

Biomarkers for Early Diagnosis of Autism Obtained Through Bioinformatics Analysis of Human Blood Samples

liuyin jin

Renmin Hospital of Wuhan University

gaohua wang (✉ wgh6402@163.com)

Renmin Hospital of Wuhan University

junyi xiang

Renmin Hospital of Wuhan University

chang shu

Renmin Hospital of Wuhan University

wenqian xu

Renmin Hospital of Wuhan University

linman wu

zhilan yang

Research Article

Keywords: Autism spectrum disorder, Differentially expressed gene, Early diagnosis, Energy metabolism, Immune regulation

Posted Date: January 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1238021/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Autism spectrum disorder (ASD) is a serious disorder that begins early in life and continues throughout life. Children diagnosed early with ASD are more responsive to therapeutic interventions and have less impairment of social and language functions whereas current diagnostic measures are mostly applied to children older than one year and lack the appropriate biological markers for early diagnosis of ASD. In this study, we used datasets GSE6575 from GEO to comprehensively analyze the mRNA expression profile of ASD, including 35 ASD samples and 12 normal control samples. Firstly, we looked for 50 different genes, and then we did an enrichment analysis of those genes. The results of enrichment analysis showed that these differential genes were mainly concentrated in the response to viruses, immune regulation of inflammation and energy metabolism. Secondly, String database was used to construct Protein-protein interaction (PPI) network of differential genes. Finally, cytoscape plug-ins Cytohubba was further used to search for Hub genes and the diagnostic value of these Hub genes was verified by subject operating characteristic curves. among which two genes IL2RB and PRF1 had good **sensitivity and specificity** for the early diagnosis of autism. the areas under ROC curve were 0.855,0.830 for IL2RB, PRF1. Data analysis using GEO database can provide some new insights into the etiology of ASD and provide some possible biomarkers and therapeutic targets for early diagnosis and treatment of ASD.

1. Introduction

Autism is a complex neurodevelopmental condition and an extremely heritable disorder, whose different forms are described in DSM-V as autism spectrum disorder (ASD) (Madore et al. 2016). The diagnosed incidence of ASD is estimated to be 1 in 68 individuals in the United States (Kosmicki et al. 2017). ASD is characterized by deficits in social communication alongside unusually narrow interests and repetitive behaviors (Tilot et al. 2016) and a heterogeneous disease with many factors leading to its occurrence. Genetic studies have demonstrated that hundreds of ASD-risk genes (Mercati et al. 2017) and copy number variations, polymorphisms and de novo mutations might be also involved in the ASD pathogenesis (Balestrieri et al. 2019). Changes in the number of microglia, dysregulated mTOR signalling and autophagy pathways (Kim et al. 2017; Huang et al. 2016) are involved in the pathogenesis of ASD. Maternal inflammation during prenatal development, environmental, immune and infectious factors are also hypothesized to be associated with the pathogenesis and severity of ASD (Theoharides et al. 2016; Irwin et al. 2019). The hypothesis for ASD pathogenesis mainly focuses on genetics and environment. However, to date, the precise pathogenesis of ASD remains undefined. (Wang et al. 2020; Chang et al. 2018)

Diagnostic assessment for ASD has evolved dramatically over the last two decades, particularly due to the creation of standardized diagnostic instruments such as the autism diagnostic observation schedule (ADOS) and the autism diagnostic interview-revised (ADI-R) (Cirnigliaro et al. 2017). ADOS is a diagnostic and assessment tool for ASD, the protocol consists of a series of structured and semi-structured tasks involving social interactions between the examiner and the person being evaluated. The examiner

observes and identifies various parts of the subject's behavior, assigns these parts to predetermined observation categories and classified observations were then combined to produce quantitative scores for analysis(Lord et al. 1989). ADI-R is a standardized test rather than an Autism scale, which was Revised by Lord C and M. Rutter et al in 1994 on the basis of ADI. The ADI-R consists of 93 items belonging to three major autism symptom categories: language blindness and communication, social interaction, and stereotyped behavior(Lord et al. 1994). However, these scales should only be used in infants and toddlers older than 30 months, none of these tests can be used to diagnose autism early. It should be noted that the first three years of life are critical for autism recovery, children diagnosed early with ASD have less impairment of social and language functions. when behavioral interventions can be more beneficial.(Smith et al. 2019) Although studies have shown that early diagnosis leads to early intervention and better outcomes, there is a significant time lag between the onset and diagnosis of autism symptoms(Roberts et al. 2012). A biomarker is a measurable indicator of some biological condition or state. Identification of biomarkers would be a key step for ASD. However, there is a lack of effective biological indicators to diagnose ASD, especially lack of effective early diagnostic biomarker.

Hence, the aim of this study was to identify biological functions and potential diagnostic biomarkers and related to ASD from the Gene Expression Omnibus (GEO)(Edgar et al. 2002). Further, the biological processes involved were analyzed using gene ontology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment(Huang et al. 2009) and gene set enrichment analysis (GSEA) (Subramanian et al. 2005) pathways for the differential genes. In addition, the hub genes screened via protein–protein interaction (PPI) network were selected for their functional similarity and their diagnostic value was assessed. Our study may provide some insights into the molecular mechanisms underlying ASD based on its pathophysiology.

2. Material And Methods

2.1 Acquisition of RNA information and data process

The GEO database includes high-throughput gene expression data submitted by research institutions around the world. It was founded in 2000 primarily to express chip data but has since expanded to include many other high-throughput data with the popularity of databases, such as: Chromatin structure, genome–protein interaction(Edgar et al. 2002). The gene expression dataset (GSE6575) (Gregg et al. 2008)was downloaded from the GEO database. (GPL570[HG-U133_Plus_2]) Affymetrix Human Genome U133 Plus 2.0 Array), by use GEO query package(Davis and Meltzer 2007), then we used limma package for difference analysis(Wettenhall and Smyth 2004), the P-value of genes was calculated using t test method, and Benjamini and Hochberg's method was used to calculate the adjusted P-value. The differentially expressed genes were screened out by the following selection criteria: $|\log_2FC|$ (an absolute \log_2 value in the fold change of the expression of the genes) >0.6 and $P\text{-value} < 0.05$ (Wang et al. 2019; Dalman et al. 2012). GGploT2 package(Walter et al. 2015) for visualization analysis, and Complex Heatmap package for thermal map visualization analysis(Gu et al. 2016). The species selected was

Homo sapiens, and the data type was microarray expression profiles. The whole blood samples included 35 ASD samples and 12 control healthy samples (CON).

