Gene expression profiling in peripheral blood mononuclear cells reveals dependency on multiple sclerosis activity

A. A. Abramov, M. R. Kabilov, O. G. Kulakova, A. E. Tupikin, V. S. Mukhina, L.V. Danilova, P. V. Mazin, O. O. Favorova , A. A. Favorov

Moscow State University, Leninskie gory, str. 12, 119234, Moscow ferk_88@list.ru

Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Lavrentiev Avenue 8, Novosibirsk, 630090, Russia, kabilov@niboch.nsc.ru

Pirogov Russian National Research Medical University, Ostrovityanova str. 1, Moscow, 117997, Russia, olga.koulakova@gmail.com, olga_favorova@mail.ru

Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkina str. 3, Moscow, 119991 , Russia, atgc@vigg.ru, favorov@sensi.org, ludmila.danilova@gmail.com

The Institute for Information Transmission Problems, Bolshoy Karetny per. 19, str.1, Moscow 127051, iaa.aka@gmail.com, atgc@vigg.ru

Johns Hopkins School of Medicine, Baltimore, MD 21205, USA, favorov@sensi.org, ludmila.danilova@gmail.com

Introduction

Multiple sclerosis (MS) is a common complex disease of the central nervous system developing with demyelination and autoimmune inflammation [1]. Pronounced clinical heterogeneity is one of the main MS features. The most common form of MS is relapsing-remitting MS (RRMS), which is characterized by recurrent relapse followed by remission (chronic inflammation). Approximately 10–15% of patients have primary-progressive MS (PPMS) that manifests as an acute inflammation, which results in continuous increase of the neurological deficit at the disease onset [2]. The transcriptome profiling in blood cells of patients with PPMS and RRMS provides a better understanding of the pathogenic mechanisms that underlying the development two main forms of MS.

Methods

We carried out a whole-genome transcriptome profiling in peripheral blood mononuclear cells (PBMC) from representative groups of 7 PPMS patients, 31 RRMS patients in remission and 7 healthy controls (HC) and performed three pairwise comparisons of data for each type of cells.

Total RNA isolation from PBMC was performed with Purelink RNA micro Kit (Life Technologies) with following ribosomal RNA depletion using Ribo-Zero Gold Kit (Epicentre). SOLiD Total RNA-Seq Kit (Life Technologies) was used to construct DNA libraries. DNA libraries were then amplified and sequenced on SOLID 5500xl platform (Life Technologies) with read length of 75 bps and exact call chemistry (ECC) module application to improve sequencing accuracy at the SB RAS Genomics Core Facility.

The analysis of the data consisted of the following steps:

- The gene-based counts for reads as they were calculated by vendor-provided CLC software were analysed by Deseq2 (1.22) R package. No reliable expression differences were revealed. We switched to more detailed analysis, which was based on the raw sequencing data.
- Producing reads were corrected with vendor supplied correction utility SAET (2.4).
- Raw reads were mapped by TopHat (2.1.0) on the human genome sequence (GRCh38.p8) with supplied Ensembl annotation [3].

- A custom annotation was created by Cufflinks package (2.2.1) from TopHat transcriptome data with supplied guide Ensembl annotation [4].
- Raw reads were again mapped by TopHat on the human genome sequence, but using the custom annotation rather than Ensemble one.
- Reads were counted in genes with htseq-count (0.6.1) [5].
- The count data tables were analysed with edgeR-package (3.18) [6,7]. We used the Benjamini–Hochberg procedure to control false discovery rate (FDR) at the level 0.05.
- SAJR (0.01) was used for analysis of differential splicing [8].

Comparisons of patients groups were performed with GLM approach followed by likelihood ratio test. Genes were considered differentially expressed if BH-corrected e p-value was below 0.05 and the absolute log2 fold change (log2FC) was above 2.

Results

Significant differences in gene expression profiles were found for all three pairwise comparisons. RRMS patients in remission were characterized by the differences in the expression of 339 genes, 80% of differentially expressed genes (DEG) were upregulated while compared with healthy controls. When comparing PPMS patients with healthy controls, 71 DEGs were identified, about half of which were upregulated. Only 10 DEGs were revealed by RRMS-PPMS comparison and half of these genes is overexpressed in RRMS patients.

Functional analysis of DEGs showed that major changes in RRMS patients compared to healthy individuals were found in genes involved in regulation of inflammation and immune response (e.g. *IL1R2*, logFC > 2; *NFKB1*, logFC < −2), nervous system development and functioning (e.g. *NBP*, logFC > 2), cell adhesion and migration (e.g. *CD44*, logFC > 2), cytoskeleton and vesicular transport functioning (e.g. *TRAPPC5*, logFC > 2; *ARF4*, logFC < −2), ubiquitin-dependent protein degradation (e.g. *ARIH1*, logFC < −2; *FBXL8*, logFC > 2), regulation of gene expression (e.g. *EGR1*, logFC < −3; *SRCAP*, logFC > 2).

PPMS patients compared to healthy controls are characterized by changes in expression of a number of immune response genes (e.g. *JCHAIN*, logFC > 2; *IGLL5*, logFC > 3), transcription factors (e.g. *EGR1*, logFC < −2.9; *JUN*, logFC < −2) and genes involved in vesicular transport regulation (e.g. *NFKB1*, logFC < −2; *ARF4*, logFC < −2), cell adhesion (e.g. *VCAN*, logFC > 2.5; *CD36*, logFC > 2), lipid biogenesis (e.g. *ARSG*, logFC > 2.5; *ELOVL5*, logFC < −2) and nervous system functioning (e.g. *DNAJC5*, logFC > 2; *TBR1*, logFC < −2).

At the same time, the direct comparison of gene expression profiles in RRMS and PPMS patients revealed DEGs associated with ubiquitin-dependent protein degeneration (e.g. *ISG15*, logFC > 3.5), apoptosis regulation (e.g. *GADD45B*, logFC > 2.5), methyltransferase activity (e.g. *THUMPD2*, logFC < −3) and expression of small RNA genes as well (e.g. *MIR7641-2*, logFC < −7.8). *THUMPD2* increased expression was observed in both PPMS comparisons (PPMS vs RRMS, PPMS vs healthy controls).

Conclusions

The genomewide differential transcriptome analysis of PBMCs from RRMS patients in remission, PPMS patients, and healthy controls identified patterns of differentially expressed genes, which were associated with chronic and/or acute inflammation. The expression levels for the majority of the DEGs were increased during the chronization of inflammatory process (in RRMS in remission).

Acknowledgements

The authors would like to thank Dr. E. V. Popova and Dr. A. N. Boyko for the recruiting and clinical characterization of the multiple sclerosis patients and N. M. Baulina, V. B. Bashinskaya and I.S. Kiselev for sample preparations.

1. A. Nylander, D.A. Hafler (2012) Multiple sclerosis. *J. Clin. Invest.* 122, 1180-1188.

2. P. Iwanowski , J. Losy (2015) Immunological differences between classical phenotypes of multiple sclerosis. *J Neurol Sci.*, 349(1-2):10-14.

3. K. Daehwan et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 14 (4): R36.

4. A. Roberts et al. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics.* 27 (17): 2325-2329.

5. S. Anders et al. (2014) HTSeq — A Python framework to work with high-throughput sequencing data. *Bioinformatics.* 31 (2): 166-169.

6. M. Robinson et al. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26 (1): 139-140.

7. J. McCarthy et al (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research.* 40 (10): 4288-4297.

8. P. Mazin et al. (2013) Widespread splicing changes in human brain development and aging. *Molecular Systems Biology.* 9 (1): 9-633.