

The Ectodermal-Brain Memory to Explain the Methylation Pattern of the ASD Gene Matrix

zehra kirdar

z.kirdar82@gmail.com

university of kirkuk

Humada Yasser Hamad

university of kirkuk

Ibtesam Ghadban Auda

Mustansiriyah University

Article

Keywords: Ectodermal brain memory, autism spectrum disorder (ASD), DNA methylation (DNAm), NGS-based whole genome bisulfite sequencing, differentially methylated regions (DMRs)

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11088809/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

The Ectodermal-Brain Memory to Explain the Methylation Pattern of the ASD Gene Matrix

Zehra Muhammed Kirdar^{1*}, Humada, Yasser Hamad^{2*}, and Ibtesam Ghadban Auda^{3*}

1*Department of Biology, College of Science, Kirkuk University, Kirkuk, Iraq.

2* Department of Biology, College of Science, Kirkuk University, Kirkuk, Iraq.

3* Department of Biology, College of Science, Al-Mustansiriyah University, Baghdad, Iraq.

1*Email: z.kirdar82@gmail.com

2*Email: yasirhamad1974@uokirkuk.edu.iq

3* Email: stbg@uomustansiriyah.edu.iq

Abstract

Background

The methylation of the genes in Autism Spectrum Disorder (ASD) is now the most controversial topic. A large cohort of ASD individuals was genetically intact. This fact was a hallmark of the epigenetic regulation of ASD, especially the promoter region of the genes. DNA methylation is a challenging subject to study, as it is influenced by multiple factors. The brain is not a touchable tissue; postmortem brain tissue could give false positives because of apoptosis bias. Therefore, the study selected the ectodermal brain memory by collecting buccal swabs.

Methods

Pre-diagnosed ten individuals with ASD and two neurotypical controls were selected for buccal swab collection, DNA extraction, and NGS-based whole genome bisulfite sequencing. To avoid sex bias, we separated male ASD as A2 and female ASD as A3, while A1 represents all ASD. All ASD cohorts were compared with neurotypical controls. To avoid sex bias and germline bias, we made comparisons between two neurotypical controls, male and female A4 (negative control).

Results

The study screened about 8209 differentially methylated regions (DMRs), 4,234 genes, and 1,121 DMR-anchored promoter region genes in the A1 cohort. 12,353 DMRs, 5,247 genes, and 1,564 DMR-anchored promoter region genes in the A2 cohort. 7,761 DMRs, 4,195 genes, and 1,329 DMR-anchored promoter region genes in the A3 cohort. 10,676 DMRs, 4,820 genes, and 1485 DMR-anchored promoter region genes in A4. SFARI overlap is consistently higher in the broader GO lists than in the promoter subsets, while A3 (female cohort) shows the highest promoter SFARI overlap at the descriptive count level. The study suggested that the following genes showed aberrant methylation overlaps in the SFARI database: high confidence like **CACNA2D3, ATRX, MECP2, SKI, CASK, GRIA2**; strong candidates **ADORA2A, CHRNA7, RPS6KA3, CACNA1B, MAOB, PIK3CG, LHX2, PITX1**; suggestive evidence **NPFFR2, PDHA1, POLR2A, KDM2A, TCEAL1**; syndromic genes **ZBTB18, HUWE1, HCFC1, RALA** and the paralogs of SFARI genes **PLCB1 (PLCB2) and WNT1 (WNT3)**. These genes, **CALML3, GNG4, MYL9, WNT10B, MSH5**, are biomarkers in DNA methylation in ASD. About 2511 genes were shared between A1, A2, and A3, which were not recorded

in the SFARI database. These genes were candidates for ASD biomarkers, including ADAM8, ASIC2, ASIC4; unclassified genes such as Chromosome Open Reading Frame C19orf44, C1orf159, C9orf3, Long Intergenic Non-Protein Coding RNA LINC00299, LINC01056; and FAM genes **FAM19A5** and **FAM171A1**.

Conclusion

Individuals with ASD have a distinct epigenetic signature from NT. A large number of genes were differentially methylated and overlapped with the SFARI database. The mutation and methylation are critical to developing ASD. Most of the genes and their paralogs were involved in synaptic differentiation. The aberrant methylation is of embryonic origin, and other environmental factors influenced the severity of the cases.

Keywords: Ectodermal brain memory, autism spectrum disorder (ASD), DNA methylation (DNAm), NGS-based whole genome bisulfite sequencing, differentially methylated regions (DMRs).

1 Introduction

Autism spectrum disorder (ASD) is a complex syndrome that can be viewed as a collection of early-onset neuroevolutionary disorders that appear to be caused by an alteration in brain connectivity, which has a domino effect on other neuropsychological processes [1, 83, 84, 107]. The majority of ASD cases result from a combination of environmental risk factors and genetic predisposition, which can be investigated using epigenetic mechanisms like DNA methylation [85, 92, 93, 116]. DNA methylation is a critical epigenetic mechanism that modifies DNA by adding a methyl group, usually at cytosine bases in CpG dinucleotides, to regulate gene expression without changing the underlying genetic sequence. In general, it works to preserve chromatin structure, suppress transposable elements, and silence genes. Normal development and differentiation depend on this process, which is controlled by DNA methyltransferases (DNMTs) [2, 59, 66].

DNA methylation and transcriptional regulation are significantly influenced by the methyl-CpG binding protein 2 (MECP2) gene, which can either activate or repress genes like brain-derived neurotrophic factor (BDNF) [3, 90, 114]. In autism spectrum disorder (ASD), decreased production of the MeCP2 protein is linked to increased methylation of the MECP2 promoter, especially in the frontal cortex of the brain. In male ASD brains, there is a correlation between elevated DNA methylation and decreased MECP2 expression, which is caused by this epigenetic dysregulation. It has been demonstrated that increased promoter methylation, particularly at specific CpG sites, causes autism-like characteristics, such as a reduction in social contact [4, 108, 109, 110].

ASD is thought to have its roots in abnormalities in neural stem cell niche regions of the developing brain that occur during pregnancy. Coordinating normal brain growth requires dynamic epigenetic mechanisms, such as DNA methylation. The degree and effects of changes in DNA methylation states in neural stem cell compartments in ASD

are unknown [5]. The WNT signaling pathway is a long-standing, extremely conserved system of signal transduction proteins that controls important biological functions such as organogenesis, migration, polarity, cell fate determination, and embryonic development [97, 98]. Dysregulation of the WNT signaling system, which frequently results in altered methylation of associated genes, impacts synaptic plasticity, neural development, and proliferation, all of which are factors in autism spectrum disorder (ASD) [7, 99, 100, 101, 102].

Oxytocin is a neuropeptide that has been conserved throughout evolution and is involved in a variety of social and reproductive behaviors. Its function in complex behavioral features and the pathophysiology of mental health issues has garnered significant interest [12]. The three primary genes in the oxytocin (OT) signaling pathway—OXTR (Oxytocin Receptor), OXT (the ligand or prepropeptide), and CD38 (which controls oxytocin secretion)—are often investigated in human social behavior and neuroscience [13, 94, 95, 96].

Through the activation of Gαq/11 or Gαi proteins, Ca²⁺ influx, and cAMP pathways, oxytocin (OXT) signaling regulates brain connectivity and synaptic function to control plasticity in dopaminergic, cholinergic, and GABAergic synapses. OXT frequently improves social behaviors by enhancing glutamatergic signaling and disinhibiting amygdala GABAergic interneurons [8, 9]. ASD is caused by dysregulated oxytocin (OT) signaling that disrupts synapse formation and E/I balance (GABA/glutamate) by combining with calcium and CAMK4-mediated cAMP pathways. Social deficiencies are correlated with decreased OT as well as impaired dopamine and cholinergic synapse function [10, 11, 112, 113].

Numerous malignancies are caused by dysregulation of the Ras/Raf/MAPK pathway, a crucial signaling cascade that controls cell division, survival, and proliferation (RAS-RAF-MEK-ERK). A similar GTPase called Rap1 controls metastasis and adhesion. Pathways, including MAPK and WNT, which are frequently stimulated by UV, are involved in melanogenesis in cancer. Although MAPK contributes to advanced disease progression, Hedgehog signaling is primarily responsible for basal cell carcinoma (BCC). RASopathies, which are hereditary conditions caused by mutations in the Ras-MAPK pathway and often result in neurodevelopmental, craniofacial, and cardiovascular abnormalities, are linked to ASD [14, 15].

Nucleotide excision repair (NER) and base excision repair (BER) are two important but different DNA repair processes. NER corrects large, helix-distorting damage (like UV-induced dimers), whereas BER corrects minor, non-helix-distorting base lesions (like oxidation and deamination). Autism spectrum disease (ASD) is increasingly associated with defective repair, which results in the accumulation of oxidative DNA damage [16, 17, 18, 19].

Circadian rhythms are intrinsic biological clocks that run around the clock and control mental, behavioral, and physical functions, including the sleep-wake cycle. The suprachiasmatic nucleus (SCN), a collection of nerve cells in the hypothalamus, serves as the body's master clock, coordinating peripheral clocks [20, 117, 118, 119]. With 50–80%

of children with autism spectrum disorder (ASD) experiencing sleep disturbances, circadian rhythm disturbance is very common in ASD. These disturbances, which have their origins in genetic, hormonal, and environmental variables, frequently show up as irregular sleep patterns and elevated melatonin/cortisol levels, which affect social, behavioral, and cognitive development [21].

ASD is closely linked to ABC (ATP-Binding Cassette) transporters, especially subfamily A (e.g., ABCA13 and ABCA7), because of their functions in blood-brain barrier maintenance, cholesterol trafficking, and neuronal development. These genes' mutations and deletions impair network function and neuronal maturation, making them possible risk factors for neurodevelopmental disorders [22, 23, 120].

Autism Spectrum Disorder (ASD) is closely associated with altered cell cycle regulation during fetal development, especially when early brain expansion is involved. Increased neural progenitor proliferation, which results in a greater number of cortical neurons, is caused by dysregulation of cell cycle genes, which are frequently downstream of genes linked to ASD, such as CHD8 and PTEN. Early juvenile brain growth is boosted in part by this aberrant cell proliferation [24, 25].

2 Materials and methods

2.1 Participants

The inclusion criteria for cases were as follows: Ten ASD samples were previously diagnosed with ASD based on the DSM-5, CARS, and M-CHAT. Two siblings, the first pair being both male and the other a male and a female. Two pairs of identical twins: one pair of identical males and one pair of identical females. Provided that one of the siblings is in a more severe condition. In addition, we selected three males and one female without siblings. Sex-matched individuals (one male and one female) with neurotypical development were included as controls. To rule out autistic symptomatology, a psychiatric assessment called the Social Responsiveness Scale (SRS) was used, along with the requirement that there be no psychiatric background. The study received official ethical and academic approval from the Institutional Scientific and Ethical Committee and Faculty Board of the affiliated special needs education centers at the University of Kirkuk (approval No: 9431-40-7, 9432-40-7, 9433-40-7). All methods and sample collection procedures were performed in strict accordance with the relevant guidelines and regulations. Informed consent was obtained from all participants' legal guardians. The sampling was done under the supervision of special needs carers and psychologists and sometimes in the presence of the parents. Some of the ASD participants were eliminated from sampling because they were incapable, and to avoid Post-Traumatic Stress Disorder (PTSD), according to the Helsinki Declaration, all the sampling steps were approved.

2.2 Sample collection and DNA extraction

The study adheres to human ethics; as we deal with a special needs community, we decided to collect non-invasive samples by scraping sterile buccal swabs to extract epithelial cells. The ectoderm is the source of both postmortem brain tissue and the

epithelial cells obtained from the mouth lining using a buccal swab. The buccal epithelium is a component of the oral epithelial lining that arises from the surface ectoderm. As the DNA extraction from buccal swabs was struggling, we used preserved buccal swabs and non-preserved samples, both of which were extracted until we reached the required concentration. The Canvax Buccal Swab Collection & Stabilization Kit is used for preserving buccal swabs. The kit kept the collected swab in a tube containing preservation solution; thus, the buccal cells can be transported and stabilized for 1 year at room temperature and indefinitely at -20°C or -80°C. For DNA extraction, we used the Canvax HigherPurity™ Buccal Swab Genomic DNA Extraction Kit and HigherPurity™ Swab DNA Extraction Mini Spin Kit. The Qubit Fluorometer was used to measure the concentration of the DNA. The total amount of the DNA should be >500 ng. The quality of all samples fulfilled all requirements for library preparation, and they were sent to Biomarker Technologies (Hong Kong) for sequencing.

2.3 Library Preparation and Sequencing Workflow

2.3.1 Library Preparation

DNA samples underwent quality control. Every eligible sample was subjected to methylation liquid-phase capture library preparation and panel sequencing. The bisulfite conversion rate was evaluated using identified CHH loci in the genome as internal controls, with a criterion of $\geq 99\%$, to guarantee the effectiveness of sequencing and analysis.

2.3.2 Sequencing

The Illumina sequencing platform was used to sequence qualified libraries once they had been pooled based on effective concentration and target data volume requirements. Sequencing by Synthesis (SBS) is the foundation of the sequencing principle. DNA polymerase, adapter primers, and four fluorescently labeled dNTPs were introduced to the flow cell for amplification. A matching fluorescence signal was released when a fluorescently labeled dNTP was incorporated during the complementary strand's extension in each sequencing cluster. These signals were recorded by the sequencer, and software transformed them into sequencing peaks to determine the fragment's sequence.

2.4 Methylome-seq library construction and sequencing

Whole genome bisulfite sequencing (WGBS) libraries were constructed using the Acegen Methylome-seq Library Prep Kit (Acegen, Cat. No. AG1209) according to the manufacturer's protocol. Briefly, $\geq 0.5 \mu\text{g}$ of genomic DNA spiked with 1% unmethylated lambda DNA was fragmented by sonication to an average size of 100–500 bp and subsequently treated with bisulfite. The heat-denatured, single-stranded DNA was then ligated to 3'-dA-tailed adapters. Next, the single-stranded DNA linked to the 3' adapter was extended to double-stranded DNA, followed by ligation of 5-methylcytosine-modified adapters. Finally, the DNA was amplified via ~10 cycles of PCR using Illumina 8-bp dual-index primers. This was followed by capture with the Acegen MethylDesign panel and subsequent PCR amplification of the captured library. The constructed WGBS libraries were analyzed using an Agilent 2100

Bioanalyzer for quality control and were ultimately sequenced with a 150×2 paired-end sequencing protocol. Following sequencing, the original data were uploaded to the NCBI SRA database ([PRJNA1397483](https://www.ncbi.nlm.nih.gov/sra/PRJNA1397483)) for additional analysis and scientific community sharing. The SRA records can be accessed at (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1397483>).

2.4.1 Library structure

5'-

AATGATACGGCGACCACCGAGATCTACACIIIIIIITCGTCGGCAGCGTCAGA
TGTGT ATAAGAGACAG-NNNNNN-
~~~CTGTCTCTTATACACATCTCCGAGCCCACGAGACIIIIIIIA  
TCTCGTATGCCGTCTTCTGCTTG-3'

IIIIIII: Index 2 (i5), 8 bases

IIIIIII: Index 1 (i7), 8 bases

-NNNNNN-: Insert sequence

~~~: tail

2.5 Bioinformatics Analysis Workflow

Species: hsa

Genome Information: hg38

Upon obtaining raw sequencing reads (Sequenced Reads), the analysis typically involves the following steps:

2.5.1 Sequencing Data Quality Assessment

Raw data was produced by high-throughput sequencing (such as the Illumina PE150) and stored in FASTQ (fq) format. Each read in the FASTQ text format, which stores biological sequences and the associated quality scores, consists of four lines

```
@ACEGEN01:289:C3Y96ACXX: 6:1101:1704:2425 1: N:0: GGCTAC
GCTCTTTGCCCTTCTCGTCGAAAATTGTCTCCTCATTCGAAACTTCTCTGT
+
@@CFFFDEHHHHFIJJ@FHGIIIEHIIJBHHHIJEGIIJJIGHIGHCCF
```

Based on the ASCII code, each base's quality score was extracted from the matching character in Line 4. The sequence's quality score, QPhred, was defined as $Q_Phred = -10 \log_{10}(e)$.

Raw Data Quality Assessment: FastQC, a commonly used quality assessment tool, was applied to raw reads. **Raw Data Filtering:** The goal of data filtering is to guarantee clean data quality by eliminating low-quality reads. Trimming was done with Trimmomatic. **Data Mapping Analysis:** To map bisulfite-converted methylation sequencing data to the reference genome, we used the BSMAP software. BSMAP achieved quick and precise alignment of bisulfite-converted sequences by combining a bitwise

masking algorithm with a genomic hash table. It allowed for quick comparison between bisulfite-converted reads and the native genome by avoiding genome sequence conversion. BSMAP is quicker, more sensitive, and more adaptable than current bisulfite alignment techniques.

Methylation Locus Detection: The methylation locus detection was carried out utilizing the BSMAP software's built-in detection and statistical tools after read alignment. The methylation levels' chromosomal distribution was shown using the Circos program. Genome-Wide Methylation Level Correlation Analysis: Pearson correlation analysis was performed after methylation levels were computed for 20 Kbp bins across sequence contexts. Genome-Wide Methylation Level Principal Component Analysis (PCA) was applied. In Genome-wide Methylation Level Clustering Analysis, the pairwise distance was estimated using the Pearson correlation coefficient, and the clustering technique used was the unweighted pair-group approach with arithmetic means (UPGMA).

Differentially Methylated Region (DMR) Detection: In order to quickly and precisely identify de novo DMRs between matched or group samples, the metilene software was utilized for DMR analysis. It used a binary segmentation algorithm in conjunction with dual statistical tests (MWU test and 2D KS test). DMRs were obtained utilizing CG loci for detection following multiple testing corrections. IGV Visualization of DMR Regions: For display in genomic data browsers, DMR data were stored in BED format. Circos plots, which show statistically significant levels (corrected p-values) of DMRs, were used to depict the genomic distribution and significance of DMRs. GO Enrichment Analysis of DMR-Related Genes, done based on the Gene Ontology (GO, <http://www.geneontology.org/>) database. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a primary public database for pathway analysis [82]. KEGG Enrichment Analysis of DMR-Related Genes Based on DMR annotation results was performed for genes overlapping with DMRs.

2.6 Statistical analysis

The data were analyzed with IBM SPSS Statistics (version 31.0.2.0), for descriptive and inferential statistical analysis. Categorical variables were presented as frequencies and percentages, and continuous variables were presented as mean $\pm$ standard deviation. Group comparisons were done with Welch's t-test, one-sample t-test, Mann-Whitney U test, exact permutation test and Fisher's exact test as per the test case. For SFARI gene analysis, Fisher's exact test was run to test whether the overlap between each gene list and the reference SFARI gene list was higher or lower than expected due to chance within a fixed background gene universe. Odds ratio values above 1 indicated enrichment, and odds ratio values below 1 indicated depletion. Multiple testing was controlled using the Benjamini-Hochberg false discovery rate (FDR) method, and $P < 0.05$ was used as the threshold for statistical significance. In some analyses, robust ANCOVA (HC3) was applied after logit transformation of methylation percentages.

3 Results

3.1 Differential Methylation Analysis

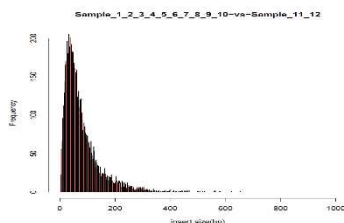
Differentially methylated regions (DMRs) are believed to have a role in transcriptional control, and the methylation status can change among the genomes of various samples (tissues, cells, and individuals). Finding DMRs makes it possible to thoroughly examine the variations in epigenetics between tissues. Cell differentiation and proliferation are known to be correlated with DNA methylation, and several DMRs have been found throughout the developmental stages, including embryonic reprogramming.