2.2 Enrichment analysis

GSEA is based on the idea of using a predefined Set of genes, often derived from functional annotations or the results of previous experiments, to sequence genes according to the degree of differential expression in two types of samples. Then check to see if the set of genes is enriched at the top or bottom of the sequencing list. Gene collection enrichment analysis detects changes in expression of collections of genes rather than individual genes, and therefore can include these subtle changes in expression and expect better results (Subramanian et al. 2005). All genetic information of ASD and CON samples were uploaded to GSEA for further analysis. We also used R package Cluster profiler for enrichment analysis of GO and KEGG combined logFC R package GOplot and ggplot2 for visual analysis, meanwhile used cytoscape plus-in cluego and metascape for further validation and visualization (Yu et al. 2012; Walter et al. 2015).

2.3 Gene cluster identification and protein-protein interaction (PPI) network analysis

The differential genes in ASD samples were uploaded to STRING database (Szklarczyk et al. 2017) to obtain the protein network interaction diagram. The result of STRING database analysis was imported into Cytoscape v.3.9.0, CytoHubba, a Cytoscape plugin (Chin et al. 2014), was used to identify the 10 top hub genes of the merged network.

2.4 Setting the cut-off score based on receiver operating characteristic curve analysis

Receiver operating characteristic (ROC curve) reflects the sensitivity and specificity of comprehensive index by composition method is to reveal the relationship between sensitivity and specificity of it by using a continuous variable set out a number of different threshold calculated a series of sensitivity and specificity and sensitivity as the ordinate 1 - specificity of abscissa are drawn. The larger the area under the curve, the higher the diagnostic accuracy (Akobeng 2007). The maximum value of the sum of specificity and sensitivity was used as the cut-off score for each hub gene. The "pROC" and "ggplot2" packages of the R software were applied for ROC curve analysis and visualization (Robin et al. 2011).

3. Results

3.1 Gene expression levels of samples were processed and differential genes were screened

According to the sample information and data matrix, 50 differential genes were screened from the ASD samples, including 38 up-regulated genes and 12 down-regulated genes. The screening criteria for differentially expressed genes were as follows: P-value < 0.05 and $|\log FC| > 0.6$. Based on the analysis of

gene expression of the samples, the volcano plot was made as shown in Figure 1. The expression of genes with high and low expression of 20 top genes respectively was made using a heatmap as shown in Figure 2.

3.2 The differentially expressed genes in the ASD samples were mainly enriched in immune response and energy metabolism.

GSEA, Cytoscape and Metascape software were used for enrichment analysis of the samples' genes. Firstly, all gene expression information in ASD and CON samples was uploaded to GSEA software, and the hallmark gene set database was used to analyze genes at the overall level of expression profile. The significantly enriched gene sets were set at a default cut-off as $P\text{-value} < 0.05$ and $FDR < 0.25$. The enrichment analysis of gene sets revealed that the gene sets were significantly enriched in immune-related functions and energy metabolism were significantly enriched in ASD samples as shown in Figure 3.

Enrichment analysis of GO and KEGG combined \log_2FC was conducted on 50 differentially expressed genes in the ASD samples by using R package ClusterProfiler. GO and KEGG combined \log_2FC enrichment analysis showed that the differentially expressed genes in ASD samples were mainly related to immune response in the biological process as shown in Figure 4. Use Cytoscape ClueGo plug-in to visualize the interaction network of biological process, as shown in Figure 5.

A total of 50 differentially expressed genes were then uploaded to Metascape, further enrichment analysis. Through Metascape analysis of the pathway enrichment of differentially expressed genes in ASD samples, it indicated that the differentially expressed genes are mainly enriched in the leukocyte mediated immunity and response to virus shown in Figure 6a and Figure 6b.

3.3 Construction of protein-protein interaction (PPI) network and further excavation of gene clusters involved in immune system-related biological pathway

In order to screen out the core genes from the differentially expressed genes in the ASD samples, 50 differentially expressed genes were uploaded to the STRING for further analysis, and 48 nodes plus 85 edges were obtained. Local clustering coefficient is 0.474 and PPI enrichment $P\text{-value} < 1 \times 10^{-16}$, and shown in Figure 7. The data file was then processed with Cytoscape as shown in Figure 8. Cytohubba was used to process the network data to identify top ten hub genes KLRF1, TBX21, CST7, NKG7, GZMH, IL2RB, RUNX3, PRF1, GZMB and GNLY shown in Figure 9.

3.4 Using hub genes for ASD diagnosis

Top ten hub genes were obtained by using Cytoscape's Cytohubba plug-ins. The accuracy of diagnosis was evaluated by ROC curve, and the areas under ROC curve were 0.745, 0.767 for KLRF1, TBX21 as shown in Fig. 10a. The areas under the ROC curves were 0.802, 0.800 for RUNX3, GZMB as shown in Fig. 10b. The areas under the ROC curves were 0.736, 0.769 for CST7, NKG7 as shown in Fig. 10c. The areas under the ROC curves were 0.855, 0.733 for IL2RB, GZMH as shown in Fig. 10d. The areas under the ROC curves were 0.830, 0.782 for PRF1, GNLY as shown in Fig. 10e.

