

## Исследование особенностей профилей экспрессии РНК у пациентов с COVID-19

Шиманский В.С.<sup>1,2\*</sup>, Попов О.С.<sup>1,2</sup>, Ключкова Т.Г.<sup>1</sup>, Апалько С.В.<sup>1,2</sup>,  
Сушенцева Н.Н.<sup>1,2</sup>, Асиновская А.Ю.<sup>1</sup>, Мосенко С.В.<sup>1,2</sup>, Сарана А.М.<sup>2</sup>,  
Щербак С.Г.<sup>1,2</sup>

<sup>1</sup> Санкт-Петербургское ГБУЗ «Городская больница № 40 Курортного района» отдела здравоохранения администрации Курортного района г. Санкт-Петербурга, Сестрорецк, Россия

<sup>2</sup> Санкт-Петербургский государственный университет Правительства Российской Федерации, Санкт-Петербург, Россия

\* shimansky.valya@yandex.ru

**Ключевые слова:** дифференциальная экспрессия генов; COVID-19; секвенирование РНК

**Мотивация и цель:** Исследования, посвященные анализу периферических транскриптомных сигнатур COVID-19, активно ведутся во всем мире. Одно из таких исследований, проведенное китайскими учеными, показало 268 генов с измененной экспрессией, что привело к выявлению ключевых регуляторов убиквитинирования белков, включая 6 факторов транскрипции и 2 микроРНК [1]. Другое исследование выявило значительные различия в экспрессии 25 482 мРНК, 23 микроРНК и 410 длинных некодирующих РНК между пациентами с COVID-19 и контрольной группой. Отмечено, что изменения в экспрессии мРНК связаны с иммунным ответом, в частности с процессингом и представлением антигена, а также с активацией Т-клеток [2]. Транскриптомный анализ предоставляет возможность оценки молекулярного ответа организма на инфекцию и может помочь в выявлении потенциальных мишеней для терапии.

Целью данного исследования является дифференциальный анализ экспрессии генов для выявления набора регуляторных генов, влияющих на ключевые молекулярные и клеточные пути, вовлеченные в патогенез COVID-19.

**Методы и алгоритмы:** В исследовании приняли участие 146 пациентов с COVID-19. Все пациенты подписали информированное добровольное согласие на участие в исследовании. Исследование было одобрено экспертным советом по этике СПб ГБУЗ «Городская больница № 40» (протокол № 171 от 18 мая 2020 г.). Пациенты были разделены на основе степени тяжести заболевания: первая группа включала случаи с легким и умеренным течением, вторая группа – с тяжелым и крайне тяжелым. Дополнительно учитывался исход заболевания (выживание/летальный исход).

РНК была выделена из образцов крови вручную с помощью фенол-хлороформенного метода экстракции. Подготовка библиотек проведена с использованием реагентов KAPA RiboErase HMR (Roche), RNA Hyper Prep (Roche), KAPA Universal Adapter (Roche), KAPA UDI Primer Mixes (Roche). Конверсия проведена с использованием реагентов High-Throughput Sequencing Primer Kit (App-C) и MGIEasy Universal Library Conversion Kit (App-A) (MGI). Секвенирование полученных библиотек произведено на секвенаторе MGISEQ-2000 методом парно-концевого прочтения длиной 100 пар оснований на ячейке DNBSEQ-G400 High-throughput Sequencing Set (PE100, 360 Гб) (MGI).

Контроль качества прочтений выполнен в FastQC. Выравнивание прочтений на референсный геном версии hg38 выполнено программой STAR. Подсчет прочтений и аннотация выполнены программой featureCounts. Дифференциальная экспрессия генов оценивалась с помощью непараметрического критерия Манна–Уитни. Расчеты проводились в программном пакете R. Далее на базе платформы [ge.genexplan.com](http://ge.genexplan.com) произведен анализ обогащения дифференциально экспрессируемыми генами метаболических путей и поиск мастер-регуляторов.

**Результаты:** Мы провели анализ дифференциальной экспрессии генов, разделив образцы на группы по нескольким критериям. Уровень статистической значимости  $p\text{-value} \leq 0.01$ .

1. Группы, выделенные по тяжести заболевания.

Мы выявили 806 генов с повышенной экспрессией и 3925 генов с пониженной экспрессией у группы с тяжелым течением COVID-19 по сравнению с группой с умеренным и легким течением. Анализ по базе данных Gene Ontology (GO) выявил следующие категории.

