AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Asian Journal of Medicine and Biomedicine

An integrative bioinformatics meta-analysis of microarray data for identifying hub genes as biomarkers of autism spectrum disorder (ASD)

Nor Azian Abdul Murad[1*], Wan Fahmi Wan Mohamad Nazarie[2], Zam Zureena Mohd Rani[1], Muhammad Redha Abdullah Zawawi[1], Norazlin Kamal Nor[3], Fara Zela Mohd Radin[1], Rahman Jamal[1], Juriza Ismail[3,4]

[1] UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia (UKM), Jalan Yaacob Latiff, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia
[2] Faculty of Science and Natural Resources, Universiti Malaysia Sabah (UMS), Kota Kinabalu, Sabah, Malaysia
[3]Department of Pediatrics, Faculty of Medicine, Universiti Kebangsaan Malaysia (UKM), Jalan Yaacob Latiff, 56000 Cheras, Kuala Lumpur, Malaysia
[4]Child Health Department, Sultan Qaboos University Hospital, Muscat, Oman

*Corresponding author: nor_azian@ppukm.ukm.edu.my

Received: 25 August 2022 **Accepted: 09 November 2022** **Published: 30 November 2022**

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder primarily affecting young children. ASD is a complex disease involving genetic and environmental factors. Environmental risk factors identified include gestational exposure to pollution, pesticides, maternal infections, and inflammation. Genetic mutations account for about 10 – 20% of ASD cases. Based on the Centre for Disease Control (CDC) in the United States, 1 in 68 children are affected with ASD. Recent advancements in genetic technologies have enabled the detection of biomarkers for the early detection of diseases and risk identification. *Aim*: This meta-analysis aims to determine the gene signatures involved in ASD. We conducted a meta-analysis to identify the differentially expressed genes (DEGs) in ASD microarray datasets comprising 122 ASD and 89 control peripheral blood mononuclear cell (PBMC) and whole blood samples from two microarray studies. We performed gene ontology, pathway enrichment, and protein-protein interaction (PPI) network analysis to identify associations between autism and altered gene expression levels. At a false discovery rate (FDR) < 0.05, we identified 1862 DEGs; 1056 genes were upregulated, and 806 genes were downregulated. DEGs revealed that dysregulated genes were significantly enriched in the "Primary immunodeficiency pathway", "Influenzae A", "Epstein-Barr virus infection pathway", and other signalling pathways from the analyses using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. Subsequently, protein-protein interactions (PPI) analysis identified *SUMO1*, *SP1*, *EGR1*, *EP300*, and *VHL* as hub genes to be the potential biomarkers for ASD. In total, eighteen differentially expressed hub genes could potentially be used as potential biomarkers for the diagnosis of ASD.

Keywords
Autism spectrum disorder, microarray, peripheral blood mononuclear cell, public repositories, whole blood.

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder related to communication and behaviour deficits. ASD is diagnosed using the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), and the clinical characteristics include lack of communication and interaction, repetitive behaviours, and limited interests [1]. It is a broad-spectrum disease, and each ASD child presents with different behavioural characteristics. Based on the Centre for Disease Control in the United State, the prevalence of ASD is estimated about 1 in 68 children [2]. Based on this figure, there are about 9,000 children born with ASD yearly in Malaysia [2,3]. For Malaysia, based on ASD screening, the most recent data from 2015 infer that 1 in 635 children have ASD [4].

The cause of ASD is still unknown. However, environmental, biological, and genetic factors play essential roles in the development of ASD [5]. Environmental risk factors include advanced parental age, prematurity, encephalopathy, low birth weight and birth complication trauma [6]. Genetic factors contribute to about 49% of all ASD cases [7]. The advancements in next-generation sequencing technology via whole-genome sequencing (WGS) and whole-exome sequencing (WES) have facilitated the identification of ASD-risk genes. To date, 600 ASD-risk genes have been identified, including *ADNP, ANK2, ARID1B, CHD8, GRIN2B* and *PTEN* [8].The synaptic formation, transcriptional modelling and chromatin remodelling pathways are the common pathways associated with ASD [9]. Integrated analysis of the brain transcriptome revealed that abnormal synaptic functions and neurological development contribute to the development of ASD. In addition, immune system dysfunction is also associated with ASD [10]. Inflammatory activity is increased in ASD children via pro-inflammatory biomarkers [11]. In a landmark paper on ASD and inflammation, 97 ASD children recruited from the Childhood Autism Risks from Genetics and Environment (CHARGE) study showed increased pro-inflammatory cytokines such as IL-1β, IL-6, IL-8 and IL-12p40 [12]. High cytokine levels linked to stereotypical behaviours suggested that immune system dysfunction affects ASD core behaviours [12].

