Published as a manuscript

**Liu Guojun**

# PATHOPHYSIOLOGIC GENE DEPENDENT MECHANISMS OF SEPARATE TYPES OF IMMUNE MEDIATED PATHOLOGY

14.03.03 – Pathological physiology

ABSTRACT

Of dissertation
for the science degree
of candidate of biological sciences

Yekaterinburg- 2019

The work was performed at the Department of Medical Biochemistry and Biophysics, Institute of Natural Sciences and Mathematics, Federal State Autonomous Educational Institution of Higher Education “Ural Federal University” named after the First President of Russia B.N. Yeltsin.

**Scientific advisers**:

Doctor of Medical Sciences, Professor, Honored Scientist of the Russian Federation — **Tuzankina Irina Aleksandrovna**

Academician of the Russian Academy of Sciences, Doctor of Medical Sciences, Professor — **Chereshnev Valery Aleksandrovich**

**Official opponents:**

Doctor of Biological Sciences, Professor, Director of the Higher Medical and Biological School, Head of Laboratory of Advanced Studies of Molecular Stress Mechanisms Federal State Autonomous Educational Institution of Higher Education “South Ural State University (national research university)” FSAEIHE SUSU (NRU) Constitutor is Ministry of Science and Higher Education of the Russian Federation. — **Zeilikman Vadim Eduardovich**

Doctor of Medical Sciences, Director of the Center for Immunology and Cellular Biotechnology, Professor, Chair of Fundamental Medicine, Medical Institute Immanuel Kant Baltic Federal University — **Litvinova Larisa Sergeevna**

**Leading organization** - Federal State Budget Scientific Institution “Institute of General Pathology and Pathophysiology” (Moscow)

Defense will be held on February, 27, 2020 at ____ hours on the meeting of thesis defense board for the degree of candidate of science, for the degree of doctor of science D 004.027.02 based on IIP UB RAS (620049, 106, Pervomayskaya, Yekaterinburg)

A dissertation and an abstract can be found at the Central Scientific Library of the Ural Branch of the Russian Academy of Sciences and on the IIP UB RAS website - http://iip.uran.ru , with an abstract - on the website of the Higher Attestation Commission of the Russian Federation http://vak2.ed.gov.ru.

Abstract sent «___» _____ 2020 года

Scientific Secretary of the thesis defense board D.004.027.02
on the basis of the Institute of Immunology and Physiology, Ural Branch of the Russian Academy of Sciences,
Doctor of Medical Sciences, Professor, Honored Scientist of the Russian Federation — I.A. Tuzankina

# GENERAL DESCRIPTION OF THE WORK

## Relevance and degree of elaboration of the research topic

Modern pathophysiology involves a systematic approach for the analysis of any immunopathological process in the body. The rapid development of life sciences, which are integrative and unifying for many areas of knowledge, indicates that people have entered the era of "big data." In this era, the requirements for clinical medical research are constantly increasing, and emerging problems are becoming increasingly complex. The interdisciplinary model (pathophysiology, immunology, genetics and bioinformatics) is gradually becoming the main trend in the development of medical and biological research.

Due to the development of genetics, it became possible to identify general patterns of network interactions within the genome, which help explain how the body responds with numerous multidirectional processes in response to a specific point impact. Moreover, the components of innate and adaptive immunity are the main executors of maintaining homeostasis.

The use of new knowledge about the pathogenetic mechanisms of the development of immune-mediated pathology will allow timely application of methods for determining disease markers for their early diagnosis, pathogenetic substantiation of new therapeutic technologies, prevention of complications of medical manipulations and the development of irreversible changes in organs and systems, predicting the course of the disease and monitoring the effectiveness of therapy.

In our study, diseases such as bladder cancer and primary immunodeficiency (PID) were selected as models of immune-dependent pathology, in which attention is paid to one of the most common forms of PID - common variable immune deficiency and one of the rarest forms - Henneckam syndrome.

Bladder cancer ranks ninth among the most common cancers worldwide with an annual record of about 430,000 new cases (Cumberbatch M., et al., 2018). It is known that micro miRNA in the extracellular medium is involved in the regulation of local processes. Recently, the study of the regulatory network of miRNAs of the miR-200 family and their target genes has been given close attention by patho-physiologists and immunologists (Pieraccioli M., et al., 2013, and Guan T., et al., 2018). MiR-200 family (miR-200a, miR-200b, miR-200c, miR-141 and miR-429) have been shown to be critical in cancer pathogenesis and metastasis (Humphries B. and Yang C., 2014), and also are necessary for the formation of CD8 + memory T cells (Guan T., et al., 2018). Researchers determined that miR-200-controlled zinc-dependent transcription factors ZEB1 and ZEB2 are involved in various processes of immune-genesis and hematopoiesis, therefore miR-200 overexpression can have a contradictory effect on the development of

various types of tumors, including affecting mesenchymal transformation. Reducing the expression of tumor-suppressing miR-200 leads to overexpression of genes encoding the so-called “immune control check-points” - immunosuppressive proteins (including PD1 ligand, Chen L., 2014), as a result of which cancer cells become immune-privileged due to apoptosis of helper and cytotoxic T cells involved in antitumor immunity (Halytskiy V., 2018). Therefore, miR-200 miRNA is a target of immunotherapy for various types of cancer, including bladder cancer, and the study of the interaction of different genes and micro RNAs is an important task, both for understanding the pathogenesis of immune-dependent diseases and for developing immunotherapy. Due to its versatility, the miR-200 family has an ambiguous effect on the tumor process. In particular, miR-200 increases in the cells of the lung tumor (miR200b) and the prostate (miR141, miR200b), but decreases in the cells of the breast tumor (miR-200bc, miR-429) and liver cancer (miR200a). Thus, miR200 cannot be unambiguously considered as antitumor factors without assessing the type of tumor and the stage of the tumor process.

Primary immunodeficiencies (PIDs) are a heterogeneous group of genetic disorders caused by malformations or dysfunction of the components of the immune system. New data characterize PID as a widespread pathology that seriously affects survival, quality of life and increases the economic burden of the family. According to recent studies, patients with PID are at increased risk for cancer (Mortaz E., et al., 2016).

Common variable immunodeficiency (CVID) is a group of PID syndromes most commonly associated with cancer, including the bladder (Kralickova P., et al., 2019). In 2009, Ortutay C., et al. 26 PID candidate genes were predicted based on annotation data of protein-protein interactions and enriched gene ontology (GO). In 2009, Keerthikumar S., et al. 1442 predicted the PID candidate gene using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the support vector algorithm. In 2015, Itan Y. and Casanova JL predicted more than 3,000 PID candidate genes that have close relationships with already known PID genes. They were able to make such a discovery by analyzing data on protein-protein interactions (PPI) and their mathematical models, which are called the Human Genome Connectome (HGC). In this study, to efficiently predict the CVID candidate genes, we applied and improved these methods using data on signaling pathways and co-expression, since genes with similar functions involved in the same signaling pathway are usually more co-expressed, and the data of protein-protein interaction and gene expression profiles of a large number of CVID samples were taken into account.

Hennekam syndrome is known as an autosomal recessive syndrome, characterized by asymmetric expansion of the lymphatic vessels, both peripheral and internal, facial dysmorphism and mild or moderate cognitive impairment resulting from malformation of

the lymphatic system. By 2015, about 50 cases have been reported with this rare syndrome in the United States. Whereas, in Russia, there are currently no statistics on Hennekam syndrome. Nevertheless, this syndrome is a unique natural model for studying the mechanisms of formation and functioning of lymphatic vessels, lymph circulation and its relationship with immunity. It is known that mutations of the CCBE1 gene (tumor suppressor, inducer of the formation of epithelial endothelium of the lymphatic vessels), the FAT4 gene (synthesis of protocadherin, participates in the generation of the endothelium of the lymphatic vessels) and the ADAMTS3 gene, but in half of the reported cases, mutations could not be identified as cause of the Hennekam syndrome. This justifies the need to identify disease genes using NGS, bioinformatics, and a database of human genetic variations. Full use of these resources can significantly reduce the difficulties of diagnosing such diseases and increase the effectiveness of therapy.