4. Discussion

Worldwide, ASD is a neurodevelopmental disease that appears early in life and lasts for a lifetime. In recent decades, the incidence of ASD has gradually increased (Duffy and Als 2012). Most patients are discovered and diagnosed with ASD when they are 2-3 years old (Chen et al. 2018). The current scale of ASD is mainly used for children around 2 years old, which lacks the method of early diagnosis. More and more researches are underway to develop a noninvasive and quantitative clinical trial. However, there are no specific and sensitive biomarkers available for the diagnosis and treatment of ASD. Therefore, to identify effective diagnostic biomarkers for ASD, we analyzed mRNA expression profile data from 35 ASD patients and 12 healthy controls. A total of 50 differential genes were identified in ASD. GSEA, GO, and KEGG enrichment analyses revealed that these enriched modules and pathways were closely related to the immune responses observed in ASD. In addition, several Hub genes found in PPI network have high functional similarity and diagnostic value of ASD.

In the first part of this study, we used the GSE6575 dataset, which included 35 ASD patients and 12 healthy controls, to screen a total of 50 differential genes. In order to study the functional pathways involved in these differential genes, we performed GSAE, GO and KEGG enrichment analyses. In the enrichment analysis of GO and KEGG functions, it was found that these different genes were mainly concentrated in the immunomodulatory response, mainly the defense response to the virus. Prenatal viral infection is considered a possible cause of autism. Some studies have shown that viral infection during the critical period of fetal neurodevelopment can lead to an increased risk of autism in offspring (Theoharides et al. 2016). Valayi S. et al found that antibodies to EBV increased significantly in children with autism through serum samples from 45 children with autism and 45 healthy controls. Possible mechanisms by which viral infections lead to autism include direct teratogenic effects and indirect effects of inflammation or maternal immune activation on the developing brain (Valayi et al. 2017). Xiaohong Li. et al found that proinflammatory cytokines (TNF α , IL-6 and GM-CSF), Th1 cytokine (IFN- γ) and chemokine (IL-8) were significantly increased in the brains of ASD patients compared with the controls (Li et al. 2009). Studying the role of viral infection in the development of ASD may be a way to improve the prognosis of ASD in the future.

In the second part of this study, we conducted a GSEA study to investigate the biological functions of differential genes associated with ASD. The results of GSEA enrichment analysis mainly focused on leukocyte mediated immune regulation pathway, substance metabolism pathway and mTORC1 signaling pathway. Translation is one of the most energy-intensive cellular processes, and the mTORC1 pathway further regulates the rate of translation by continuously sensing and integrating various environmental stressors such as uv exposure and hypoxia, which are implicated in the development of autism (Hooshmandi et al. 2020; Sonenberg and Hinnebusch 2009; Roux and Topisirovic 2018). Therefore, changes in energy metabolism may also play a key role in the etiology of ASD.

In the third part of this study, a PPI network was constructed by using Cytoscape software's Cytohubba plug-in. A total of top ten genes were found from the networks, and these genes had multiple interactions

in the network. Further study of these genes may shed light on the pathophysiological mechanisms of ASD. Among the ten differentially expressed genes, we selected six genes with the most significant expression differences, namely RSAD2, IFI27, IFIT2, KLRF1, GNLY and GZMB, among which the first three genes were up-regulated and the last three genes were down-regulated. These six genes are all related to immune regulation and inflammatory response, which further indicates that there is an imbalance of immune inflammatory regulation network in children with autism. Immunomodulatory and inflammatory responses play key roles in the etiology and pathophysiology of ASD. In terms of diagnostic value, the AUC of ten hub genes was analyzed. The AUC of two genes was greater than 0.83, namely IL2RB 0.855, PRF1 0.830 indicating that these two genes had moderate accuracy as diagnostic markers. This may be a promising target for the diagnosis of ASD(Akobeng 2007).

Finally, our study has some limitations. The first study used blood samples from 36 individuals with ASD and 12 healthy controls. In order to determine the diagnostic accuracy of ASD-related pivotal genes, it is necessary to further expand the sample size, as well as conduct external validation of the specificity and sensitivity of diagnostic indicators. Second, we only analyzed the changes in mRNA expression in the blood of individuals with ASD and we don't know exactly what the changes in the brain. The third single microarray analysis may lead to high false positive rate and unilateral results, so experimental studies such as real-time PCR, western blot, immunohistochemistry assays analysis are required to verify, so as to fully clarify the role of hub genes and the potential mechanism of ASD. Fourth, due to the high heterogeneity of autism and the lack of clinical data, we were unable to assess the relationship between severity and our predicted diagnostic indicators. Therefore, we need to include more clinical and demographic characteristics of ASD patients in future studies for further subgroup analysis.

These data indicate multiple pathophysiological pathways and several biomarkers of ASD, which are consistent with our current research and understanding of the disease. This hypothesis-generating research could provide new insights into the etiology and pathophysiology of autism and improve some possible biomarkers for its diagnosis and treatment.

5. Conclusion

In short, this study intends to comprehensively explore the molecular information mechanism of ASD through bioinformatics analysis. To determine which molecules may be involved in the development of the disease. In addition, three candidate genes were identified as potential diagnostic biomarkers by functional enrichment analysis and PPI network analysis. Further experiments are needed to verify the validity of the study.

References

Akobeng AK (2007) Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* 96 (5):644-647. doi:10.1111/j.1651-2227.2006.00178.x

