have been identified as the hub genes of ID from PPI. Additionally, eleven topological algorithms discovered only one hub protein, namely RPS27A from the protein-protein interaction (PPI) network. Through the analysis of the regulatory network, we have identified several crucial transcriptional factors (TFs) including FOXC1, FOXL1, and GATA2; microRNAs such as mir-92a-3p, and mir-16-5p were investigated by procedural data analysis. This study used scRNA-seq and transcriptomics data analysis to define unique biomarkers associated with T cell types throughout the progression of ID. Ongoing research on the activity of ID genes is contributing to a greater understanding of the pathophysiology of ID and will become more scientific and research-based in future.

Keywords Intellectual disability, Single-cell RNA-sequencing, T cell, Biomarkers, Hub genes, RPS27A, And FOXL1

Intellectual disability (ID) is defined by major obstacles in intellectual and adaptive skills that begin throughout stages of development[1]. Intellectual disability can be distinguished primarily by deficits in cognitive as well as adaptive operation; its dominance is estimated to range from 1 to 3% of the population with geographical variations[2]. This particular syndrome is commonly observed, with a prevalence of approximately 1.5% in Western countries, and potentially increasing to 4% in socioeconomically disadvantaged regions of the world[3]. The average occurrences of ID in Western nations are estimated between 1.5% and 2%, with a further 0.3–0.5% experiencing severe impairment, characterized by IQ levels below 50[4].

[1]Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh. [2]Center for Advanced Bioinformatics and Artificial Intelligence Research, Islamic University, Kushtia 7003, Bangladesh. [3]Department of Microbiology, Primeasia University, Banani, Dhaka 1213, Bangladesh. [4]Department of Life Science, Texas A&M University-San Antonio, San Antonio, TX 78224, USA. [5]Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, P.O Box 2457, 11451 Riyadh, Saudi Arabia. [6]School of Engineering, Pokhara University, 427, Lekhnath, Kaski, Nepal. [7]Kh Mujahidul Islam and Md Habibur Rahman contributed equally to this work and shared first authorship. ✉email: habib@iu.ac.bd; mahammad.humayoo@pu.edu.np

The development and transmission of synapses is one of the biological functions that is frequently discussed. Therefore, it is evident that variations in genes connected to ID affect a variety of biological processes[5]. Centrosome function, protein modification, chromatin remodeling, transcriptional and translational regulation, and the development of neural and supporting nervous system cells are some of these activities[6,7]. On the other hand, severe variants are believed to be the result of specific genetic reasons, chromosomal anomalies, or deficiencies in single genes. Considerable advancements have been achieved in understanding the genetic elements that produce severe ID over recent years. Approximately 15% of cases are caused by chromosomal abnormalities that are cytogenetically detectable[4]. There is no causal relationship that can be established between T cells and intellectual disability. However, people who have specific genetic disorders that result in intellectual disability can also have immune system anomalies, such as changes in T cell activity[8]. The vulnerability of an individual to infections and their ability to combat them may be influenced by the function of T cells. Altering CD8+T cell expressing activation markers, autism patients have dramatically altered adaptive cellular immune function, which may be a reflection of behavioral abnormalities, developmental disorders, and defective immunological activation[9].

Mild types of ID are thought to be on the low end of the normal range for IQ, which is the consequence of numerous genetic as well as nongenetic elements. As the list of genes related to ID has expanded, numerous scientists have searched for patterns in the roles that these genes encode. With hundreds of genes currently recognized to participate in the phenotype, there is a great deal of genetic variability associated with ID[10]. Several academic articles have revealed biomarkers associated with ID. Notable X-linked genes include MECP2, which was first associated with Rett syndrome and is now responsible for several male and female-specific ID symptoms[11]. There have since been two reports of mutations in the DYNC1H1 gene in other ID patients[12]. The pathophysiology of various neuropsychiatric traits is also influenced by genes related to ID. Even with the latest advancements in gene discovery, only a small percentage of ID cases are still explainable. Significant clinical and genetic heterogeneity made large-scale genetic research difficult. Through computational analysis, a novel biomarker discovery is urgently needed for a better understanding of the molecular mechanisms of ID.

RNA sequencing and next-generation sequencing (NGS) has contributed to the recent rapid progress in differential gene identification such as ID. Some recent studies[13,14] have used single-cell RNA sequencing (scRNA-seq) to determine disease pathogenesis and prognosis and have been commonly used to identify new biomarkers. We used scRNA-seq in this study to investigate transcriptional patterns at the level of gene expression[15]. Our objective was to identify key genes (KG), key microRNAs (miRNAs), and key transcriptional factors (TFs) that might be utilized for the personalized diagnosis and prognosis of ID. Therefore, the use of bioinformatics analysis to identify significant genes, miRNAs, TFs, and associated signaling pathways in the context of ID holds great potential for advancing future research in this field.

Results

Processing of collected datasets

After reading the raw dataset, we created a Seurat object so that it could be further examined. We used initial quality control during the data read process to eliminate low quality cells with less than 200 expressed features (genes) and lowly expressed features seen in less than three cells. We removed zero count instances and selected 10,000 read counts from the complete gene expression matrix to illustrate gene expression patterns prior to and during normalization. Approximately 2,000 significant features are usually found by Seurat for our dataset. We have identified 28 clusters employing the KNN algorithm. After cluster analysis, 28 scRNA-seq data clusters (have been annotated using reference-based annotation and represented in an UMAP plot (Fig. 1) Then five different sets of cell markers including T cells, B cells, natural killer cells, monocytes, and dendritic cells. Those five cell markers have provided multiple subsets. In addition, Supplementary Table S1 displays the clustering information for 28 detected and annotated clusters. We have merged only seven groups (0, 1, 4, 6, 8, 9, and 10) of T cell related marker including Th1 cells (CD4 + Tcell), Non-Vd2 gd T cells, Mait cells, and Th1 cells.

Statistical analysis and identification of shared DEGs between scRNA-seq and bulk RNA-seq datasets

By performing several statistical operations, we have found 1318, 1417, 1223, 1376, 1055, 1420, and 1246 DEGs from 0, 1, 4, 6, 8, 9, and 10 number of clusters, respectively. After removing duplicate and merging all identified DEGs from seven T cell clusters, we have obtained 3510 unique individual DEGs in ID patients. On the other hand, analyzing RNA-seq dataset (GSE46831) through the GREIN database, we identified 3459 significant DEGs of ID applying a cut-off range ($P < 0.05$, and $|log2FC| > 1$). Performing cross-match analysis between the merged DEGs of the T cells marker of scRNA-seq and the RNASeq datasets (GSE46831), total 196 shared DEGs were found for further analysis (Fig. 2). Among 196 shared DEGs, 102 genes were up-regulated, and 89 genes were down-regulated, showed in Table S2. About five DEGs were found common in both up and down-regulated (HLA-DQA1, HLA-DRA, HLA-DRB1, CSNK2B, PPP1R18).

Functional pathways enrichment analysis

For the gene ontology (GO) pathways, a common 196 DEGs were used in the FunRich software. In the biological process (BP), both signal transduction and translation are enriched with 15.7% genes, cytoplasmic translation with 15.2%, and immune response with 9% genes. In the cellular component (CC), 27.7% of genes are enriched with extracellular vesicular exosome, 13.6% with the cytosolic ribosome, 12% with focal adhesion and 3.8% with the MHC class II protein complex. In molecular function (MF), 84.4% of genes are enriched with protein binding, 20.6% of genes are enriched with RNA binding and 15.6% with the structural constituent of the ribosome (Fig. 3).
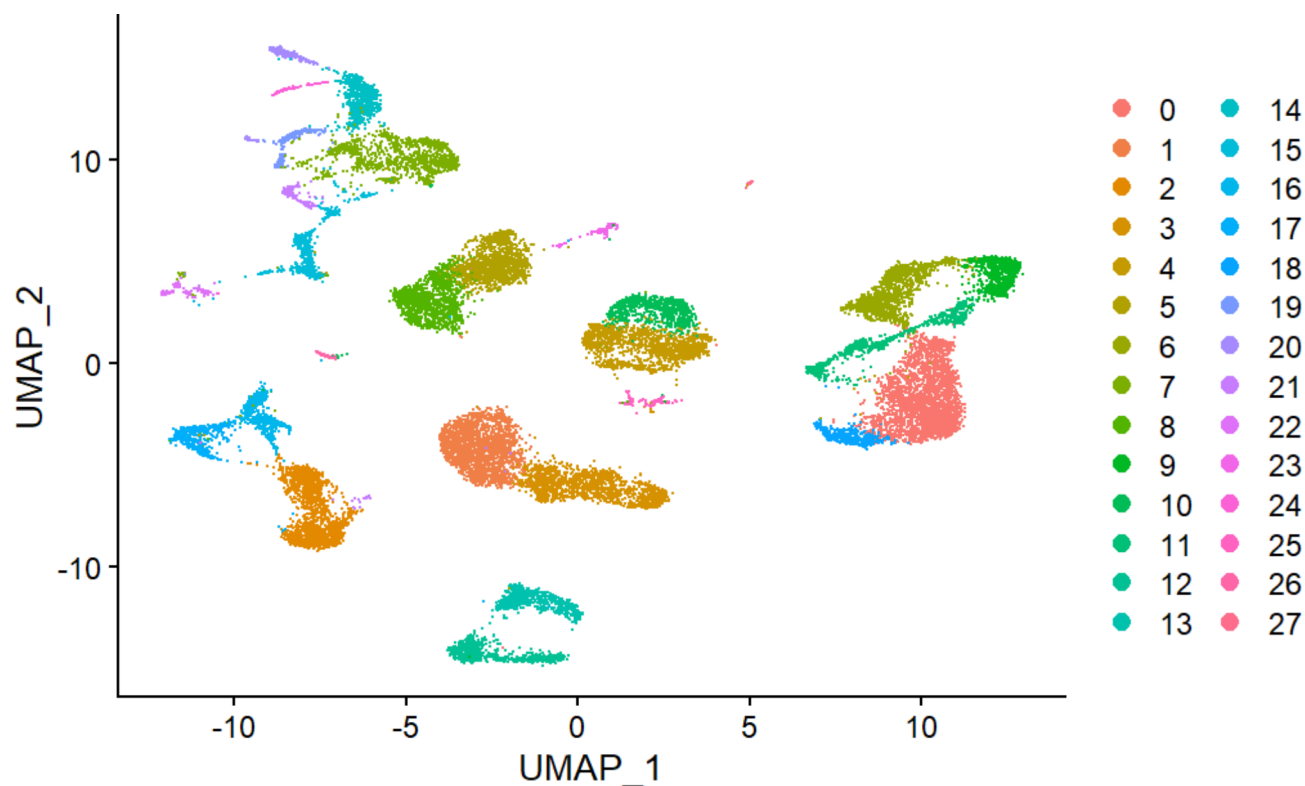
Fig. 1. UMAP shows the 28 clusters that were found. A significant number of related DEGs are shown by the region of the colors and various colors representing separate clusters. Generated using ggplot2 (v 3.5.0) in R (v 4.3.1).