**Purpose of the study –** to identify significant pathophysiological mechanisms of the formation of certain types of immune-dependent pathology based on bioinformatic analysis.

**Research objectives:**

1. To analyze the pathophysiological mechanisms of heterogeneity of muscle-invasive bladder cancer.
2. To identify prognostic biomarkers of the two main subtypes of muscle-invasive bladder cancer.
3. To evaluate the expression of immune checkpoint molecules (PD-1, PD-L1, CTLA-4, HAVCR-2, and LAG-3) in patients with different subtypes of muscle-invasive bladder cancer.
4. To analyze the gene functional mechanisms underlying the multigene nature of the common variable immune deficiency.
5. To identify candidate genes for common variable immune deficiency and highlight a list of candidate genes.
6. To identify mutations in genes that lead to the development of Hennekam syndrome.

**Scientific novelty.** As a result of scientific research for the first time in the Russian Federation, the following has been completed:

For the first time, a systematic analysis of signaling pathways that are significant for immunity, leading to heterogeneity of the basal and luminal subtypes of muscle-invasive bladder cancer, was performed.

For the first time in muscle-invasive bladder cancer, with its various subtypes, differences were revealed in the results of molecular interactions of miRNA, lncRNA, and mRNA.

For the first time, a new perspective on understanding the multigene nature of the group of primary immunodeficiency syndromes, common variable immune deficiency, is presented, due to the results of gene relationship analysis.

A new method for predicting candidate genes of the group of primary immunodeficiency syndromes “common variable immune deficiency” has been developed by incorporating co-expression factors, protein-protein interaction, and signaling pathways into the clinical bioinformatic analysis.

A list of probable pathogenic mutations in genes is presented, the combination of which can lead to the development of primary immunodeficiency “Hennekam syndrome”.

**Theoretical and practical significance of the research.** Micro RNA molecules miR-141-5p, miR-141-3p, miR-200C-3p, long non-coding RNAs (lncRNAs) AC010326.3 and AC073335.2 (controlling GATA3), MIR100HG (inhibiting CLIC4 and PALLD), mRNA CLIC4 (intracellular chloride channel 4), GATA3 (transcription factor), PALLD (Palladin, a protein associated with the cytoskeleton) are important in tumorigenesis and tumor progression. These molecules can serve as new prognostic parameters, as well as markers for the classification of subtypes of muscle-invasive bladder cancer. The fact of increased expression of genes related to the process of epithelial-mesenchymal transition, metastasis and immune system functions, has been revealed in the basal subtype of bladder cancer. The luminal subtype is characterized by a relatively increased expression of genes responsible for metabolic processes. The basal cancer subtype revealed higher values of the expression of immunosuppressive receptors (PD-1, PD-L1, CTLA-4, HAVCR-2 and LAG-3) than the luminal subtype. These data provide a better understanding of the immunological heterogeneity of bladder cancer.

The conclusions that the CVID genes are more similar to each other in their functional relationships and closely interact with each other compared to other PID genes, which helps to better understand the multi-genic nature of CVID. The proposed 172 genes for the CVID candidate must be considered for the early diagnosis of CVID whenever creating targeted therapy.

Homozygous mutation (g.125452634G> A) in the FAT4 gene (cadherin family member 4 — tumor suppressor homolog 4, unknown function, ubiquitous), heterozygous mutation (g.36575963G> A) in RAG1 gene (recombinase 1 activation gene, functions when the formation of antigen-recognizing molecules of T and B lymphocytes), a heterozygous mutation (g.9715914T> A) in the PIK3CD gene (a leukocyte-specific subunit of the enzyme phosphatidylinositol-3-kinase) and a heterozygous mutation (g.36471505C> T) in the CSF3R receptor colony stimulating factor 3, necessary for the maturation of leukocytes), together, have pathogenetic importance and are taking into

account for the understanding of the Hennekam syndrome pathogenesis, moreover, they will expand diagnostic capabilities.

**Protection provisions:**

1. Data on the expression of miRNA miR-141-5p, miR-141-3p, miR-200C-3p, lncRNA AC010326.3, AC073335.2, MIR100HG, and mRNA CLIC4, GATA3, PALLD, should be used as diagnostic markers of immune-dependent pathology of muscle-invasive bladder cancer, and for the prognosis and differential diagnosis of the basal and luminal subtype.

2. The difference in the expression of immunosuppressive molecules PD-1, PD-L1, CTLA-4, HAVCR-2 and LAG-3 allows us to distinguish between basal and luminal subtypes of muscle-invasive bladder cancer, in which, basal tumors show higher levels of expression of immunosuppressive molecules than luminal tumors subtype.

3. In this study, an extended spectrum of 172 candidate genes was identified for the molecular genetic etiology of the group of pathological syndromes of primary immunodeficiencies “common variable immune deficiency”

**The reliability of the research results** is ensured by the validity of the initial theoretical positions, a sufficient sample size , the use of modern laboratory methods, reproducibility of the results, the use of computer programs. For statistical analysis of the data were obtained and confirmed by the expert commission of the Institute of Natural Sciences and Mathematics of the Federal State Autonomous Educational Institution of Higher Education “Ural Federal University named after the first President of Russia B. N. Yeltsin ”(an act of verification of primary documentation was drawn up, an order of the director of the institute, dated December 3, 2019).

**The author’s personal contribution** consists in direct participation at all stages of the dissertation research, including creation of a basic idea, planning of scientific work, formulation of a working hypothesis, work tasks, and the definition of a research methodology. Interpretation and analysis of the results in dissertation were carried out with collaboration of supervisors.

**Putting research results into practice.** The results are implemented in the practice of the Regional Children's Clinical Hospital of the Sverdlovsk Region, were used in the work of the Regional Commission of the Ministry of Health of the Sverdlovsk Region on infant mortality. The list of CVID candidate genes has been introduced for use in the laboratory of inflammation immunology at the Institute of Immunology and Physiology of the Ural Branch of the Russian Academy of Sciences, in teaching and advisory medical practice.

**Approbation of work.** The main provisions of the thesis were presented at the III All-Russian Congress “Autoimmune and Immunodeficiency Diseases” (Moscow, 16-17,

November, 2018); X NODGO Congress “Actual Problems and Prospects of the Pediatric Hematology-Oncology in the Russian Federation” (Sochi, 27-29, April, 2019), IV Joint immunological forum (Novosibirsk, 24-29, June, 2019), 17th International Congress of Immunology (Beijing, 19-23, October, 2019), II St. Petersburg Lymphological Forum “Lymphology XXI century: new approaches and current research” (St. Petersburg, 10-11, October, 2019), the International Euro-Asian Congress on Bioethics, Molecular and Personalized Medicine “Biomed-inn-2019” (Perm, 5-8 November 2019).

**Publications** As a result of the dissertation, 7 works were published, of which 5 were published in journals indexed in the international electronic databases, Web of Science and Scopus, and 2 in journals peer-reviewed by the Higher Attestation Commission.

**Volume and structure of the dissertation.** The English version of dissertation is presented on 122 pages of typewritten text, the Russian version is of 144 pages, consists of an introduction, literature review, chapter “Materials and Methods”, 3 chapters with the results of our own research, conclusions, findings, practical recommendations, and a list of abbreviations and a list of references (204 sources, including 5 russian and 199 foreign). The work is illustrated by 12 tables, 30 figures, and 3 formulas.