- Balestrieri E, Cipriani C, Matteucci C, Benvenuto A, Coniglio A, Argaw-Denboba A, Toschi N, Bucci I, Miele MT, Grelli S, Curatolo P, Sinibaldi-Vallebona P (2019) Children With Autism Spectrum Disorder and Their Mothers Share Abnormal Expression of Selected Endogenous Retroviruses Families and Cytokines. *Front Immunol* 10:2244. doi:10.3389/fimmu.2019.02244
- Chang Q, Yang H, Wang M, Wei H, Hu F (2018) Role of Microtubule-Associated Protein in Autism Spectrum Disorder. *Neurosci Bull* 34 (6):1119-1126. doi:10.1007/s12264-018-0246-2
- Chen J, Lei L, Tian L, Hou F, Roper C, Ge X, Zhao Y, Chen Y, Dong Q, Tanguay RL, Huang C (2018) Developmental and behavioral alterations in zebrafish embryonically exposed to valproic acid (VPA): An aquatic model for autism. *Neurotoxicol Teratol* 66:8-16. doi:10.1016/j.ntt.2018.01.002
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8 Suppl 4:S11. doi:10.1186/1752-0509-8-S4-S11
- Cirnigliaro M, Barbagallo C, Gulisano M, Domini CN, Barone R, Barbagallo D, Ragusa M, Di Pietro C, Rizzo R, Purrello M (2017) Expression and Regulatory Network Analysis of miR-140-3p, a New Potential Serum Biomarker for Autism Spectrum Disorder. *Front Mol Neurosci* 10:250. doi:10.3389/fnmol.2017.00250
- Dalman MR, Deeter A, Nimishakavi G, Duan ZH (2012) Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* 13 Suppl 2:S11. doi:10.1186/1471-2105-13-S2-S11
- Davis S, Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23 (14):1846-1847. doi:10.1093/bioinformatics/btm254
- Duffy FH, Als H (2012) A stable pattern of EEG spectral coherence distinguishes children with autism from neuro-typical controls - a large case control study. *BMC Med* 10:64. doi:10.1186/1741-7015-10-64
- Edgar R, Domrachev M, Lash A (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207-10. doi:doi: 10.1093/nar/30.1.207
- Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, Croen LA, Ozonoff S, Hansen R, Pessah IN, Sharp FR (2008) Gene expression changes in children with autism. *Genomics* 91 (1):22-29. doi:10.1016/j.ygeno.2007.09.003
- Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32 (18):2847-2849. doi:10.1093/bioinformatics/btw313
- Hooshmandi M, Wong C, Khoutorsky A (2020) Dysregulation of translational control signaling in autism spectrum disorders. *Cell Signal* 75:109746. doi:10.1016/j.cellsig.2020.109746
- Huang d, W., Sherman B, Lempicki R (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44-57. doi:doi: 10.1038/nprot.2008.211

Huang WC, Chen Y, Page DT (2016) Hyperconnectivity of prefrontal cortex to amygdala projections in a mouse model of macrocephaly/autism syndrome. *Nat Commun* 7:13421. doi:10.1038/ncomms13421

Irwin JL, Yeates AJ, Mulhern MS, McSorley EM, Strain JJ, Watson GE, Grzesik K, Thurston SW, Love TM, Smith TH, Mruzek DW, Shamlaye CF, Monthy C, Myers GJ, Davidson PW, van Wijngaarden E (2019) Maternal Gestational Immune Response and Autism Spectrum Disorder Phenotypes at 7 Years of Age in the Seychelles Child Development Study. *Mol Neurobiol* 56 (7):5000-5008. doi:10.1007/s12035-018-1424-y

Kim HJ, Cho MH, Shim WH, Kim JK, Jeon EY, Kim DH, Yoon SY (2017) Deficient autophagy in microglia impairs synaptic pruning and causes social behavioral defects. *Mol Psychiatry* 22 (11):1576-1584. doi:10.1038/mp.2016.103

Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M, Karczewski KJ, Cutler DJ, Devlin B, Roeder K, Buxbaum JD, Neale BM, MacArthur DG, Wall DP, Robinson EB, Daly MJ (2017) Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet* 49 (4):504-510. doi:10.1038/ng.3789

Li X, Chauhan A, Sheikh AM, Patil S, Chauhan V, Li XM, Ji L, Brown T, Malik M (2009) Elevated immune response in the brain of autistic patients. *J Neuroimmunol* 207 (1-2):111-116. doi:10.1016/j.jneuroim.2008.12.002

Lord C, Rutter M, Goode S, J H, Jordan H, Mawhood L, E S (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior *Journal of Autism and Developmental Disorder* Vol 19

Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *A J Autism Dev Disord* (5):659-85. doi:doi: 10.1007/BF02172145

Madore C, Leyrolle Q, Lacabanne C, Benmamar-Badel A, Joffre C, Nadjar A, Laye S (2016) Neuroinflammation in Autism: Plausible Role of Maternal Inflammation, Dietary Omega 3, and Microbiota. *Neural Plast* 2016:3597209. doi:10.1155/2016/3597209

Mercati O, Huguet G, Danckaert A, Andre-Leroux G, Maruani A, Bellinzoni M, Rolland T, Gouder L, Mathieu A, Buratti J, Amsellem F, Benabou M, Van-Gils J, Beggiato A, Konyukh M, Bourgeois JP, Gazzellone MJ, Yuen RK, Walker S, Delepine M, Boland A, Regnault B, Francois M, Van Den Abbeele T, Mosca-Boidron AL, Faivre L, Shimoda Y, Watanabe K, Bonneau D, Rastam M, Leboyer M, Scherer SW, Gillberg C, Delorme R, Cloez-Tayarani I, Bourgeron T (2017) CNTN6 mutations are risk factors for abnormal auditory sensory perception in autism spectrum disorders. *Mol Psychiatry* 22 (4):625-633. doi:10.1038/mp.2016.61

Roberts JE, Hatton DD, Long AC, Anello V, Colombo J (2012) Visual attention and autistic behavior in infants with fragile X syndrome. *J Autism Dev Disord* 42 (6):937-946. doi:10.1007/s10803-011-1316-8

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi:10.1186/1471-2105-12-77

Roux PP, Topisirovic I (2018) Signaling Pathways Involved in the Regulation of mRNA Translation. *Mol Cell Biol* 38 (12). doi:10.1128/MCB.00070-18

Smith AM, King JJ, West PR, Ludwig MA, Donley ELR, Burrier RE, Amaral DG (2019) Amino Acid Dysregulation Metabotypes: Potential Biomarkers for Diagnosis and Individualized Treatment for Subtypes of Autism Spectrum Disorder. *Biol Psychiatry* 85 (4):345-354. doi:10.1016/j.biopsych.2018.08.016

Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136 (4):731-745. doi:10.1016/j.cell.2009.01.042

Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, Paulovich A, Pomeroy S, Golub T, Lander E, Mesirov J (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43). doi:doi: 10.1073/pnas.0506580102

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45 (D1):D362-D368. doi:10.1093/nar/gkw937

Theoharides TC, Tsilioni I, Patel AB, Doyle R (2016) Atopic diseases and inflammation of the brain in the pathogenesis of autism spectrum disorders. *Transl Psychiatry* 6 (6):e844. doi:10.1038/tp.2016.77

Tilot AK, Bebek G, Niazi F, Altemus JB, Romigh T, Frazier TW, Eng C (2016) Neural transcriptome of constitutional Pten dysfunction in mice and its relevance to human idiopathic autism spectrum disorder. *Mol Psychiatry* 21 (1):118-125. doi:10.1038/mp.2015.17

Valayi S, Eftekharian MM, Taheri M, Alikhani MY (2017) Evaluation of antibodies to cytomegalovirus and Epstein-Barr virus in patients with autism spectrum disorder. *Hum Antibodies* 26 (3):165-169. doi:10.3233/HAB-180335

Walter W, Sanchez-Cabo F, Ricote M (2015) GOrplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* 31 (17):2912-2914. doi:10.1093/bioinformatics/btv300

Wang H, Zhang M, Xie Q, Yu J, Qi Y, Yue Q (2019) Identification of diagnostic markers for major depressive disorder by cross-validation of data from whole blood samples. *PeerJ* 7:e7171. doi:10.7717/peerj.7171

Wang X, Ding R, Song Y, Wang J, Zhang C, Han S, Han J, Zhang R (2020) Transcutaneous Electrical Acupoint Stimulation in Early Life Changes Synaptic Plasticity and Improves Symptoms in a Valproic

Wettenhall J, Smyth G (2004) limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* 20(18):3705-6. doi:doi: 10.1093/bioinformatics/bth449

Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5):284-287. doi:10.1089/omi.2011.0118

Figures

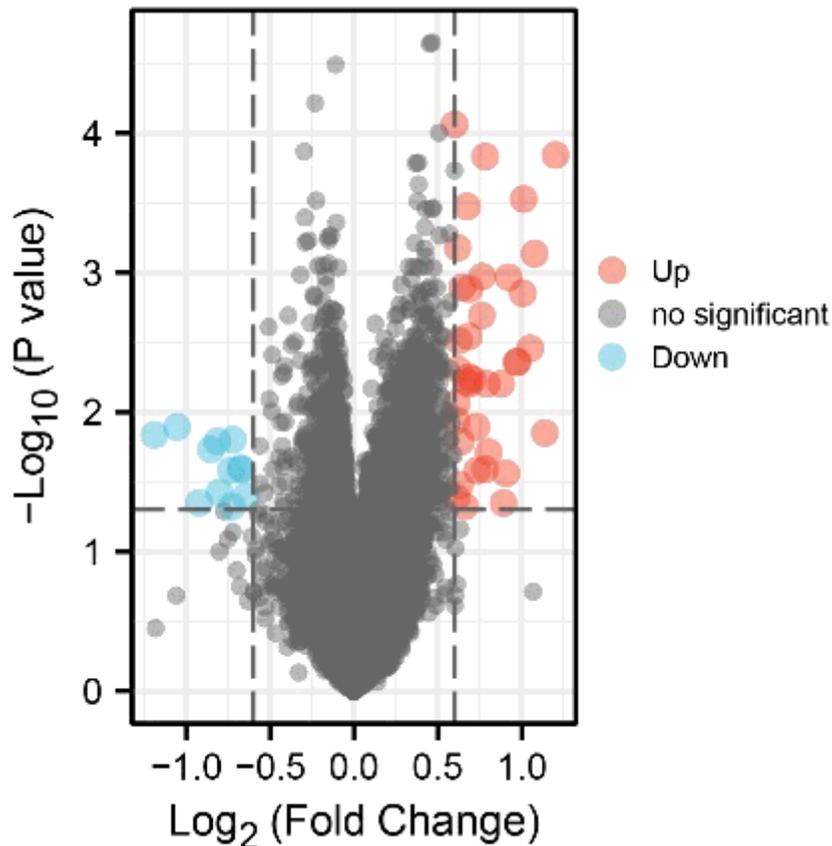


Figure 1

Differentially expressed genes in ASD and CON samples were shown in the volcano plot, with blue dots representing significantly down-regulated genes in ASD samples and red dots representing significantly up-regulated genes. CON, healthy control; ASD, Autism spectrum disorder

