

Federal State Autonomous Educational
Institution of Higher Education
«Ural Federal University
named after the first President of Russia B.N.Yeltsin»
Institute of Natural Sciences and Mathematics

Published as a manuscript

Liu Guojun

**PATHOPHYSIOLOGIC GENE DEPENDENT
MECHANISMS OF SEPARATE TYPES OF IMMUNE
MEDIATED PATHOLOGY
(English version)**

14.03.03 – Pathological Physiology

THE DISSERTATION
for the science degree
of candidate of biological sciences

Scientific advisers:
Tuzankina Irina Aleksandrovna,
Doctor of Medical Sciences, Honored Scientist of
the Russian Federation, Professor

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

Yekaterinburg – 2019

TABLE OF CONTENTS

<u>INTRODUCTION</u>	5
<u>CHAPTER 1 - REVIEW OF THE LITERATURE</u>	16
<u>1.1 - The prevalence, prognosis, treatment, molecular subtypes paka bladder</u> ...	16
<u>1.2 - Classification, prevalence, diagnosis, treatment and prevention of primary immunodeficiencies</u>	23
<u>1.3 - Problems of research of primary immunodeficiencies</u>	29
<u>1.4 - Monogenic and multigenic defects in common variable immunodeficiency</u>	32
<u>1.5 - Phenotype and genotype of Hennekam syndrome</u>	35
<u>CHAPTER 2 - MATERIALS AND METHODS USED IN WORK</u>	37
<u>2.1 -Data collection</u>	37
<u>2.2 - Assessment of the levels of infiltrating immune cells in patients with muscle-invasive bladder cancer</u>	42
<u>2.3 - Cluster analysis of data from patients with muscle-invasive bladder cancer</u>	42
<u>2.4 - Survival analysis of patients with muscle-invasive bladder cancer</u>	43
<u>2.5 - Functional enrichment analysis of genes characteristic of different subtypes of muscle-invasive bladder cancer</u>	43
<u>2.6 - Screening of representative genes between subtypes using machine learning approaches</u>	45
<u>2.7 - Determination of differences in gene expression between tumor and normal bladder tissue samples</u>	47
<u>2.8 - Construction of a subtype-dependent interaction network of mRNA-miRNA-lncRNA for muscle-invasive bladder cancer</u>	47
<u>2.9 - Assessment of gene network density and biological distance for common variable immunodeficiency genes and other primary immunodeficiency genes</u>	48
<u>2.10 -Prediction total variable immunodeficiency candidate genes</u>	50
<u>2.11 - Identification of new candidate genes for common variable immunodeficiency</u>	51
<u>2.12 – Visualization of molecular interaction networks using Cytoscape software</u>	52
<u>2.13 – Whole genome sequencing and search for candidate genes for the patient with Hennekam syndrome</u>	52

<u>CHAPTER 3 - USING MIR-141-5P, MIR-141-3P, MIR-200C-3P, AC010326.3, AC073335.2, MIR100HG, CLIC4, GATA3, PALLD AS PROGNOSTIC BIOMARKERS IN PATIENTS WITH MUSCLE-INVASIVE BLADDER CANCER</u>	54
<u>3.1 - Impact of infiltrating immune cells on the progression and survival in patients with muscle-invasive bladder cancer</u>	54
<u>This logically fits into the pathogenesis of the development of the oncological process, in which the severity of the process correlates with greater activity of tumor macrophages. The results of our research allow recommending M2 macrophages as markers for the MIBC prediction.</u>	55
<u>3.2 - Clustering analysis of RNA expression data of patients with muscle-invasive bladder cancer</u>	55
<u>3.3 - Functional enrichment analysis of data of patients with muscle-invasive bladder cancer</u>	60
<u>3.4 - Search for differentially expressed genes using DEGs method and machine learning</u>	63
<u>3.5 - Construction of a subtype-dependent mRNA-miRNA-lncRNA network for muscle-invasive bladder cancer</u>	66
<u>3.6 - Survival analysis for muscle-invasive bladder cancer</u>	68
<u>CHAPTER 4 - IDENTIFICATION OF CANDIDATE COMMON VARIABLE IMMUNODEFICIENCY (CVID) GENES</u>	72
<u>4.1 - Investigation of the multigenic nature of common variable immunodeficiency</u>	72
<u>4.2 - Analysis of the gene relationships of the group of common variable immunodeficiency syndromes</u>	72
<u>4.3 - Evaluation of identified candidate genes for common variable immunodeficiency</u>	77
<u>CHAPTER 5 - IDENTIFICATION OF MUTATIONS IN GENES, FAT4, RAG1, PIK3CD and CSF3R, LEADING TO THE DEVELOPMENT OF A DISEASE IN A PATIENT WITH PHENOTYPIC MANIFESTATIONS OF HENNEKAM SYNDROME</u>	82
<u>5.1 - Clinical features of the studied patient</u>	82
<u>5.2 - Search for mutations in the patient, which could lead to the development Hennekam syndrome</u>	85
<u>5.3 - Assessment of the probable pathogenicity mutations identified in the study</u>	85
<u>CONCLUSION</u>	93
<u>FINDINGS</u>	97

<u>PRACTICAL ADVICE</u>	98
<u>ABBREVIATIONS</u>	100
<u>REFERENCES</u>	102

INTRODUCTION

Relevance of the research topic

Modern pathological physiology involves a systematic approach to analyze any process in the body. Conventional research methods, such as the assessment of cellular and humoral immune parameters, cannot comprehensively cover all aspects of physiology and pathology of the integration mechanism of the nervous, endocrine and immune systems. Systems analysis requires the use of methods and systems biology, including bioinformatics. Due to the development of genetics, it is possible to identify general patterns of network interaction within the genome, which has allowed us to explain many complex life processes in the body. In the same mechanism, innate and adaptive immunity are the main regulators, performing processes to maintain homeostasis.

The pathology that forms when immunity is impaired is immuno-mediated. Examples of immune-mediated diseases are primary immunodeficiency, autoimmunity, autoinflammation, allergic diseases, as well as cancer pathology that has been studied for a long time.

The cause of different pathologies may be a serious defect in a single gene or genome, usually a disorder of interaction and interference, which ultimately renders known treatments for phenotypic pathology ineffective. This results in the development of complex diseases, lack of therapeutic effects, decreased quality of life and disability of patients.

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 addition, the systematic identification of genes that are significant for diseases allows revealing the pathogenetic mechanisms of their formation, helping

to develop effective personalized treatment methods and new drugs. Therefore, the study of molecular genetic mechanisms of the formation of pathology and prediction of the course of diseases has always been the global basic research topic and an urgent solving problem in the field of personalized medicine.

It is a complex and time-consuming task to identify disease-related genes by biological experiments. With the development of next-generation sequencing (NGS) methods, the accuracy, speed, and cost of sequencing are improved, which greatly promotes the development of this knowledge field. The application of bioinformatics methods in interpreting NGS results has become an important tool for detecting new genes to predetermine disease development.

The rapid development of life sciences, which are integrative and unifying for many fields 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 use of traditional research methods and evaluation of individual human functions, independently, is insufficient to interpret knowledge in the requirements of the modern paradigm of clinical medical research. The interdisciplinary model (bioinformatics, pathophysiology, genetics, and immunology) is gradually becoming the main trend in the development of clinical research.

Perceptions of long-known facts are changing. Thus, diseases that are previously treated as monogenic, are now regarded as polygenic that determines the enormous phenotypic diversity of known diseases. Significant prospects are opening up for the development of personalized treatment of known diseases, taking into account the characteristics of a particular individual, and the range of effective treatment methods for previously prognostically unfavorable pathologies is expanding.

For polygenic diseases, the main methods for identifying causally significant genes are the analysis of co-expression, protein-protein interaction and functional enrichment analysis. When identifying candidate genes for various types of

pathology, their verification is required, which requires the use of both experimental and clinical data, and analysis of independent samples to determine their role in the development of the disease. For a monogenic disease, the pathogenetic significance of the mutation is predicted primarily on the basis of whole exon or whole genome sequencing of the genetic material of the patient and other family members.

The most important manifestations of an immuno-dependent pathology are disorders of proliferation, repair, and inflammation. In our study, diseases that occur with various proliferation disorders, namely bladder cancer, a group of primary immunodeficiencies - “common variable immunodeficiency”, and the primary immunodeficiency associated with syndromic manifestations - Hennekam syndrome, were selected as models of immuno-dependent pathology.

The degree of elaboration of the research topic. The work is based on studies of various immune-dependent diseases, bladder cancer and primary immunodeficiencies, which are fundamentally united by the fact that they are immune-mediated processes, and the pathogenesis of each of them is largely determined by the expression of certain molecules of the immune system.

Bladder cancer (BC) ranks ninth among the most common cancers worldwide with an annual registration of about 430,000 new cases [60]. According to M. Wong et al., 2019, as the population ages and grows, the absolute incidence of bladder cancer in European countries may increase [181]. It was previously noted that some miRNA molecules, such as miRNA-141-5p, miRNA-141-3p, miRNA-200c-3p, and GATA3, are the most important markers of BC [43]. In addition, previous studies have shown that the low expression level of miRNA-200c and miRNA-141 is associated with the increased expression of ZEB1 [48, 61, 62]. Similarly, the low level of expression of miRNA-200c is associated with the decreased expression level of BMI-1 and E2F3 [124], which plays an important role in the processes of invasion, migration, and epithelial-mesenchymal transition of tumor cells of the bladder cancer. In recent years, the study of the regulatory

network between miR-200 family and their target genes are paying close attention by immunologists. It is known that miR-200 family (miR-200a, mi-200b, miR-200c, miR-141 and miR-429) are critical in the pathogenesis of cancer and metastasis, as well as necessary for the formation of memory CD8 + T cells [202].

Primary immunodeficiency (PID) is a heterogeneous group of genetic disorders caused by malformations or dysfunction of immune system components [154]. Recent studies have characterized PID as a widespread pathology that seriously affects the survival, quality of life, and even family economic burden [3]. Our work focuses on one of the most common forms of PID, common variable immune deficiency (CVID), as well as one of the rarest forms, Hennekam syndrome.

In 2015, Itan Y. and Casanova J.L. predicted more than 3,000 PID candidate genes that have close relationships with already known PID genes by analyzing protein-protein interaction (PPI) data. They developed a mathematical model, Human Genome Connectome (HGC), which allows us to predict the PID candidate gene [103]. Back in 2009, Shivakumar et al. predicted 1442 candidate genes using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the support vector algorithm [150]. Ortutay et al., in 2008, predicted 26 candidate genes based on annotation data of protein-protein interactions and enriched Gene Ontology (GO) [141]. In this study, co-expression was considered as key factors that can efficiently predict CVID candidate genes, since genes with similar functions involved in the same signaling pathway are usually more co-expressed [27, 37, 40, 68, 94, 121, 194, 199, 201]. Thus, full use of gene expression profiles of a large number of CVID samples allowed us to better predict candidate genes for CVID.

Hennekam syndrome is an autosomal recessive disorder characterized by the asymmetric expansion of both peripheral and internal lymphatic vessels including the intestines, facial dysmorphism and mild or moderate levels of cognitive impairment resulting from malformation of the lymphatic system [18]. In the United States, the National Institute of Health (NIH) reports at least 50 patients

with Hennekam syndrome in 2015 [160]. There are currently no statistics on Hennekam syndrome in Russia. It was previously reported that mutations in the CCBE1, FAT4, and ADAMTS3 genes can cause Hennekam syndrome (types 1, 2, and 3, respectively), but in-depth functional studies have not been conducted [18, 77, 115]. The relatively low incidence and limited information resources make Hennekam syndrome a complex disease for clinicians and researchers. This justifies the need to identify disease genes using NGS, bioinformatics, and a database of human genetic variations. Make full use of these resources that can significantly reduce the difficulties of diagnosing such diseases and increase curative effectiveness.

Purpose of the study: identification of important pathophysiological mechanisms of the formation of certain types of immune pathologies based on the bioinformatic analysis.

Research objectives:

1. To analyze the pathophysiological mechanisms of heterogeneity of muscle-invasive bladder cancer.
2. To identify prognostic biomarkers for 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 new candidate genes in the disease group “common variable immune deficiency”.
6. To identify gene mutations that lead to the development of Hennekam syndrome.

Methodology and research methods. The work was performed at the Department of Medical Biochemistry and Biophysics of the Institute of Natural Sciences and Mathematics of the Ural Federal University.

Clinical data of patients with muscle-invasive bladder cancer (MIBC) and the corresponding mRNA and lncRNA sequencing were obtained from The Cancer Genome Atlas (TCGA) data portal (<https://cancergenome.nih.gov>), and corresponding miRNA sequencing data publicly available server Broad GDAC Firehose (<https://gdac.broadinstitute.org>).

To perform studies related to muscle-invasive bladder cancer, the raw RNA sequence data of 403 patients and 19 normal tissue control samples were obtained from the TCGA data portal (<https://cancergenome.nih.gov>), the microarray gene expression data of 139 patients with muscle-invasive bladder cancer the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov>).

The microarray gene expression data of 111 patients with common variable immune deficiency and 56 relatively healthy control samples were obtained from the GEO database.

The genomic DNA of a patient with a clinical diagnosis of Hennekam syndrome was isolated from peripheral blood mononuclear cells using a FlexiGene DNA kit (Qiagen) according to the manufacturer's instructions. For sample preparation, the PCR free DNA fragmentation technique was used. Whole genome sequencing (WGS) was carried out at company Genomed (Moscow) using a paired-end sequencing technology. The analysis covers 98.5% of the whole genome, and the average coverage is at least 30x.

The analysis of the results of whole genome sequencing and identification of mutations and single nucleotide polymorphisms (SNPs) was carried out on a supercomputer provided by the center for collective use of IMM UB RAS "ИММ YpO PAH Supercomputer Center" (Yekaterinburg). Informed consent was obtained from the patient's parents for the use of anonymized research results.

Statistical analysis, bioinformatical tools, and mathematical models were performed using Python version 3.7.1 (<https://www.python.org/>) and R version 3.4.3 (<https://www.r-project.org/>) on the Linux operating system.

The degree of reliability and approbation of the results. The reliability of the research results is ensured by theoretical reliability, a sufficient number of samples, reproducibility of studies, statistical data analysis, independent testing datasets. The reliability of the results of the study was confirmed by the Commission for the verification of primary documentation №33.15-03/163 (certificate of verification of primary documents from 02.02.2019 2019). Key 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).

The personal involvement of the applicant. The applicant directly participated at all stages of the study, including the creation of the main idea, the planning of scientific work, the formulation of the working hypothesis, tasks, work assignments, and the determination of the research methodology. Interpretation and analysis of the results of the dissertation were carried out jointly with the supervisor - Honored Scientist of Russia, Professor, Doctor of Medical Sciences, Tuzankina Irina Alexandrovna and Academician, Professor, Doctor of Medical Sciences, Chereshev Valery Aleksandrovich.

In the discussion of the results was attended by, Head of the Department of Medical Biochemistry and Biophysics, Doctor of Biology, Danilova Irina

Georgievna, and, associate professor at the Laboratory of Immunology of Inflammation of the Institute of Immunology and Physiology of the Ural Branch of the Russian Academy of Sciences, candidate of Medical Science, Bolkov Mikhail Artemovich. Separate bioinformatic studies were conducted in collaboration with the Professor of Key Laboratory of Functional genomic bioinformatics of the University of Science and Technology Inner Mongolia (Baotou, China), Guoqing Liu.

The provisions for defense:

1. As diagnostic markers of immune pathology, muscle-invasive bladder cancer, as well as for the prognosis and differential diagnosis of the basal and luminal subtype, it is necessary to use data on the expression of molecules, miR-141-5p, miR-141-3p, miR-200C-3P, AC010326.3, AC073335.2, MIR100HG, CLIC4, GATA3, PALLD.

2. The difference in the expression of the immune checkpoint 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; basal tumors exhibit the higher expression levels of immune checkpoint molecules than tumors of the luminal subtype.

3. The molecular genetic etiology of the group of nosological syndromes of primary immunodeficiencies, common variable immunodeficiency, is an extended spectrum of 172 candidate genes identified in this study.

The scientific novelty of research. 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, differences in the interdependencies between molecules of mRNA, miRNA, and lncRNA were proved for different subtypes of muscle-invasive bladder cancer.

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 based on the results of the analysis of gene relationships.

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 the bioinformatic analysis.

A list of possibly 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. The molecules, miR-141-5p, miR-141-3p, miR-200C-3p, AC010326.3, AC073335.2, MIR100HG, CLIC4, GATA3, and PALLD, can be of great importance in oncogenesis 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 the functioning of the immune system, in the basal subtype of bladder cancer, was revealed. The luminal subtype is characterized by a relatively increased expression of genes responsible for metabolic processes. The higher expression level of the immune checkpoint molecules, PD-1, PD-L1, CTLA-4, HAVCR-2, and LAG-3, were revealed in the basal subtype of bladder cancer, compared with the luminal subtype. These data provide a deeper understanding of the immunological heterogeneity of bladder cancer.

Findings that the CVID genes in their functional relationships are more similar to each other and closely interact with each other compared to other PID genes will help to better understand the multi-gene nature of the common variable immune deficiency. The proposed 172 CVID candidate genes will help to carry out the early diagnosis of CVID and identify new CVID genes.

Homozygous mutation (g.125452634G>A) in FAT4 gene, heterozygous mutation (g.36575963G>A) in RAG1 gene, heterozygous mutation (g.9715914T>A) in PIK3CD gene and heterozygous mutation (g.36471505C>T) in CSF3R together, may have pathogenetic significance and will allow to expand diagnostic possibilities, taking into account understanding of pathogenesis Hennekam syndrome.

Implementation of research results in 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 consulting medical practice.

Publications. According to the results of the thesis the applicant published 7 papers indexed in international electronic databases, Web of Science, Scopus, as well as peer-reviewed Higher Attestation Commission.

Volume and structure of the thesis. The dissertation is presented on 118 pages of typewritten text, consists of the introduction, literature review, materials and methods, three chapters with the results of our own research, conclusions, practical recommendations, list of abbreviations, and list of references (204 sources, including 5 domestic and 199 foreign). The work is illustrated by 12 tables, 30 figures, and three formulas.

Acknowledgments. First and foremost, my special gratitude goes to my supervisors, Professor Tuzankina Irina Alexandrovna and Academician Chereshev Valery Aleksandrovich, for their support and guidance of my thesis. Indeed, they enriched my academic and professional skills by her fruitful scientific discussions, brilliant ideas, and unending motivations.

My deep appreciation to Professor Danilova Irina Georgievna, the head of the Department of Medical Biochemistry and Biophysics, for her assistance in carrying out this study and the many kindnesses she showed me during the whole Ph.D. period.

I am truly thankful to Bolkov Mikhail Artyomovich, associate professor at the laboratory of Immunology and inflammation of the Institute of Immunology and Physiology of the Ural Branch of the Russian Academy of Sciences, for his effective comments and illumination discussions. I did enjoy working with him.

I am grateful to Professor Evgeny Gusev, the head of the laboratory of Immunology and Inflammation of the Institute of Immunology and Physiology, Ural Branch of Russian Academy of Sciences, for a systematic approach to the problems raised in the work and important comments on the common molecular mechanisms and signaling pathways.

I am grateful to, Professor Boris Germanovich Yushkov, doctor of medical science, senior researcher of the Institute of Immunology and Physiology, Ural Branch of the Russian Academy of Sciences, and Alexey Petrovich Sarapultsev for correct and accurate observations which allowed to adjust the work in the right direction.

I express my sincere gratitude to all my friends and colleagues at the Department of Medical Biochemistry and Biophysics for the great work environment, the incredible lovely memories that will always be in mind, and their honest feedback and critical comments that helped me to improve my work.

Last and most importantly, I would like to give my special thanks to my parents, girlfriend, and younger sister, for their care and love to me. Their constant support to me is invaluable, and no words could express my feelings towards them.

CHAPTER 1 - REVIEW OF THE LITERATURE

Oncological diseases and congenital defects of the immune system represent a huge medical and social problem due to the high prevalence, complexity of diagnosis, clinical camouflage, long-term nature of the course of this pathology, as well as complications and the risk of death or disability, which significantly violates the quality of life of patients. Therefore, the problem of immuno-mediated pathology is being actively studied all over the world.

1.1 - The prevalence, prognosis, treatment, molecular subtypes paka bladder

As we know, malignant diseases take a leading place among all diseases leading to death. This class of diseases has a polygenic nature with the synergistic action of several genes. Realization of tumor development can occur within the boundaries of one organ, various organs of one system or several body systems. At least about 33% of cancer cases have evidence of the genetic nature [64].

Bladder cancer (BC) is one of the most common and fatal diseases worldwide. According to the China National Cancer Center, during 2015, BC was detected in 80,500 Chinese citizens and caused 32,900 deaths [26]. BC is the fourth most common cancer in men and the 11th in women, with 76,960 new cases and 16,390 deaths each year in the United States [171]. According to MEDLINE (Medical Literature Analysis and Retrieval System Online), the recurrence rate of BC reaches almost 80%, and this type of oncology is the most expensive for treatment per patient. The high recurrence rate initiates many problems that affect the quality of life due to its stability [81, 179]. BC is usually classified into muscle non-invasive and muscle-invasive bladder cancer (MIBC), depending on whether the cancer cells are limited by their own plate of the mucous membrane (lamina propria) or penetrate into muscularis propria [15, 22, 75]. A quarter of patients with BC suffer from a muscle-invasive form (MIBC) with a higher risk of

metastasis, in which tumor cells can spread to the regional lymph nodes of the pelvis and / or visceral region, resulting in the disease becoming incurable [102].

A large number of researchers indicated that, depending on the general patterns of RNA expression or specific genomic changes, MIBC can be further divided into two main subtypes, namely basal and luminal subtype [8, 43, 44, 85, 100, 101, 152], which are strikingly similar to the molecular subtypes first described in breast cancer [129, 145]. The basal subtype attracted much attention from researchers because it is associated with a more aggressive tumor phenotype and has a higher risk of distant metastases than the luminal one [43, 85]. Molecular biomarkers and signaling pathways involved in different subtypes of muscle-invasive bladder cancer are key steps towards understanding their differences. The first difference between the two subtypes is that they develop etiologically along different signaling pathways.

The basal subtype of MIBC is inherently more aggressive and associated with stronger metastasis, while the luminal subtype is considered to be milder. The following biological markers are described for the basal subtype of MIBC: CD44, CDH3, KRT14, KRT16, KRT5, KRT6A, KRT6B, KRT6C, STAT3. These genes are relatively highly expressed in basal subtype compared with basal subtype). The luminal subtype of MIBC is more characterized by a relative increase in the expression of genes such as CYP2J2, ERBB2, ERBB3, FGFR3, FOXA1, GATA3, GPX2, KRT18, KRT19, KRT20, KRT7, KRT8, PPARG, TRIM24 [120].

To understand the pathogenesis, it is necessary to know the differences in cell differentiation and their development, which leads to differences in the severity of the disease, duration, localization and sensitivity or tolerance to drugs, which, in turn, affects the difference in diagnosis, prognosis and treatment methods cancer, which is still unsatisfactory.

Traditional prognostic criteria used by physicians to calculate survival prognosis include histological data and tumor stage according to the TNM stage determination system, which takes into account the number or size of the tumor

and the number of metastases in the lymph nodes [156, 176, 190]. According to these traditional indicators, the prognosis of cancer patients can be very different, even if they were at the same stage and received conventional treatment. Therefore, it is especially important to find differentially expressed genes or certain mutations that are closely related to the genetic type of cancer for specific treatment. High-throughput sequencing techniques, along with improved sequencing accuracy and reduced costs, have greatly influenced the use of medical biomarkers (signatures) in early diagnosis, cancer prognosis, relapse prognosis, monitoring drug reactions, and the development of targeted treatment methods [30, 65, 84, 133, 164, 200]. All of this will significantly improve patient survival.

Sanguedolce F., in 2015, published evidence that prediction models using molecular markers are more accurate than traditional prediction criteria. Despite the fact that in recent years significant progress has been made both in the identification of molecular markers and in the creation of multifactorial tools for assessing the prognosis of BC, only a few qualified biomarkers are available for use in predicting the course of BC [128, 200]. Thus, it is necessary to identify new molecular biomarkers and include them in multifactorial predictive tools.

For many decades, surgery, chemotherapy, and radiotherapy have been the main strategies of traditional cancer treatment [153, 163]. Radical cystectomy with neoadjuvant cisplatin-based chemotherapy is the first-line multimodal treatment standard for patients with MIBC. However, about 60% of patients with MIBC do not show a significant response to treatment due to residual malignant cells and / or metastasis [89]. In addition, for many patients, the traditionally used cisplatin therapy is contraindicated due to the high toxicity of the drug [168].

Immunotherapy is becoming an attractive and trustworthy strategy among various therapeutic options. In recent years, this method has already proved to be effective in the treatment of malignant tumors, but it is necessary to take into account the balance between stimulating and inhibitory signals, which are crucial to prevent uncontrolled immune reactions that lead to the development of

complications. The effector cells of the immune system can be blocked by various inhibitors called “immune control points”. Receptors on the surface of lymphocytes, such as PD-1 and CTLA-4, are immune control points that suppress T-cell activation signals to protect normal tissues from autoimmunity. However, in cancer, these mechanisms are often activated to suppress the antitumor immune response [66, 196] (Figure 1). In 2013, Science magazine named cancer immunotherapy as "Breakthrough of the Year," and in 2018, the Nobel Prize in Physiology or Medicine was awarded to two researchers, namely James P. Allison and Tasuku Honjo, for studies that ultimately led to the use of immune control point inhibitors for cancer treatment. PD-1 / PD-L1 inhibitors such as pembrolizumab, atezolizumab, durvalumab, nivolumab, and avelumab are currently approved by the US Food and Drug Administration (FDA) for use in clinical practice for patients with advanced MIBC who have platinum-containing chemotherapy regimens was ineffective [17].

In clinical trials, nivolumab, a human immunoglobulin of the IgG4 subclass to the PD1 receptor, showed a strong immune response in less than 20% of patients with locally advanced or metastatic urothelial cancer [153]. Many studies are aimed at developing methods of using immune checkpoint inhibitors to increase their effectiveness, minimize their toxicity and expand their use for a wider range of oncology types.

Tumor mutational burden (TMB) is the number of mutations contained in 1 megabase of the coding genome of tumor cells. TMB is analyzed using whole-exon or whole genome sequencing of biopsy or paraffinized material obtained from patients. TMB is a quantitative biomarker used to predict the sensitivity of solid tumors to the immune checkpoint inhibitor therapy and is independent of PD-L1 expression, but correlates with the frequency of immunotherapy response [135].

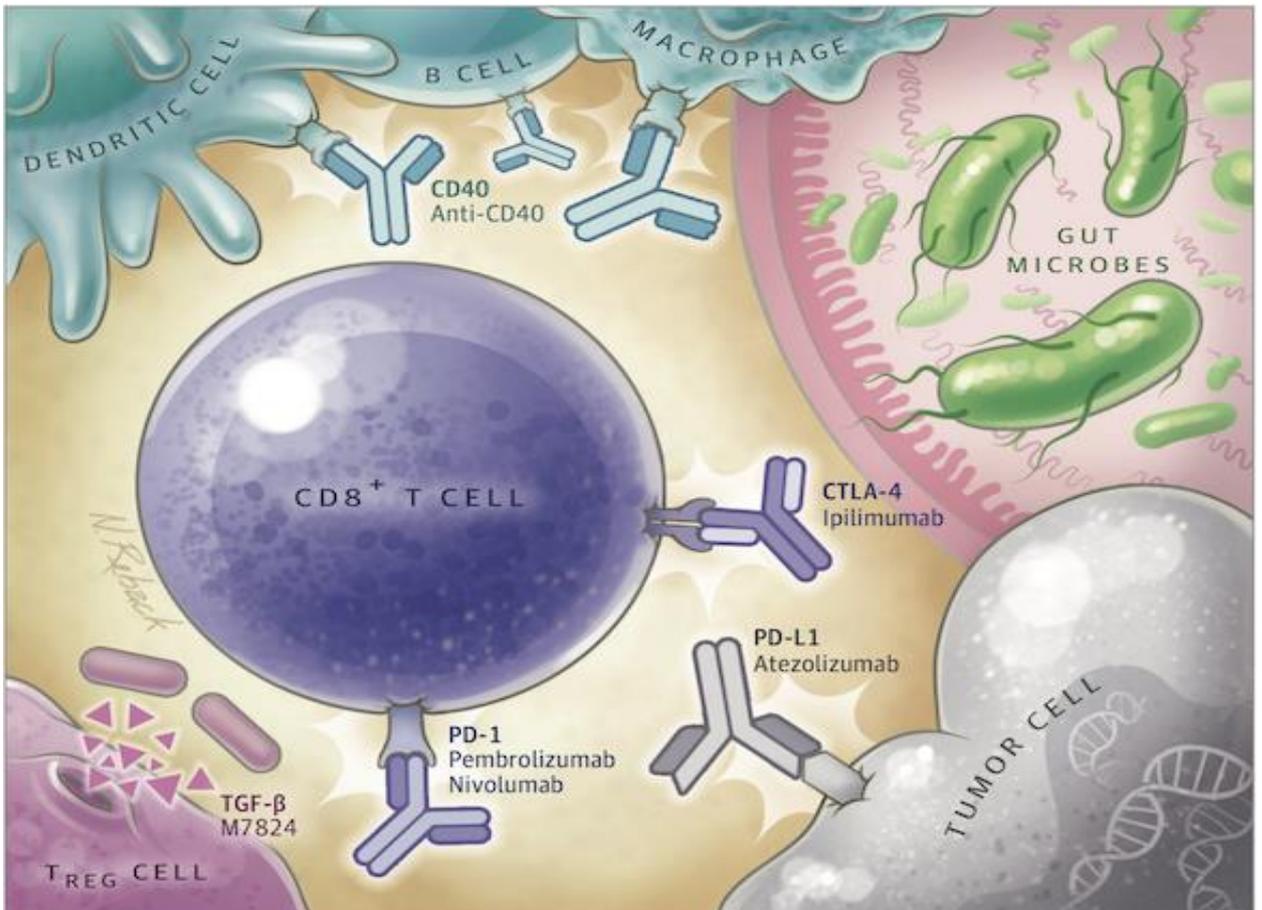


Figure 1 - Treatment of cancer by inhibiting PD1 and CTLA4 [66]

Tumor-infiltrating lymphocytes (TILs) are a favorable prognostic factor in patients with malignant tumors [34, 191]. Tumor infiltration occurs by cytotoxic T-lymphocytes, macrophages, natural killers, dendritic cells, which can have both stimulating and inhibitory effects on tumor growth, and also have different prognostic value [157, 189].