ACTN4, ASAH1, ATP6V0B, B3GNT3, BANK1, BASP1, BAZ2B, BHLHE40, CD46, CD79A, CD79B, CFLAR, CSNK2B, DDX3X, DUSP1, EBF1, EVI2B, FAM129A, FBXL5, FBXO7, FOSL2, GREM1, H3F3A, HLA-A, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DRA, HLA-DRB1, IFITM2, IGF2R, IQGAP1, ITCH, KCNJ3, KIAA1551, KLRD1, LITAF, LYN, LYST, MAP3K2, MS4A1, MTRNR2L8,, NEAT1, NEGR1, NHSL2, OSBPL8, PAX5, PHGDH, PJA2, PPP1R18, PREX1, PRF1, PRKAA2, PRR26, PTEN, PXN, RGS2, RPL9 ,RPL26, RPL35A, RPL39, RPS10, RPS27, RPS29, RPS7, RPS9, S100A11, SLC2A3, SLCO5A1, SMCHD1, SORL1, STEAP4, SYTL3, TALDO1, TAP1, TGFBR3, TMEM154, TMX4, TRIM22, TYMP, VASP, XIST, CNNM1, COX7C, CST7, DDX52, DHRS7, FCGR3A, FOS, GPR85, HINT1, HLA-F, HSH2D, JAML, MBP, MIDN, OCIAD2, PTPRE, RAB11FIP1, RPL23, RPL27, RPL31, RPL34, RPL35, RPL41, RPS15A, RPS17, RPS18, RPS20, RPS21, RPS25, RPS27A, RPS3A, RPS5, RPS6, RPS8, SMAP2, SNRPD2, SOCS3, SOD2, TPT1, USP15, VCAN, VMP1, GABRB1, GABRR3, MKRN3, MRPL3, NFIB, RNF24, RPAP2, RPL36A, SPARCL1, IKBIP, LRRN3, NFAM1, NPAS3, PASK, RPS6KA5, TUBA1A, ABCA8, ATF6B, CD52, NABP1, PROX1, CEP63, COL21A1, ENOX1, FAM118A, IL6R, INTU .
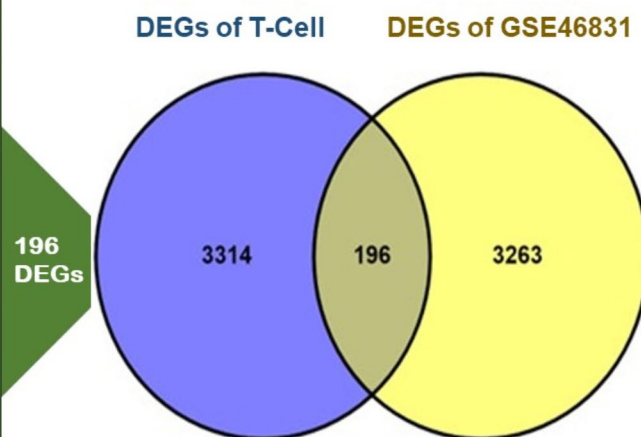
Fig. 2. The Venn diagram represents the list of 196 common DEGs between T-cell clusters of scRNA-seq and bulk RNA-seq (GSE46831) datasets; generated using Venny (v2.0.1).
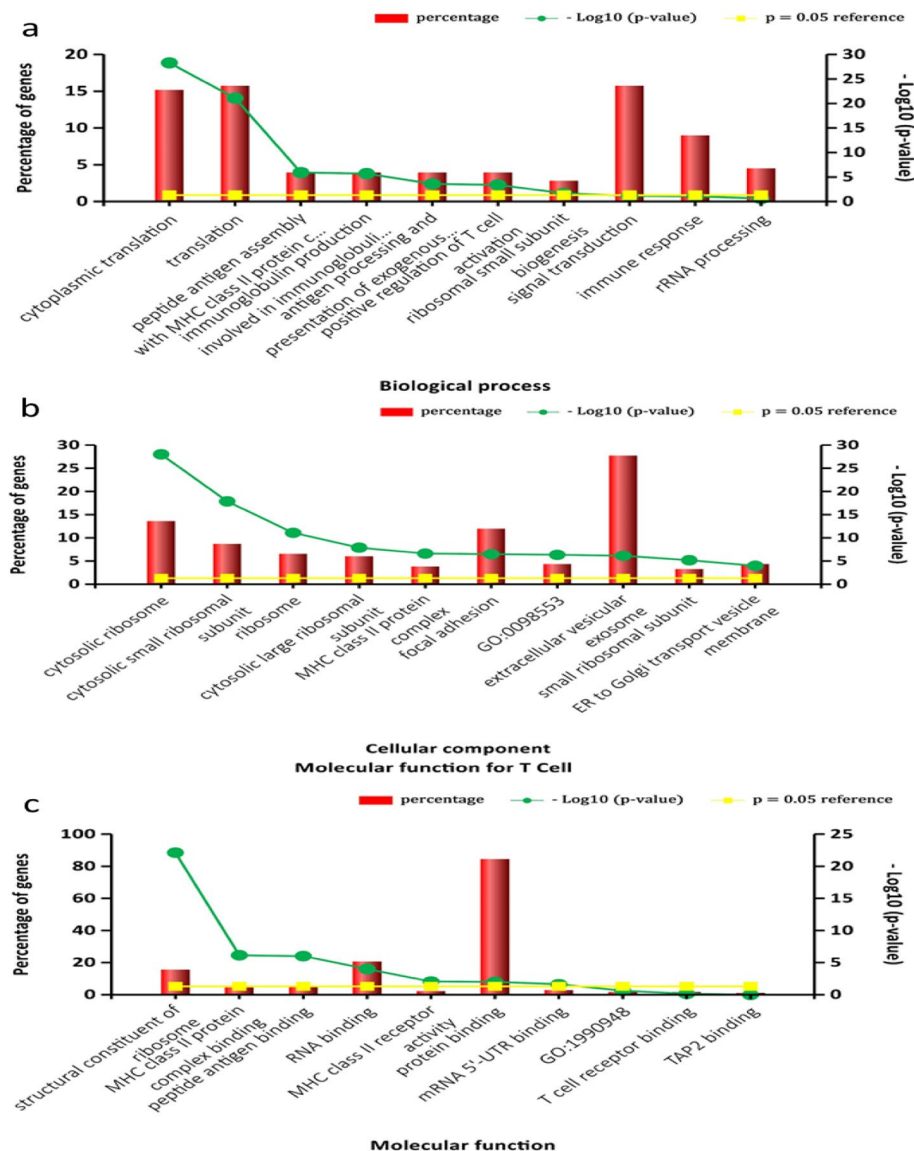
Fig. 3. Bar diagram, constructed with Funrich (v 3.4.1), illustrates the analysis of significant gene ontology (GO) pathways of ID. Based on the p-value (< 0.05), the top 10 pathways were included in the (**A**) biological process (BP), (**B**) cellular components (CC) and (**C**) molecular functions (MF) pathways, respectively.

We have operated KEGG, Reactome, WiKi, and BioCarta for metabolic pathway enrichment analysis. Based on the p-value (cut-off value < 0.05) and the enrichment score; the most important pathways were taken in this study. In KEGG analysis, Fig. 4 shows that the main enriched pathways are allograft rejection (21%), type I diabetes mellitus (18%), and graft-versus-host disease (20%). The Reactome database shows the most enriched pathways for shared DEGs of T cells including viral mRNA translation (24.32%), eukaryotic translation elongation (24.32%), and peptide chain elongation (24.15%) as the most enriched pathways. In WiKi pathways, cytoplasmic ribosomal proteins (18.77%), allograft rejection (7.47%), Ebola virus infection in host cells, and the B-cell receptor signaling pathway are some of the most enriched pathways. In BioCarta, the most enriched pathways are antigen processing and presentation (23.05%), and BCR signaling pathway (5.97%) (Fig. 4).

