COMPARATIVE EXPRESSION ANALYSIS OF HUMAN
ENDOGENOUS RETROVIRUS ELEMENTS IN BLOOD
SAMPLES OF CHILDREN WITH AUTISM SPECTRUM
DISORDER

Danail S. Minchev^{*,**,***, #}, Nikolay T. Popov^{****},
Samir I. Naimov^{*}, Ivan N. Minkov^{*****}, Tihomir I. Vachev^{*}

Received on May 17, 2020
Presented by I. Ivanov, Member of BAS, on June 30, 2020

Abstract

Autism spectrum disorder (ASD) is a highly diverse group of developmental disorders with a multicausal etiology and a strong genetic basis. About 8% of the human genome consists of human endogenous retrovirus elements (HERVs). HERV transcription appears to be altered not only in specific brain regions but also in the peripheral blood of patients with psychiatric conditions such as schizophrenia, multiple sclerosis (MS), ASD, and different types of cancer. This study aims to investigate the blood-based expression of five distinct human endogenous retrovirus elements in a patient cohort of 31 children with Autism Spectrum Disorder (ASD) and a control group of 27 children with typical development.

The comparative HERV expression analysis in 31 children with a diagnosis of ASD and 27 age- and sex-matched healthy controls was performed via a whole transcriptome sequencing approach combined with qRT-PCR analysis.

Two HERV-encoded genes: HERV-P env, and HERV-R env, were found to be significantly downregulated in the group of children with ASD. Another HERV gene, HERV-W pol appeared to be significantly upregulated. Receiver operating characteristic (ROC) analysis indicated that the observed changes in

[#]Corresponding author.

This project received funding from the European Union's Horizon 2020 Programme under grant agreement No 739582 (Project PlantaSYST).

DOI:10.7546/CRABS.2022.02.08

transcript levels of HERV-P env and HERV-R env can discriminate between ASD patients and healthy controls with considerable sensitivity and specificity.

Our results alongside a large number of expression studies conducted by other researchers suggest that human endogenous retrovirus elements may participate in important physiological processes and serve as key factors and markers for many conditions including ASD.

Key words: autism, qRT-PCR, HERV expression, relative quantitation, whole transcriptome sequencing

Introduction. The term Autism Spectrum Disorder (ASD) denotes a wide range of psychiatric conditions affecting children's behaviour mainly in three major aspects: a narrow range of social interactions, impaired verbal communication skills, and stereotypic repetitive behaviour [1]. ASD etiology is highly heterogeneous with various genetic and environmental factors contributing to autism development. Twin studies have undoubtedly revealed the strong heritability of autism with concordance rates among monozygotic twins higher than 75% and much lower concordance among the dizygotic twins [2]. Rare and common ASD-related genetic variants have previously underlined the strong genetic basis of ASD. Nevertheless, known pathology-associated genetic variants only account for 10–20% of all ASD cases, which suggests the potential importance of not only genetic but also environmental factors in ASD etiology. Indeed, advanced parental age, birth complications, vitamin D insufficiency, and environmental toxicants have all been associated with a higher risk of ASD [3].

From another perspective, human endogenous retroviruses (HERVs), may provide a missing link between environmental and genetic factors in the context of complex pathologies. HERVs originate from ancient infectious retroviruses that have integrated into the genomes of the germ-line cells of our ancestors throughout the evolution and started to be transmitted vertically across many generations [4]. HERVs often retain their original proviral structure and consist of two long terminal repeats (LTRs) and an internal region with four characteristic genes: *gag*, *pol*, *pro*, and *env*. Dysregulation of human endogenous retrovirus elements has been associated with various psychiatric and non-psychiatric conditions such as multiple sclerosis (MS), schizophrenia and bipolar disorder, autoimmune conditions, and different types of cancer [5].

Materials and methods. This study included a cohort of 31 children with idiopathic autism spectrum disorder and 27 age- and sex-matched non-autistic controls. All participants with autism spectrum disorder were examined and diagnosed by certified psychiatrists.

Total RNA was extracted from the whole blood cell lysates using a PAXgene Blood miRNA extraction kit according to the manufacturer's protocol (PreAnalytiX, Feldbachstrasse, Switzerland).

Pooled samples were prepared by mixing equivalent amounts of total RNA from each individual sample to a final quantity of 5 µg RNA per pool. Pooled

RNA samples were precipitated, frozen at -80°C , and shipped on dry ice. Whole transcriptome sequencing (RNA-seq) of the pooled samples was carried out by the Beijing Genomics Institute (Shenzhen, China) on an Illumina HiSeq 2000 sequencer.

Reads generated by the whole transcriptome sequencing (fastq files containing more than 27.5 million reads for each pooled sample) were mapped to the UCSC hg38 human assembly using the TopHat 2.0 high-throughput alignment program. The presence of HERV-derived mRNAs was detected by calculating the RNA-seq reads aligning to known HERV loci. Several tools for differential expression analysis were also used: FeatureCounts, Limma, and EdgeR.