## MAIN CONTENT OF WORK

**Materials and methods.** MRNA and long non-coding RNAs (lncRNAs) sequence data and clinical data for 403 patients with muscle-invasive bladder cancer (MIBC) and 19 relatively healthy individuals, were obtained from the publicly available Cancer Genome Atlas portal ( https://cancergenome.nih.gov), micro RNA (miRNA) sequencing data were obtained from the publicly available database Broad GDAC Firehose (https://gdac.broadinstitute.org). The relative levels of tumor-infiltrating immune cells for each sample were evaluated using CIBERSORT (https://cibersort.stanford.edu).

First, a consensus cluster analysis was carried out by using the ConsensusClusterPlus based on separated mRNA, miRNA, and lncRNA data. After that cluster analysis was performed to determine subtypes of basal and luminal bladder cancer, taking into account all three data sets. Survival analysis was performed based on the Kaplan-Meyer (KM) curve using the Survival R package. A logarithmic rank test was used to compare survival times between two subtypes. To verify the independence of samples of information about the subtypes and clinical parameters of patients with MIBC, the criterion $\chi 2$, $p < 0.001$ was used. The Wilcoxon T-test was used to test differences between the two groups. Functional enrichment analysis was performed using the GSEA software (version 3.0). Two machine learning models, Random Forest and XGBoost, were used to screen for representative RNA between the two subtypes of MIBC (*Mean decrease accuracy* > 0 and *Gain* > 0). The Ballgown R package was used to determine

differentially expressed RNA between MIBC patients and healthy control samples (FDR $p<0.05$ and $|log_2 fold\ change|> 0.57$). The Pearson correlation coefficient ($|r|> = 0.4$ and $p<0.05$) and the miRWalk2.0 database (http://mirwalk.umm.uni-heidelberg.de/) were used to study the interaction of mRNA, miRNA and lncRNA.

Gene expression data (GSE72625 and GSE51406) obtained using DNA microarrays of 111 patients with common variable immunodeficiency (CVID) and 56 relatively healthy control samples were obtained from the Gene Expression Omnibus (GEO) database (https: //www.ncbi.nlm .nih.gov). The GSE72625 dataset contained data of 20 CVID patients and 17 relatively healthy adults; the GSE51406 data set contained information of 91 patients with CVID and 39 relatively healthy adults. A list of 351 known at the time of the study PID genes, including 39 CVID genes, was obtained from the website of the European Society of Immunodeficiency (https://esid.org). The protein-protein interactions (PPIs) for the PID and CVID genes were obtained from the STRING database (https://string-db.org/). The network density provided by Horvath S., et al., 2008 was used to calculate the cohesion of the PPI networks of a group of 39 CVID genes and ten random PID groups (each group consisted of 39 randomly selected PID genes). The calculation of the "biological distance" in a group of 39 CVID genes and two random PID groups (each consists of 39 PID genes) was performed using the Human gene connectome (HGC) package provided by Itan et al., 2013. CVID candidate genes were predicted using the Pearson correlation analysis ($| r |> 0.9$ and $p <0.05$), the data of 217 160 protein-protein interaction (PPI), provided by Cheng F., and KEGG enrichment analysis (packet R «clusterProfiler », $p < 0.05$).

The results of a clinical examination of a patient with Hennekam's syndrome were obtained from the Regional Children's Clinical Hospital, with the written consent of the parents to conduct the study. The patient's genomic DNA was extracted from peripheral blood mononuclear cell lymphocytes. Full genome sequencing was performed by Genomed company (Moscow). DNA analysis was carried out by a next generation sequencing technology using pair-terminal reading. The analysis covers 98.5% of the total genome. The average coverage was at least 30x. Sequence results were aligned in accordance with the standard hg38 human genome sequence using the software « the Burrows-Wheeler Aligner (BWA)» . The SAMtools program was then used to reorder, sort, and convert SAM files. Single nucleotide variants (SNVs) and insertion/deletion (insertion) variant calls were identified with GATK4 version 4.1.2 (GATK4, http://www.broadinstitute.org/gatk/). To reduce substandard results, arising from the mapping error, we use a variant in the exon region of the depth reading>10 × and minimum display quality score of 30, using the software «VCFtools ». We then discarded all synonymous SNV, insertions and deletions without shifting the reading

frame (InDels) and variants, exon feature which is labeled as «NA» or «Unknown». Candidates SNVs and insertion of the previous step were further filtered by the connection with the phenotype, the minor allele frequency (MAF), pathogenetic, using software ANNOVAR and online server CADD (https://cadd.gs.washington.edu). The analysis of sequencing results and the identification of mutations and single nucleotide polymorphisms were carried out on a supercomputer provided by the Center for Collective Use of IMM UB RAS "Supercomputer Center IMM UrO RAS" (Yekaterinburg).

## RESULTS AND ITS DISCUSSION

**Study of the mechanism of tumor heterogeneity in muscle-invasive bladder cancer.** By analyzing complex RNA sequencing and clinical data from 403 patients with muscle-invasive bladder cancer (MIBC), levels of B-lymphocytes, dendritic cells, various macrophages, neutrophils, NK-cells, CD4 + and CD8 + T-lymphocytes were evaluated. The log-rank test and Wilcoxon's T-test showed that, among all immunological parameters, the levels of M2 macrophages strongly correlated with patient survival, grade, and pathological stage ($p<0.01$) (Figure 1). This logically fits into the pathogenesis of the development of the oncological process, in which the severity of the process correlates with the greater activity of tumor macrophages. The results of our studies allow us to recommend M2 macrophages as markers for predicting the MIBC prognosis.

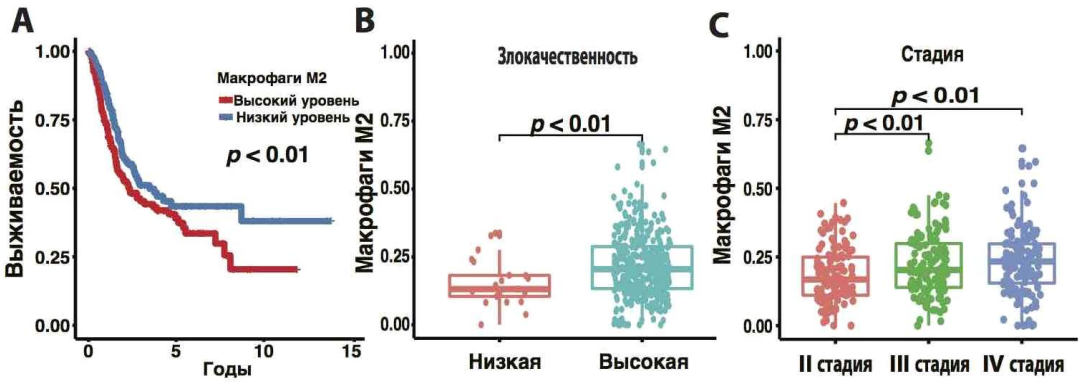


**Figure 1 - Graphs of the Kaplan-Meyer curve for M2 macrophages (A), and the level of M2 macrophages in patients with MIBC depending on grade and stage (B, C)**

Using the RNA sequencing data available in the TCGA database, we classified 403 MIBC samples into two molecular subtypes, namely, subtype-1 and subtype-2, which respectively resemble previously identified luminal and basal subtypes (Figure 2). Subtype-1 was characterized by high expression of luminal markers, such as CYP2J2, ERBB2, and KRT18, while subtype-2 was characterized by high expression of basal

markers, such as CD44, CDH3, and KRT1. Survival analysis revealed an overall 5-year survival rate for the MIBC luminal subtype of 55%, and 30 % for the basal subtype, which indicates a significant difference in the clinical prognosis (log-rank test, $p$ <0.05). The Pearson test ($\chi 2$) was used to compare clinical performance between two subtypes. Histology, cancer stage, grade, and survival status varied significantly between the two subtypes ( $p$ <0.01).