Maternal infections and fever during pregnancy increase the risk for ASD [13]. However, the association between influenza and ASD is still indefinite. A study by Zerbo and colleagues showed no association between maternal influenza infection during pregnancy and ASD risk. The risk of ASD in their children would be higher if the mother received an influenza vaccination during their first trimester. However, after multiple corrections, the association was not statistically significant [14]. Self-reported maternal influenza infection is associated with an increased risk of infantile autism adjusted HR: 2.3 [95% CI: 1.0–5.3]), even higher during prolonged fever [15]. Several viral infections have been associated with ASD, including Epstein-Barr, influenza, and the Borna disease virus [16]. However, the role of viral infection in ASD is still poorly understood. Valayi and colleagues investigated the levels of anti-Cytomegalovirus (CMV) and anti-Epstein-Barr virus (EBV) antibodies in 45 ASD children and healthy controls [17]. Anti-CMV IgG and IgM antibodies in the blood of ASD patients but not statistically significant ($P< 0.05$). In ASD patients, anti-EBV IgM antibodies were increased significantly ($P< 0.05$), but the serum IgG level against EBV was not significant. Thus, it has been postulated that EBV infection may be associated with an increased risk of ASD.

In this study, we performed an integrative bioinformatics meta-analysis of microarray data from the GEO DataSets (https://www.ncbi.nlm.nih.gov/gds) to identify the hub genes involved in ASD.

Materials and Methods

Workflow of analysis strategy

We performed an integrative bioinformatics analysis of microarray data to identify the hub genes in ASD as potential diagnostic biomarkers from the GEO DataSets (Fig. 1). Two microarray data sets were used, including GSE111176 [18] and GSE18123 [19] and the data were analysed accordingly. In total, 211 samples were included in this meta-analysis, including 122 confirmed ASD patients and 89 healthy controls.

eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb

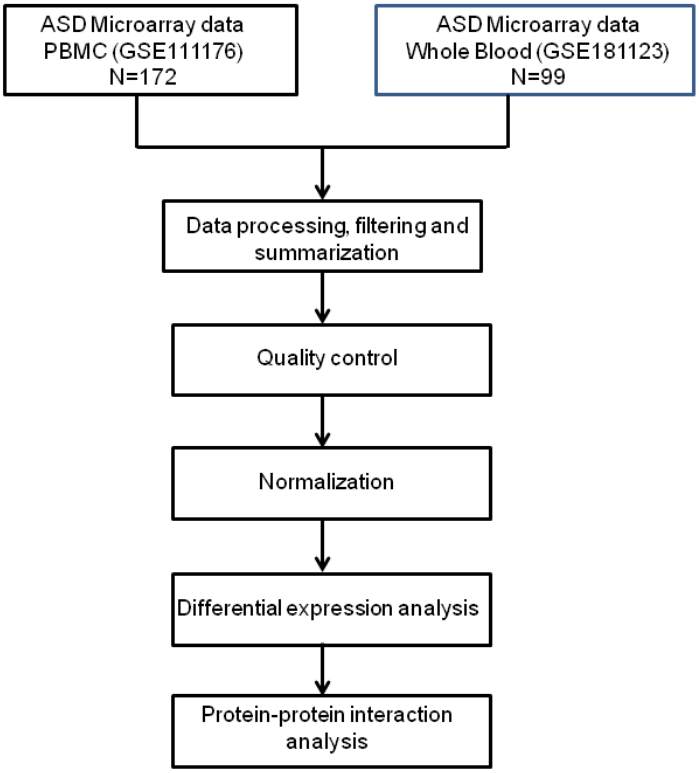

Figure 1: Workflow of analysis strategy for analysis of microarray data from the GEO DataSets to determine the differentially expressed genes and the pathways involved in ASD.

Data acquisition and data clean-up

Two microarray studies depicted in Table 1 were used to determine the differentially expressed genes from the blood of ASD patients. ASD microarray datasets were searched based on a study by Voinsky and colleagues[20]. Only studies using peripheral blood mononuclear cells (PBMC) and whole blood were included. Subsequently, the list of database searches of publications related to selected datasets was consulted to confirm the clinical diagnosis. A definitive diagnosis was ascertained before blood collection for microarray expression analysis. All searches were performed on public data repositories from the GEO DataSets[21]. The GEO accession number, sample type, platform, and matrix expression data were downloaded for each study. Each sample was assigned an identification condition and class (ASD and control) in Microsoft Excel format and then saved as a tab-delimited ".txt" file. Subsequent analysis was conducted using the web-application NetworkAnalyst version 3.0[22].

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Table 1: Characteristics of individual studies included in the ASD meta-analysis.

No	GEO Accession No.	Samples (ASD/control)	Organism	Sample source	Platform	References
1.	GSE111176	n=172, 91/56	Human	PBMC	Illumina HumanHT-12 V4.0 expression beadchip	Gazestani et al., 2019 (23)
2.	GSE18123	n=99, 31/33	Human	Whole blood	Affymetrix Human Genome U133 Plus 2.0 Array	Kong et al., 2012 (24)

Data processing, filtering, and summarization

"Multiple Gene Expression Tables" were selected and the ".txt" file from the previous step was uploaded separately for each study. A series of quality control, including ID conversion, metadata check, and visualization, was performed. Subsequently, each dataset was independently normalized using variant stabilization normalization (VSN) followed by quantile normalization before differential expression analysis. The meta-datasets were visually inspected using principal component analysis (PCA) plots to identify the outliers. The individual analysis of each dataset was carried out using the Benjamini–Hochberg's False Discovery Rate (FDR) with cut-off p-values less than 0.05. Each dataset has different probe identifiers (IDs) representing different transcripts and genes on the array chip due to different platforms used by different studies, Illumina and Affymetrix. The Entrez ID will replace the identifier based on the platform used in each study. Finally, the meta-datasets batch effects were adjusted using ComBat [23], and the batch covariate is known. A heatmap of differential expressed genes (DEGs) was generated using the network visual inspection tools of NetworkAnalyst and clustered using the single linkage method.

