University of Plovdiv “Paisii Hilendarski”

Faculty of Biology

Department of Plant Physiology and Molecular Biology

Head of Department: Prof. Valentina Toneva, PhD

Asst. Prof. Tihomir Iliev Vachev, PhD

COMPARATIVE GENOMICS, TRANSCRIPTOMICS AND PROTEOMICS RESEARCH IN NEURODEVELOPMENTAL DISORDERS

SUMMARY

for awarding the scientific degree "DOCTOR OF SCIENCES"

Field of higher education: 4"Natural Sciences, Mathematics and Informatics"

Professional field: 4. 3. "Biological Sciences"

Scientific specialty: "Molecular Biology"

Plovdiv 2020

Contents

ABSTRACT

Hundreds of millions of people around the world suffer from mental illness. The World Health Organization (WHO) estimates that 25% of the time European staff spend on sick-leave is caused by mental disorders. This includes disabling diseases such as schizophrenia, autism spectrum disorder (ASD), specific language impairment (SLI), etc., which carry enormous weight: loss of productivity of affected patients, relatives, and the health system. In many cases, these disorders and diseases are difficult to diagnose and require many different approaches and protocols for both diagnosis and treatment of the disease. To address the clinical needs that arise in the diagnosis of this type of disease, modern medicine is looking for specific diagnostic and prognostic biomarkers in diseases that can correctly identify the disease through a specific model of candidate biomarker molecules.

Thanks to modern advances in molecular genetics and biotechnology, the development and application in recent years of new next-generation sequencing technologies such as transcriptomic and proteomic analyzes have allowed the discovery of a large number of molecular biomarkers, including small RNA molecules protein-coding messenger RNAs, DNA-specific variants, proteins, and metabolites. All of this has expanded the range of potential biomarkers, including genetic profiling, transcriptome, and proteome analysis. Despite the apparent assumption that the only suitable material for the study and validation of biomarkers associated with neurodevelopmental disorders is brain tissue, a new direction in the investigation for biomarkers in these diseases is the profiling of nonprotein coding regulatory genes such as miRNA and the use of transcripts analysis that shows high sensitivity and specificity in tissues other than the CNS.

Micro RNA molecules are small regulatory RNA molecules that show a change in the expression profile in a number of diseases. The role of miRNA molecules and the advantage they offer in associating specific expression changes with a respective disease underscores their value as potential molecular biomarkers. A number of studies have demonstrated the role of miRNA molecules and in particular the complex miRNA expression profile that can be used to diagnose diseases. The specific miRNA profile shows a high diagnostic value and the analysis of body fluids, as peripheral blood offers the possibility of conducting a relatively non-invasive diagnostic test. Micro RNA molecules as potential biomarkers have objectively measurable biological characteristics that can be used as indicators of normal or pathological processes: the miRNA profile serves as a highly specific marker for diagnosing, predicting, monitoring disease progression, and predicting therapeutic response. Peripheral blood cells as a dynamically reflecting system allow the profiling of miRNA expression, which makes it practically possible to use them as a potential object of examination. From this perspective, miRNA molecules are considered a potential new class of biomarkers in a number of diseases, including central nervous system (CNS) disorders. The existence of a standardized

methodology for global characterization and profiling of miRNA molecules (Small RNA sequencing) as well as fast universally applicable quantitative analysis (quantitative RT-PCR) approaches suggest that the approach to identifying miRNAs as biomarkers from discovery to validation would be even more effective than the validation approaches of traditional protein biomarkers. Early diagnosis and assessment of disease development are essential factors for successful disease control, in which timely therapeutic interventions are essential. The application of these findings in clinical practice brings us closer to personalized medicine, improving the care of the patient, increasing the capacity of diagnostic and therapeutic procedures.

In modern molecular genetics, RNA is at the basis of processes such as the transfer of genetic information and the regulation of gene expression. With a modest size of (20 - 24 nt.) and a simplified structure of miRNA molecules, they are a clear example in terms of functional flexibility in the world of RNAIn recent years, the analysis of miRNA molecules is increasingly used in medical research, both for the identification of new prognostic biomarkers in various diseases and in the search for new therapeutic approaches. Initially, much of the research on miRNA molecules was focused on research in various cancers. Micro RNA molecules are reliable biomarkers in neuropsychiatric diseases due to their significant stability. These aspects of miRNA molecules allow their detection not only from body tissues but also from body fluids such as blood (both from circulating nucleic acids in blood plasma and/or serum and from nuclear blood cells). While the diagnostic value of single biomarkers is limited due to the lack of sensitivity and specificity, the overall miRNA profile obtained by multiplex analysis using small RNA molecule sequencing platforms characterized by high informativeness and specificity, making it possible both to identify individual miRNA and the complete miRNA expression profile with a high degree of accuracy. In recent years, however, it has become clear that approaches related to the study of miRNA molecules are very appropriate in studying the etiology and pathogenesis of multifactorial diseases, which include neurodevelopmental disorders such as schizophrenia and ASD. Expression studies of miRNA molecules in the field of molecular psychiatry have not yet been conducted in Bulgaria, which further justifies the conduct of similar studies in Bulgaria. Proteomics is another potential approach that can generate new hypotheses essential in the pathogenesis and diagnosis of diseases by identifying candidate protein biomarkers. Proteomics approaches complement genetic and genomic research and thus focus attention on protein products that are associated with genetic abnormalities associated with neurodegenerative and neurodevelopmental disorders. The presented technologies lay the foundation for the beginning of early diagnostic testing, detection and evaluation of the development of diseases, all of which are essential for successful diagnosis and therapy, especially in patients where timely therapeutic interventions are extremely critical.

AIM AND TASKS

Aim

Conducting comparative genomics, transcriptomics and proteomics research in neurodevelopmental disorders

Tasks

1. Conducting a comparative expression analysis of protein-encoding geniuses (RNA Sequencing) in ASD.

2. Perform a comparative expression analysis of small RNA molecules (Small RNA sequencing) in the ASD.

3. Conducting a comparative proteome analysis - Isobaric Tag for Relative and Absolute Quantification (ITRAQ) in ASD.

4. Carrying out large-scale exome sequencing (Whole-Exome Sequencing) in ASD.

5. Conducting a comparative expression analysis of protein-encoding geniuses (Digital Gene Expression) in schizophrenia.

6. Conducting comparative expression (qRT-PCR) analysis of protein-encoding geniuses in ASD.

7. Perform comparative expression (qRT-PCR) analysis of miRNA in ASD.

8. Conducting comparative expression (qRT-PCR) analysis of protein-coding genes in schizophrenia.

9. Conducting ROC (Receiver Operating Characteristic) analysis of differentially expressed genes and miRNAs in ASD and schizophrenia.

MATERIALS AND METHODS

Patients

A total of 30 subjects (24 males and six females) with ASD aged 3 to 11 years, and 30 Healthy children sex- and age-matched to the ASD group were included in this study. All participants were randomly selected from the family practices in the Plovdiv region. Probands were evaluated by certified psychiatrists and the diagnosis of ASD was made by clinical examination, Gilliam autism rating scale (GARS), childhood autism rating scale (CARS) and autism diagnostic interview (ADI-R), adhering to the diagnostic and statistical manual of mental disorders (DSM V) criteria. The GARS norm-referenced screening instrument was used for ASD symptom assessment. To help differentiate subjects with ASD from those with other developmental delays, we used CARS. The ADR-R is a structured interview performed with the parents of the patients. This is the golden standard for assessment of ASD patients. The DSM-V is the standard classification of mental disorders used by mental health professionals and physicians in the USA and most research teams worldwide.

The control group representing typically developing children (TDC) was assigned with the aim to match by sex and age to the ASD group. Inspection of all children in the healthy group for absence of autistic features was done by clinical examination and CARS. Children with known infectious, oncological, metabolic or genetic conditions were excluded from the study. No children were receiving any drug therapy when they were recruited.

Ethics Statement

The Institutional Review Board of the Ethics Committee of the Medical University of Plovdiv approved the methodology of the study and the written informed consent forms.

Participants and clinical assessments

30 schizophrenic patients were recruited from State Psychiatric Hospital Pazardzhik Bulgaria after signing Inform Consent form, approved by Medical University of Plovdiv Local Ethics Committee. All the patients were interviewed by certified MINI rater to evaluate diagnosis of schizophrenia by DSM IV TR criteria. Psychiatric diagnosis was also determined independently by two psychiatrists using all available information, including the hospital records and standard psychiatric clinical interview. If both disagreed about a diagnosis, the patient was not included in the study. Each patient's clinical symptoms were assessed by a certified research psychiatrist using the SCI - PANSS interview and PANSS score sheet. The patient's group was also rated by CGI-S scale. Main inclusion criteria was that the schizophrenic patient had not taken his/her medication for at least 2 weeks. After

signing Inform Consent Form, control group was recruited from healthy volunteers, age and sex matched to patient's group. Participants were interviewed by clinician for significant symptoms and clear family history for lack of second degree relatives with psychiatry diagnosis of Axis I according to DSM IV TR criteria was obtained. The participants did not receive any medication before blood sampling and they had standard breakfast. Physical examination was done to prove lack of acute or chronic medical illness and if any was found, the person was also excluded.

Blood collection, serum processing, and RNA extraction

Blood (4 ml) was drawn by experienced physicians from a peripheral vein while the probands were fasting (>3 h without food consumption). The material was collected in EDTA-containing tubes and stored at 4°C before subsequent procedures. Thereafter, the samples were separated into serum and blood cells by centrifuging at 1600 g for 10 min at 4°C. A volume of 500 µl serum was recentrifuged at 16 000 g for 10 min at 4°C to remove any residual blood cells. The clear supernatant was transferred to RNase/DNase-free microfuge tubes in 200 µl aliquots and then stored at - 80°C until use. Total serum RNA (including miRNAs) was extracted from the 200 µl samples, with the addition of 5 µl *Caenorhabditis elegans* miR-39 (100 nM). This is a synthetic spike-in exogenous control, which is considered an appropriate choice for internal standardization. The procedure for RNA extraction from total serum was performed using the PAXgene blood miRNA kit (PreAnalytiX), according to the "Manual Purification of Total RNA, Including miRNA protocol" recommended by the manufacturer, but with some minor modifications, including initiation of the purification at step 4 and elution in 30 µl BR5 buffer.

Quantification of Serum MicroRNAs

The Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA) were used for miRNA specific cDNA synthesis. MicroRNA specific cDNA (5 µl) were subjected to pre amplification with peqGOLD Taq DNA Polymerase (VWR, Radnor, PA, USA) prior to the reverse transcription-(RT-PCR) step in order to enhance sensitivity of the test. The qRT-PCR was carried out using the Maxima SYBR Green qPCR Master Mix (Thermo Scientific) and the ABI PRISM® 7500 system (Applied Biosystems, Foster City, CA, USA). All the experiments were performed in duplicate. Each sample was normalized using spiked-in control and relative quantification of miRNAs were calculated applying the $2^{-\Delta\Delta Ct}$ method. Statistical analyses were made by the Statistical Package for the Social Sciences (SPSS) software, version 20.0 (SPSS®; IBM Inc., Armonk, NY, USA). The analysis of variance (ANOVA) *t*-test on the data of Ct values was used for investigation of dysregulation of the analyzed miRNAs between Control and ASD groups. MedCalc statistical (https://medcale.org/) software was used to perform receiver operating characteristic (ROC) analysis.

Genomic DNA extraction

Genomic DNA was extracted from the peripheral blood with QIAamp DNA Blood Mini following the manufacturer's recommendations. After measuring the concentration and the integrity, the extracted DNA was stored at -80°C prior to further usage.

Horizontal agarose gel electrophoresis of RNA

Qualitative assessment of the presence and extent of RNA degradation was performed using horizontal agarose gel electrophoresis in 1% agarose gel with ethidium bromide. RNA imaging was performed on 1% agarose TBE gel electrophoresis using a 0.5X TBE buffer separation system (50mM Tris, 50mM boric acid, 1mM EDTA, pH 8.3) at 4.5 V / cm² and ethidium bromide imaging agent. To assess the integrity and size of the analyzed nucleic acids, a DNA marker was used: O'GeneRuler (Fermentas), visualized using a fixed-wavelength photographic documenting system 312 nm.

Quantitative analysis of isolated RNAs

Spectrophotometric quantification of isolated RNA was performed using the Epoch Micro-Volume Spectrophotometer System (BioTec), using 2 µl volume of each sample, which allows minimal consumption of total RNA samples. The measurement was performed at a wavelength of 260 nm. The measured A260/A280 ratio showed that all isolated samples were of a quality suitable for subsequent analysis (in the range 1.93-2.10).

RNA-Seq (quantification) analysis

Pooled samples were created by adding an equivalent amount of total RNA from each individual sample to a final concentration of 5 µg RNA samples. Pooled RNA samples were precipitated according to the service requirements, each pooled RNA sample was mixed with 1/10th volume of 3M NaOAc, pH 5.2, and 3 volume 100% ethanol, to the final volume of 400 µl. Aliquots of pooled RNAs were frozen at -80°C and shipped on dry ice. RNA integrity of pooled samples (ASD and control group) was assessed by agarose gel electrophoresis and checked by Agilent 2100 Bioanalyzer.

In this study, we used Beijing Genomics Institute (BGI) as a Certified Service Provider for sequencing service. The total RNA samples were first treated with DNase I to degrade any possible DNA contamination. Then the mRNA was enriched by using the oligo (dT) magnetic beads. Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments (about 200 bp). Then the first strand of cDNA was synthesized by using random hexamer-primer. Buffer, dNTPs, RNase H and DNA polymerase I were added to synthesize the second strand. The double-

strand cDNA was purified with magnetic beads. End reparation and 3'end single nucleotide A (adenine) addition was then performed. Finally, sequencing adaptors were ligated to the fragments. The fragments were enriched by PCR amplification. During the QC step, Agilent 2100 Bioanalyzer and ABI StepOnePlus RealTime PCR System were used to qualify and quantify the sample library. The library products were sequencing via Illumina HiSeqTM 2000.

.

Quantification of gene expression

The expression level for each gene was determined by the numbers of reads uniquely mapped to the specific gene and the total number of uniquely mapped reads in the sample. The gene expression level was calculated by using RPKM. (Reads per Kilobase per Million mapped reads) method and the formula is shown as follows:

$$RPKM = \frac{10^6 C}{NL/10^3}$$

RPKM is a method of quantifying gene expression from RNA sequencing data by normalizing for total read length and the number of sequencing reads. The RPKM method is able to eliminate the influence of different gene lengths and sequencing discrepancy on the calculation of gene expression level. Therefore, the RPKM values can be directly used for comparing the difference of gene expression among samples. If there is more than one transcript for a gene, the longest one is used to calculate its expression level and coverage.

Screening of differentially expressed genes (DEGs)

This analysis includes the screening of genes that are differentially expressed among samples and KEGG pathway enrichment analysis for these DEGs. Referring to "The significance of digital gene expression profiles", a strict algorithm to identify differentially expressed genes between two samples has been developed.

P-value corresponds to differential gene expression test. FDR (False Discovery Rate) is a method to determine the threshold of P-value in multiple tests and assume that we have picked out R differentially expressed genes in which S genes really show differential expression and the other V genes are false positive. If we decide that the error ratio „Q = V/R" must stay below a cutoff (e.g. 1%), we should preset the FDR to a number no larger than 0.01. „FDR ≤ 0.001 and the absolute value of log2 Ratio ≥ 1"as the threshold was used to judge the significance of gene expression difference.

In silico design and analysis of primer pairs

PCR primer design is one of the most important factors by which amplification of a gene can be obtained and the formation of primer-dimers can be avoided. Specific primer design software, such as PrimerExpress® or Primer3 (http://example3.ut.ee/), was used in the present work.

Stem-loop q-RT - PCR quantification

Maxima First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) and miRNA-specific stem-loop primers were used for cDNA synthesis, following the manufacturer's instructions. The reaction mixtures for reverse transcription consisted of 8 μl RNA, 4 μl miRNA-specific cDNA synthesis primer mix (stem-loop and forward primers, 100 μm each), 4 μl Maxima Enzyme Mix and 4 μl 5X Reaction Mix in a final volume of 20 μl. Then, before carrying out the qRT-PCR quantification, 5 μl of each miRNA-specific cDNA were subjected to pre-amplification with peqGOLD Taq DNA Polymerase (VWR, Radnor, PA, USA) in order to increase the sensitivity of the assay. qRT-PCR was performed using Maxima SYBR Green qPCR Master Mix (Thermo Fisher Scientific) with an ABI 7500 system (Applied Biosystems). The samples were normalized against the spiked-in synthetic *C. elegans* Cel-miR-39 miRNA control. These measurements were performed in duplicates. Finally, relative quantification (RQ) was calculated by means of the $2^{-\Delta\Delta Ct}$ methods. Amplicons were confirmed by monitoring the dissociation curves (Melting curve analysis) and by agarose gel electrophoresis (data not shown).

Statistical analyses

Statistical assessments were performed using version 20.0 of the SPSS package (SPSS Inc., Chicago, IL, USA). Nonparametric Mann-Withney U test with delta Ct values were used to examine the differences in expression levels considering the Kolmogorov-Smirnov criteria for normality. Subsequent ROC (receiver operating characteristic) curve analysis was carried out by the MedCalc statistical software package to obtain specificity and sensitivity values of the analyzed circulating miRNA biomarkers.

Pathways prediction analysis of differentially expressed miRNAs in serum

All presently validated target genes of the investigated miRNAs were obtained by the miRWalk 2.0 database: http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/. Since this tool offers a convenient search option for putative target genes, but not for validated ones, we developed our own script which uses as an input a list of validated target genes and then assigns them to pathways found in the KEGG database: http://www.genome.jp/kegg/.

Small RNA Sequencing

Pool samples (20 μg total RNA) for both ASD and CT groups were generated using equal amounts of RNA from each individual sample. Sequencing was done in Beijing Genomics Institute (BGI), China via Illumina HiSeqTM 2000 technology. Briefly, after 15% Tris-Borate-EDTA (TBE) denaturing polyacrylamide gel electrophoreses (PAGE) fractioning of the total RNA, the 15-30 bp fragments were purified. 3′ (5′-pUCGUAUGCCGUCUUCUGCUUGidT-3′) and 5′ (5′-GUUCAGAGUUCUACAGUCCGACGAUC-3′) adapters were ligated to the small RNAs using T4 RNA ligase.

First-strand cDNA was synthesized using Illumina (5′-CAAGCAGAAGACGCATACGA-3′) reverse primer, followed by cDNA amplification with (5′-CAAGCAGAAGACGGCATACGA-3′) and (5′-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA-3′). After purification, the fragments were sequenced.

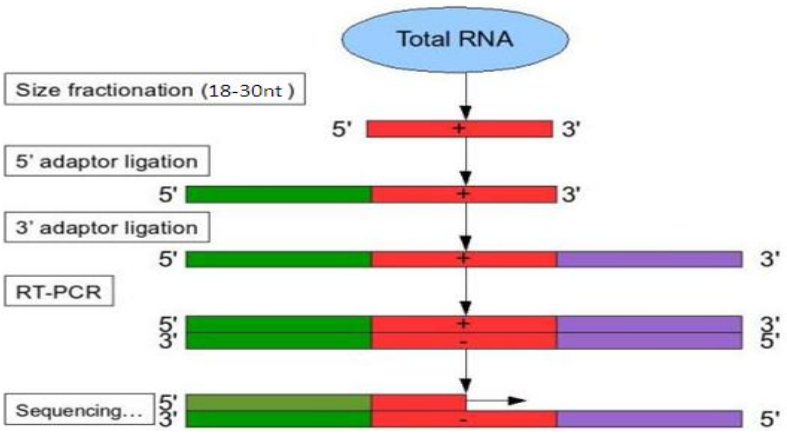


Figure 1. Schematic diagram of the protocol for isolation and analysis of small RNA molecules using new generation sequencing techniques (Small RNA-Seq).

Bioinformatics analysis of small RNA transcriptome

Standard Bioinformatics Analysis - The standard bioinformatics analysis was performed based on the data obtained from the profile of small RNA molecules (expression profile) including the steps described below as follows.

1. Removal of adapter sequences, low-quality markers, and more artifacts affecting the correct interpretation of the results of sequencing fragments.

2. Summary of the length distribution of the sequenced small RNA molecules.

3. Analysis of the general and specific sequences between the analyzed samples.

4. Investigation of the distribution of small RNA sequences by genomic localization.

5. Identification of rRNA molecules, mRNA molecules, small nuclear RNA molecules, etc. RNAs, mRNAs, snRNAs, etc., alignment to Rfam and Genebank databases.

