Analysis of lncRNA, miRNA, mRNA-associated ceRNA networks in Autism

Asal Tabatabaei Azad¹, Mohammad-Reza Mehrabi², Yasamin Zarinzad¹, Atefeh Noori³, Hossein Nemati¹, Amir Shahbazi¹, Parisa Badameh¹, Seyed Roham Mohammadnezad kenari¹, Parham Arabzadeh¹, Amir-Reza Javanmard*¹

¹-Genetics Department, Faculty of Biological Science, Tarbiat Modares University, Tehran, Iran
²-Microbial Biotechnology, School of Biology, College of Science, university of Tehran, Iran.
³-Biology Department, Faculty of Basic Science, Shahed University, Tehran, Iran.

* Corresponding Author: Amir-Reza Javanmard and Hossein Nemati, PhD in Genetics Department, Faculty of Biological Science, Tarbiat Modares University, Tehran, Iran,
J.amirreza@modares.ac.ir and Hossein.neamati@modares.ac.ir

Abstract

Autism is a multifactorial behavioral disease, which is caused by different environmental and genetic alterations. In this disease, various molecular pathways such as inflammation and cell-cell connections are affected. In this study, by analyzing RNA-seq data and bioinformatics approaches, candidate genes were selected for further analysis, and the expression of candidate genes in the blood samples of autistic people was compared to normal. Finally, the candidate genes showed potential as biological markers for prognosis in autism. Through the analysis of these candidate genes, the researchers were able to identify changes in the expression of certain genes that were associated with the disease. By understanding how these genes are involved in the disease, researchers can develop better treatments and diagnose the disease earlier.

Key words: Autism, Bioinformatics, LncRNA, miRNA
Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder and a complex syndrome consisting of a wide range of symptoms caused by abnormalities in the structure and function of the brain. This condition is characterized by difficulties communicating and social skills, delayed speech and motility, abnormal reactions to sound and other stimuli, repetitive behaviors and interests, and digestive problems. Autism is inherited in 80% of cases but environmental factors, random mutations and epigenetics can also play a role in its occurrence[1-3].

The prevalence of autism around the world has increased in recent years due to advances in screening and diagnostic methods. In the United States in 2022, the prevalence of autism spectrum disorders was about 1 in 44 children. Moreover, it was similar between races and there was no difference between black and white children. ASD is potentially affected by gender, with boys being about 4 times more likely to be diagnosed than girls.

Autism symptoms usually appear in the first three years of life and it can be diagnosed at the age of 18 months or even earlier, although the diagnosis is more reliable at 2 years of age[4]. However, some people are not diagnosed until adolescence. Early detection of autism is a vital step, for it has a significant impact on improving symptoms and raising the life quality of ASD children.

According to science, ASD gender-dependence stems from the fact that women have two x chromosomes and alterations to one chromosome can be compensated by another healthy chromosome. But for men, having only one x chromosome, defects on the x chromosome can lead to an unhealthy phenotype. For instance, a defect in the NLGN4 gene, which acts to create and maintain synaptic structures and is located on two sex chromosomes (one on the x chromosome and one on the y chromosome), has been shown to be involved in autism. NLGN4Y protein does not have the required function due to having a different single amino acid and the main function of this protein is related to NLGN4X protein. The region around this amino acid in the NLGN4X gene is a mutation-sensitive region, and different autism-associated variants have been seen in people with autism[5].
Genomic studies since 2005 have led to the identification of numerous CNVs, SNVs and SNPs, with CNVs accounting for about 15% and SNVs for 7% of autism cases. However, their association with autism is limited to a small number of people and has a very low prevalence in society[6, 7]. In addition, more than 1,000 autism-related genes are candidates in case-control studies. Most of these genes encode proteins involved in synaptic structure and function, and the others have functions in cell proliferation, chromatin rearrangement, and transcriptional regulation (released 20 July 2022, gene.sfari.org).

In humans, almost 80% of the genome is transcribed, with only 2% of these transcripts encoding proteins and the rest including non-coding RNA[8]. Two main groups of ncRNAs include microRNAs (miRNA) and long noncoding RNAs (lncRNAs), which have a regulatory role in various cellular processes. Changes in them are related to the occurrence of diseases and abnormalities[9].

To date, a large number of ncRNAs are known to be associated with neurodevelopment. Changes in their expression level are related to neurodevelopmental diseases, including autism[10]. Analysis of circulating serum samples from children with autism has led to the identification of a number of differentially expressed miRNAs[11]. The interaction between these miRNAs and the critical ASD genes indicates a potential role for the ncRNAs in the molecular pathways involved in ASD. Moreover, examining prefrontal cortex and cerebellum tissues of Postmortem brain, identified more than 200 autism-related lncRNAs with regions containing genes associated with neurodevelopmental disease[12]. LncRNAs can affect the regulation of gene expression through sponging miRNAs[13]. The study of cross-talk between miRNA-lncRNA-mRNA as a ceRNA network analysis has provided a way to identify biomarkers and molecular pathways associated with autism.