Prior to copy-DNA synthesis 1 unit of RNase-free DNase I (Promega, USA) was used per 1 μg of total RNA, for 60 min at 37°C for complete degradation of the residual DNA contaminations. Reverse transcription reactions were carried out in a final reaction volume of 20 μL using 1 μg of the total RNA samples, treated with DNase I, 4 μL 5 $\times$ Reaction mix (with the oligo-dT/random hexamer primers provided), 2 μL Maxima Enzyme Mix, and 3 μL nuclease-free water. The obtained cDNA was subsequently diluted 20 times with nuclease-free water and used as a template for Real-Time PCR. The real-time-PCR assay was performed in an ABI 7500 real-time thermal cycler (Applied Biosystems) using Maxima SYBR Green qPCR Master Mix (Thermo Scientific). Detected Ct data for every sample was normalized using the *Glyceraldehyde-3-Phosphate Dehydrogenase* gene as an endogenous control. Reactions were carried out in duplicates. Each experiment was completed with a melting curve analysis to confirm the specificity of amplification and the lack of non-specific amplification. Transcript levels of HERV-related genes were calculated by relative quantification according to the $2^{-\Delta\Delta\text{Ct}}$ method.

Results. Identification of differentially expressed HERV genes in ASD via transcriptome sequencing. The sequencing procedure produced 27 758 304 and 28 358 958 reads associated with the pooled samples from the ASD group and the control group, respectively. No significant difference in the total read counts was observed between the two pool samples. Approximately 1.38 million reads (about 5% of the total number of mapped reads) from the pool RNA sample of the group of children with autism spectrum disorder and approximately 1.5 million (5%) from the pooled sample of the control group were mapped to HERV loci, which indicates their ubiquitous expression in both pools assayed. We then extracted and re-analyzed the whole set of reads aligning only within the known HERV elements to obtain differential expression data for the HERV-related loci only. Following the RNA-seq analysis, we chose five separate HERV-encoded genes for additional qRT-PCR validation, based on their RNA-seq differential expression data.

RNA-seq data confirmation via qRT-PCR. Using the comparative Ct method described by LIVAK et al. [6] we obtained the transcript levels of five HERV-encoded genes. Technical replicates of a single sample with Ct differences

Fig. 1. Fold changes in expression of human endogenous retrovirus elements in children with ASD (Data presented in logarithmic scale)

greater than 0.2 were not accepted, which allowed expression changes as small as $\log_2(\text{foldchange}) = 0.5$ to be identified with high certainty.

Calculated mean levels for the five HERV genes are shown in Fig. 1. Relative quantitation data is presented as a relative fold difference ($\log_2$) for each gene. Differences in expression between the two groups of individuals were evaluated using a two-sample unpaired Mann–Whitney test on the ΔC_t values (Fig. 2) calculated for every separate gene in all individual samples. Differential expression at a level of $\log_2 \text{FC} \leq -0.5$ or $\log_2(\text{Fold Change}) \geq 0.5$ (translating into fold change differences lower than 0.71 or greater than 1.41) combined with a p -value from the Mann–Whitney test less than 0.05 was considered statistically significant. Our data suggested that the transcript levels of two genes, *HERV-P env* ($U = 607$; $z\text{-score} = 3.52$; $p\text{-value} = 0.00042$) and *HERV-R env* ($U = 576$; $z\text{-score} = 3.65$; $p\text{-value} = 0.0226$), were statistically lower in the group of the children with ASD compared with those in the control group. Another gene, *HERV-W pol* ($U = 232$; $z\text{-score} = -2.27$; $p\text{-value} = 0.0226$), appeared to be statistically upregulated in the group of children with ASD. The other two genes studied: *HERV-K (HML-2) gag*, and *HERV-K env*, did not show statistically significant changes between the two groups studied (with p -values higher than 0.05 and absolute fold change lower than 1.4).

In order to determine whether the expression levels of *HERV-P env*, *HERV-R env*, and *HERV-W pol* may serve as potential markers, discriminating ASD cases from healthy controls, we performed a Receiver Operating Characteristic (ROC) analysis on the qRT-PCR data. ROC curves were plotted and area under curve (AUC) values with corresponding confidence intervals (CI) were calculated as follows: AUC 0.775 (CI: 0.644–0.876) for *HERV-P env*, 0.790 (CI: 0.658–0.889) for *HERV-R env*, and 0.681 (CI: 0.540–0.802) for *HERV-W pol*. ROC analysis was

Fig. 2. ΔC_t values for the five HERV-encoded genes of interest. Data is presented in box plot diagrams with a linear scale

also performed in order to obtain the sensitivity and specificity of the predicted markers. Calculated sensitivity and specificity values at optimal cut-off were: 72.4% and 70.4% for *HERV-P env*, 74.1% and 74.1% for *HERV-R env*, and 71.4% and 61.5% for *HERV-W pol*, respectively. Taken together, these results indicate that the observed statistically significant differences in expression of *HERV-P env*, *HERV-R env*, and *HERV-W pol*, can discriminate between ASD cases and healthy controls with considerable accuracy (Fig. 3).