Identification of the hub protein from protein-protein interaction (PPI) network

In STRING, 196 shared DEGs of T cells were used in the PPI network. The analysis involves 191 nodes and 376 edges with an average degree of node of 3.94 with enrichment p-value < 1.0e-16. We analyze the physical sub-network under the confidence score 700 (high confidence) represented in Fig. 5. Based on the score, the most enriched 15 genes were selected from each of eleven topological methods (betweenness, stress, bottleneck, eccentricity, radiality, EPC, MNC, closeness, degree, DMNC, and MCC) in the cytoHubba plugin of Cytoscape. Using these genes, we created an upset plot to represent the most significant hub genes (Fig. 6). We extracted the hub genes that cover seven or more methods had represented them in the Upset plot. Therefore, we found 6 genes that are considered key hub genes (KGs) for T cells, such as RPS27A, RPS18, RPS5, RPS7, RPS21, and RPL9.
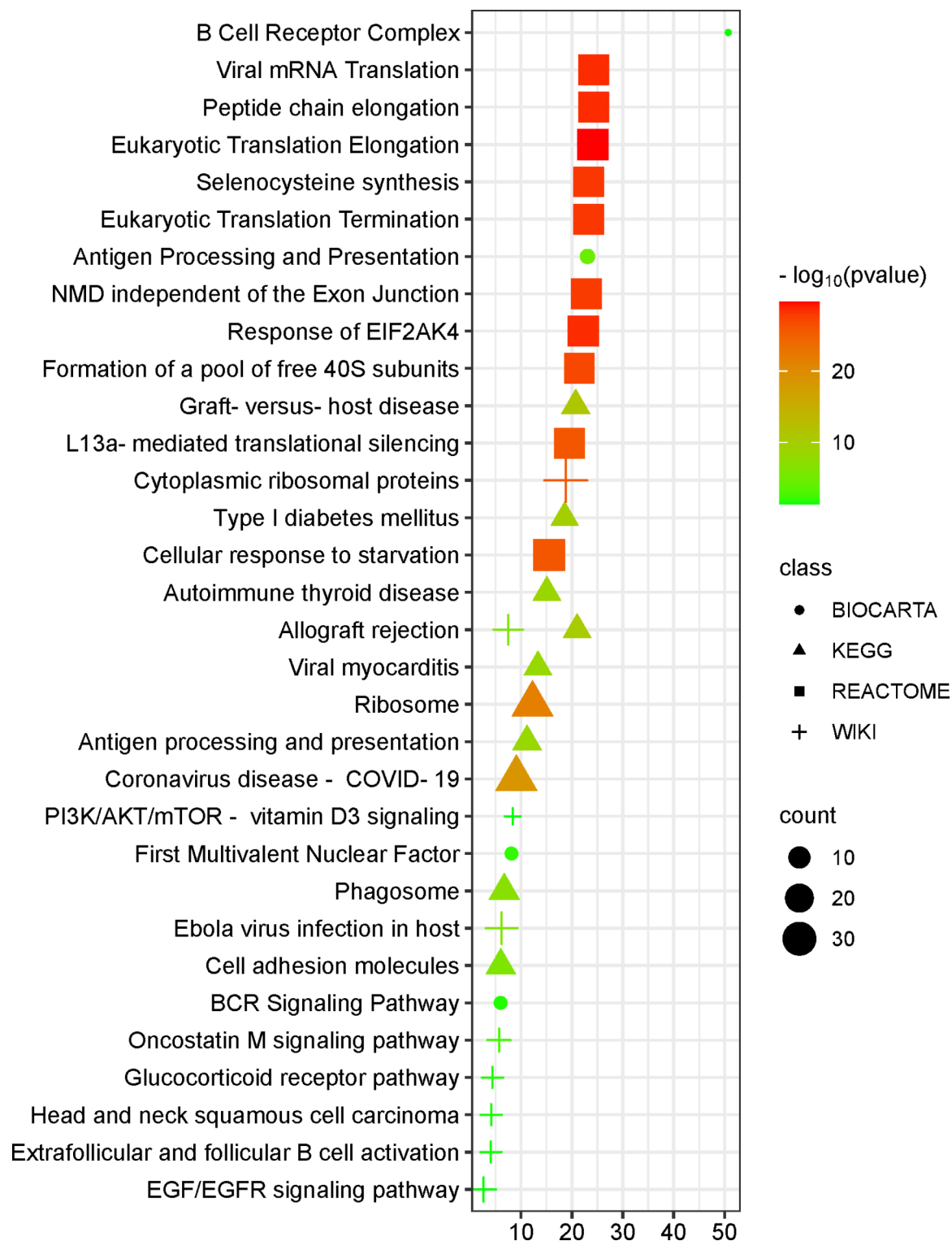
Fig. 4. The bubble graph represents the significant metabolic pathways of ID. Based on the p-value (< 0.05), the top 10 pathways, were included in the KEGG, Reactome, WiKi and BioCarta databases, respectively; where databases were accesses through DAVID (v 6.0) and visualized using the SRplot web server.
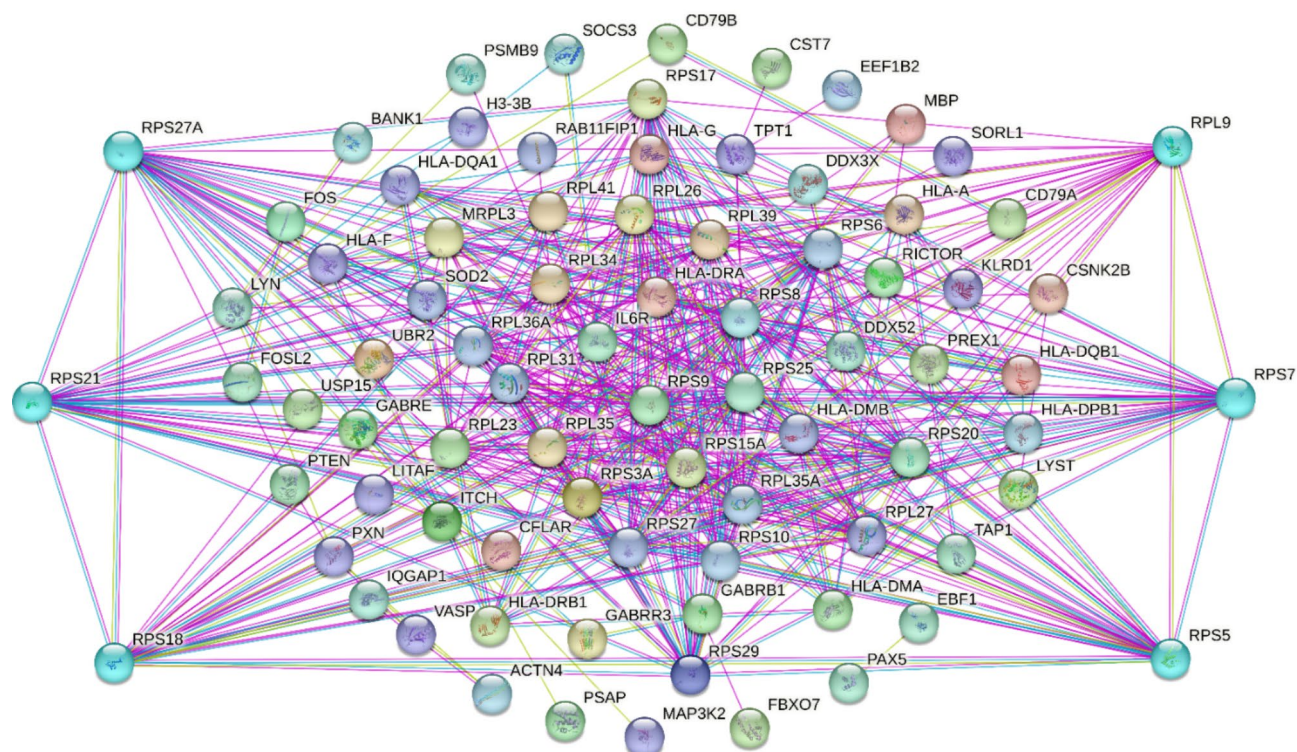
Fig. 5. Visualization of the PPI network of 196 DEGs, where the sky-blue color indicates the hub genes, the network nodes symbolize target proteins, while the edges denote the relationships between proteins. STRING (v11.0) was employed to construct the network and Cytoscape (v3.10.1) for visualization.
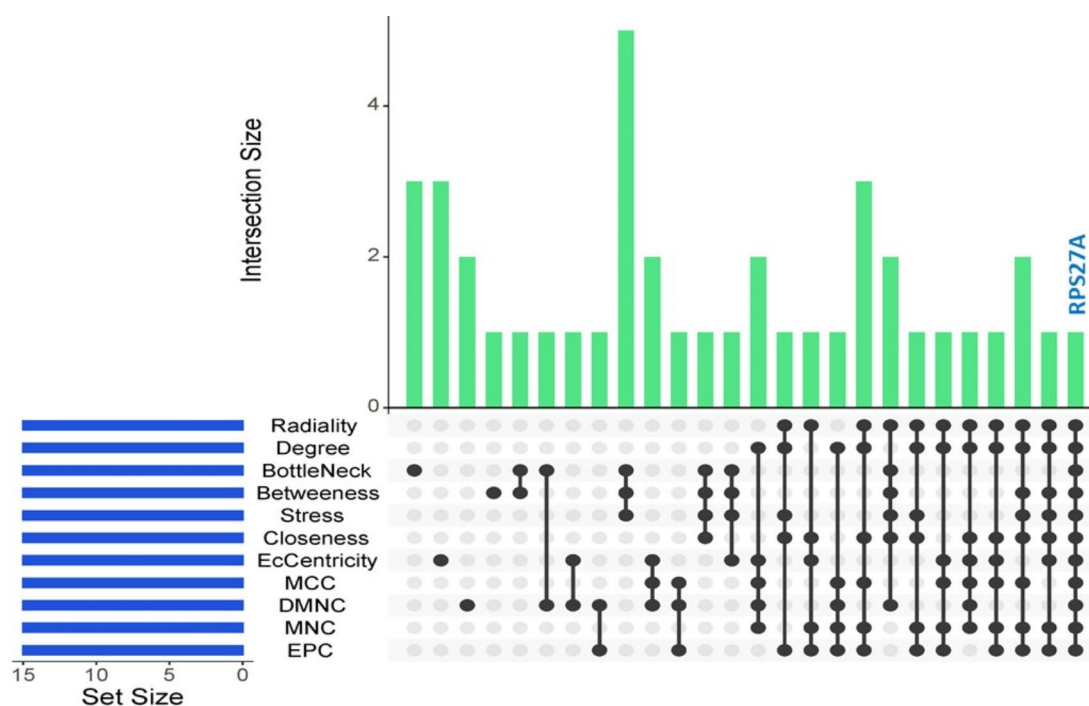

Fig. 6. Upset plot displaying hub genes identified across eleven cytoHubba topological algorithms, generated using the SRplot web server. **RPS27A** is present across all methods. The X-axis denotes the eleven cytoHubba algorithms, while the Y-axis represents interaction size.

Among these, only RPS27A hub genes were found in 11 algorithms of the cytoHubba plugin. Furthermore, Table 1 represents the biological function of the 6 potential hub genes in the human body.

Chord plot showing significant paths and key hub genes interaction

The arcs that connect the elements in a chord plot, which are arranged radially as geometric chords, indicate how the elements interact. Data groupings are distinguished from each other using different arc colors. The most enriched GO terms (BP, CC, and MF) with HUBGs including cytoplasmic translation, cytosolic ribosome, structural constituent of the ribosome, translation, ribosome, and small ribosomal subunit as depicted in Fig. 7 have strong connections with the main targets of the ID. Furthermore, the top nine molecular pathways including ribosome, Coronavirus disease - COVID-19, eukaryotic translation elongation, the response of EIF2AK4 (GCN2) to amino acid deficiency, viral mRNA translation, peptide chain elongation, and cytoplasmic ribosomal proteins were related to the core ID targets as shown in Fig. 8.

TFs-hub gene and hub genes-miRNAs interaction networks

Transcriptional and post-transcriptional regulatory networks were identified using network-based techniques to analyze the TF and miRNAs linkage networks of key hub genes. TFs and miRNAs are displayed as squares, and hub genes are displayed as circles shown in Fig. 9. Biomolecules including FOXC1, FOXL1, GATA2, TFAP2C, NR3C1, HINFP, and SREBF1 were the most significant TF regulators identified from the analysis of the JASPER database. We determined the quantity of seven miRNAs (hsa-mir-186-5p, hsa-mir-193b-3p, hsa-mir-93-5p, hsa-mir-16-5p, hsa-mir-92a-3p, hsa-mir-5011-5p, and hsa-mir-1277-5p) in our investigation from the miRTarbase database (Fig. 9). Furthermore, the biological function of reported biomolecules is represented in Table 2.

Protein chemical interaction analysis of key hub genes

Protein-chemical interactions (PCI) are essential for understanding the functions of proteins that assist molecular mechanisms within the cell; this knowledge could be of great assistance in the process of drug discovery. In this investigation, the protein–chemical interaction networks of ID were identified. Thirteen potentially interrelated chemical compounds were identified, among them chloropicrin, sodium selenite, arsenic trioxide, estradiol, enzyme inhibitors, and cupric oxide are among the highly enriched chemical agents found (Fig. 9).

Performance evaluation by ROC curve analysis

In this section, six hub genes and seven TFs were used in the ROC analysis. The area under the curve (AUC) in a ROC analysis serves as a performance benchmark. When the AUC score is between 0.5 and 1.0, the classifier is acceptable. In our test, the AUC score of the hub genes for GSE7329 stays between 0.751 (RPS5) and 0.891 (RPS27A); in the GSE25507 dataset, the range is 0.511 (RPS5) to 0.771 (RPS18) shown in Fig. 10a and b, respectively. TFs also have shown a significant AUC score in our test. In GSE7329, the range of AUC values are 0.436 to 0.818; where both FOXC1 and FOXL1 show the highest AUC value (Fig. 10c). In GSE25507, the range is 0.519 to 0.844; where GATA2 shows the highest AUC value (Fig. 10d).

Discussion

Using a cutting-edge technique of data analysis technique, single-cell RNA sequencing (scRNA-seq) uncovered previously unreported biomarkers associated with intellectual disability (ID). In summary, our research has helped identify certain genes and pathways that undergo alterations in individuals with intellectual disability (ID), suggesting their potential relevance as diagnostic markers for this condition.

Based on our raw dataset, the PBMCs clustered in 28 cellular subsets using the advanced tools of scRNA-seq [16]. In our study, we observed consistent cell proportions for each cell type in the five experimental groups, namely Natural Killer cells, Naïve B cells, T cells, Monocytes and Dendritic cells. In the context of our clustering analysis, we designate the following cell types including Th1 cells (CD4 + Tcell), non-Vd2 gd T cells, Mait cells, and Th1 cells as T cells. Subsequently, a comparison was performed between the aforementioned genes and the DEGs identified in the RNA-seq dataset (GSE46831).