Differential expression and statistical analysis

Following the normalization process, pre-processing and data integrity checks of the individual datasets were conducted. The limma package from the R package in the NetworkAnalyst was used to determine the differentially expressed genes via meta-analysis approaches [24]. The estimated significant difference of expression and differential expression meta-analysis of ASD and control samples was performed. For each study, only one contrast was used: PBMC versus control and whole blood versus control. P-value from different datasets was combined to increase the statistical power, and cross-validation with Fisher's method was applied to determine the estimated gene expression fold changes (FC) and statistical significance across two different studies. This method produced the most consistent biological results. DEGs with adjusted p-value less than 0.05 and fold change of more than 1.5 were obtained from the meta-analysis. The high confidence DEGs were used for downstream analysis, i.e., gene ontology, pathway enrichment analysis, and hub gene network analysis.

Gene ontology and pathway enrichment analysis

Gene ontology and pathway enrichment analysis were analysed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [25] based pathway enrichment identification module in the NetworkAnalyst 3.0 using the differentially dysregulated genes in ASD versus control.

Protein-protein interaction (PPI) network analysis for hub gene identification

The hub genes were identified using PPI network analysis based on the gene's interaction in the biological network. Following identifying DEGs, the hub gene network was generated using the PPI network option in Network Analyst 3.0 using the InnateDB interactive database [26]. A zero-order PPI network was created using the original seed of 894 dysregulated genes. A network of 894 nodes represented the proteins, while 2325 edges represented interaction between the proteins.

eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb

Results

Identification of differentially expressed genes from ASD related microarray dataset

Principal component analysis (PCA) of before and after normalization of these two datasets was shown in **Figure 2a.** Our analysis identified 1,862 DEGs, including 1,056 up-regulated and 806 down-regulated genes with a significance threshold of p-value less than 0.05. The most significantly dysregulated genes (n=29) are shown in **Supplementary Table 1**. However, only four genes overlap between PBMC and the whole blood expression dataset, including *AP5S1*, *RBBP6*, *ANKRD28*, and *ZNF638* (**Figure 2a**). **Figure 2b** shows the heatmap of all the DEGs across two microarray datasets. *NANOS2* (combined Tstat = 36.777, combined p-value = 0.00041468), *IGF2BP3* (combined Tstat = 36.47, combined p-value = 0.00041468) and *MRRF* (combined Tstat = 34.274, combined p-value = 0.00088271) were the most significantly up-regulated genes. While *NR3C2* (combined Tstat = −42.933, combined p-value=8.49E-05), *ANKRD28* (combined Tstat = −42.288, combined p-value= 8.49E-05) and *ZNF609* (combined Tstat = −43.759, combined p-value=8.49E-05) were the most down-regulated genes across the two microarray datasets.

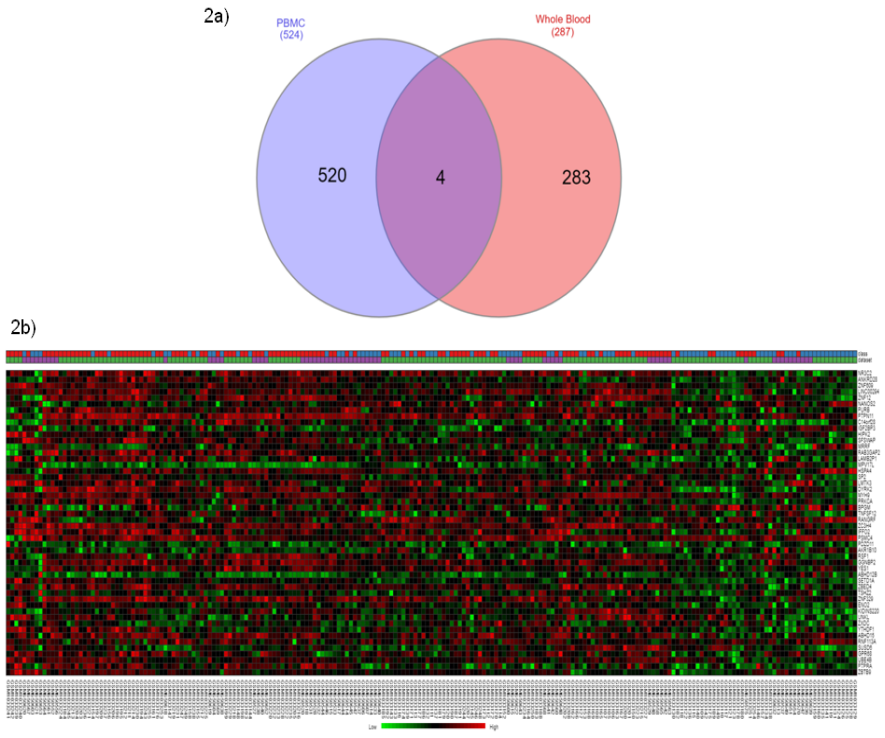

Figure 2: a) Venn diagram of DEGs. The combined meta-analysis of both datasets shows many dysregulated genes that were significantly different (a p-value of < 0.05 were considered significant). b) Heatmap of most significantly differentially expressed genes. Heatmap showing the expression of the 50 most significantly differentially expressed genes (DEGs) (p-value < 0.05) from 1,862 significant DEGs identified through the meta-analysis. In total, 205 genes were up-regulated, and 136 genes were down-regulated in ASD compared to control.