Cancers show local hypermethylation in areas such as tumor suppressor gene promoter CpG islands and global methylation loss in comparison to normal samples. This is a presentation of DMR analysis between comparable groups. The four groups we chose for differential comparison were as follows:

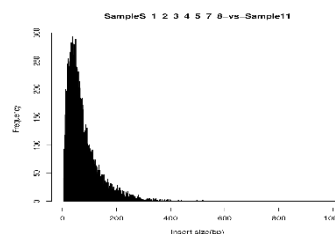
- 1- A1: Represents all ASD samples (male, female) vs. controls (male, female). (8,209 DMRs), 4,234 genes, and 1,121 DMR-anchored promoter region genes.
- 2- A2: Represents all male ASD samples vs. male control. (12,353 DMRs), 5,247 genes, and 1,564 DMR-anchored promoter region genes.
- 3- A3: Represents all female ASD samples vs. female control. (7,761 DMRs), 4,195 genes, 1,329 DMR-anchored promoter region genes
- 4- A4: Represents male control vs. female control. (10,676 DMRs), 4,820 genes, 1,485 DMR-anchored promoter region genes

Another comparison was done between ASD twins and siblings, and further against controls. All results were filtered from the A4 group to avoid tissue-specific bias, sex bias, and neurodevelopmental bias. Regional categories (e.g., promoter, exon, intron, gene body, upstream, and downstream) were used to annotate DMRs, and regional data were appended to each DMR. DMR Length Distribution, DMR Methylation Level Distribution, DMR Methylation Level Clustering Heatmap, DMR Anchored Region Distribution, and DMR Significance Circos Plots were performed.

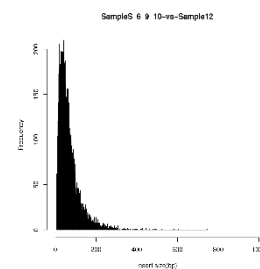
(A)



A1.



A2.



A3.

(B)

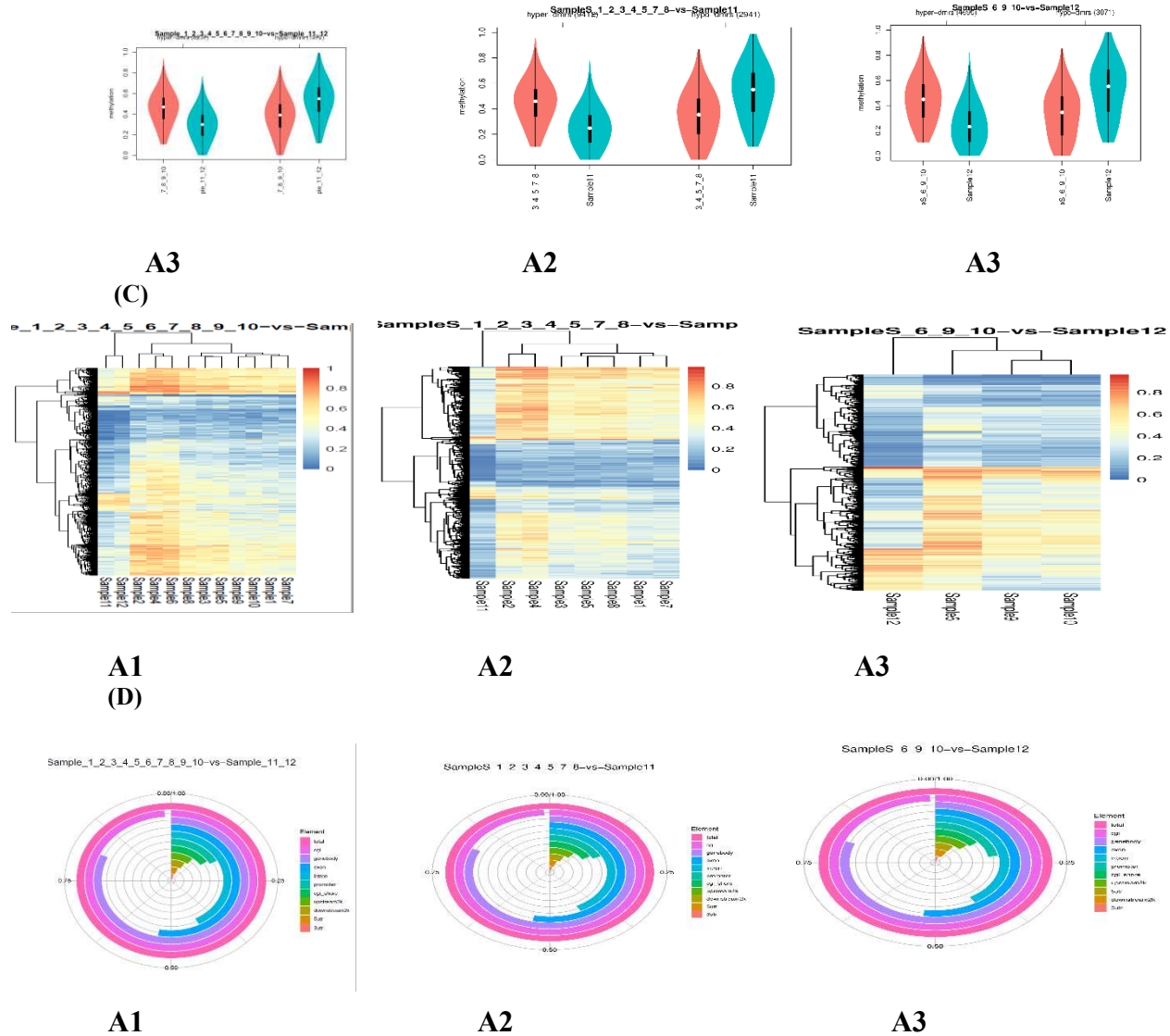
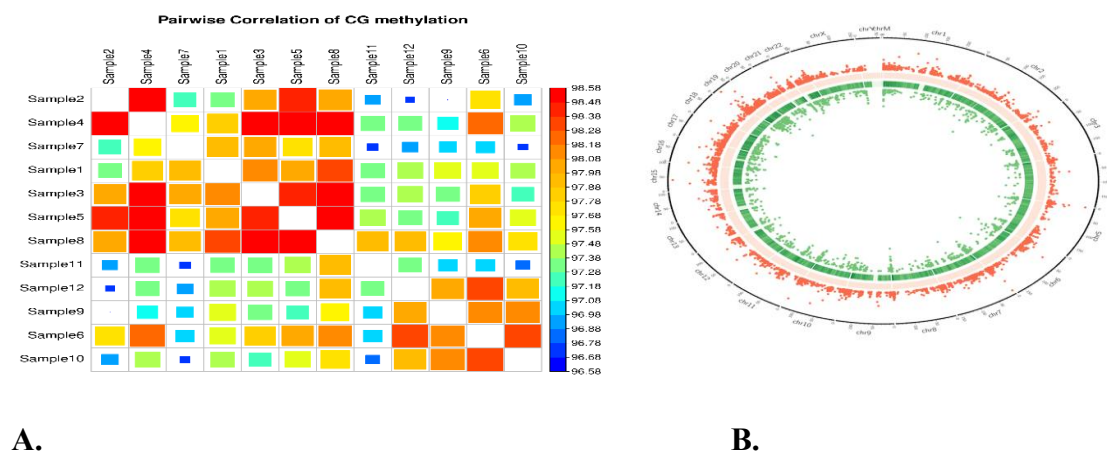
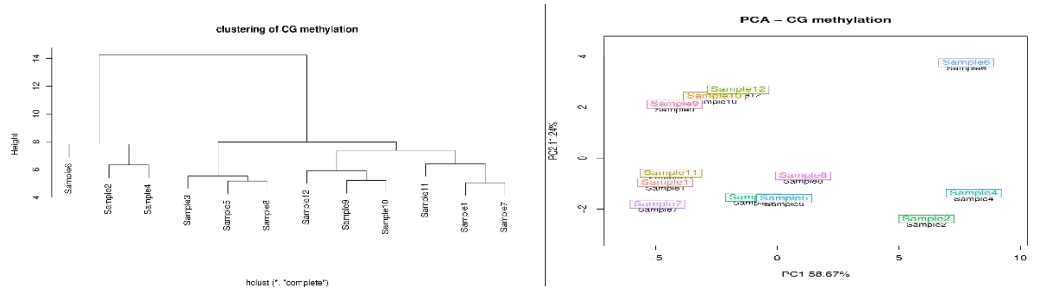


Figure (1): (A) DMR Length Distribution; (B) Average Methylation Level Distribution of DMRs; (C) Clustering Heatmap of DMR Methylation Levels; (D) Proportional Distribution of DMRs in Anchored Regions.





C. **D.**
Figure 2: (A) Sample Correlation Analysis; (B) Genome-Wide DMR Significance Distribution ;(C) Clustering Analysis Plot of Methylation Levels ;(D) PCA Analysis of Inter-Sample Methylation.

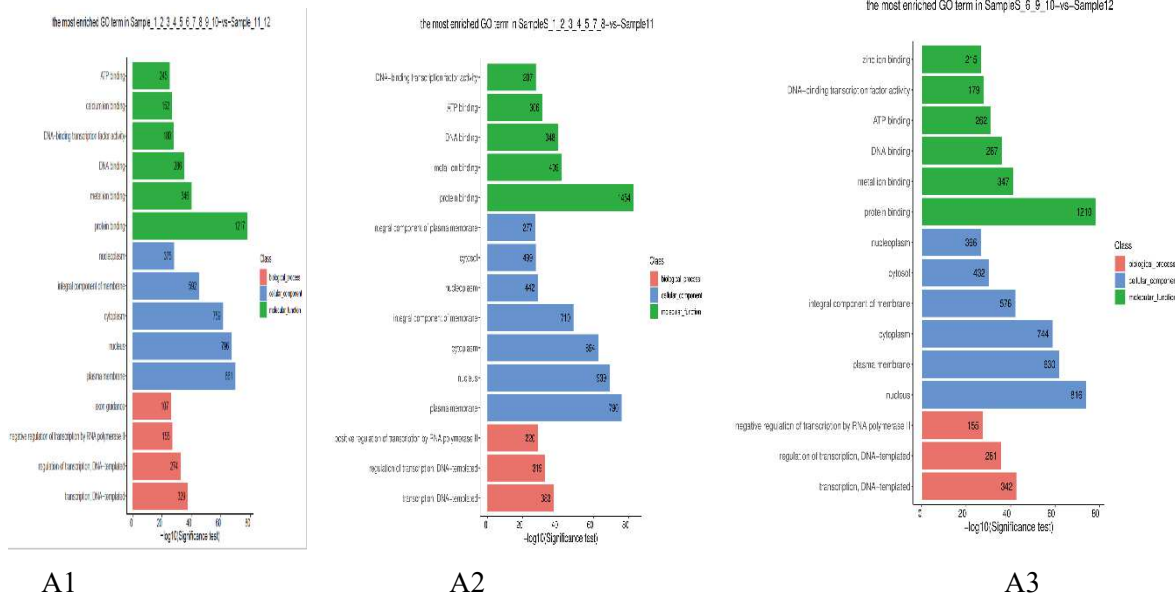
3.2 GO Enrichment Analysis of DMR-Related Genes

This analysis is based on the Gene Ontology database; the Gene Ontology Consortium was founded to develop a species-universal semantic language standard for characterizing and specifying the activities of genes and proteins. It thoroughly characterizes the characteristics of genes and gene products in organisms by describing the functions of proteins and genes within cells using dynamically controlled vocabularies. Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), which describe molecular functions, subcellular locations, and biological processes of gene products, respectively, are the three primary categories of the GO database. Gene annotation data from the GO database, including GO functional categorization and enrichment outcomes, were used to conduct gene enrichment analysis. While enrichment finds functional categories that are noticeably overrepresented in comparison to a reference gene set, functional categorization counts genes at a certain functional level.

Functional enrichment analysis was performed for genes that overlapped with DMRs in the gene body and upstream/downstream areas (upstream 2k, downstream 2k, and gene body) based on the results of DMR annotation. The criteria for selecting the gene ontology entries to be studied and to make comparisons between groups should be compatible with the following features: $\text{adjP} < 0.05$, Fold Change > 2 , and observed > 2 . Then, the common entries across all groups, significant or not, were selected to understand the differences in methylated genes between groups by filtering all genes that appear in A4.

In addition, entries that show specific significance in the female or male group were selected. Entries related to the spectrum have priority. The entries were shown in Figure 3. The DNA-binding GO were selected to be studied between cohorts.

(A)



(B)

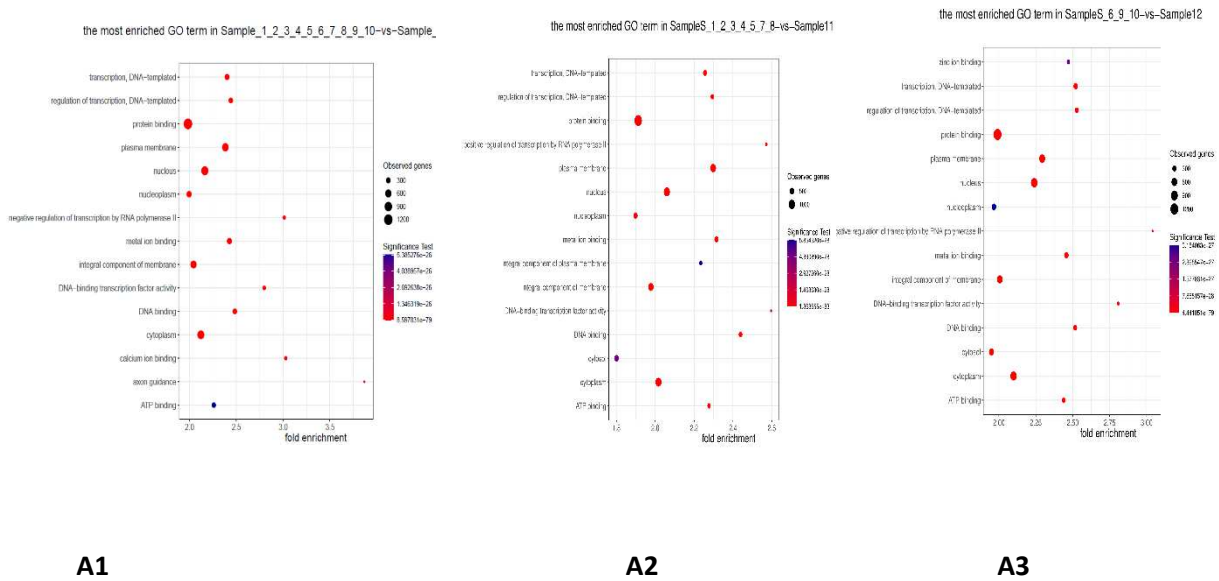


Figure (3): (A) Bar Chart of Enriched GO Terms. Horizontal bar chart of GO-related gene classifications: Y-axis = enriched GO term; X-axis = $-\log_{10}(adjP\text{-value})$. Longer bars indicate more significant statistical differences. Numbers on the right of bars denote the number of enriched genes and enrichment fold. (B) Scatter Plot of GO Enrichment. Note: Y-axis: pathway name; X-axis: fold change. Dot size indicates the number of DMR-related genes, and color corresponds to Q-value range (redder = more significant).

Table (1) GO Enrichment Analysis of DMR-Related Genes for all cohorts (DNA binding GO)

| Term Name | | | rawP | Fold Change | Expected | Observed | Category |
|---------------------------------|------|----------|----------|-------------|----------|----------|----------|
| Molecular function: DNA binding | adjp | | | | | | |
| GO:0003677 | A3 | 7.32E-33 | 7.76E-36 | 2.484499 | 115.1138 | 286 | 1548 |
| | A2 | 9.21E-38 | 7.84E-41 | 2.438985 | 142.6823 | 348 | 1548 |
| | A3 | 9.32E-34 | 1.01E-36 | 2.516365 | 114.0534 | 287 | 1548 |
| | A4 | 4.05E-35 | 4.06E-38 | 2.441386 | 131.0731 | 330 | 1548 |

All gene cohorts were filtered from A4 (negative control).

A1 unique: DUX4L7, EBF2, POLR2E, SOX15, STAT3, STAT5A, ZBTB18

A2 unique: CDKN2A, DFFB, EMX1, FOXL1, FOXO3, GMEB2, HIST1H2AA, HIST1H2BA, HIST2H3PS2, HOXD12, HOXD4, KDM2A, MEIS3P2, MLX, NR2F6, PCBP4, SCX, UBTF, ZFP36, ZKSCAN4, ZNF132, ZNF232, ZNF366, ZNF782

A3 unique: AIFM1, ATRX, BHLHE23, DNASE1L1, FOXF2, FOXO4, HOXB1, HUWE1, KLF8, LHX2, NEUROD6, PHF6, PITX1, SP8, SP9, TCEA1, ZMAT1, ZNF208, ZNF354B, ZNF41, ZNF449, ZNF81

Shared A1, A2, and A3: AKAP8, ASCL4, EEF1D, EGR3, EME2, HNF1A, LZTS1, SAFB2, SP6, TCF7L1, ZBTB32, ZNF296, ZNF701

In cohort A1

ZBTB18 zinc finger and BTB domain containing 18. Is syndromic in the SFARI GENE score S. Genes involved in neural development are transcriptionally repressed by the C2H2-type zinc finger protein that this gene produces. The encoded protein targets genes by identifying a particular sequence motif and enlisting chromatin components [32]. **EBF2** is the paralog of EBF3 (early B-cell factor 3), classified as High Confidence, Syndromic in the SFARI GENE Score 1S. **POLR2E** is the paralog of the gene **POLR2A** RNA polymerase II subunit A. It is suggestive evidence in the SFARI GENE score 3 [32].

In cohort A2

KDM2A (lysine demethylase 2A) is suggestive evidence in the SFARI GENE score3. This gene produces a member of the F-box protein family, which is distinguished by the F-box, a motif of about 40 amino acids [32].

In cohort A3

HUWE1 HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase, is syndromic in the SFARI GENE score S. E3 ubiquitin-protein ligase, which facilitates target protein ubiquitination and subsequent proteasomal destruction. catalyzes the polyubiquitination and degradation of MYCN, which controls neuronal proliferation and development [32]. **LHX2** LIM homeobox 2 is a strong candidate in the SFARI GENE Score 2. This gene produces a protein that is part of a broad family of proteins that have the LIM domain, a special zinc-binding domain rich in cysteines. The encoded protein may serve as a transcriptional regulator [32].

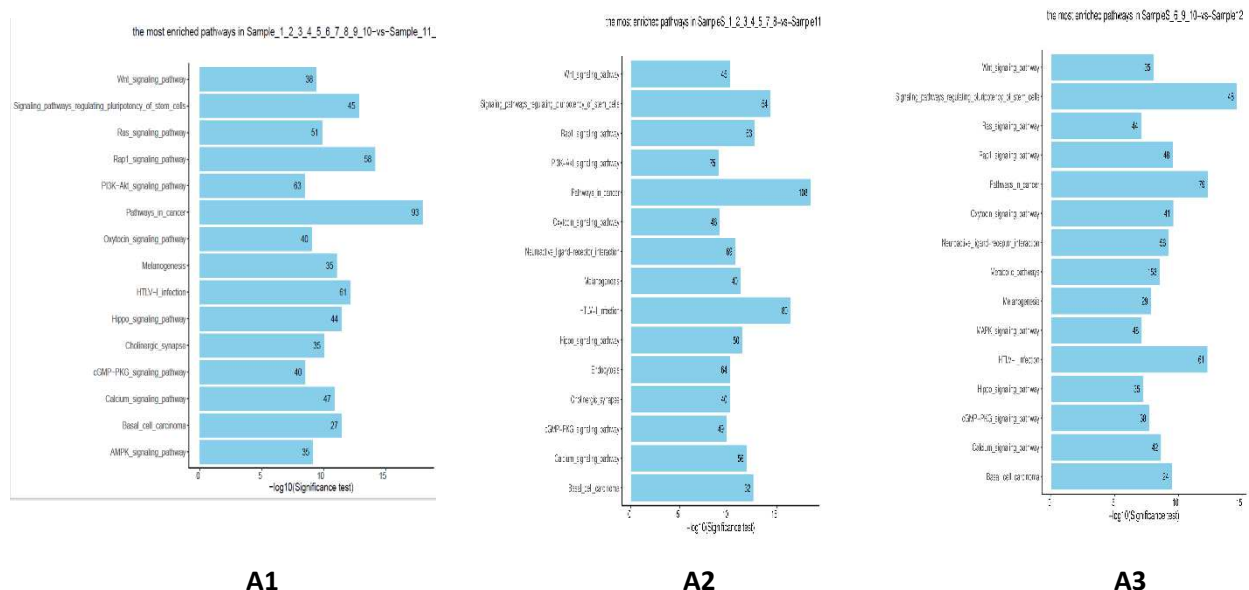
TCEAL1 transcription elongation factor A like 1 is suggestive evidence in the SFARI GENE score 3. This gene produces a member of the family of genes known as transcription elongation factor A (SII)-like (TCEAL) [32]. **PITX1** paired-like homeodomain is a strong candidate in SFARI GENE Score 2. It acts as a transcription factor [32].