In this study we aimed to identify and screen potential biomarkers for ASD. We also studied the interactions between differentially expressed RNAs in the autistic brain through the construction and analysis of a CERNA network. Based on the connections between DEmRNAs, DEmiRNAs and DElncRNAs we were able to identify a biomarker for ASD.

METHODS

Differential expression of lncRNAs, mRNAs, and microRNA

In order to investigate and find differentially expressed RNA, including mRNA, microRNA and LncRNA in autism, the raw data of sequencing (SRP174367) was taken from the SRA database (http://www.ncbi.nlm.nih.gov/sra). The selection of raw data was done based on a search for the word autism. Information files related to phenotype and expression level were extracted from primary data. Genes were annotated using HTSeq-Count and then 16 numbers of autism samples and 14 healthy samples were compared in terms of expression levels. Using the R package
edgeR with the threshold of P-value < 0.05 and |log2-fold change (FC)| > 2 dRNAs were determined and then used for further analysis.

**IncRNA-miRNA-mRNA related network**

Mircode database (Version 11; http://www.mircode.org/mircode/, accessed on 25 June 2022), targetscan (Version 8.0; http://www.mircode.org/mircode/, accessed on 25 June 2022) and starbase were used to investigate the interaction between demRNA, demiRNA and IncRNA and to predict the direct relationship between them. The mircode database makes it possible to check the interactions and binding sites of miRNAs among the complete GENCODE annotated transcriptome. Through it, the binding of Mir IncRNAs can also be checked [8]. The interactions between Mir and mRNA were explored using the Mirbase, TargetScan and Starbase databases. Starbase explores Mir targets with data obtained from CLIP-Seq and Degradome-Seq[9]. The possible targets for miR were predicted based on the complementary relationship and connection between the miR seed region and the 3UTR region of mRNA.

**The ceRNA regulatory network**

The interactive relations between the differentially expressed miRNAs, mRNAs IncRNAs were derived from pre-named databases. Then the predicted miRNA-IncRNA interactions and miRNA-mRNA interactions were visualized as a novel ceRNA network related to autism using Cytoscape software.

**Enrichment analysis**

To classify the genes based on their function either in molecular function(MF), biological processes(BP) and cellular components(CC) the Gene Ontology(GO) analysis and to map genes in pathways Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were made by the use of R.
Results

Differential expression analysis of lncRNAs, mRNAs, and miRNAs from public SRA data

Differential expression analysis led to 499 DEGs, 150 DEMs, and 550 DELs, from which 225 mRNA, 75 miRNA, 260 lncRNA were upregulated and 274 mRNA, 75 miRNA, 290 lncRNA were downregulated (table 1 & 2). All were evaluated with the criteria of logfc < 0.5 or > 2 and pv < 0.05. The expression level of differentially expressed mRNA and lncRNAs is shown in heatmap and volcano plot (fig 1).

Prediction and construction of the lncRNAs-miRNAs-mRNAs network

A lncRNA–miRNA–mRNA regulatory network was constructed based on the links between miRNA and mRNA and also between miRNA and lncRNA. Possible interactions were predicted using starbase, micode and targetscan. A competitive endogenous RNA network is built to describe lncRNA functions in competing with miRNAs in binding to mirs and also sponging them in order to regulating target genes expression (fig 2).

Functional and pathway enrichment analysis

By using enrichers, GO analysis containing Cellular Components (CC), Biological Functions (BF) and Molecular Processes (MP) and pathway analysis was performed to evaluate DEGs functions. The KEGG database was also examined to indicate the involvement of DELs in different pathways. Accordingly, 14 MF, BP, CC and pathways were associated with DEGs (p < 0.05). Top GO analysis and KEGG pathways are shown in (fig 3).
Discussion