6. Identification of repeat-associated small RNA molecules.

7. Identification of degradation fragments of mRNA molecules.

8. Identification of known miRNA molecules by comparison to miRBase sequences.

9. Annotation of small RNA molecules in priority categories.

10. Prediction of new miRNA molecules and their secondary structures with the help of Mireap from non-annotated sequences with mRNA molecules.

11. Analysis of the expression profile of annotated miRNA molecules.

12. Family analysis of known miRNA molecules.

Detailed bioinformatics analysis

1. Differential expression analysis of miRNA molecules.

2. Identification of target mRNA molecules of the identified miRNA molecules.

3. Target gene annotations GO annotation and KEGG miRNA analysis.

4. Analysis of bases editing in the miRNA sequences.

Identification of mutations in protein-coding sequences associated with ASD by Whole-Exome Sequencing

Over the last decade, Next Generation Sequencing (NGS) has shown significant improvement, establishing itself as a fast approach, with high productivity and cost-effectiveness for medical and research purposes. In particular, whole-exome sequencing has recently been developed as a large-scale sequencing technique to identify rare or de novo mutations in protein-coding sequences in the genome. Due to the extreme genetic complexity of the disorder and the association of multiple loci and the relatively weak genotype-phenotype correlation, ASD is

considered a model in which the application of whole-exome sequencing is well justified.

Based on the results of whole-exome sequencing, a model has been developed involving hundreds of genes with the potential for de novo mutations. Most of these mutations may increase the risk, but not become the "cause" of the disorder, which supports the already proposed polygenic model for the etiopathogenesis and development of ASD. Modern technological advances allow the parallel study of a huge number of markers scattered throughout the genome through a very fast and relatively inexpensive approach based on determining the nucleotide sequence of all coding regions of the currently annotated genes. These techniques provide summary information on the rare and common genetic variations in the genomic aspect that underlie a better understanding of neuropsychiatric disorders.

In addition, these new technologies changing the face of experimental design, moving from predominantly hypothetical approaches to genome-wide approaches that largely reject the preliminary hypothesis. Exome sequencing (also known as targeted exome re-sequencing) is an effective strategy for selectively sequencing coding regions (exons) in the genome, being relatively inexpensive, and an effective alternative to genome-wide sequencing approaches. The laborious approach to complete sequencing of all coding regions has the potential to become both a clinically relevant study for genetic diagnostics and the identification of DNA variants with a potentially significant pathological effect in diseases of unknown etiology. The aim of this approach is to identify functional changes in the genome that are responsible for the development of a specific pathology. Large-scale exome sequencing becoming the technique of choice to identify new genetic variations that underlie diseases, such as cancer, and a number of psychiatric diseases such as schizophrenia, ASD, and others. However, the development of effective approaches to the analysis of whole-exome sequencing data is essential to ensure clinically and analytically relevant results.

Sequencing characteristics

• Whole-exome sequencing was performed on two complex DNA samples, ASD and a control group (pools) using a next-generation sequencing platform of Illumina HiSeq 2000.

• A minimum of 91.23% of the regions were covered to a sequencing depth of at least 20 times coverage.

• The amount of information obtained from the sequencing is 16.00 GB of high-quality data, which were compared with a reference human genome. (Genome Reference Consortium human genome build 37, human genome 19).

• 281,566 variations were identified in comparison with the reference genome. 33,305 variations were identified among the analyzed samples, which are indicative of a serious mutation burden in coding sequences.

Enrichment and preparation of the library for sequencing

The samples were processed according to the Agilent SureSelect protocol; version 1.2.• Enrichment was performed in accordance with Agilent SureSelect protocols.• The concentration of each library was determined using the QPCR NGS Agilent Library Quantitative Evaluation Kit (G4880A).• Samples were pooled immediately before sequencing with a final concentration of 10 nM.

Files carrying information about the sequenced fragments (Fastq) are generated directly from the sequencing platform through specialized software of the manufacturer (OGT).

- The sequenced fragments were mapped according to the human genome version (hg19 / b37) using the Burrows-Wheeler Aligner (BWA) package, version 0.6.2.
- Local, re-mapping of fragments around potential insertion/deletion sites (INDEL) was performed using the Genome Analysis Tool (GATK) version 1.6. The use of the algorithm ensures the presence of a minimum number of mismatched bases in the sequenced fragments. The main effect of which is to reduce the presence of false-positive SNP polymorphisms, as well as accurately determine the length of INDEL mutations.
- Repeated sequenced fragments were marked using Picard version 1.98. The additional processing used removes the sequenced fragments probably obtained as a result of PCR deviations, which can introduce false-positive results in the obtained data.
- Additional processing of BAM files was performed using Samtools 0.1.18. Evaluation of the quality of the performed sequencing (Phred scale) was performed using GATK analysis.

Identification of genetic variants using the IGV browser

The identification of genetic variants of the exome sequencing was performed using an IGV browser, allowing the tracking of each identified variant by providing detailed information on 1) the genomic localization of the identified variant 2) type of mutation to which the identified variant leads, Frameshift mutations, mutations in splicing sites, synonymous mutations, etc. 3) gene in which the identified variant is observed 4) type of the variant obtained as a result of the subsequent clinical effect: deleterious, variants of unknown clinical significance and benign variants.

Digital Gene Expression (DGE): Tag Profiling

We use Beijing Genomics Institute (*BGI)* as a Certified *Service Provider for DGE service*. The RNA samples used were pooled according to their disease status and by mixing equal amounts of RNA from each individual in the group. Total RNA (1- 2 μg) was fractionated using oligo-dT magnetic beads to yield poly(A+) mRNA. mRNA bound to the beads was then used as a template for first strand cDNA synthesis primed by oligo-dT and the second strand cDNA was consequently synthesized using random primers. Next, the double stranded cDNA covalently attached to oligo-dT beads was digested with *DpnII*. The fragments that remained attached to the beads were ligated to the Illumina GEX DpnII Adapter 1, which includes a *MmeI* recognition site. Therefore, the library preparation protocol allows only one tag per RNA molecule. Digestion with *MmeI* yielded the adapter tag linked to 20 bp of cDNA including 4 bp of the *DpnII* recognition site, which was ligated to GEX Adapter 2 at the site of *MmeI* cleavage.

Digital Gene Expression (DGE) tag annotation

A four-tier procedure for annotating the sequenced tags was developed. All NlaIII restriction sites 'CATG' were identified and +/- 17 bases of flanking sequence were extracted in silico using the Ensembl Perl application programming interface from all genes annotated in Ensembl version 58 and stored in a MySQL database. The sequences were extracted from all annotated spliced and unspliced transcripts, including one kb of sequence up- and downstream of annotated genes. Additionally, SNPs from the Ensembl Variation database were applied to the sequences to include all possible single base variants of the sequences as well as new possible annotations arising due to the introduction of new NlaIII restriction sites by SNPs. Of the 49,733 genes in Ensembl, 93% contain an NlaIII restriction site, which resulted in a database of about 10 million possible unique sequences. The sequenced tags were annotated by querying the database according to the following four-tier scheme: In the first tier, the tags mapping to the canonical site (most 3' NlaIII site in a transcript) were given precedence over other sites. Each tag was matched against the canonical site in each transcript, including SNP variants, and tags mapping to only one gene were assigned annotations. In the second tier, tags that did not match any of the canonical sites were searched against any sites present in an exon of spliced transcripts. Tags that matched a single gene were assigned to that gene. In the third tier the remaining tags were searched against sites from the unspliced transcript, including the added extra flanking sequences. Finally, in the fourth tier any remaining unmatched tags were searched against all the sites in the anti-sense direction of the spliced and unspliced transcripts as well as the additional flanking sequences. Tags matching more than one gene were discarded. Each tag was normalized to tags per million (TPM) and the total expression profile for each gene was calculated by summing all tags mapped to the same gene, including intronic tags.

Quantitative Real-Time PCR (qRT-PCR) analysis of protein-coding gene in schizophrenia

qRT-PCR was used for validation of DGE levels of protein coding gene. Copy DNA specific for protein coding gene was synthesized from total RNA using oligo $(dT)_{18}$ primer using RevertAid First Strand cDNA Synthesis Kit according to the assay protocol (Thermo Scientific). Reverse transcriptase reactions contained 1 μg of total RNA samples, 1μl oligo $(dT)_{18}$ primer and nuclease free water to final volume of 12 μl, after incubation at 65°C for 5 min to the mix was added 4μl 5X RT buffer, 1μl RiboLock RNase Inhibitor (20 U/μl), 2 μl 10 mM dNTP Mix and 1 μl RevertAid MMuLV Reverse Transcriptase (200 U/μl) to final volume of 20 μl. Relative gene expression method was employed to determine gene expression levels. The reactions were set up in a 96-well format using the Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems) and Maxima SYBER Green/Rox qRT-PCR Kit (Thermo Fisher Scientific). Relative quantification (RQ) was determined with respect to the ACTIN B mRNA levels by standard $2^{-\Delta\Delta Ct}$ method.

ROC analysis of dysregulated mRNAs in schizophrenia

To investigate the characteristics of these mRNAs as potential diagnostic biomarkers in autism, MedCalc 9.6.4 software was used to construct ROC (receiver operating characteristic) curves and calculate the AUC (area under the ROC curve).

Isobaric Tag for Relative and Absolute Quantification (ITRAQ) LC-MS/MS analysis

All serum samples were collected before breakfast and processed according to standard operating procedures to minimize pre-analytical variation. Blood samples were allowed to clot for 1 h at 37°C and were centrifuged at 3000 g at 4°C for 20 min. The serum was then frozen immediately at − 80 °C.

iTRAQ discovery experiments. Total protein (100 μg) was recovered from each sample solution and digested with Trypsin Gold (Promega, Madison, WI, USA) with the ratio of protein:trypsin = 20:1 at 37°C for 4 h, followed by another trypsin digestion with the same ratio for 8 hours.

iTRAQ Labeling. After trypsin digestion, the peptides were dried by vacuum centrifugation. The peptides were redissolved in 0.5 M TEAB and processed according to the manufacturer's protocol in preparation for 8-plex iTRAQ (AB Sciex, Framingham, MA, USA). The peptides were labeled with isobaric tags and incubated at room temperature for 2 h. The labeled peptide mixtures were then pooled and dried by vacuum centrifugation.

SCX chromatography. For SCX chromatography were used the Shimadzu LC-20AB HPLC Pump system, the peptide from digestion were reconstituted with 4 ml buffer A (25 mM NaH_2PO_4 in 25% ACN, pH 2.7) and loaded onto a 4.6×250 mm

Ultremex SCX column containing 5-μm particles (Phenomenex). The peptides were eluted at a flow rate of 1 mL/min with a gradient of buffer A for 10 min, 5-35% buffer B (25 mM NaH_2PO_4, 1M KCl in 25% ACN, pH 2.7) for 11 min, 35-80% buffer B for 1 min. The system is then maintained in 80% buffer B for 3 min before equilibrating with buffer A for 10 min prior to the next injection. Elution was monitored by measuring absorbance at 214 nm, and fractions are collected every 1 min. The eluted peptides were pooled as 20 fractions, desalted by Strata X C18 column (Phenomenex) and vacuum-dried.

LC-ESI-MS/MS analysis based on Triple TOF 5600. The combined mixtures were analyzed by LC-ESI-MS/MS analysis based on Triple TOF 5600. Each fraction were resuspended in certain volume of buffer A (2% ACN, 0.1% FA) and centrifuged at 20000 g for 10 min. In each fraction, the final concentration of peptide is about 0.5μg/μl on average. 10 μl supernatant is loaded on an Shimadzu LC-20AD nano HPLC by the autosampler onto a 2 cm C18 trap column (inner diameter 200 μm) and the peptides are eluted onto a resolving 10 cm analytical C18 column (inner diameter 75 μm). The samples are loaded at 15 l/min for 4 min, then the 44 min gradient were run at 400 nl/min starting from 2 to 35% B (98% ACN, 0.1% FA), followed by 2 min linear gradient to 80%, and maintenance at 80% B for 4 min, and finally return to 2% in 1 min. Data acquisition was performed with a TripleTOF 5600 System (AB SCIEX, Concord, ON) fitted with a Nanospray III source (AB SCIEX, Concord,ON) and a pulled quartz tip as the emitter (New Objectives, Woburn, MA). Data was acquired using an ion spray voltage of 2.5 kV, curtain gas of 30 PSI, nebulizer gas of 15 PSI, and an interface heater temperature of 150°C. The MS was operated with a RP of greater than or equal to 30 000 FWHM for TOF MS scans. For information-dependent acquisition (IDA), survey scans were acquired in 250 ms and as many as 30 product ion scans were collected if exceeding a threshold of 120 counts per second (counts/s) and with a 2+ to 5+ charge-state. Total cycle time was fixed to 3.3 s. Q2 transmission window was 100 Da for 100%. Four time bins were summed for each scan at a pulser frequency value of 11 kHz through monitoring of the 40 GHz multichannel TDC detector with four-anode channel detect ion. A sweeping collision energy setting of 35±5 eV adjust rolling collision energy was applied to all precursor ions for collision-induced dissociation. Dynamic exclusion was set for 1/2 of peak width (18 s), and then the precursor was refreshed off the exclusion list.

Transcriptomic profiling of protein-coding genes in schizophrenia using Digital Gene Expression Tag Profiling (DGE))

Digital Gene Expression (DGE-tag profiling)

Pooled samples were created by adding an equivalent amount of total RNA from each individual sample to a final concentration of 2 μg RNA samples. Pooled RNA samples were precipitated according to the service requirements, each pooled RNA sample was mixed with 1/10th volume of 3M NaOAc, pH 5.2, and 3 volume

100% ethanol, to the final volume of 400 µl. Aliquots of pooled RNAs were frozen at -80°C and shipped on dry ice. Beijing Genomics Institute (BGI) was used as a Certified Service Provider for DGE service. The RNA samples used were pooled according to their disease status and by mixing equal amounts of RNA from each individual in the group. Total RNA (1- 2 µg) was fractionated using oligo-dT magnetic beads to yield poly(A+) mRNA. mRNA bound to the beads was then used as a template for first-strand cDNA synthesis primed by oligo-dT and the second strand cDNA was consequently synthesized using random primers. Next, the double-stranded cDNA covalently attached to oligo-dT beads was digested with DpnII. The fragments that remained attached to the beads were ligated to the Illumina GEX DpnII Adapter 1, which includes a MmeI recognition site. Therefore, the library preparation protocol allows only one tag per RNA molecule. Digestion with MmeI yielded the adapter tag linked to 20 bp of cDNA including 4 bp of the DpnII recognition site, which was ligated to GEX Adapter 2 at the site of MmeI cleavage.

Digital Gene Expression (DGE) tag annotation

A four-tier procedure for annotating the sequenced tags was developed. All NlaIII restriction sites 'CATG' were identified and +/- 17 bases of flanking sequence were extracted in silico using the Ensembl Perl application programming interface from all genes annotated in Ensembl version 58 and stored in a MySQL database. The sequences were extracted from all annotated spliced and unspliced transcripts, including one kb of sequence up- and downstream of annotated genes. Additionally, SNPs from the Ensembl Variation database were applied to the sequences to include all possible single base variants of the sequences as well as new possible annotations arising due to the introduction of new *NlaIII* restriction sites by SNPs. Of the 49,733 genes in Ensembl, 93% contain an *NlaIII* restriction site, which resulted in a database of about 10 million possible unique sequences. The sequenced tags were annotated by querying the database according to the following four-tier scheme: In the first tier, the tags mapping to the canonical site (most 3' *NlaIII* site in a transcript) were given precedence over other sites. Each tag was matched against the canonical site in each transcript, including SNP variants, and tags mapping to only one gene were assigned annotations. In the second tier, tags that did not match any of the canonical sites were searched against any sites present in an exon of spliced transcripts. Tags that matched a single gene were assigned to that gene. In the third tier the remaining tags were searched against sites from the unspliced transcript, including the added extra flanking sequences. Finally, in the fourth tier any remaining unmatched tags were searched against all the sites in the anti-sense direction of the spliced and unspliced transcripts as well as the additional flanking sequences. Tags matching more than one gene were discarded. Each tag was normalized to tags per million (TPM) and the total expression profile for each gene was calculated by summing all tags mapped to the same gene, including intronic tags.

RESULTS

Whole-Exome Sequencing of protein-coding sequences in the human genome associated with ASD

In order to identify genomic variants in patients diagnosed with ASD, large-scale Whole Exome Sequencing was performed in complex samples from the subjects of the study, which include the cohort of children diagnosed with ASD and the cohort from the control group of healthy children, each of which includes 40 children.

The genomic variants obtained from the whole-exome sequencing were annotated with respect to gene affiliation and gene function, respectively, using data from the Ensemble database. The results of the analysis are focused on which genes and their corresponding transcription variants are affected by the identified genetic changes, and whether these variants can cause significant functional problems.

DNA sequencing was performed on a next-generation sequencing platform (NGS) of Ilumina in the laboratory of Oxford Gene Technology (OGT). Briefly, the resulting complex genomic DNA samples were fragmented using ultrasound, ligated to multiplexed adapters (paired-end adapters), and amplified by PCR using sequencing primers labeled with specific barcodes (indexes). Sequencing was performed on the HiSeq 2000 platform, Ilumina.

Genomic variants identified after whole-exome sequencing were as follows: mutations in the DNA sequence, including insertion, deletions, single nucleotide polymorphisms (SNPs), etc. Variants that passed through discriminant filters (steps described above) were classified as: deleterious mutations (potentially pathogenic variants), variants of unknown clinical significance, (benign variants). The classification of the identified variants was performed in accordance with the American College of Medical Genetics and Genomics (ACMG) guidelines for interpretation. The two main objectives in the present study were: 1) to confirm established genomic variants characteristic of ASD in the studied cohort, showing a trend, and supporting the hypothesis of genetic heterogeneity in ASD. 2) to identify specific genomic variants in the analyzed sample of patients diagnosed with ASD, which have not been identified so far. Regarding the first goal of the study, the data obtained support the existence of pathogenetic mutations in the studied samples, some of them characteristic of ASD.

The known variants of dbSNP (Release 135) were annotated in separate datasets, which allows relatively easy identification of new variants with serious prognostic clinical consequences. The groups included in the analysis were assessed for carrying the status of genetic variants (according to the criteria listed above) by sequencing each of the coding exome regions. Exons were analyzed based on data obtained from large-scale exome sequencing. The generated data were obtained

using a sequencing depth of 50X exome coverage of each specific locus, which allows the evaluation of the observed variants in the studied groups to be performed by estimating the mutation frequency as a function of the number of sequenced fragments (reads) obtained from the respective genome region.

Among the identified candidate genes, one of the most interesting is the ANK3 (Ankyrin 3) gene. A number of large-scale studies conducted in ASD have shown a strong association of the ANK3 gene with ASD. Recently, missense mutations in the ANK3 gene have been identified using exome sequencing in four of 67 patients diagnosed with ASD. Given the fact that the ANK3 gene is strongly associated with schizophrenia and bipolar disorder, the data support the hypothesis of an association between mutations in the ANK3 gene and ASD, which supports claims of a common molecular basis between ASD and other neuropsychiatric disorders.

The final analysis of the data obtained from the whole-exome DNA sequencing demonstrates the presence of mutations in ASD leading to the formation of non-functional proteins.

Basic mutations with the potential for serious biological consequences can be divided into several separate categories:

1. **Mutations leading to the formation of termination codons**, important due to the general assumption that the production of a termination codon leads to the formation of a smaller mRNA and, accordingly, a non-functional protein.

2. **Mutations leading to the loss of termination codons**, important due to the general assumption that the change of each termination codon leads to the formation of a larger mRNA and, accordingly, a pathologically functional protein.

3. **Mutations leading to the formation of serious non-synonymous substitutions**, important due to the general assumption that any change in the amino acid sequence can result in the formation of a pathologically functional or non-functional protein.

4. **Frameshift mutation** - deletions or insertions important due to the general assumption that leading to a change in the amino acid sequence may result in the formation of a protein with a pathological function or a completely non-functional protein.

The group of children diagnosed with ASD makes impressed that mutations leading to the formation of stop codons are associated with genes such as MOB3C (MOB kinase activator 3C), associated with the binding of metal ions. Cytochrome P450 Family 4 Subfamily B Member 1 (CYP4B1), a gene encoding a member of the cytochrome P450 superfamily of enzymes. Cytochrome P450 proteins catalyze

reactions related to drug metabolism and the synthesis of cholesterol, steroids, and other lipids. Regenerating islet-derived protein 4 (REG4), possibly associated with gastrointestinal epithelium and inflammation, PDE4DIP, phosphodiesterase anchor in the centrosome, RHBG, a transmembrane protein associated with the Rh family.