Гены с повышенной экспрессией концентрируются в метаболических путях, отвечающих за адаптивный иммунный ответ, сигналинг, межклеточную коммуникацию и ряд других. Гены с пониженной экспрессией сосредоточены в молекулярных путях, отвечающих за процессы метаболизма клеточных соединений азота и гетероциклов, генную экспрессию и т.д. (рис. 1).



**Рис. 1.** Наиболее значимые результаты анализа обогащения молекулярных путей по GO сверхэкспрессированными генами (слева) и с пониженной экспрессией (справа)

Анализ мастер-регуляторов показал следующие результаты: *CDH2*, *CXCL9*, *KIT*, *IL13*, *EPHA4*, *FLT3*, *DUSP6*, *FGFR2*, *IL4* для генов с повышенной экспрессией; *RHOH*, *NFATC3*, *FXN*, *ADRB2*, *NTRK1*, *NAA30*, *ZC3H12D*, *MAF*, *KLRD1* для генов с пониженной экспрессией.

2. Группы, выделенные по исходу заболевания.

Из 9869 генов, прошедших по порогу статистической значимости, 348 показали повышенную экспрессию и 9521 пониженную у группы с летальным исходом по сравнению с группой выживших.

Проведя анализ обогащения по базе данных GO, мы выяснили, что гены с повышенной экспрессией чаще встречаются в метаболических путях, отвечающих за дегрануляцию нейтрофилов, нейтрофил-опосредованный иммунитет, ответ на внешние раздражители и ряд других. Гены с пониженной экспрессией сосредоточены в молекулярных путях, отвечающих за метаболизм клеточных соединений азота, соединений, содержащих азотистые основания, гетероциклов и т.д. (рис. 2).



**Рис. 2.** Наиболее значимые результаты анализа обогащения молекулярных путей по GO сверхэкспрессированными генами (слева) и с пониженной экспрессией (справа)

Анализ мастер-регуляторов показал следующие результаты: *MAPK3*, *MMP9*, *PRL*, *MMP2*, *FLT3*, *S100AB*, *ITGB3*, *ITGA2B*, *PPARG* для генов с повышенной экспрессией; *PRNP*, *MAPK8*, *ERBB2*, *ZC3H122D*, *ITK*, *NAA30*, *TRIM3*, *ESR1*, *SH3RF2*, *ADRB2*, *KLRC1*, *KLRD1*, *PRKCH* для генов с пониженной экспрессией.

**Выводы:** По результатам данного исследования мы выявили ряд дифференциально экспрессируемых генов, потенциально объясняющих более тяжелое течение COVID-19, а также несколько мастер-регуляторов, чьи функции и роль в путях, обогащенных дифференциально экспрессированными генами, делают их потенциальными целями для биохимических и метаисследований. Наиболее интересными находками стали гены *ALOX15*, *IL4* и *KNDC1*. *ALOX15* имеет пониженную экспрессию в группах тяжелого течения и летального исхода, что свидетельствует об идущем в организме процессе воспаления и нарушенном механизме его терминации, потенциально ведущем к более тяжелому течению заболевания и исходу [3]. Также этот ген был выявлен как мастер-регулятор, что делает его кандидатом для исследования в качестве биомаркера и цели для терапии. У группы с летальным исходом снижена экспрессия гена *IL4*, кодирующего цитокин, являющегося ключевым регулятором гуморального и адаптивного иммунного ответа. Хотя прямой ассоциации *IL4* с тяжелым течением заболевания выявлено не было, этот ген представляет интерес для дальнейшего изучения [4]. Наш анализ показал, что *IL4*, предположительно, является одним из мастер-регуляторов для сверхэкспрессированных генов в группе с тяжелым течением заболевания. Ген *KNDC1*, участвующий в сигналинге, распознавании белков и функциональной регуляции, демонстрирует сниженную экспрессию у пациентов с тяжелым течением COVID-19 и летальным исходом.

**Финансирование:** Исследование выполнено в рамках проекта Санкт-Петербургского государственного университета (Pure ID 95412780) и при взаимодействии с базой Ресурсного центра Биобанк.