HOXB1 is the paralog of **HOXA1** homeobox A1, considered syndromic in the SFARI GENE score S. The encoded protein is a transcription factor that is important for the development of the brainstem and skull [32]. **KLF8** is the paralog of the **KLF7** Kruppel like factor 7. Is suggestive evidence in the SFARI GENE score 3 [32] that belongs to the peptidase C19 family and produces a protein that matches ubiquitin-specific proteases. Despite being on the X chromosome, this gene avoids X-inactivation [32]. Other gene paralogs were identified, including **LZTS2(LZTS1)**, **NR2F1(NR2F6)**, **TCF7L2(TCF7L1)**. The results showed that the ASD female cohort has a high enrichment in the SFARI GENE, consistent with the female protective effect hypothesis.

3.3 KEGG Enrichment Analysis of DMR-Related Genes

Functional enrichment analysis was conducted for genes that overlapped with DMRs in gene body and flanking regions (upstream2k, gene body, and downstream2k) based on DMR annotation results. One of the most important public databases for pathway analysis is the Kyoto Encyclopedia of Genes and Genomes (KEGG) [82]. KEGG enrichment data are visualized using scatter plots, which quantify enrichment by fold change, Q-value, and gene count. The ratio of DMR-related genes to annotated genes in the pathway is known as fold change; a higher fold change indicates greater enrichment. The multiple testing corrected p-value (range [0,1]; closer to 0 = more significant) is known as the Q-value. A bar chart of enriched KEGG pathways is shown below:

(A)



Below is a scatter plot of enriched KEGG pathways:

(B)

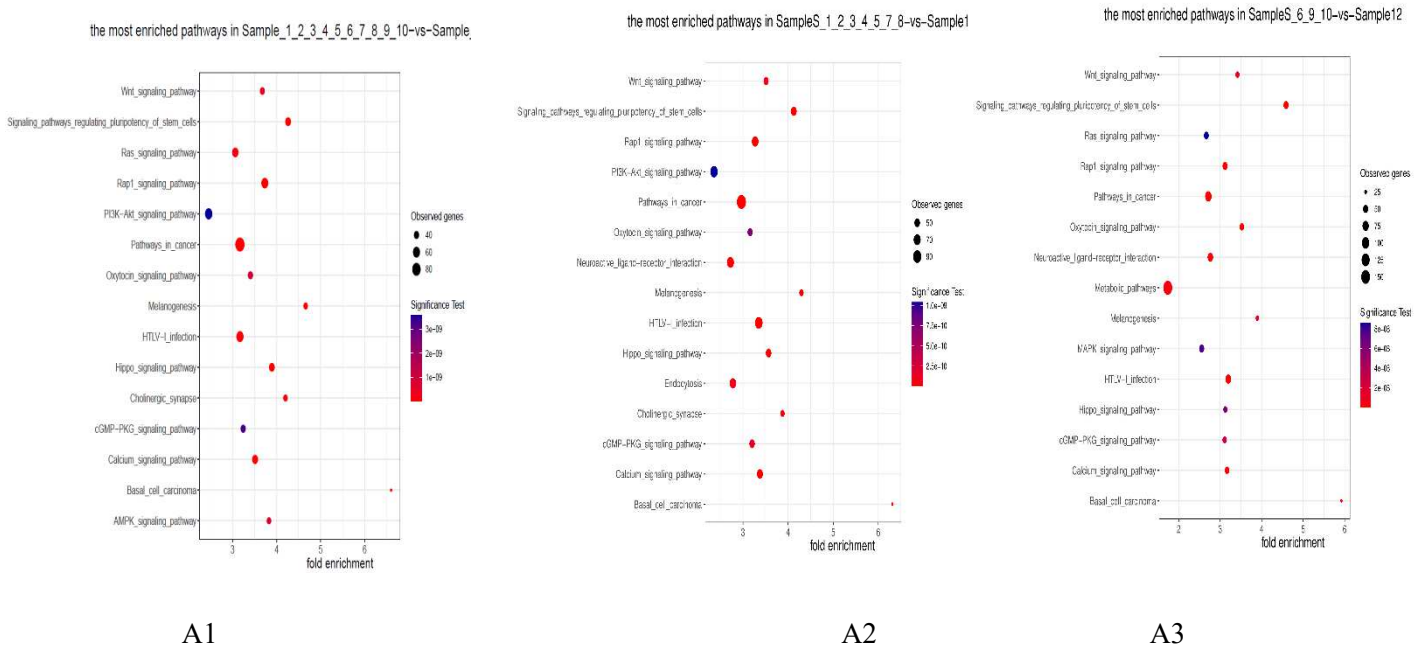


Figure (4): (A) Bar Chart of Enriched KEGG pathways. Horizontal bar chart of KEGG pathway classifications: Y-axis = enriched KEGG pathway; X-axis = $-\log_{10}(\text{adjP})$. Longer bars indicate more significant statistical differences. Numbers on the right of bars denote the number of enriched genes and enrichment fold. (B) Scatter Plot of Enriched KEGG Pathways. Note: Y-axis: pathway name; X-axis: fold change. Dot size indicates the number of DMR-related genes, and color corresponds to the Q-value range (redder = more significant).

According to the results obtained from the KEGG analysis of DMR-related genes, it is noticeable that the **oxytocin signaling pathway** was common between groups and represents a connection between other pathways, like the **calcium signaling pathway**. And plays a key role in controlling plasticity in **dopaminergic, cholinergic, and GABAergic synapses**. All cohorts' gene results were filtered from A4.

Table (2) KEGG Enrichment Analysis of DMR-Related Genes in all cohorts (oxytocin signaling pathway)

| Oxytocin signaling pathway | | | | | | |
|--|----------|----------|-----------|-------------|----------|----------|
| A1(All ASD samples vs control) | | | | | | |
| Term | Category | Observed | Expected | Fold Change | rawP | adjP |
| hsa04921 | 158 | 40 | 11.74934 | 3.404447 | 9.73E-10 | 2.12E-08 |
| A2 (Male ASD samples vs. male control) | | | | | | |
| Term | Category | Observed | Expected | Fold Change | rawP | adjP |
| hsa04921 | 158 | 46 | 14.563184 | 3.15865 | 7.45E-10 | 1.53E-08 |
| A3 (female ASD samples vs. female control) | | | | | | |
| Term | Category | Observed | Expected | Fold Change | rawP | adjP |
| hsa04921 | 158 | 41 | 11.641112 | 3.522 | 2.53E-10 | 1.59E-08 |

A1: RYR2; PLCB2; MYL9; KCNJ12; ELK1; CALML3; CACNA2D3; MYLK4; CAMK2G;

A2: RYR2; KCNJ12; ELK1; CALML3; CACNA2D3; MYLK4; CAMK2G; KCNJ3

A3: KCNJ12; CACNA2D3; RYR2; MYLK4; ELK1; PLA2G4C; CACNG7; MYL9

A1, A2, A3: RYR2—KCNJ12—ELK1—CACNA2D3—MYLK4

Unique Genes (Appearing in Only One List):

A1 only: PLCB2

A2 only: KCNJ3

A3 only: PLA2G4C—CACNG7

CACNA2D3 Calcium channel, voltage-dependent, alpha 2/delta subunit 3

was shared between all cohorts and has a high confidence SFARI Gene Score. This gene was found to be highly enriched for variants likely to impact ASD risk with a false discovery rate (FDR) of less than 0.1 in ASD whole-exome sequencing research and subsequent TADA (transmission and de novo association) analysis [31]. This gene produces a protein in the voltage-dependent calcium channel complex that belongs to the alpha-2/delta subunit family. The calcium current density and activation/inactivation kinetics of voltage-dependent calcium channels are controlled by the alpha-2/delta subunit. It functions as a regulatory subunit for the L-type (CACNA1C or CACNA1D), N-type (CACNA1B), and P/Q-type calcium channels (CACNA1A), but not the T-type (CACNA1G) [32]. The paralogs of these genes were PLCB1 (PLCB2) and CACNG2 (CACNG7).

Table (3) KEGG Enrichment Analysis of DMR-Related Genes in all cohorts (calcium signaling pathway)

| Calcium signaling pathway | | | | | | |
|----------------------------------|----------|----------|-----------|-------------|----------|----------|
| A1 Term | Category | Observed | Expected | Fold Change | rawP | adjP |
| hsa04020 | 180 | 47 | 13.385321 | 3.511309 | 1.49E-11 | 5.27E-10 |
| A2 Term | Category | Observed | Expected | Fold Change | rawP | adjP |
| hsa04020 | 180 | 56 | 16.590969 | 3.37533 | 1.36E-12 | 6.53E-11 |
| A3 Term | Category | Observed | Expected | Fold Change | rawP | adjP |
| hsa04020 | 180 | 42 | 13.262026 | 3.166937 | 2.36E-09 | 8.34E-08 |

All cohorts' gene results were filtered from A4.

A1: ADORA2A; ADRB1; CALML3; CAMK2G; GNA15; HTR5A; ITPKB; MYLK4; ORAI3; PDGFRB; PLCB2; RYR2

A2: ADORA2A; ADRA1A; ADRA1B; ADRB1; CALML3; CAMK2G; CHRNA7; GNA15; HTR5A; ITPKB; MYLK4; ORAI2; ORAI3; PDGFRB; RYR2

A3: ADRB1; CHRNA7; GNA15; HRH2; MYLK4; ORAI2; RYR2; VDAC1

A1, A2, and A3: ADRB1, GNA15, MYLK4, RYR2.

Unique to A3: HRH2, VDAC1.

Unique to A2: ADRA1A, ADRA1B

Unique to A1: PLCB2

The ADORA2A adenosine A2a receptor was shared between A1 and A2. A strong candidate in the SFARI GENE score is 2. In a Caucasian population, a genetic correlation between autism and the ADORA2A gene was discovered. The encoded protein is an adenosine receptor subtype [32].

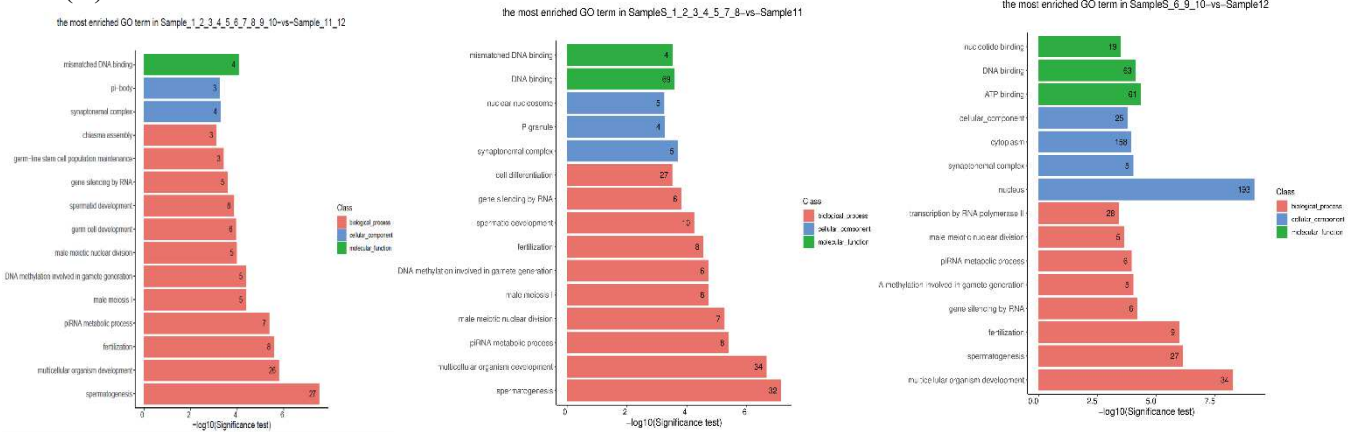
CHRNA7, cholinergic receptor, nicotinic, alpha 7

is shared by A2 and A3, is classified as a strong candidate in the SFARI GENE score of 2. Developmental delay and intellectual disability have been linked to an extremely rare deletion in the CHRNA7 gene [34]. Furthermore, two patients with autism and intellectual disability have an uncommon CHRNA7 duplication [35]. Nicotinic acetylcholine receptors (nAChRs), a superfamily of ligand-gated ion channels, facilitate quick signal transmission at synapses. It is believed that the nAChRs are heteropentamers made up of homologous subunits. This gene codes for a protein that forms a homo-oligomeric channel, has significant permeability to calcium ions, and is a key part of brain nicotinic receptors that are highly sensitive to and inhibited by alpha-bungarotoxin. After this receptor binds acetylcholine, all of its subunits experience a significant conformational shift that opens an ion-conducting channel across the plasma membrane [32]. The paralogs of these genes were ADRB2 (ADRB1) and PLCB1 (PLCB2).

3.4 GO Enrichment Analysis of DMR-Anchored Promoter Genes

Functional enrichment analysis was conducted for genes that overlapped with promoter regions (1.5K upstream to 0.5K downstream of TSS) based on DMR genomic distribution. Then, the common entries across all groups, significant or not, should be selected to understand the differences in genes between groups by filtering all genes that appear in A4. In addition, entries that show specific significance in the female or male group were selected. Entries related to the spectrum have priority. The distribution of gene counts across enriched GO terms in BP, CC, and MF is visually represented by bar charts showing GO enrichment for DMR-related genes. Plots display GO-related gene classifications in horizontal bar charts, selecting entries with adjP < 0.05, observed > 2, and fold change > 2. The enriched GO terms are presented in the bar charts below.

(A)

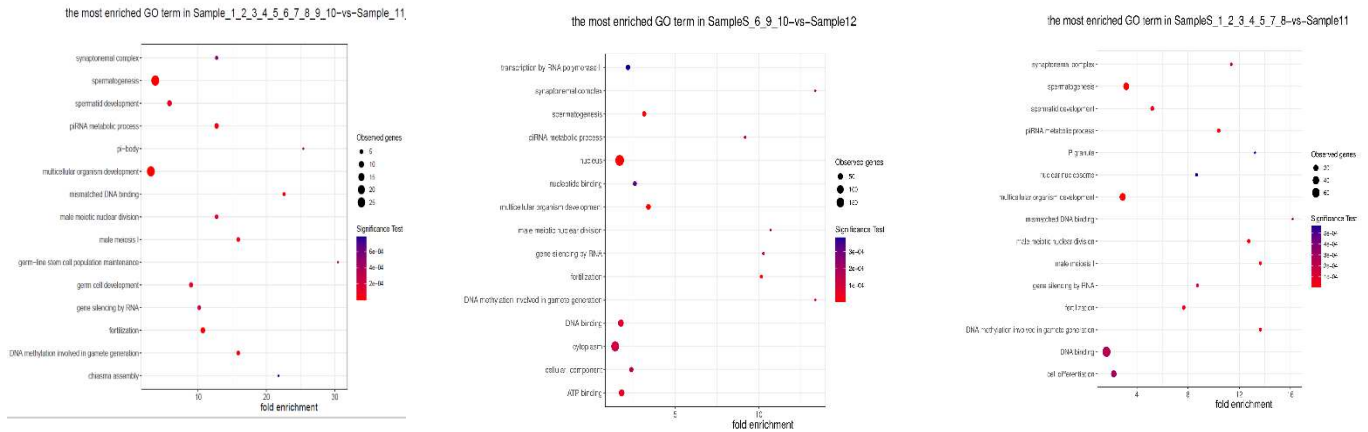


A1.

A2.

A3

(B)



A1.

A2.

A3.

Figure 5: (A) Bar Chart of Enriched GO Terms. Note: Horizontal bar chart of GO-related gene classifications: Y-axis = enriched GO term; X-axis = $-\log_{10}(\text{adjP})$. Longer bars indicate more significant statistical differences. Numbers on the right of the bars denote the number of enriched genes and the enrichment fold. (B) Scatter Plot of GO Enrichment. Note: Y-axis: pathway name; X-axis: fold change. Dot size indicates the number of DMR-related genes, and color corresponds to Q-value range (redder = more significant).

The GO terms were below in the table.