The combination of methods for assessing tumor mutational load (TMB) and tumor-infiltrating lymphocytes (TILs) helps to better determine the susceptibility of patients with some malignant neoplasms to immunotherapy [142]. Identification and study of the molecular subtypes of cancer will help to create an independent assessment of the prognosis and provide additional information to predict the response to immunotherapy [153].

Regarding the molecular drivers of cancer development, it is generally accepted that oncogenic mutations and downstream signaling pathways in a

precancerous or cancerous cell play a crucial role in tumor formation and progression. In addition, recent studies have shown that metabolic reprogramming plays a much more important role than previously thought [25]. It is possible that a large number of genomic mutations detected in cancer provide a selective advantage for cancer cells in the stressful microenvironment of the tumor by reprogramming the processes of cell metabolism [25]. Regardless of what is the main cause of cancer, it becomes clear that both oncogenic signaling pathways and metabolic pathway reprogramming affect many genes that work together in a complex network. Consequently, presenting the regulation of these processes as a gene network provides a deeper understanding of the causes of cancer.

The microRNA (miRNA), as a rule, can regulate gene expression or huge and complex regulatory gene networks for a wide range of cellular processes, such as cell proliferation, metabolism, apoptosis, antiviral immunity [39, 47, 92, 95, 104, 122, 174, 184, 202]. It is known that the miRNA not only has an inhibitory effect on gene expression but, in some cases, also contributes to gene expression [125, 167]. The long non-coding RNA (lncRNA) also have stimulating or inhibiting effects on gene networks during tumor development [159, 165]. In addition, abnormalities or disturbances in microRNA-mediated levels of mRNA and lncRNA expression have a significant correlation with serious clinical consequences, including diseases of various origins and malignant neoplasms [6, 7, 96, 111, 140, 183]. In previous publications, various researchers evaluated the relationships between MIBC-related micro-RNAs and their target genes, but the genetic heterogeneity of the MIBC subtypes was not taken into account [113, 114, 204]. Therefore, one of the objectives of our study was to identify the molecular signaling pathways characteristic of subtypes and new biomarkers, especially since the basal and luminal subtypes differ significantly from each other.

The Cancer Genome Atlas (TCGA) is a project that aims to systematize data on genetic mutations that lead to cancer. This project provides a comprehensive genomic characterization of various cancer pathologies, accelerating our

understanding of cancer biology and setting new scientific and therapeutic goals. The TCGA project, with a budget of more than \$300 million dollars, is the second-largest project after the Apollo Project. The TCGA project was launched in 2005 in conjunction with the National Cancer Institute and the National Human Genome Research Institute and has set itself the difficult task of defeating cancer with the help of researchers from around the world [188]. Today, the TCGA includes omics data from more than 30 types of cancer. Omics technologies are called technologies based on the achievements of genomics, transcriptomics, proteomics and metabolomics, so omics technology is one of the main tools of genomic and postgenomic medicine. For each of the most common types of cancer (lung, breast, prostate, bladder, etc.), the TCGA has accumulated genomic and transcriptomic portraits, as well as methylation data for more than 300-500 samples. Clinical data in the TCGA database includes more than 34,000 samples with reference to the clinical condition at the time of analysis, diagnosis-specific, diagnostic information, initial treatment protocols, and other information (Figure 2).

The increase in the number of tumor sample datasets in the TCGA project helps increase statistical power and the ability to detect molecular defects in cancer. This makes TCGA a useful source of information for evaluating gene expressions [106], molecular classification of a tumor subtype [41, 42, 45], the discovery of changes in driver genes [86], determination of prognostic markers [88, 162], and other applied directions. In addition, recent progress in complex multi-omics analysis has shed light on a new understanding of the genome of malignant neoplasms [9, 67, 87, 93]. With the accumulation of a wealth of multidimensional data, there is a great need for methods aimed at integrative analysis of multiple genomic data sources.

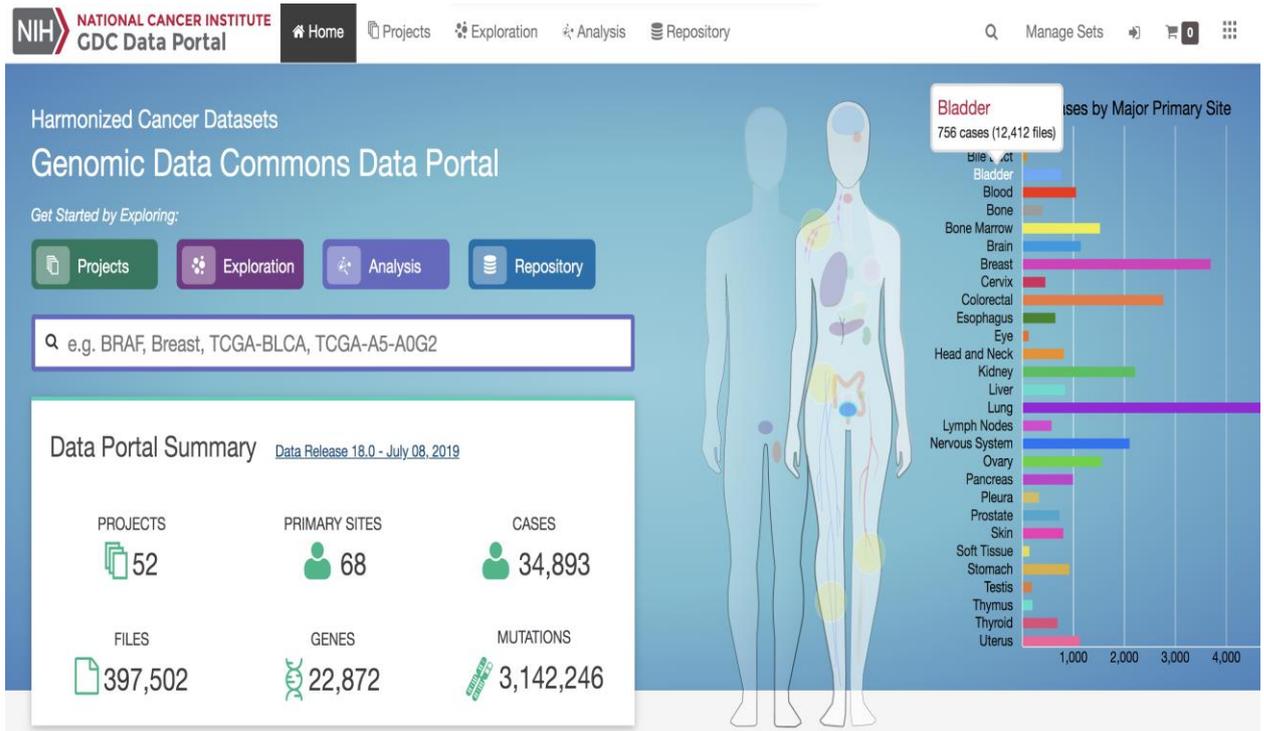


Figure 2 - Screenshot showing the home page of TCGA data portal

Note: the image shows a fragment of the first page of the website <https://portal.gdc.cancer.gov>, while the tooltip shows the number of clinical cases of bladder cancer that were uploaded to the database (756 cases, 12,412 files). On the left, given the general statistical picture of the database, you can search by keywords.

1.2 - Classification, prevalence, diagnosis, treatment and prevention of primary immunodeficiencies

Primary immunodeficiency (PID) is a heterogeneous group of hereditary diseases associated with developmental defects or dysfunction of immune system components [154]. An analysis of 10,000 American families showed that, with the exception of selective deficiency of immunoglobulin A (IgA), the prevalence of PID was close to 1 in 1200. According to Jeffrey Modell Foundation (JMF), different types of PID in varying degrees can be affected by 1-2% of the population of all people on the planet [73]. Moreover, for children with PID, the overall risk of developing malignancy is estimated at 4–25% [55]. Currently, all over the world, including the Russian Federation, during the first year of life, children die due to the lack of timely diagnosis and pathogenetically substantiated therapy [1, 2, 3, 5].

Most often, PID debuts at an early age, affecting the health of newborns and toddlers [109], but new data suggest that the majority of patients survive and grow up, with PID pathology masked under other diagnoses and has a serious burden on the health care system, manifesting itself as chronic difficult-to-treat diseases [4, 119].

One of the first attempts to classify PID was made in 1970 by a committee of the World Health Organization. The classification included 16 nosologies. By 1998, their number reached 50. Currently, their number is approaching 460 [99].

According to the report of the International Union of Immunological Societies, there are 9 PID groups [99]:

1. Combined immunodeficiencies (cellular and humoral defects, including variants of severe combined immune deficiency (SCID) and less severe than SCID combined defects).
2. Combined PID associated with syndromic manifestations (including defects in DNA repair, etc.).
3. Mostly antibody defects (in the absence, decrease or normal number of B-lymphocytes).
4. PID with immune dysregulation (hereditary hemophagocytic syndromes, regulatory T-cell defects, autoimmune lymphoproliferative syndrome, etc.).
5. Congenital defects in the number and functions of phagocytes (congenital neutropenia, defective leukocyte mobility, defective respiratory explosion).
6. Defects of innate immunity (Mendelian susceptibility to mycobacteria, epidermoplastia, susceptibility to viral infections, invasive fungal infections, etc.).
7. Auto-inflammatory syndromes (including ideopathic fevers and aseptic osteomyelitis).
8. Defects in complement.
9. PID phenocopies (caused by somatic mutations or due to autoimmunity to immune system gene products).

Different forms of PID have different pathogenetic mechanisms and can be completely different from each other. Moreover, each of the patients with the same genetically confirmed diagnosis may have different variants of onset and phenotypic manifestations of the disease, from extremely severe to asymptomatic forms. Numerous clinical cases of identifying relatives with the same mutation as in the proband, but without clinical manifestations, have been described [4].

Secondary immunodeficiencies are caused by various causes unrelated to congenital genetic defects. PID is fundamentally different from secondary immunodeficiency, which can result from causes such as deltaretrovirus infections (human T-lymphotropic virus, HTLV) or bacterial infections, malnutrition (including burn disease) and immunoglobulin, malignant tumors or use immunosuppressive drugs [155].

Despite the evidence that the prevalence of diagnosed PID in the United States is approaching 1: 1200, this number does not take into account people with mild defects in the immune system who have not received an official diagnosis due to the lack of genetic diagnosis. IgA deficiency is the most common of PID and is found in the frequency of 1: 300 to 1: 500 people [23]. Other most common forms of PID include common variable immunodeficiency (CVID), severe combined immunodeficiency (SCID), agammaglobulinemia (Bruton's disease), Wiskott-Aldrich syndrome, DiGeorge syndrome, ataxia-telangiectasia (Louis-Barr syndrome), Mediterranean family, Marshall syndrome, hereditary angioedema, etc. [4, 118].

To briefly describe the main differences between PIDs, they depend on the level of the genetic defect and the corresponding receptor or protein defects. Functionally, the immune system is divided into two main components, the innate and adaptive immune response, and depending on this, depending on which component of the immune response is primarily disturbed, two large groups of immunodeficiencies can be distinguished conditionally. However, in fact, the classification presented by the International Union of Immunological Societies

(IUIS) classification most fully reflects possible classes of immunodeficiencies, and defects of innate immunity occupy only 1 of 9 of the possible diversity [99]. One of the first described primary immunodeficiencies was X-linked agammaglobulinemia, described in 1952 by Dr. Ogden Bruton. On simple electrophoresis of plasma proteins, he noted in a boy with symptoms of repeated bacterial infections the almost complete absence of gamma globulins. Today it is known that B cells after their transformation into plasma cells are more responsible for humoral immunity and are the main producers of gamma globulins, that is, antibodies. Defects that occur at any stage of the development and maturation of B-cells lead to some form of hypo- or agammaglobulinemia. Maturation disorders of effector or helper T cells lead to various immunodeficiencies with a defect in the T cell link. But these dependencies are conditional since the production of antibodies by B cells requires a complex of interactions between various T cells and monocytes, and most defects of T cells lead to combined T and B cell insufficiency [56]. However, T- and B-cell mechanisms are the prerogative of the adaptive immune system, while innate immune responses represent the first line of defense against potential pathogens. Corresponding defects of the innate immunity receptors (for example, TLRs) and any of the components of the activation cascade of innate mechanisms lead to the same serious immunodeficiencies as a defect of T- or B-cells. The huge range of PID includes an increasing number of syndromes that are directly related to autoimmunity and immune dysregulation, while the signs of an infectious pathology are secondary and are not the main ones [4, 169].

The specific symptoms of primary immunodeficiency depend on the type of defect. The nature of the infections, as well as additional symptoms, may provide clues as to the exact nature of the immune defect [109]. However, in the diagnosis of immunodeficiencies, not so much the presence of an infectious pathology as other symptoms associated with the pathogenesis of PID is of great importance.

The most common manifestations of PID in the first year of life are:

- lymphopenia ($<2.5 \times 10^9 / l$), pancytopenia;

- thymus hypoplasia based on ultrasound scan results;
- severe skin lesions and / or diarrhea;
- malformations - heart, digestive system, skeletal system, etc .;
- reparation disorders, which can lead to the development of complications of the postoperative period after surgical interventions and other medical procedures;
- inadequate development of the inflammatory process:
 - lack of therapeutic effect,
 - lack of evidence of infectious inflammation,
 - sepsis-like syndrome and hypercytokinemia with lymphohistiocytosis;
- mycoses, mucocutaneous candidiasis;
- hypogammaglobulinemia - an extreme decrease in the concentration of immunoglobulin G in the blood serum;
- Bacillus Calmette–Guérin (BCG) and other vaccine infections;
- persistent diabetes mellitus and other autoimmune diseases;
- late fall of the umbilical cord;
- family history: cases of regressive pregnancy or early infant mortality [4].

Among the other signs of PID in adulthood is an extensive list of diverse symptoms, including birth defects in various organs, including the heart and bone system, skin and connective tissue diseases, autoimmune diseases, when the immune system attacks its own tissues, or tumors (sometimes specific forms of cancer, such as lymphoma), auto-inflammatory diseases, when the cascade of inflammatory reactions is disrupted and tissue destruction occurs due to spontaneous unregulated inflammation. Immunodeficiency patients have an increased risk of developing the malignant neoplasm transformation of tissues due to various congenital defects of the immune system associated with the recognition of cancer cells, their resistance, as well as with numerous defects of DNA repair. Moreover, due to the success of medicine, the life expectancy of patients with primary and secondary immunodeficiencies increases, which leads to an increased

risk of developing malignant neoplasms in patients who survived despite the presence of immunodeficiency.

The diagnosis of PID can be divided into three types depending on the degree of evidence. The criteria for diagnosis were agreed in 1999 [46]. There were three types of diagnosis: possible, probable and final.

The final diagnosis is made when the patient has a 98 % probability of the same diagnosis after 20 years; this level of diagnosis can be established with genetic confirmation or under very precise and specific circumstances and phenotypic manifestations.

A probable diagnosis is established when there is no genetic diagnosis, but the patient has all the other characteristics of a particular disease; the probability that the same diagnosis will be made in 20 years is estimated at 85-97%.

A possible diagnosis is established when the patient has certain signs of the disease, but not all [46].

Current treatments for PID include symptomatic support, targeted therapy, substitution therapy, and two types of radical surgery: hematopoietic stem cell transplantation (HSCT) and gene therapy. It should be noted that gene therapy is still in the experimental stage of research. For some children with PID, HSCT is the most important and even the only way to treat the disease and restore immune system functions. Moreover, the lack of genetic confirmation is not a contraindication for HSCT.

Among the so-called preventive methods for PID, neonatal screening is available. The objective of screening is to establish a preliminary diagnosis in the newborn before the onset of symptoms when there is still a "window of opportunity" for the timely application of curative measures. Currently, at least 54 diseases are included in newborn screening in the United States [137]. At the same time, screening of newborns for defects in T-and B-cell immunity by evaluating TREC and KREC molecules in dry blood spots is already being performed in all

States, in many countries of Europe and the Middle East, Southeast Asia. In Russia, this method is used as a pilot study.

1.3 - Problems of research of primary immunodeficiencies

The following facts make primary immunodeficiency a complex group of diseases for both the practitioner and the scientific researcher [147].

First, the low awareness of doctors about the rapidly developing field of medicine. This problem is facing doctors all over the world. Relatively recently, PID was considered a rare disease, so many general practitioners as well as specialists, have low alertness or awareness of PID. Many researchers of common variable immune deficiency (CVID) note that the time delay between the manifestation of symptoms and the beginning of immunoglobulin replacement therapy is 5 to 10 years [50, 53, 110, 178].

The second stumbling block is that phenotypic changes associated with PID are generally very diverse. For example, in the case of patients with Wiskott-Aldrich syndrome, the exact nature of the gene defect, for example, missense or nonsense mutation, the exact location of splice site abnormalities, can significantly change the phenotype of the syndrome. Manifestations of mutations of this gene can be from very severe to mild, for example, X-linked thrombocytopenia, B-cell lymphoma, frequent bacterial and fungal infections, eczema, small platelets, or neutropenia. Diagnosis of PID usually requires an in-depth analysis of symptoms, coupled with an assessment of history and genealogy.

The last problem of researchers and clinicians is the difficulty of finding information. Open access databases contain few samples, the diversity of which, in comparison with data on Oncology, is also limited. There are very few information resources that link clinical descriptions and functional genomic data, protein-protein interactions, and signaling pathways. Examples of specialized databases include UniProt, IntAct, STRING, and KEGG.

Several databases, registries, knowledge bases, prediction tools, and expert systems have rapid development driven by the diagnostic requirements. According to a paper published by Crina Samarghitean et al. in 2009, tools helpful for PID diagnosis can be divided into the following seven categories (Figure 3):

Firstly, the main PID resources provide a lot of information, ranked by levels from gene to protein structures, from disease models to certain groups of diagnoses, and so on.

Secondly, there are classifications of PID containing clinical signs.

Third, laboratory criteria and related tools are available.

Fourth, there are national and international PID patient registers that are supplemented by mutation databases (fifth category), of which information can be used to compare the case in question with the previously described ones.

Sixth, bioinformatics tools are available to predict or prioritize new candidate PID genes, which are also used for decision making (seventh category) in PID diagnostics [166].

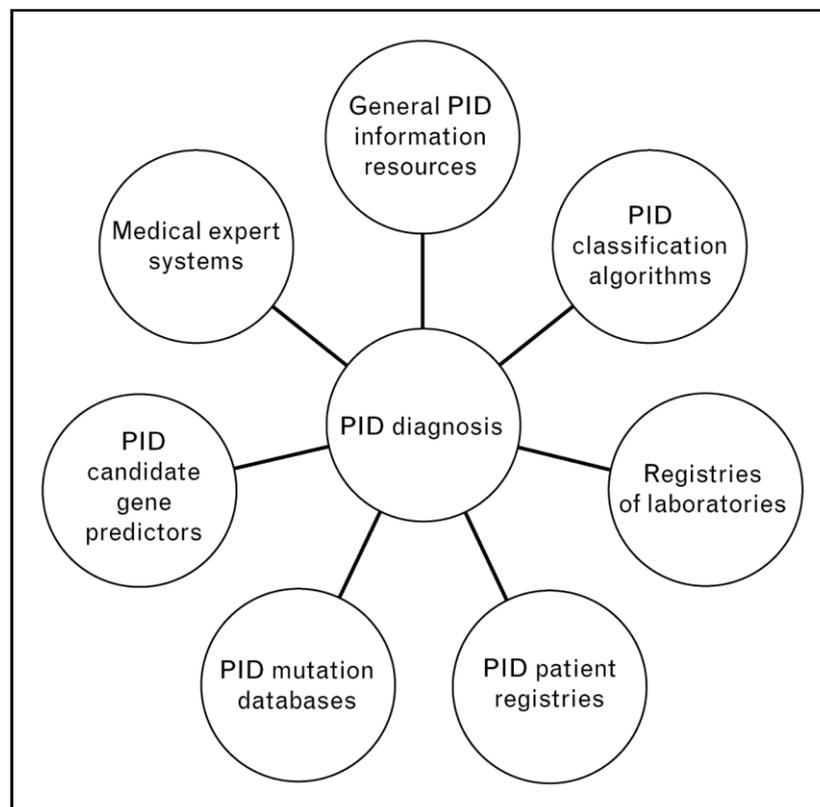


Figure 3 - Schematic grouping of bioinformatic resources and tools providing information on primary immunodeficiencies [166]

A large number of various PIDs creates difficulties for diagnosis, including at the clinical level. In addition, many diseases are so rare that it is not possible to find a sufficient number of families for linkage analysis. Screening and early detection of PID is a major challenge for physicians. A greater number of currently known genetic defects have been obtained in the last few years through high throughput sequencing. The identification of new candidate genes for PID will allow further prioritization of genes when confirmed in patients with PID, in which the exact causative gene has not yet been determined.

In 2009, Keerthikumar et al. used the support vector machine method to classify all human genes into PID genes and non-PID genes. The principle of this classification was based on the calculation of a confidence indicator for each PID candidate gene based on 69 features observed for 148 known PID genes [150]. Based on literature searches, we found that the attention of scientific researchers involved in the identification of PID genes was also focused on the integration of functional annotations from gene ontology (GO) and compilation of data sets of the network of protein-protein interactions. Ortutay et al., 2008, predicted 26 PID candidate genes by analyzing the properties of the protein-protein interaction (PPI) network of all known genes of the human immune system and their ontology (GO). Also, based on the calculation of the so-called biological distance (meaning functional interdependence), 3,110 PID candidate genes were predicted [28, 103].

Researchers wishing to participate in the research on the problem of PID, each time come across the question: are PID multigenic or monogenic? Initially, PID was considered congenital and monogenic disorders that follow the principles of Mendelian inheritance [24, 161]. Monogenic diseases are the result of changes in a single gene that occur in all cells of the body. However, progress in DNA sequencing has contributed to the identification of the multigenic and somatic causes of PID, wide phenotypic variability of these diseases has been observed [74, 192].

Understanding that most PIDs are multigenic in nature is the first step in understanding the pathogenesis of all diseases. In a multi-gene concept, PIDs are the result of complex interactions between genes. Based on this, scientists tried to find the “biological distance” between the PID genes and other human protein-coding genes, and they found that PID genes, as compared to other human genes, are usually located in the central hub of the human genome network and more closely interact with each other (Figure 4) [103]. In addition, PID genes form several closely related sub-clusters, and most of them have at least one functionally close neighbor among a wide range of biological mechanisms [28, 103, 182]. Uncovering these relationships can provide a deeper understanding of the diversity of genetic pathways that underlie PIDs, which in turn can help open up new opportunities for drug development and therapeutic approaches.

1.4 - Monogenic and multigenic defects in common variable immunodeficiency

Common variable immunodeficiency (CVID) is the most common form of PID among adult patients, characterized by low levels of serum immunoglobulins and a deficiency of specific antibodies with the exclusion of other causes of hypogammaglobulinemia. The estimated prevalence of CVID is from 1:50,000 to 1:25,000 in the population [32]. In most patients with CVID, immunodeficiency is detected in the second, third or fourth decade of life after several pneumonia, however, the debut of the diagnosis is also possible in children and the elderly. The main symptoms may be bacterial, viral, fungal and parasitic infections. However, in most patients, non-infectious manifestations are leading and patients are observed in rheumatologists, gastroenterologists, and oncologists with other diagnoses. About 50% of patients have autoimmune manifestations. There is an increased risk of malignancy.

Most CVIDs are multigenic, their cause may be different mutations of many genes, which result in the same or different phenotypes of CVID. Monogenic causes of CVID were found in about 10% of cases [32].

The monogenic causes of CVID include the following mutations:

- I. Recessive inherited mutations in IL21, IL21R, LRBA, ICOS, PRKCD, CD19, CD20, CD21, CD27, CD81, RAC2;
- II. Dominantly inherited mutations in CTLA4, TNFSF12, NFKB1, PLCG2, NFKB2, PIK3CD, PIK3R1, VAV1, BLK, IKZF1, IRF2BP2;
- III. Mono-allelic or biallelic mutations in TNFRSF13B and TNFRSF13C [71].

For example, the B-cell co-receptor complex consists of CD19, CD21, CD81, and CD225, which together lower the threshold for B-cell activation by antigen binding to the B-cell receptor. Recent studies have identified mutations in CD19, CD81, and CD21 receptor genes in autosomal recessive forms of common variable immunodeficiency [12].

The B-lymphoid tyrosine-protein kinase (BLK) plays an important role in signaling from the B-cell receptor (BCR) and in interfacing with T cells. Previously, a heterozygous mutation (mutation with loss of gene function) in the BLK gene was detected in two related patients with CVID [59]. NFKB1 encodes the mature subunit P52 and its precursor P105 transcription factor NF- κ B in B lymphocytes. And for this protein, corresponding clinical cases have also been found where mutations of this gene in sporadic or related cases led to the appearance of CVID [76, 144]. Mutation of the TNFRSF13B gene leads to deterioration of already impaired Toll-like receptor (TLR) signaling or causes defects in TLR signaling [134], which has also been found in many patients with CVID [71].

Despite the fact that all cases of CVID are believed to be the result of genetic defects, the vast majority of the exact cause of the disease is unknown [98].

There are numerous examples of multigene (or polygenic) causes of CVID, in which mutation variants in several genes can contribute to the formation of the same or different phenotypes of this disease [16, 63, 105].

Why is CVID a more multi-gene cohort compared to other PIDs? One possible reason is the fact that the multigene nature of CVID is the result of complex interactions between genes.

Understanding the mechanism of their interaction should help physicians and researchers gain insight into the pathophysiology of PID, which will enhance diagnosis and change treatment approaches. Itan Y. and Casanova J.L. in 2015, by calculating the biological functional distance for all protein-coding genes, it was shown that when visualizing the human genome network, the CVID genes are usually located in the central hub of the network (Figure 4) [103].

Thus, a closer interaction of the CVID genes with each other was revealed, compared with other PID genes. This study partially helps to better understand the multigenic nature of CVID and other diseases.

The main mechanisms underlying the network of protein-protein interactions (PPI) of the CVID genes remain unclear; there is no systematic level of interpretation. With the recent accumulation of new gene expression data in CVID [16, 71], reliable molecular interactions data [21, 79, 83], and state-of-the-art computational methods [20, 21, 149], there is an urgent need to identify the candidate genes of common variable immunodeficiency. The use of methods of systems biology and bioinformatics will accelerate and increase the accuracy of the detection of new CVID genes, which will allow a deeper understanding of the pathogenetic mechanisms of this pathology. In addition, it is an inexpensive and quick method that helps the clinical doctor in diagnosing patients with the CVID phenotype and an unknown genetic cause.

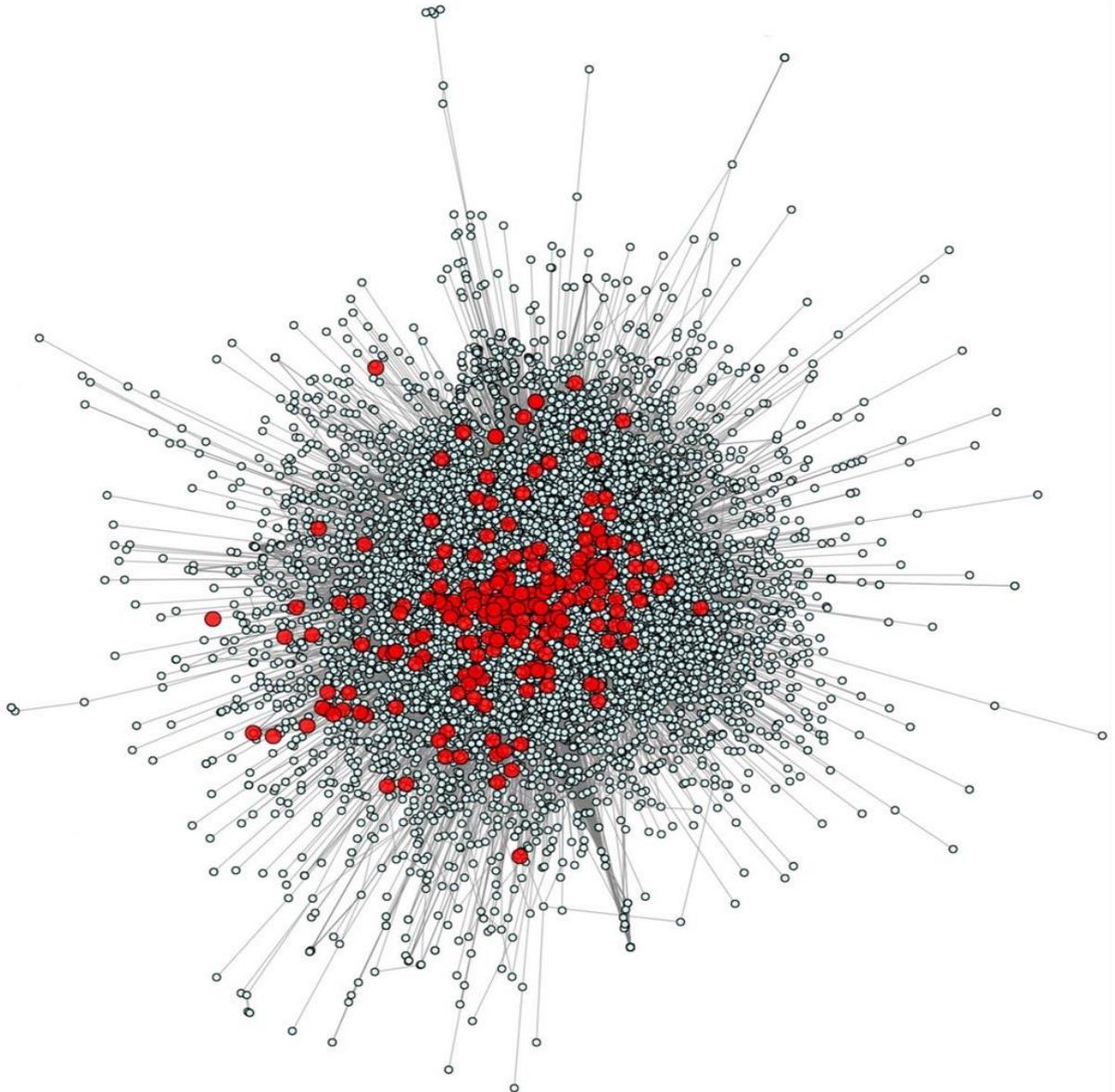


Figure 4 - Network of the human genome and PID genes

Note: 229 known at the time of the study PID genes (red) and 14131 protein-coding genes were shown, for which information on the so-called “biological distance” was available [103].