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Gene ontology and pathway enrichment analysis of DEGs in ASD

Based on the Kyoto Encyclopedia of Genes (KEGG) pathways, the up-regulated DEGs in the meta-analysis of ASD datasets were enriched in "Primary immunodeficiency pathway", "Influenzae A", and "Epstein-Barr virus infection pathway" with $p < 0.05$. While the downregulated DEGs were enriched in the "Sphingolipid signalling pathway", "Phospholipase D signalling pathway", and "Spliceosome"**.**

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

A protein-protein interaction network analysis was conducted to identify the important hub genes to extract more biologically relevant information. A degree above 30 was set as the cut off criterion, and we identified 18 hub genes **(Table 2)** and the top five genes were *SUMO1*, *SP1*, *EGR1*, *EP300*, and *VHL.* In the NetworkAnalyst, the "Zero-order" network interaction generated 894 nodes and 2,325 edges where the node represents protein while edges are the interaction between proteins (**Figure 4**). Small Ubiquitin-like Modifier 1 (*SUMO1*) with betweenness centrality = 81255.76; degree = 124; expression=20.821, Sp1 Transcription Factor (*SP1*) with betweenness centrality = 41381.92; degree = 76; expression = -14.729, E1A Binding Protein P300 (*EP300*) with betweenness centrality = 37982.03; degree = 71; expression = -21.558, Early Growth Response 1 (*EGR1*) with betweenness centrality = 40490.8; degree = 67, and Von Hippel-Lindau Tumor Suppressor (*VHL*) with betweenness centrality = 30043.04; degree = 55; expression= -16.519. A full list of hub genes based on network topology scores was shown in **Supplementary Table 1**.

Table 2: The list of 18 hub genes related to ASD.

No	Id	Label	Degree	Betweenness	Expression
1	P63165	*SUMO1*	124	81255.76	20.821
2	P08047	*SP1*	76	41381.92	-14.729
3	Q09472	*EP300*	71	37982.03	-21.558
4	P18146	*EGR1*	67	40490.8	16.638
5	P40337	*VHL*	55	30043.04	-16.519
6	Q92793	*CREBBP*	47	17243.04	-17.537
7	P35222	*CTNNB1*	44	23373.1	19.881
8	P62158	*CALM3*	43	22681.42	-22.686
9	P31749	*AKT1*	38	22188.12	-14.821
10	O15379	*HDAC3*	37	12010.95	-18.694
11	P05161	*ISG15*	35	17012.03	14.907
12	P34932	*HSPA4*	34	13777.56	32.667
13	P22681	*CBL*	33	15814.78	-16.919
14	Q12906	*ILF3*	32	12401.2	-14.901
15	P16104	*H2AFX*	32	10565.3	14.969
16	P27986	*PIK3R1*	31	13934.52	-19.518
17	O60216	*RAD21*	30	9983.29	-17.269
18	P24928	*POLR2A*	30	8772.93	-18.578

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Figure 4: Network analysis of the highly dysregulated genes. The differentially expressed genes (DEGs) (ASD versus controls) were integrated into the Network Analyst web application to visualize the protein-protein interactions network. A 'zero-order' interaction network with 894 nodes and 2325 edges was created. The most highly ranked nodes across the dataset based on network topology measures were SUMO1 (betweenness centrality = 81255.76; degree = 124), and SP1 (betweenness centrality = 41381.92; degree = 76).

Discussion

Based on the gene expression profiles of the GSE60438 dataset, a total of 1862 dysregulated genes were identified, including 1056 up- and 806 down-regulated DEGs. GO, and KEGG pathway enrichment analyses of DEGs suggested that they were significantly enriched in disease-related ASD. PPI network generation and identification of hub genes were performed to determine 18 significant hub genes with a degree value > 30. Our results showed that these 18 differentially expressed hub genes could be used as potential biomarkers for the diagnosis of ASD. From the results, we could classify the genes into two categories. First, associated genes and functions are directly related to ASD pathways, for example, *RBBP6*, *SUMO1* and *EP300*. Second, genes that will indirectly impact the ASD pathways, such as *NANOS2*, *MRRF*, and *SP1*.

Small Ubiquitin Like Modifier 1 *(SUMO1)* is associated with cellular processes such as nuclear transport, transcriptional regulation, apoptosis, and protein stability [27]. *SUMO1* is involved in signalling pathways: androgen receptor, interferon-gamma, RNA binding and ubiquitin-protein ligase binding [28]. Prasad and colleagues used a high-resolution CGH microarray to identify CNV in 696 unrelated ASD cases. They have identified a novel recurrent 24.7 kb duplication in 3/696 ASD cases and 1 in 5139 control [29]. In addition, a 50.8 kb duplication was also observed in other ASD studies [30].