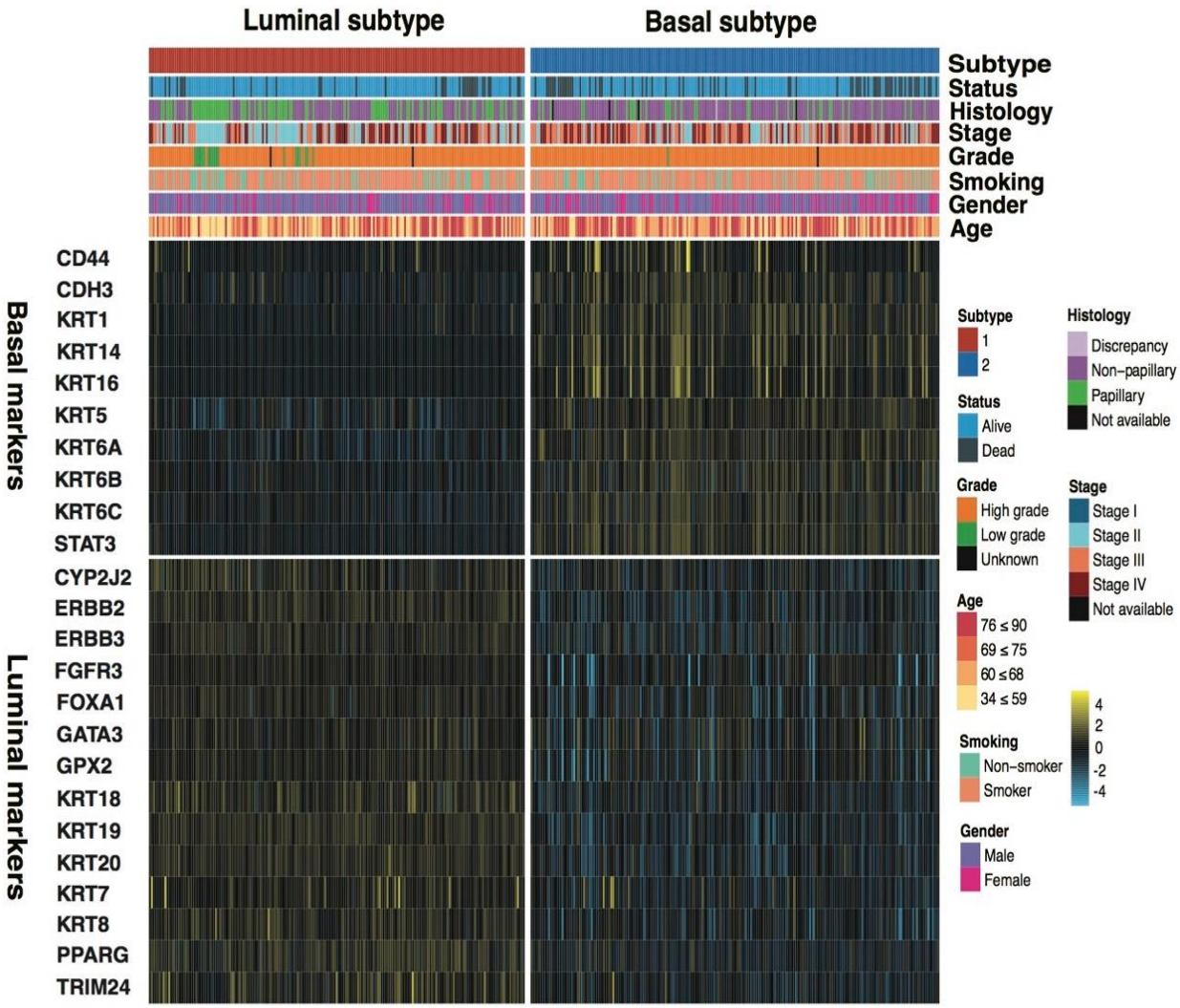


**Figure 2 - Separation of muscle-invasive bladder cancer into subtype-1 and subtype-2**

*Note:* the heat map shows the expression profiles of basal (top) and luminal (bottom) biomarkers in subtype-1 (left) and subtype-2 (right). Annotations and corresponding colors indicate some clinical features. Yellow and turquoise colors correspond to high and low relative levels of expression.

The functional gene set enrichment analysis (GSEA) revealed highly expressed signaling pathways involved in the metabolism in the luminal subtype. At the same time, highly expressed pathways in the basal subtype are mainly associated with immune system processes (organization of the extracellular structure, allograft rejection,

mTORC1 signaling, TNF- α signaling via NF-kB), metastasis (focal adhesion, interaction cytokine-cytokine receptors), and epithelial-mesenchymal transition.

More importantly, miRNA-dependent cross-interaction of mRNA and lncRNA was detected. The cross-dependency network consists of GATA3, CLIC4, PALLD, miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 and MIR100HG. It was revealed that gene expression in this interdependence significantly differed between the basal and luminal subtypes and had a close relationship with the survival prognosis ($p$<0.05) (Figure 3). MiR-200c-3p, miR-141-3p, miR-141-5p can be key modulators of the immune response and tumor development in MIBC. These results provide a deeper understanding of the pathophysiological mechanisms of muscle-invasive bladder cancer.

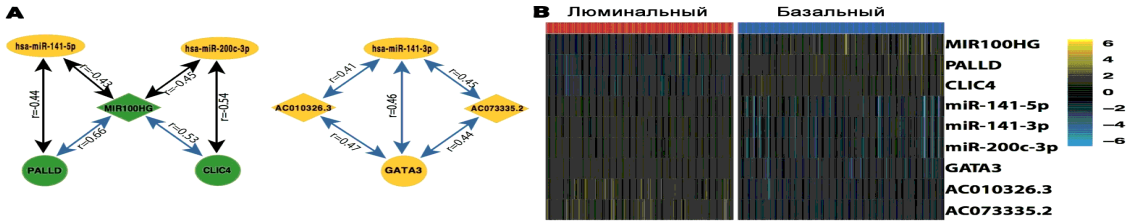


**Figure 3 - Characteristic of miRNA-dependent mRNA-lncRNA cross-relationship in MIBC subtypes**

*Note:* (A) miRNA-mediated mRNA-lncRNA cross-relationship; green indicates molecules with reduced expression in the tumor samples compared to normal samples; yellow represents an increased expression. Blue lines represent positive correlations, and black lines negative. (B) The heat map displays the expression level of nine genes in the basal and luminal subtypes, yellow and turquoise colors correspond to high and low relative expression levels, respectively.

**Assessment of the ability of different subtypes of MIBC to respond to immunotherapy.** The relationship between subtypes and immunotherapy markers was investigated. It was found that the expression values of the molecules of the control points of the immune response (PD-L1, PD-1, CTLA-4, HAVCR-2 and LAG-3) make it possible to distinguish between basal and luminal subtypes of MIBC. In this case, basal tumors showed higher levels of expression of control point molecules of the immune response than the luminal subtype (Wilcoxon T-test, p <0.01), which must be taken into account when prescribing tumor immunotherapy (Figure 4).

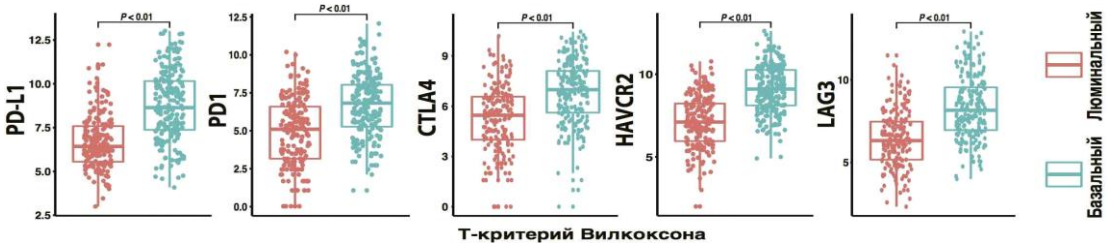


**Figure 4 - Expression levels of immune checkpoint molecules in the basal and luminal subtypes**

*Note:* the yellow and cyan colors correspond to high and low relative expression values.