1.5 - Phenotype and genotype of Hennekam syndrome

Hennekam syndrome is an autosomal recessive disease, and it is one of the rarest forms of primary immunodeficiency. In fact, the Hennekam syndrome refers to malformations of the lymphatic system.

Lymph vessel abnormalities block lymph circulation, and fluid accumulation affects various areas of the body, including the face, limbs, and internal organs.

Facial features are characterized by a flattened appearance, a wide nose bridge, hypertelorism, epicanthus and other anomalies [18].

Hennekam syndrome is divided into three main types, depending on the genetic cause affecting the VEGF-C / VEGFR-3 signaling pathways [116, 132, 177]. Type 1 Hennekam syndrome, also known as CCBE1-associated Hennekam syndrome, was originally described by Dutch physician Raul Hennekam in 1989 [18].

The main molecular mechanism of lymphedema in type 1 Hennekam syndrome is considered to be the reduction in the ability of the mutated CCBE1 (collagen and calcium-binding protein 1 containing a protein similar to epidermal growth factor) to accelerate and concentrate the activation of primary lymphangiogenic growth factor VEGF-C [132].

For type 2 Hennekam syndrome, the cause is a homozygous or compound heterozygous mutation in the FAT4 gene on chromosome 4q28. Interestingly, a mutation in the FAT4 gene can cause Van Maldergem syndrome (VMLDS2), another disease in which some symptoms overlap with the phenotype of Hennekam syndrome [77].

In a 2017 study, a team of authors led by P. Brouillard identified type 3 Hennekam syndrome, which revealed a heterozygous mutation in the ADAMTS3 gene on chromosome 4q13. More importantly, the researchers emphasized the close functional relationship between the ADAMTS3 and CCBE1 proteins in the activation of the VEGFR3 molecule, which is the cornerstone for the differentiation and functioning of lymphatic endothelial cells [115].

However, mutations in these genes are found only in some patients, and the genetic etiology of most patients with Hennekam syndrome remains unknown largely because the syndrome is genetically heterogeneous.

CHAPTER 2 - MATERIALS AND METHODS USED IN WORK

2.1 -Data collection

Clinical data of patients with MIBC and mRNA and lncRNA sequencing data normalized by the FPKM method (fragments per kilobase of exon model per million mapped reads) were obtained from the publicly available The Cancer Genome Atlas server (TCGA) from the web address <https://cancergenome.nih.gov>.

The miRNA sequencing data normalized by the RPM (reads per million) method was downloaded from the publicly available Broad GDAC Firehose server (<https://gdac.broadinstitute.org>). Samples designated as muscle-invasive urothelial carcinoma at pT2 or higher were selected according to clinical data, and we excluded samples that did not have long-term follow-up data because it was necessary to evaluate patient survival. The presentation style clinical data is presented as an example in Table 1.

Table 1 - Presentation style of clinical data for patients with MIBC

Sample \ Clinical	Gender	Stage	Satus	Time
TCGA-UY-A78P	FEMALE	Stage II	Alive	398
TCGA-LT-A8JT	FEMALE	Stage II	Alive	1522
TCGA-CF-A7I0	MALE	Stage II	Alive	641
.....
.....
.....

The columns presented identifiers of the tumor tissue sample, diagnosis, tissue, tumor type, stage, metastasis, patient race, age, gender, whether the patient smoked or not, and whether the patient was alive at the time of the study or not.

The mRNA sequencing dataset of patients was a table of expression levels for each of 19,181 mRNAs for 422 tissue samples (403 tumor and 19 normal tissue samples).

The presentation style of patient mRNA expression data is presented in Table 2. The columns show sample identifiers and expression levels for each mRNA.

Table 2 - Presentation style of mRNA data from patients with MIBC

mRNA \ Sample	TCGA-DK-AA6R	TCGA-DK-AA76	TCGA-DK-AA6P
SAMD11	0.90622308	0.47413391	0.98372
NOC2L	15.914834	42.7130759	23.0476527154
KLHL17	2.3565348	3.85048108	2.32950289
.....
.....

The lncRNA dataset of patients with MIBC consisted of the expression values of 14,376 different lncRNAs and 422 samples (403 tumor and 19 normal tissue samples). The presentation style of patient lncRNA expression data is presented in Table 3. The columns contain sample identifiers and expression levels for each lncRNA.

Table 3 - Representation style of lncRNA expression data from patients with MIBC

lncRNA \ Sample	TCGA-DK-AA6R	TCGA-DK-AA76	TCGA-DK-AA6P
LINC01128	0.47886052	0.72885094	0.39689252
LINC00115	0.39191886	0.90750898	0.29728013
FAM41C	0.01870541	0	0
.....

The miRNA dataset of patients with MIBC contained expression values of 2,588 different mRNAs and 422 samples (403 tumor and 19 normal tissue samples). The presentation style of patient miRNA expression data is presented in Table 4. The columns contain sample identifiers and expression levels for each miRNA.

Table 4 - Representation style of miRNA expression data from patients with MIBC

miRNA \ Sample	TCGA-DK-AA6R	TCGA-DK-AA76	TCGA-DK-AA6P
hsa-let-7a-3p	54.639048	48.253177	37.221532
hsa-let-7a-2-3p	9.446938	7.659235	2.788584
.....
.....
hsa-miR-32-5p	68.937116	43.274676	23.884825
hsa-miR-32-3p	1.787259	1.148885	0.242486

DNA microarray data sets for patients with MIBC and CVID were obtained from the GEO database (Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov>).

For studies of genes associated with MIBC, the GSE13507 and GSE31684 datasets were downloaded. After sampling, marked as muscle-invasive bladder cancer (pT2 stage or higher), the GSE31684 dataset contained 78 samples, GSE13507 61 samples, which were subsequently used to predict genes associated with the disease based on gene expression levels.

For studies of genes associated with CVID, the GSE72625 and GSE51406 datasets were downloaded. The GSE72625 dataset contained data from 20 patients with CVID and 17 relatively healthy samples, the GSE51406 dataset contained information from 91 patients with CVID and 39 relatively healthy samples. The preprocessing procedure for GSE72625 and GSE51406 consisted of removing

genes containing more than 70–75% of the missing values and log₂ transformation of the gene expression values.

The list of 351 known PID genes at the time of the study, including 39 CVID genes, was obtained from the website of the European Society of Immunodeficiency (<https://esid.org>).

The following is a list of 351 PID genes used in our study:

ACP5, ACTB, ADA, ADAM17, ADAR1, AICDA, AIRE, AK2, AP1S3, AP3B1, AP3D1, APOL1, APOLLO, ARPC1B, ATM, ATP6AP1, B2M, BACH2, BAFF-R, BCL10, BCL11B, BLM, BLNK, BTK, C1QA, C1QB, C1QC, C1R, C1S, C2, C2BP1, C3, C4A, C4B, C5, C6, C7, C8A, C8B, C8G, C9, CARD11, CARD11, CARD14, CARD15, CARD9, CASP10, CASP8, CCBE1, CD19, CD247, CD27, CD3D, CD3E, CD3G, CD40, CD40LG, CD46, CD55, CD59, CD70, CD79A, CD79B, CD81, CD8A, CDCA7, CEBPE, CECR1, CFB, CFD, CFH, CFHR1-5, CFI, CFP, CFTR, CHD7, CIAS1, CIITA, CLCN7, CLPB, COPA, CORO1A, CR2, CSF2RA, CSF2RB, CSF3R, CTC1, CTLA4, CTPS1, CTSC, CXCR4, CYBA, CYBB, DCLRE1B, DCLRE1C, DKC1, DNAJC21, DNASE1L3, DNASE2, DNMT3B, DOCK2, DOCK8, ELANE, EPG5, ERCC6L2, EXTL3, FAAP24, FADD, FASLG, FAT4, FCGR3A, FCN3, FERMT3, FOXP3, FPR1, G6PC3, G6PD, G6PT1, GATA2, GCS1, GFI1, GINS1, GOF, HAX1, HELLS, HMOX, HOIL1, HOIP1, HYOU1, ICOS, IFIH1, IFNAR2, IFNGR1, IFNGR2, IGHM, IGKC, IGLL1, IKBA, IKBKB, IKBKG, IKZF1, IL10, IL10RA, IL10RB, IL12B, IL12RB1, IL17F, IL17RA, IL17RC, IL1RN, IL21, IL21R, IL2RA, IL2RG, IL36RN, IL7R, INO80, IRAK1, IRAK4, IRF2BP2, IRF3, IRF7, IRF8, ISG15, ITCH, ITGB2, ITK, JAGN1, JAK1, JAK3, KDM6A, KMT2D, LAMTOR2, LAT, LCK, LIG1, LIG4, LPIN2, LRBA, LYST, MAGT1, MALT1, MAP3K14, MASP2, MCM4, MEFV, MKL1, MLL2, MOGS, MS4A1, MSH6, MSN, MTHFD1, MVK, MYD88, MYSM1, NBAS, NBS1, NCF1, NCF2, NCF4, NCSTN, NEMO, NFAT5, NFKB1, NFKB2, NFKBIA, NHEJ1, NHP2, NLRC4, NLRP1, NLRP12, NLRP3, NOD2, NOP10, NSMCE3, ORAI1, OSTM1, OTULIN, PARN, PEPD, PGM3,

PIK3CD, PIK3R1, PLCG2, PLEKHM1, PMS2, PNP, POLA1, POLE, POLE2, PRF1, PRKCD, PRKDC, PSEN, PSENN, PSMB8, PSTPIP1, PTEN, PTPRC, PYPAF1, RAB27A, RAC2, RAG1, RAG2, RANBP2, RASGRP1, RBCK1, RECQL3, RELB, RFX5, RFXANK, RFXAP, RHOH, RLTPR, RMRP, RNASEH2A, RNASEH2B, RNASEH2C, RNF168, RNF31, RNU4ATAC, RORC, RPSA, RTEL1, SAMD9, SAMD9L, SAMHD1, SBDS, SEMA3E, SERPING1, SH2D1A, SH3BP2, SLC29A3, SLC35C1, SLC46A1, SMARCAL1, SMARCD2, SNM1, SNX10, SP110, SPINK5, SRP54, STAT1, STAT1, STAT2, STAT3, STAT5B, STIM1, STK4, STN1, STX11, STXBP2, TACI, TAP1, TAP2, TAPBP, TAZ, TBK1, TBX1, TCF3, TCIRG1, TCN2, TERC, TERT, TFRC, THBD, TICAM1, TINF2, TIRAP, TLR3, TMC6, TMC8, TMM173, TNFAIP3, TNFRSF11A, TNFRSF13B, TNFRSF13C, TNFRSF1A, TNFRSF4, TNFRSF5, TNFRSF6, TNFSF11, TNFSF12, TNFSF5, TNFSF7, TPP1, TPP2, TRAC, TRAF3, TRAF3IP2, TREX1, TRNT1, TTC37, TTC7A, TYK2, UNC13D, UNC93B1, UNG, USB1, USP18, VPS13B, VPS45, WAS, WDR1, WIPF1, WRAP53, XIAP, ZAP70, ZBTB24.

A list of 39 genes associated with COVID, used in this study:

ACVR2A, BCL2L1, BTK, CD19, CD81, CLEC16A, CLU, CORO1B, CTLA4, FAS, FCGR2A, FUS, GATA2, IKZF1, IL10RA, IL21R, IL2RG, ITGAM, ITPKB, KCNN4, LRBA, MS4A1, NFKB1, NFKB2, NLRP12, PIK3CD, PIK3R1, PLCG2, PRKCD, PTPRCAP, RAC2, RAG1, RPS6KB2, STAT1, TENM3, TNFAIP3, TNFRSF13B, TNFSF12, VAV1.

The DNA sample from a patient with a diagnosis of Hennekam syndrome was extracted from peripheral blood mononuclear cells using a FlexiGen DNA isolation kit (Qiagen) according to the manufacturer's instructions. For sample preparation, the PCR free DNA fragmentation technique of sample preparation technology for analysis was used. The whole genome sequencing was performed by a company, Genomed (Moscow), using a paired-end sequencing technique. The analysis covers 98.5% of the total genome. The average coverage was at least 30x.

The study protocol was approved by the ethics committee of the Institute of Natural Sciences and Mathematics of Ural Federal University, informed consent was obtained from the parents.

2.2 - Assessment of the levels of infiltrating immune cells in patients with muscle-invasive bladder cancer

Relative levels of 22 tumor infiltrating immune cells (including Eosinophils, neutrophils, monocytes, plasma cells, naive B-cells and B-cell memory, resting and activated dendritic cells, resting and activated mast cells, resting and activated NK, macrophages M0, macrophages M1, macrophages M2, CD8 + T-cells, Treg, Naive CD4 + T-cell, follicular helper T-cells, $\gamma\delta$ T-cells, resting CD4 + memory T-cells, activated CD4 + T-cells) for each sample was evaluated by using CIBERSORT, with matrix signature LM22 and 1,000 permutations (<https://cibersort.stanford.edu>). The sum of all fractions of the evaluated infiltrating immune cells is equal to one for each tumor sample, and all estimates, therefore refer to the total content of leukocytes.

2.3 - Cluster analysis of data from patients with muscle-invasive bladder cancer

Prior to cluster analysis, we selected variable genes and performed a logarithmic normalization. For mRNA, miRNA, and lncRNA datasets from patients with muscle-invasive bladder cancer (MIBC), 25% of the genes that had low arithmetic mean of the gene in the samples were first removed, and then they retained 3000 mRNA, 300 miRNA, and 3000 lncRNA that have higher expression variance. Subsequently, all gene expression values were added by 1 and then log₂-normalized.

In previous studies, a combination of the Cluster of Cluster (CoC) method and consensus cluster analysis has proven to be a reliable way to cluster data for

cancer research [97, 112, 131]. Consensus cluster analysis refers to a situation in which different (input) clustering was obtained by analyzing different data sets, but it is necessary to find one (consensus) clustering. Cluster Clustering analysis is a method of integrating primary clustering results into final cluster assignments. Initially, a consensus cluster analysis was performed on the basis of mRNA, miRNA, and lncRNA data, respectively, then the CoC was analyzed to determine the subtypes of basal and luminal bladder cancer, taking into account all three datasets.

2.4 - Survival analysis of patients with muscle-invasive bladder cancer

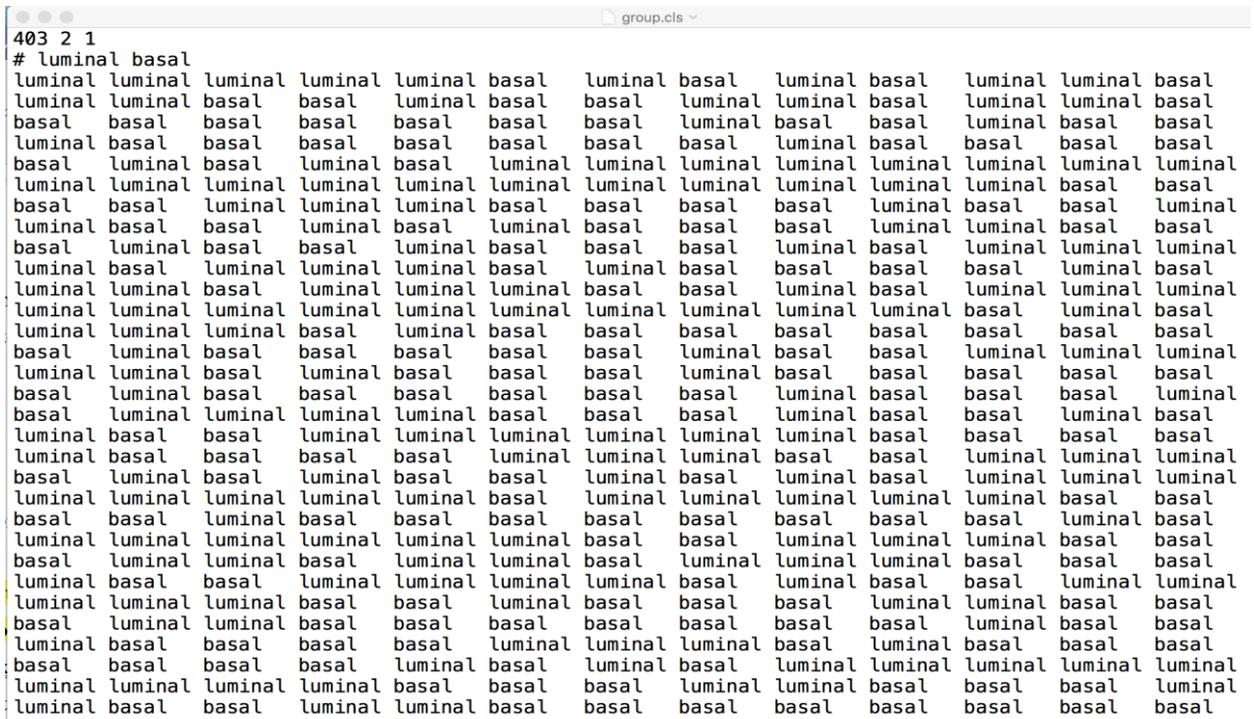
Survival analysis was carried out on the basis of the Kaplan–Meyer (K-M) curve using the “survival” R package (program) [187]. A log-rank test was used to compare survival times between two subtypes. We also evaluated whether the expression of the involved genes correlates with the overall survival of patients with MIBC. To assess the prognostic value of the gene, patient samples were divided into high and low expression groups with median levels of gene expression as a cut-off value. To compare the survival time between the two groups, the log-rank test was used. Results with a coefficient of $p < 0.05$ were taken into account.

2.5 - Functional enrichment analysis of genes characteristic of different subtypes of muscle-invasive bladder cancer

Gene Set Enrichment Analysis (GSEA) is a set of methods for associating a set of genes with a phenotype change [70]. Such methods often use databases of previously annotated gene sets to formalize existing phenotype data (for example, the Gene Ontology (GO) project terms: molecular functions, biological processes, or cellular components [130]). The result of applying the method (issuing a program) in this case is a set of pre-annotated sets that help determine whether the ordered list of genes depends on the phenotype or whether they are simply randomly distributed [69]. Such pre-annotated sets are called overrepresented (if

the frequency is higher than the background) or underrepresented (if the frequency is lower than the background). The enrichment rate (ES) is a statistic similar to Kolmogorov-Smirnov, representing the amount to which genes are overrepresented at the top or underrepresented at the bottom of a ranked list of genes.

Enrichment analysis was performed using the GSEA software (version 3.0), based on the Gene Ontology Biological Process (GO-BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), and the Hallmark annotation data set to correlate the MIBC subtypes with biological processes reported in these databases. Three genetic data sets (“h.all.v5.1.symbols.gmt”, “c2.cp.kegg.v5.1.symbols.gmt” и “c5.bp.v5.1.symbols.gmt”) were used as input data. The mRNA expression dataset consisted of 19181 mRNAs and 403 tumor samples. The phenotype label (cls) file defines the phenotypes under study and associates each sample in the dataset with one of these phenotypes. The resulting file is shown in Figure 5.



```

403 2 1
# luminal basal
luminal luminal luminal luminal luminal basal luminal basal luminal basal luminal luminal basal
luminal luminal basal basal luminal basal basal luminal basal basal luminal basal luminal basal
basal basal basal basal basal basal basal basal luminal basal basal luminal basal basal
luminal basal basal basal basal basal basal basal basal basal luminal basal basal basal
basal luminal basal luminal basal luminal luminal luminal luminal luminal luminal luminal luminal
luminal luminal luminal luminal luminal luminal luminal luminal luminal luminal luminal basal basal
basal basal luminal luminal luminal basal basal basal basal basal luminal basal basal luminal
luminal basal basal luminal basal luminal basal basal basal basal basal luminal luminal basal basal
basal luminal basal basal luminal basal basal basal basal basal luminal basal basal luminal
luminal basal basal luminal luminal luminal luminal luminal luminal luminal luminal basal basal
luminal luminal luminal luminal luminal basal luminal luminal luminal luminal luminal basal basal
basal basal luminal basal basal basal basal basal basal basal luminal luminal luminal basal basal
luminal basal basal luminal luminal luminal luminal luminal luminal basal luminal basal luminal
luminal luminal luminal basal basal luminal basal basal basal basal luminal luminal basal basal
basal luminal luminal basal basal basal basal basal basal basal basal luminal basal basal
luminal basal basal basal basal basal luminal basal luminal basal luminal luminal luminal luminal
basal basal basal basal luminal basal luminal basal luminal luminal luminal luminal luminal
luminal luminal luminal luminal basal basal basal luminal luminal basal basal basal luminal
luminal basal basal luminal luminal basal basal basal basal basal basal basal basal basal

```

Figure 5 - An example of phenotype label file (cls)

Note: the figure shows a fragment of the file that was obtained after analysis (clinical and survival) of various tissue samples. As a result, all samples were divided into either the luminal or basal subtype, which was subsequently used in the calculations.

The following parameters were set for the algorithm in GSEA: permutations=1000, permutation type=gene_set, enrichment statistic=weighted, metric for ranking genes=Signal2Noise (Figure 6).

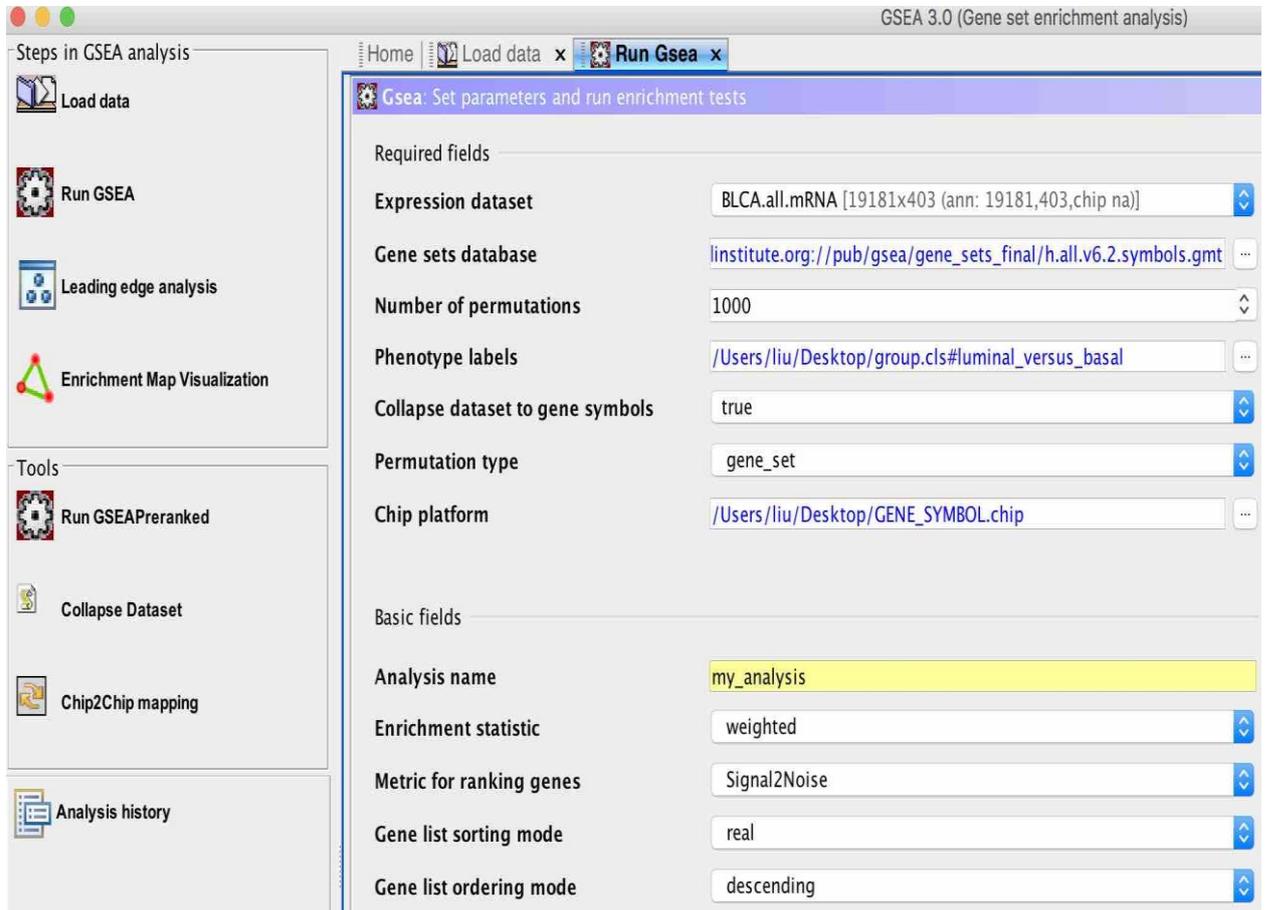


Figure 6 - Screenshot showing parameter settings for GSEA in this study
Note: setting parameters in the GSEA program according to the necessary criteria of repetition, sampling, enrichment analysis method, etc.

2.6 - Screening of representative genes between subtypes using machine learning approaches

Many machine learning techniques are widely used in many fields of biology, such as prediction of nucleosome-enriched regions and depleted regions, classification of gene families, prediction of hepatotoxicity, prediction of RNA methylation, prediction and classification of oncogenes [31, 54, 57, 117, 126, 151, 195]. As suggested in previous studies, two machine learning models - Random

Forest (RF) and eXtreme Gradient Boosting (XGBoost or XG) are powerful models for classifying cancer [13, 38, 80].

For machine learning algorithms RF and XGBoost, mRNA data sets (3,000 mRNA and 403 tumor samples), a miRNA data set (300 miRNA and 403 tumor samples) and a lncRNA data set (3,000 lncRNA and 403 tumor samples) were used. The work was carried out using the corresponding R packages “randomForest” and “XGBoost” [29, 108]. Randomly, 403 MIBC samples were divided into 60% for the training set and 40% of the test data set.

Fine-tuning the predictive model of machine learning is an important first step to improve the accuracy of classification results. The quality of the classification was evaluated by receiver operating characteristic (ROC) analysis. ROC-curve is a graph that allows you to assess the quality of binary classification, displays the ratio between the proportion of objects from the total number of carriers of the trait, correctly classified as carriers of the trait (true positive rate, TPR) and the proportion of objects from the total number of objects that do not carry a sign, mistakenly classified as bearing a sign (false-positive rate, FPR). A quantitative interpretation of ROC is given by the AUC (area under the ROC curve) indicator. The larger the area under the curve, the better the classifier, while the value $AUC = 0.5$ demonstrates the unsuitability of the selected classification method (corresponds to random divination). A value of less than 0.5 indicates that the classifier works worse; if the value of more than 0.5 indicates that the classifier will work better.

Then we used trained models to predict the classification of MIBC subtypes based on test data sets, and AUC was used to evaluate their performance. After predicting the subtypes using trained models, we wanted to find out which genes are more representative in the model. In the RF model, we used *Mean Decrease Accuracy* > 0 as the limit value for screening representative genes. In the XG model, we used *Gain* > 0 as the cut-off value for screening representative genes.

2.7 - Determination of differences in gene expression between tumor and normal bladder tissue samples

To determine the differences in gene expression in the data set of patients with muscle-invasive bladder cancer (MIBC), the standard Differentially expressed gene (DEG) method was performed by using the Ballgown R package. Differences in gene expression were detected between the tumor and normal bladder tissue samples. As parameters, the frequency of FDR false discoveries was taken into account with $p\text{-value} < 0.05$ and $|\log_2\text{fold change}| > 0.57$ as cut-off value [19]. False Discovery Frequency (FDR) is a method for conceptualizing the level of Type I errors in testing null hypotheses in multiple comparisons. *Log₂fold change* is the cut-off value, which is important for calculating the difference between the expression levels.

Then, the genes detected by DEG were compared with those representative genes differing in expression that were detected by machine learning RF and XG, thus, it was possible to overlap data to identify genes that in patients with MIBC always have a difference in expression, which was used for further analysis.

Then the differentially expressed genes determined by traditional method were compared with representative genes that differ in the expression of two subtypes identified by Random Forest and XGBoost. Thus, it is possible to identify overlapped genes that always have a difference in expression between two subtypes and relatively control samples. overlapped genes were used for construction of mRNA-miRNA-lncRNA interaction network.

2.8 - Construction of a subtype-dependent interaction network of mRNA-miRNA-lncRNA for muscle-invasive bladder cancer

Pearson's correlation coefficient (PCC), also called Pearson's r , is a statistical formula that measures the strength of the correlation between two variables x and y . Pearson correlation coefficient was calculated by the formula

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}} \quad (1)$$

where n is the number of samples (patients), and x and y are the expression values of gene x and y . The coefficient r can take values between $+1$ and -1 , where 1 is the total positive linear correlation, 0 is the absence of linear correlation, and -1 is the total negative linear correlation. PCC analysis is performed for genes with different levels of expression. For the assessment, pairs of lncRNA and miRNA, miRNA and mRNA, lncRNA and mRNA with $|r| \geq 0.4$ and $p < 0.05$ were considered to be co-expressed gene pairs. If both elements in a co-expressed lncRNA-mRNA pair are simultaneously co-expressed with a certain miRNA, it is defined as a miRNA-dependent lncRNA-mRNA co-expressed interaction.

The miRWalk2.0 is an integration of six widely used databases (miRWalk, miRanda, miRDB, miRNAMap, RNA22, and Targetscan), which provides the largest collection of predicted and experimentally confirmed miRNA target interaction [127].

Our inferred co-expressed interactions including mRNA-miRNA and lncRNA-miRNA interactions were compared to those derived from miRWalk2.0. An mRNA is considered to be a true target of miRNA if their interaction occurs in at least four databases among miRWalk, miRanda, miRDB, miRNAMap, RNA22, and Targetscan. A lncRNA is considered to be a true target of miRNA if their interaction is supported in at least one database among miRWalk, miRanda, and Targetscan.