Based on the SFARI Gene score system, which linked all evidence towards ASD risk and gene category, *SUMO1* scored 2, suggesting it is a strong candidate for ASD [31]. *SUMO1* interacts with an androgen receptor (AR), which could suppress the RORA expression [32]. *SUMO1* also interact with the *ARX* gene, which is involved in autistic disorders [33]. Numerous genomic studies have revealed the CBP and its paralog EP300 as the critical hubs in ASD-associated protein [34]. E1A-binding protein p300 (EP300) G21S mutation in exon two was identified in a patient with ASD [35]. This gene encodes for the adenovirus E1A-associated cellular p300 transcriptional co-activator protein. It acts as a histone acetyltransferase which regulates transcription via chromatin remodelling. EP300 plays a significant role in cellular proliferation and differentiation. Fei Zheng and colleagues performed an animal study, showing that CBP and CH1 domains are essential for ASD [36]. Mice

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

with CBP CH1 (TAZ1) deletion (*CBPΔCH1/ΔCH1*) appear to have an RTS-like phenotype. These include ASD-relevant repetitive behaviours, hyperactivity, social interaction deficits, motor dysfunction, impaired memory recognition, and abnormal synaptic plasticity. These results suggested that CBP is crucial in maintaining normal motor function, cognition, and social behaviour.

Ankyrin Repeat Domain 28 *(ANKRD28)* is a putative regulatory subunit of protein phosphatase 6 (PP6), which is involved in phosphoprotein substrates recognition and histone acetylation [37]. *ANKRD28* is not associated with ASD. However, the other members of *ANKRD*, namely *ANKRD 11* and *ANKRD17*, have been associated with ASD. Christian Marshall and colleagues performed a genome-wide assessment for structural abnormalities in 427 unrelated ASD cases and identified a novel locus at *ANKRD11* [38]. In an animal study, Danis and colleagues revealed that *ANKRD11* knockdown in developing mice or human cortical neural precursors leads to reduced cellular proliferation, neurogenesis, and abnormal neuronal positioning [39]. In addition, Yoda mice with *ANKRD11* point mutation showed similar phenotypes and ASD-like behaviours. This study shows that *ANKRD11* was associated with chromatin and HDAC3 colocalization. The expression of the *ANKRD11* target gene and histone acetylation was altered in Yoda neural precursors. Inhibition of the histone acetyltransferase activity or HDAC expression 3 restored the *ANKRD11* knockdown-mediated decrease in precursor proliferation. This study proved that *ANRKD11* is crucial in chromatin regulation that controls histone acetylation and gene expression during neuronal development linked to ASD [39].

The nuclear receptor subfamily 3 group C member 2 *(NR3C2)* encodes for the mineralocorticoid receptor [40]. NR3C2 protein is vital for activating their target genes by binding to the mineralocorticoid response elements [40]. It acts in the hypothalamic-pituitary-adrenal axis and is associated with stress and anxiety, which are the common features of autistic individuals. *NR3C2* mutation has been identified in ASD whole-exome sequencing study and (transmission and de novo association (TADA) analysis as a gene strongly enriched for variants likely to affect ASD risk [9]. Holly N Cukier and colleagues identified a stop gain mutation (p.Q919X) in the HPA-Axis Gene NR3C2 in three brothers with ASD [41]. Ruzzo and colleagues performed genome sequencing on 2,308 ASD individuals from families with multiple ASD children [42]. They identified a structural variant (SV) affecting non-coding regions, implicating recurrent deletions in the promoters of NR3C2. In zebrafish, loss of *NR3C2* function disrupts sleep and social function, corresponding with human ASD-related phenotypes [42].

PTPN11 is encoded by the protein tyrosine phosphatase (PTP) family [43]. The phospho-tyrosine binding domains mediate interaction with its substrates. *PTP* is widely expressed in most tissues and is crucial in cell signalling events for cellular processes [43]. Cell growth and migration, transcription regulation, differentiation, mitogenic activation, and oncogenic transformation are common cellular processes related to *PTPN11* [44]. Hani and colleagues determined the role of long non-coding RNA (lncRNA)-associated competing endogenous RNAs (ceRNAs) in the peripheral blood (PB) of ASD samples to understand the molecular regulatory processes in ASD using bioinformatics tools [45]. They identified four potential DElncRNA-miRNA-DEmRNA axes in ASD pathogenesis, including, *LINC00472/hsa-miR-221-3p/PTPN11*, *ANP32A-IT1/hsa-miR-182-5p/S100A2* , *LINC00472/hsa-miR-132-3p/S100A2*, and *RBM26-AS1/hsa-miR-182-5p/S100A2*. "JAK-STAT signalling pathway" and "Adipocytokine are among the enriched signalling pathway related to the immune-related DEmRNAs. In another study, the author sequenced 208 candidate genes from >11,730 patients and >2,867 controls. They showed ten (13%) genes, including *TANC2*, *TRIO*, *COL4A3BP*, *TBL1XR1*, *PPP2R5D*, *DLGAP1*, *SRGAP3*, *PTPN11*, *ADCY5* and *ITPR1*, which are unique to the MIS30 category for probands. Their results highlighted some interesting trends. First, there is twofold enrichment of missense variants with high CADD scores (>30) in probands, which includes autism risk genes (*PTPN11, CACNA1G, TRIP12*, and *PTK7*) and genes of interest (*SUPT16H* and *SCN3A*). Second, severe de novo missense mutations are important in ASD risk prediction [46] and autism [47]. Turner and colleagues investigated the importance of *de novo* mutations (DNMs) in 516 autism families with 2064 ASD individuals. Probands with ASD have significant changes in CNVs and SNVs. There was twofold enrichment of missense variants in *PTPN11, CACNA1G, TRIP12*, *PTK7, SUPT16H* and *SCN3A*, indicating the importance of severe *de novo* missense mutations in ASD [48].