We analyzed the set of genes that show differential expression and were found to be shared by the cluster and datasets. Meanwhile, we discerned noteworthy Gene Ontology (GO) within the domains of biological process (BP), cellular component (CC), and molecular function (MF). The Gene Ontology (GO) technology is

Gene symbol	Gene Function	Description	Uniport ID
RPS27A	Ribosomal protein S27a	The genetic sequence considered is responsible for the synthesis of a hybrid protein composed of ubiquitin at the amino (N) terminus and the ribosomal protein S27a at the carboxyl (C) terminus.	P62979
RPS18	Ribosomal protein S18	Multiple processed pseudogenes of this gene are dispersed throughout the genome, which is a common characteristic observed in genes that encode ribosomal proteins.	P62269
RPS21	Ribosomal protein s21	This protein is a member of the S21E ribosomal protein group and is found in the cytoplasm.	P63220
RPS7	Ribosomal protein S7	This protein is a member of the S7E family of ribosomal proteins.	P62081
RPS5	Ribosomal protein S5	A small subunit called 40 S and a large subunit called 60 S make up ribosomes, which are organelles that facilitate the process of protein synthesis.	P46782
RPL9	Ribosomal protein L9	The gene in question is responsible for encoding a ribosomal protein, which serves as a constituent of the 60 S subunit.	P32969

Table 1. Evaluated the six key hub genes through a variety of regulatory processes.
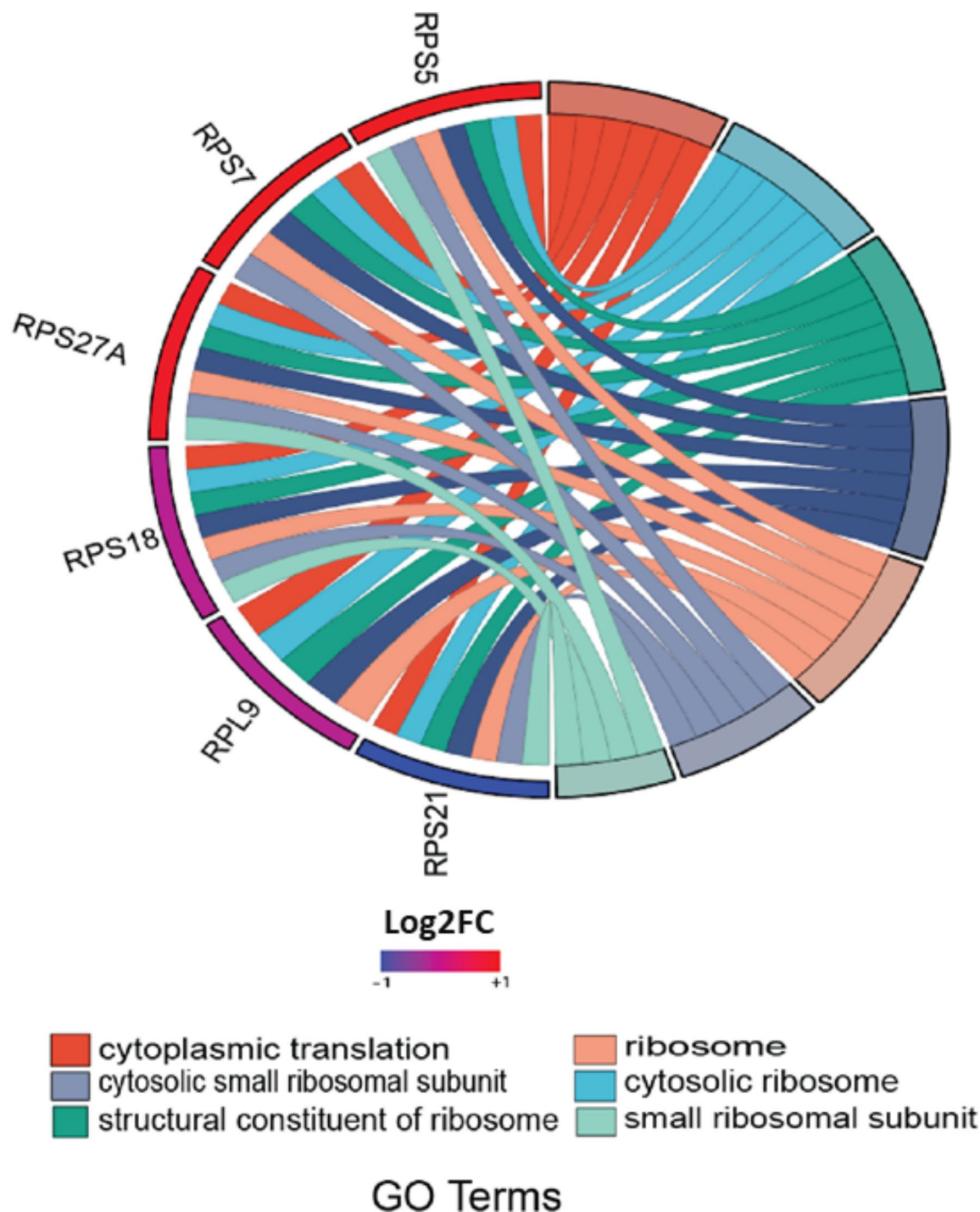
Fig. 7. Mapping the terms of the gene ontology and their association with the key hub genes in the chord plot, visualized through the chord plot module on the SRplot web server. Based on the log2FC value of key hub genes, different colors indicate the different pathways.

specifically designed to support the computational representation of biological systems which provides relevant information concerning the function of gene[17]. Based on our analysis, it has been determined that translation is among the prominent gene ontology (BP) concepts associated with biological processes. This particular process is responsible for initiating the translation of RNA molecules about specific biomolecules[18]. The process of small subunit biogenesis holds significant importance in individuals with intellectual disability, as evidenced by previous research[19]. The B cell receptor signaling pathway has been suggested to exhibit potential connections with gene mutations associated with intellectual disability[20]. In the context of form CC, it is observed that the most enriched terms exhibit a strong correlation with the ID. The binding of major histocompatibility complex (MHC) class II protein complexes and the activity of MHC class II receptors have been found to indicate a
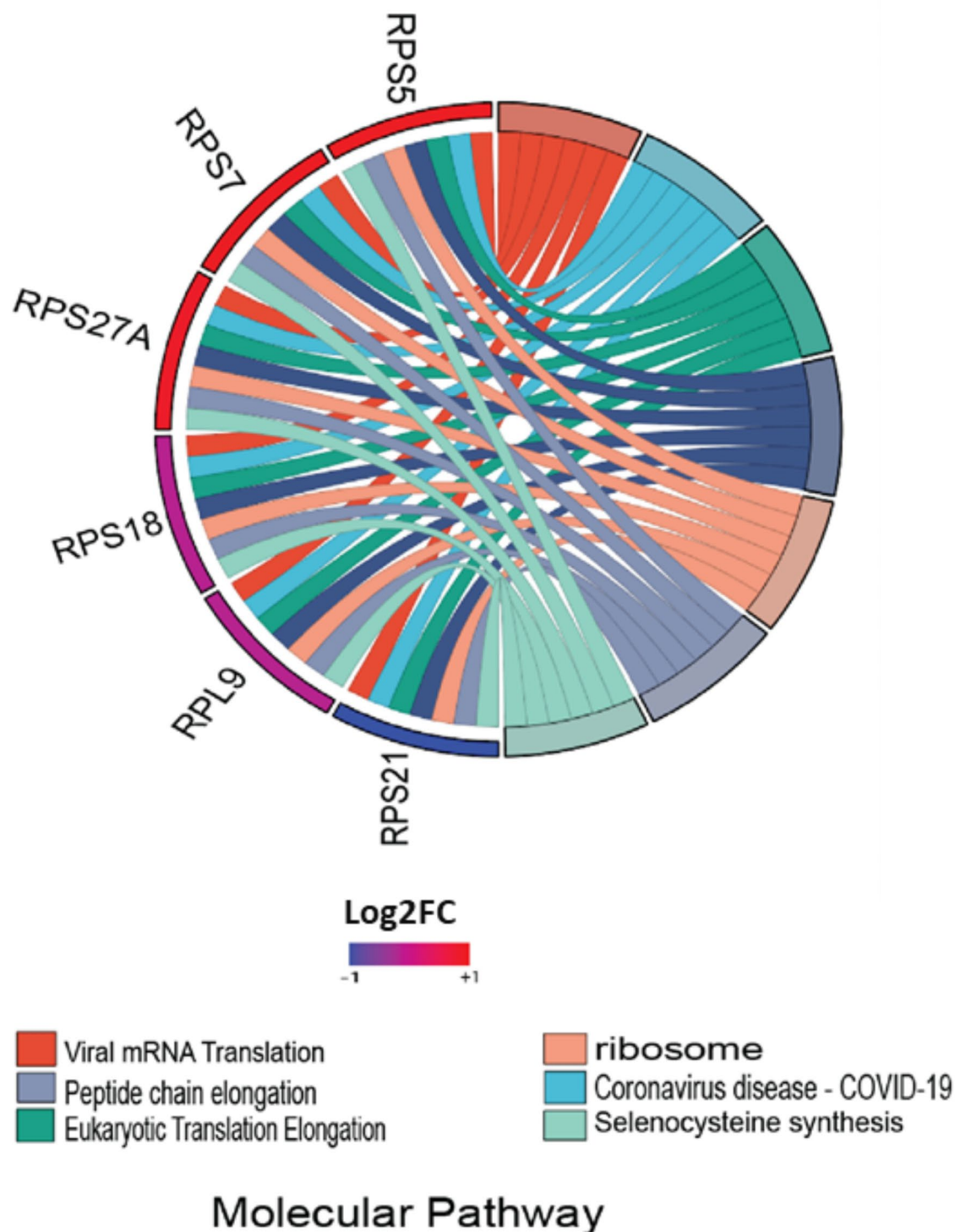
Fig. 8. Mapping the most enriched molecular pathways and their association with the hub genes in the chord plot, visualized through the chord plot module on the SRplot web server. Based on the log2FC value of key hub genes, different colors indicate the different pathways.