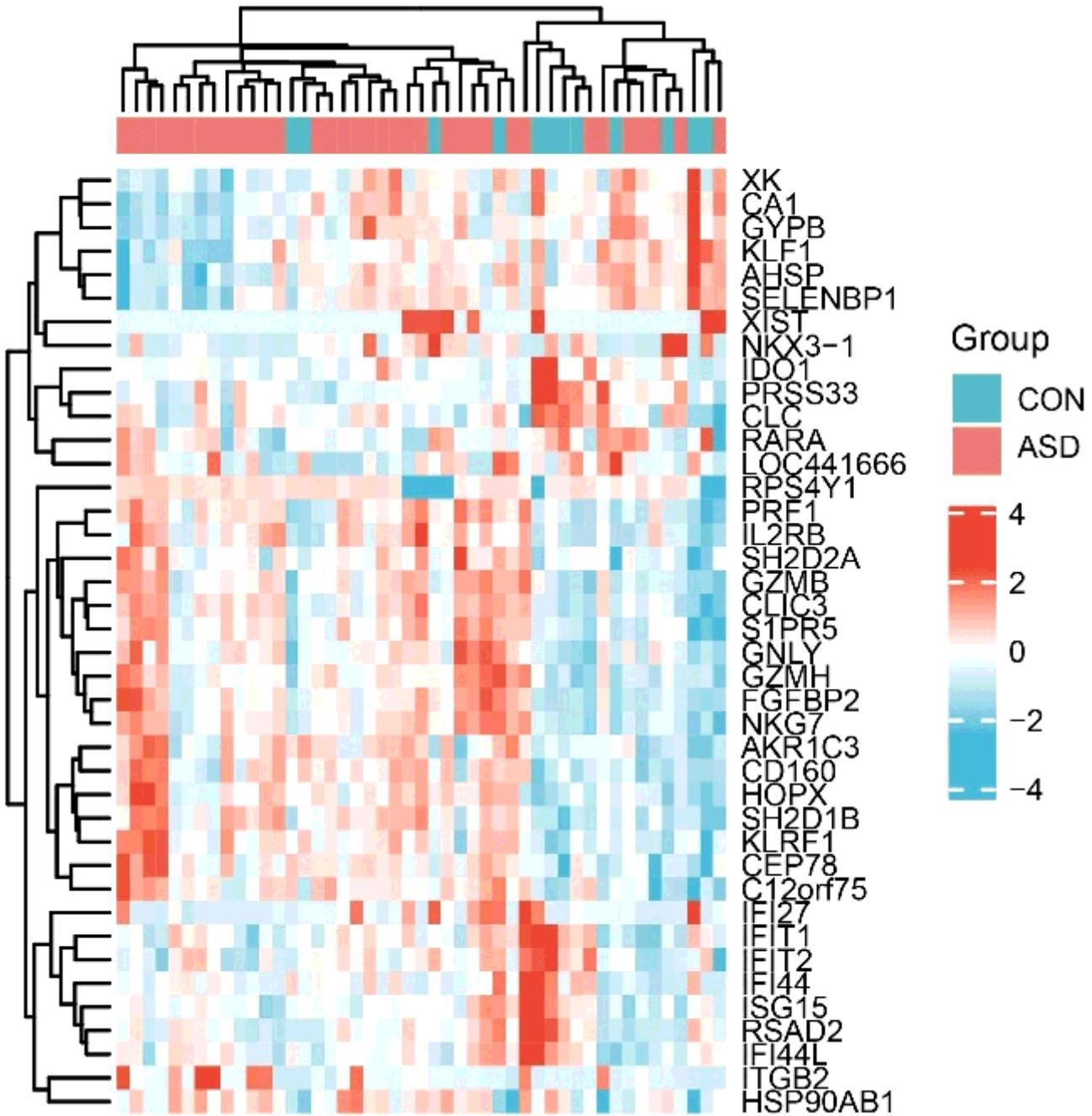


Figure 2

heatmaps were used to display the expression of top 20 genes with high and low expression, blue represents the CON and red represents the ASD

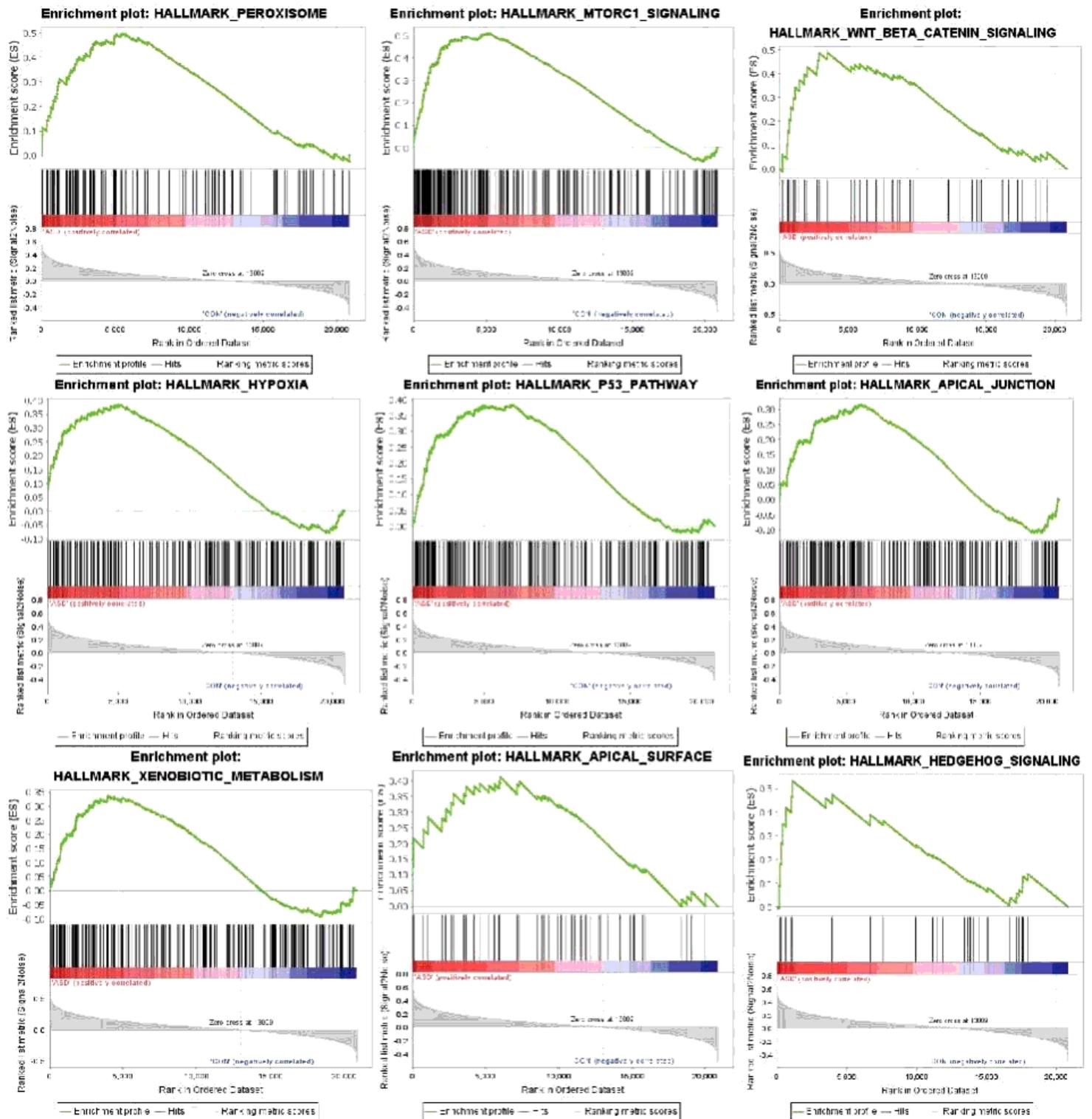


Figure 3

h.all.v 6.2.symbols.gmt [Hallmarks] gene set database was used to analyze the whole gene expression value of the ASD and CON samples. GSEA first filtered the gene set according to the number of genes contained in the gene set, with the minimum number of 15 genes and the maximum number of 500 genes by default. Significant gene sets were cut-off by FDR < 0.25 and P-value < .05