**Identification of candidate genes for common variable immunodeficiency.** Common variable immunodeficiency has deep phenotypic and genotypic heterogeneity and is caused by both monogenic and complex multigenic causes. By analyzing the network density and the biological distance density, we found that the CVID genes are functionally more similar to each other and closely interacted with each other compared to other PID genes. This is consistent with the notion that CVID is a more multi-gene cohort of disease than most PID. On the other hand, this once again confirms the idea that CVID is rather a diagnosis of exclusion and combines a huge list of various pathologies with more or less common phenotypic manifestations, in which the genetic prerequisites for the development of the pathology still remain undetected and the corresponding targeted therapeutic effects are not determined.

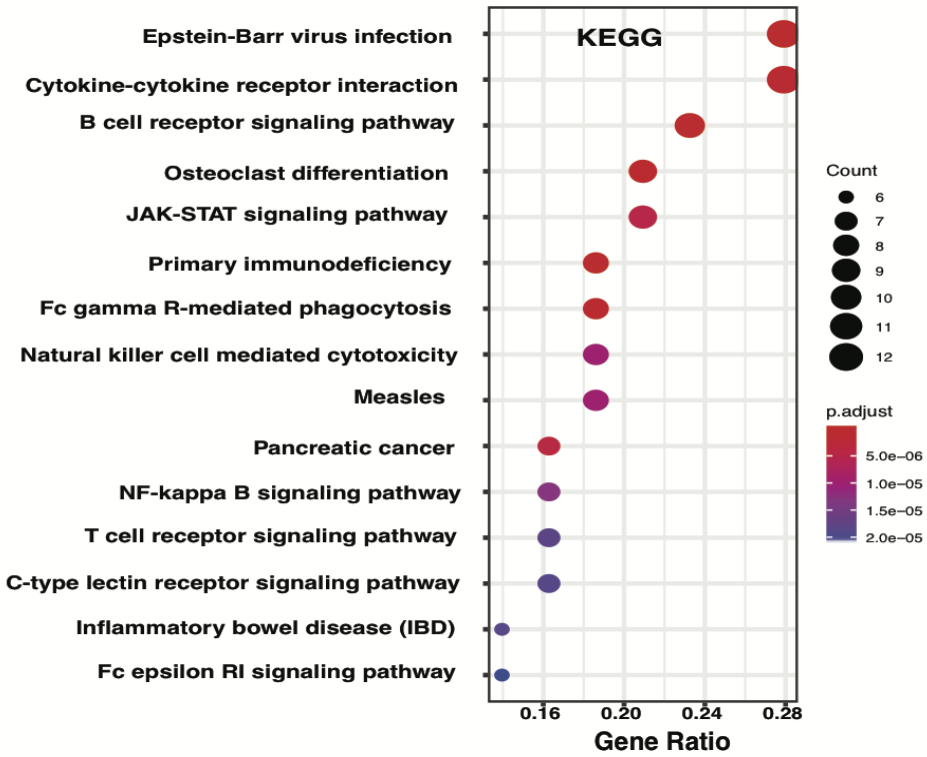


**Figure 5 - Functional enrichment of CVID genes with the biological value from the KEGG database**

*Note:* the size of the bubble corresponds to the number of genes, the color the adjusted p-value.

Based on Pearson correlation analysis and data from protein-protein interactions provided by Cheng F., et al., 2018, 2,751 CVID-specific interactions were obtained, including 1,716 candidate genes. Using the clusterProfiler R package, we performed the so-called KEGG functional enrichment analysis of the known CVID genes (in fact, the binding of genes to their biological functions) and we found that there are a total of 15 signaling pathways described in the KEGG database (for example, Epstein-Barr virus infection, cytokine-cytokine receptor interaction and B-cell receptor signaling pathways ) were statistically significant ( $p$ <0.05, Figure 5). To identify candidate genes that are

functionally similar to the known CVID genes, a screening analysis of certain genes that have any functional relation to the problem under study and associated with at least one of the 15 signaling pathways according to the KEGG database was performed.

The above analysis revealed 172 new CVID candidate genes and 414 CVID-specific protein-protein interactions between the CVID gene products. The result of constructing a network of protein-protein interactions for the known CVID genes and candidate CVID genes is shown in Figure 6.

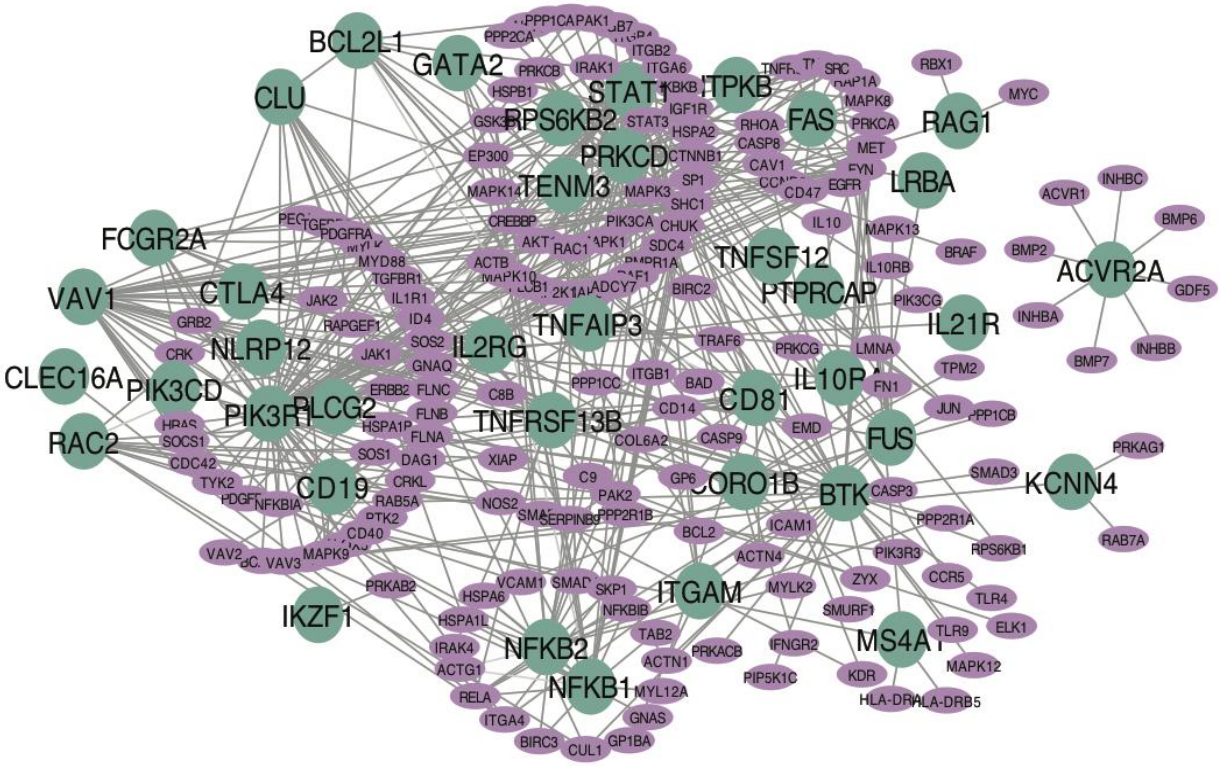


**Figure 6 - The network of protein-protein interactions (PPI) of the candidate CVID genes (light purple) and the known CVID genes (dark green)**

A list of 172 identified CVID candidate genes is presented in practical guidelines. Having identified new candidate genes, we conducted a study of the quality of the forecast, conducting a review of recent studies. In particular, eight CVID candidate genes (AKT1, AKT3, RELA, SOCS1, STAT3, XIAP, CD40 and CASP8), which were not included in our initial list, but were predicted by our method, and also were identified by other authors in experimental studies. This demonstrates the significance of the identified CVID candidate genes.