OR10X1, G-associated olfactory-associated receptor, FMO6P, flavin-containing monooxygenase pseudogene, CAPN8 associated with calcium-dependent cysteine endopeptidase activity and neuronal ceroid lipofuscinosis. TSSC1, tumor suppressor factor associated with Beckwith Weidman syndrome, Wilms tumor, rhabdomyosarcoma, adrenocortical carcinoma, and others.

The identified genes with mutations leading to the loss of stop codons include genes such as Fasciculation and elongation protein zeta 2 (FEZ2), associated with neuronal connectivity, Cytosolic beta-glucosidase (GBA3), connected to the hydrolysis of the glycosides, *Alcohol dehydrogenase 1C (class I), gamma polypeptide* (ADH1C), alcohol dehydrogenase, *matrilin 2* (MATN2), associated with the formation of filamentous bonds in the extracellular matrix, *N-Acetylneuraminate Synthase* (NANS), associated with sialic acid biosynthesis, *Actinin alpha* 3 (ACTN3), associated with a neuronal connection, *Neurofibromin 1* (NF1), associated with neurofibromatosis type 1 and syndrome autism.

Genes with mutations leading to non-synonymous substitutions in the group of patients with ASD have also been identified, such as *Spermatogenesis associated 3* (SPATA3) associated with spermatogenesis, *Solute Carrier Family 22 Member 1* (SLC22A1), a transmembrane protein associated with dopamine transport, *AarF Domain Containing Kinase 5* (ADCK5), possibly with tyrosine kinase activity, *NLR Family CARD Domain Containing* 5 (NLRC5), a gene that plays a role in the cytokine response and antiviral immunity by inhibiting NF-kappa-B activation and negative regulation of interferon type I signaling pathways, *Apolipoprotein B receptor* (APOBR), NUDT11, belonging to the phosphohydrolase family.

There were also 537 mutations in genes leading to a frameshift mutation in genes such as *Cub Domain-Containing Protein 2* (CDCP2), *Caspase 9* (CASP9), from the group of caspases, *Hornerin* (HRNR), *Ubiquitin-like -conjugating enzyme ATG3* (ATG3), ubiquitin-like conjugate enzyme, NOP16, encoding a nucleolar protein and *Fibroblast Growth Factor Receptor 4* (FGFR4-fibroblast growth factor 4).

Bioinformatics Analysis Procedures of iTRAQ quantitative proteomics

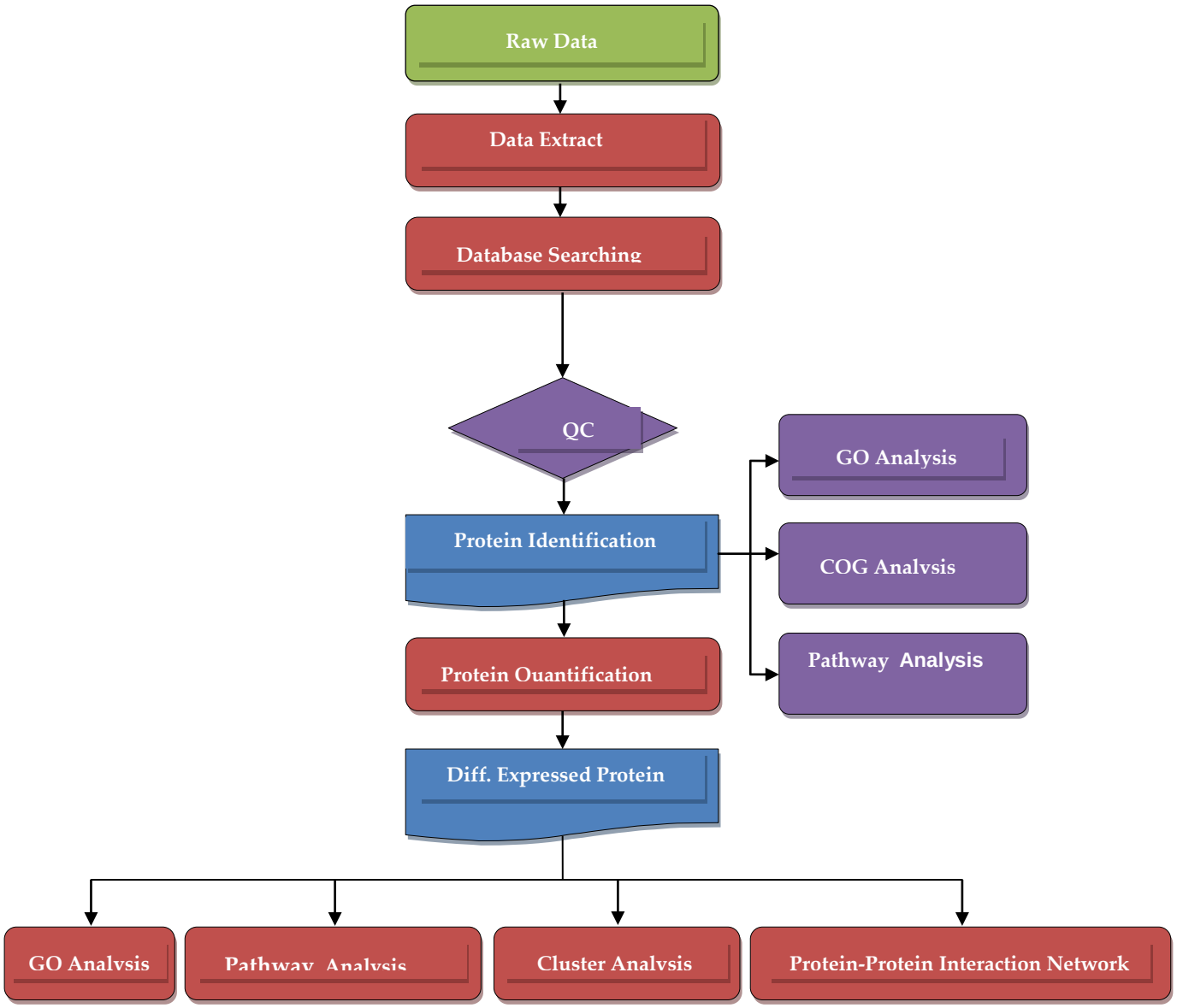


Figure 2. Bioinformatics Analysis Procedures. This figure shows the basic information Procedures. Firstly, identify the peak of the Raw Data, get the peak list. Then, eatablish the database and identify the peptide and protein. At last, compare the relationship of relative amount between the different samples, consequently get some important protein of interest.

Comparative analysis of differentially expressed proteins between the analyzed groups (ASD vs. Control group) with reduced expression

A total of 351 proteins were identified from iTRAQ-LC-MS/MS analysis data. Compared to the control group, a total of 60 differentially expressed proteins were identified, including 24 with increased expression and 36 with decreased expression.

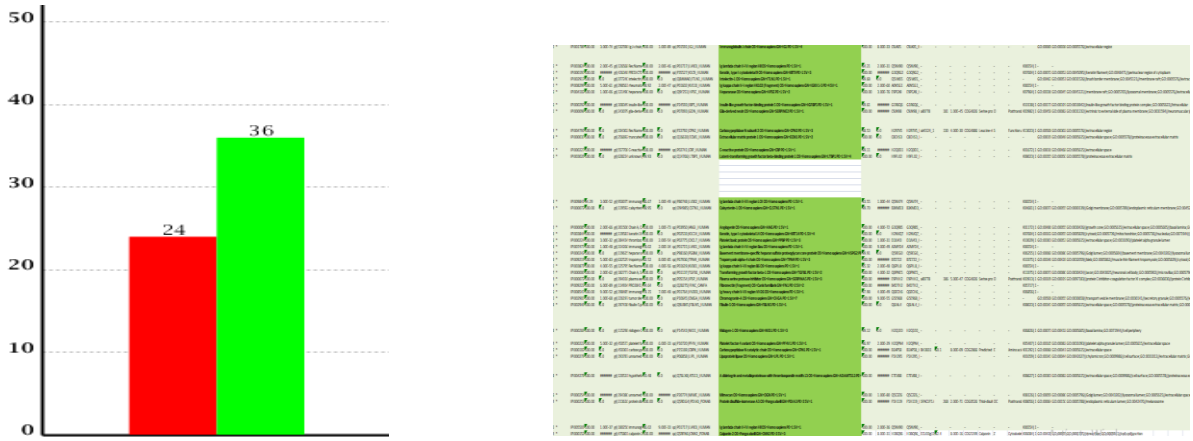


Figure 3. Differentially Expressed Protein Statistics X-axis: names of comparable group; Y-axis: the number of differentially expressed protein. Red means the number of up-regulated protein, green means the number of downregulated protein.

Proteins with reduced expression include:

SERPINE2 Glia-derived nexin (GDN)

Serine protease inhibitor with activity toward thrombin, trypsin, and urokinase. Promotes neurite extension by inhibiting thrombin. Binds heparin. Neural elongation and branching are key cellular events during brain development, as they underlie the formation of the proper construction of neural networks.

Intelectin 1

Omentin (intelectin-1, intestinal lactoferin receptor, endothelian lectin HL-1, galactofuranosebinding lectin) is newly identified secretory protein that is highly and selectively expressed in visceral adipose tissue relative to subcutaneous adipose tissue (adipokine).

Immunoglobulin J chain (Joining chain of multimeric IgA and IgM)

Serves to link two monomer units of either IgM or IgA. In the case of IgM, the J chain-joined dimer is a nucleating unit for the IgM pentamer, and in the case of IgA it induces larger polymers. It also help to bind these immunoglobulins to secretory component.

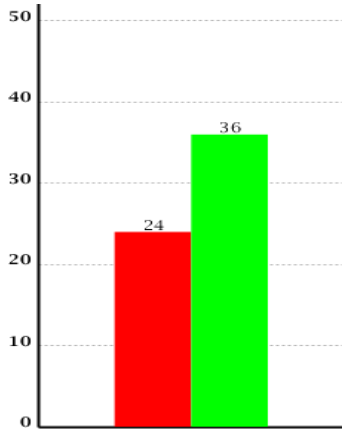
Angiogenin

Angiogenin binds to actin on the surface of endothelial cells; once bound, angiogenin is endocytosed and translocated to the nucleus. The protein encoded by this gene is an exceedingly potent mediator of new blood vessel formation. In addition, the mature peptide has antimicrobial activity against some bacteria and fungi, including *S. pneumoniae* and *C. albicans*.

FBLN5 (Fibulin-5)

The protein encoded by this gene is a secreted, extracellular matrix protein. It promotes adhesion of endothelial cells through interaction of integrins. Diseases associated with FBLN5 include Neuropathy, Hereditary, With Or Without Age-Related Macular Degeneration.

Comparative analysis of differentially expressed proteins between the analyzed groups (ASD vs. Control group) with increased expression



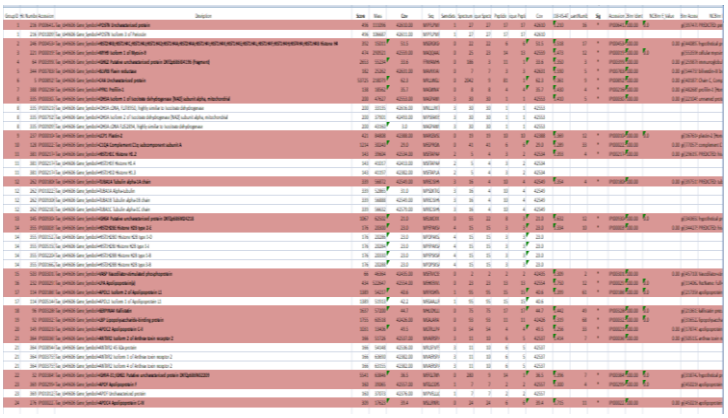


Figure 4. Differentially Expressed Protein Statistics X-axis: names of comparable group; Y-axis: the number of differentially expressed protein. In red are indicated the number of proteins (24) showed increased expression.

Apolipoprotein C-II (APOC2)

The protein is secreted in plasma where it is a component of very low density lipoprotein. This protein activates the enzyme lipoprotein lipase, which hydrolyzes triglycerides and thus provides free fatty acids for cells

Apolipoprotein C4 (APOC4)

This gene encodes a lipid-binding protein belonging to the apolipoprotein gene family. The protein is thought to play a role in lipid metabolism. Polymorphisms in this gene may influence circulating lipid levels and may be associated with coronary artery disease risk.

APOF (Apolipoprotein F)

This protein forms complexes with lipoproteins and may be involved in transport and/or esterification of cholesterol.

Complement C4A

Non-enzymatic component of C3 and C5 convertases and thus essential for the propagation of the classical complement pathway. Among its related pathways are Immune response Lectin induced complement pathway and Complement Pathway.

Complement C1q Subcomponent Subunit A

C1q associates with the proenzymes C1r and C1s to yield C1, the first component of the serum complement system. Diseases associated with C1QA include C1q Deficiency and Immunodeficiency Due To a Classical Component Pathway.

POSTN (Periostin)

This gene encodes a secreted extracellular matrix protein that functions in tissue development and regeneration, including wound healing, and ventricular remodeling following myocardial infarction. The encoded protein binds to integrins to support adhesion and migration of epithelial cells.

Myosin-9

Cellular myosin that appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion. During cell spreading, plays an important role in cytoskeleton reorganization.

LCP1 (Plastin-2)

Plastins are a family of actin-binding proteins Plays a role in the activation of T-cells in response to costimulation through TCR/CD3 and CD2 or CD28. Modulates the cell surface expression of IL2RA/CD25 and CD69.

Annotation of Clusters of Orthologous Groups of proteins (COGs)

In order to extract maximum information from the rapidly accumulating data from the obtained proteomic data, it is necessary to classify all proteins in accordance with their homologous relationships. From the COG annotation attention is drawn to certain functional classes in which a high number of differentially expressed proteins is observed: Protein metabolism and posttranslational modification, chaperones (**O**), with 98 proteins involved, General functions (only predicted) (**R**), 73 proteins and Cytoskeleton (**Z**), 28 proteins. While in the other categories the number of differentially expressed proteins is significantly lower Energy metabolism (**C**), 14 proteins, Metabolism and transport of amino acids (**E**), 14 proteins, Metabolism and transport of carbohydrates (**G**), 16 proteins, Translation, the structure of ribosome and biogenesis (**J**), 10 proteins

Gene Ontology functional enrichment analysis for differentially expressed proteins

Gene Ontology (GO) is an international standardized gene functional classification system which offers a dynamic-updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component and biological process. The basic unit of GO is GO-term. Every GO-term belongs to a type of ontology. GO functional analysis provides GO functional classification annotation for DEGs as well as GO functional enrichment analysis for DEGs.

1. Cellular component. A cellular component is just that, a component of a cell, but with the proviso that it is part of some larger object; this may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer). 2. Molecular function. Molecular function describes activities, such as catalytic or binding activities, that occur at the molecular level. GO molecular function terms represent activities rather than the entities (molecules or complexes) that perform the actions, and do not specify where or when, or in what context, the action takes place. 3. Biological process. A biological process is series of events accomplished by one or more ordered assemblies of molecular functions. It can be difficult to distinguish between a biological process and a molecular function, but the general rule is that a process must have more than one distinct steps.

First, mapping all differentially expression genes to each term of Gene Ontology database (http://www.geneontology.org/) and calculating the gene numbers each GO term has. We get a gene list and gene numbers for every certain GO term, then using hypergeometric test to find significantly enriched GO terms in DEGs comparing to the genome background.

In the corresponding GO aspect cellular component impressive is the concentration of differentially expressed genes in the elements such as the extracellular regions of 10.40% and a cell membrane 8.80% while others remain poorly represented. In the GO aspect, molecular function impressive is the concentration of differentially expressed proteins in elements such as binding 49.30%, in which element is concentrated approximately 50% of the differentially expressed proteins. Other elements with a concentration of differentially expressed proteins are the elements: activity 24.23%, catalytic activity 24.23%, and regulation of enzyme activity 7.34%. In the GO aspect, biological process, the concentration of differentially expressed proteins is observed in elements such as metabolic processes 8.35%, regulation of biological processes 7.53%, organization of cellular processes and biogenesis 4.35%, cellular processes 9.71% and others.

Identification of differentially expressed small RNA molecules of peripheral blood in ASD

The analysis of the results obtained from the large-scale expression analysis of small RNAs allowed the detection of specific changes in gene expression in ASD by helping to functionally differentiate key regulatory pathways and networks.

Standard bioinformatics analysis of small RNA showing coincidence with the reference genome

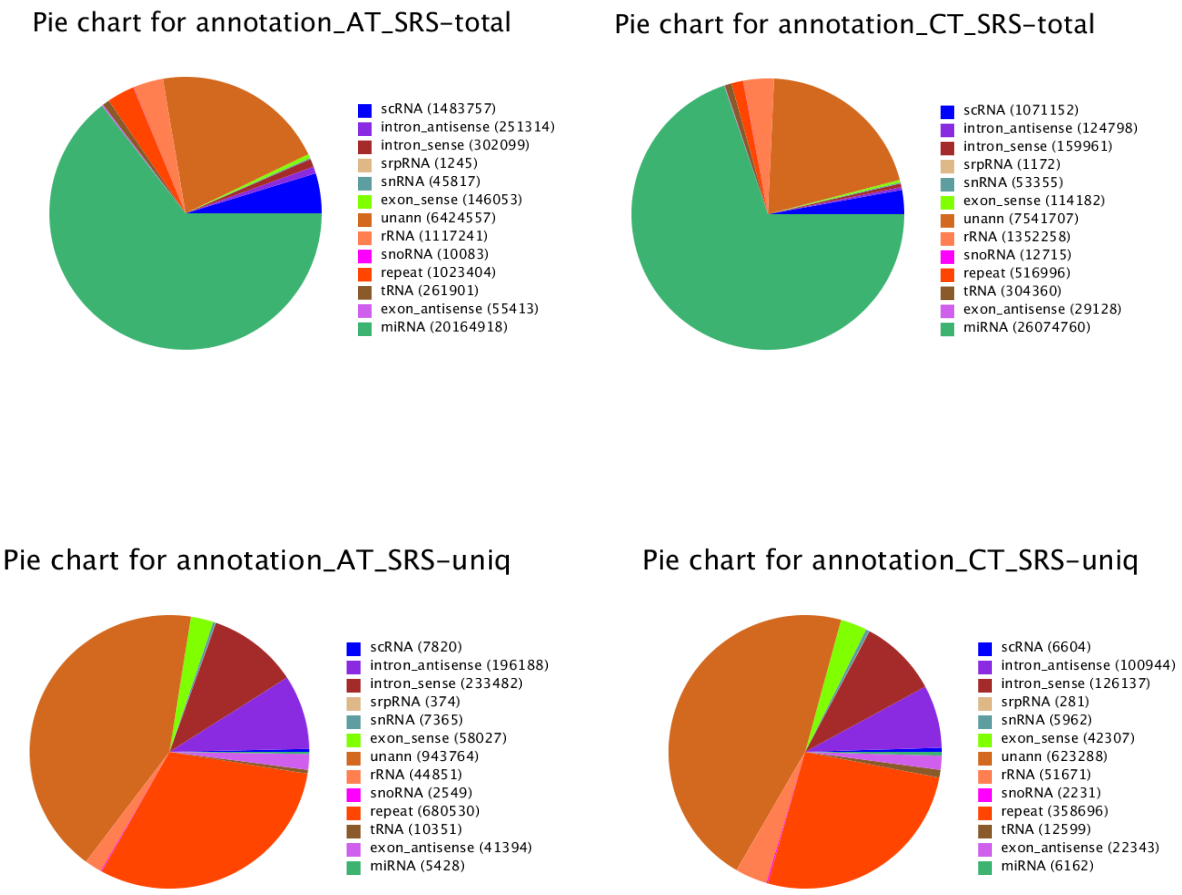


Figure 5. Distribution of annotated small RNA. The whole blood miRNAs signatures identified by Illumina Hiseq2000 sequencing. RNA species in Control group and ASD samples.