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

MYH9 gene encodes a conventional non-muscle myosin [49]. The encoded protein is crucial in several processes, including cytokinesis, cell motility and maintenance of cell shape [50]. Non-syndromic sensorineural deafness, autosomal dominant type 17, Epstein syndrome, and Alport syndrome with macrothrombocytopenia are some diseases associated with this gene [51]. Marchani and colleagues, analyzed 47 members of a multigeneration pedigree with 11 cases of ASD using three different platforms such as multiallelic linkage marker panel, dense diallelic marker panel and exome sequencing [52]. *MYH9* variant was observed in heterozygous state in four carriers with risk haplotype but not presented in one affected subject without the risk haplotype. It is a rare variant since it was reported only once in other databases. However, de novo missense variants have been identified in ASD probands from several studies, including the Simons Simplex Collection [46], the SPARK cohort [53], and the Deciphering Developmental Disorders 2017 [54].

PRKCA is involved in many molecular functions by phosphorylating targets such as *RAF1, BCL2, CSPG4, TNNT2/CTNT*, or by activating signalling cascade *MAPK1/3 (ERK1/2)* and *RAP1GAP. PRKCA* is involved in the positive and negative regulation of cell proliferation, apoptosis, differentiation, migration and adhesion, tumorigenesis, cardiac hypertrophy, angiogenesis, platelet function and inflammation [54]. *De novo* missense variants have been identified in three ASD probands [9, 46] and two probands with unspecified developmental disorders [55]. PRKCA missense mutation was identified in an integrated meta-analysis of *de novo* mutation data from a combined dataset of 10,927 individuals with neurodevelopmental disorders [55]. Turner and colleagues used several methods to identify candidate mutations in eight families. ASD patients showed a significant ($p = 0.03$) enrichment of de novo mutations within fetal CNS DNase I hypersensitive sites. This effect was only observed within 50 kb of ASD risk genes, including genes where dosage sensitivity has been recognized by recurrent disruptive *de novo* protein-coding mutations *(ARID1B, SCN2A, NR3C2, PRKCA*, and *DSCAM*) [56].

Conclusion

Thismeta-analysis identified eighteen differentially expressed hub genes that could potentially be used as potential biomarkers for the diagnosis of ASD. These genes may be directly or indirectly associated with ASD pathways. However, further confirmation studies such as validation in larger sample sizes *and in-vitro* and *in-vivo* studies are needed to verify these findings.

Acknowledgement

The authors would like to thanks Gazestani et al., 2019 (23) and Kong et al., 2012 (24).

References:

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th edition). 2013.
2. Press Release. Centers for Disease Control and Prevention (CDC). March 27, 2014.
3. The National Autism Society of Malaysia (NASOM). Online access. https://www.nasom.org.my/.
4. Lim, JM. Living with Autism in Malaysia. 2015.
5. Almandil NB, Alkuroud DN, Abdul Azeez S, AlSulaiman A, Elaissari A. Environmental and Genetic Factors in Autism Spectrum Disorders: Special Emphasis on Data from Arabian Studies. Int J Environ Res Public Health. 2019; 16(4): 658-673.
6. Modabbernia A, Velthorst E, and Reichenberg A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. Mol Autism. 2017; 8:13-28.

eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb

7. Rylaarsdam L and Guemez-Gamboa A. Genetic Causes and Modifiers of Autism Spectrum Disorder. Front Cell Neurosci. 2019; 13: 385-399.
8. Xiong J, Chen S, Pang N, Deng X, Yang L, et al. Neurological Diseases with Autism Spectrum Disorder: Role of ASD Risk Genes. Front Neurosci, 2019. 13: 349-356.
9. De Rubeis S, He X, Goldberg AP, Poultney CS, and Samocha A. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515(7526): 209-215.
10. Marchezan J, Winkler Dos Santos EGA, Deckmann I, and Riesgo RDS. Immunological Dysfunction in Autism Spectrum Disorder: A Potential Target for Therapy. Neuroimmunomodulation. 2018: 25(5-6); 300-319.
11. Xu N, Li X, and Zhong Y. Inflammatory cytokines: potential biomarkers of immunologic dysfunction in autism spectrum disorders. Mediators Inflamm. 2015. 2015:531518:1-10.
12. Ashwood P, Krakowiak P, Hertz-Picciotto I, Hansen R, Pessah Isacc, et al. Elevated plasma cytokines in autism spectrum disorders provide evidence of immune dysfunction and are associated with impaired behavioral outcome. Brain Behav Immun. 2011; 25(1): 40-45.
13. Gardener H, Spiegelman D, and Buka SL. Perinatal and neonatal risk factors for autism: a comprehensive meta-analysis. Pediatrics. 2011; 128(2): 344-355.
14. Zerbo O, Qian Y, Yoshida C, Fireman B, Klein NP, et al. Association Between Influenza Infection and Vaccination During Pregnancy and Risk of Autism Spectrum Disorder. JAMA Pediatr. 2017; 171(1): p. e163609.
15. Atladóttir HO, Henriksen TB, Schendel DE, and Parner ET. Autism after infection, febrile episodes, and antibiotic use during pregnancy: an exploratory study. Pediatrics. 2012; 130 (6): e1447-54.
16. Libbey JE, Sweeten TL, McMahon WM, and Fujinami RS. Autistic disorder and viral infections. J Neurovirol. 2005; 11(1): 1-10.
17. Valayi, S, Eftekharian MM, Taheri M, and Alikhani MY. Evaluation of antibodies to cytomegalovirus and Epstein-Barr virus in patients with autism spectrum disorder. Hum Antibodies. 2017; 26(3): 165-169.
18. Gazestani VH, Pramparo T, Nalabolu S, Kellman BP, Murray S, et al. A perturbed gene network containing PI3K-AKT, RAS-ERK and WNT-beta-catenin pathways in leukocytes is linked to ASD genetics and symptom severity. Nat Neurosci. 2019; 22(10): 1624-1634.
19. Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, et al. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One. 2012; 7(12): e49475.
20. Voinsky I, Bennuri SC, Svigals J, Frye RE, Rose S, et al. Peripheral Blood Mononuclear Cell Oxytocin and Vasopressin Receptor Expression Positively Correlates with Social and Behavioral Function in Children with Autism. Scientific Reports. 2019; 9(1):13443-13452.
21. https://www.ncbi.nlm.nih.gov/gds. GEO DataSets, National Library of Medicine, National Center for Biotechnology Information. Online access.
22. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019; 47(W1): W234-W241.
23. Behdenna A, Haziza J, Azencott C-A, and Nordor A. pyComBat, a Python tool for batch effects correction in high throughput molecular data using empirical Bayes methods. bioRxiv preprint .2021; 2021: 1-6. doi: https://doi.org/10.1101/2020.03.17.995431.
24. Matthew ER, Phipson B, Wu D, Hu Y, Law CW, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015: 43 (7):e47.
25. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, et al. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. Nucleic Acids Res. 2013; 41(Database issue): D1228-33.
26. Kanehisa M, Furumichi M, Tanabe M, Sato Ym Morishima K, et al. KEGG: new perspectives on genomes, pathways, diseases, and drug. Nucleic Acids Res. 2017; 45(Database issue): D353-361.
27. Li P, Jing H, Wang Y, Yuan L, Xiao H, et al. SUMO modification in apoptosis. J Mol Histol. 2021; 52(1): 1-10.

eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb

28. Wilkinson KA, Nakamura Y, and Henley JM. Targets, and consequences of protein SUMOylation in neurons. Brain Res Rev. 2010; 64(1): 195-212.
29. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, et al. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. G3 (Bethesda). 2012; 2(12): 1665-85.
30. Lionel AC, Crosbie J, Barbosa N, Goodake T, Thiruvahindrapuram B, et al. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med. 2011; 3(95): 95ra75.
31. SUMO1. SFARI Gene. https://gene.sfari.org/search?search=SUMO1. Online access.
32. Sarachana T and Hu VW. Differential recruitment of coregulators to the RORA promoter adds another layer of complexity to gene (dys) regulation by sex hormones in autism. Molecular Autism. 2013; 4:39-54.
33. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011; 39(Database issue): D561-8.
34. Bedford DC, Kasper LH, Fukuyama T and Brindle PK. Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. Epigenetics. 2010; 5(1): 9-15.
35. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, et al. Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet. 2012; 90(1): 133-141.
36. Zheng F, Kasper LH, Bedford DC, Lerach S, Teubner BJW, et al. Mutation of the CH1 Domain in the Histone Acetyltransferase CREBBP Results in Autism-Relevant Behaviors in Mice. PLoS One. 2016; 11(1): e0146366.
37. Chopra M, McEntagart M, Smith JC, Platzer K, Shukla A, et al. Heterozygous ANKRD17 loss-of-function variants cause a syndrome with intellectual disability, speech delay, and dysmorphism. Am J Hum Genet. 2021; 108(6): 1138-1150.
38. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet. 2008; 82(2): 477-488.
39. Gallagher D, Voronova A, Zander MA, Cancino GI, Bramall A, et al. Ankrd11 is a chromatin regulator involved in autism that is essential for neural development. Dev Cell. 2015; 32(1): 31-42.
40. NR3C2. SFARI Gene. https://gene.sfari.org/search?search= NR3C2. Online access.
41. Cukier HN, Griswold AJ, Hofmann NK, Gomez L, Whitehead PL, et al. Three brothers with Autism carry a Stop-Gain Mutation in the HPA-Axis Gene NR3C2. Autism Research. 2020; 13(4): 523-531.
42. Ruzzo EK, Cano LP, Jung JY, Wang LK, Haghighi DK, et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell. 2019; 178(4): 850-866.e26.
43. PTPN11. SFARI Gene. https://gene.sfari.org/search?search= PTPN11. Online access.
44. PTPN6. NCBI. https://www.ncbi.nlm.nih.gov/gene/119858545. Online access.
45. Sabaie H, Moghaddam MM, Moghaddam MM, Amirinejad N, Asadi MR, et al. Long non-coding RNA-associated competing endogenous RNA axes in the olfactory epithelium in schizophrenia: a bioinformatics analysis. Sci Rep. 2021; 11(1): 24497-24505.
46. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515: 216-221.
47. Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. 2017; 49(4): 515-526.
48. Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, et al. Genomic Patterns of De Novo Mutation in Simplex Autism. Cell. 2017; 171(3): 710-722 e12.
49. Pecci A, Ma X, Savoia A, and Adelstein RS. MYH9: Structure, functions, and role of non-muscle myosin IIA in human disease. Gene. 2018; 664: 152-167.
50. Juarez GA, González CL, and Manzanares MV. Linking the Landscape of MYH9-Related Diseases to the Molecular Mechanisms that Control Non-Muscle Myosin II-A Function in Cells. Cells. 2020; 9(6): 1458.