Table (4) GO enrichment results for DMR-anchored promoter region genes for all cohorts.

| Term Name | Gene List | | | | | | |
|---|--|-------------|-------------|----------|----------|----------|------------|
| Biological process: pi RNA metabolic process GO:0034587 | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.002029344 | E-064.27 | 12.697814 | 0.551276 | 7 | 28 | GO:0034587 |
| | A1: TDRD12;PIWIL2;DDX4;TDRD9;ASZ1;MOV10L1;FKBP6 | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| GENE DIFFERENCES (2) DDX4;ASZ1 | 0.003495105 | E-064.12 | 10.401349 | 0.769131 | 8 | 28 | |
| | A2: MOV10L1;DDX4;FKBP6;TDRD12;TDRD9;ASZ1;PIWIL2;PIWIL1 | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.028210517 | 0.000110705 | 9.180432 | 0.653564 | 6 | 28 | |
| A3: PIWIL1;MOV10L1;TDRD9;FKBP6;PIWIL2;TDRD12 | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.038541329 | 0.00019874 | 8.216015 | 0.730281 | 6 | | |
| | A4: TDRD9;TDRD12;FKBP6;MOV10L1;PIWIL1;PIWIL2 | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| Biological process: gene silencing by RNA GO:0031047 | 0.044628659 | 0.000258 | 10.158245 | 0.492211 | 5 | 25 | GO:0031047 |
| | A1:FKBP6;TDRD12;PIWIL2;ASZ1;TDRD9 | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.043242056 | 0.000152739 | 8.737135 | 0.686724 | 6 | 25 | |
| Gene Differences (1) ;ASZ1 | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.028210517 | E-056.43 | 10.282071 | 0.58354 | 6 | 25 | |
| | A3: PIWIL1;TDRD12;TDRD9;FKBP6;ER11;PIWIL2 | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| A4: PIWIL1;PIWIL2;TDRD9;ER11;TDRD12;FKBP6 | 0.037381524 | 0.000116168 | 9.201946 | 0.652036 | 6 | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.07557 | 0.000529 | 12.6978 | 0.315015 | 4 | 16 | GO:0000795 |
| | Cellular component: | | | | | | |

| | | | | | | | | |
|--|------------------------------------|-------------|-------------|-------------|-----------|----------|----------|------------|
| synaptonemal complex
GO:0000795

GENE DIFFERENCES(1)
MSH5 | A1: MSH5;FKBP6;MSH4;SYCE1 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.049867867 | 0.000195714 | 11.376487 | 0.439503 | 5 | 16 | |
| | A2: MSH4;SYCE1;MSH5;FKBP6;SYCP2 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| Biological process: chiasma
assembly GO:0051026

GENE DIFFERENCES(2)
MSH5;RNF212 | | 0.02821051 | E-059.34 | 13.388135 | 0.373465 | 5 | 16 | |
| | A3: FKBP6;SYCE1;SYCP2;MSH5;MSH4 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.038221895 | 0.000154859 | 11.981702 | 0.417303 | 5 | | |
| | A4: MSH4;SYCP2;HORMAD2;SYCE1;FKBP6 | | | | | | | |
| Biological process: mismatch repair GO:0006298

GENE DIFFERENCES(4)
MSH5-SAPCD1;MSH5;MSH3;
TDG | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.099307998 | 0.000784 | 21.767681 | 0.137819 | 3 | 7 | GO:0051026 |
| | A1: MSH5;RNF212;MSH4 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.427877108 | 0.022743955 | 10.401336 | 0.192283 | 2 | 7 | |
| Biological process: detection of
chemical stimulus involved in
sensory perception of smell
GO:0050911

GENE DIFFERENCES(3)
OR10H2;OR2B11; OR2L13 | A2: MSH5;MSH4 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.332448992 | 0.016868226 | 12.240576 | 0.163391 | 2 | 7 | |
| | A3: MSH5;MSH4 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| Molecular function: nucleosome
binding GO:0031491

GENE DIFFERENCES(1)
L3MBTL1 | | 0.416187259 | 0.186249547 | 5.477351 | 0.18257 | 1 | | |
| | A4:MSH4 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.318865114 | 0.005034 | 6.348904 | 0.63003 | 4 | 32 | GO:0006298 |
| | A1: MSH5;LIG1;MSH5-SAPCD1;TDG | | | | | | | |
| Biological process: dopaminergic
neuron differentiation
GO:0071542

GENE DIFFERENCES(1)
LMX1A | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.325393706 | 0.002937227 | 5.688237 | 0.879007 | 5 | 32 | |
| | A2: MSH5-SAPCD1;MSH5;TP73;LIG1;TDG | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.332448992 | 0.045314217 | 4.016435 | 0.746931 | 3 | 32 | |
| Biological process: chiasma
assembly GO:0051026

GENE DIFFERENCES(2)
MSH5;RNF212 | A3: MSH5-SAPCD1;MSH5;MSH3 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.442210252 | 0.214019989 | 2.396337 | 0.834607 | 2 | | |
| | A4: LIG1;TP73 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| Biological process: detection of
chemical stimulus involved in
sensory perception of smell
GO:0050911

GENE DIFFERENCES(3)
OR10H2;OR2B11; OR2L13 | | 0.339245425 | 0.006180835 | 0.120932 | 8.26914 | 1 | 420 | GO:0050911 |
| | A1:OR1M1 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.427877108 | 0.022059568 | 0.346712 | 11.536962 | 4 | | |
| | A2: OR10AC1;OR1M1;OR10H2;OR2B11 | | | | | | | |
| Molecular function: nucleosome
binding GO:0031491

GENE DIFFERENCES(1)
L3MBTL1 | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.142567178 | 0.001411201 | 0.102005 | 9.803467 | 1 | | |
| | A3: OR2L13 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.236272779 | 0.002697774 | 0.182578 | 10.954213 | 2 | | |
| Biological process: dopaminergic
neuron differentiation
GO:0071542

GENE DIFFERENCES(1)
LMX1A | A4: OR10AC1;OR1M1 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.399096 | 0.0291468 | 8.465214 | 0.236261 | 2 | 12 | GO:0031491 |
| | A1: L3MBTL1;NOC2L | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| Biological process: chiasma
assembly GO:0051026

GENE DIFFERENCES(2)
MSH5;RNF212 | | 0.226993776 | 0.00441225 | 10.710499 | 0.280099 | 3 | | |
| | A3::L3MBTL1;NOC2L;SMARCA1 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.380612499 | 0.048087929 | 6.390226 | 0.312978 | 2 | | |
| | A4: SMARCA1;NOC2L | | | | | | | |
| Biological process: dopaminergic
neuron differentiation
GO:0071542

GENE DIFFERENCES(1)
LMX1A | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.399096 | 0.041931 | 6.772177 | 0.295326 | 2 | 15 | GO:0071542 |
| | A1: LMX1A;DMRTA2 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.561186995 | 0.35195322 | 2.426984 | 0.412034 | 1 | 15 | |
| Biological process: chiasma
assembly GO:0051026

GENE DIFFERENCES(2)
MSH5;RNF212 | A2: DMRTA2 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | | 0.332448992 | 0.056502 | 5.712262 | 0.350124 | 2 | 15 | |
| | A3: MSH5;RNF212 | | | | | | | |
| | | adjP | rawP | Fold Change | Expected | Observed | Category | Term |

| | | | | | | | |
|--|---|-------------|-------------|-------------|----------|----------|------------|
| | A3: DMRTA2;LMX1A
adjP rawP Fold Change Expected Observed
0.514648319 0.337791273 2.556093 0.391222 1
A4: DMRTA2 | | | | | | |
| Molecular function: chromatin binding GO:0003682 | adjP | rawP | Expected | Fold Change | Observed | Category | Term |
| GENE DIFFERENCES(16)
MECP2; TOX; L3MBTL1;
TCF7; HOXD13; STAG2;
NKAP; ELK1; HCFC1;
ATRX;SKI; NKX6-1; ANKRD2;
KAT2A; ATR;CEBPB; | E-126.07 | E-143.38 | 2.86118 | 27.96044 | 80 | 376 | GO:0003682 |
| | A1: AIRE;MBD1;RBPJL;NOC2L;TOX;BHMGI;CEBPB;L3MBTL1;MEIOB;TCF7;CTBP2;POLE; | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.427877108 | 0.007262474 | 1.936422 | 10.328328 | 20 | 376 | |
| | A2: CEBPB;HMGA1;AIRE;MTA3;GATA6;NKX6-1;ARX;TCF7;NSD1;MEIOB;TP73;BHMGI;HHEX;MBD1;TOX;CTBP2;KAT2A;POLE;ANKRD2;NKX2-5 | | | | | | |
| Cellular component: actin cytoskeleton GO:0015629

GENE DIFFERENCES(3)
DMTN;NPFFR2;SEPT9; | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.135825101 | 0.001094738 | 2.278829 | 8.776437 | 20 | 376 | |
| | A3: NSD1;HOXD13;MECP2;RBPJL;POLE;STAG2;NKAP;BHMGI;ELK1;POLA1;NOC2L;L3MBTL1;MBD1;MEIOB;MTA3;CTBP2;HCFC1;UXT;ATRX;SKI; | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.205459042 | 0.001967564 | 2.141409 | 9.806628 | 21 | | |
| Cellular component: actin cytoskeleton GO:0015629

GENE DIFFERENCES(3)
DMTN;NPFFR2;SEPT9; | A4: MBD1; NONO; ARX; TWIST2; TP73; NKX2-5; GATA6; AIRE; BHMGI; POLE; CTBP2; HMGA1; MTA3; HHEX; RBMX; NSD1; MEIOB; NOC2L; RBPJL; UXT; POLA1; | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.615349 | 0.441655 | 1.309053 | 3.819555 | 5 | 194 | GO:0015629 |
| | A1: LSP1;DMTN;NPFFR2;SEPT9;KLHL17 | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| Cellular component: actin cytoskeleton GO:0015629

GENE DIFFERENCES(3)
DMTN;NPFFR2;SEPT9; | 0.933354349 | 0.823461765 | 0.750613 | 5.328978 | 4 | 194 | |
| | A2: SEPT9;ACTL7B;DMTN;CORO6 | | | | | | |
| | A3: LSP1;DMTN;NPFFR2;SEPT9;KLHL17; | | | | | | |
| | adjP | rawP | Fold Change | Expected | Observed | Category | Term |
| | 0.767738177 | 0.648406497 | 1.185817 | 5.059803 | 6 | | |
| Cellular component: actin cytoskeleton GO:0015629

GENE DIFFERENCES(3)
DMTN;NPFFR2;SEPT9; | A4: CORO6;LSP1;FLNA;ACTL7B;KLHL17;ARHGAP6 | | | | | | |

The accumulated genes from the table are MSH5 (Frequency = 3); ASZ1 (Frequency = 2); L3MBTL1 (Frequency = 2); DDX4; RNF212; MSH5-SAPCD1; MSH3; TDG; OR10H2; OR2B11; OR2L13; LMX1A; MECP2; TOX; TCF7; HOXD13; STAG2; NKAP; ELK1; HCFC1; ATRX; SKI; NKX6-1; KAT2A; ANKRD2; CEBPB; DMTN; NPFFR2; and SEPT9.

ATRX alpha thalassemia/mental retardation syndrome X-linked. Has high confidence in the SFARI GENE score 1. This gene encodes a protein that is a member of the SWI/SNF family of chromatin remodeling proteins because it has an ATPase/helicase domain. Alpha-thalassemia (ATRX) syndrome frequently coexists with X-linked mental retardation (XLMR) syndrome, which is linked to mutations in this gene. It has been demonstrated that these mutations result in a variety of alterations to the DNA methylation pattern, which may establish a connection between chromatin remodeling, DNA methylation, and gene expression throughout developmental processes. There have been several reports of alternatively spliced transcript variants that encode different isoforms [32]. **HCFC1**, host cell factor C1, has a syndromic SFARI GENE score of S. involved in cell cycle regulation [32].

MECP2 Methyl CpG binding protein 2. Has high confidence and is syndromic in the SFARI GENE score 1S. Rett syndrome is an autistic spectrum illness caused by mutations in the MECP2 gene. While some research has linked autism to uncommon changes in the MECP2 gene, other investigations have found no alterations in individuals with autism. EGR2 and MECP2 seem to be

able to control each other's expression [36]. The encoded protein exhibits transcriptional repressor activity that is reliant on methylation. Additionally, it has a role in controlling RNA splicing [32]. **SKI** proto-oncogene. Has a high confidence in the SFARI GENE Score 1. The avian sarcoma viral (v-ski) oncogene's nuclear proto-oncogene protein homolog is encoded by this gene. It suppresses TGF-beta signaling and may be involved in muscle differentiation and neural tube formation [37, 32].

NPFFR2 neuropeptide FF receptor 2. Considered as suggestive evidence in the SFARI GENE score3. A member of the G-protein-coupled neuropeptide receptor subfamily is encoded by this gene. The neuropeptides A-18-amide (NPAF) and F-8-amide (NPFF) activate this protein, which may be involved in opioid system regulation and pain modulation [32]. The paralogs of these genes KAT2B (KAT2A), STAG1 (STAG2), TCFL7L2 (TCFL7), LMX1B (LMX1A), and NKX2-2 (NKX6-1) were appeared.

Female-specific GO entries were selected according to p-value and unique appearance in the female cohort.

Table (5) GO Enrichment Results for DMR-Anchored Promoter Region Genes for Female Cohort A3.

| Term | Category | Observed | Expected | Fold Change | rawP | adjP | Term Name | Gene List |
|------------|----------|----------|----------|-------------|----------|----------|--|--|
| GO:0035240 | 9 | 3 | 0.210074 | 14.28068 | 0.002246 | 0.180645 | Molecular function: dopamine binding | DRD4; GPR143; DRD5 |
| GO:0046811 | 2 | 2 | 0.046683 | 42.84215 | 0.003034 | 0.214853 | Molecular function: histone deacetylase inhibitor activity | UCN; SKI |
| GO:0042417 | 11 | 3 | 0.256757 | 11.6842 | 0.003591 | 0.223188 | Biological process: dopamine metabolic process | DRD4; ATP7A; HPRT1 |
| GO:0048011 | 272 | 15 | 6.348912 | 2.36261 | 0.004139 | 0.226994 | Biological process: neurotrophin TRK receptor signaling pathway | ARHGEF3; FGF6; ABR; IRA K1; ELK1; ARHGEF18; NRG1; FGD1; RPS6KA3; FOXO4; NTRK2; FGF16; NDN; BCL2 L11; AP2A2 |
| GO:0031547 | 3 | 2 | 0.070025 | 28.56123 | 0.00498 | 0.226994 | Biological process: brain-derived neurotrophic factor receptor signaling pathway | NTRK2; SLC9A6 |
| GO:0045964 | 3 | 2 | 0.070025 | 28.56123 | 0.00498 | 0.226994 | Biological process: positive regulation of dopamine metabolic process | HPRT1; MAOB |
| GO:0043984 | 13 | 3 | 0.303441 | 9.886601 | 0.005339 | 0.237498 | Biological process: histone H4-K16 acetylation | HCFC1; MSL3; OGT |
| GO:0043981 | 15 | 3 | 0.350124 | 8.568393 | 0.007519 | 0.2572 | Biological process: histone H4-K5 acetylation | JADE3; OGT; HCFC1 |
| GO:0043982 | 15 | 3 | 0.350124 | 8.568393 | 0.007519 | 0.2572 | Biological process: histone H4-K8 acetylation | JADE3; HCFC1; OGT |

Female gene cohort A3: (2) DRD4; GPR143; DRD5; SKI; UCN; ATP7A; HPRT1; ARHGEF3; FGF6; ABR; ELK1; ARHGEF18; NRG1; RPS6KA3; FOXO4; FGF16; AP2A2; SLC9A6; MAOB; MSL3; OGT; JADE3; HCFC1.

RPS6KA3 Ribosomal protein S6 kinase, 90 kDa, polypeptide 3, has a strong candidate, syndromic in the SFARI GENE score 2S. A member of the RSK (ribosomal S6 kinase) family of serine/threonine kinases is encoded by this gene. Members of the mitogen-activated kinase (MAPK) signaling pathway are among the substrates phosphorylated by this kinase, which has two distinct kinase catalytic domains. This protein's function has been linked to regulating cell division and proliferation. X-linked 19 (MRX19) [MIM:300844], a non-syndromic form of mild-to-moderate mental retardation, and Coffin-Lowry syndrome (CLS) [MIM:303600], an X-linked

mental retardation linked to facial and digital dysmorphisms, progressive skeletal malformations, growth retardation, hearing impairment, and paroxysmal movement disorders [32].

The paralog of these genes also appears in female cohorts. DRD1,2,3 (DRD4,5), FGF14 (FGF6), ARHGEF2 (ARHGEF3).

Male specific GO entries were selected according to p-value and unique appearance in this cohort.

Table (6) GO Enrichment Results for DMR-Anchored Promoter Region Genes for male cohort A2.

| Term | Category | Observed | Expected | Fold Change | rawP | adjP | Term Name | Gene List |
|------------|----------|----------|----------|-------------|-------------|-------------|---|----------------------------|
| GO:0021524 | 3 | 2 | 0.08 | 24.3 | 0.00678479 | 0 | Biological process: visceral motor neuron differentiation | TBX20; ISL2 |
| GO:0004459 | 5 | 2 | 0.14 | 14.6 | 0.01374778 | 0 | Molecular function: L-lactate dehydrogenase activity | LDHAL6B; LDHAL6A |
| GO:0048312 | 5 | 2 | 0.14 | 14.6 | 0.01374778 | 0 | Biological process: intracellular distribution of mitochondria | BHLHA15; KAT2A; |
| GO:0072593 | 32 | 4 | 0.88 | 4.55 | 0.01528708 | 0 | Biological process: reactive oxygen species metabolic process | NDUFA13; GLS2; SFTPD; PDK4 |
| GO:0044295 | 12 | 2 | 0.33 | 6.07 | 0.05263846 | 0 | Cellular component: axonal growth cone | AATK; FKBP4 |
| GO:0048843 | 9 | 2 | 0.247221 | 8.089928 | 0.033538982 | 0.427877108 | Biological process: negative regulation of axon extension involved in axon guidance | SEMA3F; WNT3 |
| GO:0030517 | 12 | 2 | 0.329627 | 6.067464 | 0.052638 | 0.427877108 | Biological process: negative regulation of axon extension | NTN1; AATK |

Male cohort genes A2: TBX20, ISL2, LDHAL6B, KAT2A, GLS2, SFTPD, PDK4, (2) AATK, SEMA3F, WNT3, NTN1. Male cohorts contain only the paralog of these genes: WNT1 (WNT3), TBX22 (TBX20), PDK2 (PDK4).

A1 cohort specific GO entries were selected according to p-value and unique appearance in this cohort.

Table (7) GO Enrichment Results for DMR-Anchored Promoter Region Genes for Cohort A1.

| Term | Category | Observed | Expected | Fold Change | rawP | adjP | Term Name | Gene List |
|------------|----------|----------|----------|-------------|----------|----------|--|-----------|
| GO:0003908 | 1 | 1 | 0.019688 | 50.79236 | 0.038277 | 0.399097 | Molecular function: methylated-DNA-[protein]-cysteine S-methyltransferase activity | MGMT |

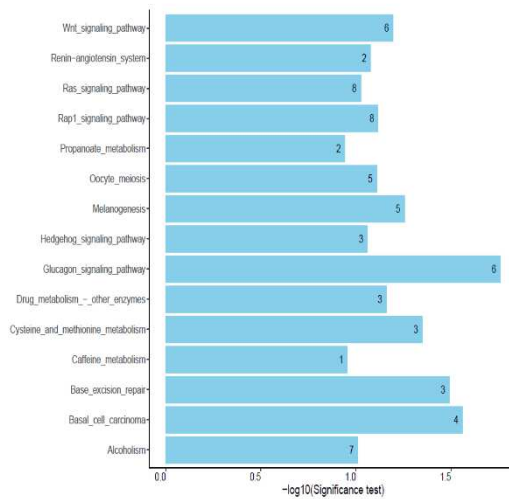
A1 cohort gene was **MGMT**, O-6-Methylguanine-DNA Methyltransferase, which is a protein coding gene. Enzymes known as protein arginine methyltransferases facilitate the movement of methyl groups from S-adenosylmethionine (SAM) to the arginine residues on histones and other proteins. Certain tumors emerge as a result of this methylation's dysregulation [38, 103, 104, 105, 106].

3.5 KEGG Enrichment Analysis of DMR-Anchored Promoter Genes

Based on DMR genomic distribution, functional enrichment analysis was performed for genes overlapping with promoter regions (1.5K upstream to 0.5K downstream of TSS). KEGG pathway significance enrichment identifies primary biochemical and signal transduction pathways of DMR-related genes [82]. A bar chart of enriched KEGG pathways is shown below:

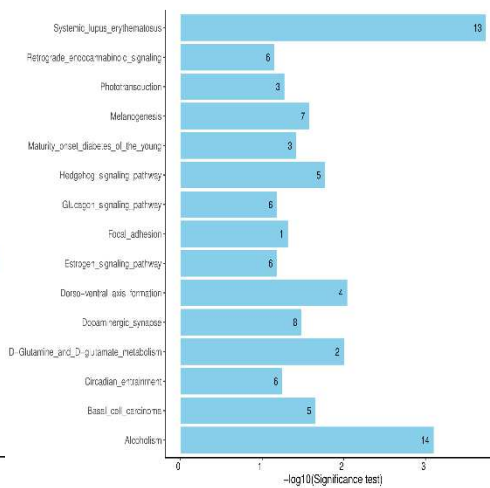
(A)

the most enriched pathways in Sample_1_2_3_4_5_6_7_8_9_10-vs-Sample_11_1:



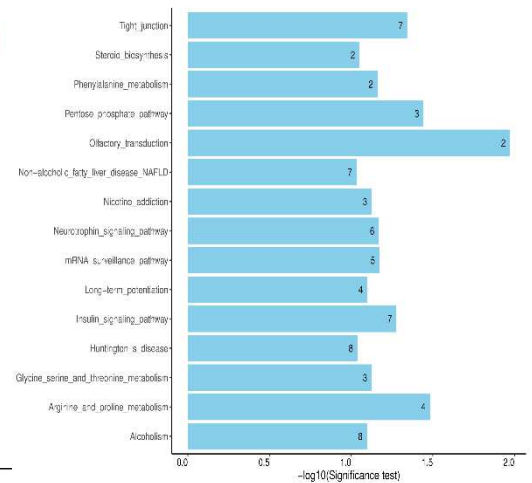
A1.

the most enriched pathways in SampleS_1_2_3_4_5_7_8-vs-Sample11



A2.