Identification of differentially expressed annotated miRNA molecules in the analyzed groups by small RNA sequencing

Differentially expressed miRNA molecules in ASD

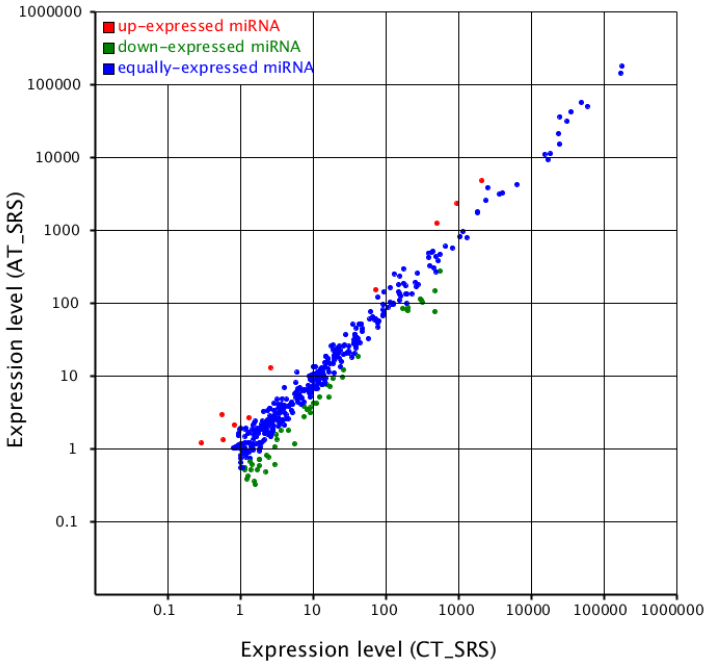


Figure 6. Scatter plot represent the expression profiles of differentially expressed miRNAs between Control and ASD samples. Each point in the figure represents a miRNA. The X and Y axis represent expression level of miRNAs in Control and ASD samples, respectively. The numbers of differentially expressed miRNAs Control and ASD samples. Up regulated (red) and down-regulated (green) miRNAs are summarized. Red points represent miRNAs with ratio>2; Blue points represent miRNAs with 1/2<ratio<2; Green points represent miRNAs with ratio<1/2. Ratio =normalized expression of the ASD sample/normalized expression of the Control sample.

Differentially expressed miRNA molecules in ASD

Small RNA expression profile in ASD compared with a control group of children in peripheral blood shows the presence of specific changes characteristic of the disorder. All this clearly demonstrates the potential of the method and the use of peripheral blood for the purpose of identifying potential biomarkers in ASD. Following the expression analysis, a significant number of miRNA molecules with statistically significant differential expression were identified as follows: let-7i-3p, miR-103a-2-5p, miR-106b-5p, miR-1249, miR-1307-5p, miR-134-5p, miR-139-5p, miR-142-3p, miR-145-5p, miR-15a-5p, miR-18b-3p, miR-193b-3p, miR-197-5p, miR-20b-3p, miR-210-5p, miR-29c-5p, miR-301a-3p, miR-3064-5p, miR-3620-3p, miR-365a-3p, miR-

378d, miR-487b-3p, miR-3909, miR-424-5p, miR-4707-3p, miR-500a-5p, miR-500b-5p, miR-5010-3p, miR-501-5p, miR-5187-5p, miR-550a-5p, miR-5690, miR-584-3p, miR-589-3p, miR-619-5p, miR-664a-3p, miR-664b-3p, miR-671-3p, miR-6799-3p miR-6850-5p, miR-96-5p, miR-183-5p, miR-199a-5p, miR-4734, miR-6849-3p, miR-223-5p, miR-3135a, miR-328-3p, miR-3674, miR-3687, miR-4489, miR-504-5p, miR-576-5p, miR-8052 *и* miR-937- 3p).

The present study provides evidence that peripheral blood is a suitable and readily available tissue (material) to examine the dynamics in the expression of candidate biomarkers in ASD.

Individual expression analysis of serum miRNAs in children with ASD

In order to identify serum circulating miRNAs related to the development of ASD, an initial small RNA-seq experiment (data not published), comparing the target miRNA profiles of healthy and ASD-affected children, was followed by qRT-PCR verification of the best candidates in pooled samples from the same groups. Relative expression levels were determined using, Cel-miR-39 as an exogenous control used to calculate changes in the miRNA expression using $2^{-\Delta\Delta Ct}$ method. In the second study, a total of 42 miRNAs were assessed, of which 29 were found to be downregulated in ASD-patients, 11 were upregulated and 2 failed to demonstrate expression changes. However, pooling can significantly reduce the sensitivity of the approach and provides information only for the average value for the population sample, which masks the variation. Thus, a confirmation step including an evaluation of the expression of samples from individual patients is advisable. The four most negatively dysregulated miRNAs, namely miR-500a-5p, miR-197-5p, miR-424-5p and miR-664a-3p, were selected for such an analysis. Their expression was measured in 38 children diagnosed with ASD and 28 healthy controls by stem-loop quantitative real-time PCR (qRT-PCR). **Figure 7** presents box-plot diagrams of their Ct-value distributions relative to the endogenous controls.

As it can be seen in **Figure 7**, the same tendency for ASD-related downregulation is manifested for all investigated miRNAs. On average, the difference of the Ct-values for miR-424-5p from the exogenous Cel-miR-39 used for normalization (delta Ct) was around 1.6 cycles higher in ASD-samples than in the respective controls (**Figure 7A**). Similarly, the same parameter is 1.43 cycles increased for miR-500a-5p (**Figure 7B**); 0.86 cycles for miR-197-5p (**Figure 7C**) and 0.67 cycles for miR-664a-3p (**Figure 7D**). Subsequent analysis by a nonparametric Mann-Whitney U test demonstrated that the observed reduction of the miRNAs expression is statistically significant for all of the studied miRNAs, with the respective p-values as follows: $p < 0.0001$ for miR-424-5p; $p < 0.0001$ for miR-197-5p; $p < 0.0001$ for miR-500a-5p and $p < 0.0001$ for miR-664a-3p. The same experimental pipeline was applied also for miR-365a-3p, which was one of the upregulated miRNAs in our previous tests with pooled samples. However, we were not able to reproduce the initial results (data not shown) and therefore miR-365a-3p was omitted from further analyses.

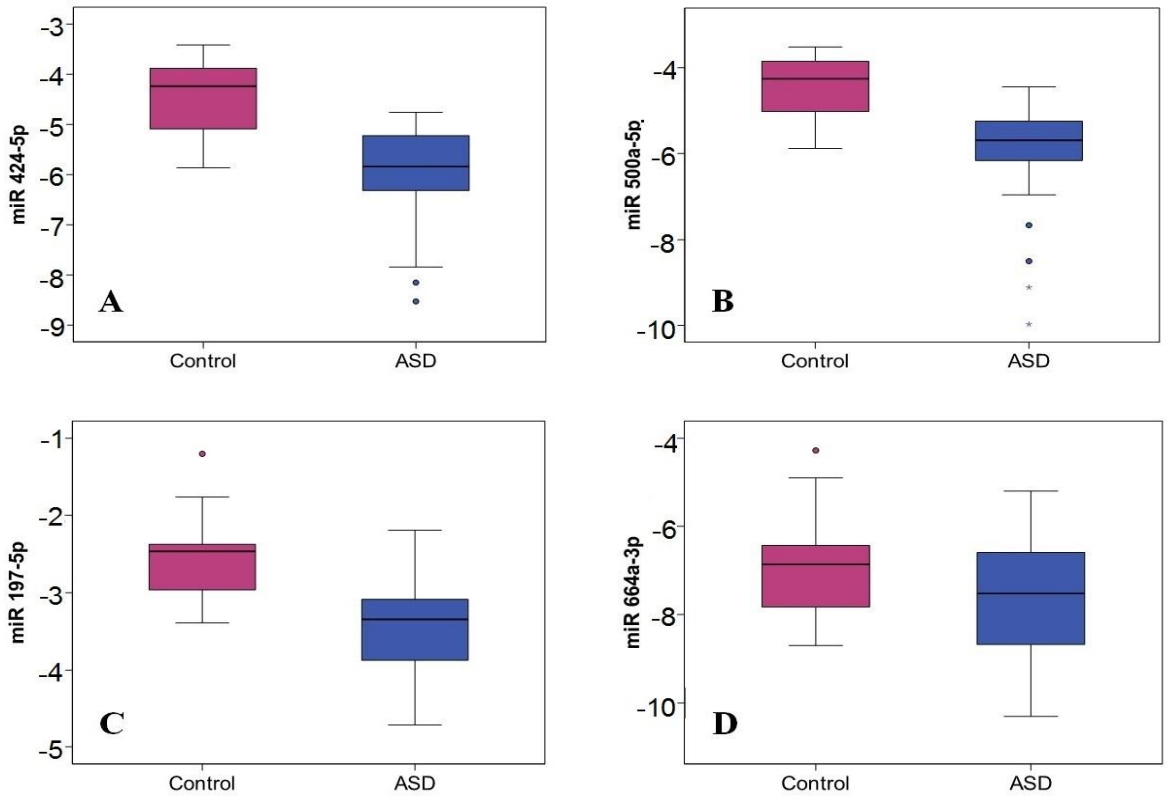


Figure 7. Differential expression of serum miRNAs in ASD patients. MiRNA specific stem-loop quantitative RT-PCR analysis of circulating miR-424-5p (Panel A), miR-500a-5p (Panel B), miR-197-5p (Panel C) and miR-664a-3p (Panel D) levels was conducted in case (n = 38) and control groups (n = 28) individually for each proband. Expression levels were normalized to spiked-in Cel-miR-39 control. The y-axis on the box plot denotes the differences of the Ct cycle for each miRNA from the Cel-miR-39 control, with lower difference, i.e. higher expression, closer to the top of the plot. The line represents the median value. Significance was calculated by a nonparametric Kolmogorov–Smirnov test. Outliers are plotted as individual dots.

Of the four remaining miRNAs, miR-424-5p and miR-500a-5p are characterized with the highest downregulation in ASD-patients, only 19,5% and 21,1% of their respective values in the controls (**Figure 8**). In turn, miR-197-5p is around two times (46% on average) less expressed than in healthy children, while for miR-664a-3p this effect is less pronounced - 60%. The observed expression pattern appears to be fairly stable across the different individuals, as indicated by the small variability of the data (the SD error bars). Only in the case of miR-664a-3p is observed a wider range of the relative quantity distribution.

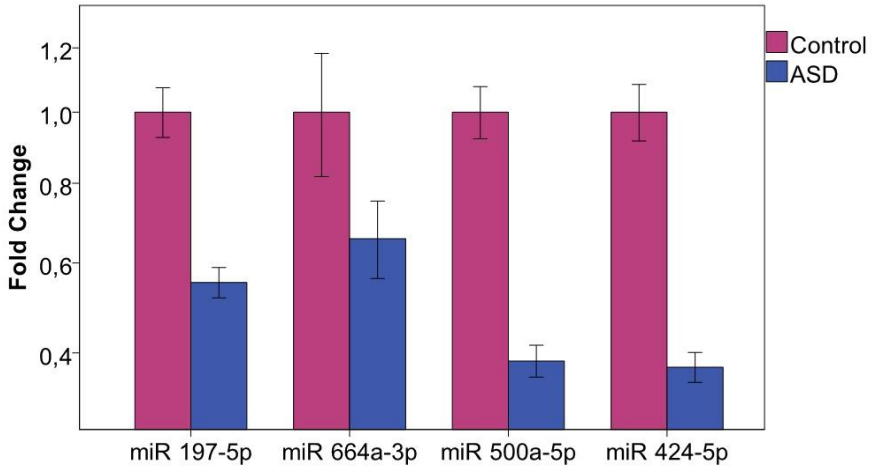


Figure 8. Fold change difference of the four investigated miRNAs between the ASD and control groups. Data are expressed as fold change of mean $2^{-\Delta\Delta Ct}$ for each miRNA after being normalized with spike-in Cel-miR-39 control. Measurements are carried out in duplicates ± SD.

In other studies performed on serum samples in ASD, the other three miRNA molecules showed differential expression from NGS data, miRNA-3135a, miRNA-328-3p, and miRNA-619-5p. Expression analysis of serum miRNA-619-5p **(Figure 11)**, miRNA-3135a, and miRNA-328-3p was performed using the miRNA-specific stem-loop qRT-PCR method. The data obtained demonstrate changes in the relative expression levels of miRNA-3135a and miRNA-328-3p, which are significantly lower in patients with ASD compared to the sample from the control group of healthy children. (**Figure 9**). The relative serum levels of investigated miRNA can differentiate ASD from healthy control patients (**Figure 10**).

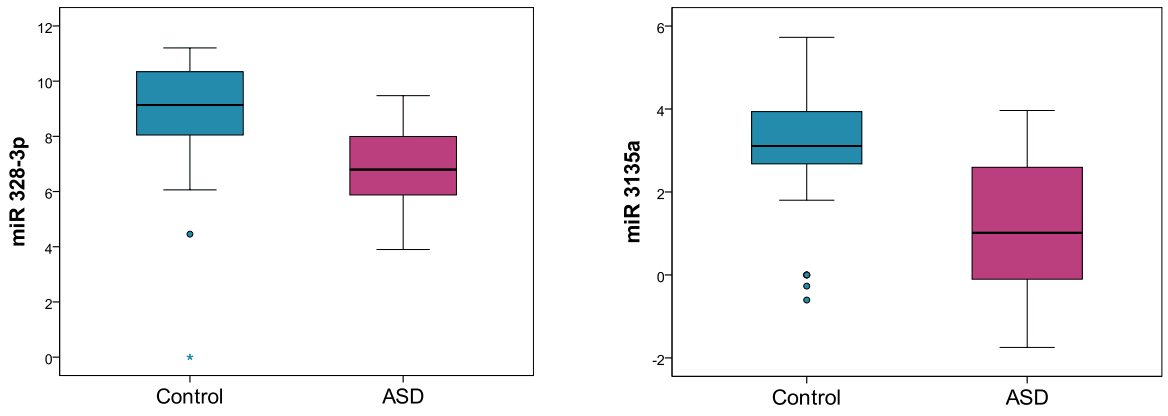


Figure 9. Differential expression of serum miRNAs in ASD patients. Quantitative RT-PCR analysis of miR-3135a and miR-328-3p levels. The circulating serum miRNAs signatures were identified by miRNA-specific stem-loop qRT-PCR analysis in the ASD and control groups. Expression levels of the analyzed miRNAs were normalized to spiked-in Cel-miR-39 control and expressed in relation to controls.

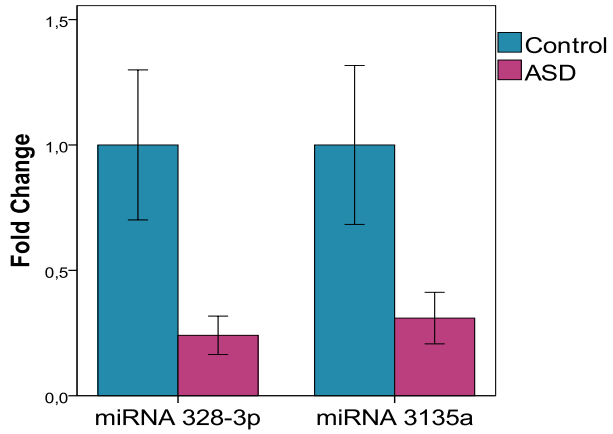


Figure 10. Fold change difference of two down-regulated serum miRNAs between the ASD and Control group. Data are expressed as fold change of mean $2^{-\Delta\Delta Ct}$ for each miRNA after being normalized with spike-in Cel-miR-39 control.

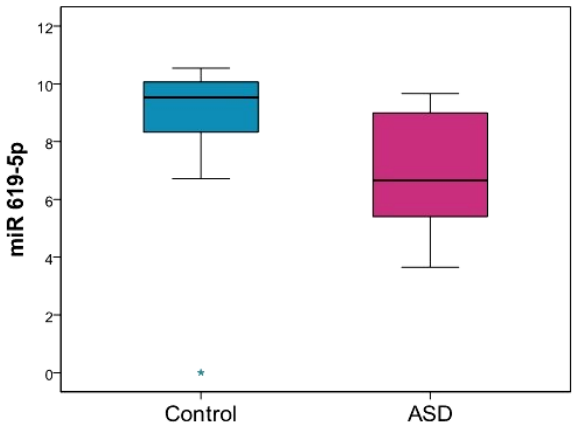
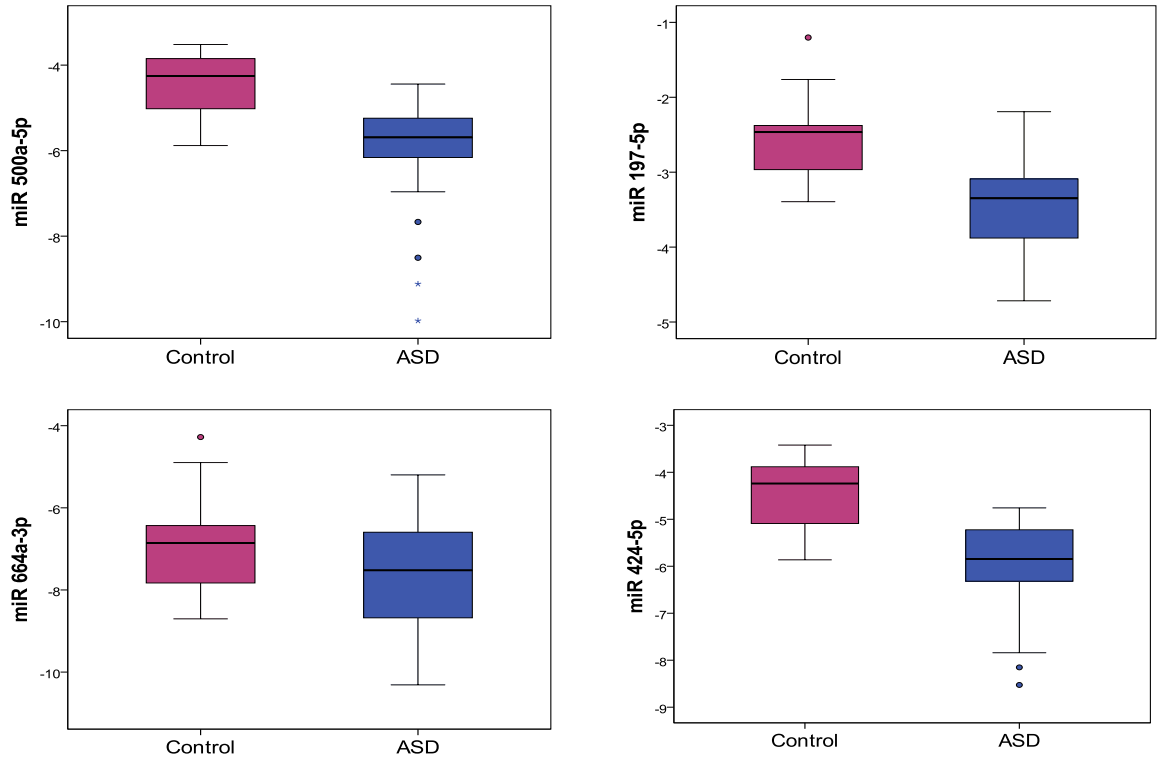


Figure 11. Differential expression of serum miR-619-5p in ASD patients. Expression levels of the analyzed miR-619-5p were normalized to spiked-in Cel-miR-39 control and expressed in relation to controls.

Individual expression profile of miRNAs in peripheral blood in children with ASD

Micro RNA-specific stem-loop qRT-PCR analysis was also used to study the expression of candidate miRNA molecules in peripheral blood samples in ASD. Relative expression levels were determined using U6 as an endogenous control used to calculate changes in miRNA expression using the $2^{-\Delta\Delta Ct}$ method. A statistically significant change in peripheral blood expression was found for miRNA-424-5p and



miRNA-500a-5p.

Figure 12. Differential expression of serum miRNAs in ASD patients. MiRNA specific stem-loop quantitative RT-PCR analysis of circulating miR-424-5p, miR-500a-5p, miR-197-5p and miR-664a-3p levels was conducted individually for each proband. Expression levels were normalized to U6 as an endogenous control. Outliers are plotted as individual dots.