2.9 - Assessment of gene network density and biological distance for common variable immunodeficiency genes and other primary immunodeficiency genes

These protein-protein interactions (PPI) and for PID CVID genes were obtained from the data base STRING (<https://string-db.org/>). STRING database

does not contain information on the mechanism of protein-protein interactions, but also about what time of the cell cycle can take place this reaction, it depends on the external environment and when it has tissue specifically. However STRING contains information about all possible protein-protein interactions in the body, including the information to predict with a certain authenticity that makes STRING complete resource of protein-protein interactions that are available today, and is especially useful for finding information about the proteins have not been studied experimentally[175]. At the base of the sample we did on the criterion of Homo Sapiens and set the minimum required rate of interaction 0.4. PPI data network was visualized using Cytoscape software (version 3.5.1) [52]. Network density ($D_{network}$) is the most widely used concept in the regulation of genes and studies networks PPI networks and can be used to determine whether or not a dense network. Network density ($D_{network}$) defined by the formula [82]

$$D_{network} = \frac{\sum_{i=1}^n \sum_{j \neq i} a_{ij}}{n(n-1)} \quad (2)$$

where a_{ij} is pairwise adjacency, $\sum_{j \neq i} a_{ij}$ represents the connectivity (the unweighted network connectivity equals the number of genes that are directly linked to gene i) of the i th gene, and n is the number of genes in the network. Note that $a_{ij} = 1$ if the interaction of gene i and gene j occurs in the STRING database, whereas $a_{ij} = 0$ otherwise. The PPI data of a CVID group (39 literature-based CVID genes) and ten random groups (each group consists of 39 PID genes) were respectively converted into the symmetric adjacency matrix (a_{ij} , $i, j = 1, \dots, n$) using the "igraph" R package [49]. Network density was used to compare their network cohesion or tightness. The greater the network density is of a group, the tighter the interaction of the genes in the group.

The concept of biological distance ($B_{i,j}$) was first introduced by Itan Y. et al., in 2013. By biological distance, researchers studying the functional relationships of genes in a network of genomic interactions do not mean the actual distance

between genes in a DNA molecule or on a chromosome, but functional proximity between pairs genes or within a group of genes [182]. Using the value of the biological distance, Itan Y. et al. showed that PID genes, as a rule, are located in the center of the human genomic network and form several closely interconnected subgroups according to various biological mechanisms [103]. The biological distance ($B_{i,j}$) is determined by the formula

$$B_{i,j} = \begin{cases} \frac{C}{S_{i,j}} & \text{if } C = 1 \\ \frac{C}{S_{i,1} + S_{1,2} + S_{2,3} + \dots + S_{C-2,C-1} + S_{C-1,j}} & \text{if } C > 1 \end{cases} \quad (4)$$

where $S_{i,j}$ is the combined score between gene i and gene j provided by the STRING database, and C is the number of direct connections between gene i and gene j . The smaller the biological distance is of a group, the closer the biological interrelatedness between genes in the group. We calculated the biological distance of a CVID group (39 CVID genes) and two random groups (each group consists of 39 PID genes) with the human gene connectome (HGC) Python package provided by Itan et al [182].

2.10 -Prediction total variable immunodeficiency candidate genes

The three following steps were taken to predict the CVID candidate genes:

I. Pearson's correlation analysis (PCC) was performed on the expression values of 39 CVID genes and each protein-coding gene (or, that is, candidate gene) based on the GSE51406 and GSE72625 datasets. Using $|r| > 0.9$ and $p < 0.05$ as cut-off values, the candidate genes were acquired from each of the two datasets. The overlapping candidate genes obtained from the two datasets were used for the subsequent analysis.

II. The PPI data for all human protein-coding genes were obtained from Cheng et al [136], including 217160 interactions provided by eleven databases (e.g., BioGRID [173], HI-II-14_Net [10], HPRD [83], Instruct [91], InnateDB [90],

IntAct [79], MINT [123], PINA [148], Signalink2.0 [172], KinomeNetworkX [158], and PhosphositePlus [146]). A candidate gene was then retained if the interaction between the CVID gene and candidate gene obtained in the previous step occurred in the PPI data.

III. Kyoto Gene and Genomic Encyclopedia (KEGG) analysis was performed using the R package "clusterProfiler" for CVID genes to estimate their biological function enrichment [36]. KEGG analysis was then performed on the remaining candidate genes, and a gene was defined as a true CVID candidate gene if the candidate gene was enriched in the same pathway as the CVID gene.

2.11 - Identification of new candidate genes for common variable immunodeficiency

In order to determine whether our method is suitable for predicting CVID candidate genes, we calculated the biological distances ($B_{i,j}$) of the predicted CVID candidate genes and compared them with 39 of the known CVID genes.

Subsequently, "functional genomic alignment" (FGA) and phylogenetic cluster analysis were carried out. These steps were performed using the APE package, available in R, to evaluate the biological correlation between the CVID candidate genes and the known CVID genes [143, 182]. In particular, we first created a biological distance matrix between the CVID genes and the CVID candidate genes, and then we applied the neighbor-joining algorithm (function nj) to generate a phylogenetic fan-shaped tree showing hierarchical clustering of the CVID candidate genes and the known CVID genes. If the CVID candidate genes were evenly distributed across the entire range of known CVID genes, this means that these CVID candidate genes were closely related to the known CVID genes. If the CVID candidate and the known CVID genes were divided into two or more groups, this means the opposite.

2.12 – Visualization of molecular interaction networks using Cytoscape software

Cytoscape is an open source bioinformatics software platform designed to visualize molecular interaction networks and integrate with gene expression profiles and other data. Cytoscape software (version 3.5.1) was used to visualize the miRNA-dependent lncRNA-mRNA network of patients with muscle-invasive bladder cancer and protein-protein interaction network of primary immunodeficiency genes (Figure 7).

2.13 – Whole genome sequencing and search for candidate genes for the patient with Hennekam syndrome

The sequencing results were aligned with the standard sequence of the hg38 human genome using Burrows-Wheeler Aligner (BWA) software [107]. Then “SAMtools” program was used to change the order, sort and convert SAM files [185].

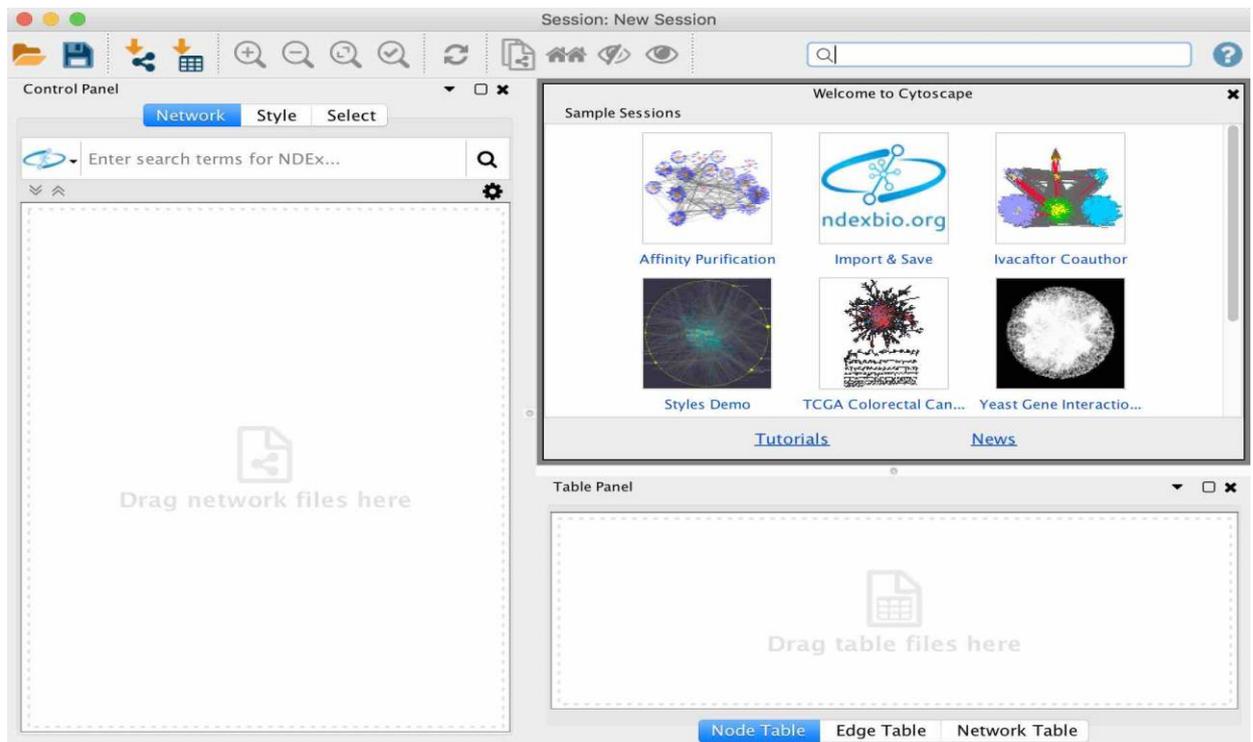


Figure 7 - Screenshot showing the Cytoscape software used in this study

Single nucleotide variants (SNVs) and insertion / deletion (insertion) variant calls were identified with The Genome Analysis Toolkit version 4.1.2.0 (GATK4, <http://www.broadinstitute.org/gatk/>) [180].

Only variants in the exonic region with a read coverage (or depth) of $>10\times$ and a minimum mapping quality score of 30 were maintained using the VCFtools software to reduce the low-quality calls resulting from mapping errors [186]. We then discarded all synonymous SNVs, non-frameshift insertions and deletions (Indels), and variants whose exonic function labeled as “NA” or “unknown”. Candidate SNVs and Indels obtained by previous steps were further filtered for SNVs and Indels in genes related to primary immunodeficiency (PID).

A candidate SNV (or Indel) was then retained and supposed to be rare or non-common if the minor allele frequency (MAF) of the SNV (or Indel) was less than 0.01 in all of Exome Aggregation Consortium (ExAC), 1000 Genomes (1000g), and The Genome Aggregation Database (gnomAD).

A SNVs (or Indels) were further considered to be a deleterious one if the SNVs (or Indels) were nominated as deleterious in at least one model among Functional Analysis through Hidden Markov Models (FATHMM), Protein Variation Effect Analyzer (PROVEAN), and Combined Annotation Dependent Depletion (CADD). FATHMM and PROVEAN were performed with the ANNOVAR software [193], and CADD was performed using an online server (<https://cadd.gs.washington.edu/snv>, version: GRCh38-v1.5).

All potential pathogenic SNVs and Indels were manually inspected using the Single Nucleotide Polymorphism Database (dbSNP) (<https://www.ncbi.nlm.nih.gov/snp/>) and Integrative Genome Viewer (IGV) software version 2.4.5. If two MAF of an SNV (or Indel) respectively obtained from ANNOVAR and dbSNP were ambiguous, the MAF obtained from dbSNP were considered as a true one. A candidate mutation was defined as a true one if a mutation identified by GATK4 was found in IGV.

CHAPTER 3 - USING MIR-141-5P, MIR-141-3P, MIR-200C-3P, AC010326.3, AC073335.2, MIR100HG, CLIC4, GATA3, PALLD AS PROGNOSTIC BIOMARKERS IN PATIENTS WITH MUSCLE-INVASIVE BLADDER CANCER

3.1 - Impact of infiltrating immune cells on the progression and survival in patients with muscle-invasive bladder cancer

22 types of tumor infiltrating immune cell levels were evaluated by CIBERSORT program. We also evaluated whether the immune cell levels are correlated with overall survival in patients with muscle-invasive bladder cancer (MIBC). We used the median as the cut-off to separate the patient samples into the high and low groups and compared the survival time of the two groups using log-rank test. Then Wilcoxon test was used to compare differences in immune cells between the two groups. The log-rank test and Wilcoxon t-test showed that, among all of the immunological parameters, tumor-infiltrated macrophage levels (M2 subtype) are significantly correlated with patient survival, grade and pathological stage ($P < 0.01$) (Figure 8).

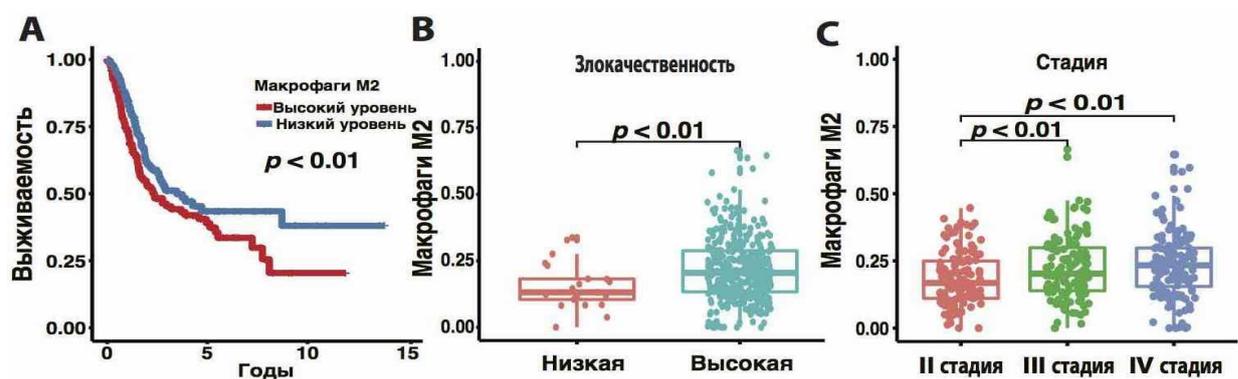


Figure 8 - Charts Kaplan-Meier curve for M2-macrophages (A) and the level M2 macrophages from patients with muscle invasive bladder cancer (B, C)

This logically fits into the pathogenesis of the development of the oncological process, in which the severity of the process correlates with greater activity of tumor macrophages. The results of our research allow recommending M2 macrophages as markers for the MIBC prediction.

3.2 - Clustering analysis of RNA expression data of patients with muscle-invasive bladder cancer

The following steps were performed to classify the molecular subtypes of MIBC.

To determine the appropriate clusters, the Consensus Clustering analysis was performed for the mRNA, miRNA and lncRNA datasets using the R programming package “ConsensusClusterPlus”, in which the parameters were configured as follows: $reps=1000$, $innerLinkage=complete$, $clusterAlg=hc$, $k=6$, $distance=pearson$ [198]. As an example, the clusterization data set is shown in Table 5. In the columns, in addition to the sample identifier, the cluster group to which the sample belonged when divided into two, three or more different clusters, was indicated.

Table 5 - The clusters generated using consensus cluster analysis

Sample\Cluster	Cluster=2	Cluster=3	Cluster=6
TCGA-2F-A9KO	2	2	1
TCGA-2F-A9KP	2	3	2
.....
.....
TCGA-4Z-AA7Y	2	1	3
TCGA-4Z-AA80	2	2	6

Average silhouette width (ASW) estimation algorithms were performed to estimate optimal clusters (K) using the R "fpc" package [170]. Usually, we can

deduce the optimal K based on the highest value of ASW [78]. The optimal K proposed by ASW in the mRNA, miRNA and lncRNA datasets is 2 (Figure 9).

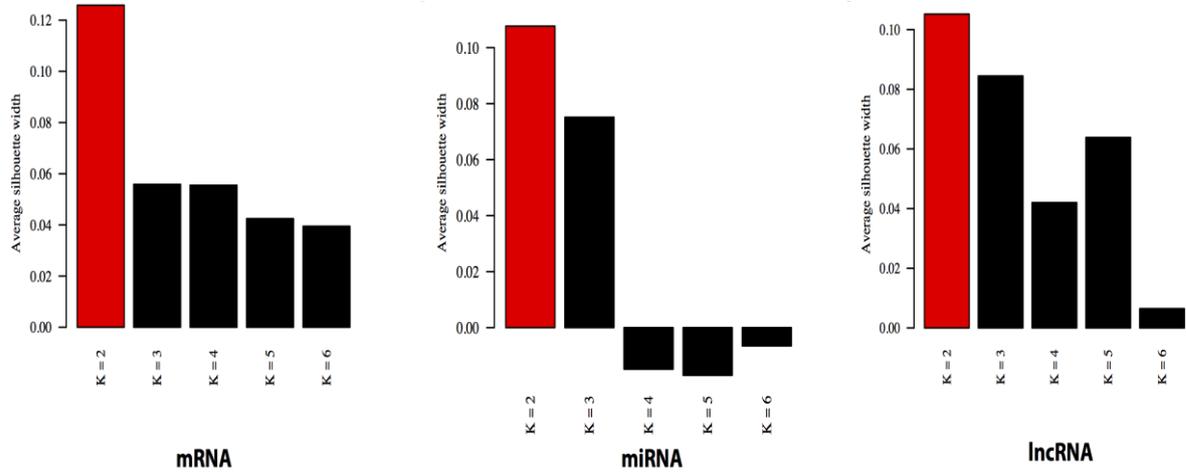


Figure 9 - Columns show the result of the evaluation by the ASW algorithm.

ASW allows you to derive the optimal number of clusters in a given dataset

Note: The column $K = 2$ is shown in red, which means that for the corresponding data sets, the optimal number of clusters is two.

Cluster of Cluster (CoC) analysis uses the cluster assignments obtained from the analysis of individual molecular datasets to obtain a single solution, thus combining the results of the analysis of mRNA, miRNA, and lncRNA datasets. First, we prepared a binary data set based on the results of clustering when $K=2$ (Table 6).

Table 6 - Binary dataset used for CoC analysis

Sample\Dataset	mRNA1	mRNA2	miRNA1	miRNA2	lncRNA1	lncRNA2
TCGA-2F-A9KO	0	1	0	1	1	0
TCGA-2F-A9KP	0	1	0	1	0	1
TCGA-2F-A9KQ	1	0	1	0	0	1
.....		
.....		
.....		
TCGA-4Z-AA7Y	0	1	1	0	1	0
TCGA-4Z-AA80	1	0	1	0	1	0

Note: 0 - number of the first cluster, 1 - second

We then performed consensus clustering for the binary dataset with $K=2$, to generate the final cluster assignment (Table 7).

Table 7 – Final clusters obtained by CoC analysis

Sample	Cluster
TCGA-2F-A9KO	1
TCGA-2F-A9KP	2
TCGA-2F-A9KQ	1
.....
.....
.....
.....
.....
TCGA-4Z-AA7Y	2
TCGA-4Z-AA80	1

As a result of cluster analysis, 403 MIBC samples were divided into two stable subtypes. A hierarchically clustered heat map of subtype-1 and subtype-2 is shown in Figure 10A.

Kaplan-Meier (K-M) survival analysis showed that the overall 5-year survival rate for subtype-1 (red curve) was 55%, and for subtype-2 (blue curve) 30%, which indicates a significant difference in the clinical prognosis (log-rank test, $p < 0.05$, Figure 10B).

The result in Figure 10C shows a detailed map, where each of the 403 samples was divided into two subclasses, with markers and clinical indicators for the basal and luminal subtypes. 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.

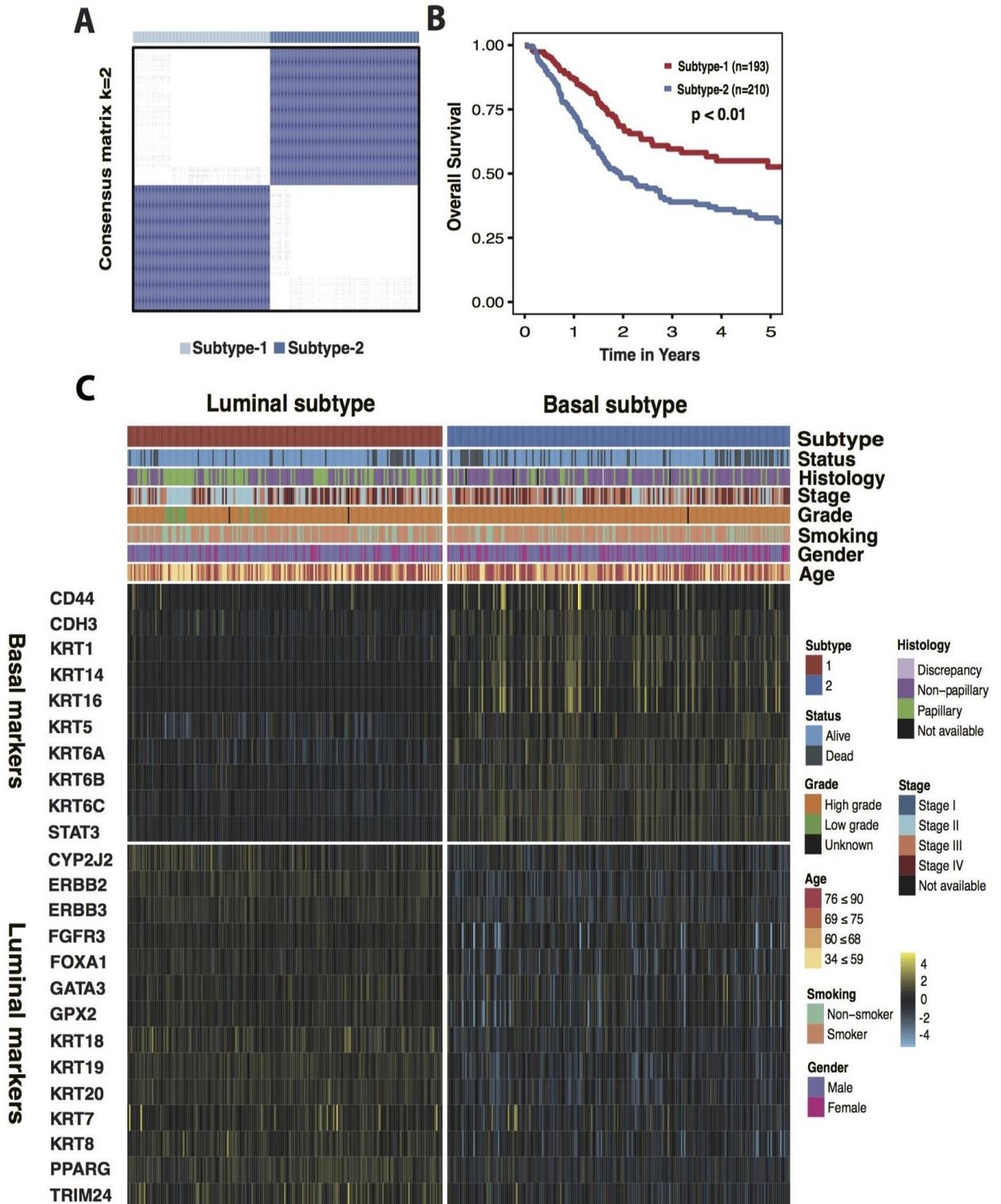


Figure 10 - Classification of MIBC samples into subtype-1 and subtype-2

Note: (A) A hierarchically clustered heat map for the CoC dataset (K=2). (B) K-M plot for overall 5-year survival of subtype-1 and subtype-2 (subtype-1=193, subtype-2=210, log-rank test, $p < 0.05$). (C) The heat map shows the expression profiles of basal (up) and luminal (bottom) biomarkers in subtype-1 (left) and subtype-2 (right). Covariate annotation tracks and corresponding colors show some clinical features. Yellow and turquoise colors correspond to high and low relative levels of expression. The original expression value was log₂-transformed.

The Pearson test (χ^2) was used to compare clinical performance between the two subtypes. Histology, cancer stage, malignancy grade, and other criteria significantly differed between the two subtypes; no significant differences were shown for gender ($p=0.06$) (Table 8).

Subtype-1 and subtype-2 resemble the luminal and basal subtype, respectively, both in survival curves and in biomarkers and clinical parameters [85]. Thus, subtype-1 and subtype-2 were redefined as luminal and basal subtypes, respectively.

Table 8 - Relationship between clinical characteristics and MIBC subtypes (χ^2 test).

Variables	Cohort	<i>p</i>-value
Subtype		
basal	210	
luminal	193	
Stage		<0.001
I	2	
II	129	
III	138	
IV	132	
Unavailable	2	
Age		0.4498
34≤59	87	
60≤68	112	
69≤75	95	
76≤90	109	
Gender		0.0612
Man	297	
Female	106	
Smoking history		0.3345
Smoker	294	
Non-smoker	109	
Status		0.0173
Alive	298	
Dead	105	
Histology		<0.001
Pillary	128	
Non-pillary	270	
Unavailable	5	
Grade		<0.001
High	380	
Low	20	
Unavailable	3	

The relationship between the subtypes and immune check-point markers, which are important for determining the effectiveness of immunotherapy, has been investigated. It was found that expression patterns of immune check-point molecules (e.g. PD-1, PD-L1, CTLA-4, HAVCR-2, and LAG-3) distinguish between basal and luminal subtypes of bladder cancer, and basal tumors exhibit higher expression levels of immune response check-point molecules than the luminal subtype (Figure 11).

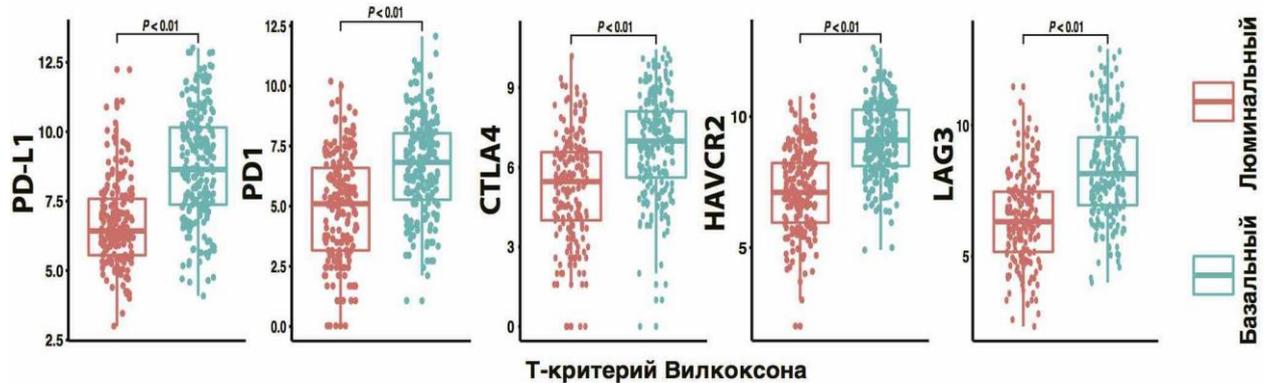


Figure 11 - Expression levels of immune checkpoint markers in the basal and luminal subtypes

Note: yellow and turquoise colors correspond to high and low relative expression values, respectively. The original expression value was log₂-transformed.

3.3 - Functional enrichment analysis of data of patients with muscle-invasive bladder cancer

The gene set enrichment analysis (GSEA) showed that highly expressed signaling pathways in the luminal subtype are mainly involved in metabolism (e.g., oxidative phosphorylation, cytochrome P450, and fatty acid metabolism) (Table 9). At the same time, highly expressed pathways in the basal subtype are mainly associated with processes in the immune system (for example, extracellular structure organization, allograft rejection, mTORC1 signaling, and TNF- α signaling via NF- κ B), metastasis (focal adhesion and cytokine interaction -

cytokine receptor) and the epithelial-mesenchymal transition ($|NES| > 1$ and FDR $q\text{-val} < 0.25$) (Table 10).

Table 9 - Top-ranked terms of GO-BP, KEGG and Hallmark gene sets for the luminal subtype.

Gene Set Name	Size	NES	FDR q Value
GO-BP			
GO monocarboxylic acid catabolic process	95	2.5151	0
GO oxidative phosphorylation	73	2.4333	0
GO fatty acid catabolic process	80	2.4284	0
GO fatty acid beta oxidation	51	2.3211	0
GO electron transport chain	78	2.3116	0
GO organic acid catabolic process	92	2.2781	2.26E-04
GO mitochondrial respiratory chain complex assembly	42	2.1788	0.0014
GO lipid oxidation	63	2.1586	0.0015
GO mitochondrial respiratory chain complex i biogenesis	199	2.1449	0.0018
GO establishment of protein localization to endoplasmic reticulum	70	2.1285	0.0022
KEGG			
KEGG ribosome	87	2.3289	0
KEGG alpha linolenic acid metabolism	19	2.077	5.29E-04
KEGG metabolism of xenobiotics by cytochrome p450	68	2.0424	5.36E-04
KEGG valine leucine and isoleucine degradation	44	1.9767	0.00137
KEGG drug metabolism cytochrome p450	70	1.9727	0.0012
KEGG oxidative phosphorylation	116	1.9372	0.00209
KEGG peroxisome	78	1.932	0.00214
KEGG fatty acid metabolism	42	1.9184	0.00229
KEGG retinol metabolism	63	1.8697	0.00391
KEGG linoleic acid metabolism	29	1.8393	0.00482
Hallmark gene sets			
Hallmark oxidative phosphorylation	198	1.5145	0.0583
Hallmark bile acid metabolism	112	1.411	0.07668
Hallmark peroxisome	103	1.4095	0.05174
Hallmark adipogenesis	191	1.3794	0.05125
Hallmark fatty acid metabolism	156	1.2596	0.11892

Note: NES=normalized enrichment score; GO-BP= Gene Ontology Biological Process; KEGG= Kyoto Encyclopedia of Genes and Genomes. Size is the number of genes in the gene set. A positive NES means that genes over-represented in the gene set are upregulated in luminal subtypes.

Table 10 - Top-ranked categories of GO-BP, KEGG and Hallmark gene sets for the basal subtype

Gene Set Name	Size	NES	FDR q Value
GO-BP			
GO extracellular structure organization	297	-2.8256	0
GO antigen processing and presentation of exogenous peptide antigen via mhc class i	65	-2.7258	0
GO antigen processing and presentation	206	-2.6334	0
GO antigen processing and presentation of peptide antigen	170	-2.6246	0
GO antigen processing and presentation of peptide antigen via mhc class i	90	-2.6134	0
GO chondroitin sulfate biosynthetic process	25	-2.6008	0
GO collagen fibril organization	36	-2.5958	0
GO regulation of innate immune response	349	-2.5825	0
GO positive regulation of defense response	360	-2.5802	0
GO cytokine mediated signaling pathway	440	-2.5675	0
KEGG			
KEGG focal adhesion	197	-2.6862	0
KEGG cytokine cytokine receptor interaction	257	-2.5127	0
KEGG ecm receptor interaction	84	-2.512	0
KEGG proteasome	43	-2.4802	0
KEGG leishmania infection	69	-2.4718	0
KEGG viral myocarditis	68	-2.4178	0
KEGG hematopoietic cell lineage	85	-2.4134	0
KEGG regulation of actin cytoskeleton	211	-2.3911	0
KEGG allograft rejection	35	-2.3902	0
KEGG autoimmune thyroid disease	50	-2.3778	0
Hallmark gene sets			
Hallmark epithelial-mesenchymal transition	197	-3.2473	0
Hallmark inflammatory response	197	-3.019	0
Hallmark interferon gamma response	197	-2.9964	0
Hallmark interferon alpha response	94	-2.9491	0
Hallmark allograft rejection	199	-2.901	0
Hallmark G2M checkpoint	194	-2.6389	0
Hallmark E2F targets	196	-2.6177	0
Hallmark TNF-a signaling via NF-kB	198	-2.5512	0
Hallmark complement	195	-2.5475	0
Hallmark mTORC1 signaling	198	-2.441	0

Note: All abbreviations are the same as in Table 9. A negative NES value indicates that genes over-represented in the gene set are upregulated in the basal subtype.