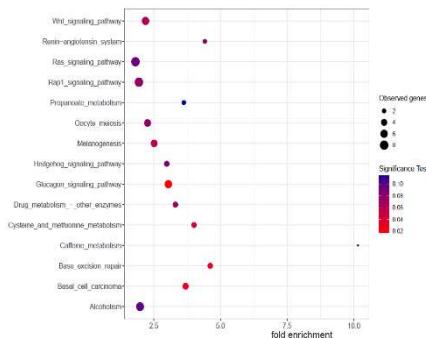
the most enriched pathways in SampleS_6_9_10-vs-Sample12



A3.

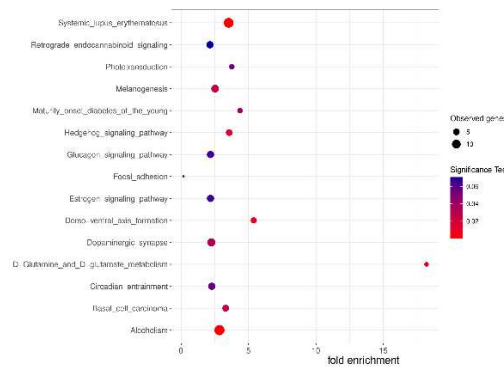
(B)

the most enriched pathways in Sample_1_2_3_4_5_6_7_8_9_10-vs-Sample_11_1:



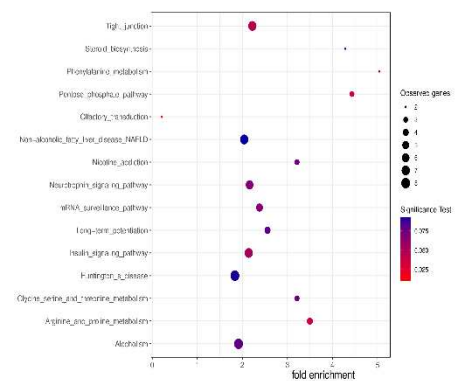
A1.

the most enriched pathways in SampleS_1_2_3_4_5_7_8-vs-Sample11



A2.

the most enriched pathways in SampleS_6_9_10-vs-Sample12



A3.

Figure (6): (A) Bar Chart of Enriched KEGG Pathways for A1, A2, and A3. Horizontal bar chart of KEGG pathway classifications: Y-axis = enriched KEGG pathway; X-axis = $-\log_{10}(\text{adjP})$. Longer bars indicate more significant statistical differences. Numbers on the right of bars denote the number of enriched genes and enrichment fold. (B) Scatter Plot of Enriched KEGG Pathways. Note: Y-axis: pathway name; X-axis: fold change. Dot size indicates the number of DMR-related genes, and color corresponds to Q-value range (redder = more significant).

Methylation in the promoter region is crucial in determining whether the gene is expressed or silenced. Because the KEGG pathways were highly divergent between the cohorts, the selection criteria for the KEGGs were based on their importance in the development of autism in individuals, as supported by prior studies, and on the biological and molecular functions of the related genes. The pathways are presented below:

Table (8) KEGG Enrichment List of DMR-Anchored Promoter Region-Related Genes for all cohorts.

| | Term name | Term | Cat
ego
ry | Obs
erve
d | rawP | adjP | List of genes |
|----|---|----------|------------------|------------------|-----------------|-----------------|---|
| A1 | WNT signaling pathway | hsa04310 | 139 | 6 | 0.063662 | 1 | WNT10A;CTBP2;WIF1;PRKACG;
;APC;TCF7 |
| A2 | Gene differences
APC;TCF7;WNT10B;PORC
N;WNT3 | | 139 | 7 | 0.1173988 | 0.9046853
76 | TCF7;PRKACG;WNT10A;WNT10
;B;CTBP2;WIF1;WNT3 |
| A3 | | | | 3 | 1 | 1 | ;PORCN;GPC4;CTBP2 |
| A4 | | | | 6 | 0.189689 | 0.887323 | CTBP2;GPC4;WIF1;FZD9;PRKA
CG;WNT10A |
| A1 | Hedgehog signaling pathway | hsa04340 | 51 | 3 | 0.086767 | 1 | ;HHIP;WNT10A;PRKACG |
| A2 | GENEDIFFERENCE
WNT10B;HHIP;WNT3; | | | 5 | 0.0169677
8 | 0.7703462 | WNT10B;PRKACG;HHIP;WNT3;
;WNT10A |
| A3 | - | - | - | - | - | - | - |
| A4 | | | | 2 | 0.39162 | 0.904781 | WNT10A;PRKACG |
| A1 | Hippo signaling pathway | hsa04390 | 153 | 4 | 0.549757 | 1 | ;TCF7;RASSF1;APC;WNT10A |
| A2 | GENE DIFFERENCES
TCF7; WNT10B;WNT3;
CALML3;RALA;FARP2;PFN
3;LAT;CALML5;PDGFA;AP
C | | 152 | 5 | 0.6166231
15 | 1 | TCF7;WNT10A;TP73;WNT10B;W
;NT3 |
| A3 | | | | 1 | 0.2718956
76 | 0.8719808
81 | CALML3;RALA;FARP2;PFN3;LA
;T;CALML5;PDGFA |
| A4 | | | | 5 | 0.604752 | 0.987456 | FZD9;TP73;RASSF1;WNT10A;D
LG3 |
| A1 | Ras signaling pathway | hsa04014 | 224 | 8 | 0.0942337
34 | 1 | RASSF1;LAT;FGF16;GNG4;PDGF
;A;CALML3;PRKACG;CALML5 |
| A2 | GENE DIFFERENCES
GNG4;RALA;CALML3;FOX
O4; ;ELK1; FGF6; PIK3CG | | | 7 | 0.6804083
37 | 1 | CALML5;LAT;GNG4;RALA;CAL
;ML3;PRKACG;PDGFA |
| A3 | | | | 7 | 0.3777607
78 | 0.9129670
01 | PIK3CG;FOXO4;CALML5;LAT;E
;LK1;FGF16;FGF6 |
| A4 | | | | 8 | 0.904781 | 0.396696 | RASGRP1;FGF16;RASSF1;CALM
L5;PDGFA;LAT;PRKACG;PIK3R5 |
| A1 | Rap1 signaling pathway | hsa04015 | 209 | 8 | 0.0762066
9 | 1 | CALML5;PDGFA;CALML3;FGF1
;6;PFN3;LAT;FARP2;LCP2 |
| A2 | GENE DIFFERENCES
CALML3 ;FGF6; RALA;
PIK3CG; LCP2 | | | 7 | 0.5254968
79 | 1 | CALML3;RALA;FARP2;PFN3;LA
;T;CALML5;PDGFA |
| A3 | | | | 6 | 0.6421325
16 | 1 | PIK3CG;PFN3;CALML5;LAT;FG
;F16;FGF6 |

| | | | | | | | |
|----|--|----------|-----|---|-------------|-------------|--|
| A4 | | | | 9 | 0.1311 | 0.81299 | FGF16;GNAS;PDGFA;CALML5;LAT;PFN1;FARP2;PFN3;PIK3R5 |
| A1 | cGMP-PKG signaling pathway | hsa04022 | 166 | 4 | 0.574289703 | 1 | MYL9;SLC25A31;CALML5;CALML3 |
| A2 | GENE DIFFERENCES
PIK3CG;SLC25A5;MYL9;IRS4; ADRB3;CALML3 | | | 5 | 0.809546458 | 1 | CALML5;CACNA1F;SLC25A31;CALML3;ADRA2B |
| A3 | | | | 7 | 0.121504062 | 0.826674845 | ATP2B3;ADRB3;CALML5;PIK3CG;SLC25A5;MYL9;IRS4 |
| A4 | | | | 7 | 0.217544 | 0.887323 | ATP2B3;ADRA1D;ADRA2B;SLC25A31;CACNA1F;PIK3Rp5;CALML5 |
| A1 | AMPK signaling pathway | hsa04152 | 123 | 1 | 0.73616974 | 1 | ;CPT1C |
| A2 | GENE DIFFERENCES
PIK3CG;IRS4; CPT1C | | | 1 | 0.26866861 | 1 | ;CPT1C |
| A3 | | | 2 | 1 | 1 | | ;PIK3CG;IRS4 |
| A4 | | | | | 0.382136 | 0.904781 | PIK3R5 |
| A1 | cAMP signaling pathway | hsa04024 | 199 | 6 | 0.29921363 | 1 | DRD5;MYL9;HHIP;CALML5;PRKACG;CALML3 |
| A2 | GENE DIFFERENCE | | | 6 | 0.826035025 | 1 | HHIP;CALML3;CACNA1F;CALML5;PRKACG;GRIA2 |
| A3 | HHIP;CALML3; GRIA2; DRD5;MYL9; PIK3CG | | | 5 | 0.811945342 | 1 | PIK3CG;DRD5;MYL9;CALML5;ATP2B3 |
| A4 | | | | 6 | 0.654152 | 0.987456 | CALML5;GNAS;ATP2B3;PIK3R5;CACNA1F;PRKACG |
| A1 | Neurotrophin signaling pathway | hsa04722 | 119 | 2 | 1 | 1 | ;CALML3;CALML5 |
| A2 | GENE DIFFERENCE
CALML3;PIK3CG; RPS6KA6;RPS6KA3 | | | 3 | 1 | 1 | ;CALML3;TP73;CALML5 |
| A3 | | | | 6 | 0.067804813 | 0.826674845 | IRAK1;NTRK2;PIK3CG;CALML5;RPS6KA6;RPS6KA3 |
| A4 | | | | 5 | 0.250219 | 0.887323 | NTRK2;TP73;IRAK1;CALML5;PIK3R5 |
| A1 | Estrogen signaling pathway | hsa04915 | 100 | 4 | 0.143049868 | 1 | CALML5;PRKACG;FKBP4;CALML3 |
| A2 | GENE DIFFERENCES
CALML3;PIK3CG; KCNJ3 | | | 6 | 0.065802458 | 0.904685376 | PRKACG;KCNJ3;FKBP4;HSPA6;CALML5;CALML3 |
| A3 | | | | 4 | 0.304728875 | 0.908167489 | FKBP4;HSPA6;CALML5;PIK3CG |
| A4 | | | | 6 | 0.054212 | 0.710776 | FKBP4;GNAS;PRKACG;CALML5;PIK3R5;HSPA6 |
| A1 | Dopaminergic synapse | hsa04728 | 130 | 5 | 0.193431143 | 1 | DRD5;GNG4;PRKACG;CALML5;CALML3 |
| A2 | GENE DIFFERENCES
DRD5;GNG4;CALML3; KCNJ3;DRD4;GRIA2;CACNA1B;ATP2B3;ADRB3;PIK3CG;SLC25A5;MYL9;IRS4 | | | 8 | 0.03297406 | 0.904685376 | GNG4;CALML3;PRKACG;KCNJ3;DRD4;CALML5;GRIA2;CACNA1B |
| A3 | | | | 6 | 0.13558563 | 0.826674845 | ATP2B3;ADRB3;CALML5;PIK3CG;SLC25A5;MYL9;IRS4 |

| | | | | | | | |
|----|--|----------|-----|---|-------------|-------------|---|
| A4 | | | | 4 | 0.586724 | 0.982034 | PRKACG;GNAS;MAOA;CALML5 |
| A1 | GABAergic synapse | hsa04727 | 89 | 2 | 0.696106598 | 1 | ;PRKACG;GNG4 |
| A2 | GENE DIFFERENCES
CACNA1B; GNG4;
GABRQ;GABRE | | | 5 | 0.108212612 | 0.904685376 | GNG4;CACNA1B;CACNA1F;PR
;KACG;GLS2 |
| A3 | | | | 3 | 0.469978821 | 0.968882703 | ;GABRQ;CACNA1B;GABRE |
| A4 | | | | 3 | 0.508958 | 0.925708 | GLS2;PRKACG;CACNA1F |
| A1 | Cholinergic synapse | hsa04725 | 112 | 3 | 0.489631738 | 1 | ;KCNJ12;GNG4;PRKACG |
| A2 | GENE DIFFERENCE
KCNJ12;GNG4; KCNJ3;
CACNA1B;PIK3CG | | | 5 | 0.246771979 | 1 | GNG4;CACNA1F;PRKACG;KCN
;J3;CACNA1B |
| A3 | | | | 2 | 1 | 1 | ;CACNA1B;PIK3CG |
| A4 | | | | | 0.769075 | 0.998825 | PIK3R5;CACNA1F;PRKACG |
| A1 | Circadian entrainment | hsa04713 | 96 | 1 | 0.129244172 | 1 | GNG4;PRKACG;CALML5;CALM
;L3 |
| A2 | GENE DIFFERENCES
GNG4; ;CALML3;KCNJ3;
GRIA2 | | | 6 | 0.056687529 | 0.904685376 | GNG4;CALML3;KCNJ3;PRKAC
;G;GRIA2;CALML5 |
| A3 | | | | 1 | 0.729181067 | 1 | ;CALML5 |
| A4 | | | | 5 | 0.250219 | 0.887323 | PRKACG;CACNA1G;GNAS;CAL
ML5 |
| A1 | Base excision repair | hsa03410 | 33 | 3 | 0.032128485 | 1 | ;TDG;LIG1;POLE |
| A2 | GENE DIFFERENCES
TDG | | | 3 | 0.07117262 | 0.904685376 | ;LIG1;TDG;POLE |
| A3 | | | | 1 | 0.543812901 | 1 | ;POLE |
| A4 | | | | 2 | 0.223417 | 0.887323 | POLE;LIG1 |
| A1 | Nucleotide excision repair | hsa03420 | 47 | 2 | 0.244536206 | 1 | ;LIG1;POLE |
| A2 | GENE DIFFERENCES
A4: 3 genes while A1,A2,A3
have 2 genes | | | | 0.378440504 | 1 | ;POLE;LIG1 |
| A3 | | | | 2 | 0.308201636 | 0.908167489 | ;POLE;CETN2 |
| A4 | | | | 3 | 0.13435 | 0.81299 | LIG1;POLE;CETN2 |
| A1 | Mismatch repair | hsa03430 | 23 | 1 | 0.37390336 | 1 | LIG1 |
| A2 | Gene differences
MSH3 | | | 1 | 0.478290061 | 1 | LIG1 |
| A3 | | | | 1 | 0.425388678 | 0.968346096 | MSH3 |
| A4 | | | | | 0.461097 | 0.909726 | LIG1 |
| A1 | Glucagon signaling pathway | hsa04922 | 100 | 6 | 0.017373528 | 1 | CALML3;PRKACG;CALML5;CP
;T1C;LDHAL6B;LDHAL6A |
| A2 | Gene difference | | | 6 | 0.065802458 | 0.904685376 | LDHAL6B;PRKACG;CPT1C;CAL
;ML5;LDHAL6A;CALML3 |

| | | | | | | | |
|----|--|----------|-----|----|-------------|-------------|---|
| | CALML3;PDHA1;CPT1C;LDHAL6B | | | | | | |
| A3 | | | | 3 | 0.511009397 | 0.982221444 | ;PHKA2;PDHA1;CALML5 |
| A4 | | | | 7 | 0.01992 | 0.672552 | PHKA2;PRKACG;PHKA1;GNAS;LDHAL6A;PGAM2;CALML5 |
| A1 | Alcoholism | hsa05034 | 179 | 7 | 0.097364115 | | H2AFB1;H3F3C;CALML3;HIST1;H3E;CALML5;GNG4;HIST1H4G |
| A2 | GENE DIFFERENCES
;CALML3; HIST1H4G;
H2AFB1; HIST1H2AA;
HIST1H2BA; MAOB; ;GNG4 | | | 14 | 0.000793287 | 0.090038098 | H2AFB1;HIST1H2AA;HIST1H4G;HIST1H4I;HIST1H2BK;HIST3H3;CALML5;HIST1H2AH;HIST1H2BA;CALML3;H3F3C;HIST1H3E;GNG4;H2BFWT |
| A3 | | | | 8 | 0.079913396 | 0.826674845 | H2BFWT;CALML5;ARAF;NTRK2;HIST3H3;H3F3C;MAOA;MAOB |
| A4 | | | | 16 | 4.81E-05 | 0.011363 | H2AFB3;HIST1H4I;NTRK2;H3F3C;HIST1H2AH;H2BFM;HIST1H2BK;HIST3H3;H2AFB2;HDAC6;GNAS;ARAF;HIST1H3E;MAOA;CALML5;H2BFWT |
| A1 | Systemic lupus erythematosus | hsa05322 | 134 | 5 | 0.199314367 | | HIST1H3E;H3F3C;H2AFB1;C8G;HIST1H4G |
| A2 | GENE DIFFERENCES
H2AFB1; HIST1H4G;
HIST1H2BA; HLA-DPA1;
HIST1H2AA; | | | | 0.000182287 | 0.041379234 | H3F3C;HIST1H3E;HIST1H2BA;H2BFWT;HLA-DPA1;HIST1H4I;HIST1H2BK;HIST3H3;HIST1H2AH;HIST1H2AA;HIST1H4G;H2AFB1;C8G |
| A3 | | | | 4 | 0.560878544 | 1 | ;H2BFWT;C8G;H3F3C;HIST3H3 |
| A4 | | | | 11 | 0.001309 | 0.154474 | HIST1H4I;H3F3C;HIST3H3;H2AFB3;H2AFB2;HIST1H2AH;H2BFM;HIST1H2BK;C8G;H2BFWT;HIST1H3E |
| A1 | Toll-like receptor signaling pathway | hsa04620 | 105 | 1 | 0.726637302 | | ;IRF7 |
| A2 | GENE DIFFERENCES
TOLLIP; PIK3CG | | | 2 | 1 | 1 | ;IRF7;TOLLIP |
| A3 | | | | 2 | 1 | 1 | ;PIK3CG;IRAK1 |
| A4 | | | | 3 | 0.756243 | 0.997058 | IRF7;PIK3R5;IRAK1 |
| A1 | ABC_transporters | hsa02010 | 44 | 2 | 0.223087751 | | ;ABCA1;ABCC9 |
| A2 | GENE DIFFERENCES
ABCA3; ABCA1; | | | 3 | 0.130943651 | 0.904685376 | ;ABCD1;ABCA3;ABCC9 |
| A3 | | | | 2 | 0.282819152 | 0.880324404 | ;ABCD1;ABCA1 |
| A4 | | | | | 0.327272 | 0.904781 | ABCD1;ABCC9 |
| A1 | Propanoate metabolism | hsa00640 | 28 | 2 | 0.113962585 | | ;LDHAL6B;LDHAL6A |
| A2 | GENE DIFFERENCES
LDHAL6B | | | 2 | 0.19114358 | 1 | ;LDHAL6A;LDHAL6B |
| A3 | | - | - | - | - | - | - |
| A4 | | | | 1 | 0.526206 | 0.933718 | LDHAL6A |

| | | | | | | | |
|----|--|----------|-----|---|-----------------|-----------------|--|
| A1 | Tight junction | hsa04530 | 153 | 3 | 0.7512907
46 | | ;CLDN4;CLDN10;MYL9 |
| A2 | Gene differences
CLDN10;MYL9;
CASK; ;SHROOM2 | | | 3 | 1 | 1 | ;CLDN4;CLDN3;CLDN10 |
| A3 | | | | 7 | 0.0453832 | 0.8266748
45 | CLDN23;CASK;CLDN4;SHROO
;M4;CLDN10;MYL9;SHROOM2 |
| A4 | | | | 4 | 0.782977 | 0.998825 | SHROOM4;CLDN4;CLDN23;CL
DN3 |

Accumulated genes from the table after filtering A4 from all cohorts; the number before the gene means its frequency of appearance:

(2)APC; (2)TCF7; (3)WNT10B; (1)PORCN; (3)WNT3; (2)HHIP; (11)CALML3; (10)PIK3CG; (6)GNG4; (3)RALA; (1)FARP2; (1)LCP2; (2)SLC25A5; (4)MYL9; (3)IRS4; (3)ADRB3; (2)CPT1C; (3)GRIA2; (2)DRD5; (1)RPS6KA6; (1)RPS6KA3; (4)KCNJ3; (1)DRD4; (2)CACNA1B; (1)ATP2B3; (1)GABRQ; (1)GABRE; (1)KCNJ12; (1)TDG; (1)MSH3; (1)PDHA1; (1)LDHAL6B; (2)H2AFB1; (2)HIST1H2AA; (2)HIST1H2BA; (1)MAOB; (2)HIST1H4G; (1)HLA-DPA1; (1)TOLLIP; (1)ABCA3; (1)ABCA1; (1)CLDN10; (1)CASK; (1)SHROOM2; (1)PFN3; (1)LAT; (1)CALML5; (1)PDGFA; (1)FOXO4; (2)FGF6; (2)IRS1; (1)ELK1.