Involvement of the studied serum miRNAs in relevant biological processes

In all the 162 genes denoted as unique, validated targets in miRWalk, 71 appeared to participate in biological pathways described in the KEGG database. One of the target genes we obtained, the amyloid β-precursor protein (*APP*), is involved in synaptic pathways. This gene encodes a membrane protein that undergoes proteolytic processing. In serotonergic synapses, the soluble APP β-fragment interacts with the cyclic adenosine monophosphate (cAMP) signal transduction protein exchange factor directly activated by cAMP (*EPAC*) to promote

neuroprotection. Another gene, solute carrier family eight member A1 (*SLC8A1*), participates in a specific exteroceptive transduction pathway. The SLC8A1 protein is a Na^+/Ca^{2+} (K^+) antiporter. In olfactory neurons, it has a role in membrane repolarization and annihilates the consequences of a previously occurred action potential. Another six different target genes showed involvement in neurodegenerative diseases: Huntington's, Parkinson's or Alzheimer's. Two genes: *APP* and *BACE1* directly involved in pathogenesis of Alzheimer's disease. The amyloid precursor protein encoded by *APP* is cleaved by ◎-secretase (encoded by *BACE1*) that leaves a particular peptide responsible for amyloid plaque formation. Two proteins (DNAL1 and POLR2I) are described to take part in Huntington's disease pathogenesis in an indirect manner, while interacting with or being influenced by, the product of the *Huntingtin* (*HTT)* gene. As a result nonspecific changes in cytoskeleton organization or gene expression occur. Two other genes whose protein products take part in the mitochondrial respiratory chain NADH: ubiquinone oxidoreductase subunit A1 (*NDUFA1)* and NADH: ubiquinone oxidoreductase subunit V3 (*NDUFV3*) are described to be important not only for Alzheimer's but also for Huntington's and Parkinson's diseases. A correlation between the Alzheimer's syndrome and the reduced expression of energy metabolism genes has been well established [Lang et al., 2008]. However, the precise mechanisms in which *NDUFA1* and *NDUFV3* contribute to a specific neurodevelopment condition and their regulatory roles is yet to be clarified. The mRNA targets of the studied miRNAs were obtained by the database miRWalk, their association with relevant biological pathways is presented in **Figure 13.** The search in the MirWalk database identified 671 unique validated target genes for all of the four differentially expressed miRNAs (35 of which are regulated by more than one miRNA). 287 of these validated targets were found to participate in 478 different biological pathways described in the KEGG database. Most of the obtained mRNA partners are related to signal transduction, metabolism and cancer diseases. However, 19 validated target genes are shown to participate in synaptic pathways (in cholinergic, dopaminergic, GABAergic or glutamatergic synapses). Among them are four separate genes that encode G-protein subunits and thus play an important role in intracellular signal integration: *gnal*, *gnaq*, *gnb1*, and *gng12*. Additional three genes (*grin2b*, *gabrg2*, and *gabarapl1*) code for subunits of receptors for neuroactive ligands or receptor-associated proteins. Finally, two other targets participate in glutamate reuptake: *slc1a2*, *slc1a1*.Despite miR-664a-3p has numerous validated target genes, their functional profile is dispersed in a large array of biological processes not directly related to the CNS and known key aspects of ASD behavior. Somewhat similar is the situation with miR-197-5p, with the distinction that the panel of known affected mRNAs is smaller. Moreover, miR-197-5p actually scores 2 hits in Alzheimer's disease category. In contrast, miR-424-5p and miR-500a-5p have a much more pronounced involvement in nerve tissue-specific pathways, including enrichment of target genes affecting different kinds of synapses, axon guidance, neuroactive ligand-receptor interactions and some pathologies (**Figure 13**).

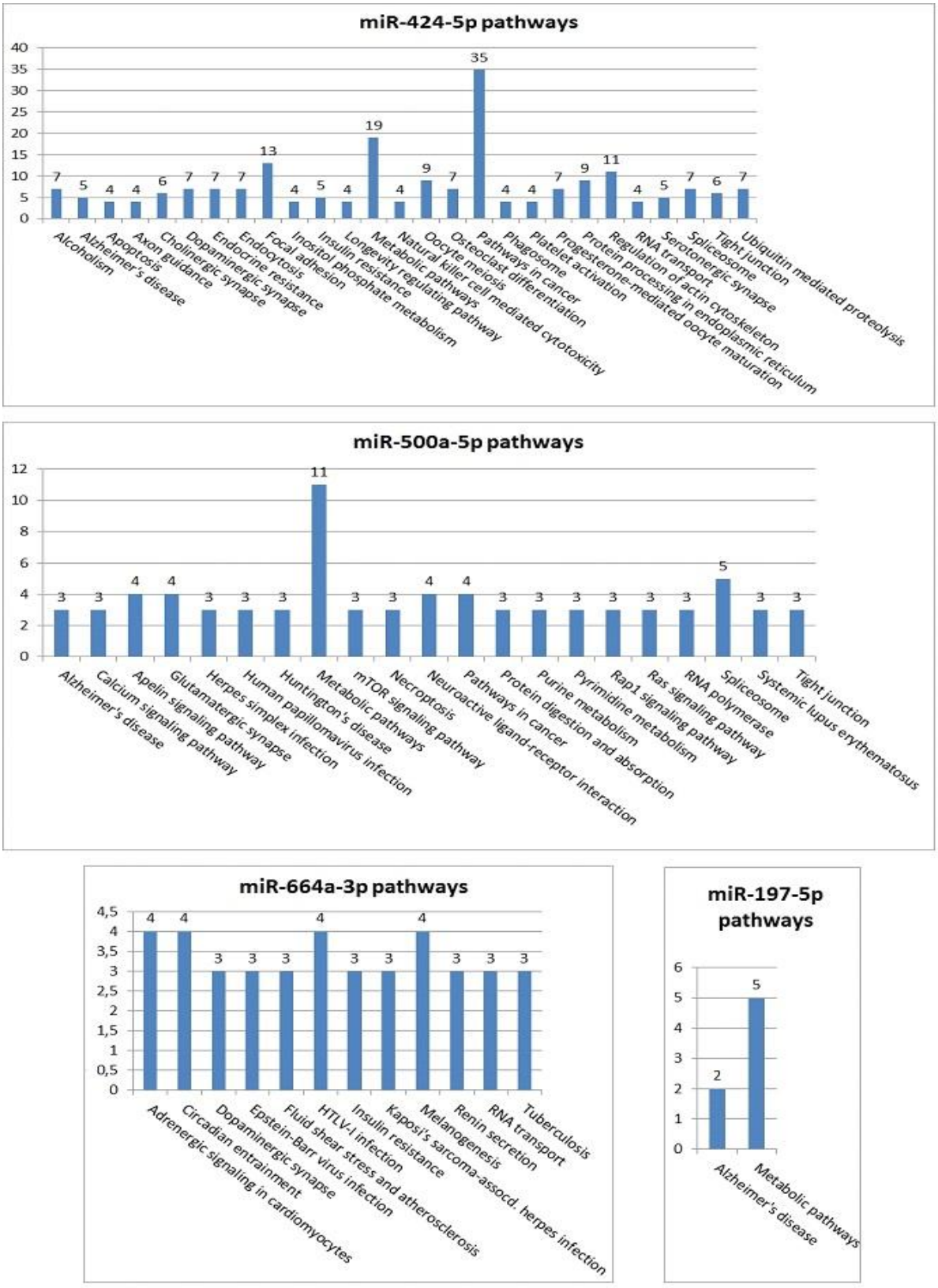


Figure 13. Enrichment of biological pathways in the mRNA target pools of the four studied miRNAs: miR-424-5p; miR-500a-5p; miR-664a-3p and miR-197-5p.

In recent years, there is growing evidence to support the hypothesis that dysregulation of the regulatory network linking epigenetic and miRNA-mediated regulation may significantly contribute to the initiation and development of neurodevelopmental disorders. The interest of the scientific community in the development of diagnostic procedures based on circulating miRNAs is justified by the many advantages that this approach has: availability of the tested tissues, low complexity and high stability of the target molecules, speed of the experimental procedure, relatively low cost, the potential for automation, advances in identification and quantitative analysis techniques, etc. However, some issues need to be considered when developing biomarkers based on miRNA molecules.

First, the miRNA regulatory network is extremely complex, where a single miRNA molecule is able to coordinate the expression of hundreds of genes, while the opposite is also true, where a single gene can be controlled by multiple miRNAs. Therefore, the dysregulation of certain miRNAs and their binding to certain biological processes in the brain is very challenging. In addition, the plasticity of the miRNA system makes it highly unlikely that major miRNA switches will exist, whose normal function will be disrupted in all or almost all cases of a pathological event. Therefore, it is advisable to examine several miRNAs simultaneously to improve the diagnostic ability of the developed tests for candidate ASD biomarkers. Another obstacle may be the low amount of extracellular miRNAs in body fluids, which on the one hand can affect the reproducibility of the obtained results and on the other hand, require a highly efficient extraction method.

Identification of differentially expressed protein-coding genes in the analyzed groups by RNA-Sequencing

To identify differentially expressed genes, the p-value was used as an indicator of differential gene expression, the FDR value (False Discovery Rate), as a method for determining the threshold of the p-value and the absolute value of the ratio log2 ≥ 1, as a threshold for estimating the difference in gene expression. All RNA molecules (transcripts) analyzed in the transcriptome analysis of children with ASD were compared with those of the control group. As a result, twenty-two RNA transcripts showed differential expression with statistical significance $p < 0.001$, FDR ≤ 0.001, and log2 ≥ 1 were identified.

Gene	Regulation	Fold Change	P-value	False Discovery Rate
SLC8A2	Up-regulation	5.18	2.7869e-10	2.517354515625e-08
ANKRD22	Up-regulation	1.78	2.7132e-11	2.90463133333333e-09
SCAND1	Up-regulation	1.66	9.2933e-06	0.000363823254853273
RAP1GAP	Up-regulation	1.43	6.07386e-10	5.2147997019802e-08
DPM3	Up-regulation	1.23	1.31106e-05	0.000497542966739606
CRB2	Up-regulation	1.11	3.15268e-06	0.00013703491037594
TPX2	Up-regulation	1.10	2.43966e-07	1.35178988434505e-05
HLA-DQA2	Up-regulation	1.06	3.29222e-11	3.50288168466258e-09
CD177	Up-regulation	1.05	2.71846e-07	1.49196999303797e-05
LTF	Up-regulation	1.00	1.533878e-16	2.41836783218182e-14
TM4SF1	Down-regulation	-7.31	5.80188e-11	5.78287384137931e-09
MAGEA4	Down-regulation	-6.26	1.077206e-05	0.0004151551924
FLNC	Down-regulation	-5.22	1.34649e-13	1.74269970671642e-11
IGF2BP1	Down-regulation	-4.41	9.62384e-08	5.75538817655172e-06
FOSL1	Down-regulation	-2.13	1.387236e-08	9.94166692066116e-07
OTOF	Down-regulation	-1.76	2.35062e-05	0.000816969993186373
XIST	Down-regulation	-1.38	4.07256e-16	6.25047859115044e-14
TRIM17	Down-regulation	-1.31	1.145212e-06	5.33908917096774e-05
TMEM40	Down-regulation	-1.22	2.79288e-08	1.8417079026616e-06
SERPINE1	Down-regulation	-1.20	4.7984e-07	2.46939617804154e-05
BIVM	Down-regulation	-1.13	1.9136e-05	0.000687113142857143
BCYRN1	Down-regulation	-1.07	4.79516e-06	0.000199909759326923

Table 1. Differentially expressed genes between the analyzed groups obtained from RNA transcriptome analysis.

Gene ontology analysis of differentially expressed genes in ASD

To categorize the identified differentially expressed genes, gene ontology analysis was used, as an international standardized classification for gene function, which offers a controlled and dynamically-updated database with strictly defined properties of genes and their products.

Functional classification of differentially expressed genes in the gene ontology aspect cellular component

Three of the analyzed protein-encoding genes are associated with function in a specific part of the cell: FLNC gene with decreased expression and TPX2, and DPM3 with increased expression. One gene showed a link to the extracellular region - SERPINE1, with reduced expression. Four of the genes are involved in macromolecular protein complexes: HLA-DQA2, TPX2, and DPM3 with increased expression; in ribonucleoprotein complex - IGF2BP1 with reduced expression. Two of the genes are associated with the cell membrane: TM4SF1 with reduced expression and DPM3 with increased expression. Three genes are associated with organelles: cytoskeleton - FLNC with decreased expression and TPX2 with increased expression; endoplasmic reticulum - DPM3 with increased expression.

Functional classification of differentially expressed genes in the gene ontology aspect molecular function

The molecular function of six of the differentially expressed genes is associated with - "binding" - the binding process (selective, non-covalent, often stoichiometric, the interaction of the molecule with one or more specific sites of another molecule) - FOSL1 gene; DNA binding and transcription factor activity; IGF2BP1 binding to RNA; FLNC and TPX2 binding to cytoskeletal proteins; CRB2 - showed receptor activity. Seven genes with catalytic function: - regulatory activity - SERPINE1, IGF2BP1, and RAP1GAP; hydrolase activity - SERPINE1, ANKRD22, and LTF; ligase activity TRIM17; transferase activity - DPM3. Two genes have a molecular function as enzyme regulators - SERPINE1 (inhibitor of enzyme activity) and RAP1GAP (regulates the activity of small GTPase). The molecular function of a gene - FOSL1, as a transcription factor. Two other genes have a molecular function as a structural molecule (the action of a molecule that contributes to the structural integrity of a complex inside or outside the cell) - FLNC and TPX2 genes. SLC8A2 is associated with transmembrane transport activity.

Functional classification of differentially expressed genes in the gene ontology aspect biological process

Two of the genes are associated with apoptosis-inducing processes - IGF2BP1 and MAGEA4, both with reduced expression. In processes related to biological regulation are SERPINE1, FOSL1 with reduced expression and RAP1GAP, and CRB2

with increased. FLNC is involved in the organization of cellular components and has reduced expression. Nine genes are associated with involvement in cellular processes - SLC8A2, RAP1GAP, CRB2, DPM3, and TPX2 with increased expression and FOSL1, MAGEA4, FLNC, and IGF2BP1 with reduced expression.Four genes are involved in developmental processes - MAGEA4, FLNC, and IGF2BP1 with decreased expression and CRB2 with increased. Gene ontology analysis showed an association of two genes with immune system processes - CRB2 and HLADQA-2 with increased expression. A link to the GO term localization (any process in which a cell, substance, or cell unit, such as a protein complex or organelle, is transported to or maintained at a specific site) has been shown for the genes SLC8A2, CRB2 and LTF with increased expression and IGF2BP1 with decreased. In metabolic processes - RAP1GAP, CRB2, ANKRD22, DPM3, and LTF, which are overexpressed and FOSL1, SERPINE1, IGF2BP1, and TRIM17 with decreased. IGF2BP1 and OTOF showed an association with processes characteristic of multicellular organisms. In stimulus-response processes, association shows the IGF2BP1 gene, which shows reduced expression.

Performance of an ASD-prediction model using the four studied serum miRNAs

In order to evaluate the strength of miR-424-5p, miR-500a-5p, miR-197-5p and miR-664-3p as potential serum biomarkers for ASD, a Receiver Operating Characteristic (ROC) curve analysis was performed with the data from the qRT-PCR experiment. Subsequently, the area under the ROC curve (AUC), as well as the diagnostic sensitivity and specificity of each serum miRNA, were calculated. The expression data of the four serum miRNAs described in this study were simultaneously taken into account to conduct a combined ROC analysis. When calculated separately for each serum miRNA, the diagnostic sensitivities are as follows: 86.1% for miR-197-5p; 77.8% for miR-500a-5p; 25% for miR-664-3p and 88.9% for miR-424-5p. In turn, the corresponding specificities are 78.6%, 92.9%, 100% and 75%, respectively. The AUC parameters for miR-197-5p, miR-500a-5p and miR-424-5p are 0.825, 0.796 and 0.756, respectively, while for miR-664-3p the AUC is below 0.7. According to these results, miR-197-5p, miR-500a-5p and miR-424-5p are good ASD predictors, with sufficient sensitivities (i.e. true positive rates) of near or over 80%, and similarly favourable specificities (i.e. true negative rates) and AUC values (i.e. accuracy). MiR-664-3p is characterized with a worse diagnostic performance since it has a much lower probability of ASD detection – only 25% sensitivity and an AUC < 0.7. However, the expression pattern of this miRNA in serum can be used to accurately rule out the possibility for ASD in healthy individuals, due to the extremely high (100%) specificity. Integrating the results for the miR-197-5p and miR-500a-5p markers increased the AUC to 0.975. The sensitivity and specificity values of the new refined model were also impressive since both of them surpassed 90%. However, miR-424-5p and miR-664-3p expression did not improve the fit of this model and thus they were excluded. Together, these findings indicate that when

considered alone, three of the selected serum miRNAs, namely miR-197-5p, miR-500a-5p, and miR-424-5p can discriminate between ASD cases and control cases with a sufficiently high accuracy. The best diagnostic power for ASD is achieved by a model which is built with the combination of miR-197-5p and miR-500a-5p expression data. In another study of serum samples in ASD, two other miRNA molecules miRNA-3135a and miRNA-328-3p) were subjected to expression analysis, showing differential expression from NGS data. The ROC analysis of the expression data showed good diagnostic characteristics of the studied predictors. The ombined ROC analysis (LOGREGR) of the differentially expressed miRNA molecules (miRNA-3135a and miRNA-328-3p) showed a diagnostic sensitivity of the combined classifiers of 78.9%, with a corresponding specificity of 88.9%. From the obtained data it is clear that the complex characterization of several miRNA molecules as potential biomarkers in ASD shows better diagnostic characteristics compared to the use of single ones. Additionally, two other candidate miRNAs were subjected to ROC analysis, miRNA-619-5p, and miRNA-365a-3p. The conducted ROC analysis of miRNA-619-5p showed good diagnostic sensitivity of 100%, but with a corresponding low specificity of 55.6%.. However, analysis of miRNA-365a-3p did not show satisfactory results.

Performance of an ASD-prediction model (receiver operating characteristic) using the four studied miRNAs in peripheral blood of ASD patients

In order to evaluate the strength of miR-424-5p, miR-500a-5p, miR-197-5p and miR-664-3p as potential peripheral blood-based biomarkers for ASD, a Receiver Operating Characteristic (ROC) curve analysis was performed. The conducted ROC analysis of miR-500a-5p showed good diagnostic sensitivity of 77.8%, with a corresponding specificity of 92.9%. ROC analysis of miR-197-5p diagnostic sensitivity of 86.1% with a corresponding specificity of 78.6%. The diagnostic sensitivities are as follows: 88.9% for miR-424-5p; 25.0% for miR-664-3p. In turn, the corresponding specificities are 75.0% and 100%, respectively.

Confirmation of differentially expressed genes from the transcriptomic (RNA-Seq) analysis by quantitative Real-Time PCR analysis

Data validation from the transcriptome analysis was performed using quantitative Real-Time PCR analysis. Individual quantitative Real-Time PCR analysis was subjected to the top 10 differentially expressed genes (FDR and p-value) in all samples from children diagnosed with ASD and a control group of healthy children. The change in gene expression was analyzed using the $2^{-\Delta\Delta Ct}$ method.

Result of quantitative expression analysis (RT-qPCR) of silencing genes (AGO2, AGO3, AGO4, and Drosha) in ASD

Interest in the present work was also focused on conducting a comparative expression analysis of genes involved in the process of miRNA biogenesis in children with ASD.

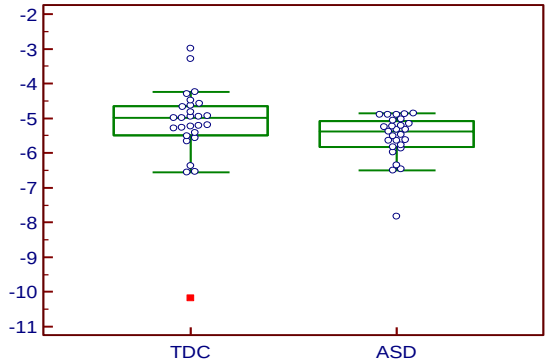


Figure 14. Box plot diagram showing differential expression of Ago2 gene in ASD patients. The performed statistical analysis showed decreased expression of the Ago2 gene in ASD, but without statistically significant change Ago2 P = 0,2342.

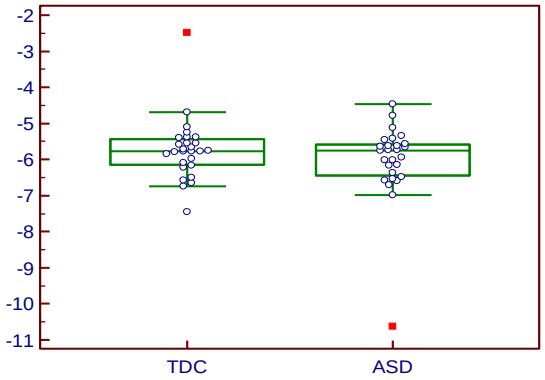


Figure 15. Box plot diagram showing differential expression of Ago3 gene in ASD patients. The performed statistical analysis showed decreased expression of the Ago3 gene in ASD, but without statistically significant change, *P* = 0,2594.