Discussion. In the present study, we describe significantly lower transcript levels of both *HERV-P env* and *HERV-R env* genes in the ASD cohort in comparison with those in the control group. Conversely, the *HERV-W pol* gene appeared to be significantly upregulated in the ASD group. Our results are consistent with other studies that also suggest small but statistically significant alterations in HERV expression [7,8]. These expression studies together with our experimental results demonstrate that even modest changes in HERV gene expression may be considered statistically valid.

Since various mechanisms can contribute to the HERV transcription regulation, the results we observed can be due to many factors that modify HERV transcription. Generally, HERV elements are actively suppressed throughout

Fig. 3. Receiver operating characteristic (ROC) curve analysis of differential expression data for HERV-P *env*, HERV-R *env*, and HERV-W *pol*. ROC curves were calculated using MedCalc commercial statistical software

the genome as a result of CpG-dependent DNA methylation within the 5' LTR which carries specific promoter and enhancer elements of particular importance for HERV regulation [9,10]. Numerous trans-acting transcription factors that bind to specific HERV regulatory sequences have also been discovered: YY1, Sp-1, Oct-1, GATA, GCM1. Only one of them, the GCM1 transcription factor, exhibits a change greater than 1.4-fold between the two groups according to our RNA-seq data. Also, our RNA-seq data suggests no differences in the expression of genes that code for the major de novo DNA methyltransferases. On this basis, we can speculate that the observed differences in the HERV transcript levels can be due to the accumulation of minor changes in various gene regulation processes. Further efforts are still required for a complete characterization of HERV regulation.

Conclusions. Our results show HERV-P *env* and HERV-R *env* downregulation and HERV-W *pol* upregulation in ASD, which suggests the potential role of these elements as markers for the ASD condition.

Acknowledgements. The authors are most grateful to the families of the probands for their collaboration. We thank Prof. V. Stoyanova, Asst. Prof. H. Ivanov and Prof. I. Ivanov (Department of Pediatrics and Medical Genetics, Medical University of Plovdiv, Plovdiv, Bulgaria) for their technical insights and support of this study.

REFERENCES

- [1] American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (DSM-5) (5th ed.), Washington, DC, American Psychiatric Association.
- [2] TICK B., P. BOLTON, F. HAPPÉ, M. RUTTER, F. RIJSDIJK (2016) Heritability of autism spectrum disorders: a meta-analysis of twin studies, *J. Child Psychol. Psychiatry*, **57**(5), 585–595.

- [3] MODABBERNIA A., E. VELTHORST, A. REICHENBERG (2017) Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses, *Mol. Autism*, **8**, 13.
- [4] SHIN W., J. LEE, S. SON, K. AHN, H. KIM et al. (2013) Human-specific HERV-K insertion causes genomic variations in the human genome, *PloS One*, **8**(4), e60605.
- [5] CHRISTENSEN T. (2016) Human endogenous retroviruses in neurologic disease, *APMIS*, **124**(1–2), 116–126.
- [6] LIVAK K., T. SCHMITTGEN (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method, *Methods*, **25**(4), 402–408.
- [7] BHETARIYA P., J. KRIESEL, K. FISCHER (2017) Analysis of human endogenous retrovirus expression in multiple sclerosis plaques, *J. Emerg. Dis. Virol.*, **3**(2), 10.16966/2473-1846.133, 9 pp.
- [8] YAO Y., J. SCHRÖDER, C. NELLÅKER, C. BOTTMER, S. BACHMANN et al. (2008) Elevated levels of human endogenous retrovirus-W transcripts in blood cells from patients with first episode schizophrenia, *Genes Brain Behav.*, **7**(1), 103–112.
- [9] WENTZENSEN N., J. COY, H. KNAEBEL, M. LINNEBACHER, B. WILZ et al. (2007) Expression of an endogenous retroviral sequence from the HERV H group in gastrointestinal cancers, *Int. J. Cancer*, **121**(7), 1417–1423.
- [10] CHENG Y., B. RICHARDSON, M. HUBERT, S. HANDWERGER (2004) Isolation and characterization of the human syncytin gene promoter, *Biol. Reprod.*, **70**(3), 694–701.

**Department of Plant Physiology
and Molecular Biology
University of Plovdiv “Paisii Hilendarski”
24 Tzar Assen St
Plovdiv, Bulgaria
e-mails: naimov0@uni-plovdiv.bg
tihomirvachev@uni-plovdiv.bg*

***Department of Medical Biology
Faculty of Medicine
Medical University-Plovdiv
15A Vassil Aprilov Blvd
4000 Plovdiv, Bulgaria
e-mail: danail.minchev@mu-plovdiv.bg*

****Division of Molecular and
Regenerative Medicine
Research Institute at
Medical University of Plovdiv
15A Vasil Aprilov Blvd
4000 Plovdiv, Bulgaria*

*****Psychiatric Ward for
Active Treatment of Men
State Psychiatry Hospital Pazardzhik
28 Bolnichna St
4400 Pazardzhik, Bulgaria
e-mail: nikpopov@abv.bg*

******Institute of Molecular Biology
and Biotechnologies (IMBB)
105 Ruski Blvd
4000 Plovdiv, Bulgaria
e-mail: minkov@plantgene.eu*