CASK is a calcium/calmodulin-dependent serine protein kinase and has high confidence in the SFARI Gene Score 1. The encoded protein belongs to the membrane-associated guanylate kinase (MAGUK) protein family. These proteins are scaffold proteins, and the brain's synapses include the encoded protein. FG syndrome 4 (OMIM 300422) and mental retardation and microcephaly with pontine and cerebellar hypoplasia (OMIM 300749) are linked to mutations in this gene [32]. **CACNA1B** calcium voltage-gated channel subunit alpha1 B, is a strong candidate in the SFARI Gene Score 2. This gene encodes the pore-forming subunit of an N-type voltage-dependent calcium channel, which regulates the release of neurotransmitters from neurons. The high-voltage-activated channel is created when the encoded protein combines with beta, delta, and alpha-2 subunits. This channel is insensitive to dihydropyridines but responsive to omega-conotoxin-GVIA and omega-agatoxin-IIIa [32].

MAOB (monoamine oxidase B) is a strong candidate in the SFARI Gene Score2. This gene codes for a protein that is a member of the flavin monoamine oxidase family. It is an enzyme found in the outer membrane of the mitochondria. It is involved in the metabolism of neuroactive and vasoactive amines in the central nervous system and peripheral tissues, and it catalyzes the oxidative deamination of biogenic and xenobiotic amines. Benzylamine and phenylethylamine are preferentially broken down by this protein [32]. **RALA** (RAS like proto-oncogene) is considered syndromic in the SFARI Gene score S. Gene expression, cell migration, cell proliferation, neoplastic transformation, and membrane trafficking are among the biological processes in which this multifunctional GTPase is engaged [32].

GRIA2 (glutamate ionotropic receptor AMPA type subunit 2) has a high confidence in the SFARI Gene Score 1. This gene produces a glutamate receptor that serves as a ligand-gated ion channel in the central nervous system and is crucial for excitatory synaptic transmission. In the central nervous system, L-glutamate functions as an excitatory neurotransmitter at several synapses. When the excitatory neurotransmitter L-glutamate is bound, a conformational shift occurs that opens the cation channel, converting the chemical signal into an electrical impulse. After that, the receptor quickly becomes desensitized and goes into a brief inactive state even in

the presence of a bound agonist. The glutamate receptors that are responsive to alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) include this gene product [32].

PDHA1 (pyruvate dehydrogenase E1 subunit alpha 1). It is Considered as suggestive evidence in SFARI Gene Score 3. The nuclear-encoded mitochondrial multienzyme complex known as the pyruvate dehydrogenase (PDH) complex serves as the main connection between glycolysis and the tricarboxylic acid (TCA) cycle. It catalyzes the complete conversion of pyruvate to acetyl-CoA and CO₂. Pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2), and lipoamide dehydrogenase (E3) are the three enzyme components that make up the PDH complex in multiple copies. Two alpha and two beta subunits make up the heterotetramer that is the E1 enzyme. This gene is essential for the PDH complex's operation and codes for the E1 alpha 1 subunit, which contains the E1 active site. X-linked Leigh syndrome and pyruvate dehydrogenase E1-alpha impairment are connected to mutations in this gene [32].

PIK3CG (phosphoinositide-3-kinase, catalytic, gamma polypeptide), which is classified as a Strong Candidate in the SFARI Gene Score 2. The protein that is encoded exhibits phosphatidylinositol 3-kinase activity [32]. Additionally, this list includes the second appearance of RPS6KA3.

Furthermore, paralogs of these genes were identified, **ATP2B2 (ATP2B3), ADRB2 (ADRB3), FGF13, FGF14 (FGF6, FGF1), KCNJ10, KCNJ15 (KCNJ3, KCNJ12), PRKCA, PRKCB (PRKCG), HLA-DPB1 (HLA-DPA1), WNT1 (WNT3), SLC25A12, SLC25A39, SLC25A27 (SLC25A5).**

According to all results we obtained, there are about six high-confidence genes, eight strong candidates, five suggestive evidence and four syndromic. These genes were recorded in the SFARI Genes database. Also, we notice a high frequency for **CALML3** in ASD cohorts compared to **CALML5** in neurotypicals. This makes it the candidate gene of the recent study, followed by **WNT10B** as appearing in ASD towards **WNT10A** in neurotypicals. Also, **GNG4, MYL9** and **WNT3** frequently appeared in this study.

3.6 The overlap between the SFARI GENE database and the study genes within all cohorts

Cohort A1 comprises 1,121 promoter genes, of which 20 are SFARI-matched, and 4,234 GO genes, of which 273 are SFARI-matched. Cohort A2 comprises 1,564 promoter genes with 24 SFARI-matched genes, and 5,247 GO genes with 323 SFARI-matched genes. Cohort 3 comprises 1,329 promoter genes with 57 SFARI-matched genes and 4,195 GO genes with 297 SFARI-matched genes. Cohort A3 exhibits the highest Promoter-SFARI overlap at the absolute count level, although SFARI overlap is consistently larger in the GO lists than in the Promoter subsets across all cohorts.

Using Fisher's exact test (two-sided) under a consistent background gene universe, determine if the observed SFARI overlap is larger or smaller than would be predicted by chance. The universe is the union of GO genes across the three cohorts (Universe n = 6,441), It has 403 SFARI genes that can be tested for enrichment. In cohort A1, the Promoter list exhibits considerable depletion (OR = 0.234, p = 1.745e-14), whereas the GO list is not significantly enriched (OR = 1.101, p = 0.4162). In cohort A2, the Promoter list exhibits considerable depletion (OR = 0.185, p = 1.231e-23), whereas the GO list is not significantly enriched (OR = 0.913, p = 0.467). Group 3 exhibits statistically significant depletion (OR = 0.617, p = 0.0007257) in the Promoter list and statistically significant enrichment (OR = 1.538, p = 0.0001543) in the GO list. Overall, the GO list in Group

3 exhibits the strongest SFARI enrichment signal, although promoter lists for SFARI genes with the same background definition are underrepresented in all cohorts.

The remaining genes from all cohorts were filtered from the A4 gene cohorts. The identified genes between cohorts were ADAM8, ASIC2, ASIC4, ANO1, ANK1, APBB1IP, BHLHE40, ASCL4, ATP13A2, AFAP1L2, ANAPC2, GRM3 , GRM4 , GRM6, GABRB1 , GABRA5, CHRNA4, CHRNA2, KCNQ1 , KCNQ5, ASIC1 , ASIC2, CDH4 , CDH18, JAK2 , JAK3, GSTP1, AK3, ANK1 with unclassified genes like Chromosome Open reading Frame C19orf44, C1orf159, C9orf3 and C16orf74 and Long Intergenic Non-Protein Coding RNA LINC00299 , LINC01056 and FAM genes FAM19A5 , FAM171A1.

4 Discussion

The study is based on the ectodermal memory as a developmental mirror, depending on the fact that during fetal development, the promoters were activated to synthesize the epithelial tissue in the oral cavity and the nervous tissue. The aim is to identify the methylated genes in all ASD cohorts and define them as biomarkers.

These genes, **CACNA2D3**, **ATRX**, **MECP2**, **SKI**, **CASK**, **GRIA2** overlap with DMRs and have a high confidence score in SFARI.

The study revealed an aberrant methylation in the MECP2 promoter within the female cohort, while another study concluded that male brain samples from autistic individuals have higher levels of **MECP2** promoter methylation, suggesting a possible abnormal spreading of methylation into the MECP2 promoter [39, 87, 88, 89, 91]. Other studies suggested that MECP2 expressed in the fetal brain and in the blood of ASD was found to be female-specific [111, 115].

CACNA2D3 was screened in DMRs of the current study, consistent with the findings of previous research that indicates an interesting possibility in the investigation of autism spectrum disorder (ASD) is the **CACNA2D3** gene, which codes for a regulatory subunit of voltage-gated calcium channels. Although Numerous studies have concentrated on its genetic alterations, data also suggest that the disorder may be influenced by its epigenetic regulation, particularly through DNA methylation [40]. A prior study concluded that site-specific DNA methylation was substantially enriched in genes associated with ASD risk. **CACNA2D3** is a member of a larger group of calcium signaling genes whose methylation levels may affect neurodevelopmental outcomes [41].

GRIA2 is also among high confidence genes in the study with a differential methylation pattern in consistency with prior study which demonstrated that GRIA2 is a crucial component of the "Excitatory/Inhibitory (E/I) imbalance" paradigm of ASD. A reduced expression of GRIA2 (typically through epigenetic silencing) leads to the development of calcium-permeable AMPA receptors, which might result in neuronal hyperexcitability [42].

The genes **ADORA2A**, **CHRNA7**, **RPS6KA3**, **CACNA1B**, **MAOB**, **PIK3CG**, **LHX2**, **PITX1** were strong candidates in SFARI score overlaps with the DMRs in the current study.

Heterozygous variations in **CACNA1B** have been linked to ASD and other neurodevelopmental problems. Children with autism and developmental delays have been found to have genomic duplications affecting the **CACNA1B** gene on the 9q34 region. Dysfunction of **CACNA1B**, a voltage-gated calcium channel that aids in controlling neurotransmitter release, can result in abnormal synaptic connections and neuronal transmission, according to previous studies [32, 52, 53]. **PIK3CG** (Phosphoinositide-3-Kinase Catalytic Subunit Gamma) is a component of the PI3K signaling pathway that is involved in synapse formation, plasticity, and neurodevelopmental disorders. It is genetically linked to autism spectrum disorder (ASD). The

genetic vulnerability and heterogeneity of ASD have been linked to variations in PIK3CG, which have been identified as possible risk factors, according to prior studies [32, 54, 55].

MAOB is associated with autism spectrum disorder (ASD) due to genetic variations that impact dopamine metabolism and serotonin levels, especially affecting the behavioral severity in males. The **MAOB** gene, which is located on the X chromosome (Xp11.3), is thought to be a candidate for ASD susceptibility because of its function in controlling neurotransmitters. Certain polymorphisms (such as rs10521432 and rs1799836) are linked to an elevated risk of ASD and behavioral characteristics, according to prior studies [56,57,58].

The genes **NPFFR2**, **PDHA1**, **POLR2A**, **KDM2A**, **TCEAL1** were suggestive evidence in the SFARI gene score overlaps with the DMRs in the current study.

Studies showed that Functional screens of *de novo* mutations have shown **NPFFR2** as a putative, rare Loss-of-Function (LoF) gene variant linked to autism spectrum disorder (ASD). Along with other possible genes, the gene is a component of a system that controls neuropeptides in the central nervous system. Mutations in this gene may contribute to the neurodevelopmental and behavioral abnormalities associated with ASD [38, 60, 61].

A germline missense variant in the KDM2A gene was found in brain tissue from an ASD brain donor from the Harvard Brain Tissue Resource Center [32, 67]. *De novo* variants in the KDM2A gene have been found in ASD probands, including a *de novo* missense variant (p. Arg449Lys) in a proband from the Simons Simplex Collection [32, 62, 63, 64, 65]. Functional evaluation linked the p. Arg449Lys variant to ASD. Another study evaluated humanized flies carrying the Arg449Lys missense variant in *Drosophila* using a rescue-based approach. In a behavioral scenario, the Arg449Lys mutation demonstrated shorter copulation times than the humanized reference [32,60]. Recent studies concluded that autism spectrum disorder (ASD) is associated with mutations in and dysregulation of **PDHA1** (pyruvate dehydrogenase E1 subunit alpha 1 subunit) due to altered energy metabolism and mitochondrial malfunction. **PDHA1** variants, including splicing and frameshift mutations, have been found in ASD cases. Some children with autism exhibit reduced PDH activity, elevated plasma pyruvate, and mitochondrial inhibition [32, 68, 69].

Also, the study reveals the presence of ASD related syndromic genes **ZBTB18**, **HUWE1**, **HCFC1**, and **RALA**.

Prior studies demonstrated that **ZBTB18**, a transcription repressor essential for brain development, is found on chromosome 1q43-q44. ZBTB18-related disorder, a rare neurodevelopmental illness that frequently manifests as intellectual disability, speech delay, abnormalities of the corpus callosum, epilepsy, and occasionally autism spectrum disorder (ASD), is caused by *de novo* mutations [70, 71, 72].

According to a study done on a cohort of ten cases with *de novo* protein-altering variants in the **RALA** gene presented with a neurodevelopmental disorder marked by developmental delay/intellectual disability, delayed or absent speech, hypotonia, incapacity to walk, facial dysmorphism, short stature, decreased body weight, and abnormal brain MRI; 3/3 individuals in this cohort for whom information was available also presented with ASD [73, 32]. Later, a study done on a Japanese patient with autism spectrum disorder who presented with a Noonan syndrome-like phenotype also had the identical *de novo* missense mutation (c.73G>A; p.Val25Met) that was found in all three ASD instances in the first study [74,73].

Mental retardation and macrocephaly syndrome, commonly known as mental retardation syndromic X-linked Turner type (MRXST) [MIM:300706], is caused by defects in **HUWE1**. Macrocephaly and varied contractures are two characteristics linked to MRXST's clinical variability. Mental retardation X-linked type 17 (MRX17) [MIM:300705], also known as mental retardation X-linked type 31 (MRX31), is caused by a chromosomal microduplication involving **HUWE1** and HSD17B10 [32]. According to another study, a male ASD proband had a *de novo* missense mutation in HUWE1, although the proband's brother, who was less seriously afflicted, did not have it [75].

In the MRX3 pedigree, a variant in the 5'-UTR of the **HCFC1** gene was found to fully segregate with intellectual impairment. This variation, which was located in the transcription factor YY1's binding site, prevented YY1 binding and increased **HCFC1** expression. Follow-up exome sequencing of probands from unresolved families with intellectual impairment revealed two more distinct **HCFC1** mutations, one of which was demonstrated to segregate with ID according to a prior study [76, 86].

The current study showed the presence of the paralog of genes were recorded in SFARI database these were KAT2B (KAT2A), STAG1(STAG2), 2TCFL7L2(TCFL7), LMX1B(LMX1A), NKX2-2(NKX6-1), POLR2A(POLR2E), HOXB1(HOXA1), KLF7(KLF8), EBF3(EBF2), LZTS2(LZTS1), NR2F1(NR2F6), 2 PLCB1(PLCB2), CACNG2(CACNG7), ADRB2(ADRB1), DRD1,2,3 (DRD4,5), FGF14 (FGF6), ARHGEF2 (ARHGEF3), 2WNT1(WNT3), TBX22(TBX20), PDK2(PDK4), ATP2B2(ATP2B3), ADRB2(ADRB3), FGF13,14(FGF6,1), KCNJ10,15(KCNJ3,12), PRKCA, B(PRKCG), HLA-DPB1(HLA-DPA1), SLC25A12,39,27(SLC25A5).

Because of its involvement in the WNT/ β -catenin signaling pathway, which controls neurodevelopment, brain growth, and synaptic plasticity, the **WNT3** gene is linked to autism spectrum disorder (ASD). Individuals with ASD have been found to have rare variations in **WNT3**, and the pathophysiology of the illness is linked to its overactivation (elevated expression) in the prefrontal cortex [45, 46, 47].

The current study concluded that these genes, **CALML3**, **GNG4**, **MYL9**, **WNT10B**, **MSH5**, **ELK1** as biomarkers in DNA methylation in ASD, as they have not previously been recorded in SFARI before.

CALML3 (Calmodulin Like 3) is a Protein- Coding gene. Because of its capacity to bind calcium ions and control downstream signaling, CALML3 is recognized as a participant in the Alzheimer's disease pathway [38, 79, 80]. Involvement in calcineurin-mediated signaling and localization in the distal axon, two crucial regions impacted by synaptic loss in Alzheimer's, are predicted roles for CALML3 in the nervous system [77, 78].

Due to their involvement in neurotransmitter pathways, common genetic variants in the **GNG4** gene, which codes for G-protein subunit gamma-4, are linked to cognitive decline and the risk of Alzheimer's disease (AD). GNG4 is highly expressed in the hippocampus, however its expression declines with age and is downregulated in AD models [43, 44].

The WNT/ β -catenin signaling cascade is triggered by **WNT10B**, a secreted glycoprotein belonging to the WNT family, which regulates cell fate, proliferation, stemness, and pluripotency. Bone, adipocytes, skin, hair, muscle, the placenta, and the immune system are among the tissues in which WNT10B is involved. Numerous disorders, including osteoporosis, obesity, split-hand/foot malformation (SHFM), fibrosis, dental abnormalities, and cancer, are driven by aberrant WNT10B signaling [48].

The CD69 molecule is known to bind to myosin light chains 9 and 12a and 12b (MyL9/12). They produce "**MyL9** nets," which are frequently found in inflammatory sites and are essential for controlling the recruitment and retention of effector cells that express CD69 in inflammatory tissues [49,50]. The cytoplasmic protein encoded by the **IRS4** gene has several possible phosphorylation sites, including tyrosine and serine/threonine. The cytoplasmic signaling molecules with SH2 domains interact with the tyrosine-phosphorylated IRS4 protein. When the insulin receptor is stimulated, the insulin receptor tyrosine kinase phosphorylates IRS4. IRS4 serves as a link between intracellular signaling molecules with SH2 domains and other growth factor receptors, such as the insulin receptor, IGF1R, and FGFR1 [51]. The **MSH5** gene is involved in meiotic recombination and DNA mismatch repair. Making it easier for homologs to cross over during meiosis (by similarity) [38].

The study showed that about 2,511 genes were shared between A1, A2, and A3, which were not recorded in the SFARI database. All these were filtered by removing the similar genes in the A4 cohort. These genes were candidates to be ASD biomarkers. Including **ADAM8, ASIC2, ASIC4, ANO1, ANK1, APBB1IP, BHLHE40, ASCL4, ATP13A2, AFAP1L2, ANAPC2, GRM3, GRM4, GRM6, GABRB1, GABRA5, CHRNA4, CHRNA2, KCNQ1, KCNQ5, ASIC1, ASIC2, CDH4, CDH18, JAK2, JAK3, GSTP1, AK3, ANK1** with unclassified genes like Chromosome Open Reading Frame **C19orf44, C1orf159, C9orf3, C16orf74** and Long Intergenic Non-Protein Coding RNA **LINC00299, LINC01056** and FAM genes **FAM19A5, FAM171A1**.

The long intergenic noncoding RNA **LINC00299** has been identified as a multifunctional regulator in human biology and disease. Evidence indicates that LINC00299 is highly expressed in the brain and is necessary for typical neural development, and its disruption has been connected to neurodevelopmental disorders [81].

According to studies, the brain-specific secretory protein and chemokine **FAM19A5** (Family with sequence similarity 19 member A5), often referred to as **TAF α 5**, is essential for controlling synapse function, neuroinflammation, and nervous system development. It functions as a new adipokine connected to metabolic syndrome, obesity, and cardiovascular diseases, and is closely linked to neurological conditions such as Alzheimer's disease [26, 27, 28, 29, 30]. Limitations of the study: Because of the high costs of WGBS, the sample size was limited. Further studies should consider expanding brain banks supporting the ASD research worldwide so the application can be done directly on postmortem brain tissue, and then subsequent comparisons with buccal swabs results to avoid tissue-specific bias and any other complications in the results analysis. No database supported the methylated genes for ASD as a reference for researchers, and SFARI supported the mutation and copy number variants.