significant correlation with immune disorders. Insufficiency of this particular protein has been identified as a potential etiological factor that can contribute to intellectual disability (ID)[21]. Furthermore, there is a substantial correlation between 5'-UTR mRNA binding and RNA binding, particularly in the context of fragile X syndrome, a specific type of intellectual disability (ID)[22]. Certain genes associated with autism spectrum disorder (ASD) can cause an inappropriate immunological response, which is also related to intellectual disability (ID)[23].
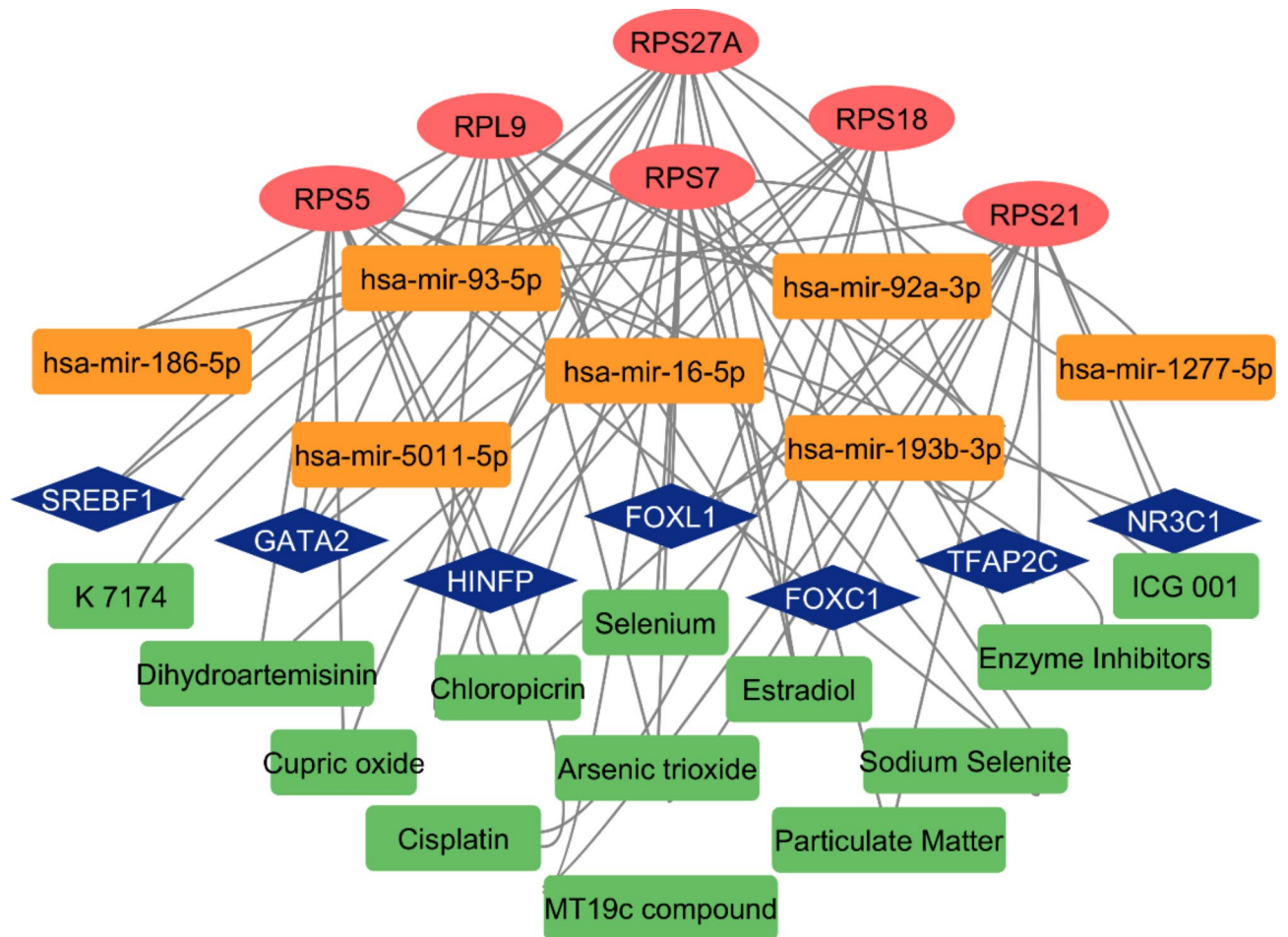
Fig. 9. Interaction of key hub genes regulatory and therapeutic interaction network of transcription factors, microRNA, and chemicals. In the figure, the oval shape indicates key hub genes, TFs denote an angle shape, miRNAs denote a rectangular shape and a deep yellow color, and chemicals are differentiated by green colors and rectangular shapes. The interaction network was constructed with NetworkAnalyst (v 3.0) and visualized in Cytoscape (v3.10.1).

In contrast, we discovered the KEGG, Reactome, WiKi, and BioCarta pathways. We found that the most important KEGG pathways were related to allograft rejection, graft-versus-host disease, asthma, type I diabetes mellitus, autoimmune thyroid disease, and coronavirus disease. A recent study shows that people with ID are at increased risk for respiratory problems and asthma and demand special care from a trained caregiver[24]. The risk of diabetes is prevalent within the population of people with intellectual disability (ID)[23]. There is a documented correlation between autoimmune thyroid disease and individuals with intellectual disabilities, particularly in the pediatric population[25]. Individuals with intellectual disabilities are more likely to contract coronavirus disease 2019 (COVID-19) and have negative consequences from it[26]. The primary expression of Reactome pathways was found in SARS-CoV-1, which affects the host translation machinery, CD22-mediated BCR control, ZAP-70 translocation to the immunological synapse, selenocysteine synthesis and peptide chain elongation. The results indicate that selenium is synthesized and incorporated into selenoproteins. SEPSECS mutations have been associated with selenium insufficiency for severe intellectual disabilities[27]. We discovered that significant pathways were largely expressed mainly in the B cell receptor complex, allograft rejection, antigen processing and the presentation, and BCR signaling pathway using the WiKi pathway. A large number of biological activities, including signal transduction and transcriptional control, are based on protein-protein interactions[28]. The PI3K/AKT/mTOR- vitamin D3 signaling pathway exhibited notable enrichment within the BioCarta pathway. A comprehensive study has revealed the pathogenic involvement of mTOR signaling in various neurological disorders, including epilepsy, autism (ASD), intellectual disability (ID), dementia, traumatic brain injury, brain tumors, and hypoxic-ischemic injury[29].

A PPI network facilitates the formation of protein complexes and regulates many cellular processes such as signaling, regulation, and transport[30]. In our genomic analysis, we successfully identified six highly connected central hub genes, namely RPS27A, RPS21, RPS18, RPS7, RPS9, and RPL9. This identification was achieved by utilizing eleven distinct computational methods available within the Cytoscape CytoHubba plug-in. All of our hub genes are related to the ribosomal protein (RP) family. These RPs play a crucial role in various

TFs	Description	Function
Transcriptional regulatory biomolecules (TFs)		
FOXC1	Forkhead box	The activity of DNA-binding transcription factors and the specificity of RNA polymerase II.
GATA2	GATA binding protein 2	The activity of transcription factors that bind to DNA and their association with chromatin.
FOXL1	Forkhead box L1	The transcription factor is essential for the appropriate proliferation and differentiation processes within the gastrointestinal epithelium.
TFAP2C	Transcription Factor AP-2 Gamma	A DNA-binding protein with sequence specificity that engages with inducible viral and cellular enhancer elements.
NR3C1	Nuclear Receptor Subfamily 3 Group C Member 1	Aids in the degradation of messenger RNAs (mRNAs) quickly by binding to their 5' untranslated regions (UTRs).
HINFP	Histone H4 Transcription Factor	The transcriptional repressor exhibits binding affinity to the consensus sequence 5'-CGGACGTT-3' as to well as the RB1 promoter.
SREBF1	Sterol Regulatory Element Binding Transcription Factor 1	The precursor of the transcription factor form, known as processed sterol regulatory element binding protein 1 (SREBP-1) is localized within the membrane of the endoplasmic reticulum.
miRNAs	Description	Function
Post-transcription regulatory biomolecules (miRNAs)		
mir-16-5p	MicroRNA 16	Prevent replication of multiple viruses (EV71).
mir-193b-3p	MicroRNA 193b	Demonstrates reciprocal interaction with MYC and inhibits the growth and metastasis.
mir-93-5p	MicroRNA 93	Contributes to the genesis of different diseases.
mir-92a-3p	MicroRNA 92a	Reduces PTEN and prevents Eca-109 cells from being phosphorylated and inhibited by Akt.
mir-186-5p	MicroRNA186	Inhibit tumorigenesis of glioblastoma multiforme both in vitro and in vivo.

Table 2. Potential transcriptional and post-transcription regulatory biomolecules of ID.

biological processes, including ribosome biogenesis, protein synthesis, cellular growth, development, and programmed cell death (apoptosis)[31]. Immunometabolism is the process by which immune cell activity and metabolic pathways are combined, and ribosomal proteins are involved in this process. During activation, T cells undergo metabolic reprogramming, changing from oxidative phosphorylation to glycolysis. This needs increased ribosome biogenesis and function[32]. Certain ribosomal proteins have direct interactions with RNA molecules or affect transcription factors' activities, regulating the expression of genes essential for T cell growth, differentiation, and function[33]. Ribosome biogenesis plays a crucial role in regulating cell development and proliferation. Dysregulation of this process can lead to abnormal cell growth and pathological conditions such cancer and metabolic diseases[32]. Some metabolic conditions, especially those that affect the nerve system or brain, might cause behavioral problems, cognitive problems, or developmental delays that resemble autism spectrum disorders[34].

A study related to perisylvian polymicrogyria shows that RPS27A has been highly enriched when genes were extracted from the ID patient[35]. RPS27A dysfunction impacts ribosome assembly, protein synthesis, and the Ubiquitin-Proteasome System (UPS), resulting in cellular stress and immunological dysfunction. These impacts may affect neurodevelopment, potentially leading to ID[36]. In our study, RPS6, although not classified as one of our central genes, exhibited expression in T-cell genes. Multiple studies have discovered a significant increase in RPS6 expression among individuals diagnosed with intellectual disability (ID)[37]. Numerous investigations have substantiated the involvement of the aforementioned protein in the translation of 5' terminal oligopyrimidine tract (TOP) mRNAs, alongside its role in governing cell size and proliferation regulation[38]. RPS7 mutants have active matrix metalloproteinase (MMP) family genes, suggesting that improper cell migration may be occurring, leading to further dysfunctional development and RPS5 knockdown leads to brain abnormalities in Zebrafish[39]. Further investigation may reveal that mutations in RPS5 and RPS7 could be biomarkers of ID. According to a recent study, significant changes in the mRNA levels of RPL9 in both individuals with autism spectrum disorder (ASD) were found to be consistent in the same direction through quantitative polymerase chain reaction (qPCR) validation[40]. RPL10 is present in people with intellectual impairment, dysmorphism, and autism[41]. The two additional hub genes, RPS18 and RPS21, exhibit limited significance in the identification process. RPS18 has been used as the reference gene in certain studies on autism spectrum disorder (ASD) as the housekeeping gene[42]. It is concluded that T cells are essential for the immune response and play an important role in neuroinflammation. Excessive T cell activation can cause prolonged neuroinflammatory states, disrupting normal brain development and function. Disruptions to neuronal growth, synapse formation, and cognitive processes can lead to ID[43].