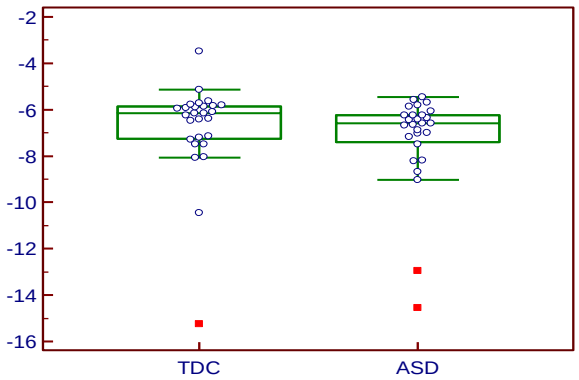


Figure 16. Box plot diagram showing differential expression of Ago4 gene in ASD patients. The performed statistical analysis showed decreased expression of the Ago4 gene in ASD, but without statistically significant change, *P* = 0,4127.

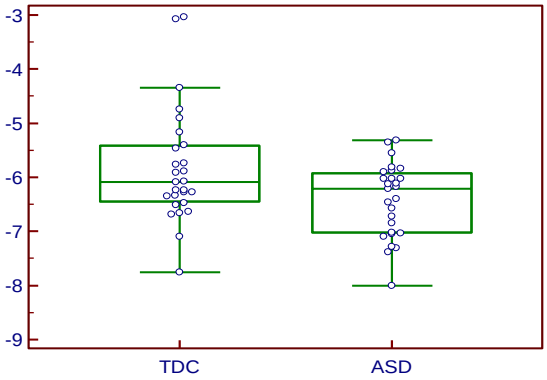


Figure 17. Box plot diagram showing differential expression of Drosha gene in ASD patients. The conducted statistical analysis showed decreased expression of Drosha in ASD, with a statistically significant difference p = 0,0171.

KEGG analysis of biological pathways enriched in differentially expressed genes in ASD

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is the main publicly available database related to data on biological pathways and gene interactions. Analysis of biological pathways involving differentially expressed genes in ASD identifies metabolic and signaling pathways associated with differentially expressed genes. The **Table 2** presents the biological pathways of statistical significance.

#	Pathway	DEGs (17)	*P*-value	Pathway ID
1	Calcium signaling pathway	3 (17.65%)	0.002537809	ko04020
2	Amphetamine addiction	2 (11.76%)	0.007751596	ko05031
3	Leishmaniasis	2 (11.76%)	0.007983234	ko05140
4	GABAergic synapse	2 (11.76%)	0.008100223	ko04727
5	Retrograde endocannabinoid signaling	2 (11.76%)	0.008575949	ko04723
6	MAPK signaling pathway	3 (17.65%)	0.009102987	ko04010
7	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	2 (11.76%)	0.01196498	ko05412
8	Cholinergic synapse	2 (11.76%)	0.01253102	ko04725
9	GnRH signaling pathway	2 (11.76%)	0.01506537	ko04912
10	Glutamatergic synapse	2 (11.76%)	0.01553389	ko04724
11	Serotonergic synapse	2 (11.76%)	0.01813789	ko04726
12	Dopaminergic synapse	2 (11.76%)	0.01864557	ko04728
13	Steroid biosynthesis	1 (5.88%)	0.02947291	ko00100
14	Wnt signaling pathway	2 (11.76%)	0.0308568	ko04310
15	Leukocyte transendothelial migration	2 (11.76%)	0.03652306	ko04670
16	Hypertrophic cardiomyopathy (HCM)	2 (11.76%)	0.04232472	ko05410
17	Dilated cardiomyopathy	2 (11.76%)	0.04449439	ko05414
18	Asthma	1 (5.88%)	0.04538188	ko05310
19	Influenza A	2 (11.76%)	0.04547207	ko05164
20	Carbohydrate digestion and absorption	2 (11.76%)	0.04895749	ko04973
21	Endocrine and other factor-regulated calcium reabsorption	2 (11.76%)	0.04921018	ko04961

Table 2. Biological pathways involving differentially expressed genes in ASD.

Summary for the award of the scientific degree "DOCTOR OF SCIENCES"

Identification of a novel mitochondrial mutation in the cytochrome c oxidase III gene in children with ASD using next generation RNA-sequencing

The recent advances in the molecular genetics field are related to the unprecedented progress in the development and implementation of novel approaches for next-generation sequencing of nucleic acids. One of the many applications of these cutting-edge technologies is the opportunity for identification of a substantial number of single nucleotide polymorphisms (SNPs) from peripheral blood samples. This can be achieved by the utilization of transcriptomic data, which are characterized with high sensitivity and specificity as a function of the number of reads obtained from a particular genomic region. The analytical procedure presented in this work is suitable for identification of novel DNA-variants in diseases and disorders in which heterogeneity in certain DNA or RNA loci is observed. A good example thereof is the illustrated mutation in the mitochondrial DNA.

Several selected target loci of interest, annotated as missense substitutions, were subsequently visualized with Integrated Genome Viewer – IGV. The most serious candidate with the highest detection quality (QUAL = 13461.3) that emerged following this procedure was the chrM:9214 locus. The identified mutation leads to the substitution of histidine with arginine at the third position of the molecule of the mitochondrially encoded protein CO3. The substation m. 9214 A>G is predominantly present in the pooled sample of ASD patients (in 22% of all mapped fragments). Overall, in the current study, we identified a point mutation in the 9214 position of the mitochondrial genome in children with ASD by using a full transcriptome sequencing method. The substation m. 9214 A>G is predominantly present in the pooled sample of ASD patients (in 22% of all mapped fragments), while in the sample from healthy children it is almost completely absent. The *in silico* analysis of the possible phenotype effect determined the mutation as potentially detrimental. These observations support a hypothesis that m. 9214 may have a putative role in the mechanisms of etiopathogenesis of single ASD cases.

Results from the transcriptome profiling of protein encoding genes in peripheral blood of schizophrenia.

Following a large-scale expression analysis (DGE), a significant number of differentially expressed genes were identified in schizophrenia. The genes showed decreased expression were 1012, while those showed increased were significantly more 2582 genes (**Figure 18** and **Figure 19**).

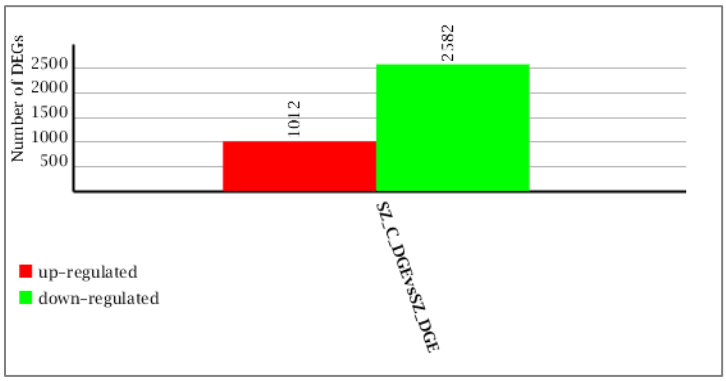


Figure 18. Expression profiles of differentially regulated genes between control and schizophrenia pooled samples. The numbers of differentially expressed genes samples. Up regulated (red) and down-regulated (green) genes are summarized.

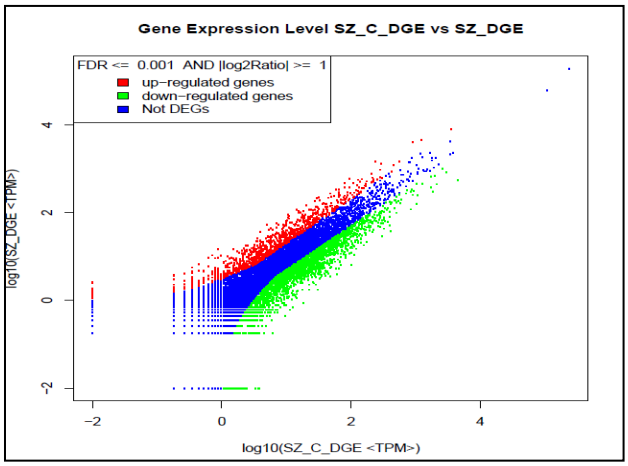


Figure 19. Differentially expressed genes in patients diagnosed with schizophrenia. Differentially expressed genes are presented in red (with increased expression) and green (with reduced expression). Genes that do not show changes in expression are listed in blue; DEG, differentiated expressed genes; TPM; tags per million; FDR, false discovery rate.

Up-regulated genes	Log2 Ratio	Down-regulated genes	Log2 Ratio
PCYT1A	5.45	SPON2	-3.12
SLC6A19	5.45	ANK3	-6.83
DRD4	6.30	DICER1	-1.98
		HTR2A	-6.72
		LZTS1	-5.12
		GRIN2D	-4.16

Table 3. Fold changes in the expression of protein coding genes (log2) in peripheral blood from patients with schizophrenia compared with healthy controls.

Dysregulated protein-coding genes in schizophrenia

Dopamine (DA) D4 receptor (DRD4)

The dopamine (DA) D4 receptor (DRD4) could play a role in mediating dopaminergic activity. The DRD4 is primarily expressed on pyramidal neurons and interneurons in the prefrontal cortex, but there is also support for DRD4 localization on medium spiny neurons in the basal ganglia (striatum, Str; and nucleus accumbens core, NAc), throughout the limbic system and in the thalamus of rodents. Dopamine receptor responsible for neuronal signaling in the mesolimbic system of the brain, an area of the brain that regulates emotion and complex behavior. Activated by dopamine, but also by epinephrine and norepinephrine, and by numerous synthetic agonists and drugs.

Fasciculation and elongation protein ζ-1 (FEZ1)

Fasciculation and elongation protein ζ-1 (FEZ1) is one of the first identified binding partners of Disrupted-in Schizophrenia 1 (DISC1), a susceptibility gene for major mental disorders including schizophrenia in yeast two-hybrid screen of an adult human brain library. Moreover, proteomic techniques revealed that the FEZ1 protein interacts with various intracellular partners, such as motor signaling, and structural proteins, one of which is DISC1.

Dicer 1, ribonuclease III (DICER1)

The DICER1 gene provides instructions for making a protein that plays a role in regulating the activity (expression) of other genes. The Dicer protein aids in the production of a molecule called microRNA (miRNA). MicroRNAs are short lengths of RNA, a chemical cousin of DNA. Dicer cuts (cleaves) precursor RNA molecules to produce miRNA. MicroRNAs control gene expression by blocking the process of

protein production. In the first step of making a protein from a gene, another type of RNA called messenger RNA (mRNA) is formed and acts as the blueprint for protein production.

ASTN2 astrotactin 2

This gene encodes a protein that is expressed in the brain and may function in neuronal migration, based on functional studies of the related astrotactin 1 gene in human and mouse. A deletion at this locus has been associated with schizophrenia.

CASP3 caspase 3, apoptosis-related cysteine peptidase

This gene encodes a protein that is a member of the cysteine-aspartic acid protease (caspase) family. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis.

CDT1 chromatin licensing and DNA replication factor 1

The protein encoded by this gene is involved in the formation of the pre-replication complex that is necessary for DNA replication. The encoded protein can bind geminin, which prevents replication and may function to prevent this protein from initiating replication at inappropriate origins. Phosphorylation of this protein by cyclin A-dependent kinases results in the degradation of the protein.

ANK3 ankyrin 3, node of Ranvier (ankyrin G)

Ankyrins are a family of proteins that are believed to link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and play key roles in activities such as cell motility, activation, proliferation, contact, and the maintenance of specialized membrane domains.

Performance of a schizophrenia-prediction model using the differentially expressed genes (Receiver Operating Characteristic analysis)

To evaluate the characteristics of the differentially expressed genes, DICER, FEZ1, DRD4, GRIN2 as potential biomarkers in schizophrenia, a ROC analysis was performed to determine the diagnostic accuracy of the studied protein-coding genes. The diagnostic sensitivity of the DICER, FEZ1, DRD4, and GRIN2D genes was determined at 72.4%, 55.2%, 86.2%, and 48.3%, respectively, with a corresponding specificity estimated at 83.3%, 87.5%, 91.7%, and 83.3%.

DISCUSSION

The research presented in the dissertation thesis is the first of its kind complex scientific study covering comparative genomic transcriptomic and proteomic studies performed on a clinically selected group of Bulgarian patients diagnosed with ASD compared to a control group of healthy individuals. One of the

biggest challenges in conducting genomic, transcriptome, and proteomic studies in neurodevelopmental disorders is the collection of samples from the primary target tissue, ie. brain. Due to the impossibility of performing brain tests *in vivo*, for this purpose, mainly retrospective studies with post-mortem brain samples are conducted. The reason for the choice of this type of material is justified and is expressed in the fact that diseases such as schizophrenia and ASD affect brain function and it would be most argued that any transcriptional and proteomic studies should be aimed at the affected tissues and involved in the disorders. Such an approach would be of great value in identifying etiopathogenetic factors in the development of disorders/diseases that currently have an unclear etiology.

Some miRNAs and other small RNAs show coexpression in peripheral blood and brain, as well as miRNAs present in brain tissues with higher expression but also expressed in peripheral blood. The present study aimed to characterize the global profile of peripheral blood miRNA expression, with a focus on the identification of specific miRNA molecules with characteristics of potential biomarkers in neurodevelopmental disorders (ASD and schizophrenia). In the present experimental work, the general expression profile of miRNA molecules in peripheral blood from patients with ASD was studied in detail. By using the small RNA sequencing, a change in the expression of a large number of miRNA molecules was detected. The approach used in the present work offers an opportunity to take a step in understanding the role of miRNA molecules and related biological processes in the etiopathogenesis of ASD.

Due to the presence of these complicating factors, a new emphasis in the study of biomarkers in schizophrenia and ASD is the approach based on the study of the molecular profile of biomolecules from peripheral tissues and especially miRNA molecules in peripheral blood, as a dynamic and objectively changing system, reflecting the physiological state. In the present study, a miRNA-specific qRT-PCR analysis was used to identify miRNA molecules with a potential role in the pathogenesis of ASD. Despite the fact that we do not know whether the observed dysregulation of miRNA molecules is also characteristic of the CNS of the studied patients, most of the analyzed miRNAs show expression in the brain according to literature data.

Due to the presence of these complicating factors, a new emphasis in the study of biomarkers in schizophrenia and ASD is the approach based on the study of the molecular profile of biomolecules from peripheral tissues and especially miRNA molecules in peripheral blood, as a dynamic and objectively changing system, reflecting the physiological state. In the present study, a miRNA-specific qRT-PCR analysis was used to identify miRNA molecules with a potential role in the pathogenesis of ASD. Despite the fact that we do not know whether the observed dysregulation of miRNA molecules is also characteristic of the CNS of the studied

patients, most of the analyzed miRNAs show expression in the brain according to literature data.

The changes in posttranscriptional regulatory mechanisms in peripheral blood observed in patients with schizophrenia are not only relevant to the pathophysiology of the disease but also shed light on a suitable site for identifying and validating potential biomarkers also in ASD. Another aspect of the guidelines for studying the links between miRNA molecule dysregulation and neurodevelopmental disorders is mutational changes, i.e., the genetic link that may result in a change in the transcription and processing of miRNA molecules. miRNA genes are for the most part transcribed by RNA polymerase II-specific promoter sequences, so they are also regulated by transcription factors and chromatin structure in a manner similar to messenger RNA molecules.

Micro RNA molecules are usually transcribed in the form of non-coding RNA molecules or are transcribed from intron sequences of protein-coding RNA molecules. As such, the degree of transcription of miRNA molecules would be affected both by the prevailing regulatory environment and by the specific epigenetic profile that is characteristic of each protein-encoding gene. Transcription of miRNA molecules also depends on the integrity of conservative sequences (motifs) in their regulatory (promoter and enhancer) elements. After transcription, processing to produce a precursor and a mature miRNA molecule will also be a function of the sequence integrity in the primary transcript that directs the enzyme system to the biogenesis of mature miRNA molecules. With all of the above, the study of the detailed mechanisms of regulation of miRNA expression remains a promising prospect in research in the field of neurodevelopmental disorders.

Discussion of the results obtained from the whole-exome sequencing for the study of protein-coding sequences in the human genome associated with ASD

Using large-scale molecular tools, such as NGS, with proven efficacy, Codina-Solà et al. in combination with blood transcriptome analysis (RNAseq) in a selected group of men diagnosed with ASD to detect putative causal genetic variants. The authors use segregation analysis in combination with expression studies to better discriminate between putative pathogenic variants from clinically insignificant (harmless) rare ones. Using WES analysis, the authors also identified several cases of probable monogenic etiology in ASD, including four patients with de novo variants in high-potential candidate-autosomal genes (11%) and two patients with inherited X-linked mutations (5.6%). Autosomal mutations with loss of function (LoF) have been identified in the SCN2A, MED13L, and KCNV1 genes. SCN2A is one of the few genes found to mutate in patients who show no link to ASD and intellectual disability. Another gene, MED13L has previously been associated with intellectual disabilities and heart defects, while *de novo* mutations have also been described in patients diagnosed with autism. Recently, de novo deletions affecting the coding

regions of the MED13L gene have been found in girls showing a phenotype very similar to the patient the authors report, including facial dysmorphism, hypotension, and developmental delay, along with ASD. The authors' findings correspond to the role of the MED13L gene in neurodevelopmental disorders. The third *de novo* and possibly pathogenic variant is a variant in the KCNV1 gene encoding a potassium channel subunit, mainly expressed in the brain and involved in the regulation of two other potassium channels (KCNB1 and KCNB2). Potassium channel defects have been associated with a variety of neuropsychiatric disorders, including bipolar disorder, schizophrenia, and ASD. Therefore, the authors' data support the role of potassium channels in the etiology of ASD. The authors also found a *de novo* missense mutation in the CUL3 gene, which is one of the periodically de novo mutant genes in patients with ASD. Regarding X-linked mutations, the authors also identified changes in two genes (MAOA and CDKL5) previously associated with ASD and intellectual disability. A splicing mutation was detected in the MAOA in a multiplex family in which an X-linked inheritance pattern was observed, with two affected males and history of the mother with a psychiatric illness. MAOA encodes the enzyme monoamine oxidase A, which breaks down neurotransmitters such as dopamine, norepinephrine, and serotonin. Biochemical changes in the catecholamine pathway have been found in both affected individuals, which are much milder in their mother, which maintains the pathogenicity of the mutation. A protein-shortening mutation in the MAOA gene described in the Dutch family in 1993 was first identified, and recently a second mutation with loss of function was found in a family with ASD and behavioral problems. The work of the authors further strengthens the link between MAOA and ASD. Another established X-linked mutation affects a highly conserved amino acid in CDKL5, which has been predicted to be delegated by various algorithms. Ongoing efforts to determine the extent of variation in gene expression in a large number of healthy controls, such as the Geuvadis project, which aims to pool knowledge and resources on genome sequence at the European level and allow researchers to develop and test new hypotheses about the genetic basis of diseases. As well as the Genotype-Tissue Expression (GTEx) consortium, which aims to catalog genetic variations and their influence on gene expression in and between all major tissues in the human body. The project aims to build a resource database and related tissue biobank to investigate the relationship between genetic variation and gene expression in all major human tissues in 1000 individuals, which will help to better elucidate the effects of genes that show a change in patients with ASD. The analysis of the data obtained from this type of project will allow establishing standards in the procedures for interpretation of data in connection with the manifestation of the clinical phenotype.

The whole-exome sequencing results presented in the dissertation thesis includes an additional level of complexity. As pool samples are subjected to analysis, the interpretation of the results is further complicated. The presented analysis can be considered as a preliminary pilot study which would require expansion of the

sample size as well as subsequent individual validation of the identified candidate genetic variants in ASD. Recently, there has been a significant enrichment of information about the spectrum and frequency of genetic changes in the genome associated with ASD. However, despite the progress made, much of the genetic contribution to the etiopathogenesis of the disease remains unclear.

In the present study, a significant number of genetic variants in the genome were found in patients with a diagnosis of ASD compared to clinically healthy children using the new generation exome sequencing method.

The high informativeness of the algorithm used in the present dissertation thesis for the evaluation of genetic variants could serve to detail our understanding of the etiology and pathogenesis of the disease, as well as in the future to improve the diagnosis of ASD. Regardless of the success of the bioinformatics analyzes used in the development to assess the possible effect of the identified genetic variants in the genome, in order to prove the results obtained, additional detailed functional analyzes are needed.

In this regard, future research in the field could focus on:

- Conducting an individual analysis of de novo identified pathological variants in children with ASD.
- Expanding the scope of the studied groups of patients with ASD and clinically healthy children from the Bulgarian population.
- Carrying out complex functional analyzes of the de novo identified, pathological variants, and of the variants of unknown clinical significance in the genome of patients with ASD.