3.4 - Search for differentially expressed genes using DEGs method and machine learning

Differential expression analysis showed that 4167 mRNAs (2314 with increased expression and 1853 with reduced), 208 miRNA (148 with increased expression and 60 with reduced), and 2488 lncRNAs (1402 with increased expression and 1086 with reduced) were differentially expressed between tumor and normal bladder tissue samples (Figure 12).

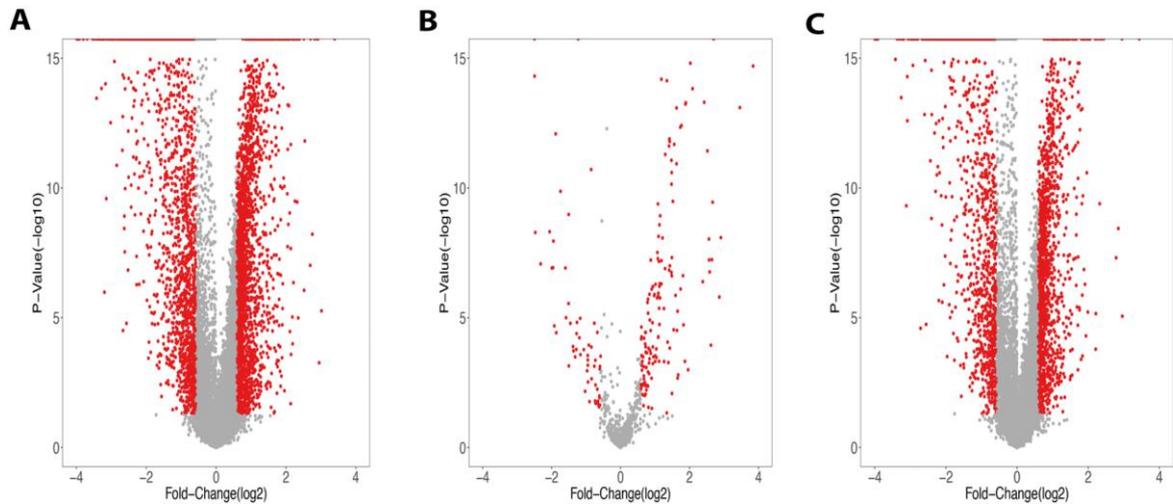


Figure 12 - The software package “Ballgown” identified DEGs between MIBC patients and healthy controls, including a 4167 mRNA (A), 208 miRNA (B) and 2488 lncRNA (C)

Note: corrected value $p < 0.05$ and $|\log_2 \text{fold change}| > 0.57$.

Machine learning is the most effective method for identifying critical genes and classifying them into subtypes depending on the expression patterns of mRNA, miRNA, and lncRNA. The first step in machine learning is to train your existing machine learning models. We used two models, namely Random Forest (RF) and XGBoost (XG).

For RF, we adjusted the optimal “mtry” parameter according to the large area under the ROC curve (AUC), and then adjusted the parameter “ntree” according to the small “Obb Error Rates”. We have configured the “iter” (number of iterations) parameter of the XGBoost model based on the 10-fold cross-validation method. In addition, the following important parameters were used for XGBoost: $\text{gamma}=1$, $\text{min_child_weight}=1$, $\text{max_depth}=14$, $\text{nrounds}=2000$.

In the mRNA dataset, we obtained $mtry=9$ and $ntruel=300$ as optimal parameters for the RF model and $iter=2000$ as optimal parameters for the XGBoost model (Figure 13).

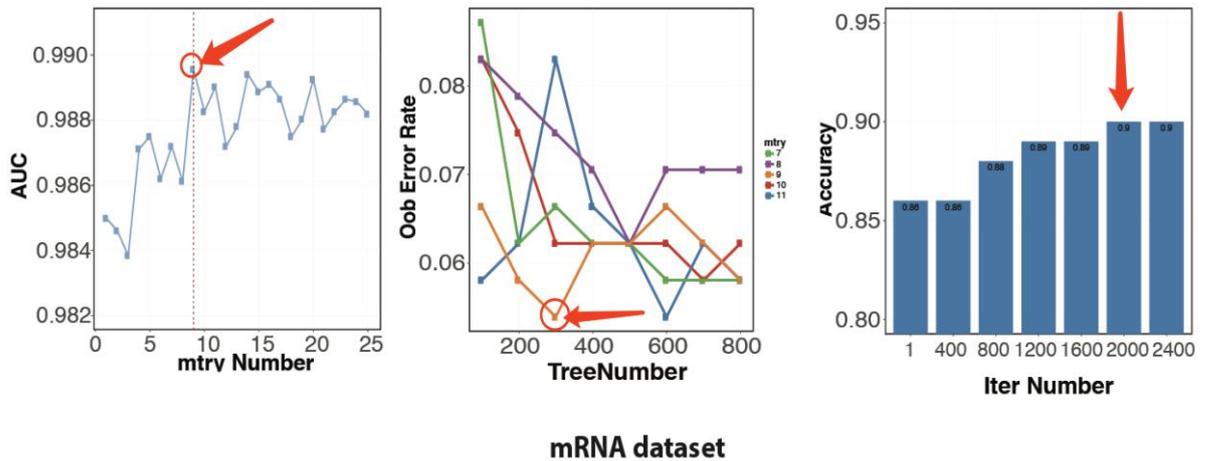


Figure 13 - Selection of parameters for RF and XGBoost in mRNA dataset

In the microRNA data set, we obtained $mtry=7$, $ntruel=300$ as the optimal parameters of the RF model and $iter=2000$ as the optimal parameter of the XGBoost model (Figure 14).

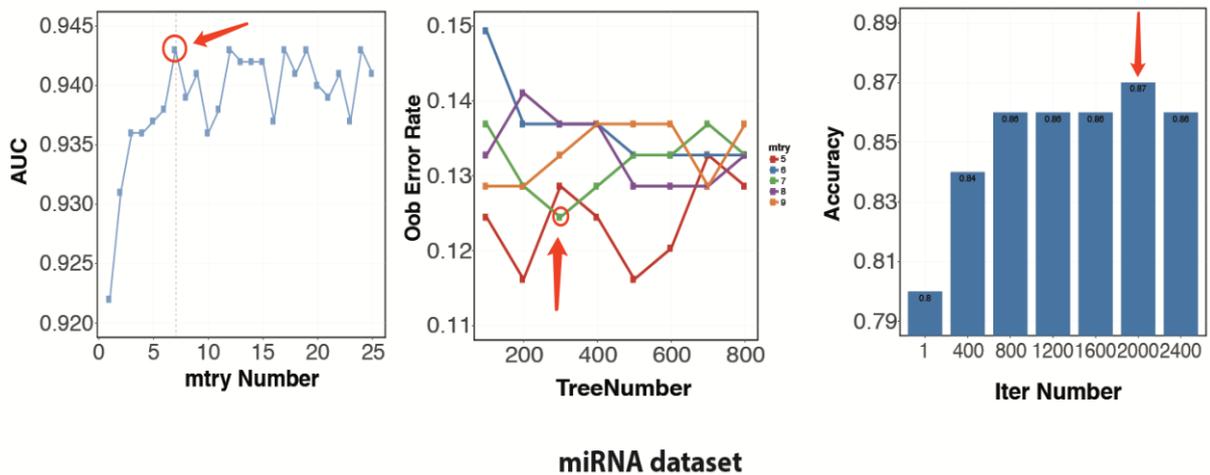


Figure 14 - Selection of parameters for RF and XGBoost in the miRNA dataset

In the lncRNA dataset, we obtained $mtry=16$, $ntruel=200$ as optimal parameters for the RF model and $iter=1600$ as the optimal parameter for the XGBoost model (Figure 15).

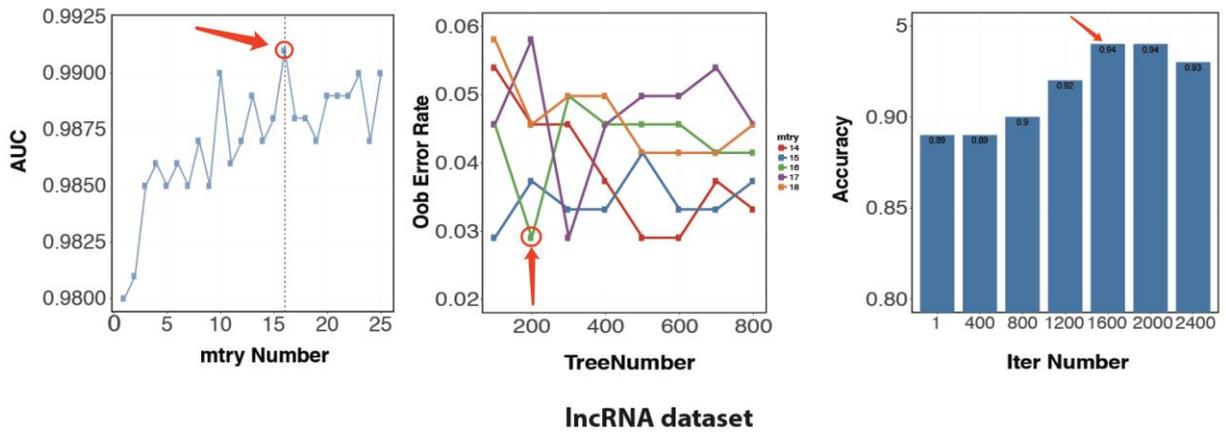


Figure 15 - Selection of parameters for RF and XGBoost in the lncRNA dataset

The area under the ROC curve showed that the RF and XGBoost models had AUC values of more than 90% in three data sets, which indicates the high quality of MIBC sample classification (Figure 16).

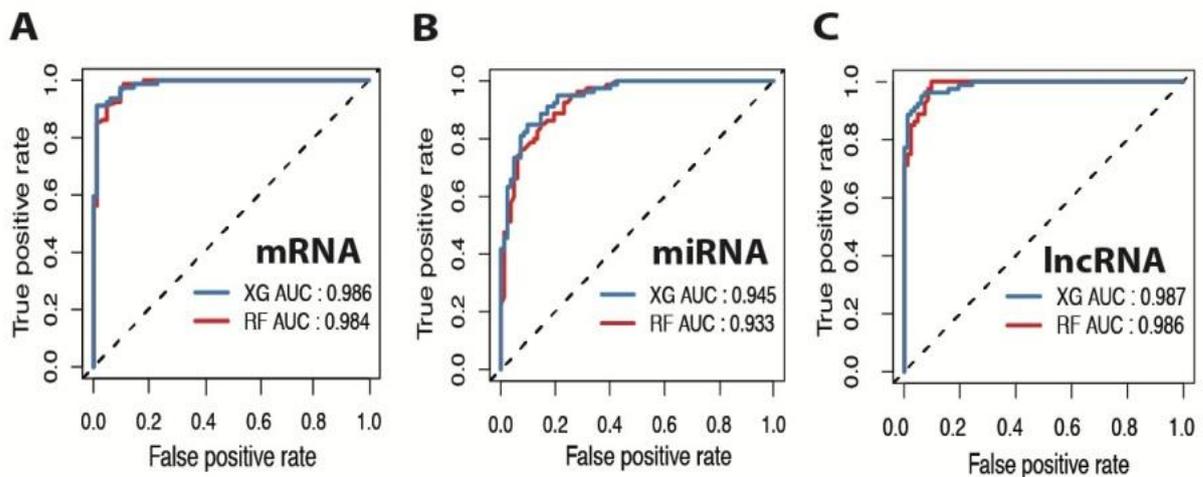


Figure 16 - ROC-curves for XGBoost and RF classifiers in the (A) mRNA, (B) miRNA, and (C) lncRNA datasets

The “Mean Decrease Accuracy” of RF and “Gain” of XGBoost are used to identify representative genes. The overlap between the genes in DEGs and representative genes at the output of the two types of machine learning included 57 lncRNAs, 120 miRNAs and 278 mRNAs, which were used for subsequent analysis (Figure 17).

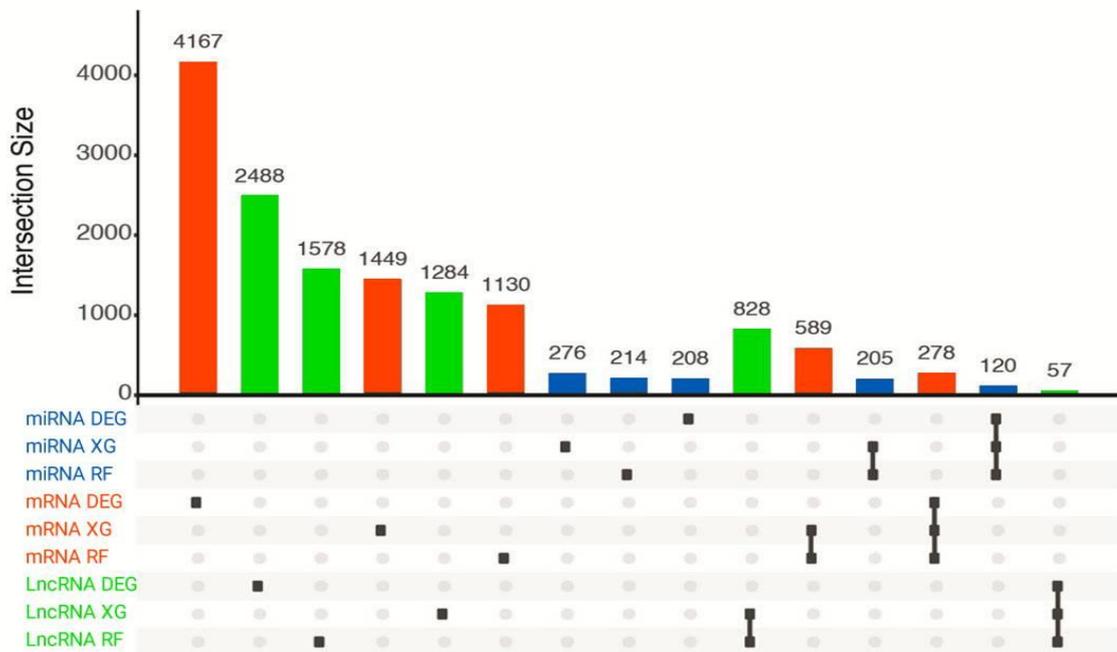


Figure 17 - UpSet diagram for overlapping data between DEG and MIBC-specific genes obtained using XG (XGBoost) and RF machine learning methods

3.5 - Construction of a subtype-dependent mRNA-miRNA-lncRNA network for muscle-invasive bladder cancer

Pearson's correlation coefficient ($|r| \geq 0.4$ and $p < 0.05$) and Cytoscape software allowed to construct a miRNA-dependent mRNA-lncRNA co-expression network, which consisted of 90 mRNA, 22 miRNA and 14 lncRNA (Figure 18A). Using the miRWalk database, we found 4 overlapping interdependencies between mRNA, miRNA, and lncRNA. In particular, the two co-expressed lncRNA-mRNA pairs, namely AC010326.3-GATA3 and AC073335.2-GATA3, are positively upregulated by miR-141-3p. Other lncRNA-mRNA pairs, such as MIR100HG-CLIC4 and MIR100HG-PALLD, are negatively regulated by miR-200c-3p and miR-141-5p, respectively (Figure 18B). A “heat map” type graph showed that these nine genes differ significantly in their expression between the basal and luminal subtypes (Figure 18C).

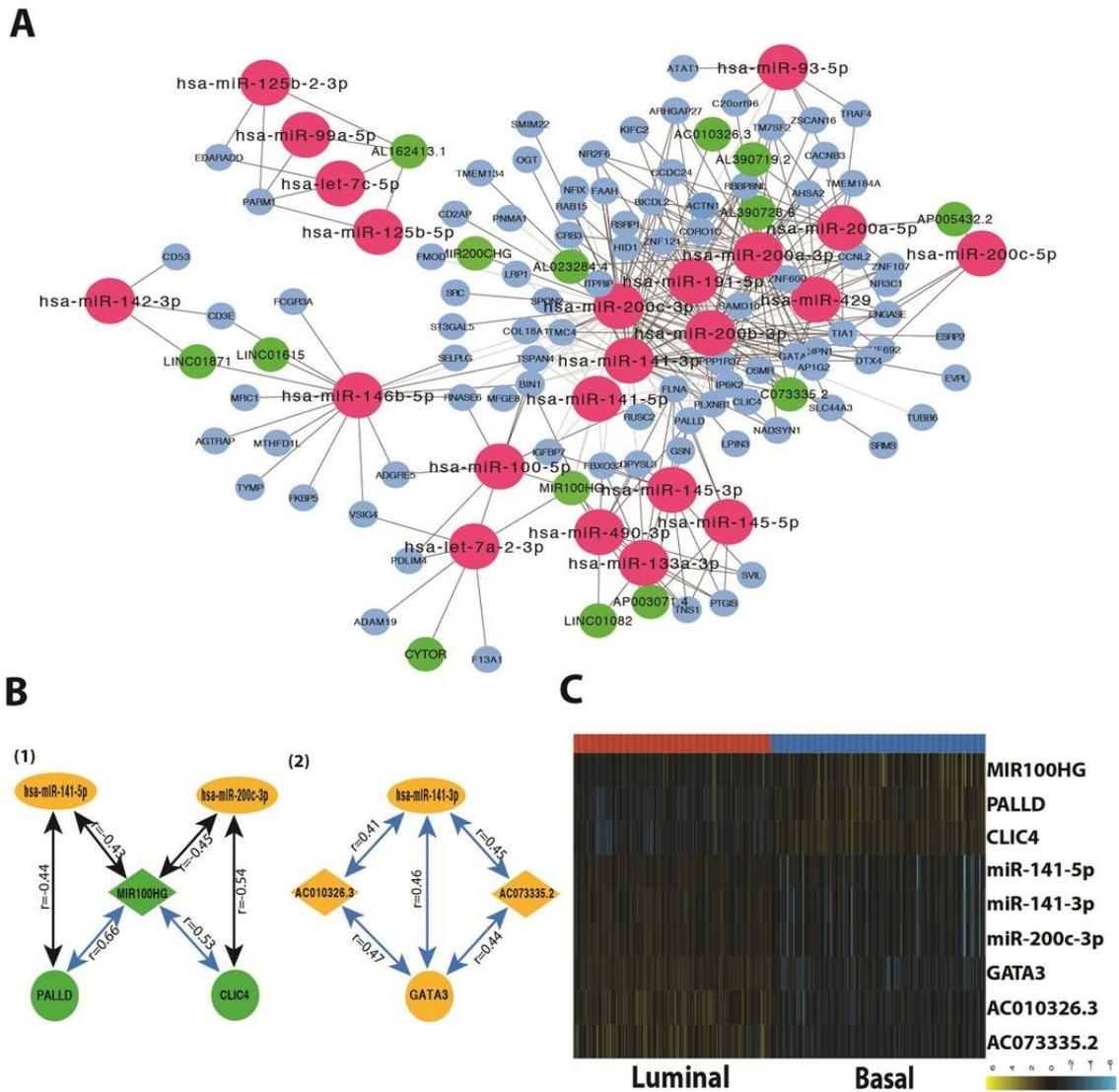


Figure 18 - Characteristic of the miRNA-dependent mRNA-lncRNA cross-relationship of cancer subtypes

Note: (A) a co-expression network for miRNA-mediated mRNA-lncRNA interactions. Red-colored nodes are miRNA, green nodes are lncRNA, and sky-blue nodes are mRNA. (B) miRNA-mediated mRNA-lncRNA intersecting interdependencies, confirmed by mirWalk 2.0. Green indicates molecules with reduced expression in the tumor compared to a normal sample, while yellow indicates increased expression in the tumor. Blue lines represent positive correlations, and black lines represent negative ones. (C) A 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. The original expression value was log₂-transformed.

3.6 - Survival analysis for muscle-invasive bladder cancer

Survival analysis showed that each of the identified RNA molecules involved in the identified cross-dependence is closely related to the prognosis of survival. In particular, higher expression levels of miR-141-5p, miR-141-3p, AC010326.3, AC073335.2, miR-200c-3p and GATA3 predict a favorable prognosis, indicating that they may function as tumor inhibitors (Figure 19 B, C, D, E, F and H). In contrast, higher expression levels of MIR100HG, PALLD and CLIC4 are associated with an unfavorable prognosis, suggesting that they may play an oncogenic role (Figure 19 A, G and I).

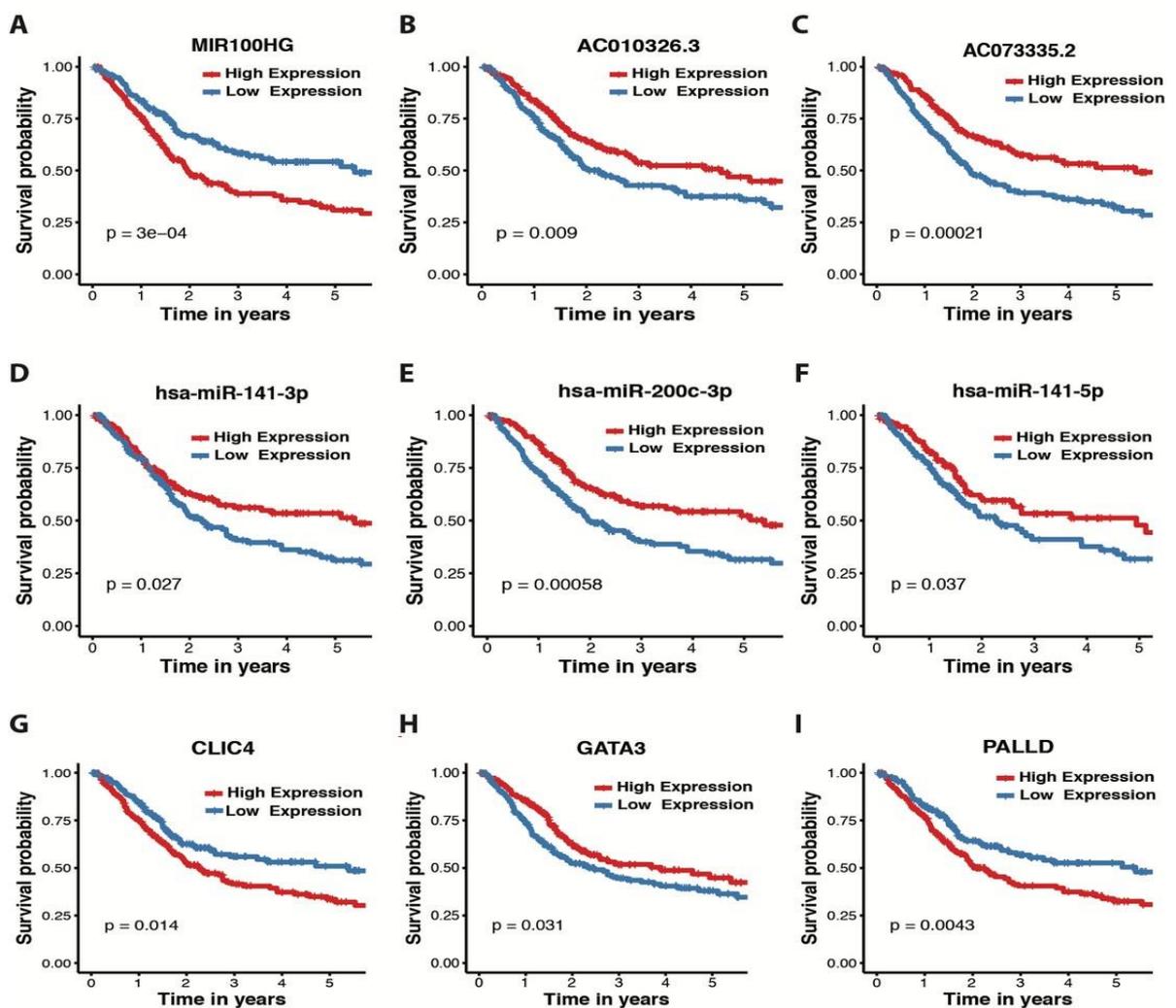


Figure 19 - Graphs of the Kaplan-Meier curve for nine interdependent RNAs from the TCGA dataset (A-I).

Note: red lines represent high RNA expression, while blue lines represent corresponding low expression. P-value was calculated using the log-rank test, where $p < 0.05$ represents statistical significance.

To test the prognostic effects of the three mRNAs (CLIC4, PALLD and GATA3) involved in cross-relationship dependent on miR-200c and miR141, we used two microarray data sets, GSE13507 (61 patients with MIBC) and GSE31684 (78 patients with MIBC). Further evidence is needed to confirm the prognostic role of lncRNA and miRNA.

Our results showed that GATA3 can be used as a marker of tumor suppressors, and CLIC4 and PALLD can be used as oncogenes. Due to cross-dependence, miR-200c and miR-141 can be used as effective clinical biomarkers for predicting MIBC (Figure 20 A-F).

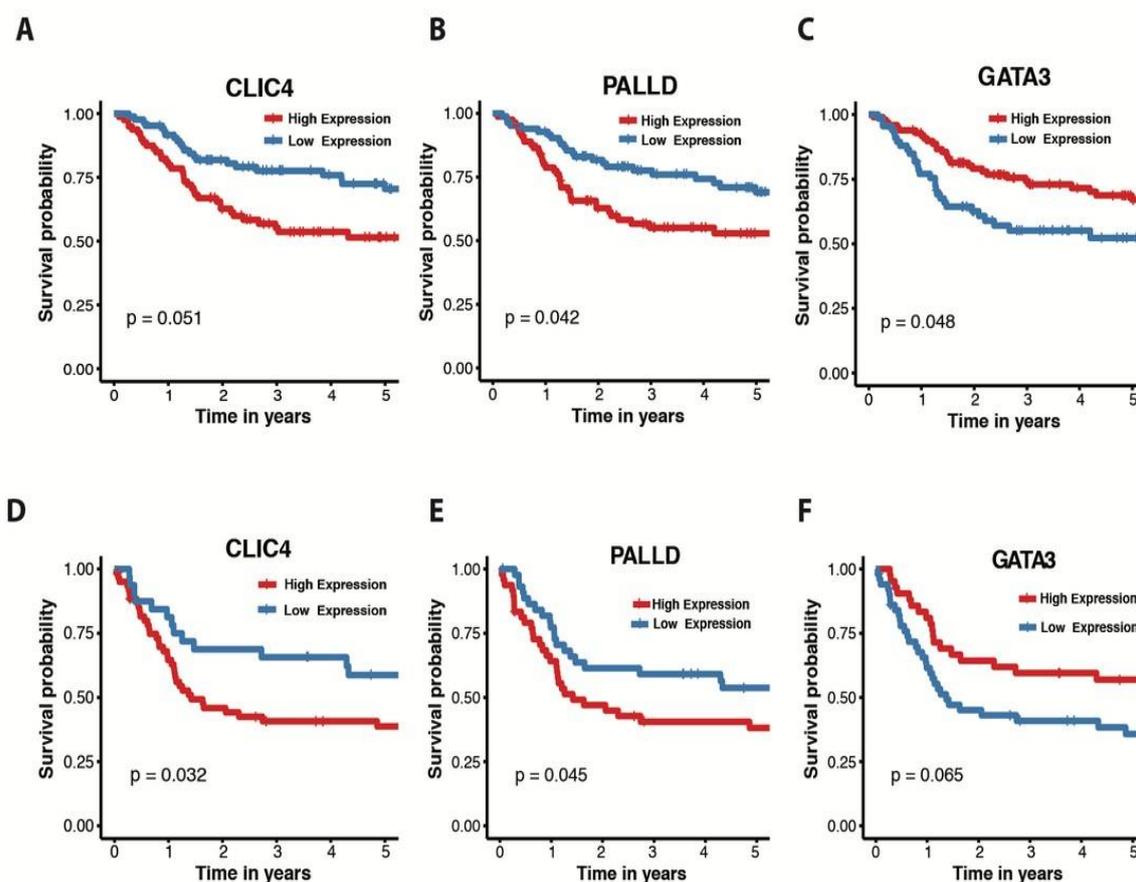


Figure 20 - Kaplan-Meier plots for CLIC4, PALLID and GATA3 in the GSE13507 and GSE31684 dataset

Note: (A-C) K-M survival curves show overall survival according to the high expression and low expression of CLIC4, PALLID and GATA3 in the GSE13507 dataset. (D-F) K-M survival curves show overall survival according to the high expression and low expression of CLIC4, PALLID, and GATA3 in the GSE31684 dataset.

Summary

In this chapter, we used RNA sequencing data available in the TCGA database to study tumor heterogeneity in muscle-invasive bladder cancer (MIBC), as well as to predict prognostic biomarkers. We classified 403 MIBC samples into two molecular types, which respectively resemble the previously identified basal and luminal subtypes.

It was revealed that the basal subtypes of bladder cancer are characterized by increased expression of genes related to the process of epithelial-mesenchymal transition, metastasis and the functioning of the immune system. Luminal subtypes are characterized by relatively increased expression of genes responsible for signaling processes. These results provide a deeper understanding of the heterogeneity of bladder cancer.

More importantly, we identified the miRNA-dependent mRNA and lncRNA cross-interactions, which consists of GATA3, CLIC4, PALLD, MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, AC010326.3, AC073335.2 and MIR100HG. It was revealed that the expression level of gene in this cross-dependence differs significantly between the basal and luminal subtypes and has a close relationship with the prognosis of MIBC.

These results indicate that the mRNA and lncRNA cross-interactions caused by miR-200c and miR-141 may play a significant role in tumor oncogenesis and progression. GATA3, CLIC4, PALLD, MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, AC010326.3, AC073335.2 and MIR100HG can serve as prognostic biomarkers of muscle-invasive bladder cancer, and can also be used as markers for subtype classification.