Autism spectrum disorder is a type of neurodevelopmental spectrum disorder that is characterized by abnormal communication and verbal behaviors\[14\]. The symptoms of this disorder usually appear before the age of 1 to 3 years and its cause is still unknown, but it is believed that genetic factors and the environment are involved in it. The social, economic, lifestyle, and education of parents do not affect this disorder\[15\]. Because of the heterogeneity among ASD patients, the diagnosis of ASD depends on the assessment of the patient's behavior. Therefore, identification specific molecular markers may help to understand the pathophysiological mechanism of ASD, thereby increasing the chances of correct diagnosis and improving therapeutic interventions \[16\]. Many studies have shown that females with ASD show different characteristics than males with ASD. In addition, clinical studies have also reported that according to the different characteristics of ASD and camouflage, females with ASD are diagnosed much later than males with ASD\[17\]. Many genes are affected by risk mutations in ASD encoded proteins for synaptic, transcriptional and Chromatin remodeling pathways, which involve voltage ion channels regulating the release of action potentials, Acceleration and excitability-transcription coupling as well as histone modifying and chromatin remodeling enzymes\[18\]. In recent studies, it has been found that the molecular pathways of the inflammatory response\[19\] and PI3K\[20\], MAP Kinase\[21\], and TGFB play significant roles in the development of autism\[3\]. in the process of brain development, PI3K involves in different cellular functions, including cell migration, propagation, and axonal guidance\[22\]. Inactivation of PTEN (an upstream negative regulator of PI3K-AKT/mTOR pathway) causes upregulation of PI3K-AKT/mTOR pathway that leading to axonal dysregulation, megalencephaly, alteration of neuron size, protein synthesis, cerebral cell proliferation, and neuronal circuit connectivity in different regions of the brain\[20\]. transforming growth factor beta (TGFB) consists of three closely related isoforms TGFB1-TGFB2-TGFB3. These three isoforms are expressed in several central nervous system CNS cell types including neurons, astrocytes, and microglia\[3\]. In the past, it has been observed that the expression level of TGFB1 in the plasma of autistic children is significantly reduced\[23\]. The MAPK pathway regulates neural progenitor biogenesis, learning and memory, and mRNA translation by phosphorylation of TSC2 and eIF4E through MAPK-
interacting kinase 1 and 2 (MNK1 and MNK2). Besides, The deletion of MAPK3 (mitogen-activated protein kinase 3) gene that is located in 16p11.2 locus is related to ASD patients[24]. Moreover, ERK-MAPK signaling pathway in neurons has a pivotal function in regulating of synaptic plasticity[25] the MAPK pathway in patients with ASD syndromic is upregulated[26]. In addition, it has been found that LncRNAs and MicroRNAs, which are ncRNAs[27]. play a significant role in physiological processes[28]. Cellular and also in causing or preventing various diseases including cancers, diabetes, cardiovascular diseases, and nerve diseases[29]. For example, IncRNAGAS5 causes autism by affecting mir-221 on the PI3K signaling pathway[30]. Also in the early stage of brain development, Lin28A mediates increased protein synthesis that is crucial for neural progenitor cell, proliferation and maintenance. Furthermore, exogenous LIN28A promoted the proliferation of neural progenitor cells in vitro and induced early survival effect on primary cortical neurons, possibly via Upregulation of IGF2[31]. Our bioinformatics analysis revealed a number of lncRNAs, microRNAs, and mRNAs involved in pathways such as inflammation, WNT, TGFB, and PI3K, which are significant pathways in this disease. Such as the genes LIN28A[32], CLDN6[33], COL6A3[34], FBLN5[34], and LncRNAs HOXA-A3, DANCR, LINC00461, LINC00261, and microRNAs miR-103, miR-494, miR-520, miR-141, miR-214, miR-33.

Conclusion

The construction of ceRNA network introduces Several genes that are regulated by various factors through the influence they have on each other via multiple signaling pathways. As a final conclusion, it can be suggested that the identified genes in this study may serve as novel biomarkers for the diagnosis of autism and may provide insights into how the interaction of environmental factors with the genetic of an individual lead to the development of autism. This could potentially help to create more accurate diagnostic tests that can be used to identify autism in its early stages. Furthermore, this research may open the door to new treatment and prevention strategies for autism, enabling individuals to better manage and live with this condition.

Declarations
Ethics approval and consent to participate

This study was approved by the Ethical/Scientific Committee of Tarbiat Modares University (IR.MODARES.REC.1400.122).

Competing interests

The authors declare that they have no competing interests.

Funding

There was no funding for this study.

Author contributions

ASTA, MM, YZ & AN performed the experiments. ARJ designed the experiments and supervised the study. AS, SRK, PA & PB helped with lab works.

Acknowledgments

Authors are thankful for the help and advice of the 4402-laboratory members at the Department of Genetics, TMU. Tehran-Iran

Reference

3. Khulood, A.K. and N.I. Wijdan, GENE POLYMORPHISM OF TRANSFORMING GROWTH FACTOR Beta 1 (TGF-β1) IN AUTISM SPECTRUM DISORDER ASD IN BASRAH.


**Legends**

**Fig1**) PCA plot and heat map diagram for dysregulated genes. A) The PCA diagram shows that the autism samples are completely normal. B) The Heatmap diagram shows that the autism samples are completely normal. C) The volcano plot shows the distribution of dysregulated mRNAs. D) The PCA diagram shows that the autism samples are completely normal. E) The Heatmap diagram shows that the autism samples are completely normal. F) The volcano plot shows the distribution of dysregulated LncRNAs.