## Expression profile analysis of COVID-19 patients

Shimansky V.S.<sup>1, 2\*</sup>, Popov O.S.<sup>1, 2</sup>, Klochkova T.G.<sup>1</sup>, Apalko S.V.<sup>1, 2</sup>,  
Sushentseva N.N.<sup>1, 2</sup>, Asinovskaya A.Yu.<sup>1</sup>, Mosenko S.V.<sup>1, 2</sup>, Sarana A.M.<sup>2</sup>,  
Shcherbak S.G.<sup>1, 2</sup>

<sup>1</sup> City Hospital No. 40 of Health Department of the St. Petersburg Kurortniy District Administration, Sestroretsk, Russia

<sup>2</sup> St. Petersburg State University, Government of the Russian Federation, St. Petersburg, Russia

\* shimansky.valya@yandex.ru

**Key words:** differential expression; RNAseq; COVID-19

*Motivation and Aim:* Studies dedicated to the analysis of peripheral transcriptomic signatures of COVID-19 are actively being conducted worldwide. One such study by Chinese scientists identified 268 genes with altered expression, leading to the discovery of key regulators of protein ubiquitination, including 6 transcription factors and 2 microRNAs [1]. Another study from China revealed significant differences in the expression of 25,482 mRNAs, 23 microRNAs, and 410 long non-coding RNAs between patients with COVID-19 and a control group. It was noted that changes in mRNA expression are associated with the immune response, particularly with antigen processing and presentation, as well as T-cell activation [2]. Transcriptomic analysis provides an opportunity to assess the molecular response of the body to infection and may assist in identification of potential targets for therapy.

The aim of this study is to perform a differential gene expression analysis to identify a set of regulatory genes that affect key molecular and cellular pathways involved in the pathogenesis of COVID-19.

*Methods and Algorithms:* The study involved 146 patients with COVID-19. All patients signed an informed voluntary consent to participate in the study. The research was approved by the ethics committee of the St. Petersburg State Budgetary Healthcare Institution “City Hospital No. 40” (protocol No. 171 of May 18, 2020). Patients were divided based on the severity of the disease: the first group included cases with mild and moderate progression, while the second group included severe and critically severe cases. Additionally, the outcome of the disease (survival/lethal outcome) was considered.

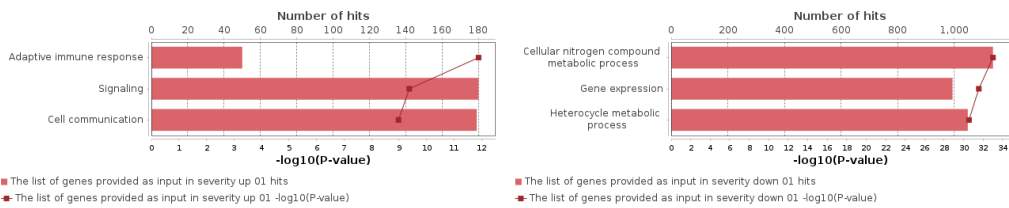
RNA was manually extracted from blood samples using the phenol-chloroform extraction method. Library preparation was conducted using KAPA RiboErase HMR (Roche), RNA Hyper Prep (Roche), KAPA Universal Adapter (Roche), and KAPA UDI Primer Mixes (Roche) reagents. Conversion was performed using High-Throughput Sequencing Primer Kit (App-C) and MGIEasy Universal Library Conversion Kit (App-A) (MGI) reagents. The sequencing of the obtained libraries was carried out on the MGISEQ-2000 sequencer by the paired-end reading method with a length of 100 base pairs on the DNBSEQ-G400 High-throughput Sequencing Set (PE100, 360 Gb) (MGI). Read quality control was performed in FastQC. Reads were aligned to the reference genome version hg38 using the STAR aligner. Read counts and annotation were performed with featureCounts. Differential gene expression was assessed using the Mann-Whitney nonparametric test. Calculations were conducted in the R software package. Subsequently, an analysis of the enrichment of differentially expressed genes in metabolic pathways and the search for master regulators were carried out on the [ge.genexplain.com](http://ge.genexplain.com) platform.

*Results:* We performed an analysis of differential gene expression, dividing the samples into groups based on several criteria. The level of statistical significance  $p$ -value is  $\leq 0.01$ .

#### 1. Groups divided by disease severity.

We identified 806 genes with increased expression and 3,925 genes with decreased expression in the group with severe COVID-19 compared to the group with moderate and mild progression. Gene Ontology (GO) enrichment analysis revealed the following categories. Genes with increased expression are concentrated in metabolic pathways responsible for adaptive immune response, signaling, intercellular communication, and others. Genes with decreased expression are focused on molecular pathways responsible

for the metabolism of cellular nitrogen compounds and heterocycles, gene expression, etc. (Fig. 1).