The relationship between subtypes and immunotherapy markers was investigated. It has been found that expression patterns of immune response checkpoint molecules (e.g., PD-1, PD-L1, CTLA-4, HAVCR-2 and LAG-3) distinguish between basal and luminal subtypes of bladder cancer. The basal subtype of MIBC revealed higher expression values of the immune response

checkpoint molecules than the luminal subtype, which must be taken into account when prescribing tumor immunotherapy.

List of works published by the 3rd Chapter

1. 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.

2. 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.

CHAPTER 4 - IDENTIFICATION OF CANDIDATE COMMON VARIABLE IMMUNODEFICIENCY (CVID) GENES

4.1 - Investigation of the multigenic nature of common variable immunodeficiency

The study of protein-protein interactions (PPI) of genes for common variable immunodeficiency (CVID) in the human genome is key to understanding the multi-gene nature of CVID and further identifying the ideal candidate CVID gene. PPI data for PID genes were obtained from the STRING database and visualized using Cytoscape software. It was found that the CVID genes tend to be centralized in the network, figuratively speaking, in the hub. This suggests that the CVID genes may interact with each other more often than other PID genes (Figure 21).

4.2 - Analysis of the gene relationships of the group of common variable immunodeficiency syndromes

To further investigate the phenomenon of complex interactions of CVID genes, the PPI data of the CVID group consisting of 39 CVID genes were compared with the PPI of ten random groups (each group consisted of 39 randomly selected PID genes).

To measure and compare the cohesion and density of the PPI networks of each group, the network density ($D_{network}$) was calculated. It is generally accepted that the higher the network density for a group, the closer the interaction of genes within the group.

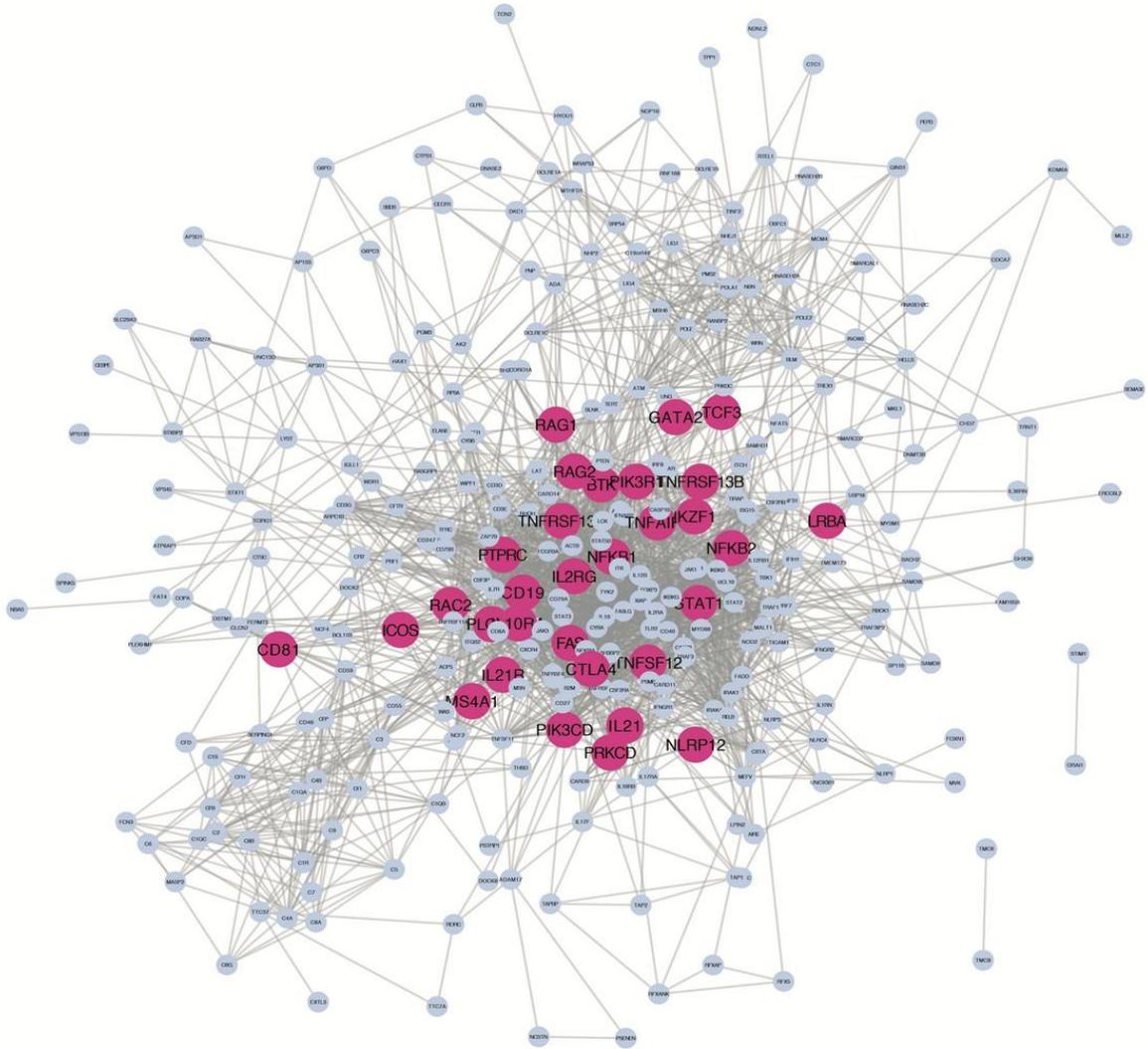


Figure 21 - Protein-protein interaction network (PPI) of primary immunodeficiency genes

Note: PPI data for PID genes obtained from the STRING database were visualized using Cytoscape software. Red nodes represent CVID genes; light blue nodes represent other PID genes.

The results showed that the CVID group has a higher network density than ten random groups, which indicates that the CVID genes actually interact more closely with each other than any other PID genes (Figure 22A). The biological distance ($B_{i,j}$) of the CVID group and two random groups (each group consisted of 39 PID genes) was calculated and the distribution of the biological distance density was compared. When interpreting the results, we had in mind that the smaller the biological distance for the group, the closer the functional significance between the genes in the group. The results showed that the density distribution of the CVID

group had a median value of 8.8, while the median values of the random groups were about 14, indicating a closer functional relationship between the CVID genes (Figure 22B).

The results obtained by above two methods indicate that the CVID genes interact more closely with each other and have a closer biological relationship than other PID genes.

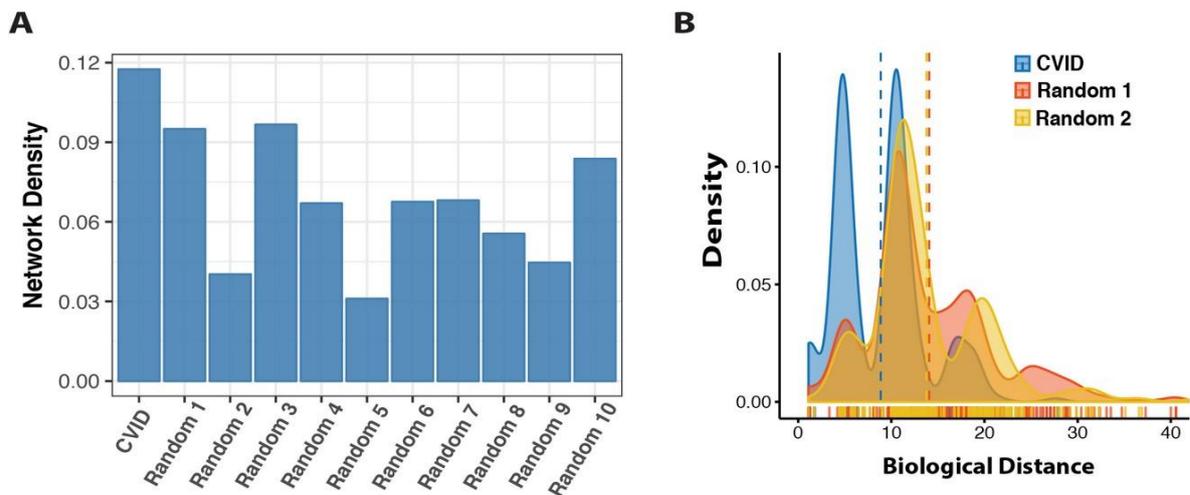


Figure 22 - Network densities and biological distances of the CVID group and random groups

Note: (A) a graph of the network density of the CVID group and ten random groups. (B) Density plot of biological distances of the CVID group and two random groups.

Based on Pearson's correlation analysis (PCC) and protein-protein interactions provided by Cheng F. et al. (2018) [136], 2751 CVID-specific interactions were obtained, including 1716 candidate genes. Using KEGG data, we performed a so-called analysis of the functional enrichment of the known CVID genes (in fact, linking the genes from the list to their biological functions) and found that 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 pathway) were statistically significant ($p < 0.05$, Figure 23).

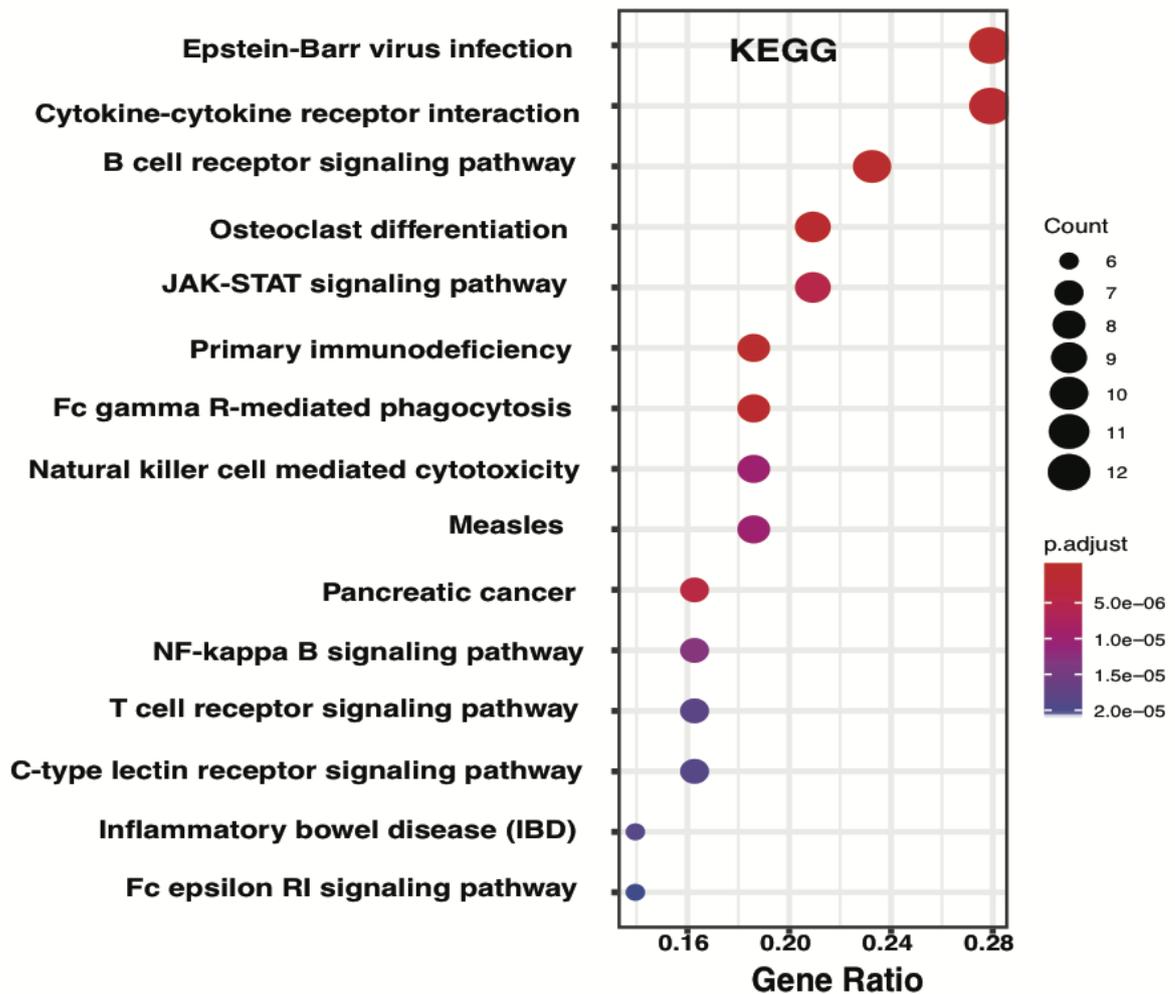


Figure 23 - Functional enrichment analysis of the CVID genes based on KEGG database

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

In order to identify candidate genes that are functionally similar to the known CVID genes, additional screening was carried out for certain candidate genes enriched with at least one of the above 15 signaling pathways according to the KEGG database. The analysis of 414 CVID-specific interactions revealed 172 new CVID candidate genes having a biological and functional relationship. The result of constructing a PPI network of interactions for the well-known CVID genes and candidate CVID genes is shown in Figure 24.

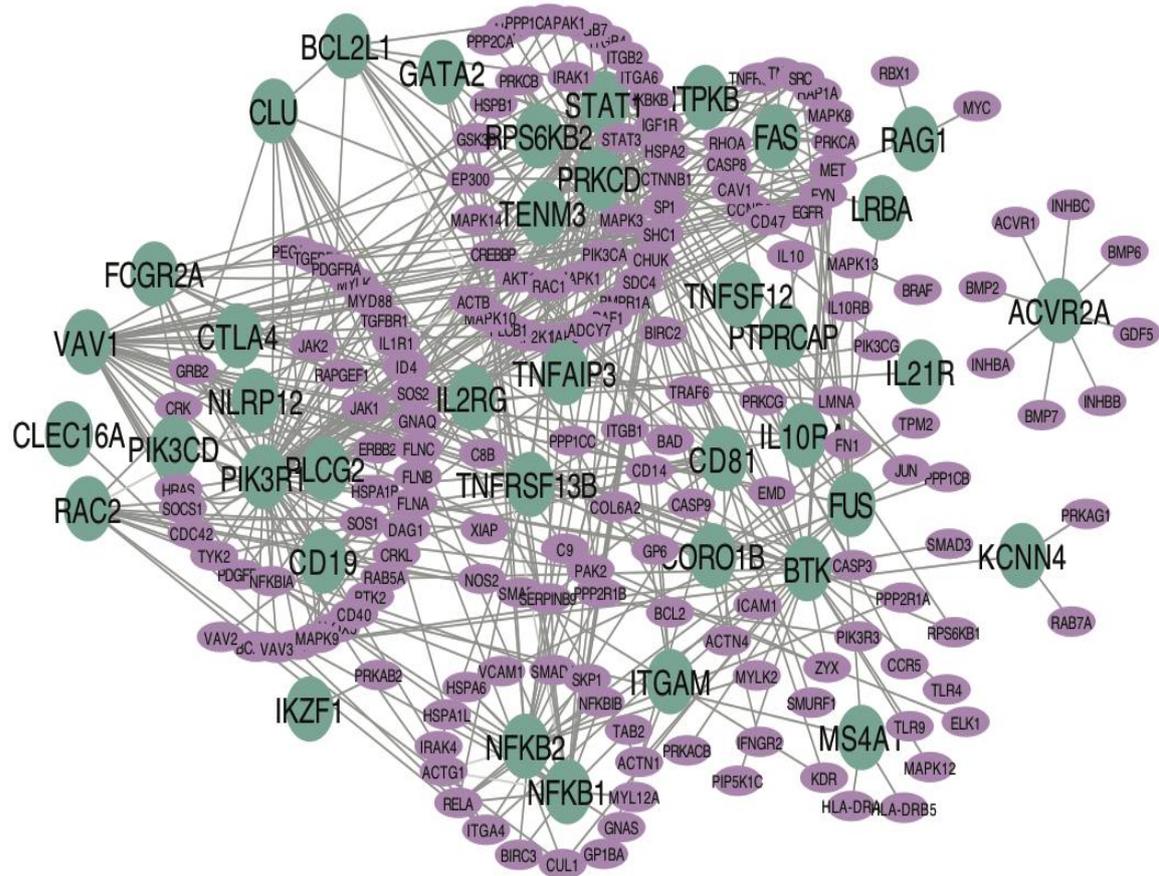


Figure 24 - Network of protein-protein interactions (PPI) of the CVID candidate genes (light purple) and the known CVID genes (dark green)

List of identified candidate genes for common variable immunodeficiency: 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.

4.3 - Evaluation of identified candidate genes for common variable immunodeficiency

To evaluate our predictions, the biological distance of 172 CVID candidate genes was calculated and compared with 39 known CVID genes. As a result, the average biological distance between the CVID candidate genes was 6.01, which is similar to the distance between the known CVID genes, and indicates a strong biological relationship between the CVID candidate genes and the CVID genes (Figure 25).

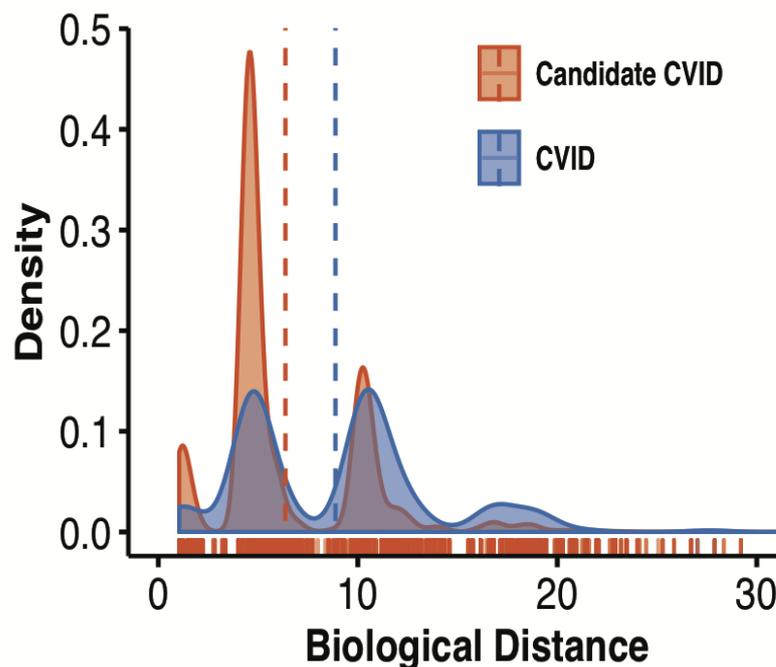


Figure 25 - Density graph of biological distances of the known CVID genes and predicted CVID candidate genes

Having identified the new candidate genes, we conducted a study of the quality of the forecast, having reviewed some of the studies. In particular, eight CVID candidate genes (AKT1, AKT3, RELA, SOCS1, STAT3, XIAP, CD40 and CASP8) that were not included in our initial list of CVID genes obtained from ESID, but were predicted by us, were identified in clinical cases of common variable immunodeficiency. This demonstrates the significance of the identified CVID candidate genes (Table 11).

Table 11 - List of recently published CVID candidate genes with identified association with CVID

Gene symbol	Description	Aliases	References
AKT1	AKT serine/threonine kinase 1	AKT, CWS6, PKB, PKB-ALPHA, PRKBA, RAC, RAC-ALPHA	PMID: 27664934
AKT3	AKT serine/threonine kinase 3	MPPH, MPPH2, PKB-GAMMA, PKBG, PRKBG, RAC-PK-gamma, RAC-gamma, STK-2	PMID: 26081581
RELA	RELA proto-oncogene, NF-kB subunit	NFKB3, P65	PMID: 27461466
SOCS1	Suppressor of cytokine signaling 1	CIS1, CISH1, JAB, SOCS-1, SSI-1, SSI1, TIP-3, TIP3	PMID: 29618830
STAT3	Signal transducer and activator of transcription 3	ADMIO, ADMIO1, APRF, HIES	PMID: 26360251 PMID: 27379089
XIAP	X-linked inhibitor of apoptosis	API3, BIRC4, IAP-3, ILP1, MIHA, XLP2, hIAP-3, hIAP3	PMID: 27492372
CD40	CD40 molecule	Bp50, CDW40, TNFRSF5, P50	PMID: 23305827 PMID: 30464201 PMID: 28756897
CASP8	Caspase 8	ALPS2B, CAP4, Casp-8, FLICE, MACH, MCH5	PMID: 28756897 PMID: 25814141

Summary

Common variable immunodeficiency is a type of PID that has deep phenotypic and genotypic heterogeneity and is caused by both monogenic and complex multigenic causes. In this chapter, by analyzing the network density and biological distance density, we found that CVID genes are functionally more similar to each other and interact closely with each other compared to other PID genes. This is consistent with the notion that CVID is more multigenic compared to most PID disease cohorts. On the other hand, this once again confirms the idea that CVID is rather a diagnosis of exclusion and unites a huge cohort of diverse nosologies with more or less common phenotypic manifestations for which a genetic cause and corresponding targeted treatment have not yet been determined.

In addition, we have developed methods that combine co-expression data, protein-protein interactions, as well as data characterizing gene functions to predict causally significant genes for the development of CVID. We identified 172 new CVID candidate genes that are interdependent with the known CVID genes in the same biological pathways and show a high biological correlation with the known CVID genes.

Eight CVID candidate genes that were not included in our initial list of CVID genes, obtained from ESID, but were predicted by us, were found in clinical cases of common variable immunodeficiency. This demonstrates the significance of the identified CVID candidate genes.

In conclusion, it can be said that this study contributed to a better understanding of the molecular mechanisms of CVID genes and gave new meaning to the list of CVID candidate genes.

List of papers published by the 4th chapter

1. 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: тез. докл. III всерос. конгресса. – 2018. – Vol. 56, № 3. – P. 5.

2. 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. – Vol. 6, № 2. – P. 47.

3. 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.

CHAPTER 5 - IDENTIFICATION OF MUTATIONS IN GENES, FAT4, RAG1, PIK3CD and CSF3R, LEADING TO THE DEVELOPMENT OF A DISEASE IN A PATIENT WITH PHENOTYPIC MANIFESTATIONS OF HENNEKAM SYNDROME

This study was conducted due to the fact that under our supervision there was a patient with phenotypic manifestations of Hennekam syndrome, but with clinical features that allowed to doubt the correctness of this diagnosis. After conducted a whole genome sequencing in a commercial laboratory, the diagnosis was not confirmed. In view of this, it was decided to apply new methods to identify mutant genes associated with the phenotype of the patient and confirm the diagnosis.

This section of the study has allowed to show the effectiveness and opportunities of applied techniques and the application of basic research. In addition, in this case it turned out to identify the mutation in the pathogenic significance of genes, making it possible to justify the practical application of new therapeutic technologies tested abroad, but not included for use in Russia.

5.1 - Clinical features of the studied patient

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 27A);
- warty lesions on the skin of the palm, forefinger, and thumb of the right hand (Figure 27B);
- 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 27B);
- cystic defects of the mandible involving bone tissue - magnetic resonance imaging (MRI) of 11.10.2017;
- 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 27B);

In history - increase of urea, renal hypoplasia.

In addition, heredity -

- the presence of a regressive pregnancy interrupted at 12 weeks of gestation
- for several generations, primary vegetative-vascular dystonia was observed against the background of connective tissue dysplasia, which occurred with severe vascular changes in the form of varicose veins, hemorrhoids, sub-compensated tonsillitis. On the maternal side, the presence of a constitutionally caused (primary) vegetative-vascular dystonia accompanied the entire pregnancy - hypotension and rhinitis with the use of vasoconstrictor drops throughout pregnancy, polyhydramnios;
- the birth of large fruit – the sibling, 13 years old, was born with a weight of 4150g years, but a usual girl - 4750g with a length of 57cm;
- the mother does not exclude autoimmune hypothyroidism and the weakness of labor during the first birth, which served as the basis for operative delivery by cesarean section;

- on the paternal side - the presence of pathology of the digestive system in the form of pancreatitis and calculous cholecystitis, for which an operation was performed.

A**B**

Figure 27 - 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).

5.2 - Search for mutations in the patient, which could lead to the development Hennekam syndrome

Analysis of the whole genome sequencing data of the patient was carried out using the BWA, GATK4, VCFtools software. An analysis of the so-called “candidate variant filtering” was performed using the ANNOVAR software and the Combined Annotation Dependent Depletion (CADD) database, and its results are schematically presented in Figure 28. The first filtration step was to remove all synonymous SNV, non-frames InDels and embeddings are marked as “NA” or “unknown”. A total of 10302 were identified variants or INDEL SNV. Then, the identified variants were filtered by overlaying on the known 351 PID genes and 2 known Hennekam syndrome genes, not included in the list, FAT4 and CSF3R. Selected 111 variants were retained to search for more possible ones. After eliminating the common variants, whose Minor allele frequency (MAF)>0.01 for The Exome Aggregation Consortium (ExAC), 1000g and The Genome Aggregation Database (gnomAD), a total of 6 rare variants remained. To select pathogenic mutations, CADD, the Functional Analysis through Hidden Markov Models (FATHMM), and Protein Variation Effect Analyzer (PROVEAN) models were used, and finally, four mutations that are likely to lead to the development of the disease in this patient were predicted.

5.3 - Assessment of the probable pathogenicity mutations identified in the study

In particular, a homozygous mutation (g.125452634G>A) was identified in the FAT4 gene, which was reported to have both biallelic and monoallelic mutations associated with Hennekam syndrome. 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 (Table 12).

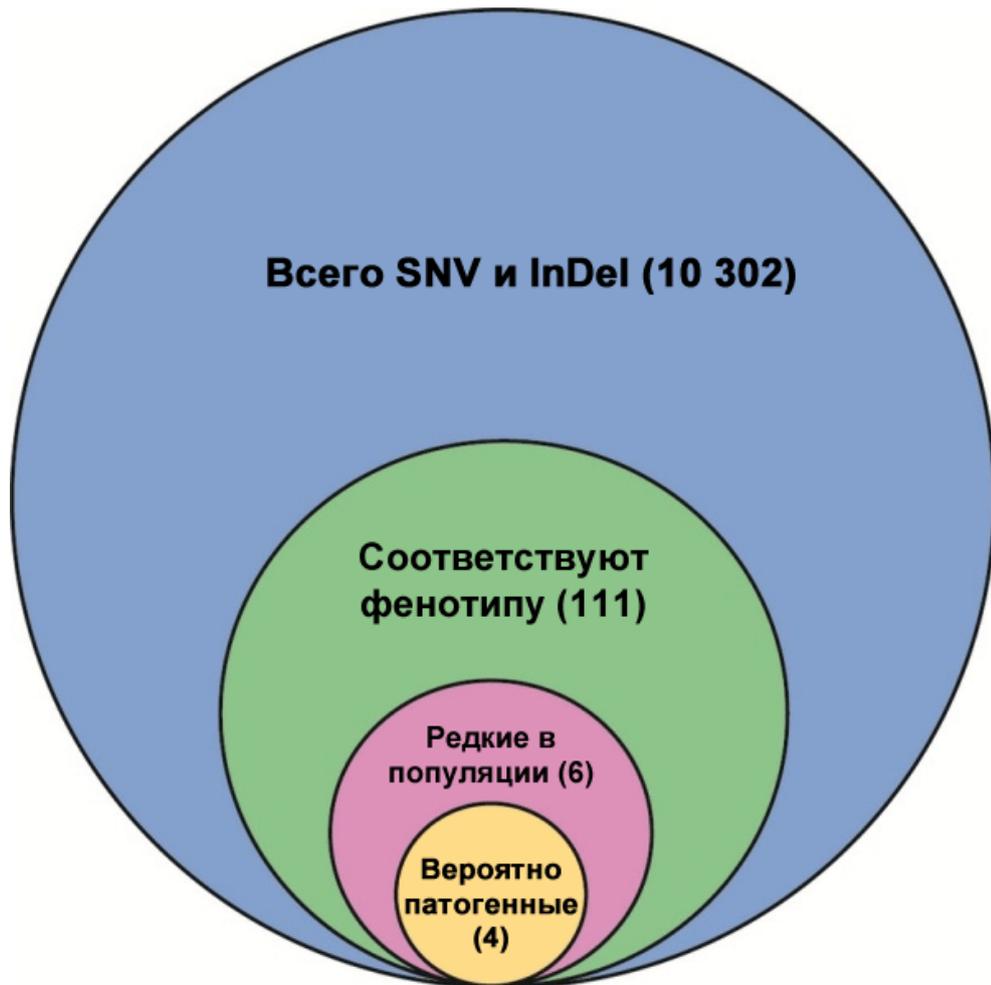


Figure 28 - Diagram illustrating all four main filtering steps to identify potentially pathogenic mutations

Note: the number of filtered genes is shown in parentheses.

All candidate mutations were confirmed by examining the data in the “integrated genomics viewer” (IGV) software, the results are shown in Figure 29.

Regarding the homozygous mutation g.125452634G> A, the FAT4 gene contains conflicting information. According to the dbSNP database of the National Center for Biotechnological Information (NCBI) of the United States, a mutation is rare and benign, not associated with any phenotype of the disease (data dated January 8, 2016).

Table 12. Summary of the SNV results of the molecular screening of disease-related genes, including the location of mutations, pathogenicity predictions, and population data

Gene	dbSNP ID	Characteristic phenotype described	Localization on chromosome	Homo - or heterozygous	Mutation	Amino acid substitution	Population frequency (1000g/gnomAD / ExAC)	The probability of pathogenicity (FATHMM/PROVEAN/CADD)
FAT4	rs12650153	Hannekama syndrome	Chr4 (q28.1)	Hom	g.125452634G>A	p.Ser3875Asn	0.009/0.007/0.002	T/N/20.1
RAG1	rs4151034	various PID	Chr11 (p12)	Het	g.36575963G>A	p.Asp887Asn	0.001/0.006/0.004	D/N/23.5
PIK3CD	rs142285826	CVID	Chr1 (p36.22)	Het	g.9715914T>A	p.Phe146Ile	0.001/0.003/0.002	T/N/23.4
CSF3R	rs3918019	Congenital neutropenia	Chr1 (p34.3)	Het	g.36471505C>T	p.Glu405Lys	0.005/0.002/0.006	T/N/17.7

Note: 1000g, gnomAD, ExAC (score<0.01 indicated rare); FATHMM (T, tolerated; D, deleterious); PROVEAN (N, neutral; D, deleterious); CADD (score>15 implied deleterious); CN indicated Congenital neutropenia, Het indicated heterozygous, and Hom indicated homozygous.