**Identification of gene mutations (FAT4, RAG1, PIK3CD and CSF3R), leading to the development of the disease in a patient with phenotypic manifestations of Hennekam syndrome.** In the section, our research object was a 5-year-old patient with phenotypic manifestations of Hennekam syndrome , but with some features that made it possible to doubt the correctness of the diagnosis. After whole genome sequencing, no pathogenic mutations were detected in the commercial

laboratory. It was decided to design new methods to predict genes associated with the patient's phenotype.

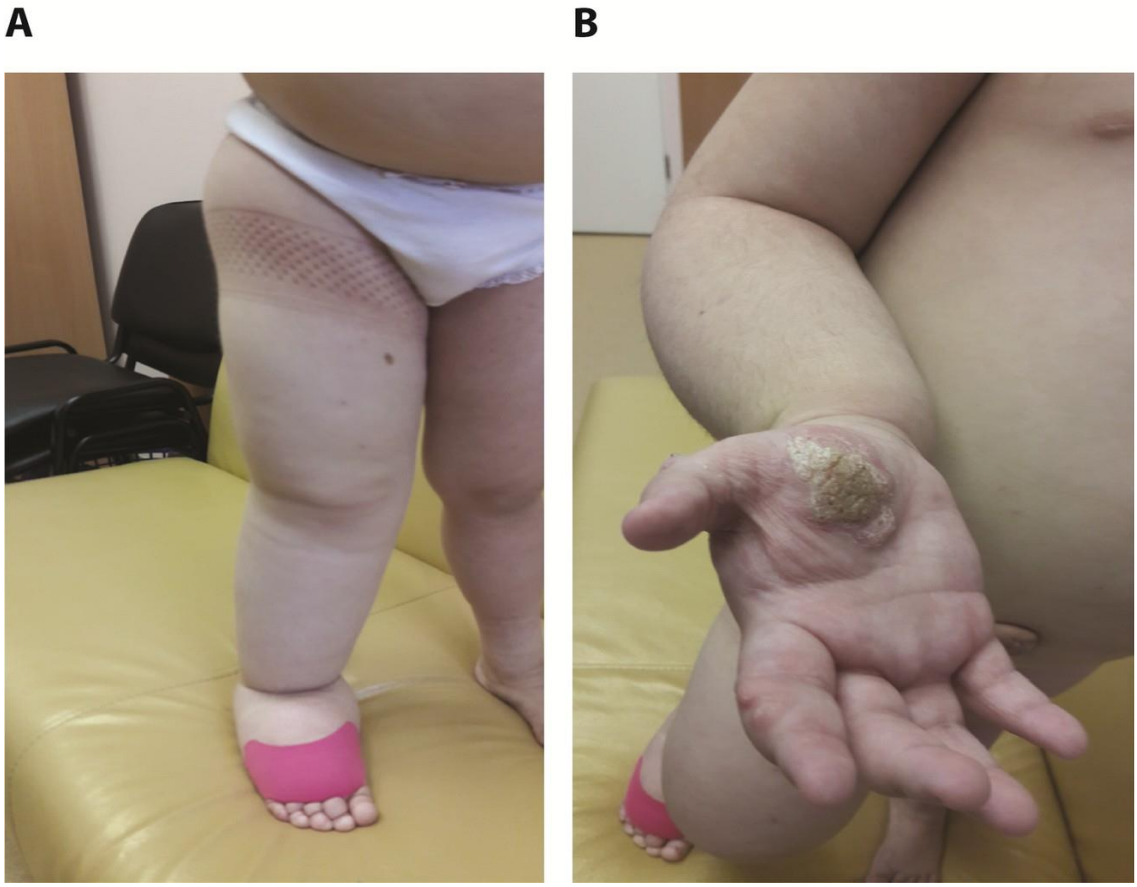


**Figure 7 - Photograph of a patient with Hennekam syndrome**

*Note:* (A) lymphedema of the right half of the body (in particular, the lower limb) and (B) warty formations on the right palm, ascites, lymphedema of the right upper limb, hirsutism (on the forearms).

The patient diagnosed with Hennekam syndrome was a five-year-old girl. Her clinical manifestations included the following symptoms of the disease:

- asymmetric lymphedema with a predominant lesion on the right side first detected in the perinatal period at 34 weeks of gestation (Figure 7A);
- warty lesions on the skin of the palm, forefinger, and thumb of the right hand (Figure 7B);
- lymphadenopathy - lymphangiectasia, which was complicated by the addition of a skin pathological process, presumably papillomavirus etiology;
- chylous ascites detected from the prenatal period, which remains to date (Figure 7B);
- cystic defects of the mandible involving bone tissue;
- right-sided hydrothorax;
- internal biventricular open hydrocephalus;

- lymphadenopathy of all groups of cervical nodes (larger on the right), supraclavicular, subclavian, parasternal, and also lymph nodes of the back at the thoracic level, a single lymph node of the right lumbar region up to 5 mm in diameter, multiple lymph nodes of the left ileal and inguinal groups enlarged to 17 mm, in present - small lymph nodes up to 0.3 cm, mobile, palpation determined mainly in the axillary groups;
- hirsutism expressed on the upper limbs (Figure 7B);

A history of increased urea, kidney hypoplasia. In addition, signs of immuno-dependent pathology in several generations along both hereditary lines were noted in heredity.

Basic steps of filtering potentially pathogenic mutations are represented schematically in figure 8. Initially, we excluded synonymous SNV, non-frames InDels and embodiments are marked as “NA” or “Unknown”. Then, the identified variants were filtered out through by the responding known 351 PID gene and 2 Hennekam syndrome genes (FAT4 and CSF3R). After eliminating the common variants of MAF> 0.01 by ExAC, 1000g and gnomAD, 6 rare mutation variants remained. The CADD, FATHMM, and PROVEAN models were used to select pathogenic mutations, and finally, 4 mutations were identified that are “probably pathogenic,” that might lead to the development of the disease in this patient with high probability.


**Figure 8 - Diagram illustrating all four main filtering steps to identify potentially pathogenic mutations**
*Note:* the number of filtered genes is shown in parentheses.

In particular, a homozygous mutation (g.125452634G> A) of the FAT4 gene was revealed, which has already been reported that biallelic and monoallelic mutations are associated with Hennekam syndrome ( Alders M., et al ., 2014). In addition, the following mutations were revealed in the patient: heterozygous mutation (g.36575963G> A) in RAG1, heterozygous mutation (g.9715914T> A) in PIK3CD, heterozygous mutation (g.36471505C> T) in CSF3R.

Regarding the homozygous mutation g.125452634G> A in the FAT4 gene, that contains conflicting information. According to the dbSNP database of the National Center for Biotechnological Information (NCBI) of the United States, the mutation is rare and benign, not associated with any phenotype of the disease (data dated January 8, 2016). Whereas, according to the latest version of the CADD database using the human reference genome Ch38 (GRCh38-v1.5), the mutation is potentially pathogenic, therefore, its etiological significance in the case we estimate could not be completely denied.

The Sanger sequencing was carried out for checking the pathogenicity of this mutation for this patient. When comparing the DNA of parents, sibs, and probands, it was found that all relatives are carriers of the same homozygous mutation (g.125452634G> A) in FAT4, as was in the girl. While the phenotype of Hennekam syndrome was observed only in the patient. This means that G> A substitution for Chr4: 125452634 in FAT4 may not be pathogenic. However, identical mutations in the same gene can lead to different symptoms of a hereditary human disease due to changes in penetrance and variable expressivity, which can be induced by factors such as gene interactions, environmental factors, allelic variations and stochasticity (Cooper DN, et al., 2013). Therefore, the question of whether a homozygous mutation (g.125452634G> A) in the FAT4 gene can lead to different symptoms in different people deserves further analysis.