Discussion of the results obtained from the transcriptomic (RNA-Seq) studies of protein-coding genes in ASD

In the present dissertation, differentially expressed protein-encoding genes in peripheral blood from patients diagnosed with ASD have been identified using a set of techniques. The study presented in this dissertation is the first of its kind comprehensive study conducted on a selected group of Bulgarian patients. To date, a limited number of whole transcriptome analysis have been reported in the literature in children with ASD (**Table 3**).

The number of samples tested in the literature, as well as the tissues from which the test material was isolated, differed significantly. Three of the studies were performed in samples from different parts of the brain, but the number of patients studied is relatively small, as one of the biggest challenges in the study of expression changes in neuropsychiatric disorders is the collection of samples from the primary target tissue, ie. brain. Due to the apparent impossibility of performing *in vivo* gene expression studies in patients' brains, retrospective studies with *post-mortem* brain

samples are mainly performed. These experiments are also limited to some extent by access to high-quality, well-phenotyped specimens from individuals diagnosed with ASD. Moreover, the limited number of brain samples that can be examined makes it difficult to detect and validate significant associations with the disorder. In all studies cited above focus on brain tissue, the biological material used was taken from post-mortem material.

In addition, a number of data have been published in the scientific literature indicating a correlation of gene expression between peripheral blood cells and brain tissue, which further justifies the use of peripheral blood samples in neurodevelopmental disorders (**Table 3**).

The search for biomarkers for the diagnosis and prevention of ASD in clinical practice makes brain tissue not very suitable for this purpose. Performing a brain biopsy is unthinkable for a healthy individual in order to conduct research and determine the risk of developing a neurodevelopmental disorder.

Several studies of the expression profile in patients diagnosed with ASD are currently available. When discussing the most objective material for examining the expression profile, lymphoblastic cell lines should also be considered. However, the use of lymphoblastic cell lines for expression assays is also not without limitations, such as the fact that cell line experiments do not reflect the exact characteristics of the disease. This is one of the most important issues addressed in the study of gene expression in neurodevelopmental disorders in general. In recent years, a number of authors have also reported studies of the expression profile of protein-encoding genes in peripheral blood from patients diagnosed with ASD with promising results. As a biological material, peripheral blood is one of the most readily available and most commonly used materials in biomedical research. Peripheral blood samples can be taken repeatedly, at different stages of a disease, with minimal discomfort and risk of side effects in the patient.

Due to all the above advantages, the technique is defined as a "revolutionary tool" for the purposes of transcriptome analysis and is included as the main approach in the implementation of the present experimental work.

The present dissertation thesis reviews data from a whole transcriptome analysis revealing differences in the expression of protein-encoding genes in children with ASD, suggesting that some genetic aspects of the disorder are shared by most of the children in the sample. Some of the differentially expressed genes are involved in key processes related to cellular signaling, the immune system, and metabolic processes. The results obtained reveal that the processes such as synaptic plasticity, disturbances in the orientation of axons, neuronal survival, differentiation, and inflammation have a role in the pathophysiology of ASD. Some of the differentially

expressed genes identified in the study were associated with ASD in other studies. (**Table 3**).

Gene	Linkage study	CNV analysis	Exome sequencing	Expression study
ANKRD22	+ (Buxbaum 2004)			
BCYRN1	+ (Lauritsen, 2006)			
BIVM		A case of deletion (Pinto, 2010)		
FLNC		Two case of deletion (Marshall, 2008; Sanders, 2011)		Increased expression (Garbett et al. (2008)
FOSL1		Three case of deletion (Bucan, 2009)		
HLA-DQA2		A case of deletion (Pinto, 2010)		
RAP1GAP		Three case of deletion (Bucan, 2009)		
SCAND1		One case of duplication (Sanders, 2011)		
SERPINE1		One case of duplication (Sanders, 2011)		
SLC8A2		One case of deletion (Pinto, 2010)		
TM4SF1				Two expression study (Garbett et al. 2008) (Voineagu et al. 2011)
TMEM40		One case of		

		duplication (Sanders, 2011)
TPX2	+ (Allen-Brady, 2008)	Frameshift mutation (Iossifov, 2012)

Table 3. Previous studies citing identified differentially expressed genes.

Some of the genes identified in the study as CD177; CRB2; DPM3; IGF2BP1; LTF; MAGEA4; OTOF; are new candidate genes.

CD177 (CD177 molecule) is a gene encoding a protein - glycosyl-phosphatidylinositol (GPI) - a linked surface glycoprotein that plays a role in activating neutrophils. The protein binds platelet adhesion endothelial cell molecule-1 (PECAM-1) and has a function in neutrophil migration. PECAM-1 binds in a way that facilitates homophilic, adhesive, transendothelial migration of leukocytes. Altered gene expression may lead to altered leukocyte migration into the CNS. There is sufficient evidence for the key role of immunity in the development of RAS. In this context, the CD177 gene could be considered as a potential candidate relevant to the development of the disorder.

CRB2 is a protein-coding gene. Data obtained from GO annotations associate the protein with calcium ion binding. It is thought that the protein may be involved in cell morphogenesis (http://www.genecards.org/).

DPM3 is a protein-coding gene that encodes a dolichol phosphate subunit of mannosyltransferase and acts as a stabilizer in complex formation. The protein is involved in the biosynthesis of N-glycan precursors. N-glycoproteins play a key role in embryonic development, differentiation, and maintenance of cellular functions (http://www.genecards.org/).

IGF2BP1 (insulin-like growth factor 2, mRNA binding protein 1) is a gene encoding a protein. This gene encodes an mRNA-binding protein, a member of the insulin-like growth factor 2 family. Its action is mediated by the binding of mRNA to certain genes, such as insulin-like growth factor 2, β-actin, and the regulation of their translation. It is also involved in Wnt-mediated β-catenin signaling and is thought to be involved in synaptic plasticity and hippocampal development. The study linked its reduced expression, due to hypermethylation, low birth weight, and memory changes. In the data presented, the gene also shows reduced expression, suggesting its involvement in the pathogenesis of ASD.

LTF (Lactotransferrin) is a gene encoding a protein that is a member of the transferrin family. The protein is involved in the regulation of iron homeostasis, protecting against microbial infections, anti-inflammatory activity, regulating cell growth and differentiation, and protecting against cancer development and

metastasis (http://www.genecards.org). Increased number of activated cells leading to increased LTF production may also explain to some extent its accumulation. A number of hypotheses have been proposed regarding the protective function of LTF as an inhibitor of tissue damage through its anti-inflammatory and antioxidant functions. However, the question remains whether the increased expression of LTF is a consequence of the development of ASD, or is it an attempt by the body to respond to pathological immune processes in the brain and has a protective function?

MAGEA4 (Melanoma Antigen Family A4) is a protein-encoding gene. The protein is thought to be involved in the process of embryonic development (http://www.genecards.org).

OTOF (Otoferlin) is a gene encoding a multi C2 domain protein involved in the binding of Ca2 +, the fusion of vesicles with the plasma membrane, and in the control of the release of neurotransmitters at synapses. It may also play a role in endosome recycling (http://www.genecards.org). The C2 domain of the protein is similar to that of protein kinase C and is thought to have the same function. Autosomal recessive mutations in the gene lead to nonsyndromic deafness and sound (auditory) neuropathy.

A number of scientific pieces of evidence suggest that otoferlin may replace synaptotagmin as a calcium-sensitive sensor. As well as the presence of sufficient experimental evidence showing its direct interaction with syntaxin 1A, SNAP-25, as well as with voltage-gated calcium channels (Cav1.3). In addition, otoferlin is required in calcium-dependent exocytosis/neurosecretion at some synapses [Roux et al., 2006]. The result obtained for reduced expression in the presented study can be considered as a potential mechanism of imbalance at synapses involved in the pathogenesis of ASD.

Discussion of the results obtained from gene ontology analysis of differentially expressed genes and KEGG analysis of biological pathways

A number of studies have evaluated changes in gene expression in postmortem brain tissue. These studies have repeatedly shown the ASD transcriptome shows altered expression in different brain areas compared to healthy individuals. Differential expressed genes are associated with pathways involved in synapses, immune response/apoptosis], and neurotransmitters. The presented analysis confirms previous studies in the field confirming that in ASD patients a change in pathways associated with synaptic plasticity, disturbances in axon orientation, neuronal experience, cell differentiation, immunity and inflammation has been found.

Although the number of studies trying to integrate genetic findings in ASD is significantly smaller than the number of studies that aim to look for altered genes, we

believe that through integrative functional analysis of genetic changes, common pathways, and mechanisms underlying the etiopathogenesis of ASD can be found. The results of the performed gene ontology analysis show biological connections and common pathways between the differentially expressed genes involved in structural components of the cell, apoptosis, developmental, metabolic and regulatory processes, immunity, and intercellular signaling.

The obtained data demonstrate: 1) the exceptional heterogeneity in the pathogenesis of this disorder; 2) the presence of evidence for the multifactorial model of ASD; 3) the connection of the immune system with ASD; 4) attitude to this disorder of genes involved in intercellular signaling; 5) dysregulation of basic cellular processes may contribute to the development of the disorder. Genes usually interact with each other to perform specific biological functions. KEGG is the main publicly available database related to the biological pathway and gene interaction data. The analysis identifies metabolic and signaling pathways such as endocannabinoid signaling, calcium signaling, and others. in which the differentially expressed genes obtained in the present study are involved.

Calcium signaling pathway

Dysregulation of calcium signaling pathway is interesting since recent collaborative effort of meta-analysis of five major neuropsychiatric disorders including ASD suggests that two calcium channel coding genes – CACNA1C and CACNB2 – are significantly associated with all five diseases. Calcium signals culminate by initiating the fusion of synaptic vesicles into the neuronal presynaptic membrane, and the sum of the evidence is that pre- and postsynaptic membrane changes recognized in several fundamental processes of synaptic plasticity and learning are calcium sensitive and are altered in models of ASD. Recent findings in the etiopatogenesis of the ASD-causing Angelman, Prader-Willi, Rett, tuberous sclerosis suggest an inability of neurons to generate adaptive responses via calcium-regulated gene expression.

Glutamatergic and GABAergic synapse

Critical balance between excitatory glutamate and inhibitory GABA neurotransmitter is essential and crucial for proper development and functioning of brain. GABAergic (gamma aminobutyric acid) and glutamatergic interneurons maintain excitability, integrity and synaptic plasticity. Several recent evidences implicated relative loss of inhibitory GABA with corresponding glutamate mediated hyper-excitation in development of ASD.

MAPK signaling pathway

Dysregulation in Ras/mitogen-activated protein kinase (Ras/MAPK) pathway genes lead to a class of disorders known as RASopathies which includes

neurofibromatosis type 1 (NF1), Costello syndrome (CS), Noonan syndrome (NS), and cardio-facio-cutaneous syndrome (CFC). Potential genetic and phenotypic overlap between dysregulation of Ras/MAPK signalling and ASD were suggested in previous studies. Higher prevalence and severity of ASD traits in RASopathies compared to unaffected siblings propose that dysregulation of Ras/MAPK signalling during development may be implicated in ASD risk.

Wnt signaling pathway

The Wnt pathway is involved in various well-known cellular processes including differentiation and migration, especially during nervous system development and cell proliferation. Given its various functions, dysfunction of the canonical Wnt pathway probably exert adverse effects on neurodevelopment and therefore leads to the pathogenesis of ASD. In the last years an increased amount of evidence has shown that components of Wnt pathways are involved in major psychiatric disorders. A literature review supports the contention that modification of genes affecting the activity of the Wnt pathway could contribute to individual forms of ASD.

Endocanabinoid signalization

In the presented study, the analysis of biological pathways revealed dysregulation in endocannabinoid signaling due to the established differential expression of RAP1GAP and OTOF genes in children diagnosed with ASD. Endocannabinoids are with key significance for modulating synaptic function. By activating cannabinoid receptors expressed in the CNS, these lipid mediators can regulate neuronal function and behavior. Retrograde endocannabinoid signaling (from postsynapse to presynapse) is a key modulator of synaptic plasticity and cognitive functions. Retrograde signaling is the main way in which endocannabinoids mediate short-term and long-term forms of plasticity at excitatory and inhibitory synapses. Deletions and point mutations in the gene encoding neuroligin 3 (NLG 3) are associated with ASD. Földy et al. examined synaptic signaling in NLG3 "knockout" and NLG 3 (R451C) mice to determine whether these non-syndromic ASD models had a common synaptic phenotype.

Another study in mice with a model of Fra X syndrome PAC (FMR1- / Y), which showed that deletion of the FMR1 gene leads to a greater endocannabinoid-mediated response to GABAergic synapses in the dorsal striatum and CA1 region of hippocampus. In addition, loss of FMR1 protein results in a deficiency in mGluR5-dependent release of endocannabinoids at excitatory synapses.

RAS/MAPK pathway

The RAS/MAPK (mitogen-activated protein kinase) signaling pathway mediates the transmission of signals from cell surface receptors to their cytoplasmic

and nuclear targets. Different groups of molecular adapters bind to RAS and Rap1 (RAS-linked protein 1) and initiate signaling along the ERK (extracellular signal-regulated kinase). This pathway mediates various cellular functions such as proliferation, migration, differentiation, and cell survival., also a diverse spectrum of neuronal manifestations such as synaptic plasticity, long-term potentiation, and suppression (LTP and LTD), and memory formation.. Three CNVs that are associated with ASD deletions in 16p11.2 and duplications in 7q11.23 and 22q11.2 reveal genes associated with Ras / MAPK-dependent signaling.

Serotoninergic synapse

Serotonergic synapses are one of the most common human synapses, developing in the early stages of embryonic development. Most serotonergic neurons are located in the medial and dorsal part of the nuclei raphe. Serotonergic neurons can be detected in the human brain, from the 5th week of gestation, and in the following months, they grow and multiply intensively. The early appearance of the serotonergic pathway, as well as its intensive activity during the first stages of development, shows its essential role in the developmental process. Serotonin is thought to affect the process of neurogenesis and/or neuronal removal, neuronal differentiation, and synaptogenesis. Evidence for the involvement of serotonin in the pathogenesis of RAS has been obtained from genetic, imaging and, neuropathological studies. The presented data obtained from the conducted bioinformatics analysis show that the RAP1GAP and OTOF genes take part in the dysregulation of the serotonergic signaling.

Glutamatergic synapse

Glutamate is the main excitatory neurotransmitter in vertebrates. Its effects are mediated by the metabotropic (mGlu) and ionotropic (iGLu) glutamate receptors located in the cell membranes of neurons and glia. Metabotropic receptors belong to the superfamily G-protein-coupled receptors and are divided into three groups: group I (mGluR1 and mGluR5), group II (mGluR2 and mGluR3) and group III (mGluR4 and mGluR6-8) according to their pharmacological agonists, primary sequence and G-protein effector. Changes in the function of metabotropic glutamate receptors, especially mGluR dependent long-term depression (LTD), have been reported in several models of RAS in mice.

Increased incidence of epilepsy in patients with ASD shows that abnormally elevated glutamatergic signaling may contribute to the development of some of the observed autistic traits. There is a positive correlation between plasma glutamate levels and the severity of symptoms in ASD. Elevated levels of EAAT1 and AMPA 1 expression have been found in the cerebellum of ASD patients, which increase the extracellular concentration of glutamate and improve the postsynaptic glutamatergic activity.

Cholinergic synapse

In the central nervous system, acetylcholine (ACh) is essential in evaluating stimuli from the environment with the nature of reward or threat. According to several authors, cholinergic signaling is involved in the regulation of behaviors, attention, cognitive plasticity, social interactions, and stereotype behavior. At the neuronal level, thalamo-cortical cholinergic projections modulate the release of the excitatory neurotransmitter glutamate. In the cerebral cortex, acetylcholine has both an excitatory and an inhibitory effect on different layers, depending on the receptor type on which it acts. Acetylcholine also regulates the development of the CNS - growth, differentiation and plasticity, and has an effect on the differentiation of excitatory and inhibitory synapses. Therefore, acetylcholine plays an important role in modulating the balance between excitation and inhibition in the brain. One of the hypotheses about the pathogenesis of ASD is focused on the imbalance between excitation and inhibition in the brain.

Several studies have associated the cholinergic system with ASD. Kemper и Bauman found altered size, number, and structure of cholinergic neurons in the basal ganglia of the forebrain in ASD patients. In another study, Friedman and colleagues. also found reduced levels of the precursor acetylcholine and nicotine-cholinergic receptor agonist in ASD patients. Sokol and colleagues. low levels of choline in the cytosol correlate with the severity of symptoms in ASD. Several immunohistochemical studies report reduced nicotine-cholinergic receptor subunits in post-mortem specimens of the neocortex, cerebellum, thalamus, and striatum in patients with autism. From the obtained data we can assume that the dysregulation of the cholinergic signaling is related to the established changes in the expression of the RAP1GAP and OTOF genes.

Dopaminergic synapse

Dopamine (DA) is an important neurotransmitter in the brain, where it controls various functions such as motor activity, endocrine regulation, learning, memory, motivation and reward. Upon release from presynaptic axon terminals, dopamine interacts with at least five CNS receptor subtypes. Dopamine receptors affect intracellular signaling processes through a variety of cAMP and $Ca2^{+}$ dependent mechanisms. Thus, dopamine affects neuronal activity, synaptic plasticity, and behavior. There is ample evidence for an association between dopaminergic signaling and ASD. Several genetic studies have demonstrated the mutations in dopaminergic pathway genes: Dopamine transporter (DAT); dopamine receptors. Enzymes involved in dopamine synthesis (DOPA decarboxylase, DDC) and catabolic processes (catechol-O'Methyltransferase, COMT and monoamine oxidase MAO-A and - B). Dopamine is thought to be a key neurotransmitter in motor activity, and may contribute to the "motor" symptoms seen in ASD, such as speech, social behavior, and behavioral perseverations (compulsive repetition of words,

movements, thoughts, etc.). This neurotransmitter is critical to the reward system and may mediate the social reward/deficit aspect observed in ASD. From the obtained data we can assume that the observed dysregulation is associated with the change in the expression of RAP1GAP and OTOF genes.

Transendothelial leukocyte migration

The leukocytes migration from the blood into the tissues is vital for the immune system and the inflammatory response. Through diapedesis, leukocytes bind to endothelial cell adhesion molecules (CAMs) and then migrate through the vascular endothelium. Leukocyte passage results in the activation of a number of endothelial cell signals that stimulate endothelial cell dissociation and calcium influx in the endothelial cell, which are required for ICAM-1-dependent leukocyte migration. (www.genome.jp/kegg). In the data obtained, the disruption of this pathway may be due to altered calcium signaling due to a change in RAP1GAP expression and altered intracellular protein kinase C (PKC) activity due to OTOF. Although ASD mainly affects brain functions, it is not known to what extent other organs and systems are also affected. A number of studies have been published identifying widespread changes in the immune system in children with ASD, both at the systemic and cellular levels. Brain samples from ASD patients show signs of active, ongoing inflammation, as well as changes in immune signaling and immune function pathways. In addition, many genetic studies have revealed a link between ASD and genes associated with the nervous and immune systems. Changes in these pathways may affect function in both systems. Taken together, these reports, as well as the results of our study, suggest that ASD may in fact be considered a systemic disorder associated with abnormal immune responses.

Gonadotropin-releasing factor signaling pathway

Gonadotropin-releasing factor (GnRH) is secreted by the hypothalamus and acts on its receptor in front of the pituitary gland to regulate the production and release of gonadotropins LH and FSH. GnRH binds to GQ / 11 proteins and activates phospholipase C, which transmits a signal to diacylglycerol (DAG) and inositol 1,4, 5-triphosphate (IP3). In the present case, this signal pathway is disrupted again due to the influence of RAP1GAP, leading to a change in the voltage-dependent calcium channels type L. Consequently, it has been clarified that these results were compromised by the presence of nuclear pseudogenes of mitochondrial origin in the sequences of which the reported mutations have been actually found. In the current work, such a risk of pseudogene contamination is avoided, since the initial data comes from RNA-expression and not genomic DNA samples. Even more indicative is the fact that the assessed locus has coverage of over 1000 individual sequenced reads in both complex samples, which demonstrates that the gene of interest is actively transcribed, a feature not characteristic of pseudogene expression.