5 Conclusion

Decline towards oxidative stress genes was observed across the NER, BER and mismatch repair pathways in the ASD cohorts in the study. The clear appearance of cancer-related genes suggests a genetic overlap between ASD and cancer, enhancing the pleiotropic function of the genes. The female cohort showed aberrant methylation in the MECP2 promoter region, while other studies suggested male ASD. X-inactivation in females is protected by MECP2 activity. The aberrant methylation in the MECP2 is crucial for females to develop ASD, besides the genetic overload damaging the female protective effect. The same SFARI genes appeared in the study and the paralogs of these genes confirms the role of these genes in ASD, methylated or mutated. The pleiotropy of genes

and the high presence of the paralogs of SFARI genes in the results could be antagonistic pleiotropy, increasing ASD cases.

Evolution in human brains creates more paralogs for higher cognitive abilities, increasing the possibility of mutations. The epigenetic divergence plays a role in silencing one paralog and activating another irregularly, causing synaptic noise. The variations in the severity of ASD cases are further influenced by methylation caused by environmental factors. The methylated genes matrix was a combination of epigenetic regulators, transcriptional regulators, cell cycle regulators, synapse formation and transmission, neurotransmitter regulators, cellular oxidative stress and DNA damage. Utilizing ectodermal memory provides strong evidence that this aberrant methylation is embryonic in origin or occurs during gestation.

6 The ASD Hexamer Gene Matrix Hypothesis (AHGM):

Based on all the findings derived from the research, for an individual to develop or to be diagnosed with autism spectrum disorder (ASD), they must possess a differentially methylated pattern in at least 273 genes (including paralogs), accompanied by at least 20 differentially methylated genes (including paralogs) within the promoter region. The severity and divergence in mental and social skills depend on the number and function of these genes, which supports the umbrella effect of autism. The methylated genes should be according to the following gene matrix:

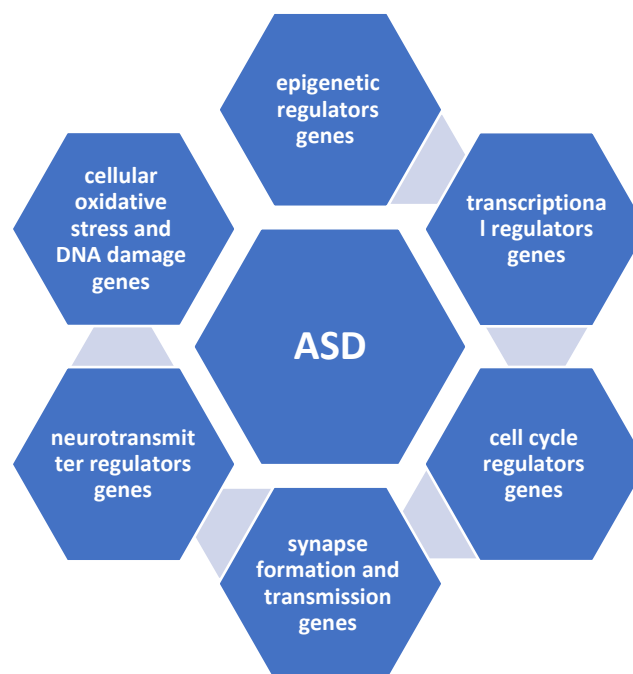


Figure (7): Schematic diagram of the **AHGM** hypothesis

Ethics Approval and Consent to Participate

Ethical approval to conduct this study was granted by the Ethics Committee affiliated with Kirkuk University. Prior to participation, all participants or their legal guardians gave their informed consent. Additionally, Quillbot (<https://quillbot.com/>) and Grammarly (<https://www.grammarly.com/>) are two AI-assisted technologies that were employed to enhance the language and readability of several sections of this work. All information was thoroughly vetted and edited, and these technologies were applied under human supervision and control. SFARI and Gene Cards databases were used as a reference for gene information.

Consent for Publication

All participants or their legal guardians gave their informed agreement for the anonymized data to be published.

Availability of data and materials

The raw sequencing data generated in this study are available at NCBI SRA under accession number([PRJNA1397483](https://www.ncbi.nlm.nih.gov/sra/PRJNA1397483)). The SRA records will be accessible with the following link (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1397483>).

Competing interests

No conflicting interests exist.

Funding

It was a self-funded project.

Acknowledgements

The authors would like to express their gratitude to the participants and their families for their cooperation. The authors also acknowledge the technical support and resources provided to conduct this study.

Authors' Contributions

Z.M.K. designed the study, contributed to sample collection, laboratory work, data interpretation, data analysis and manuscript drafting. The study was advised by H.Y.H. as advisor and by I.G.A. as a co-advisor.

Supplementary material

All supplementary files, and data tables supporting the findings of this study, are available online at [insert DOI link or Journal URL].

Abbreviations

ASD: Autism Spectrum Disorder

DMRs: Differentially Methylated Regions

KEGG: Kyoto Encyclopedia of Genes and Genomes

GO: Gene Ontology

SFARI: Simons Foundation Autism Research Initiative

References

- 1- Kana, R. K., Uddin, L. Q., Kenet, T., Chugani, D., & Müller, R. A. (2014). Brain connectivity in autism. *Frontiers in human neuroscience*, 8, 349.
- 2-Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1), 23-38.
- 3-Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J., & Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science (New York, N.Y.)*, 320(5880), 1224–1229. <https://doi.org/10.1126/science.1153252>.
- 4-Nagarajan RP, Hogart AR, Gwye Y, Martin MR, LaSalle JM. Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics*. 2006 Oct-Dec;1(4):e1-11. doi: 10.4161/epi.1.4.3514. PMID: 17486179; PMCID: PMC1866172.
- 5- Corley, M. J., Vargas-Maya, N., Pang, A. P., Lum-Jones, A., Li, D., Khadka, V., ... & Maunakea, A. K. (2019). Epigenetic delay in the neurodevelopmental trajectory of DNA methylation states in autism spectrum disorders. *Frontiers in Genetics*, 10, 907
- 6-Salcedo-Arellano, M. J., Cabal-Herrera, A. M., Punatar, R. H., Clark, C. J., Romney, C. A., & Hagerman, R. J. (2021). Overlapping Molecular Pathways Leading to Autism Spectrum Disorders, Fragile X Syndrome, and Targeted Treatments. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, 18(1), 265–283. <https://doi.org/10.1007/s13311-020-00968-6>
- 7- Komiya, Y., & Habas, R. (2008). Wnt signal transduction pathways. *Organogenesis*, 4(2), 68–75. <https://doi.org/10.4161/org.4.2.5851>.
- 8-Bakos, J., Srancikova, A., Havranek, T., & Bacova, Z. (2018). Molecular Mechanisms of Oxytocin Signaling at the Synaptic Connection. *Neural plasticity*, 2018, 4864107. <https://doi.org/10.1155/2018/4864107>.
- 9-Farsetti, E., Amato, S., Aversa, M., Guidolin, D., Pedrazzi, M., Maura, G., Agnati, L. F., Cervetto, C., & Marcoli, M. (2025). Dual Oxytocin Signals in Striatal Astrocytes. *Biomolecules*, 15(8), 1122. <https://doi.org/10.3390/biom15081122>.
- 10-Pekarek, B. T., Kochukov, M., Lozzi, B., Wu, T., Hunt, P. J., Tepe, B., Hanson Moss, E., Tantry, E. K., Swanson, J. L., Dooling, S. W., Patel, M., Belfort, B. D. W., Romero, J. M., Bao, S., Hill, M. C., & Arenkiel, B. R. (2022). Oxytocin signaling is necessary for synaptic maturation of adult-born neurons. *Genes & development*, 36(21-24), 1100–1118. <https://doi.org/10.1101/gad.349930.122>.
- 11-Lal, N., Song, B., Zhang, C., Al-Shehari, W. A., Jabeen, S., Ahmed, N., Edrees, W. H., Gupta, R., Soomro, S., Cui, F., & Qasem, E. A. (2025). An overview of oxytocin integrative mechanisms in autism spectrum disorder. *Discover mental health*, 5(1), 185. <https://doi.org/10.1007/s44192-025-00331-1>.
- 12-Quintana, D. S., Rokicki, J., van der Meer, D., Alnæs, D., Kaufmann, T., Córdova-Palomera, A., Dieset, I., Andreassen, O. A., & Westlye, L. T. (2019). Oxytocin pathway gene networks in

the human brain. *Nature communications*, 10(1), 668. <https://doi.org/10.1038/s41467-019-08503-8>.

13-Feldman, R., Monakhov, M., Pratt, M., & Ebstein, R. P. (2016). Oxytocin pathway genes: evolutionary ancient system impacting on human affiliation, sociality, and psychopathology. *Biological psychiatry*, 79(3), 174-184.

14-Bahar, M. E., Kim, H. J., & Kim, D. R. (2023). Targeting the RAS/RAF/MAPK pathway for cancer therapy: from mechanism to clinical studies. *Signal transduction and targeted therapy*, 8(1), 455.

15-Zhang, Y. L., Wang, R. C., Cheng, K., Ring, B. Z., & Su, L. (2017). Roles of Rap1 signaling in tumor cell migration and invasion. *Cancer biology & medicine*, 14(1), 90–99. <https://doi.org/10.20892/j.issn.2095-3941.2016.0086>.

16-Cipollini, M., Figlioli, G., Maccari, G., Garritano, S., De Santi, C., Melaiu, O., ... & Landi, S. (2016). Polymorphisms within base and nucleotide excision repair pathways and risk of differentiated thyroid carcinoma. *DNA repair*, 41, 27-31.

17-Attia, S. M., Al-Hamamah, M. A., Ahmad, S. F., Nadeem, A., Attia, M. S., Ansari, M. A., ... & Al-Ayadhi, L. Y. (2020). Evaluation of DNA repair efficiency in autistic children by molecular cytogenetic analysis and transcriptome profiling. *DNA repair*, 85, 102750.

18-Kolbanovskiy, M., Aharonoff, A., Sales, A. H., Geacintov, N. E., & Shafirovich, V. (2021). Base and Nucleotide Excision Repair Pathways in DNA Plasmids Harboring Oxidatively Generated Guanine Lesions. *Chemical research in toxicology*, 34(1), 154–160. <https://doi.org/10.1021/acs.chemrestox.0c00463>.

19-Markkanen, E., Meyer, U., & Dianov, G. L. (2016). DNA Damage and Repair in Schizophrenia and Autism: Implications for Cancer Comorbidity and Beyond. *International Journal of Molecular Sciences*, 17(6), 856. <https://doi.org/10.3390/ijms17060856>.

20-Reddy S, Reddy V, Sharma S. Physiology, Circadian Rhythm. [Updated 2023 May 1]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK519507/>.

21-Zhang, Y., Chen, Y., Li, W., Tang, L., Li, J., & Feng, X. (2024). Targeting the circadian modulation: novel therapeutic approaches in the management of ASD. *Frontiers in Psychiatry*, 15, 1451242.

22-Nakato, M., Shiranaga, N., Tomioka, M., Watanabe, H., Kurisu, J., Kengaku, M., ... & Ueda, K. (2021). ABCA13 dysfunction associated with psychiatric disorders causes impaired cholesterol trafficking. *Journal of Biological Chemistry*, 296.

23-Iritani, S., Torii, Y., Habuchi, C., Sekiguchi, H., Fujishiro, H., Yoshida, M., ... & Ozaki, N. (2018). The neuropathological investigation of the brain in a monkey model of autism spectrum disorder with ABCA13 deletion. *International Journal of Developmental Neuroscience*, 71, 130-139.

24-Li, C. V., & Knoblich, J. A. (2025). Advancing autism research: Insights from brain organoid modeling. *Current Opinion in Neurobiology*, 92, 103030.

- 25-Pramparo, T., Lombardo, M. V., Campbell, K., Barnes, C. C., Marinero, S., Solso, S., ... & Courchesne, E. (2015). Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Molecular systems biology*, 11(12), 841.
- 26- Wesołek-Leszczynska, A., Pastusiak, K., Bogdański, P., & Szulińska, M. (2024). Can adipokine FAM19A5 be a biomarker of metabolic disorders? *Diabetes, Metabolic Syndrome and Obesity*, 17, 1651–1666. <https://doi.org/10.2147/DMSO.S460226>
- 27- Kim, H. B., Yoo, S., Kwak, H., Ma, S. X., Kim, R., Lee, M., Ha, N., Pyo, S., Kwon, S. G., Cho, E. H., Lee, S. M., Jang, J., Kim, W. K., Park, H. C., Baek, M., Park, Y., Park, J. Y., Park, J. W., Hwang, S. W., Hwang, J. I., ... Seong, J. Y. (2025). Inhibition of FAM19A5 reverses synaptic loss and cognitive decline in mouse models of Alzheimer's disease. *Alzheimer's research & therapy*, 17(1), 168. <https://doi.org/10.1186/s13195-025-01813-8>
- 28- Park, S., Shahapal, A., Yoo, S., Kwak, H., Lee, M., Lee, S. M., Hwang, J. I., & Seong, J. Y. (2024). FAM19A5 Deficiency Mitigates the A β Plaque Burden and Improves Cognition in Mouse Models of Alzheimer's Disease. *Experimental neurobiology*, 33(4), 193–201. <https://doi.org/10.5607/en24017>
- 29- Han, K. M., Tae, W. S., Kim, A., Kang, Y., Kang, W., Kang, J., ... & Ham, B. J. (2020). Serum FAM19A5 levels: a novel biomarker for neuroinflammation and neurodegeneration in major depressive disorder. *Brain, behavior, and immunity*, 87, 852-859.
- 30- Seong, J. Y., Kim, H., Yoo, S., Kwak, H., Ma, S., Kim, R., ... & Hwang, J. (2023). Antibody-based immunotherapy targeting FAM19A5 reverses synaptic loss and improves cognitive function in Alzheimer's disease.
- 31-De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Shih-Chen, F., Aleksic, B., Biscaldi, M., Bolton, P. F., Brownfeld, J. M., Cai, J., Campbell, N. G., ... Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526), 209–215. <https://doi.org/10.1038/nature13772>
- 32-Simons Foundation. (2026, 1,14). *SFARI Gene*. Retrieved from <https://gene.sfari.org/>
- 33-Freitag, C. M., Staal, W., Klauck, S. M., Duketis, E., & Waltes, R. (2010). Genetics of autistic disorders: Review and clinical implications. *European Child & Adolescent Psychiatry*, 19(3), 169–178. doi.org.
- 34-Mikhail, F. M., Lose, E. J., Robin, N. H., Desmond, R. A., Rutledge, S. L., Rutledge, S. C., Korf, B. R., & Carroll, A. J. (2011). Clinically relevant chromosomal imbalances in 225 patients with unexplained intellectual disability or developmental delay, including autism spectrum disorders. *Genetics in Medicine*, 13(5), 429–438. doi.org.
- 35- Leblond, C. S., Heinrich, J., Delorme, R., Proepper, C., Betancur, C., Huguet, G., Konyukh, M., Chaste, P., Ey, E., Rastam, M., Anckarsäter, H., Nygren, G., Gillberg, C., Melke, J., Toro, R., Regnault, B., Fauchereau, F., Mercati, O., Lemièrre, N., ... Bourgeron, T. (2012). Genetic and functional analyses of *SHANK2* mutations provide evidence for a multi-hit model of autism spectrum disorders. *PLOS Genetics*, 8(5), e1002721. doi.org
- 36- Swanberg, S. E., Mefford, H. C., Fransen, J., De Kovel, C. G., & Eichler, E. E. (2009). Reciprocal co-regulation of *EGR2* and *MECP2* is disrupted in Rett syndrome and autism. *Human Molecular Genetics*, 18(3), 525–534. <https://doi.org/10.1093/hmg/ddn380>

- 37- Doyle, A. J., Bowdin, S., Abot, A. L., Rokosh, R., Cunningham, M. L., Pauly, R. R., ... & Loeys, B. L. (2012). Mutations in the TGF- β repressor *SKI* cause Shprintzen-Goldberg syndrome with aortic aneurysm. *Nature Genetics*, 44(11), 1249–1254. doi.org
- 38- GeneCards. (15/1/2026). *MGMT Gene - Methylguanine-DNA methyltransferase*. Weizmann Institute of Science. Retrieved March 7, 2026, from www.genecards.org
- 39-Nagarajan, R. P., Patzel, K. A., Martin, M., Yasui, D. H., Swanberg, S. E., Hertz-Picciotto, I., Hansen, R. L., Van de Water, J., & Lasalle, J. M. (2008). *MECP2* promoter methylation and X chromosome inactivation in autism. *Autism Research*, 1(3), 169–178. doi.org.
- 40-Shao, W., Zheng, H., Zhu, J., Li, W., Li, Y., Hu, W., Zhang, J., Jing, L., Wang, K., & Jiang, X. (2023). Deletions of *Cacna2d3* in parvalbumin-expressing neurons leads to autistic-like phenotypes in mice. *Neurochemistry international*, 169, 105569. <https://doi.org/10.1016/j.neuint.2023.105569>
- 41-Tremblay, M. W., & Jiang, Y. H. (2019). DNA Methylation and Susceptibility to Autism Spectrum Disorder. *Annual review of medicine*, 70, 151–166. <https://doi.org/10.1146/annurev-med-120417-091431>
- 42- Cai, Q., Zhou, Z., Luo, R., Yu, T., Li, D., Yang, F., & Yang, Z. (2022). Novel *GRIA2* variant in a patient with atypical autism spectrum disorder and psychiatric symptoms: a case report. *BMC pediatrics*, 22(1), 629. <https://doi.org/10.1186/s12887-022-03702-7>
- 43- Bonham, L. W., Evans, D. S., Liu, Y., Cummings, S. R., Yaffe, K., & Yokoyama, J. S. (2018). Neurotransmitter Pathway Genes in Cognitive Decline During Aging: Evidence for *GNG4* and *KCNQ2* Genes. *American journal of Alzheimer's disease and other dementias*, 33(3), 153–165. <https://doi.org/10.1177/1533317517739384>
- 44- Bayraktar, A., Lam, S., Altay, O., Li, X., Yuan, M., Zhang, C., Arif, M., Turkez, H., Uhlén, M., Shoaie, S., & Mardinoglu, A. (2021). Revealing the Molecular Mechanisms of Alzheimer's Disease Based on Network Analysis. *International Journal of Molecular Sciences*, 22(21), 11556. <https://doi.org/10.3390/ijms222111556>
- 45-Martin, M., Vermeiren, S., Bostaille, N., Eubelen, M., Spitzer, D., Vermeersch, M., ... & Vanhollebeke, B. (2022). Engineered Wnt ligands enable blood-brain barrier repair in neurological disorders. *Science*, 375(6582), eabm4459.
- 46- Bae, S. M., & Hong, J. Y. (2018). The Wnt Signaling Pathway and Related Therapeutic Drugs in Autism Spectrum Disorder. *Clinical psychopharmacology and neuroscience : the official scientific journal of the Korean College of Neuropsychopharmacology*, 16(2), 129–135. <https://doi.org/10.9758/cpn.2018.16.2.129>
- 47- Caracci, M. O., Ávila, M. E., & De Ferrari, G. V. (2016). Synaptic Wnt/GSK3 β signaling hub in autism. *Neural plasticity*, 2016(1), 9603751.
- 48- Perkins RS, Singh R, Abell AN, Krum SA and Miranda-Carboni GA (2023) The role of *WNT10B* in physiology and disease: A 10-year update. *Front. Cell Dev. Biol.* 11:1120365. doi: 10.3389/fcell.2023.1120365
- 49- Yokoyama M, Kimura MY, Ito T, Hayashizaki K, Endo Y, Wang Y, Yagi R, Nakagawa T, Kato N, Matsubara H and Nakayama T (2021) Myosin Light Chain 9/12 Regulates the

Pathogenesis of Inflammatory Bowel Disease. *Front. Immunol.* 11:594297. doi: 10.3389/fimmu.2020.594297

50- Hayashizaki, K., Kimura, M. Y., Tokoyoda, K., Hosokawa, H., Shinoda, K., Hirahara, K., ... & Nakayama, T. (2016). Myosin light chains 9 and 12 are functional ligands for CD69 that regulate airway inflammation. *Science immunology*, 1(3), eaaf9154-eaaf9154.