Mutations in RAG1, PIK3CD and CSF3R are also potentially important for assessing significance in the development of pathology in this patient. The recombination activation gene 1 (RAG1) plays a significant role in the formation of a protein complex for the V(D)J recombination process in lymphocytes. Defects of RAG1 or RAG2 lead to various primary immunodeficiencies, including T and B-SCID, increased susceptibility to cytomegalovirus infection, autoimmune diseases, Omenn syndrome, etc. Mutations in the CSF3R gene are important for the intracytoplasmic domain of G-CSFR, which were discovered in patients with congenital neutropenia. It is also known that these mutations can manifest as somatic mutations associated with the progression of leukemia in congenital neutropenia. Mutations with increased function of the PIK3CD gene, which encodes the catalytic subunits P110δ and P85 of phosphoinositide-3-kinase-delta, lead to the development of the so-called Pi3 kinase activation syndrome (previously considered one of the CVID), in the phenotype of which lymphoproliferation is one of the expressed symptoms of the disease, up to B-lymphomas. In addition, this syndrome is characterized by hypogammaglobulinemia, recurrent viral infections, bronchiectasis, and herpesvirus infections. More importantly, the patient is treated with everolimus, a mTOR signaling pathway inhibitor whose role in lymphangiectasia, which has been reported by many researchers (Pollack S., et al., 2015 and Altın Z., et al., 2018).

## CONCLUSION

From the “Human Genome Project” to the “Precision Medicine Initiative”, the scientific and medical community is moving step by step in accordance with the established goals to “individualized medicine”. With the development of high-performance next-generation sequencing technology, the generation and accumulation of huge volumes of genomic data began, which necessitates their interpretation and the solution of more complex technical problems. A sensible interpretation of these large-scale and complex multidimensional data is becoming a key step in modern studies of

various human pathologies , including immuno-dependent diseases, including oncological pathology, auto-inflammatory, autoimmune diseases and primary immunodeficiencies.

The conducted study allows us to look at the processes of system biology from a pathophysiological point of view, and identify general patterns. Both primary immunodeficiencies and oncological pathology can be presented in a single scheme at the level of common typical pathological processes, for example, tissue stress and its participation in the general signaling and metabolic pathways.

Due to studies of the results of RNA sequencing, the concept was confirmed that muscle invasive cancer of the bladder is heterogeneous by its genetic nature and, based on the results of our studies, can be classified into two main subtypes, basal and luminal subtypes. Methods for predicting the effectiveness of immunotherapy for this type of cancer were identified and biomarkers for detecting subtypes were proposed.

In addition, methods have been developed that combine data on co-expression and protein-protein interactions, as well as data characterizing the functions of genes, allowing to identify among them causally significant for the development of PID. In our study, 172 new CVID candidate genes were identified, which are interdependent with the known CVID genes, have strong correlation between them and function at the level of the same signaling pathways. Several of our analysis tests, as well as the confirmation in 8 experimental studies of the identified CVID genes predicted in our study, confirm the significance of the predicted candidate genes in the development of pathology. Our study contributed to a better understanding of the molecular genetic mechanisms of CVID and allowed us to expand the list of candidate genes for this PID group.

Many of the CVID genes predicted and proven in experimental studies have broad functions and the complete shutdown of some of them can lead to death of a person in the embryonic period.

However, it is known that CVID quite often makes its debut in people over the age of 20-40. The multigene nature of CVID is explained, in particular, by the fact that the debut of the disease requires the combined effect of many factors leading to the fact that functionally weak proteins, under certain conditions, induce clinical manifestations of the disease.

Using bioinformatic analysis methods tested in theoretical studies, we analyzed the case of pathology in a patient with lymphangiectasia-lymphedema, warty formations, hirsutism, hydrothorax, and lymphatic ascites. For a better understanding of pathogenesis and possible target exposure points, whole-genome sequencing was performed. As a result, four mutations were identified in genes known to be pathogenic in PID, including Hennekam syndrome. Determining their role in the future will allow us to approach a

more accurate understanding of pathogenesis and justify treatment for the patient's pathogenetic, as well as determine the origin of these mutations and investigate the molecular mechanisms of the disease formation.

**Prospects for further development of the topic.** In our study, a statistically significant increased expression of signaling molecules of cytokine-cytokine receptor interactions was revealed, both in the basal type of MIBC and CVID. Continuing research in this area will help to reveal the general mechanisms of development of immunopathology of primary immunodeficiencies in cancers,

The approach used to identify the mechanisms of the pathogenesis of CVID can be borrowed to study the formation of the phenotype of other syndromes.

Identification of specific genes for CVID syndromes and determination of the molecular basis of immunopathology will allow us to study pathogenesis, differentiate pathological in the vast heterogeneous group of CVID.

The process of verification of primary immunodeficiency genes can be improved by the development of software for predicting candidate genes of various immuno-pathologies, which provides the possibility of its effective application in clinical trials.

The revealed patterns of molecular genetic foundations of immunopathology during the formation of various subtypes of muscle-invasive bladder cancer can be further continued by experimental studies. The interactions of miR-200c and miR-141 mi-RNAs detected by computational analysis should also be confirmed by biological experimental studies in the future.

The integrated approach used to identify the patterns of formation of immunopathology can be used to study any other disorders of the immune system.

Regarding the identification of the development of syndromes with malformation of the lymphatic vessels and hereditary lymphedema, it is necessary to take into account other forms of transmission of hereditary information, including autosomal dominant, co-dominant and co-expressive, as well as DNA variations outside exons. This determines the need to continue the search for the pathogenetic basis for the formation of immunopathology, taking into account new data on the pathogenesis mechanisms and the involved signaling pathways, as well as conducting full-scale studies of the human genome at the higher level.

## FINDINGS

1. Muscle-invasive bladder cancer, being heterogeneous by genetic nature, has two main subtypes, namely basal and luminal subtypes, which differ in gene expression profile.

2. Expression levels of GATA3, PALLD, CLIC4, miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 and MIR100HG differ significantly in the two subtypes of muscle-invasive bladder cancer and are closely related to its prognosis.

3. Expressions of immune response checkpoint molecules, PD-1, PD-L1, CTLA-4, HAVCR-2, and LAG-3, distinguish between basal and luminal subtypes of muscle-invasive bladder cancer, in which basal tumors exhibit higher levels of immune response checkpoint molecule expression than luminal subtype tumors.

4. Common variable immunodeficiency genes are functionally similar to each other and interact more closely at the level of signaling pathways, compared to other primary immunodeficiency genes.

5. Combining data on co-expression, protein-protein interactions, characterizing gene functions, it was possible to identify 172 candidate genes for common variable immunodeficiency, expanding the current list of known genes.

6. Probable pathogenic mutations in the FAT4, RAG1, PIK3CD, and CSF3R genes have been identified, which may have clinical significance and lead to the phenotype of Hennekam syndrome.

## PRACTICAL ADVICE

1. Tumor macrophages belonging to the M2 family significantly correlated with patient survival, grade, the stage of the muscle-invasive bladder cancer, and can be markers of the tumor process in muscle-invasive bladder cancer.

2. The cross-interference of various coding (GATA3, PALLD, CLIC4) and non-coding RNAs (miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 and MIR100HG) could play a significant role in oncogenesis and tumor progression, as well as identified biomarkers are important in determining subtypes of muscle-invasive bladder cancer and predicting the development of pathology.

3. Expression of the immunosuppressive molecules PD-1, PD-L1, CTLA-4, HAVCR-2, and LAG-3 allows us to distinguish between basal and luminal subtypes of muscle-invasive bladder cancer, in which basal tumors exhibit higher levels of expression of immunosuppressive molecules than luminal subtype tumors.

4. The multigene nature of common variable immune deficiency is based on the functional community of genes and their closer interaction between themselves at the level of signaling pathways, in comparison with other genes of primary immunodeficiencies.