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

51. Heath KE, Barros AC, Toren A, Granot GR, Carlsson LE, et al. Nonmuscle myosin heavy chain IIA mutations define a spectrum of autosomal dominant macrothrombocytopenias: May-Hegglin anomaly and Fechtner, Sebastian, Epstein, and Alport-like syndromes. Am J Hum Genet. 2001; 69(5): 1033-1045.
52. Marchani EE, Chapman NH, Cheung CYK, Ankenman K, Stanaway IB, et al. Identification of rare variants from exome sequence in a large pedigree with autism. Hum Hered, 2012; 74(3-4): 153-164.
53. Wang T, Hoekzema K, Vecchio D, Wu H, Sulovari A, Coe BP, et al. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. Nat Commun. 2020; 11(1): 4932.
54. Deciphering Developmental Disorders. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017; 542(7642): 433-438.
55. Coe BP, Stessman HAF, Sulovari A, Geisheker MR, Bakken TE, et al., Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. Nat Genet. 2019; 51(1):106-116.
56. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am J Hum Genet. 2016; 98(1): 58-74.

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Supplementary table 1. The top 29 most dysregulated genes.

No	EntrezID	Gene Name	Combined Tstat	Combine P-value	Regulation
1	4306	*NR3C2*	-42.933	8.49E-05	Down-regulated
2	23243	*ANKRD28*	-42.288	8.49E-05	Down-regulated
3	23060	*ZNF609*	-43.759	8.49E-05	Down-regulated
4	283267	*LINC00294*	-40.467	0.00015181	Down-regulated
5	7559	*ZNF12*	-39.112	0.00023152	Down-regulated
6	339345	*NANOS2*	36.777	0.00041468	Up-regulated
7	5814	*PURB*	-37.177	0.00041468	Down-regulated
8	5781	*PTPN11*	-36.425	0.00041468	Down-regulated
9	122525	*C14orf28*	-36.767	0.00041468	Down-regulated
10	10643	*IGF2BP3*	36.47	0.00041468	Up-regulated
11	28996	*HIPK2*	-34.933	0.00076387	Down-regulated
12	6433	*SFSWAP*	-34.538	0.00084412	Down-regulated
13	92399	*MRRF*	34.274	0.00088271	Up-regulated
14	25782	*RAB3GAP2*	-33.96	0.00095044	Down-regulated
15	22973	*LAMB2P1*	-33.072	0.001349	Down-regulated
16	255027	*MPV17L*	32.575	0.0015042	Up-regulated
17	3308	*HSPA4*	32.667	0.0015042	Up-regulated
18	6668	*SP2*	-32.127	0.0016234	Down-regulated
19	114783	*LMTK3*	-32.204	0.0016234	Down-regulated
20	8445	*DYRK2*	-32.068	0.0016234	Down-regulated

AJMB AN OPEN ACCESS JOURNAL
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2022.6.2.262
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

21	4627	*MYH9*	-31.742	0.0016459	Down-regulated
22	5578	*PRKCA*	-31.888	0.0016459	Down-regulated
23	669	*BPGM*	31.788	0.0016459	Up-regulated
24	8742	*TNFSF12*	-31.14	0.0018245	Down-regulated
25	29098	*RANGRF*	31.054	0.0018245	Up-regulated
26	23211	*ZC3H4*	-31.03	0.0018245	Down-regulated
27	126917	*IFFO2*	-31.08	0.0018245	Up-regulated
28	5704	*PSMC4*	31.177	0.0018245	Up-regulated
29	51248	*PDZD11*	31.326	0.0018245	Up-regulated