TFs play an important role in controlling transcription, which means that their amounts can be used to find people with diseases like ID. Our study shows that the TFs that regulate DEGs through the TFs-Hub genes interaction network play a key role in the development of ID. The TFs are FOXC1, FOXL1, GATA2, TFAP2C, NR3C1, HINFP, and SREBF1 were identified in our study. Two separate case report studies have established a correlation between the presence of FOXC1 on ring chromosome 6 and intellectual disability (ID), short stature, and various facial deformities[44]. The whole genome a GATA2 mutation produces a rare Syndromic Congenital Neutropenia with ID, according to sequencing[45]. FOXL1 is a Forkhead box protein (FOX) whose dysregulation encourages the Wnt/b-catenin signaling pathway[45]. There is a potential correlation between TFAP2C and the
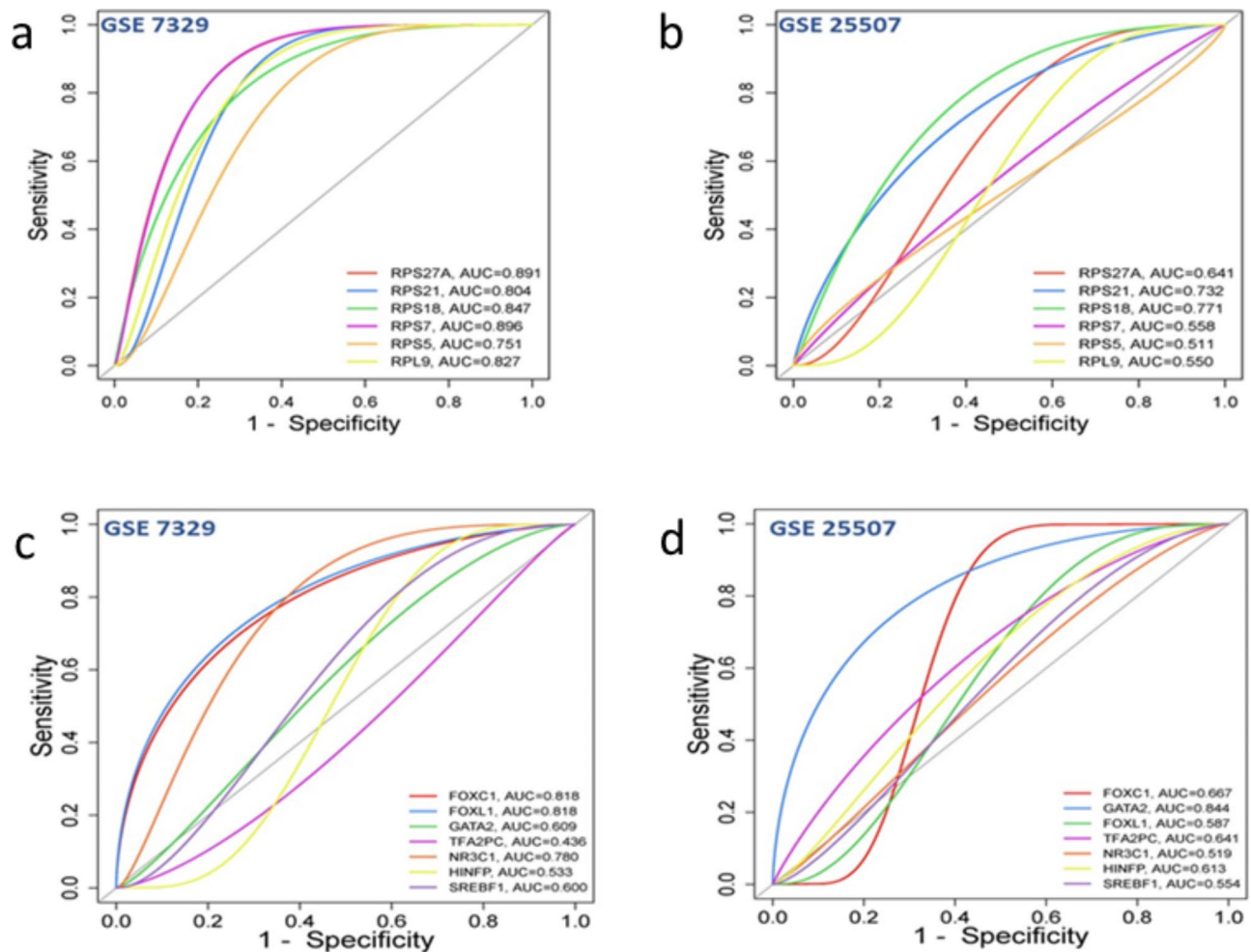
Fig. 10. ROC curve of potential biomarkers (hub genes and TFs). The ROC curve of six Hub genes was represented in two GEO profiles (**a**) GSE7329 and (**b**) GSE25507. On the other hand, seven TFs were represented in two GEO profiles (**c**) GSE7329 and (**d**) GSE25507. The ROC curve in constructed via pROC package (v 1.18.5) in R (v 4.3.1).

development of intellectual disability, as suggested in previous research[46]. Nuclear receptor subfamily 3, group C, member 1 (NR3C1) has been identified by the ingenuity pathway analysis (IPA) as an upstream transcriptional regulator for ID[47]. Previous research shows that HINFP and SREBF1 do not have any significant relevance in studies related to intellectual disability (ID). However, it is suggested that these factors may be of value in future investigations.

Using the miRTarBase v8.0 database, we investigated the interaction between miRNAs and hub genes. MicroRNAs control the post-transcriptional regulation of gene expression, as opposed to transcription factors, making them a potentially valuable tool for diagnostic testing and biomarker research[48]. Based on topological analysis, the following miRNAs have been identified as significant including hsa-mir-186-5p, hsa-mir-193b-3p, hsa-mir-93-5p, hsa-mir-16-5p, hsa-mir-92a-3p, hsa-mir-5011-5p, and hsa-mir-1277-5p. The microRNA mir92a-3p shows a significant correlation with the structure and function. Furthermore, it has been recognized as a biomarker in peripheral blood for schizophrenia[49]. Furthermore, dysregulation of mir-92a-3p was found in peripheral blood of patients in a study on gene regulation linked to autism in the Chinese population[50]. Furthermore, miR-16-5p has previously been described in rat neurons as a negative regulator of dendritic complexity and mediates BDNF-induced dendritogenesis by regulating the translation of the BDNF mRNA itself, supporting the hypothesis that miR-16-5p plays a role in neuronal development[51].

We identified chloropicrin, sodium selenite, arsenic trioxide, estradiol, enzyme inhibitors, and cupric oxide via protein-chemical interaction analysis. Some studies have indicated that sodium selenite treatment improves cognitive performance in triple transgenic Alzheimer's disease mice by reducing lipid peroxidation[52]. A retrospective cohort study was conducted to investigate the association between soil arsenic concentration and intellectual disability (ID) in pregnant women. The findings revealed that elevated soil arsenic concentration was only significantly associated with increased odds of diagnosed ID during the first trimester of pregnancy[53]. In the final phase, we evaluated the diagnostic effects of our biomarkers with the help of ROC analysis. Our
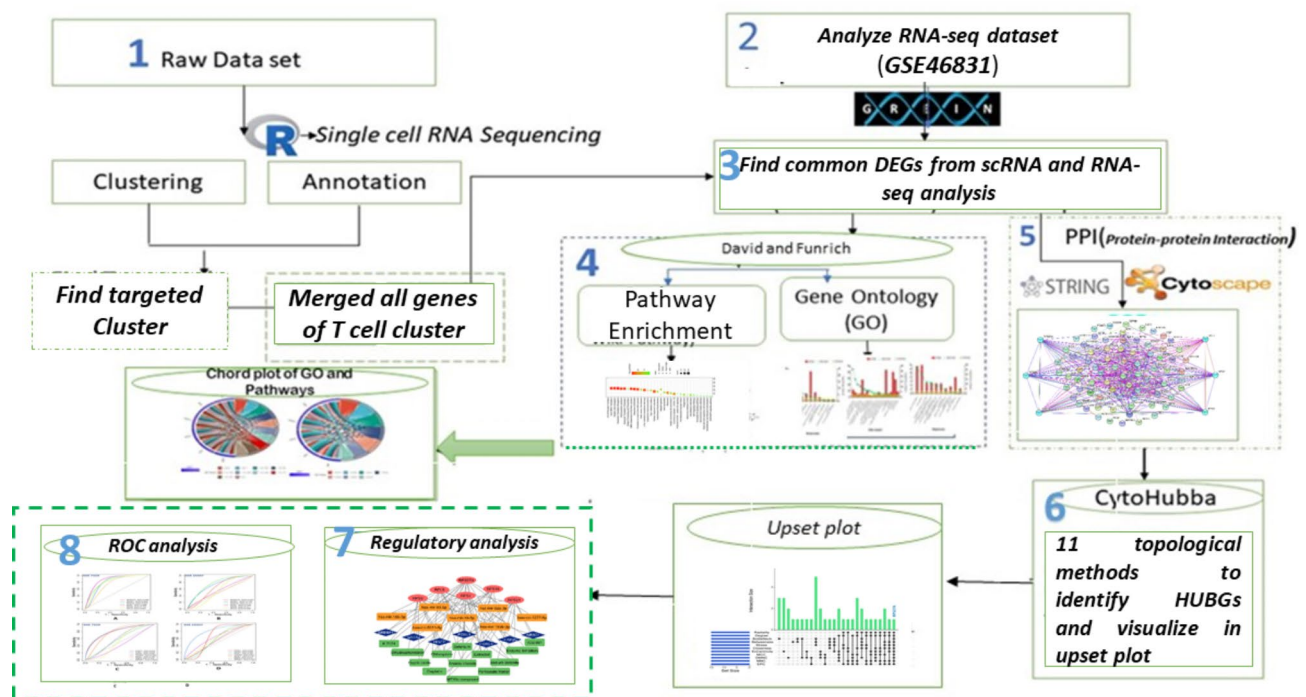
Fig. 11. Graphical illustration of the workflow in our study, constructed with Adobe Illustrator CC (v 2019).

predicted biomarkers (hub genes and TFs) have shown decent performance during analysis. Our atlas offers a valuable framework for improving our understanding of this intricate immune-mediated disease. Despite several limitations including, lower number of dataset and human-mediated sample, excluding cell line data, total sample number less than ten, and wet laboratory validation of potential biomarkers, we provide a thorough overview of the variations. In summary, we believe that our approach is scientifically valid and provides meaningful insights into the role of CTCF mutations in Intellectual Disability. As a result, we can serve as a resource to guide future studies aimed at identifying the most effective approach for managing or potentially curing disease.

Materials and methods

A graphical representation of the combined system biology and analytical technique approach to identify biological markers and pathways in the blood tissues of ID patients is presented in Fig. 11.

Parameters of single cell RNA-seq dataset analysis

Collection and pre-processing

The "Single Cell Portal" was used to access publicly available scRNA-seq data (PMID: 28212749), which was collected from patient peripheral blood (Single Cell Comparison: PBMC data - Single Cell Portal (broadinstitute.org))[54]. Data processing was performed with the help of the Seurat package 4.3.0 version 4.3.1 of R[55]. Expression data in this study are represented as logarithmic counts per 10,000. This representation is derived from unique molecular identifier (UMI) counts for all methods, except for Smart-seq2, where it is based on read count[56]. The data processing was carried out with the help of Seurat package 4.3.0 version 4.3.1 of R. Initially, we converted the raw dataset into a Seurat object.

Quality control, normalization, and feature selection

Three cells and a minimum of 200 features with nonzero counts were needed for the gene retrieval process to provide quality control[55]. By using a function that determines the percentage of counts derived from a given set of parameters, 51 mitochondrial quality control (QC) measures can be obtained[57]. The filtered data from the previous step was normalized[58] to simplify the following analysis; a log transformation is applied to the previously processed data.

To account for and minimize cell-specific bias, which could affect subsequent uses such as detection of differential gene expression, normalization of the data is necessary[58]. The purpose of normalization is to correct for measurement discrepancies between samples and/or features (genes, for example) that are caused by undesirable biological effects (batch effects, for example) or technical artifacts rather than biological effects of interest[59]. Next, we count how many features in the sample are highly variable from cell to cell, which means that those are particularly expressed in specific cells but not in the remains[60].

Cell clustering, annotation, and visualization

When identifying groups of cells with similar patterns of gene expression, clustering is a useful technique. Initially, a graph-based clustering algorithm based on K-nearest neighbor (KNN) is used. Seurat v3 implements a clustering methodology based on graphs, this process is also used in previous studies[61]. Our approach to cell clustering involves using modularity optimization techniques, such as the Louvain algorithm, to systematically group cells together. The maximum standard modularity function is the aim[62]. The reference-based cluster annotation is implemented in a separate module SingleR[62] and started by converting the information from Seurat data to its format. Finally, annotation based on external references enables side-by-side evaluation with filtered data done by R's Monaco package using the MonacoImmune Data function[63]. A well-liked method to reduce dimensionality while maintaining the overall structure of the data is called the uniform manifold approximation and projection (UMAP). Complex datasets can be visualized in an easy-to-understand and instructive way using UMAP plots, which facilitate the identification of patterns, clusters, and outliers within the data[64].