51- Mapmygenome. (10..3.2026). *Decode Your DNA for Longevity, Wellness*. IRS4 Gene: Function, Mutations, Research & More

52-Liao, X., & Li, Y. (2020). Genetic associations between voltage-gated calcium channels and autism spectrum disorder: a systematic review. *Molecular brain*, 13(1), 96. <https://doi.org/10.1186/s13041-020-00634-0>

53-Gomathi, M., & Pradeepkumar, M. (2024). CACNA1B Gene Mutation Associated with Typical Rett Syndrome Patients from India. *Biomedical and Biotechnology Research Journal*, 8(Abstract Supplement), NA-NA.

54- Serajee, F. J., Nabi, R., Zhong, H., & Mahbubul Huq, A. H. (2003). Association of INPP1, PIK3CG, and TSC2 gene variants with autistic disorder: implications for phosphatidylinositol signalling in autism. *Journal of medical genetics*, 40(11), e119. <https://doi.org/10.1136/jmg.40.11.e119>

55- Vaishnavi, V., Manikandan, M., & Munirajan, A. K. (2014). Mining the 3'UTR of autism-implicated genes for SNPs perturbing microRNA regulation. *Genomics, proteomics & bioinformatics*, 12(2), 92–104. <https://doi.org/10.1016/j.gpb.2014.01.003>

56- Chakraborti, B., Verma, D., Karmakar, A., Jaiswal, P., Sanyal, A., Paul, D., ... & Rajamma, U. (2016). Genetic variants of MAOB affect serotonin level and specific behavioral attributes to increase autism spectrum disorder (ASD) susceptibility in males. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 71, 123-136.

57- Verma, D., Chakraborti, B., Karmakar, A., Nandi, A., Debarshi, P., Sinha, A., ... & Rajamma, U. (2017). Genetic effect of monoamine oxidase B (MAOB) gene on ASD associated behavior phenotypes. *Canadian Journal of Biotechnology*, 1(Special), 77.

58- Bortolato, M., Godar, S. C., Alzghoul, L., Zhang, J., Darling, R. D., Simpson, K. L., Bini, V., Chen, K., Wellman, C. L., Lin, R. C., & Shih, J. C. (2013). Monoamine oxidase A and A/B knockout mice display autistic-like features. *The international journal of neuropsychopharmacology*, 16(4), 869–888. <https://doi.org/10.1017/S1461145712000715>

59-Yang Y, Wen X, Wang L. Advancements in DNA methylation technologies and their application in cancer diagnosis. *Epigenetics*. 2025 Dec;20(1):2539995. doi: 10.1080/15592294.2025.2539995. Epub 2025 Jul 28. PMID: 40726108; PMCID: PMC12309559.

60- Marcogliese, P. C., Deal, S. L., Andrews, J., Harnish, J. M., Bhavana, V. H., Graves, H. K., ... & Yamamoto, S. (2022). Drosophila functional screening of de novo variants in autism uncovers damaging variants and facilitates discovery of rare neurodevelopmental diseases. *Cell Reports*, 38(11).

- 61- Davinelli, S., Medoro, A., Siracusano, M., Savino, R., Saso, L., Scapagnini, G., & Mazzone, L. (2025). Oxidative stress response and NRF2 signaling pathway in autism spectrum disorder. *Redox biology*, 83, 103661. <https://doi.org/10.1016/j.redox.2025.103661>
- 62- Lossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y. H., ... & Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), 216-221. doi.org
- 63- C Yuen, R. K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R. V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., Pellecchia, G., Buchanan, J. A., Walker, S., Marshall, C. R., Uddin, M., Zarrei, M., Deneault, E., D'Abate, L., Chan, A. J., Koyanagi, S., ... Scherer, S. W. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature neuroscience*, 20(4), 602–611. <https://doi.org/10.1038/nn.4524>
- 64- Turner, T. N., Coe, B. P., Dickel, D. E., Hoekzema, K., Nelson, B. J., Zody, M. C., Kronenberg, Z. N., Hormozdiari, F., Raja, A., Pennacchio, L. A., Darnell, R. B., & Eichler, E. E. (2017). Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell*, 171(3), 710–722.e12. <https://doi.org/10.1016/j.cell.2017.08.047>
- 65- Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J. Y., ... & Wigler, M. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, 180(3), 568-584. <https://doi.org/10.1016/j.cell.2019.12.036>
- 66-Li L, Chen R, Zhang H, Li J, Huang H, Weng J, Tan H, Guo T, Wang M and Xie J (2024) The epigenetic modification of DNA methylation in neurological diseases. *Front. Immunol.* 15:1401962. doi: 10.3389/fimmu.2024.1401962.
- 67-Woodbury-Smith, M., Lamoureux, S., Begum, G., Nassir, N., Akter, H., O'Rielly, D. D., Rahman, P., Wintle, R. F., Scherer, S. W., & Uddin, M. (2022). Mutational Landscape of Autism Spectrum Disorder Brain Tissue. *Genes*, 13(2), 207. <https://doi.org/10.3390/genes13020207>
- 68-Gong, K., Xie, L., Wu, Z. S., Xie, X., Zhang, X. X., & Chen, J. L. (2021). Clinical exome sequencing reveals a mutation in PDHA1 in Leigh syndrome: A case of a Chinese boy with lethal neuropathy. *Molecular genetics & genomic medicine*, 9(4), e1651. <https://doi.org/10.1002/mgg3.1651>
- 69-Luo, T., Pan, J., Zhu, Y., Wang, X., Li, K., Zhao, G., Li, B., Hu, Z., Xia, K., & Li, J. (2024). Association between de novo variants of nuclear-encoded mitochondrial-related genes and undiagnosed developmental disorder and autism. *QJM : monthly journal of the Association of Physicians*, 117(4), 269–276. <https://doi.org/10.1093/qjmed/hcad249>
- 70-Mahmoud, L., Shaw, J., Croke, B., & Fleischer, J. (2022). eP104: Two patients with autism secondary to ZBTB18 loss-of-function pathogenic variants. *Genetics in Medicine*, 24(3), S67.
- 71-Yang, F., Ding, Y., Wang, Y., Zhang, Q., Li, H., Yu, T., ... & Wang, X. (2024). A de novo variant in ZBTB18 gene caused autosomal dominant non-syndromic intellectual disability 22 syndrome: A case report and literature review. *Medicine*, 103(2), e35908.

- 72-More, R. P., Warrier, V., Brunel, H., Buckingham, C., Smith, P., Allison, C., ... & Baron-Cohen, S. (2023). Identifying rare genetic variants in 21 highly multiplex autism families: the role of diagnosis and autistic traits. *Molecular Psychiatry*, 28(5), 2148-2157.
- 73- Hiatt, S. M., Neu, M. B., Ramaker, R. C., Hardigan, A. A., Prokop, J. W., Hancarova, M., ... & Cooper, G. M. (2018). De novo mutations in the GTP/GDP-binding region of RALA, a RAS-like small GTPase, cause intellectual disability and developmental delay. *PLoS Genetics*, 14(11), e1007671.
- 74- Okamoto, N., Takata, A., Miyake, N., & Matsumoto, N. (2019). RALA mutation in a patient with autism spectrum disorder and Noonan syndrome-like phenotype. *Congenital anomalies*, 59(6), 195–196. <https://doi.org/10.1111/cga.12327>
- 75- Nava, C., Lamari, F., Héron, D., Mignot, C., Rastetter, A., Keren, B., ... & Depienne, C. (2012). Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. *Translational psychiatry*, 2(10), e179-e179.
- 76- Huang, L., Jolly, L. A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., ... & Gecz, J. (2012). A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *The American Journal of Human Genetics*, 91(4), 694-702.
- 77- Ibarreta, D., Tao, J., Parrilla, R., & Ayuso, M. S. (1997). Mutation analysis of chromosome 19 calmodulin (CALM3) gene in Alzheimer's disease patients. *Neuroscience letters*, 229(3), 157–160. [https://doi.org/10.1016/s0304-3940\(97\)00453-9](https://doi.org/10.1016/s0304-3940(97)00453-9)
- 78- O'Day, D. H., Eshak, K., & Myre, M. A. (2015). Calmodulin Binding Proteins and Alzheimer's Disease. *Journal of Alzheimer's disease : JAD*, 46(3), 553–569. <https://doi.org/10.3233/JAD-142772>
- 79- O'Day D. H. (2019). Alzheimer's Disease: A short introduction to the calmodulin hypothesis. *AIMS neuroscience*, 6(4), 231–239. <https://doi.org/10.3934/Neuroscience.2019.4.231>
- 80- O'Day D. H. (2023). Alzheimer's Disease beyond Calcium Dysregulation: The Complex Interplay between Calmodulin, Calmodulin-Binding Proteins and Amyloid Beta from Disease Onset through Progression. *Current issues in molecular biology*, 45(8), 6246–6261. <https://doi.org/10.3390/cimb45080393>
- 81-Talkowski, M. E., Maussion, G., Crapper, L., Rosenfeld, J. A., Blumenthal, I., Hanscom, C., Chiang, C., Lindgren, A., Pereira, S., Ruderfer, D., Aldinger, K. A., Kirby, A., Parris, J., Rosenfeld, J. A., Shaffer, L. G., Eichler, E. E., Girard, S. L., Fombonne, E., Rouleau, G. A., . . . Gusella, J. F. (2012). Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. *The American Journal of Human Genetics*, 91(6), 1128–1134. doi.org
- 82- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., & Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. *Nucleic acids research*, 36(Database issue), D480–D484. <https://doi.org/10.1093/nar/gkm882>

- 83- Narzisi, A., Muratori, F., Calderoni, S., Fabbro, F., & Urgesi, C. (2013). Neuropsychological profile in high functioning autism spectrum disorders. *Journal of autism and developmental disorders*, 43(8), 1895-1909.
- 84- Masini, E., Loi, E., Vega-Benedetti, A. F., Carta, M., Doneddu, G., Fadda, R., & Zavattari, P. (2020). An Overview of the Main Genetic, Epigenetic and Environmental Factors Involved in Autism Spectrum Disorder Focusing on Synaptic Activity. *International Journal of Molecular Sciences*, 21(21), 8290. <https://doi.org/10.3390/ijms21218290>
- 85- Mouat, J. S., & LaSalle, J. M. (2022). The promise of DNA methylation in understanding multigenerational factors in autism spectrum disorders. *Frontiers in Genetics*, 13, 831221.
- 86- Jolly, L. A., Nguyen, L. S., Domingo, D., Sun, Y., Barry, S., Hancarova, M., Plevova, P., Vlckova, M., Havlovicova, M., Kalscheuer, V. M., Graziano, C., Pippucci, T., Bonora, E., Sedlacek, Z., & Gecz, J. (2015). HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain. *Human molecular genetics*, 24(12), 3335–3347. <https://doi.org/10.1093/hmg/ddv083>
- 87- Lu, Z., Liu, Z., Mao, W., Wang, X., Zheng, X., Chen, S., ... & Li, J. D. (2020). Locus-specific DNA methylation of Mecp2 promoter leads to autism-like phenotypes in mice. *Cell death & disease*, 11(2), 85.
- 88- Zhou, Q., Tian, Y., Xu, C., Wang, J., & Jin, Y. (2021). Prenatal and postnatal traffic pollution exposure, DNA methylation in Shank3 and MeCP2 promoter regions, H3K4me3 and H3K27me3 and sociability in rats' offspring. *Clinical Epigenetics*, 13(1), 180.
- 89- Wang, T., Guo, H., Xiong, B., Stessman, H. A., Wu, H., Coe, B. P., ... & Eichler, E. E. (2016). De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nature communications*, 7(1), 13316.
- 90- Cheng, T. L., Chen, J., Wan, H., Tang, B., Tian, W., Liao, L., & Qiu, Z. (2017). Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. *Scientific reports*, 7(1), 42790.
- 91- Orefice, L. L., Zimmerman, A. L., Chirila, A. M., Sleboda, S. J., Head, J. P., & Ginty, D. D. (2016). Peripheral mechanosensory neuron dysfunction underlies tactile and behavioral deficits in mouse models of ASDs. *Cell*, 166(2), 299-313.
- 92- Kubota, T., & Mochizuki, K. (2016). Epigenetic effect of environmental factors on autism spectrum disorders. *International journal of environmental research and public health*, 13(5), 504.
- 93- Tran, N. Q. V., & Miyake, K. (2017). Neurodevelopmental disorders and environmental toxicants: epigenetics as an underlying mechanism. *International journal of genomics*, 2017(1), 7526592.
- 94- Francis, S. M., Kistner-Griffin, E., Yan, Z., Guter, S., Cook, E. H., & Jacob, S. (2016). Variants in adjacent oxytocin/vasopressin gene region and associations with ASD diagnosis and other autism related endophenotypes. *Frontiers in neuroscience*, 10, 195.
- 95- Egawa, J., Watanabe, Y., Endo, T., Kitamura, H., & Someya, T. (2014). Possible association between the oxytocin receptor gene and N-acetylaspartate of the right medial temporal lobe in autism spectrum disorders. *Psychiatry & Clinical Neurosciences*, 68(1).

- 96- Romano, A., Tempesta, B., Micioni Di Bonaventura, M. V., & Gaetani, S. (2016). From autism to eating disorders and more: the role of oxytocin in neuropsychiatric disorders. *Frontiers in neuroscience*, 9, 497.
- 97- Xue, C., Chu, Q., Shi, Q., Zeng, Y., Lu, J., & Li, L. (2025). Wnt signaling pathways in biology and disease: mechanisms and therapeutic advances. *Signal Transduction and Targeted Therapy*, 10(1), 106.
- 98- Tollefsbol, T., Balasinor, N. H., Parte, P., & Singh, D. (2020). *Epigenetics and Reproductive Health* (Vol. 21). Academic Press.
- 99- Bou Najm, D., Alame, S., & Takash Chamoun, W. (2025). Unraveling the role of wnt signaling pathway in the pathogenesis of autism spectrum disorder (ASD): A systematic review. *Molecular Neurobiology*, 62(4), 4971-4992
- 100- Kumar, S., Reynolds, K., Ji, Y., Gu, R., Rai, S., & Zhou, C. J. (2019). Impaired neurodevelopmental pathways in autism spectrum disorder: a review of signaling mechanisms and crosstalk. *Journal of neurodevelopmental disorders*, 11(1), 10.
- 101- Kwan, V., Unda, B. K., & Singh, K. K. (2016). Wnt signaling networks in autism spectrum disorder and intellectual disability. *Journal of neurodevelopmental disorders*, 8(1), 45.
- 102- Langton, P. F., Kakugawa, S., & Vincent, J. P. (2016). Making, exporting, and modulating Wnts. *Trends in cell biology*, 26(10), 756-765.
- 103- Morandi, L., Franceschi, E., de Biase, D., Marucci, G., Tosoni, A., Ermani, M., Pession, A., Tallini, G., & Brandes, A. (2010). Promoter methylation analysis of O6-methylguanine-DNA methyltransferase in glioblastoma: detection by locked nucleic acid based quantitative PCR using an imprinted gene (SNURF) as a reference. *BMC cancer*, 10, 48.
<https://doi.org/10.1186/1471-2407-10-48>
- 104- Qian, X. C., & Brent, T. P. (1997). Methylation hot spots in the 5' flanking region denote silencing of the O 6-methylguanine-DNA methyltransferase gene. *Cancer research*, 57(17), 3672-3677.
- 105- Stupp, R., Mason, W. P., Van Den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., ... & Mirimanoff, R. O. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New England journal of medicine*, 352(10), 987-996.
- 106- Hegi, M. E., Diserens, A. C., Gorlia, T., Hamou, M. F., De Tribolet, N., Weller, M., ... & Stupp, R. (2005). MGMT gene silencing and benefit from temozolomide in glioblastoma. *New England Journal of Medicine*, 352(10), 997-1003.
- 107- Perini, S., Filosi, M., & Domenici, E. (2023). Candidate biomarkers from the integration of methylation and gene expression in discordant autistic sibling pairs. *Translational Psychiatry*, 13(1), 109.
- 108- Morales-Marín, M. E., Castro Martínez, X. H., Centeno Cruz, F., Barajas-Olmos, F., Náfate López, O., Gómez Coteró, A. G., ... & Nicolini Sánchez, H. (2023). Differential DNA methylation from autistic children enriches evidence for genes associated with ASD and new candidate genes. *Brain Sciences*, 13(10), 1420.

- 109-Loke, Y. J., Hannan, A. J., & Craig, J. M. (2015). The role of epigenetic change in autism spectrum disorders. *Frontiers in neurology*, 6, 107.
- 110-Zhang, B., Hu, X., Li, Y., Ni, Y., & Xue, L. (2022). Identification of methylation markers for diagnosis of autism spectrum disorder. *Metabolic brain disease*, 37(1), 219-228.
- 111-Mordaunt, C. E., Jianu, J. M., Laufer, B. I., Zhu, Y., Hwang, H., Dunaway, K. W., ... & LaSalle, J. M. (2020). Cord blood DNA methylome in newborns later diagnosed with autism spectrum disorder reflects early dysregulation of neurodevelopmental and X-linked genes. *Genome medicine*, 12(1), 88.
- 112-Sohal, V. S., & Rubenstein, J. L. (2019). Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Molecular psychiatry*, 24(9), 1248-1257.
- 113-Hegarty, J. P., Weber, D. J., Cirstea, C. M., & Beversdorf, D. Q. (2018). Cerebro-cerebellar functional connectivity is associated with cerebellar excitation–inhibition balance in autism spectrum disorder. *Journal of autism and developmental disorders*, 48(10), 3460-3473.
- 114-Gholamalizadeh, H., Amiri-Shahri, M., Rasouli, F., Ansari, A., Baradaran Rahimi, V., & Reza Askari, V. (2024). DNA methylation in autism spectrum disorders: biomarker or pharmacological target?. *Brain Sciences*, 14(8), 737.
- 115-Gui, A., Jones, E. J., Wong, C. C., Meaburn, E., Xia, B., Pasco, G., ... & BASIS Team. (2020). Leveraging epigenetics to examine differences in developmental trajectories of social attention: A proof-of-principle study of DNA methylation in infants with older siblings with autism. *Infant Behavior and Development*, 60, 101409.
- 116- Stoccoro A, Conti E, Scaffei E, Calderoni S, Coppedè F, Migliore L, Battini R. DNA Methylation Biomarkers for Young Children with Idiopathic Autism Spectrum Disorder: A Systematic Review. *Int J Mol Sci*. 2023 May 23;24(11):9138. doi: 10.3390/ijms24119138. PMID: 37298088; PMCID: PMC10252672.
- 117- Zha, K., Mi, B., Xiong, Y., Wu, S., Lu, L., Zhang, S., Lu, X., Mak, H. C., Huang, J., Panayi, A. C., Knoedler, S., Chen, L., Liu, G., & Lin, S. (2025). Circadian Rhythm: Biological Functions, Diseases, and Therapeutic Targets. *MedComm*, 6(11), e70435. <https://doi.org/10.1002/mco2.70435>
- 118- Morin, R., Forest, G., & Imbeault, P. (2025). Circadian rhythms revealed: unraveling the genetic, physiological, and behavioral tapestry of the human biological clock and rhythms. *Frontiers in Sleep*, 4, 1544945.
- 119- Liu, Y., & Sancar, A. (2025). Biochemical mechanism of the mammalian circadian clock. *FEBS letters*, 10.1002/1873-3468.70150. Advance online publication. <https://doi.org/10.1002/1873-3468.70150>
- 120- Falguières T. (2022). ABC Transporters in Human Diseases: Future Directions and Therapeutic Perspectives. *International journal of molecular sciences*, 23(8), 4250. <https://doi.org/10.3390/ijms23084250>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialforarticle2.docx](#)
- [FISHERTESTlastversion.docx](#)