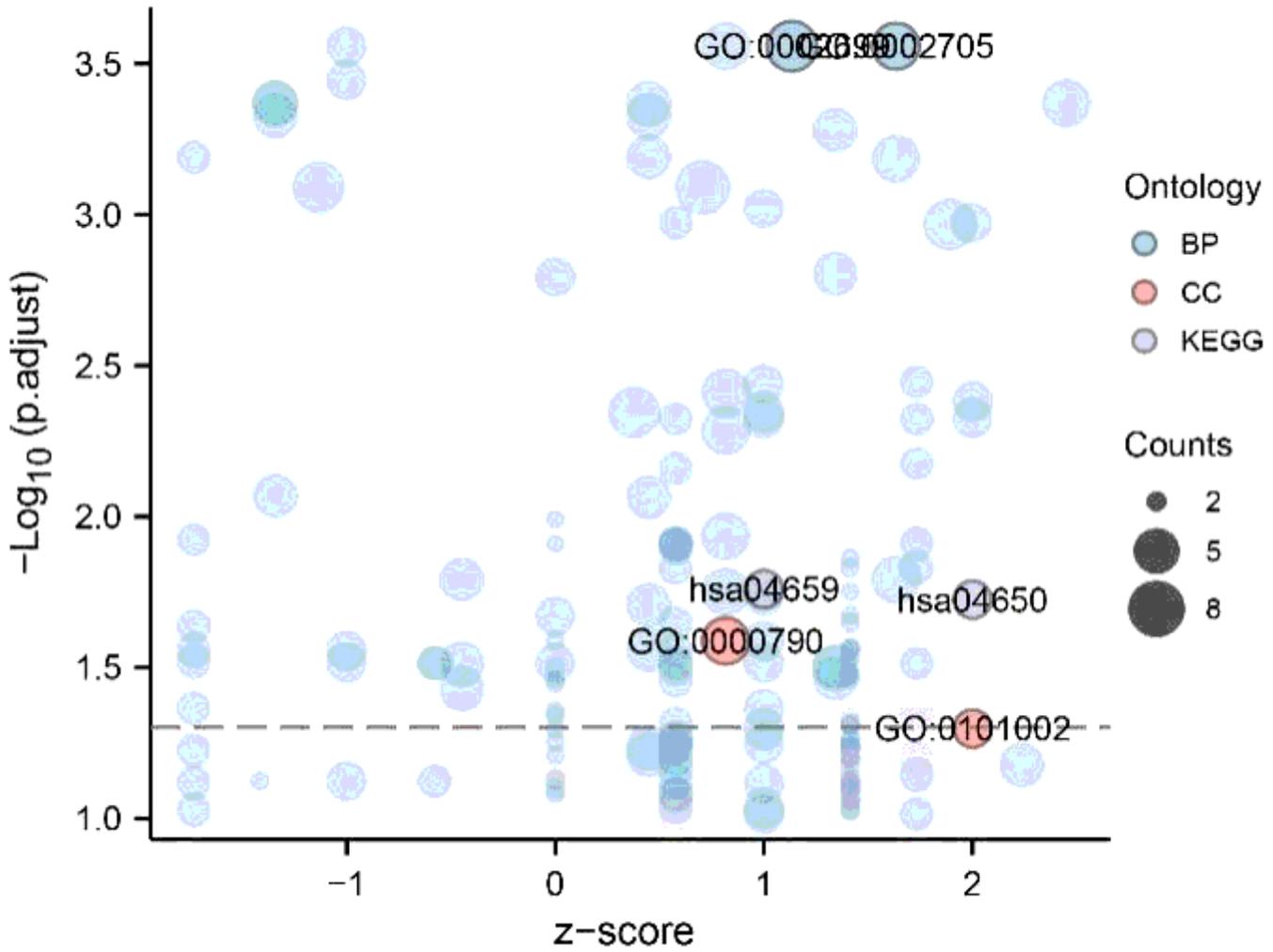


Figure 4

Bubble diagram analysis by GO and KEGG combined logFC, Blue dots represent pathways enriched in BP, pink dots represent pathways enriched in CC and purple dots represent pathways enriched in KEGG. The size of the dot represents the number of genes enriched in the pathway. Gene enrichment pathway map obtained using the plug-in cluego