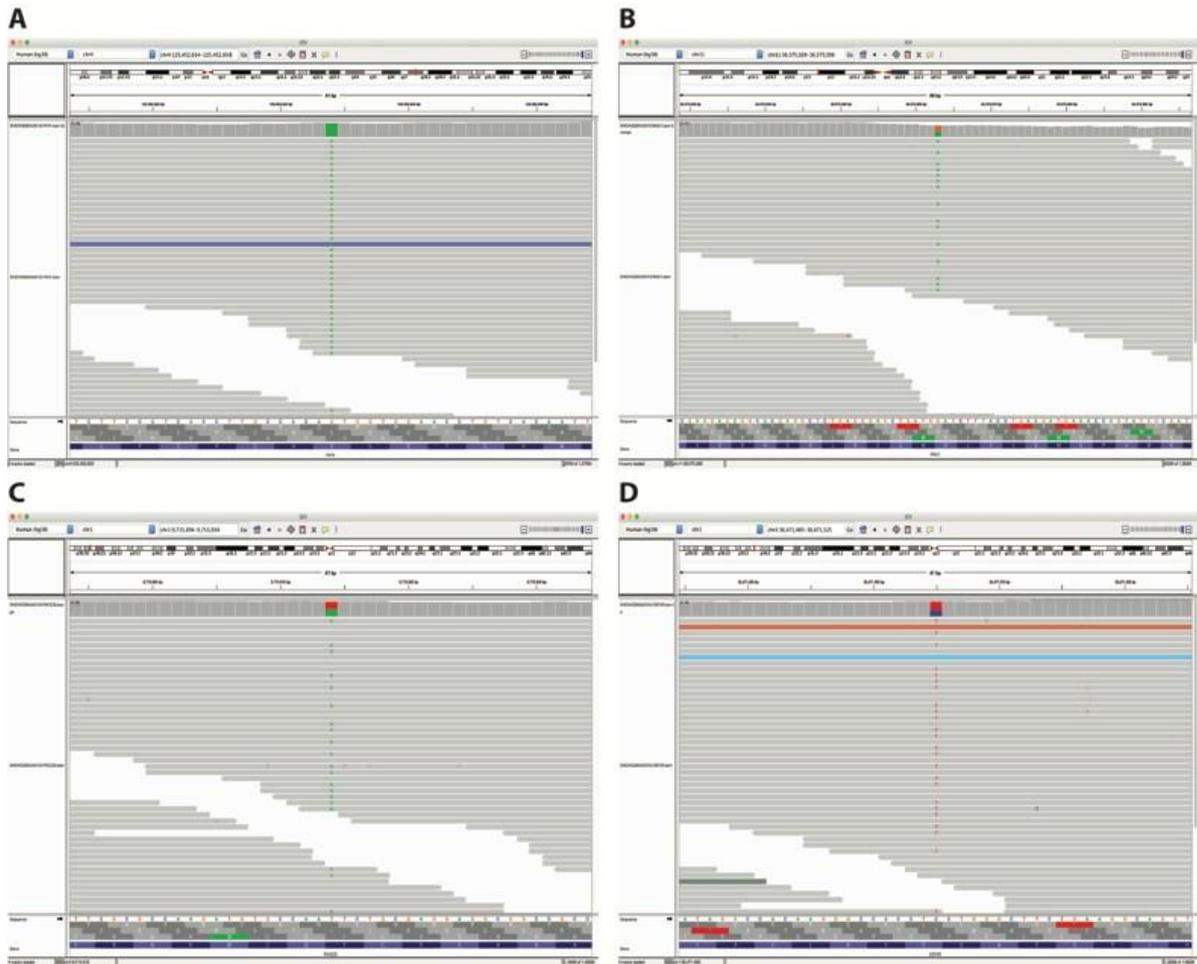


Figure 29 - Visualization of the four mutations candidates in Integrative genomics viewer (IGV)

Note: (A) G>A substitution on Chr4: 125452634 in FAT4. (B) G>A substitution on Chr11:125452634 in RAG1. (C) T>A substitution on Chr1:9715914 in PIK3CD. (D) C>T substitution on Chr1:36471505 in CSF3R. One cell on the coverage track represents a heterozygous mutation and two cells on coverage track represent a homozygous mutation.

According to the latest version of the CADD database using the human reference genome Ch38 (GRCh38-v1.5), the mutation is potentially pathogenic. Given that mutations in FAT4 were previously determined to be significant for Hennekam syndrome, this mutation was considered potentially significant in this study and its etiological significance could not be completely denied. It is known that in addition to Hennekam's syndrome, the allelic mutation at the FAT4 locus leads to the development of Van Maldergam syndrome, which is manifested by some signs of Hennekam syndrome, such as facial dysmorphism and cognitive

impairment, but without malformations of the lymphatic system with symptoms of lymphedema and lymphangiectasia [77]. In the studied patient, functional studies of the FAT4 gene variants were not previously performed.

Checking the pathogenicity of this mutation for this patient was carried out by the Sanger sequencing method. When comparing the results of parent and sibling, it turned out that all subjects are carriers of the same homozygous mutation FAT4 (g.125452634G> A) as in the patient, while the phenotype of Hennekam syndrome was observed only in the patient. It was concluded that the G>A substitution for Chr4:125452634 in FAT4 might be non-pathogenic (Figure 30).

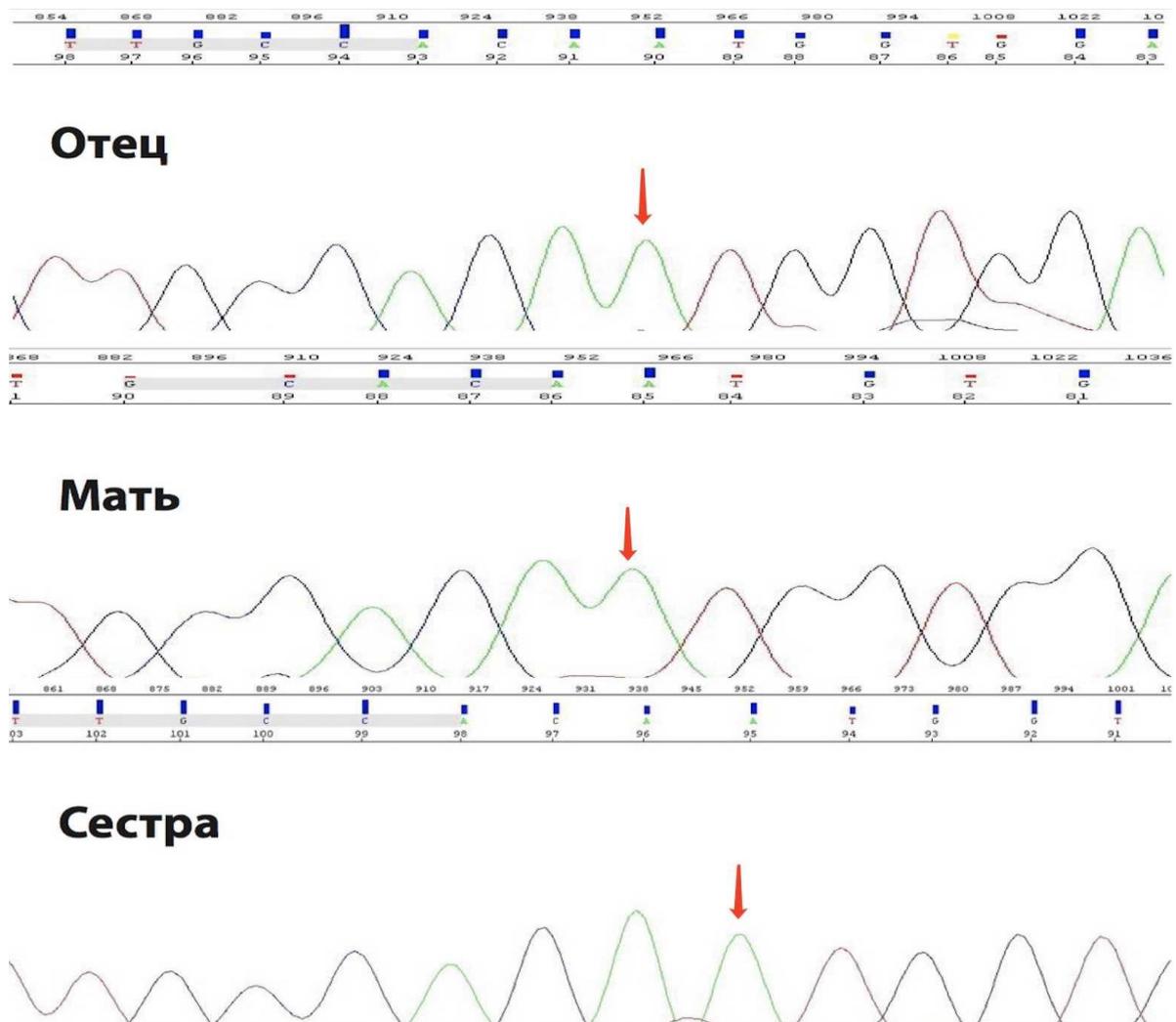


Figure 30 - Analysis of the nucleotide sequence of the FAT4 gene in father, mother and sister

Note: a solid line with an arrow indicates the replacement g.125452634G> A.

However, identical mutations in the same gene may lead to different symptoms of hereditary human diseases due to decrease of the variable penetrance and expressivity which are caused by factors such as the interaction of genes and environmental factors, allelic variations and stochasticity [197]. Consequently, the question of whether homozygous mutation (g.125452634G> A) in gene FAT4 can lead to different symptoms in different people, deserve further analysis. In addition, the complex genetic and environmental factors that have led to reduced penetrance and variable expressivity, should be explored by conducting a cohort study.

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 (or B) cell-negative severe combined immunodeficiency (SCID), increased susceptibility to severe cytomegalovirus infection, autoimmune diseases, Omenn's syndrome, etc. [138]. Mutations with the enhanced function of the PIK3CD gene, which encodes the catalytic subunits P110 δ and P85 phosphoinositide-delta, lead to the development of the so-called PI-3 kinase activation syndrome (previously considered one of the CVID), in the phenotype of which one of the expressed symptoms of the disease is lymphoproliferation, which can lead to B-lymphomas. In addition, this syndrome is characterized by hypogammaglobulinemia, recurrent viral infections, bronchiectasis, and herpesvirus infections [11, 35, 51, 139].

In the treatment of phosphoinositide 3-kinase δ activation syndrome, along with immunoglobulin replacement therapy, mTOR inhibiting drugs are used, for example, rapamycin (sirolimus, everolimus).

Previously, this drug was used to prevent allograft rejection during kidney and liver transplantation. In clinical immunology, this drug began to be used in the last 10 years in the treatment of phosphoinositide 3-kinase δ activation syndrome. However, recent reports of researchers indicate that mTOR pathway inhibitors can

be successfully used for malformations of the lymphatic vessels [14]. Everolimus was prescribed to this patient, which also proved to be effective, eliminated warty growths on the skin and reduced the rate of increase in lymphatic edema, which allowed avoiding traumatic and painful manipulations of lymphatic drainage. The effect of the drug is explained by the participation of the mTOR pathway and the PIK3CD gene in the development of human lymphatic vessels.

Mutations in the CSF3R gene are important for the intracytoplasmic domain of G-CSFR, which were discovered in patients with congenital neutropenia [58]. It is also known that these mutations sometimes manifest as somatic mutations associated with the progression of leukemia in congenital neutropenia [33, 58, 72].

Summary

In Chapter 5, we analyzed the case of a patient with lymphangiectasia-lymphedema, warty formations, hirsutism, hydrothorax and lymphatic ascites. To better understand pathogenesis and provide accurate treatment options, whole genome sequencing was performed. Using bioinformatics tools, several mutations have been identified that may be significant for a patient with primary immunodeficiency, including with Hennekam syndrome.

Four mutations were identified in genes known to be pathogenic in PID, including Hennekam syndrome. Identified mutations will help to approach a more accurate diagnosis and substantiate the patient's pathogenetic treatment, as well as determine the origin of these mutations and investigate the molecular mechanisms of disease formation.

List of papers published by chapter 5

1. 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. – 2019. – прил. 2, окт. – С. 51.

Liu, G. 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.

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, huge amounts of genomic data will be generated, 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 cancers, auto-inflammatory, autoimmune diseases, and primary immunodeficiencies.

The pathophysiological gene-dependent mechanisms of certain types of immune-mediated pathology, muscle-invasive bladder cancer, and a number of primary immunodeficiencies, were studied.

The studies conducted allowed us to look at the processes of system biology from a pathophysiological point of view and find general patterns. Both primary immunodeficiencies and onco-pathology can be presented in a single scheme at the level of common typical pathological processes, such as tissue stress and its participation in the general signaling and metabolic pathways.

The genes with increased expression revealed in the studies encode molecules whose activity is characteristic of the epithelial-mesenchymal transition, intercellular adhesion, regulation of oxidative phosphorylation, activation of transcription and blocking of apoptosis. All molecules are physiological and work under normal conditions in cells in many body tissues. However, they change their meaning in the context of tumor growth and are witnesses of a typical pathological process.

Our findings on 403 cases of MIBC allowed us to classify MIBC into basal and luminal subtypes. The basal subtype of MIBC revealed higher expression values of the immune checkpoint molecules than the luminal subtype, which

should be taken into account when prescribing tumor immunotherapy. According to our studies, the basal subtypes of bladder cancer are characterized by increased expression of genes related to the process of epithelial-mesenchymal transition, metastasis, and the functioning of the immune system. Luminal subtypes were characterized by relatively increased expression of genes responsible for signaling processes. These results provide a deeper understanding of the heterogeneity of bladder cancer.

The overall clinical picture is formed with the participation of a huge network of interactions of regulatory mechanisms, including non-coding RNAs such as miRNA and lncRNA. We identified the cross-interaction of mRNA, miRNA and lncRNA, including GATA3, CLIC4, PALLD, MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, AC010326.3, AC073335.2 and MIR100HG. The level of gene expression in this cross-dependence was significantly different between the basal and luminal subtypes and was closely related to the prognosis of MIBC. These results indicate that the cross-interaction of non-coding RNAs can play a significant role in tumorigenesis and tumor progression, and the corresponding molecules can serve as prognostic biomarkers of muscle-invasive bladder cancer used to classify pathology subtypes.

More importantly, we have identified miRNA-dependent mRNA and lncRNA cross-interaction, including GATA3, CLIC4, PALLD, MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, AC010326.3, AC073335.2 and MIR100HG. We found that the expression level of gene in this cross-dependence was significantly different between the basal and luminal subtypes and was closely related to the prognosis of MIBC.

Exploring the network density and the so-called the biological distance for the genes related to primary immunodeficiencies (PID) and common variable immunodeficiency (CVID), in particular, it was found that CVID genes are functionally more similar to each other and interact closely with each other compared to other PID genes. This is consistent with the idea that CVID is a more

multi-gene nosological syndrome than most diseases among PIDs. On the other hand, this once again confirms the opinion that CVID is rather an exclusion diagnosis and combines a huge cohort of diverse nosologies with more or less common phenotypic manifestations, for which a genetic cause has not yet been determined and appropriate targeted treatment has not been selected.

In addition, we have developed methods that combine co-expression data, protein-protein interaction data, as well as data characterizing gene functions, allowing us to predict causally significant genes for the development of CVID. 172 new CVID candidate genes were identified that are interdependent with the known CVID genes in the same biological pathways and show a high biological correlation with the known CVID genes. Several testing methods and confirmation of several predicted genes in our study in patients, that together strongly demonstrate the importance of predicted candidate genes.

In addition, we have developed methods that combine data on co-expression, protein-protein interactions, as well as data characterizing the functions of genes, allowing us to predict causally significant genes in the development of CVID. 172 new CVID candidate genes were identified that are interdependent with the known CVID genes in the same biological pathways and demonstrate a high biological correlation with the known CVID genes. Several test validation methods and confirmation of several predicted genes in our study in other clinical studies also demonstrate the importance of predicted candidate genes.

This study contributed to a better understanding of the molecular mechanisms of CVID genes by expanding the range of candidate genes.

Using bioinformatics analysis methods tested in theoretical studies, a specific case was analyzed with a rare combination of phenotypic signs in one symptom complex - lymphangiectasia-lymphedema, warty formations, hirsutism, hydrothorax and lymphatic ascites. To better understand the pathogenesis and justify the necessary therapeutic interventions, whole genome sequencing was performed. As a result of bioinformatics analysis, four mutations in the genes

known as pathogenic in PID, including Hennekam syndrome, were identified. Identified mutations will help to approach a more accurate diagnosis and substantiate the patient's pathogenetic treatment, as well as determine the origin of these mutations and investigate the molecular mechanisms of the formation of the disease.

In addition, studies of rare cases of human pathology make it possible to solve general pathological issues of the formation of diseases, enriching pathophysiological science with knowledge about the patterns of functioning of the human body, about the molecular mechanisms of pathology formation, becoming objective justifications for the development and application of targeted therapeutic tactics, which opens up the prospect of creating new targeted drugs.

Prospects for further development of the theme

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 with CVID. Continuing research in this area will help to reveal the general mechanisms of development of such immunopathology as cancer, including that developed as a complication of primary immunodeficiencies.

The approach used to identify the mechanisms of immunopathogenesis of the group of syndromes "common variable immune deficiency" can be extrapolated to study the formation of the phenotype of other syndromes of primary immunodeficiencies.

Identification of specific genes for CVID syndromes and determination of the molecular basis of immunopathology will allow us to study pathogenesis, differentiate nosological forms from the vast heterogeneous group of CVID, and bring closer the creation of specific targeted therapy, which is much more effective than traditional replacement therapy with intravenous or subcutaneous immunoglobulins. This will level out the clinical manifestations of pathology, improve the prognosis and quality of life of patients.

The process of verification of primary immunodeficiency genes can be improved by the development of software for predicting candidate genes of various immunopathologies, 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 micro-RNAs detected by computational analysis should also be confirmed by biological experimental studies in the future.

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

Regarding the identification of the causes of the development of syndromes of malformation of the lymphatic vessels and hereditary lymphedema in the interpretation of the results obtained, it must be borne in mind that other forms of transmission of hereditary information are possible, including autosomal dominant, co-dominant and co-expressive. This determines the need to continue the search for the pathogenetic foundations of the formation of immunopathology, taking into account new data on the pathogenesis mechanisms and the signaling pathways involved, using bioinformation analysis and accumulated clinical experience in identifying a rare pathology associated with congenital immunity errors. In addition, according to recent studies, DNA variations outside exons can also lead to genetic abnormalities, prompting full-scale studies of the human genome at the international level.

Thus, our studies have allowed us to draw the following conclusions.

FINDINGS

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

1. Tumor-infiltrated macrophages relating to the M2 family significantly correlated with MIBC patient's survival, grade, tumor stage, therefore, can be markers of tumor at muscle-invasive bladder cancer.

2. cross-interaction among GATA3, PALLD, CLIC4, miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 and MIR100HG, plays a significant role in oncogenesis and progression of the tumor, and identified biomarkers are important in determining the subtypes of muscle-invasive bladder cancer and prognosis of disease.

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, and 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, as well as data 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. Before starting immunotherapy for patients with muscle-invasive bladder cancer, it is extremely necessary to determine the subtype of cancer (basal or luminal) by determining 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, it is necessary to study the following biomarkers: GATA3, PALLD, CLIC4, miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 and MIR100HG.

3. When conducting research on predicting candidate genes for common variable immunodeficiency, it is necessary to include co-expression factors, protein-protein interactions, and signaling pathways in the 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 172 two 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.

ABBREVIATIONS

1000g	1000 Genomes
ASW	Average silhouette width
AUC	Area under the ROC curve
BC	Bladder cancer
BLK	B-lymphoid tyrosine-protein kinase
BWA	Burrows-Wheeler Aligner
CADD	Combined Annotation Dependent Depletion
CoC	Cluster of Cluster
CVID	Common variable immune deficiency
dbSNP	Single Nucleotide Polymorphism Database
DEG	Differentially expressed gene
ExAC	Exome Aggregation Consortium
FATHMM	Functional Analysis through Hidden Markov Models
FDR	False Discovery Frequency
FGA	Functional genomic alignment
GATK	The Genome Analysis Toolkit
GEO	Gene Expression Omnibus
gnomAD	The Genome Aggregation Database
GO	Gene Ontology
GSEA	Gene set enrichment analysis
HGC	Human Genome Connectome
HSCT	Hematopoietic stem cell transplantation
IgA	Immunoglobulin A
IGV	Integrative Genome Viewer
iter	Number of iterations
IUIS	International Union of Immunological Societies
K-M	Kaplan–Meyer
KEGG	Kyoto Encyclopedia of Genes and Genomes

lncRNA	Long non-coding RNA
MAF	Minor allele frequency
MIBC	Muscle-invasive bladder cancer
miRNA	MicroRNA
NES	Normalized enrichment score
NGS	Next-generation sequencing
PCC	Pearson's correlation analysis
PID	Primary immunodeficiency
PPI	Protein-protein interaction
PROVEAN	Protein Variation Effect Analyzer
RF	Random Forest
ROC	Receiver operating characteristic
SCID	Severe combined immune deficiency
SNPs	Single nucleotide polymorphisms
SNVs	Single nucleotide variants
TCGA	The Cancer Genome Atlas
TILs	Tumor-infiltrating lymphocytes
TLR	Toll-like receptor
TMB	Tumor mutational burden
WGS	Whole genome sequencing
XG	eXtreme Gradient Boosting; XGBoost

REFERENCES

1. Латышева, Е.А. Первичные иммунодефициты: состояние проблемы на сегодняшний день. JMF-центры в России / Е.А. Латышева // Вопросы современной педиатрии. – 2013. – Т. 12, № 6. – С. 73-77.
2. Современные возможности скрининга и диагностики первичных иммунодефицитных состояний в педиатрии / Е.В. Чернышова, [и др.] // Педиатрия. – 2016. – Т. 95, № 3. – С. 124-128.
3. Тузанкина, И.А. К вопросу диагностики иммунопатологии / И.А. Тузанкина // Медицинская иммунология. – 2010. – Т. 12, № 6. – С. 485-496.
4. Тузанкина, И.А. Первичные иммунодефициты в раннем возрасте / И.А. Тузанкина, С.С. Дерябина, М.А. Болков. – Москва, 2018. – 176 с.
5. Щербина, А.Ю. Первичные иммунодефициты-реалии XXI века / А.Ю. Щербина // Вопросы гематологии/онкологии и иммунопатологии в педиатрии. – 2016. – Т. 15, № 1. – С. 8-9.
6. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? / L. Salmena, [et al.] // Cell. – 2011. – Vol. 146, № 3. – P. 353-358.
7. A long noncoding RNA activated by TGF- β promotes the invasion-metastasis cascade in hepatocellular carcinoma / J. Yuan, [et al.] // Cancer cell. – 2014. – Vol. 25, № 5. – P. 666-681.
8. A molecular taxonomy for urothelial carcinoma / G. Sjö Dahl, [et al.] // Clinical Cancer Research. – 2012. – Vol. 18, № 12. – P. 3377-3386.
9. A multivariate approach to the integration of multi-omics datasets / C. Meng, [et al.] // BMC Bioinformatics. – 2014. – Vol. 15, № 1.
10. A proteome-scale map of the human interactome network / T. Rolland, [et al.] // Cell. – 2014. – Vol. 159, № 5. – P. 1212-1226.
11. Activating PI3K δ mutations in a cohort of 669 patients with primary immunodeficiency / M. Elgizouli, [et al.] // Clinical and Experimental Immunology. – 2016. – Vol. 183, № 2. – P. 221-229.
12. An antibody-deficiency syndrome due to mutations in the CD19 gene

/ M.C. Van Zelm, [et al.] // *New England Journal of Medicine*. – 2006. – Vol. 354, № 18. – P. 1901-1912.

13. An extensive comparison of recent classification tools applied to microarray data / J.W. Lee, [et al.] // *Computational Statistics and Data Analysis*. – 2005. – Vol. 48, № 4. – P. 869-885.

14. An updated review on activated PI3 kinase delta syndrome (APDS) / A. Singh, [et al.]. // *Genes & Diseases*. – 2019.

15. Anastasiadis A. Best practice in the treatment of nonmuscle invasive bladder cancer / A. Anastasiadis, T.M. de Reijke // *Therapeutic Advances in Urology*. – 2012. – Vol. 4, № 1. – P. 13-32.

16. Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders / P.A. van Schouwenburg, [et al.] *Clinical Immunology*. – 2015. – Vol. 160, № 2. – P. 301-314.

17. Approved checkpoint inhibitors in bladder cancer: which drug should be used when? / P. Ghatalia, [et al.] // *Therapeutic advances in medical oncology*. – 2018. – Vol. 10. – P. 1758835918788310.

18. Autosomal recessive intestinal lymphangiectasia and lymphedema, with facial anomalies and mental retardation / R.C.M. Hennekam, [et al.] // *American Journal of Medical Genetics*. – 1989. – Vol. 34, № 4. – P. 593-600.

19. Ballgown bridges the gap between transcriptome assembly and expression analysis / A.C. Frazee [et al.]. // *Nature biotechnology*. – 2015. – Vol. 33, № 3. – P. 243.

20. Bioinformatics for cancer immunology and immunotherapy / P. Charoentong, [et al.] // *Cancer Immunology, Immunotherapy*. – 2012. – Vol. 61, № 11. – P. 1885-1903.

21. Bioinformatics tools and databases for analysis of next-generation sequence data / H.C. Lee, [et al.] // *Briefings in Functional Genomics*. – 2012. – Vol. 11, № 1. – P. 12-24.

22. Bladder cancer / A.M. Kamat, [et al.] // *Lancet*. – 2016. – Vol. 6736, № 16. – P. 1-15.
23. Boyle J.M. Population Prevalence of Diagnosed Primary Immunodeficiency Diseases in the United States / J.M. Boyle, R.H. Buckley // *Journal of Clinical Immunology*. – 2007. – Vol. 27, № 5. – P. 497-502.
24. Brief report Novel HAX1 mutations in patients with severe congenital neutropenia reveal isoform-dependent genotype-phenotype associations / M. Germeshausen, [et al.] // *Blood*. – 2008. – Vol. 111, № 10. – P. 4954-4957.
25. Cairns R.A. Regulation of cancer cell metabolism / R.A. Cairns, I.S. Harris, T.W. Mak // *Nature Reviews Cancer*. – 2011. – Vol. 11, № 2. – P. 85.
26. Cancer statistics in China, 2015 / W. Chen, [et al.] // *CA: A Cancer Journal for Clinicians*. – 2016. – Vol. 66, № 2. – P. 115-132.
27. Care M.A. Parsimonious Gene Correlation Network Analysis (PGCNA): a tool to define modular gene co-expression for refined molecular stratification in cancer / M.A. Care, D.R. Westhead, R.M. Tooze // *NPJ systems biology and applications*. – 2019. – Vol. 5, № 1. – P. 13.
28. CDG: An online server for detecting biologically closest disease-causing genes and its application to primary immunodeficiency / D. Requena, [et al.] // *Frontiers in Immunology*. – 2018. – Vol. 9, № JUN. – P. 1340.
29. Chen T. Xgboost: a scalable tree boosting system In: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* / T. Chen, C. Guestrin // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. – 2016. – P. 785–794.
30. Circular RNA is enriched and stable in exosomes: A promising biomarker for cancer diagnosis / Y. Li, [et al.] // *Cell research*. – 2015. – Vol. 25, № 8. – P. 981.
31. Classification of Small GTPases with Hybrid Protein Features and Advanced Machine Learning Techniques / Z. Liao, [et al.] // *Current*

Bioinformatics. – 2017. – Vol. 13, № 5. – P. 492-500.

32. Clinical picture and treatment of 2212 patients with common variable immunodeficiency / B. Gathmann, [et al.] // *Journal of Allergy and Clinical Immunology*. – 2014. – Vol. 134, № 1. – P. 116-126.

33. Clinical Relevance of Point Mutations in the Cytoplasmic Domain of the Granulocyte Colony-Stimulating Factor Receptor Gene in Patients With Severe Congenital Neutropenia / N. Tidow, [et al.] // *Blood*. – 1997. – Vol. 89, № 7. – P. 2369-2375.

34. Clinical significance of immune cell infiltration within gallbladder cancer / Y. Nakakubo, [et al.] // *British Journal of Cancer*. – 2003. – Vol. 89, № 9. – P. 1736-1742.

35. Clinical spectrum and features of activated phosphoinositide 3-kinase δ syndrome: A large patient cohort study / T.I. Coulter, [et al.] // *Journal of Allergy and Clinical Immunology*. – 2017. – Vol. 139, № 2. – P. 597-606.

36. ClusterProfiler: An R package for comparing biological themes among gene clusters / G. Yu, [et al.] // *OMICS A Journal of Integrative Biology*. – 2012. – Vol. 16, № 5. – P. 284-287.

37. Coexpression analysis of human genes across many microarray data sets / H.K. Lee, [et al.] // *Genome Research*. – 2004. – Vol. 14, № 6. – P. 1085-1094.

38. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data / B. Wu, [et al.] // *Bioinformatics*. – 2003. – Vol. 19, № 13. – P. 1636-1643.

39. Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma / M. Huang, [et al.] // *Oncotarget*. – 2016. – Vol. 7, № 30. – P. 47186-47200.

40. Comprehensive bioinformatics analysis of the characterization and determination underlying mechanisms of over-expression and co-expression of

genes residing on 20q in colorectal cancer / D. Li, [et al.] // *Oncotarget*. – 2017. – Vol. 8, № 45. – P. 78642-78659.

41. Comprehensive molecular characterization of gastric adenocarcinoma / A.J. Bass, [et al.] // *Nature*. – 2014. – Vol. 513, № 7517. – P. 202-209.

42. Comprehensive molecular characterization of human colon and rectal cancer / D.M. Muzny, [et al.] // *Nature*. – 2012. – Vol. 487, № 7407. – P. 330-337.

43. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer / A.G. Robertson, [et al.] // *Cell*. – 2017. – Vol. 171, № 3. – P. 540-556.

44. Comprehensive molecular characterization of urothelial bladder carcinoma / J.N. Weinstein, [et al.] // *Nature*. – 2014. – Vol. 507, № 7492. – P. 315-322.

45. Comprehensive molecular portraits of human breast tumours / D.C. Koboldt, [et al.] // *Nature*. – 2012. – Vol. 490, № 7418. – P. 61-70.

46. Conley, M.E. Diagnostic Criteria for Primary Immunodeficiencies / M.E. Conley, L.D. Notarangelo, A. Etzioni // *Clinical Immunology*. – 1999. – Vol. 93, № 3. – P. 190-197.

47. Construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma / S. Yang, [et al.] // *Oncotarget*. – 2016. – Vol. 7, № 52. – P. 85728-85740.

48. Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer / E.D. Wiklund, [et al.] // *International Journal of Cancer*. – 2011. – Vol. 128, № 6. – P. 1327-1334.

49. Csardi, G. The igraph software package for complex network research / G. Csardi, T. Nepusz // *InterJournal, Complex Systems*. – 2006. – Vol. 1695, № 5. – P. 1-9.

50. Cunningham-Rundles, C. How I treat common variable immune deficiency / C. Cunningham-Rundles // *Blood*. – 2010. – Vol. 116, № 1. – P. 7-15.

51. CVID-associated tumors: Czech nationwide study focused on epidemiology, immunology, and genetic background in a cohort of patients with CVID / P. Kralickova, [et al.] // *Frontiers in Immunology*. – 2019. – Vol. 10, № JAN. – P. 3135.

52. Cytoscape: A software Environment for integrated models of biomolecular interaction networks / P. Shannon, [et al.] // *Genome Research*. – 2003. – Vol. 13, № 11. – P. 2498-2504.

53. Delayed diagnosis and complications of predominantly antibody deficiencies in a cohort of Australian adults / C.A. Slade, [et al.] // *Frontiers in Immunology*. – 2018. – Vol. 9. – P. 694.

54. Developing a Multi-Dose Computational Model for Drug-induced Hepatotoxicity Prediction based on Toxicogenomics Data / R. Su, [et al.] // *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. – 2018. – Vol. 16, № 4.– P. 1.

55. Development of cancer in patients with primary immunodeficiencies / K. Salavoura, [et al.] // *Anticancer Research*. – 2008. – Vol. 28, № 2B. – P. 1263-1269.

56. Disturbed canonical nuclear factor of κ light chain signaling in B cells of patients with common variable immunodeficiency / B. Keller, [et al.] // *Journal of Allergy and Clinical Immunology*. – 2017. – Vol. 139, № 1. – P. 220-231.

57. DNA physical properties outperform sequence compositional information in classifying nucleosome-enriched and -depleted regions / G. Liu, [et al.] // *Genomics*. – 2019. – Vol. 111, № 5. – P. 1167-1175.

58. Dong F. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia / F. Dong, L. Hoefsloot // *Proceedings of the National Academy of Sciences*. – 1994. – Vol. 91, № 10. – P. 4480-4484.

59. Dysfunctional BLK in common variable immunodeficiency perturbs B-cell proliferation and ability to elicit antigen-specific CD4⁺ T-cell help /

E.B. Compeer, [et al.] // *Oncotarget*. – 2015. – Vol. 6, № 13. – P. 10759-10771.

60. Epidemiology of bladder cancer: a systematic review and contemporary update of risk factors in 2018 / M. Cumberbatch, [et al.] // *European urology*. – 2018. – Vol. 74, № 6. – P. 784-795.

61. Epithelial-mesenchymal transition, a novel target of sulforaphane via COX-2/MMP2, 9/Snail, ZEB1 and miR-200c/ZEB1 pathways in human bladder cancer cells / Y. Shan, [et al.] // *Journal of Nutritional Biochemistry*. – 2013. – Vol. 24, № 6. – P. 1062-1069.

62. Evaluation of miR-141, miR-200c, miR-30b expression and clinicopathological features of bladder cancer / A. Mahdavinezhad, [et al.] // *International journal of molecular and cellular medicine*. – 2015. – Vol. 4, № 1. – P. 32.

63. Exome Sequencing Analysis Reveals Variants in Primary Immunodeficiency Genes in Patients with Very Early Onset Inflammatory Bowel Disease / J.R. Kelsen, [et al.] // *Gastroenterology*. – 2015. – Vol. 149, № 6. – P. 1415-1424.

64. Familial risk and heritability of cancer among twins in Nordic countries / L. Mucci, [et al.] // *Jama*. – 2016. – Vol. 315, № 1. – P.68-76.

65. Floyd, E. Development and Use of Biomarkers in Oncology Drug Development / E. Floyd, T.M. Mcshane // *Toxicologic Pathology*. – 2004. – Vol. 32, № suppl_1.– P. 106-115.

66. Friedrich, M.J. Immunotherapy 2.0: Improving the Response to Checkpoint Inhibitors / M.J. Friedrich // *Jama*. – 2019. – Vol. 321, № 2. – P. 131-133.

67. Frost, H.R. A multi-omics approach for identifying important pathways and genes in human cancer / H.R. Frost, C.I. Amos // *BMC Bioinformatics*. – 2018. – Vol. 19, № 1. – P. 479.

68. Gene co-expression analysis for functional classification and gene-disease predictions / S. van Dam, [et al.] // *Briefings in bioinformatics*. – 2018. –

Vol. 19, № 4. – P. 575-592.

69. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles / A. Subramanian, [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 2005. – Vol. 102, № 43. – P. 15545-15550.

70. Gene set enrichment analysis: Performance evaluation and usage guidelines / J.H. Hung, [et al.] // Briefings in Bioinformatics. – 2012. – Vol. 13, № 3. – P. 281-291.

71. Genes associated with common variable immunodeficiency: One diagnosis to rule them all? / D.J.A. Bogaert, [et al.] // Journal of Medical Genetics. – 2016. – Vol. 53, № 9. – P. 575-590.

72. Germeshausen M. Implications of Mutations in Hematopoietic Growth Factor Receptor Genes in Congenital Cytopenias / M. Germeshausen, M. Ballmaier, K. Welte // Annals of the New York Academy of Sciences. – 2001. – Vol. 938, № 1. – P. 305-321.

73. Global overview of primary immunodeficiencies: a report from Jeffrey Modell Centers worldwide focused on diagnosis, treatment, and discovery / V. Modell, [et al.] // Immunologic Research. – 2014. – Vol. 60, № 1. – P. 132-144.

74. Guidelines for genetic studies in single patients: Lessons from primary immunodeficiencies / J.L. Casanova, [et al.] // Journal of Experimental Medicine. – 2014. – Vol. 211, № 11. – P. 2137-2149.

75. Guidelines on non-muscle-invasive (TA, T1, CIS) bladder cancer / Babjuk, M. [et al.] // Eur Urol. – 2011. – Vol. 59. – № 4. – P. 584-594.

76. Haploinsufficiency of the NF- κ B1 Subunit p50 in Common Variable Immunodeficiency / M. Fliegau, [et al.] // American Journal of Human Genetics. – 2015. – Vol. 97, № 3. – P. 389-403.

77. Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome / M. Alders, [et al.] // Human Genetics. – 2014. – Vol. 133, № 9. – P. 1161-1167.

78. Hennig, C. *fpc: Flexible Procedures for Clustering* / C. Hennig // R package version. – 2010. – Vol. 2, № 2. – P. 0-3.
79. Hermjakob, H. *IntAct: an open source molecular interaction database* / H. Hermjakob // *Nucleic Acids Research*. – 2004. – Vol. 32, № suppl_1. – P. 452D - 455.
80. High-dimensional variable selection for survival data / H. Ishwaran, [et al.] // *Journal of the American Statistical Association*. – 2010. – Vol. 105, № 489. – P. 205-217.
81. Hong, Y.M. Economic impact of tumor markers in bladder cancer surveillance / Y.M. Hong, K.R. Loughlin // *Urology*. – 2008. – Vol. 71, № 1. – P. 131-135.
82. Horvath, S. Geometric interpretation of gene coexpression network analysis / S. Horvath, J. Dong // *PLoS Computational Biology*. – 2008. – Vol. 4, № 8. – P. e1000117.
83. Human Protein Reference Database - 2009 update / T.S. Keshava Prasad, [et al.] // *Nucleic Acids Research*. – 2009. – Vol. 37, № suppl_1. – P. 767-772.
84. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer / F. Crea, [et al.] // *Oncotarget*. – 2014. – Vol. 5, № 3. – P. 764-774.
85. Identification of Distinct Basal and Luminal Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline Chemotherapy / W. Choi, [et al.] // *Cancer Cell*. – 2014. – Vol. 25, № 2. – P. 152-165.
86. Identification of druggable cancer driver genes amplified across TCGA datasets / Y. Chen, [et al.] // *PLoS ONE*. – 2014. – Vol. 9, № 5. – P. e98293.
87. Identification of ovarian cancer driver genes by using module network integration of multi-omics data / O. Gevaert, [et al.] // *Interface Focus*. – 2013. – Vol. 3, № 4. – P. 20130013.
88. Identification of prognostic gene signatures of glioblastoma: A study

based on TCGA data analysis / Y.W. Kim, [et al.] // *Neuro-Oncology*. – 2013. – Vol. 15, № 7. – P. 829-839.

89. Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy / R. Seiler, [et al.] // *European Urology*. – 2017. – Vol. 72, № 4. – P. 544-554.

90. InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation / K. Breuer, [et al.] // *Nucleic Acids Research*. – 2013. – Vol. 41, № D1. – P. D1228-D1233.

91. Instruct: A database of high-quality 3D structurally resolved protein interactome networks / M.J. Meyer, [et al.] // *Bioinformatics*. – 2013. – Vol. 29, № 12. – P. 1577-1579.

92. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma / B.J. Raphael, [et al.] // *Cancer Cell*. – 2017. – Vol. 32, № 2. – P. 185-203.

93. Integrated multi-omics analysis of oligodendroglial tumours identifies three subgroups of 1p/19q co-deleted gliomas / A. Kamoun, [et al.] // *Nature Communications*. – 2016. – Vol. 19, № 7. – P. 11263.

94. Integrating coexpression networks with GWAS to prioritize causal genes in maize / R.J. Schaefer, [et al.] // *Plant Cell*. – 2018. – Vol. 30, № 12. – P. 2922-2942.

95. Integrative analysis of lncRNAs and miRNAs with coding RNAs associated with ceRNA crosstalk network in triple negative breast cancer / N. Yuan, [et al.] // *OncoTargets and therapy*. – 2017. – Vol. 10. – P. 5883–5897.

96. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks / X. Zeng, [et al.] // *Briefings in bioinformatics*. – 2015. – Vol. 17, № 2. – P. 193-203.

97. Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome / M.R. Aure, [et al.] // *Breast Cancer*

Research. – 2017. – Vol. 19, № 1. – P. 1-18.

98. International Consensus Document (ICON): Common Variable Immunodeficiency Disorders / F.A. Bonilla, [et al.] // *The Journal of Allergy and Clinical Immunology: In Practice*. – 2016. – Vol. 4, № 1. – P. 38-59.

99. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity / C. Picard, [et al.] // *Journal of Clinical Immunology*. – 2018. – Vol. 38, № 1. – P. 96-128.

100. Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer / W. Choi, [et al.] // *Nature Reviews Urology*. – 2014. – Vol. 11, № 7. – P. 400.

101. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology / J.S. Damrauer, [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2014. – Vol. 111, № 8. – P. 3110-3115.

102. Invasive bladder cancer: genomic insights and therapeutic promise / J. Kim, [et al.] // *Clin Cancer Res*. – 2015. – Vol. 21, № 20. – P. 4514-4524.

103. Itan, Y. Novel primary immunodeficiency candidate genes predicted by the human gene connectome / Y. Itan, J.L. Casanova // *Frontiers in Immunology*. – 2015. – Vol. 6. – P. 142.

104. Johnston, R.J. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans* / R.J. Johnston, O. Hobert // *Nature*. – 2003. – Vol. 426, № 6968. – P. 845-849.

105. Kienzler, A.K. The role of genomics in common variable immunodeficiency disorders / A.K. Kienzler, C.E. Hargreaves, S.Y. Patel // *Clinical and Experimental Immunology*. – 2017. – Vol. 188, № 3. – P. 326-332.

106. Large scale comparison of gene expression levels by microarrays and rnaseq using TCGA data / Y. Guo, [et al.] // *PLoS ONE*. – 2013. – Vol. 8, № 8. – P. e71462.

107. Li, H. Fast and accurate short read alignment with Burrows-Wheeler

transform / H. Li, R. Durbin // *Bioinformatics*. – 2009. – Vol. 25, № 14. – P. 1754-1760.

108. Liaw, A. Classification and regression by randomForest / A. Liaw, M. Wiener // *R news*. – 2002. – Vol. 2, № 3. – P. 18-22.

109. Lim M.S. The molecular pathology of primary immunodeficiencies / M.S. Lim, K.S.J. Elenitoba-Johnson // *The Journal of molecular diagnostics*. – 2004. – Vol. 6, № 2. – P. 59.

110. Lin, L.J. Clinical and immunological features of common variable immunodeficiency in China / L.J. Lin, Y.C. Wang, X.M. Liu // *Chinese Medical Journal*. – 2015. – Vol. 128, № 3. – P. 310-315.

111. LincRNAs: Systemic Computational Identification and Functional Exploration / H. Hu, [et al.] // *Current Bioinformatics*. – 2017. – Vol. 12, № 1. – P. 34-42.

112. Liu, G. Integrated multiple “-omics” data reveal subtypes of hepatocellular carcinoma / G. Liu, C. Dong, L. Liu // *PLoS ONE*. – 2016. – Vol. 11, № 11. – P. 1-15.

113. Long non-coding RNA urothelial cancer-associated 1 promotes bladder cancer cell migration and invasion by way of the hsa-miR-145-ZEB1/2-FSCN1 pathway / M. Xue, [et al.] // *Cancer Science*. – 2016. – Vol. 107, № 1. – P. 18-27.

114. Long noncoding RNA in prostate, bladder, and kidney cancer / E. Martens-Uzunova, [et al.] // *European urology*. – 2014. – Vol. 65, № 6. – P. 1140-1151.

115. Loss of ADAMTS3 activity causes Hennekam lymphangiectasia-lymphedema syndrome 3 / P. Brouillard, [et al.] // *Human molecular genetics*. – 2017. – Vol. 26, № 21. – P. 4095-4104.

116. Lymphedema-lymphangiectasia-mental retardation (Hennekam) syndrome: A review / I.D.C. Van Balkom, [et al.] // *American journal of medical genetics*. – 2002. – Vol. 112, № 4. – P. 412-421.

117. Machine learning applications in cancer prognosis and prediction / K. Kourou, [et al.] // Computational and structural biotechnology journal. – 2015. – Vol. 13. – P. 8-17.

118. McCusker C. Primary immunodeficiency / C. McCusker, J. Upton, R. Warrington // Allergy, Asthma & Clinical Immunology. – 2018. – Vol. 14, № S2. – P. 61.

119. Mccusker, C. IMMUNOLOGY Primary immunodeficiency / C. Mccusker, R. Warrington // Allergy, Asthma & Clinical Immunology. – 2011. – Vol. 7, № suppl_1. – P. S11.

120. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use / V. Dadhania, [et al.] // EBioMedicine. – 2016. – Vol. 12. – P. 105-117.

121. Michalak, P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes / P. Michalak // Genomics. – 2008. – Vol. 91, № 3. – P. 243-248.

122. MicroRNAs: A new paradigm in the treatment and diagnosis of heart failure? / V. Oliveira-Carvalho, [et al.] // Arquivos Brasileiros de Cardiologia. – 2012. – Vol. 98, № 4. – P. 362-369.

123. MINT: The Molecular INTeraction database / A. Chatr-aryamontri, [et al.] // Nucleic Acids Research. – 2006. – Vol. 35, № suppl_1. – P. D572-D574.

124. MiR-200c Inhibits invasion, migration and proliferation of bladder cancer cells through down-regulation of BMI-1 and E2F3 / L. Liu, [et al.] // Journal of Translational Medicine. – 2014. – Vol. 12, № 1. – P. 305.

125. miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110 / R. Song, [et al.] // Nature. – 2014. – Vol. 510. – № 7503. – P. 115.

126. miRClassify: An advanced web server for miRNA family classification and annotation / Q. Zou, [et al.] // Computers in Biology and Medicine. – 2014. – Vol. 45, № 1. – P. 157-160.

127. MiRWalk - Database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes / H. Dweep, [et al.] // *Journal of Biomedical Informatics*. – 2011. – Vol. 44, № 5. – P. 839-847.
128. Molecular markers in bladder cancer: Novel research frontiers / F. Sanguedolce, [et al.] // *Critical Reviews in Clinical Laboratory Sciences*. – 2015. – Vol. 52, № 5. – P. 242-255.
129. Molecular portraits of human breast tumours / C.M. Perou, [et al.] // *Nature*. – 2000. – Vol. 406, № 6797. – P. 747-752.
130. Mooney, M.A. Gene set analysis: A step-by-step guide / M.A. Mooney, B. Wilmot // *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*. – 2015. – Vol. 168, № 7. – P. 517-527.
131. Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma / F. Chen, [et al.] // *Cell Reports*. – 2016. – Vol. 14, № 10. – P. 2476-2489.
132. Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans / M. Alders, [et al.] // *Nature Genetics*. – 2009. – Vol. 41, № 12. – P. 1272-1274.
133. Nagata, M. Molecular biomarkers in bladder cancer: novel potential indicators of prognosis and treatment outcomes / M. Nagata, S. Muto, S. Horie // *Disease Markers*. – 2016. – Vol. 2016. – P. 1-5.
134. Naturally occurring mutation affecting the MyD88-binding site of TNFRSF13B impairs triggering of class switch recombination / M.B. Almejun, [et al.] // *European Journal of Immunology*. – 2013. – Vol. 43, № 3. – P. 805-814.
135. Neoadjuvant PD-1 blockade in resectable lung cancer / P.M. Forde, [et al.] // *New England Journal of Medicine*. – 2018. – Vol. 378, № 21. – P. 1976-1986.
136. Network-based approach to prediction and population-based validation of in silico drug repurposing / F. Cheng, [et al.] // *Nature Communications*. – 2018. – Vol. 9, № 1. – P. 1-12.
137. Newborn screening: Toward a uniform screening panel and system -

Executive summary / M.S. Watson, [et al.] // *Pediatrics*. – 2006. – Vol. 117, № 5. – P. S296-S307.

138. Novel RAG1 mutation in a case of severe combined immunodeficiency / J. Zhang, [et al.] // *Pediatrics*. – 2005. – Vol. 116, № 3. – P. e445-e449.

139. Ochs, H.D. Common variable immunodeficiency (CVID): New genetic insight and unanswered questions / H.D. Ochs // *Clinical and Experimental Immunology*. – 2014. – Vol. 178, № S1. – P. 5-6.

140. Orang, A.V. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation / A.V. Orang, R. Safaralizadeh, M. Kazemzadeh-Bavili // *International journal of genomics*. – 2014. – Vol. 2014.

141. Ortutay, C. Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies / C. Ortutay, M. Vihinen // *Nucleic acids research*. – 2008. – Vol. 37, № 2. – P. 622-628.

142. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy / R. Cristescu, [et al.] // *Science*. – 2018. – Vol. 362, № 6411. – P. eaar3593.

143. Paradis, E. APE: Analyses of phylogenetics and evolution in R language / E. Paradis, J. Claude, K. Strimmer // *Bioinformatics*. – 2004. – Vol. 20, № 2. – P. 289-290.

144. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy / B. Lo, [et al.] // *Science*. – 2015. – Vol. 349, № 6246. – P. 436-440.

145. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer / A. Prat, [et al.] // *Breast Cancer Research*. – 2010. – Vol. 12, № 5. – P. R68.

146. PhosphoSitePlus: A comprehensive resource for investigating the

structure and function of experimentally determined post-translational modifications in man and mouse / P.V. Hornbeck, [et al.] // *Nucleic Acids Research*. – 2012. – Vol. 40, № D1. – P. D261-D270.

147. PIDO: the primary immunodeficiency disease ontology / N. Adams, [et al.] // *Bioinformatics*. – 2011. – Vol. 27, № 22. – P. 3193-3199.

148. PINA v2.0: Mining interactome modules / M.J. Cowley, [et al.] // *Nucleic Acids Research*. – 2012. – Vol. 40, № D1. – P. D862-D865.

149. PMTDS: a computational method based on genetic interaction networks for Precision Medicine Target-Drug Selection in cancer / V. Vasudevaraja, [et al.] // *Quantitative Biology*. – 2017. – Vol. 5, № 4. – P. 380-394.

150. Prediction of candidate primary immunodeficiency disease genes using a support vector machine learning approach. / S. Keerthikumar, [et al.] // *DNA research*. – 2009. – Vol. 16, № 6. – P. 345-351.

151. Prediction of central nervous system embryonal tumour outcome based on gene expression / S.L. Pomeroy, [et al.] // *Nature*. – 2002. – Vol. 415, № 6870. – P.436.

152. Prevalence and co-occurrence of actionable genomic alterations in high-grade bladder cancer / G. Iyer, [et al.] // *Journal of Clinical Oncology*. – 2013. – Vol. 31, № 25. – P. 3133-3140.

153. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy / P. Sharma, [et al.] // *Cell*. – 2017. – Vol. 168, № 4. – P. 707-723.

154. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015 / C. Picard, [et al.] // *Journal of Clinical Immunology*. – 2015. – Vol. 35, № 8. – P. 696-726.

155. Primary vs. secondary antibody deficiency: Clinical features and infection outcomes of immunoglobulin replacement / S.S. Duraisingham, [et al.] // *PLoS ONE*. – 2014. – Vol. 9, № 6. – P. e100324.

156. Prognostic factors in survival of patients with stage Ta and T1 bladder urothelial tumors: the role of G1-S modulators (p53, p21Waf1, p27Kip1, cyclin D1, and cyclin D3), proliferation index, and clinicopathologic parameters / A. Lopez-Beltran, [et al.] // *American journal of clinical pathology*. – 2004. – Vol. 122, № 3. – P. 444-452.

157. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma / C.G. Clemente, [et al.] // *Cancer*. – 1996. – Vol. 77, № 7. – P. 1303-1310.

158. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy / F. Cheng, [et al.] // *Oncotarget*. – 2014. – Vol. 5, № 11. – P. 3697-3710.

159. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus / Q. Sun, [et al.] // *Science*. – 2013. – Vol. 340, № 6132. – P. 619-621.

160. Rao, B.S. Hennekam lymphangiectasia syndrome / B.S. Rao, M. Sree Vani, B.S. Kanth // *International Journal of Research in Medical Sciences*. – 2015. – Vol. 3, № 2. – P. 516-519.

161. Revisiting human primary immunodeficiencies / J.L. Casanova, [et al.] // *Journal of internal medicine*. – 2008. – Vol. 264, № 2. – P. 115-127.

162. Ricketts, C.J. Tumor-specific hypermethylation of epigenetic biomarkers, including SFRP1, predicts for poorer survival in patients from the TCGA kidney renal clear cell carcinoma (KIRC) project / C.J. Ricketts, V.K. Hill, W.M. Linehan // *PLoS ONE*. – 2014. – Vol. 9, № 1. – P. e85621.

163. Rius, M. Epigenetic cancer therapy: Rationales, targets and drugs. Vol. 31 / M. Rius, F. Lyko // *Oncogene*. – 2012. – Vol. 31, № 39. – P. 4257.

164. RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics / M.G. Best, [et al.] // *Cancer Cell*. – 2015. – Vol. 28, № 5. – P. 666-676.

165. RNF12 initiates X-chromosome inactivation by targeting REX1 for

degradation / C. Gontan, [et al.] // *Nature*. – 2012. – Vol. 485, № 7398. – P. 386.

166. Samarghitean, C. Bioinformatics services related to diagnosis of primary immunodeficiencies / C. Samarghitean, M. Vihinen // *Current Opinion in Allergy and Clinical Immunology*. – 2009. – Vol. 9, № 6. – P. 531-536.

167. Sayed, D. Micromas in development and disease / D. Sayed, M. Abdellatif // *Physiological Reviews*. – 2011. – Vol. 91, № 3. – P. 827-887.

168. Second-line single-agent versus doublet chemotherapy as salvage therapy for metastatic urothelial cancer: A systematic review and meta-analysis / D. Raggi, [et al.] // *Annals of Oncology*. – 2016. – Vol. 27, № 1. – P. 49-61.

169. Seidel, M.G. Autoimmune and other cytopenias in primary immunodeficiencies: Pathomechanisms, novel differential diagnoses, and treatment / M.G. Seidel // *Blood*. – 2014. – Vol. 124, № 15. – P. 2337-2344.

170. Şenbabaoğlu, Y. Critical limitations of consensus clustering in class discovery / Y. Şenbabaoğlu, G. Michailidis, J.Z. Li // *Scientific Reports*. – 2014. Vol. 4. – P. 6207.

171. Siegel, R.L. Cancer statistics, 2016. - PubMed - NCBI / R.L. Siegel, K.D. Miller, A. Jemal // *CA: A Cancer Journal for Clinicians*. – 2016. – Vol. 66, № 1. – P. 7-30.

172. Signalink 2 - a signaling pathway resource with multi-layered regulatory networks / D. Fazekas, [et al.] // *BMC systems biology*. – 2013. – Vol. 7. – P. 7.

173. Stark, C. BioGRID: a general repository for interaction datasets / C. Stark // *Nucleic Acids Research*. – 2006. – Vol. 34, № 90001. – P. D535-D539.

174. Stem cell division is regulated by the microRNA pathway / S.D. Hatfield, [et al.] // *Nature*. – 2005. – Vol. 435, № 7044. – P. 974.

175. STRING 8—a global view on proteins and their functional interactions in 630 organisms 2008 / L.J. Jensen, [et al.] // *Nucleic acids research*. – 2008. – Vol. 37, № suppl_1. – P. 412-416.

176. Superficial bladder cancer: an update on etiology, molecular

development, classification, and natural history,. / E. Pasin, [et al.] // *Reviews in urology*. – 2008. – Vol. 10, № 1. – P. 31-43.

177. Syndrome de Hennekam / T. Erkan, [et al.] // *Archives de Pediatrie*. – 1998. – Vol. 5, № 12. – P. 1344-1346.

178. Takahashi, N. Common variable immunodeficiency / N. Takahashi, T. Morio // *Japanese Journal of Clinical Immunology*. – 2008. – Vol. 31, № 1. – P. 9-16.

179. The economics of bladder cancer: Costs and considerations of caring for this disease / R.S. Svatek, [et al.] // *European urology*. – 2014. – Vol. 66, № 2. – P. 253-262.

180. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data / A. McKenna, [et al.] // *Genome Research*. – 2010. – Vol. 20, № 9. – P. 1297-1303.

181. The global epidemiology of bladder cancer: a joinpoint regression analysis of its incidence and mortality trends and projection / M. Wong, [et al.] // *Scientific reports*. – 2018. – Vol. 8, № 1. – P. 1129.

182. The human gene connectome as a map of short cuts for morbid allele discovery / Y. Itan, [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2013. – Vol. 110, № 14. – P. 5558-5563.

183. The multilayered complexity of ceRNA crosstalk and competition / Y. Tay, [et al.] // *Nature*. – 2014. – Vol. 505, № 7483. – P. 344.

184. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation / J. Chen, [et al.] // *Nature genetics*. – 2006. – Vol. 38, № 2. – P. 228.

185. The Sequence Alignment/Map format and SAMtools / H. Li, [et al.] // *Bioinformatics*. – 2009. – Vol. 25, № 16. – P. 2078-2079.

186. The variant call format and VCFtools / P. Danecek, [et al.] // *Bioinformatics*. – 2011. – Vol. 27, № 15. – P. 2156-2158.

187. Therneau, T. A Package for Survival Analysis in S. version 2.38 /

T. Therneau // R package. – 2015.

188. Tomczak, K. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge / K. Tomczak, P. Czerwińska, M. Wiznerowicz // Contemporary oncology. – 2015. – Vol. 19, № 1A. – P. A68.

189. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial / S. Loi, [et al.] // Annals of Oncology. – 2014. – Vol. 25, № 8. – P. 1544-1550.

190. Tumor size predicts the survival of patients with pathologic stage t2 bladder carcinoma / L. Cheng, [et al.] // Cancer. – 1999. – Vol. 85, № 12. – P. 2638-2647.

191. Tumour-infiltrating lymphocytes are correlated with higher expression levels of PD-1 and PD-L1 in early breast cancer / A. Kitano, [et al.] // ESMO Open. – 2017. – Vol. 2, № 2. – P. e000150.

192. Uses of next-generation sequencing technologies for the diagnosis of primary immunodeficiencies / M. Seleman, [et al.] // Frontiers in immunology. – 2017. – Vol. 8. – P. 847.

193. Wang, K. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data / K. Wang, M. Li, H. Hakonarson // Nucleic Acids Research. – 2010. – Vol. 38, № 16. – P. e164-e164.

194. Wang, K. A brief procedure for big data analysis of gene expression / K. Wang, W. Wang, M. Li // Animal Models and Experimental Medicine. – 2018. – Vol. 1, № 3. – P. 189-193.

195. Wei, L. M6APred-EL: A Sequence-Based Predictor for Identifying N6-methyladenosine Sites Using Ensemble Learning / L. Wei, H. Chen, R. Su // Molecular Therapy - Nucleic Acids. – 2018. – Vol. 12. – P. 635-644.

196. Wei, S.C. Fundamental mechanisms of immune checkpoint blockade therapy / S.C. Wei, C.R. Duffy, J.P. Allison // Cancer discovery. – 2018. – Vol. 8, № 9. – P. 1069-1086.

197. Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease / D.N. Cooper, [et al.] // *Human genetics*. – 2013. – Vol. 132, № 10. – P. 1077-1130.

198. Wilkerson, M.D. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking / M.D. Wilkerson, D.N. Hayes // *Bioinformatics*. – 2010. – Vol. 26, № 12. – P. 1572-1573.

199. Williams E.J.B. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana* / E.J.B. Williams, D.J. Bowles // *Genome Research*. – 2004. – Vol. 14, № 6. – P. 1060-1067.

200. Williams S.G. Molecular markers for diagnosis, staging, and prognosis of bladder cancer / S.G. Williams, M. Buscarini, J.P. Stein // *Oncology*. – 2001. – Vol. 15, № 11. – P. 1461.

201. Yuan Y. CNNC: Convolutional neural networks for co-expression analysis / Y. Yuan, Z. Bar-Joseph // *BioRxiv*. – 2018. – P. 365007.

202. ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8⁺ T cell fates / T. Guan, [et al.] // *Journal of Experimental Medicine*. – 2018. – Vol. 215, № 4. – P. 1153-1168.

203. Zhao Y. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis / Y. Zhao, E. Samal, D. Srivastava // *Nature*. – 2005. – Vol. 436, № 7048. – P. 214.

204. Zhong Z. Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma / Z. Zhong, M. Lv, J. Chen // *Scientific Reports*. – 2016. – Vol. 6. – P. 30919.