RNA-seq data collection and preprocessing

The gene expression profiles for the ID-linked RNA sequence dataset GSE46831 are sourced from the GEO (Gene Expression Omnibus) maintained by the National Center for Biotechnology Information (NCBI). The source of the cells in the sample is blood lymphocytes. We have found 11 samples in total, where there are 8 controls and 3 affected by intellectual disabilities. The RNAseq data represents three ID patients with de novo CTCF mutation that have become considered substantial contributions to intellectual[65].

After analyzing the dataset, differentially expressed genes (DEGs) were extracted by employing GREIN (a widely used web-based tool for RNA-seq data analysis) (http://www.ilincs.org/apps/grein/) online database. GREIN uses the GEO RNA-seq experiment processing pipeline (GREP2) to process data such as DESeq2 for differential expression analysis[66]. In our study, we analyzed the data based on phenotypes; where the experimental group was 'Intellectual disability' and control group was 'control'.

Statistical analysis of collected datasets and identification of DEGs

For testing differential expression, a negative binomial distribution is commonly employed, with the false discovery rate (FDR). It is controlled using techniques such as the Benjamini-Hochberg correction procedure, which accounts for multiple comparisons. To reduce the possibility of misidentifying false positives as significantly differentially expressed genes, the raw p-values of each test are adjusted to control the false discovery rate (FDR). Statistically significant differentially expressed genes (DEGs) of each corresponding cluster and RNAseq datasets were identified based on the specified cutoff value (P-Value < 0.05 and absolute value log2FC > 1). Furthermore, a P-value < 0.05 and absolute log2FC > 1 was set for upregulated while absolute log2FC < −1 was used for the identification of downregulated DEGs.

In addition, we employed a cross-matching analysis between the scRNA-seq and RNA-seq datasets to identify DEGs in our targeted immune cell (T cells) related ID. T cells are the specific cellular targets of ID. Several studies demonstrated that these cells exhibit a significant association with ID[28,67]. To identify genes that showed a significant match between scRNA-seq (T cell cluster merged genes) and RNA-seq datasets (GSE46831), a VENN analysis was performed using the Venny 2.0.1 tool, which is available at (bioinfogp.cnb.csic.es/tools/venny/)[68]. These shared DEGs were used as fuel in the further analysis of our study.

Pathways and functional enrichment analysis of DEGs

Reliable software 'FunRich' (version 3.4.1) and its 63 integrated database[69] were used to analyze and visualize gene ontology (GO) enrichment. An online bioinformatics tool DAVID v6.8[70] was used to analyze KEGG[71], Reactome[72], and WiKi[73] pathways. We also used SRplot (https://www.bioinformatics.com.cn/en) for GO pathway visualization. In this study, we fixed the threshold p-value below 0.05 to determine functional pathway that was considered as significant. In addition to annotating the enrichment analysis of bioterm classifying methods and gene clusters, pathways and GO were utilized to provide a proper tool for prediction.

Protein-protein interaction (PPI) network construction and analysis of hub genes

The study of pathogenic mechanisms and disease progression is facilitated by networks, as they provide valuable insight into the molecular mechanisms that underlie the activity of cells. In this particular investigation, the STRING v11.0 (https://string-db.org/)[74] was used to build a PPI network of DEGs. Then Cytoscape (version 3.10.1)[75] was used to determine and visualize the PPI networks of DEGs. We have explored the most significant gene of T cells by considering the genes that are common in at least seven out of eleven topological methods, including radiality, degree, bottleneck, betweenness, stress, closeness, eccentricity, maximum clique centrality (MCC), density of maximum neighborhood component (DMNC), maximum neighborhood component (MNC), edge percolated component (EPC) in cytoHubba plugin[76] of Cytoscape (version 3.10.1). We have also constructed an Upset plot[77] using the SRplot online (https://www.bioinformatics.com.cn/en) server to visualize the hub genes that are used in eleven cytoHubba methods. Those genes reported are identified as the key hub genes (KGs) of ID[78].

Correlation of most significant pathways and hub genes

A chord plot[79] visually represents the co-relation between the log2FC (logarithm of fold change) values of potential hub genes and the most significant pathways that are associated with ID. The GO chord plot displayed a comprehensive selection of the top seven GO terms, which are included in the top CC, BP and MF pathways. In addition, this plot also depicts the nine distinct pathways sourced from KEGG, Reactome, WiKi, and BioCarta databases. The corresponding pathways, along with their hub genes and log2FC, are subsequently inputted into

the SRplot web-based server, resulting in the generation of a chord plot showing the interactions between the GO terms[80].

Regulatory biomolecules (TFs and miRNAs) identification of hub genes

Vital regulatory biomolecules such as transcriptional factors (TFs) and microRNAs (miRNAs) can cause substantial modifications in transcriptional processes and gene expression. We used the latest version of the NetworkAnalyst (version 3.0)[81] online tool to predict TFs-Hub genes and Hub genes-miRNAs interactions through Jasper[82] and miRTarbase v8.0 [83] databases, respectively. Then, nonmajor signature molecules were filtered out of the interaction networks using degrees centrality and betweenness values of 2, and 6 for TFs-hub genes interaction and 2, and 6 for miRNAs-hub genes interaction, respectively.

Analyze protein-chemical compounds of hub genes

The examination of protein-chemical compounds is employed to identify the specific chemical compounds that play a role in the interaction of proteins in comorbidities. The Comparative Toxicogenomics Database (CTD) was used to analyze those proteins that are connected with hub genes[84]. We managed to discover protein-chemical interaction (PCI) with the aid of NetworkAnalyst[81].

ROC analysis of potential biomarkers

In biomedical and bioinformatics applications, receiver operating characteristic (ROC) curves are useful tools for assessing classifier performance. ROC curves, useful tools for evaluating classifier performance, are employed in biomedical and bioinformatics applications. The classifier uses the test result to decide whether a patient has a positive or negative disease state[85]. We used GSE25507 and GSE7329 as reference datasets to evaluate the performance of our biomarkers (hub genes and TFs). We used the pROC package-based tool to generate the ROC curve[86]. It uses true positive values and false positive values; a true positive rate (TPR) is plotted on the Y axis and a false positive rate (FPR) on the X axis. Analyzing our results on test sets, we can obtain a pair of (TPR and FPR) points for a given classifier.

Conclusions

Intellectual disability (ID) is an abnormality that has far-reaching social consequences; it affects not just individuals who suffer from it but also their families and society in its entirety. Both the scRNA-seq and RNA-seq approaches were used to identify biomarkers as indicators of intellectual disability in this research. Differentially expressed genes (DEGs) have been applied in pathway analysis, as well as in the study of protein-protein interactions, transcription factors, post-transcriptional factors, and potential therapeutic compounds. Therefore, we have identified several molecular signatures and pathways that demonstrate frequent dysregulation in the tissue of cells involved in ribosomal synthesis. Multiple transcription factors (TFs) and microRNAs (miRNAs) have been identified as candidate regulatory factors associated with the hub genes. Comprehensive treatment will be vital for individuals with ID. Currently, it is imperative that pharmacists or nursing staff immediately inform the physician regarding any instances of medication-related adverse effects. Therefore, the application of an integrated strategy is considered the most effective method of providing treatment to individuals with intellectual disabilities. In contrast, our observations will provide insight into the molecular etiology of intellectual disability and help identify potential therapeutic interventions.

Data availability

PBMC data for scRNA-seq analysis was collected from the Single Cell portal (Single Cell Comparison: PBMC data - Single Cell Portal (broadinstitute.org)). The Gene Expression Omnibus (GEO) database from NCBI (https://www.ncbi.nlm.nih.gov) was used to access the GSE46831, GSE7329 and GSE25507 datasets. All the data produced in this research can be available on request to the corresponding author (Dr. Md Habibur Rahman).

References

1. Ropers, H. H. Genetics of intellectual disability. *Curr. Opin. Genet. Dev.* **18**, 241–250. https://doi.org/10.1016/j.gde.2008.07.008 (2008).
2. Patel, D. R., Cabral, M. D., Ho, A. & Merrick, J. A clinical primer on intellectual disability. *Transl. Pediatr.* **9**, S23–S35 (2020).
3. Salvador-Carulla, L. & Bertelli, M. 'Mental retardation' or 'intellectual disability': Time for a conceptual change. *Psychopathology* **41**, 10–16. https://doi.org/10.1159/000109950 (2007).
4. Leonard, H. & Wen, X. The epidemiology of mental retardation: Challenges and opportunities in the new millennium. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 117–134. https://doi.org/10.1002/mrdd.10031 (2002).
5. Vissers, L. E. L. M. et al. A de novo paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).
6. Melom, J. E. & Littleton, J. T. Synapse development in health and disease. *Curr. Opin. Genet. Dev.* **21**, 256–261 (2011).
7. Guilmatre, A. et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch. Gen. Psychiatry.* **66**, 947–956 (2009).
8. Roa-Bautista, A., López-Duarte, M., Paz-Gandiaga, N., Arribas, D. S. S. & Ocejo-Vinyals, J. G. Deletion in the BCL11B gene and intellectual developmental disorder with speech delay, dysmorphic facies, and T-cell abnormalities–a case report. *EJIFCC* **33**, 325 (2022).
9. Ashwood, P. et al. Altered T cell responses in children with autism. *Brain Behav. Immun.* **25**, 840–849 (2011).
10. Hamdan, F. F. et al. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011).
11. Pascual-Alonso, A., Martínez-Monseny, A. F., Xiol, C. & Armstrong, J. MECP2-related disorders in males. *Int. J. Mol. Sci.* **22**, 9610 (2021).

12. Willemsen, M. H. et al. Mutations in DYNC1H1 cause severe intellectual disability with neuronal migration defects. *J. Med. Genet.* **49**, 179–183 (2012).

13. Bruel, A. L. et al. Next-generation sequencing approaches and challenges in the diagnosis of developmental anomalies and intellectual disability. *Clin. Genet.* **98**, 433–444. https://doi.org/10.1111/cge.13764 (2020).

14. Li, D. et al. Exome and RNA-Seq analyses of an incomplete penetrance variant in USP9X in female-specific syndromic intellectual disability. *Am. J. Med. Genet. Part. A* **188**, 1808–1814 (2022).

15. Luo, Z. G., Peng, J., Li, T. & Single-Cell, R. N. A. Sequencing reveals cell-type-specific mechanisms of neurological diseases. *Neurosci. Bull.* **36**, 821–824 (2020).

16. Saliba, A. E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860. https://doi.org/10.1093/nar/gku555 (2014).

17. Thomas, P. D. The gene ontology and the meaning of biological function. In *Methods in Molecular Biology* vol. 1446 15–24 (Humana Press Inc., 2017).

18. Wan, D., Zhang, Z. C., Zhang, X., Li, Q. & Han, J. X chromosome-linked intellectual disability protein PQBP1 associates with and regulates the translation of specific mRNAs. *Hum. Mol. Genet.* **24**, 4599–4614 (2015).

19. Saghi, M., InanlooRahatloo, K., Alavi, A., Kahrizi, K. & Najmabadi, H. Intellectual disability associated with craniofacial dysmorphism due to POLR3B mutation and defect in spliceosomal machinery. *BMC Med. Genom.* **15** (2022).