**Fig3**) Predicted cellular and molecular pathways involved in autism. A&B) KEGG analysis predicted pathways involved in autism.
Figure 1
Figure 2
<table>
<thead>
<tr>
<th>Index</th>
<th>Name</th>
<th>P-value</th>
<th>Adjusted P-value</th>
<th>Odds Ratio</th>
<th>Combined score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>elastic fiber (GO:0071953)</td>
<td>0.002248</td>
<td>0.02473</td>
<td>624.59</td>
<td>3808.55</td>
</tr>
<tr>
<td>2</td>
<td>supramolecular fiber (GO:0099512)</td>
<td>0.008519</td>
<td>0.06247</td>
<td>138.70</td>
<td>660.97</td>
</tr>
<tr>
<td>3</td>
<td>rough endoplasmic reticulum (GO:0005791)</td>
<td>0.01298</td>
<td>0.07137</td>
<td>89.12</td>
<td>387.19</td>
</tr>
<tr>
<td>4</td>
<td>collagen-containing extracellular matrix (GO:0062023)</td>
<td>0.0005250</td>
<td>0.01155</td>
<td>26.01</td>
<td>196.46</td>
</tr>
<tr>
<td>5</td>
<td>basement membrane (GO:0005604)</td>
<td>0.02316</td>
<td>0.07523</td>
<td>48.87</td>
<td>184.02</td>
</tr>
<tr>
<td>6</td>
<td>extracellular membrane-bounded organelle (GO:0065010)</td>
<td>0.02492</td>
<td>0.07523</td>
<td>45.31</td>
<td>167.28</td>
</tr>
<tr>
<td>7</td>
<td>extracellular vesicle (GO:1903561)</td>
<td>0.02624</td>
<td>0.07523</td>
<td>42.96</td>
<td>156.38</td>
</tr>
<tr>
<td>8</td>
<td>cytoplasmic stress granule (GO:0010494)</td>
<td>0.02888</td>
<td>0.07523</td>
<td>38.92</td>
<td>137.96</td>
</tr>
<tr>
<td>9</td>
<td>bicellular tight junction (GO:0005923)</td>
<td>0.03456</td>
<td>0.07523</td>
<td>32.33</td>
<td>108.78</td>
</tr>
<tr>
<td>10</td>
<td>P-body (GO:0000932)</td>
<td>0.03544</td>
<td>0.07523</td>
<td>31.51</td>
<td>105.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Index</th>
<th>Name</th>
<th>P-value</th>
<th>Adjusted P-value</th>
<th>Odds Ratio</th>
<th>Combined score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ECM-receptor interaction</td>
<td>0.0006754</td>
<td>0.01823</td>
<td>66.13</td>
<td>482.76</td>
</tr>
<tr>
<td>2</td>
<td>African trypanosomiasis</td>
<td>0.01653</td>
<td>0.08100</td>
<td>69.29</td>
<td>284.26</td>
</tr>
<tr>
<td>3</td>
<td>Type II diabetes mellitus</td>
<td>0.02051</td>
<td>0.08100</td>
<td>55.41</td>
<td>215.34</td>
</tr>
<tr>
<td>4</td>
<td>Glycosaminoglycan biosynthesis</td>
<td>0.02360</td>
<td>0.08100</td>
<td>47.93</td>
<td>179.56</td>
</tr>
<tr>
<td>5</td>
<td>Regulation of lipolysis in adipocytes</td>
<td>0.02448</td>
<td>0.08100</td>
<td>46.15</td>
<td>171.21</td>
</tr>
<tr>
<td>6</td>
<td>Focal adhesion</td>
<td>0.003454</td>
<td>0.04663</td>
<td>28.42</td>
<td>161.07</td>
</tr>
<tr>
<td>7</td>
<td>Adipocytokine signaling pathway</td>
<td>0.03063</td>
<td>0.08100</td>
<td>36.62</td>
<td>127.66</td>
</tr>
<tr>
<td>8</td>
<td>Small cell lung cancer</td>
<td>0.04065</td>
<td>0.08100</td>
<td>27.34</td>
<td>87.55</td>
</tr>
<tr>
<td>9</td>
<td>Human papillomavirus infection</td>
<td>0.009104</td>
<td>0.06992</td>
<td>17.08</td>
<td>80.24</td>
</tr>
<tr>
<td>10</td>
<td>Amoebiasis</td>
<td>0.04498</td>
<td>0.08100</td>
<td>24.62</td>
<td>76.35</td>
</tr>
</tbody>
</table>