Horvath et al., observed an intricate clinical picture (sensorineural hearing loss, myopathy, and ataxia) in two patients carrying heteroplasmic *de novo* mutations in the genes for CO2 and CO3 subunits. Furthermore, Debray et al., identified a *de novo* heteroplasmic mutation (m.7402delC) causing a shift in the reading frame in CO1, in a patient with non-convulsive status epilepticus, temporary cortical blindness, muscle weakness, hearing loss and cognitive impairment. A third example is the work of Kytovuori et al., who described a case of a patient with cognitive impairment, epilepsy, psychosis and sensorineural hearing loss, in which a frameshift mutation (m.8156delG) in the subunit II of the cytochrome oxidase complex has been found. In summary, the list of differentially regulated genes is enriched with pathways associated with nervous system development and function, and immune system and most of them seem to be around core networks such as those involved in kinase and/or signaling networks. Therefore, our results support the involvement of various genetic factors (heterogeneity) in the development of ASD, while suggesting these different factors can be converging at, or diverging from central networks such as signaling networks. Findings from our current study demonstrate that there are clear and significant abnormalities in the gene expression of peripheral blood samples obtained from children with ASD compared to healthy controls. This promising work, while far from being definitive, gives further proof to the recently emerging principle that peripheral blood is a potentially useful source of diagnostic biomarkers for disorders of the brain and other inaccessible tissues. If the results of this work are confirmed in future studies and the identified changes in the study group are individually validated by us or by others in other independent cohorts, we can assume that the differentially expressed genes may help clarify the etiology and pathogenesis of this disorder and the dysregulated pathways may also provide targets for the experimental treatments in ASD.

Discussion of the results obtained from ASD expression analyses of miRNA

The interest of the scientific community in the development of diagnostic tests based on circulating miRNA molecules is justified by many advantages that this approach offers: the accessibility of the studied tissues, the relatively low complexity of analysis and high stability of examined molecules, the time for conducting experimental procedure, relatively low cost, potential for development of identification and quantitative analysis techniques, the presence of state-specific expression profiles and the relative constancy of expression between individuals. However, some issues need to be considered in the development of biomarkers based on miRNA molecules. First, regulatory networks involving miRNAs are inextricably linked, where a mingle miRNA is able to coordinate the expression of hundreds of genes, while the opposite is also true, a mingle gene can be controlled by multiple miRNA molecules. Therefore, the binding of a specific miRNA molecule to certain processes in the brain is an extremely challenging work. In addition, the plasticity in the expression of miRNA molecules makes it unlikely the existence of

basic miRNA regulators, whose normal function will be impaired in all or almost all cases of the pathological condition. Thus, it is advisable to analyze several molecules simultaneously in order to increase the diagnostic power of developed test. Another limitation may be the low amount of extracellular molecules in body fluids, which on the one hand may affect the reproduction of the results obtained, on other hand requires an optimized approach for extraction and the use of a highly efficient analytical method. Finally, the problem with the quantitative analysis of circulating miRNA molecules for clinical purposes and the need to develop generally accepted

In the preliminary expression studies performed on complex (pool) samples from analyzed groups, 42 miRNA molecules showed encouraging results as candidate biomarkers for initial analysis. Of these, a total of 29 miRNAs with decreased expression were identified (miRNA-589-3p, miRNA-6849-3p, miRNA-3135a, miRNA-15a-5p, miRNA-328-3p, miRNA-183-5p, miRNA-3674, miRNA -96-5p, miRNA-3687, miRNA-6799, miRNA-3p, miRNA-587-3p, miRNA-504-5p, miRNA-576-5p, miRNA-486-3p, miRNA-let-7i-3p, miRNA -29c-5p, miRNA-301a-3p, miRNA-3064-5p, miRNA-145-5p, miRNA-424-5p, miRNA-193b-3p, miRNA-487b-3p, miRNA-197-5p, miRNA-500a-5p, miRNA-664b-3p, miRNA-20b-3p, miRNA-671-3p and miRNA-199a-5p), 11 showed increased expression (miRNA-4489, miRNA-4489, miRNA-106b-5p, miRNA-1423p, miRNA-3620-3p, miRNA-365a-3p, miRNA-664a-3p, miRNA-374b-5p, miRNA-18b-3p, miRNA-619-5p and miRNA-210-5p), while two showed no changes in expression in children with ASD compared to control group. The next step in the experimental work was to confirm the expression of the eight most highly unregulated serum miRNAs using qRT-PCR analysis of complex samples from the studied groups. The results obtained in summary are as follows. One, miRNA-365a-3p, no statistically significant difference in expression was found. For the other four: siRNA-500a-5p, siRNA-197-5p, siRNA-424-5p, and siRNA-664a-3p, decreased expression was found, initially observed in pooled ASD samples and subsequently confirmed in the individual expression analysis of ASD patients.

It should be noted that miRNA-424-5p and miRNA-500a-5p mediated regulation of gene expression appear to have a significant effect on CNS processes, assumed that their validated mRNAs are actively involved in brain processes. In contrast, no such apparent association was found between miRNA-197-5p, miRNA-664a-3p, and the CNS, which can be partly explained by the smaller number of mRNA molecules affected and their weaker involvement in more specific KEGG pathways.

MiRNA-500a-5p shows involvement in various malignancies and more specifically liver cancer. In 2009, miRNA-500a-5p was proposed as a diagnostic marker for hepatocellular carcinoma (HCC) following conclusions about its increased expression in cancer cell lines, HCC tissue, and serum. Of particular interest is the fact that serum levels of miRNA-500a-5p return to normal values in patients with HCC after surgery for removing cancer. Several other studies have demonstrated

abnormal levels of miRNA-500a-5p and its association with tumor pathology and overall survival rate in cases of non-small cell lung cancer (NSCLC) Similarly, increased miRNA-500a-5p expression has been associated with increased cell proliferation and poor prognosis in prostate cancer and gastric cancer also.

In contrast to the studies performed with miRNA-500a-5p and the amount of data available, studies that are specifically related to miRNA-197-5p are significantly limited. A search for publications in scientific databases with the keyword "miRNA-197-5p" (January 2020) shows limited results. Remarkably, two of the studies revealed the role of siRNA-197-5p in neurological diseases. In other study, miRNA-197-5p showed significantly decreased expression in plasma exososomes in patients with mesial temporal lobe epilepsy and hippocampal sclerosis and was therefore one of the proposed potential therapeutic targets and diagnostic biomarkers. In addition, the parental miRNA-197, is considered to be one of the major human oncomiRNAs, acting as either a tumor promoter or suppressor, through its effect on active or inactive oncogenes.

Like the two miRNA molecules described above, miRNA-424-5p is not associated with ASD in previous studies. In many studies miRNA-424-5p has been associated with a number of different cancers. In some cases, as in colon cancer, adenocarcinoma, gastric carcinoma and oral cancer, where miRNA-424-5p shows increased expression. In other cases, as in cervix adenocarcinoma and ovarian carcinoma, miRNA-424-5p shows typical tumor suppressive function and in this cases, its expression is decreased in affected tissues, in connection with abnormal cell proliferation. Obviously, the involvement of miRNA-424-5p in controlling the rate of cell division in various tissues may also play an important role in the development of the pathology observed in ASD. Recently, Wu and colleagues suggest that the evolutionary role of some miRNAs in the primate brain has been associated with inhibition of excessive cell proliferation, a phenomenon also observed in ASD. Therefore, it can be concluded that decreased expression of miRNA-424-5p in ASD patients may reflect an insufficient ability to limit early postnatal brain growth and development.

The most significant study of siRNA-664a-3p for CNS is devoted to Alzheimer's disease and major depressive disorder. In the study, authors performed a systematic review of dysregulated miRNA molecules in both conditions and found that only seven are common, one of which miRNA-664a-3p. Although, in the studies presented in the dissertation, the decreased expression of this miRNA in ASD is statistically significant, it shows the lowest change in the expression level and the lowest prognostic value.

Many miRNA regulated genes have been implicated in neurological diseases and mental disorders. Three separate studies found an increase in GNB1 protein expression in the schizophrenia prefrontal cortex. Other authors describe decreased

transcription levels of guanine nucleotide-binding (G protein), alpha q polypeptide (gnaq) in mice with chronic stress. Mutations in another gene, guanine nucleotide-binding protein (G protein), a gene showing expression mainly in the brain, are among the main causes of primary torsional dystonia and craniocervical dystonia. A change in gnal expression has also been reported in schizophrenia. One of the two identified miRNA target genes encoding glutamate transporters, slc1a1, has been extensively studied in epilepsy, OCD, MS, Alzheimer's disease and schizophrenia. Significant overexpression of SLC1A1 protein was observed in post-mortem specimens in patients with epilepsy as well as in rats with pilocarpine-induced epilepsy. Significant increase in SLC1A1 protein expression in patients with epilepsy has also been reported by other researchers. However, decreased SLC1A1 protein expression has also been reported. Increased expression of the slc1a1 gene at the mRNA and protein levels has been demonstrated in patients diagnosed with schizophrenia. More and more data suggest the involvement of the SLC1A1 gene in the development of multiple sclerosis, however definitive evidence are not available.

In conclusion, the data presented in the dissertation demonstrate that the studied miRNA molecules: miRNA-500a-5p, miRNA-197-5p, miRNA-424-5p and miRNA-664a-3p, show decreased serum levels in ASD, and that these miRNA molecules have the potential to be used as readily easily accessible and objectively measurable biomarkers in this disorder. The diagnostic power of a test based on miRNA-500a-5p and miRNA-197-5p will be highest if these molecules are evaluated en block. Undoubtedly, additional evidence from larger independent studies will be needed prior to the possible clinical application of the studied miRNA molecules as biomarkers in ASD. In addition, further studies of these miRNA molecules may reveal details of the regulation mechanism and the exact consequences of dysregulation in ASD patients, also whether their abnormal expression is related to the etiopathogenesis of the disorder or it is an indirect result of other processes.

Discussion of the results from Digital Gene Expression (DGE-tag profiling)

Despite the identification of numerous schizophrenia susceptibility genes, the pathobiology of schizophrenia remains unknown. Using DGE to identify mRNA expression genome-wide has the capacity to add a novel dimension to our understanding of transcriptional regulation in schizophrenia. Peripheral blood is an ideal substitute tissue as it has the potential to reflect responses to changes in the immediate and distant environments by alterations of gene expression levels. The change in the expression of a miRNA molecules could be attributed in part to a dysregulation of miRNA processing genes like DICER1, a gene that was shown to be differentialy expressed in our study. Moreover, thus, it is possible that DICER might contribute to the changes in the expression of the protein-coding gene by modulating miRNA processing.

Abnormalities in dopaminergic signaling are usually observed in patients with schizophrenia. Variations in the dopamine system are not only affected by dopamine itself but also by dopamine receptors. The diverse physiological functions of dopamine are mediated by five different dopamine receptors, encoded by the genes DRD1, DRD2, DRD3, DRD4, and DRD5. The expression of dopamine receptors is well documented in the brain but very few studies have determined their expression in other organ tissues including blood. However, alterations in gene expression of dopamine receptors have been reported in different cells in certain diseases of the nervous systems. The dopamine (DA) D4 receptor (DRD4) could play a role in mediating dopaminergic activity. The DRD4 is primarily expressed on pyramidal neurons and interneurons in the prefrontal cortex, but there is also support for DRD4 localization on medium spiny neurons in the basal ganglia (striatum, Str; and nucleus accumbens core, NAc), throughout the limbic system and in the thalamus of rodents. Glutamate (Glu) is the main neurotransmitter for multiple connections in this circuit and serves as the neurotransmitter of the pyramidal cells**.** A key factor in glutamatergic neurotransmission in the N-methyl-D-aspartate (NMDA) receptors, which are involved in brain development, excitatory neurotransmission, synaptic plasticity, and memory formation. The NMDA receptors are composed of multiple subunits including at least one NR1subunit (encoded by the GRIN1 gene) and one or more NR2 subunits (encoded by the GRIN2A-D genes), and less commonly, an NR3 subunit (encoded by the GRIN3A-B genes). Several lines of evidence support that hypofunction of the NMDA receptors may be involved in the pathophysiology of schizophrenia.

Discussion of the results of a quantitative proteome analysis - Isobaric Tag for Relative and Absolute Quantification (ITRAQ) in the ASD

Using two-dimensional iTRAQ-based LC-MS/MS profiling, a number of differentially expressed serum proteins were identified from patients diagnosed with ASD compared to typically developing children in the control group. Using the iTRAQ method for quantitative proteomic analysis, 60 proteins were identified in the present study, in which 24 proteins with increased expression and 36 proteins with reduced. In summary, the proteomic approach used in the present study helped to identify a number of differentially expressed peptides that correspond to known proteins. The obtained results support several previous studies that demonstrate differences in circulating immune proteins in patients with ASD. The resulting data reflect differences in immune molecules, including those of the complement system, which could indirectly affect the developing brain of the ASD. Identical or similar molecules in patients' brains can also show abnormal expression and thus contribute directly to abnormal brain development and ASD. Among the differentially expressed proteins of the assay are those which; involved in a cluster of proteins with a role in the regulation of the Notch signaling pathway, proteins involved in the innate immune response (complement activation), including those involved in the complement cascade, complement C4-A protein and complement C1q subcomponent

subunit A. Nucleosome assembly and positioning, axons guidelines, proteins involved in cholesterol metabolism and lipoprotein transport such as apolipoprotein C-II, apolipoprotein C4 (APOC4) and apolipoprotein F as well as proteins activating T cells involved in the immune response, cholesterol, and cholesterol the process of positive regulation of fatty acid biosynthesis, show significant differential regulation. Among the differentially expressed proteins of the assay are proteins which; involved in a cluster of proteins with a role in the regulation of the Notch signaling pathway, proteins involved in the innate immune response (complement activation), including those involved in the complement cascade, complement C4-A protein and complement C1q subcomponent subunit A. Nucleosome assembly and positioning, axons guidelines, proteins involved in cholesterol metabolism and lipoprotein transport such as apolipoprotein C-II, apolipoprotein C4 (APOC4) and apolipoprotein F, as well as proteins involved in the activation of T cells, immune response, cholesterol efflux and in the process of positive regulation of fatty acid biosynthesis, show significant differential regulation. Apolipoproteins (Apo) are involved in the transport of lipids, cholesterol, and vitamin E into the extracellular circulation and play a common role in maintaining lipid homeostasis. Apolipoprotein C-II plays an important role in lipoprotein metabolism as an activator of lipoprotein lipase. Apolipoprotein C4 (APOC4) is a lipid-binding protein belonging to the apolipoprotein gene family. Apolipoprotein F, this protein forms complexes with lipoproteins and may be involved in the transport and/or esterification of cholesterol. The observed changes in serum protein levels, tropomyosin alpha-4 chain showing decreased expression, and MYH9 isoform 1 of myosin-9 showing increased expression in individuals with ASD complement the evidence for the involvement of oxidative stress in ASD and provide a possible alternative for assessing the early risk of developing ASD. Another study examining the activity of antioxidant enzymes in ASD revealed evidence of decreased expression. It is possible that the mechanism underlying the increase in enzyme levels is activated as a result of the need to compensate for the observed decreased enzyme activity. Further studies in this direction would help to clarify the role of stress-associated markers and the mechanisms of their regulation in ASD. Another important role of complement is that of the system in the brain. Complement plays a direct role in the nervous system and is likely to be involved in cell death and cell apoptosis, both during normal development and in CNS diseases such as Alzheimer's, Huntington's disease, stroke, and many others. Most of the complement proteins found in the blood are also expressed in the brain, mainly by microglia and to a lesser extent by damaged or stressed astrocytes and neurons. In the context of the observed change in complement serum proteins in individuals with ASD in the present study, it is interesting to note the study by Vargas et al. which provides evidence of inflammatory processes in the brains of patients with ASD. The authors found evidence of activation of microglia in the cortex, subcortical white matter, and cerebellum of patients with ASD associated with evidence of a chronic and prolonged neuroglial inflammatory response. It has been suggested that microglial

activation may cause complement activation, which would contribute to brain damage. Although the proteomic approach seems promising in detecting putative biological markers, the data also highlight the main difficulties facing previous and future studies. The small changes observed in the levels of specific proteins may be an indication that there are specific subgroups in the sample that take into account the changes found in the group as a whole; or that the detected changes in serum are indeed small and would require well-controlled future confirmatory studies. One of the main challenges for proteomic research relates to the criteria used to identify proteins. Proteomic studies are one of the analytical studies in which there is a concern for limited reproducibility of results and the need to unify and standardize the analytical methods used. In addition, methods based on the introduction of a certain amount of external substance, a chemical analog as an internal standard (i.e. 'spiking' of standard reference material), hide many unknowns. The presented study demonstrates the potential of the proteomic approach by providing encouraging preliminary data that should help stimulate the continued search for etiopathology in ASD and new therapeutic approaches through peripheral blood testing.

CONCLUSIONS

In recent years, there has been reported an increasing number of abnormal expression signatures of circulating miRNAs which have been demonstrated to have potential as biomarkers for neuropsychiatric disorders. The research on circulating miRNA, though, is still in its beginning stages. With the progress of novel techniques and further research investigations, circulating miRNAs undoubtedly show a great promise in diagnosing CNS disorders and evaluating related therapies. The role of miRNAs in psychiatric disorders and ASD will be further elaborated using continuously improved relevant approaches. In addition, a meta-analysis of miRNAs, covering genetic variation, expression, and biological function will provide valuable information for the potential role of miRNA in ASD, and this could help the diagnosis and prognosis of ASD and psychiatric disorders. Moreover, miRNA biomarkers could be very useful in distinguishing of different subtypes of psychiatric disorder. Finally, the results contribute to the new course of miRNA research in ASD biology but it is only a small part of the long validation processes of miRNA dysregulation in ASD patients. In conclusion, the presented findings support the fact that levels of circulating miRNAs can be dysregulated under certain disease conditions. The present work suggests that this could reflect a metabolic imbalance of this miRNA in vivo and these miRNAs could be used as biomarkers of ASD. A greater sample size of ASD patients is still needed in order to support these conclusions. Transcriptomic data based on the deep RNA-Seq approach can provide valuable information on differential gene expression. It is believed that transcriptomic profiling based on the RNA-Seq approach offers significant promise towards precision medicine and systems diagnostics. In conclusion, this study revealed the utility of whole-exome sequencing and a bioinformatics analysis process

for identifying possible causative variants of ASD. These results also suggest that in the near future diagnostic whole-exome sequencing would be an efficient primary diagnostic method for ASD patients. Using DGE to identify mRNA expression genome-wide has the capacity to add a novel dimension to our understanding of transcriptional regulation in schizophrenia.

LIST OF PUBLICATIONS

1. Circulating miRNAs as a novel class of potential diagnostic biomarkers in neuropsychiatric disorders. Tatyana M. Kichukova, Nikolay T. Popov, Hristo Y. Ivanov, **Tihomir I. Vachev** *Folia Medica 2016; 57(3&4): 159-172.*

2. Autism Spectrum Disorder – A Complex Genetic Disorder. Hristo Y. Ivanov, Vili K. Stoyanova, Nikolay T. Popov, **Tihomir I. Vachev** *Folia Medica* 57(1) (2015).

3. Profiling of circulating serum microRNAs in children with autism spectrum disorder using stem-loop qRT-PCR assay. Tatyana M. Kichukova, Nikolay T. Popov, Ivan S. Ivanov, **Tihomir I. Vachev** *Folia Medica* 2017; 59(1).

4. Identification of a novel mitochondrial mutation in the cytochrome c oxidase III gene in children with autistic sprectrum disorders using next generation RNA-sequencing. Danail Minchev, Nikolay Popov, Veselin Petrov, Ivan Minkov, **Tihomir Vachev,** *Comptesrendus de l'Académiebulgare des Science***. 2019** ***IN PRESS***

5. Epigenetic aspects in schizophrenia etiology and pathogenesis. Nikolay T. Popov, Vili K. Stoyanova, Nadezhda P. Madzhirova, **Tihomir I. Vachev** *Folia Medica* 2012; 54(2): 12-16

6. Investigation of fasciculation and elongation protein zeta-1 (FEZ1) gene expression changes in schizophrenia. **Tihomir Iliev Vachev**, Vili Krasteva Stoyanova, Ivan Minkov, Nikolay Todorov Popov *BJMG* **18 (1), 2015** l 31-38

7. Investigation of circulating serum microRNAs miR-328-3p and miR-3135a expression as a promising novel biomarkers for ASD. Nikolay Popov, Danail Minchev, Ivan S. Ivanov, Vili Stoyanova, **Tihomir Vachev** *BJMG* **21 (2) 2018**

8. Blood-Based Gene Expression in children with Autism spectrum disorder. Hristo Y. Ivanov1, Vili K. Stoyanova, Nikolay T. Popov, **Tihomir I. Vachev,** *Biodiscovery* 2015; **17**: 2; DOI: 10.7750

9. Comparative expression analysis of miR-619-5p in serum and PBMCs as a promising candidate biomarker for autism spectrum disorder. Nikolay T. Popov, Veselin D. Petrov, Danail S. Minchev, Ivan N. Minkov, Tihomir I. Vachev, *Comptesrendus de l'Académiebulgare des Science.* 2019 *IN PRESS.*