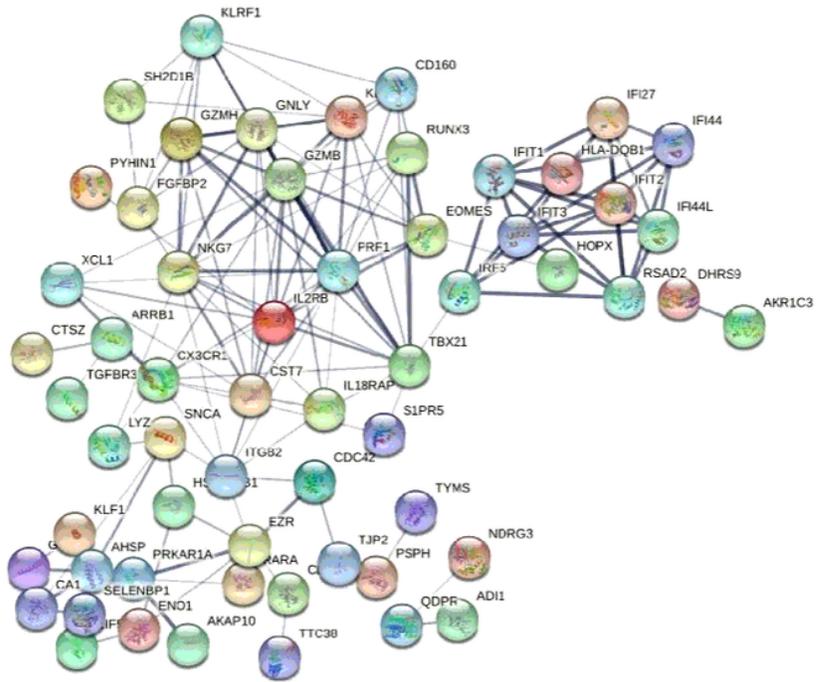


Figure 7

The PPI network analysis using STRING shows the interaction between 50 differential genes

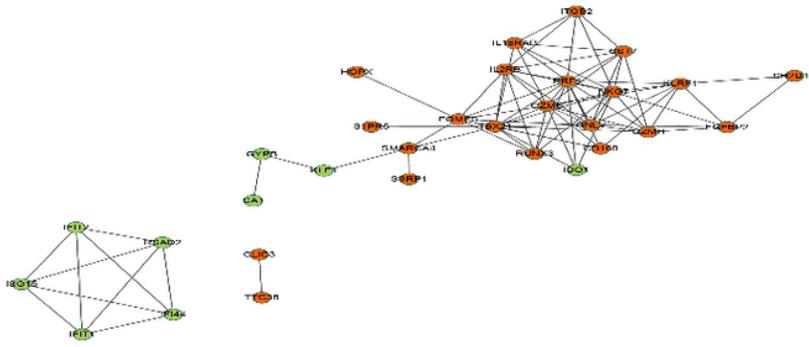


FIGURE 7 The PPI network analysis using STRING then processed with Cytoscape shows the interaction between 50 differential genes; the red dots represent up-regulated genes and the green dots represent down-regulated genes

Figure 8

The PPI network analysis using STRING then processed with Cytoscape shows the interaction between 50 differential genes, the red dots represent up-regulated genes and the green dots represent down-regulated genes

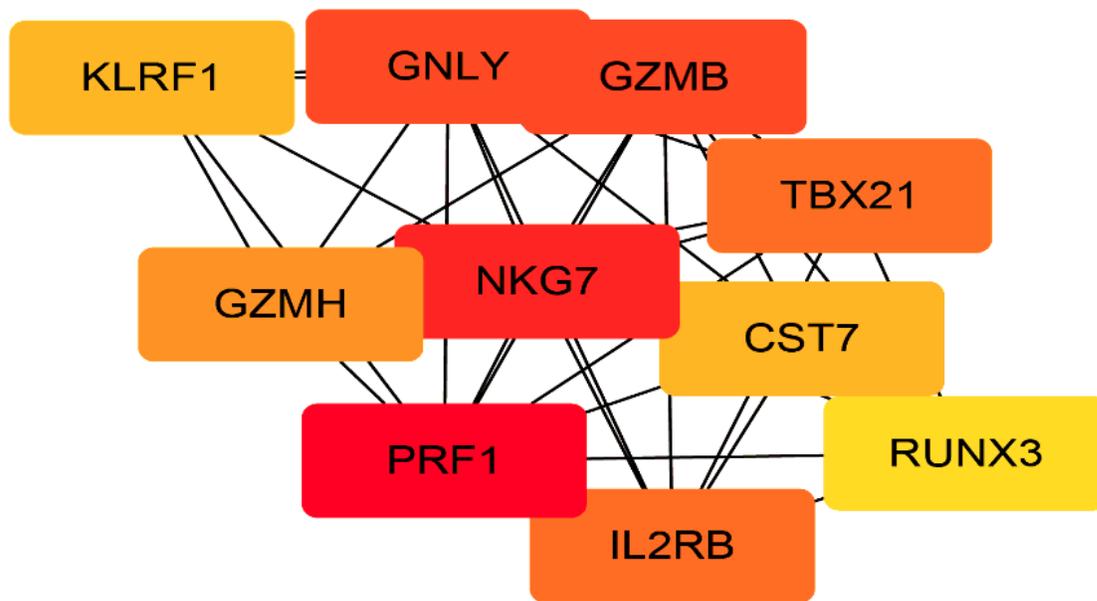


FIGURE 8 Top ten hub genes were identified by processing the PPI network data using Cytohubba.

Figure 9

Top ten hub genes were identified by processing the PPI network data using Cytohubba.

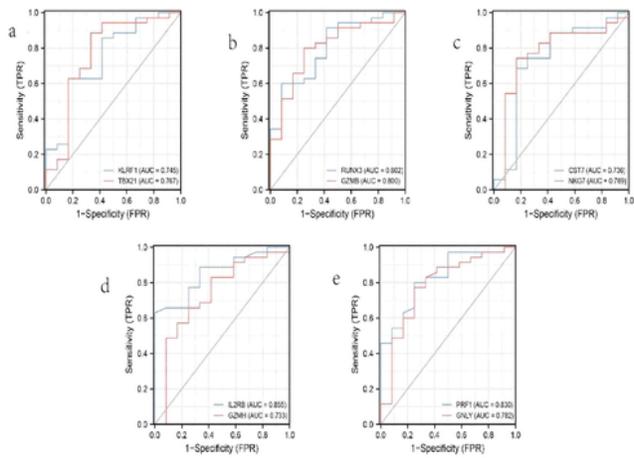


FIGURE 9 Validation of diagnostic value of the hub genes for ASD. (A) (B) Receiver operating characteristic curve of the hub genes for diagnosis of ASD.

Figure 10

Validation of diagnostic value of the hub genes for ASD. (a-e) Receiver operating characteristic curve of the hub genes for diagnosis of ASD.