5. Prediction of candidate genes of the group of primary immunodeficiencies “common variable immune deficiency”, provided by a comprehensive analysis of the development of immunopathology, data on co-expression and gene interactions,

information on their biological role, which made it possible to identify 172 candidate genes.

6. Probably pathogenic mutations in the FAT4, RAG1, PIK3CD, and CSF3R genes have been identified, which may have clinical significance and lead to the phenotype of primary immunodeficiency of Hennekam syndrome.

## PRACTICAL RECOMMENDATIONS

1. Before starting immunotherapy for muscle-invasive bladder cancer, the subtypes (basal or luminal) of cancer should be determined by detecting biomarkers, which will reduce the cost of treatment and prevent the development of side effects.

2. To determine the subtype of muscle-invasive bladder cancer, a study of the following biomarkers of the disease is necessary: GATA3, PALLD, CLIC4, miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 and MIR100HG.

3. After determining the clinical phenotype of primary immunodeficiency in studies on the prognosis of candidate genes for common variable immunodeficiency, it is necessary to include co-expression factors, protein-protein interactions, and signaling pathways analysis.

4. For differential diagnosis of CVID, in addition to the genes listed on the ESID site, it is necessary to include a list of these 172 genes identified in our study as candidate genes: LMNA, BAD, PRKAB2, CRK, CRKL, CASP8, CTNNB1, STAT3, HRAS, PTK2, JAK1, BCL2, PDGFRB, CHUK, JAK2, GRB2, SRC, RELA, NFKBIA, EGFR, CREBBP, EP300, MAPK1, IKBKB, PIK3CA, SHC1, PIK3CB, IL10, ERBB2, PPP1CA, TRAF6, INHBA, GNAQ, VAV2, TGFBR1, TGFBR2, BCAR1, NOS2, FYN, SOCS1, SMAD4, TAB2, TNFRSF1A, TNF, HLA-DRA, FLNA, ITGB1, ACTN1, AKT1, SP1, PIK3CG, BRAF, INHBC, CD47, MET, PIP5K1C, SOS1, GP6, JUN, ICAM1, GP1BA, PRKACA, PRKCA, PRKAG1, TYK2, KDR, ACTB, MAPK3, PLCB1, RAC1, ADCY7, RAF1, SDC4, IGF1R, ITGA6, ITGB4, ITGB7, PAK1, AKT3, CDC42, SMAD1, CCR5, CASP3, ZYX, ITGB2, RAP1A, PDGFRA, MAPK8, VAV3, CASP9, GSK3B, IRAK1, MYD88, TLR9, PRKCB, BMP7, RHOA, NFKBIB, MAPK14, BMP6, INHBB, GDF5, TLR4, CAV1, CD40, ITGA4, C8B, C9, IL10RB, PECAM1, PPP2CA, CUL1, IL1R1, CD14, HSPA1B, BIRC3, HSPA1L, IRAK4, BMP2, SOS2, MYLK, PIK3R3, RAPGEF1, MAPK13, GNAS, HSPA2, MAPK9, ELK1, MYC, PPP1CB, CCND3, MYLK2, HSPA6, BIRC2, DAG1, FLNB, FLNC, ID4, XIAP, ACTN4, SMAD3, MAPK12, PPP2R1A, PRKACB, MAPK10, MAP2K1, MAP2K6, MAP3K7, BMPR1A, HSPB1, PRKCG, ACVR1, PPP1CC, RPS6KB1, HLA-DRB5, SERPINB9, FN1, COL6A2, VCAM1, RAB7A, ACTG1, MYL12A, PPP2R1B, SKP1, ALOX5, RAB5A, PAK2, RBX1, SMURF1, EMD, TPM2, IFNGR.

## LIST OF PUBLICATIONS ON THE TOPIC OF THE DISSERTATION

***Publications in journals peer-reviewed by the Higher Attestation Commission, indexed in the international Scopus and Web of Science databases***

1. A co-expression network for differentially expressed genes in bladder cancer and a risk score model for predicting survival / Z. Chen, *Guojun Liu*, I. Tuzankina, M. Bolkov, Guoqing Liu, W. Tang // Hereditas. – 2019. – Vol. 156, № 1. – P. 24 (IF WoS - 0,896; Scopus CiteScore - 1).
2. Identification of candidate disease genes in patients with common variable immunodeficiency / *Guojun Liu*, M. Bolkov, I. Tuzankina, I. Danilova // Quantitative Biology. – 2019. – Vol. 7, № 3. – P. 190-201 (Scopus CiteScore - 0,84).
3. Identification of miR-200c and miR141-mediated lncRNA-mRNA crosstalks in muscle-invasive bladder cancer subtypes / *Guojun Liu*, Z. Chen, I.G. Danilova, M.A. Bolkov, I.A. Tuzankina, Guoqing Liu // Frontiers in genetics. – 2018. – Vol. 9. – P. 422 (IF WoS - 3,517; Scopus CiteScore - 3,6).
4. A novel mutation in FAT4 caused a mild form of Hennekam syndrome with warty formations / *G. Liu*, M. Bolkov, I. Tuzankina // European journal of immunology. Abstracts of IUIS 2019 Beijing 17th International Congress of Immunology. – 2019. – Vol. 49, № suppl_3. – P. 1228-1229 (IF WoS - 4,695; Scopus CiteScore -3,83).
5. Постмортальный анализ иммунозависимой патологии у детей раннего возраста / *Г. Лю*, М.А. Болков, С.С. Дерябина, Я.М. Крохалева, И.А. Тузанкина // Аутоиммунные и иммунодефицитные заболевания: научно-практ. ревматология: тез. докл. III Всерос. – Москва, 2018. – Т. 56, № 3, прил. 3. - С 5 (Scopus CiteScore - 0,27; РИНЦ - 1,306).
6. Функциональные взаимоотношения между генами общей вариабельной иммунной недостаточности в геноме человека / М.А. Болков, *Г. Лю*, И.А. Тузанкина, И.Г. Данилова // Российский журнал детской гематологии и онкологии. – 2019. – Т. 6, № 2. – С. 47 (РИНЦ - 0,81).
7. Молекулярно-генетическая природа синдрома Хеннекама / Лю Г., Тузанкина И.А., Болков М.А., Долгих М.А. // Трансляционная медицина: II Санкт-Петербургский лимфологический форум: Лимфология XXI века: новые подходы и актуальные исследования. – Санкт-Петербург, 2019. – прил. 2, окт. – С. 51 (РИНЦ 0,283).

## ABBREVIATIONS

| | |
|---|---|
| BWA | Burrows-Wheeler Aligner |
| CADD | Combined Annotation Dependent Depletion |
| CVID | common variable immune deficiency |
| ExAC | The Exome Aggregation Consortium |
| FATHMM | Functional Analysis through Hidden Markov Models |
| GATK4 | The Genome Analysis Toolkit version 4 |
| GEO | Gene Expression Omnibus |
| gnomAD | The Genome Aggregation Database |
| GO | Gene Ontology |
| GSEA | gene set enrichment analysis |
| HGC | Human Genome Connectome |
| InDel | Insertions/deletions |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| MAF | Minor allele frequency |
| MIBC | muscle-invasive bladder cancer |
| NCBI | National Center for Biotechnological Information |
| PID | Primary immunodeficiency |
| PPI | protein-protein interactions |

| | |
|---|---|
| SAM | Sequence Alignment Map |
| PROVEAN | Protein Variation Effect Analyzer |
| SCID | severe combined immunodeficiency |
| SNVs | single nucleotide variants |
| TCGA | The Cancer Genome Atlas |

**Liu Guojun**

# PATHOPHYSIOLOGIC GENE DEPENDENT MECHANISMS OF SEPARATE TYPES OF IMMUNE MEDIATED PATHOLOGY

14.03.03 – Pathological physiology

ABSTRACT

Of dissertation
for the science degree
of candidate of biological sciences