20. Smith, C. I. E. & Burger, J. A. Resistance mutations to BTK inhibitors originate from the NF-κB but not from the PI3K-RAS-MAPK arm of the B cell receptor signaling pathway. *Front. Immunol.* **12**. https://doi.org/10.3389/fimmu.2021.689472 (2021).

21. Farrokhi, S. et al. MHC class II deficiency: Report of a novel mutation and special review. *Allergol. Immunopathol.* **46**, 263–275. https://doi.org/10.1016/j.aller.2017.04.006 (2018).

22. McAninch, D. S. et al. Fragile X mental retardation protein recognizes a G quadruplex structure within the survival motor neuron domain containing 1 mRNA 5′-UTR. *Mol. Biosyst.* **13**, 1448–1457 (2017).

23. Jones, K. L. et al. Autism with intellectual disability is associated with increased levels of maternal cytokines and chemokines during gestation. *Mol. Psychiatry* **22**, 273–279 (2017).

24. Davis, S. Asthma in intellectual disability: Are we managing our patients appropriately?. *Breathe* **12**, 310–317 (2016).

25. Nandi-Munshi, D. & Taplin, C. E. Thyroid-related neurological disorders and complications in children. *Pediatric Neurol.* **52**, 373–382. https://doi.org/10.1016/j.pediatrneurol.2014.12.005 (2015).

26. Buono, S. et al. Clinical management of individuals with intellectual disability: The outbreak of Covid-19 pandemic as experienced in a clinical and research center research in developmental disabilities. *Res. Dev. Disabil.* **110** (2021).

27. Schoenmakers, E. & Chatterjee, K. Human disorders affecting the selenocysteine incorporation pathway cause systemic selenoprotein deficiency. *Antioxid. Redox Signal.* **33**, 481–497. https://doi.org/10.1089/ars.2020.8097 (2020).

28. Yang, S. et al. Mutant BCL11B in a patient with a neurodevelopmental disorder and T-cell abnormalities. *Front. Pediatr.* **8** (2020).

29. Trifonova, E. A., Klimenko, A. I., Mustafin, Z. S., Lashin, S. A. & Kochetov, A. V. The mTOR signaling pathway activity and vitamin D availability control the expression of most autism predisposition genes. *Int. J. Mol. Sci.* **20** (2019).

30. Childs, L. M. & Larremore, D. B. Network models for malaria: Antigens, dynamics, and evolution over space and time. In *Systems Medicine: Integrative, Qualitative and Computational Approaches* vol. 1–3, 277–294 (Elsevier, 2020).

31. Saha, A. et al. Genome-wide identification and comprehensive expression profiling of ribosomal protein small subunit (RPS) genes and their comparative analysis with the large subunit (RPL) genes in rice. *Front. Plant. Sci.* **8** (2017).

32. Geiger, R. et al. L-arginine modulates T cell metabolism and enhances survival and anti-tumor activity. *Cell* **167**, 829–842 (2016).

33. Marchingo, J. M. & Cantrell, D. A. Protein synthesis, degradation, and energy metabolism in T cell immunity. *Cell. Mol. Immunol.* **19**, 303–315 (2022).

34. Ceravolo, F. et al. Neurological involvement in inherited metabolic diseases: An overview. *J. Pediatr. Biochem.* **6**, 3–10 (2016).

35. Amrom, D. et al. Duplication 2p16 is associated with perisylvian polymicrogyria. *Am. J. Med. Genet. Part. A* **179**, 2343–2356 (2019).

36. Wang, W. et al. Ribosomal proteins and human diseases: Pathogenesis, Molecular mechanisms, and therapeutic implications. *Med. Res. Rev.* **35**, 225–285 (2015).

37. Pelorosso, C. et al. Somatic double-hit in MTOR and RPS6 in hemimegalencephaly with intractable epilepsy. https://doi.org/10.1093/hmg/ddz194/5549779

38. Meyuhas, O. Chapter 1 Physiological roles of ribosomal protein S6: One of its kind. *Int. Rev. Cell Mol. Biol.* **268**, 1–37. https://doi.org/10.1016/S1937-6448(08)00801-0 (2008).

39. Uechi, T. et al. Ribosomal protein gene knockdown causes developmental defects in zebrafish. *PLoS One* **1** (2006).

40. Kuwano, Y. et al. Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with autism and healthy women having autistic children. *PLoS One* **6** (2011).

41. Lezzerini, M. et al. Ribosomal protein gene RPL9 variants can differentially impair ribosome function and cellular metabolism. *Nucleic Acids Res.* **48**, 770–787 (2020).

42. Sjaarda, C. P. et al. Exome sequencing identifies de novo splicing variant in XRCC6 in sporadic case of autism. *J. Hum. Genet.* **65**, 287–296 (2020).

43. Estes, M. L. & McAllister, A. K. Maternal immune activation: Implications for neuropsychiatric disorders. *Science (80-)* **353**, 772–777 (2016).

44. Pace, N. P., Maggouta, F., Twigden, M. & Borg, I. Molecular cytogenetic characterisation of a novel de novo ring chromosome 6 involving a terminal 6P deletion and terminal 6Q duplication in the different arms of the same chromosome. *Mol. Cytogenet.* **10**, 1–6 (2017).

45. Shrestha, S. & Offer, S. M. Epigenetic regulations of GABAergic neurotransmission: Relevance for neurological disorders and epigenetic therapy. *Med. Epigenetics.* **4**, 1–19 (2016).

46. Butler, M. G., Usrey, K. M., Roberts, J. L., Manzardo, A. M. & Schroeder, S. R. Case Report 20q13.2-q13.33 deletion syndrome: A case report (2013). https://doi.org/10.3233/PGE-13065

47. Renaud, J. et al. Identification of intellectual disability genes showing circadian clock-dependent expression in the mouse hippocampus. *Neuroscience* **308**, 11–50 (2015).

48. Hasan, M. I., Hossain, M. A., Bhuiyan, P., Miah, M. S. & Rahman, M. H. A system biology approach to determine therapeutic targets by identifying molecular mechanisms and key pathways for type 2 diabetes that are linked to the development of tuberculosis and rheumatoid arthritis. *Life Sci.* **297**, 120483 (2022).

49. Ma, J. et al. Identification of miR-22-3p, miR-92a-3p, and miR-137 in peripheral blood as biomarker for schizophrenia. *Psychiatry Res.* **265**, 70–76 (2018).

50. Huang, F. et al. Investigation of gene regulatory networks associated with autism spectrum disorder based on MiRNA expression in China. *PLoS One* **10** (2015).

51. Antoniou, A. et al. The dynamic recruitment of TRBP to neuronal membranes mediates dendritogenesis during development. *EMBO Rep.* **19** (2018).

52. Wu, H. et al. Supplementation with selenium attenuates autism-like behaviors and improves oxidative stress, inflammation and related gene expression in an autism disease model. *J. Nutr. Biochem.* **107** (2022).

53. Tolins, M., Ruchirawat, M. & Landrigan, P. The developmental neurotoxicity of arsenic: Cognitive and behavioral consequences of early life exposure. *Ann. Glob. Health* **80**, 303–314. https://doi.org/10.1016/j.aogh.2014.09.005 (2014).

54. Ding, J. et al. Systematic comparative analysis of single cell RNA-sequencing methods. https://doi.org/10.1101/632216

55. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
56. Levin, J. Study: Single Cell Comparison: PBMC data.
57. Song, Z. et al. Identification of five hub genes based on single-cell RNA sequencing data and network pharmacology in patients with acute myocardial infarction. *Front. Public. Heal* **10** (2022).
58. Gopal, S. & Patro, K. & Kumar Sahu, K. *Normalization: A Preprocessing Stage*.
59. Cole, M. B. et al. Performance assessment and selection of normalization procedures for single-cell RNA-Seq. *Cell. Syst.* **8**, 315–328 (2019).
60. Lin, X., Ho, J., & Qian, G. Fundamentals of scRNA-Seq analysis.
61. Xu, C. & Su, Z. Identification of cell types from single-cell transcriptomes using a novel clustering method.
62. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
63. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
64. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).
65. Gregor, A. et al. De novo mutations in the genome organizer CTCF cause intellectual disability. *Am. J. Hum. Genet.* **93**, 124–131 (2013).
66. Mahi, N., Al, Najafabadi, M. F., Pilarczyk, M., Kouril, M. & Medvedovic, M. GREIN: An interactive web platform for re-analyzing GEO RNA-seq data. *Sci. Rep.* **9**, 1–9 (2019).
67. Nguyen, L. S. et al. Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. *Mol. Psychiatry* **17**, 1103–1115 (2012).
68. Jia, A., Xu, L. & Wang, Y. Venn diagrams in bioinformatics. *Brief. Bioinform.* **22** (2021). https://doi.org/10.1093/bib/bbab108
69. Pathan, M. et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **15**, 2597–2601 (2015).
70. Huang, D. W. et al. The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8** (2007).
71. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
72. Fabregat, A. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
73. Kelder, T. et al. WikiPathways: Building research communities on biological pathways. *Nucleic Acids Res.* **40**, D1301–D1307 (2012).
74. von Mering, C. et al. STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).
75. Kohl, M., Wiese, S. & Warscheid, B. Cytoscape: Software for visualization and analysis of biological networks. In *Data Mining in Proteomics* 291–303 (Springer, 2011).
76. Chin, C. H. et al. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* https://doi.org/10.1186/1752-0509-8-S4-S11 (2014).
77. Camilo Gomez, A. V. G. & Soulakova, J. N. Constructing UpSet plot for survey data with weights using SAS and R software. *Commun. Stat. Simul. Comput.* **52**, 2320–2326 (2023).
78. Hossain, M. A. et al. Study of kaempferol in the treatment of COVID-19 combined with Chikungunya co-infection by network pharmacology and molecular docking technology. *Inf. Med. Unlocked* **40** (2023).
79. Walter, W., Sánchez-Cabo, F. & Ricote, M. GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914 (2015).
80. Zhang, X. et al. Identification of key pseudogenes in nasopharyngeal carcinoma based on RNA-Seq analysis. *BMC Cancer* **21**, 1–14 (2021).
81. Zhou, G. et al. NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234–W241 (2019).
82. Fornes, O. et al. JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
83. Hsu, S. et al. MiRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **39**, 163–169 (2011).
84. Grondin, C. J. et al. Advancing exposure science through chemical data curation and integration in the comparative toxicogenomics database. *Environ. Health Perspect.* **124**, 1592–1599 (2016).
85. Nahm, F. S. ROC curve: Overview and practical use for clinicians. *Korean J. Anesthesiol.* **75**, 25–36 (2022).
86. Turck, N. et al. pROC: An open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinform.* **8**, 12–77 (2011).

Acknowledgements

Author contributions

K.M.I: Methodology, software, validation, writing original draft; M.H.R: Conceptualization, Methodology, Software, Validation, Writing - reviewing & editing, and Supervision; M.A.H: Software, Data curation, Writing, Validation, Writing, Original Draft; M.K: Data Curation; M.M.R: Writing, Review & Editing; P.S.A: Writing - Review & Editing; M.I.H: Data curation, writing, original draft; S.A.A: Writing - Review & Editing; M.H: Writing - Review & Editing.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This research requires no approval from any ethics committee as this research uses data that is publicly available.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-85162-4.

Correspondence and requests for materials should be addressed to M.H.R. or M.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.