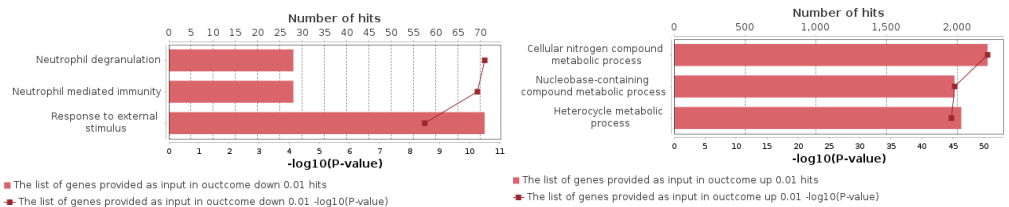


**Fig. 1.** The most significant results of the GO enrichment analysis of molecular pathways for overexpressed genes (on the left) and underexpressed genes (on the right)

The analysis of master regulators showed the following results: *CDH2*, *CXCL9*, *KIT*, *IL13*, *EPHA4*, *FLT3*, *DUSP6*, *FGFR2*, *IL4* for genes with increased expression; *RHOH*, *NFATC3*, *FXN*, *ADRB2*, *NTRK1*, *NAA30*, *ZC3H12D*, *MAF*, *KLRD1* for genes with decreased expression.

2. Groups divided by outcome of the disease.

Out of 9,869 genes that passed the threshold of statistical significance, 348 showed increased expression and 9,521 showed decreased expression in the group with lethal outcomes compared to the survivors' group. By conducting an enrichment analysis using the GO database, we found that genes with increased expression are more frequently involved in metabolic pathways responsible for neutrophil degranulation, neutrophil-mediated immunity, response to external stimuli, and others. Genes with decreased expression are concentrated in molecular pathways responsible for the metabolism of cellular nitrogen compounds, nucleobase-containing compound metabolic process, heterocycles, etc. (Fig. 2).



**Fig. 2.** The most significant results of the GO enrichment analysis of molecular pathways for overexpressed genes (on the left) and underexpressed genes (on the right)

The master regulator analysis revealed the following results: *MAPK3*, *MMP9*, *PRL*, *MMP2*, *FLT3*, *S100AB*, *ITGB3*, *ITGA2B*, *PPARG* for genes with increased expression; *PRNP*, *MAPK8*, *ERBB2*, *ZC3H122D*, *ITK*, *NAA30*, *TRIM3*, *ESR1*, *SH3RF2*, *ADRB2*, *KLRC1*, *KLRD1*, *PRKCH* for genes with decreased expression.

**Conclusion:** Based on the results of this study, we have identified a number of differentially expressed genes that potentially explain the more severe progression of COVID-19, as well as several master regulators whose functions and roles in pathways enriched with differentially expressed genes make them potential targets for biochemical and meta-research. The most interesting findings were the genes *ALOX15*, *IL4*, and *KNDC1*. *ALOX15* has decreased expression in groups with severe progression and lethal outcomes, indicating an ongoing inflammatory process in the body and a disrupted mechanism of its termination, potentially leading to a more severe course of the disease

and outcome [3]. This gene was also identified as a master regulator, making it a candidate for research as a biomarker and therapeutic target. The group with lethal outcomes showed decreased expression of the *IL4* gene, which encodes a cytokine that is a key regulator of humoral and adaptive immune responses. Although no direct association of *IL4* with severe disease progression was identified, this gene is of interest for further study [4]. Our analysis suggests that *IL4* is presumably one of the master regulators for the overexpressed genes in the group with severe disease progression. The *KNDC1* gene, which is involved in signaling, protein recognition, and functional regulation, shows decreased expression in patients with severe COVID-19 and lethal outcomes.

**Funding:** The study was supported by St. Petersburg State University, project ID: 95412780. The study was carried out in cooperation with the Core facility center Biobank.

### *Список литературы/References*

1. Huang J. et al. Transcriptome analysis of peripheral blood mononuclear cells response in patients with severe COVID-19 reveals crucial genes regulating protein ubiquitination. *Med Sci Monitor*. 2022;28:e937532. doi 10.12659/MSM.937532
2. Li C. et al. Whole-transcriptome RNA sequencing reveals significant differentially expressed mRNAs, miRNAs, and lncRNAs and related regulating biological pathways in the peripheral blood of COVID-19 patients. *Mediators Inflamm*. 2021;2021:6635925. doi 10.1155/2021/6635925
3. Tian R. et al. ALOX15 as a suppressor of inflammation and cancer: lost in the link. *Prostaglandins Other Lipid Mediat*. 2017;132:77-83
4. Chang Y., Mengru B., Qinghai Y. Associations between serum interleukins (IL-1 $\beta$ , IL-2, IL-4, IL-6, IL-8, and IL-10) and disease severity of COVID-19: a systematic review and meta-analysis. *BioMed Res Int*. 2022;30:2022:2755246. doi 10.1155/2022/2755246