Shinwari Khyber

NOVEL GENE VARIANTS IN THE EVALUATION OF INBORN ERRORS OF IMMUNITY: RBCK1 DEFICIENCY, CONGENITAL NEUTROPENIA, HENNEKAM SYNDROME

3.2.7 - Immunology

ABSTRACT

Of dissertation

for the science degree

of candidate of biological sciences

Ekaterinburg – 2023

The work was performed at the Immunochemistry Department of the Chemical Engineering Institute of the Federal State Educational Institution of Higher Professional Education "Ural Federal University named after the first President of Russia B.N. Yeltsin" and at the Laboratory of Inflammation Immunology of the Federal State Budgetary Scientific Institution of the Institute of Immunology and Physiology, Ural Branch of RAS.

Supervisors:

Doctor of Medical Sciences, Professor, Honored Scientist of Russian Federation

Tuzankina Irina Aleksandrovna

Academician of the RAS, Doctor of Medical Sciences, Professor

Chereshnev Valery Alexandrovich

Official opponents:

Doctor of Medicine, Professor, Professor of the Department of Organization and Management in the Circulation of Medicines of the Institute of Professional Education of the First Moscow State Medical University named after I.M. Sechenov. I.M. Sechenov First Moscow State Medical University of the Russia (Sechenov University)

Kozlov Ivan Genrikhovich

Doctor of Biology, Director of the Center of Immunology and Cell Biotechnology, Professor of the Fundamental Medicine Department of the Medical Institute of the Federal State Educational Institution of Higher Professional Education "Immanuel Kant Baltic Federal University

Litvinova Larisa Sergeevna

Leading organization - Federal Budgetary Scientific Institution "St. Petersburg Pasteur Institute of Epidemiology and Microbiology" of the Federal Service for Surveillance for Consumer Rights Protection and Human Welfare (St. Petersburg, Russia).

The defense will be held on June 27, 2023 at _____ at the meeting of the Council for thesis defense for the degree of Candidate of Sciences, for the degree of Doctor of Sciences 24.1.063.01 on the basis of the Institute of Philosophy of Ural Branch of RAS (620049, Ekaterinburg, Pervomaiskaya str., 106).

The thesis and the abstract are available at the Central Scientific Library of the Ural Branch of RAS and at the site of the Ural Branch of RAS - http://iip.uran.ru, with the abstract - at the site of the Higher Attestation Commission of the Russian Federation http://vak2.ed.gov.ru.

Author's abstract sent "___" _____ 2023

Scientific Secretary of the Council 24.1.063.01
at the Institute of Immunology and Physiology UB RAS,
Candidate of biological science   Yu.A. Zhuravleva

GENERAL CHARACTERISTICS OF WORK

Relevance and degree of development of the research topic. The immune system is a complex biological system that has evolved to fight against foreign antigens, recognize both external and internal antigens, destroy infected and abnormally developing cells, and maintain tolerance to autoantigens and commensal microbiota, thus fulfilling its crucial role in the preservation of the species.

However, inborn errors of Immunity (IEI) or primary immunodeficiencies (PID) can increase susceptibility to infections, autoimmune processes, autoinflammatory diseases, malignancies, and allergies. The main causes of these disorders are genetic changes occurring at both the genome and individual gene level, which encode protein molecules involved in immune mechanisms.

Although, inborn errors of Immunity (IEI) were once considered rare diseases, their collective impact can be significant. Furthermore, due to advancements in diagnosis and the development of next-generation sequencing technologies, the reported prevalence of primary immunodeficiencies (PIDs) has increased to 40 per 100,000 population in recent years (Rubin Z. et al., 2018).

To develop new methods for the diagnosis and treatment of immunopathologies, a thorough understanding of the implementation of immune functions in the body is necessary. The emergence of high-performance biological methods has led to an unprecedented understanding of the molecular mechanisms underlying the functioning of the immune system and its interconnections with other body systems. However, the analysis of parameters of enormous complexity, covering several orders of spatial and temporal scales, can only be achieved through computational immunology. In particular, computational approaches are necessary for processing and modeling large immunological datasets.

In our work, we utilized bioinformatics analysis methods to investigate individual syndromes that arise from inborn errors of immunity. We selected three pathologies for this purpose: congenital neutropenia (one of the most common forms of PID), Hennekam syndrome (one of the rarest), and RBCK1 deficiency, which is associated with autoinflammatory syndromes of PID but with increased susceptibility to pyogenic infections. RBCK1 deficiency was first described in 2012 (Boisson B. et al., 2011), Hennekam syndrome in 1989 (Hennekam R.C. et al., 1989), and the first discoveries in the genetics of congenital neutropenia date back to 1999 (Horwitz M. et al., 1999). However, the diagnosis of each of these syndromes remains challenging, new gene variants that lead to the phenotypes of these diseases continue to be identified, and the exact mechanisms underlying the pathology of Hennekam syndrome and RBCK1 deficiency remain the subject of ongoing discussion.

In addition to identifying gene variants, it is crucial to demonstrate their impact on the final product, i.e., the protein, which can be evaluated using *in silico* tools to determine its destabilization. This approach would expedite the assessment of the gene variant's pathogenicity and facilitate the inclusion of identified variants in the list of causally significant factors, accelerating diagnosis and guiding the development of pathogenetic or gene therapy methods. These goals represent the ultimate objective of investigating human congenital pathology.

In our study, we aimed to investigate significant pathophysiological mechanisms underlying the development of certain types of immune-dependent diseases using bioinformatics analysis methods. To accomplish this, we selected subjects with congenital immune disorders, such as RBCK1 deficiency autoinflammatory syndrome, congenital neutropenia, and Hennekam syndrome as models of immune-dependent pathology.

RBCK1 deficiency is an autoinflammatory syndrome that is associated with an increased susceptibility to infections. The RBCK1 protein, also known as HOIL-1, is involved in the assembly of the linear ubiquitin chain complex (LUBAC). It plays a crucial role in the activation of the classical NF-kB signaling pathway, inflammation prevention, and participation in apoptosis (Gerlach B. et al., 2011). The deficiency of RBCK1 is characterized by glycogen metabolism disorders, which leads to the accumulation of glycogen in the muscles, a condition known as amylopectinosis. Patients with RBCK1 deficiency exhibit a wide range of clinical manifestations, including fever, infectious syndrome (various skin inflammations, recurrent bacterial infections, and in severe cases sepsis), as well as myopathies, cardiomyopathies, and encephalopathies (Nilsson J et al., 2013).

Congenital neutropenia is a condition that can occur in various inherited immune system disorders. Patients with congenital neutropenia suffer from acute and chronic infectious conditions such as otitis, gingivitis, skin infections, pneumonia, deep abscesses, and sepsis. The disease can present in the neonatal period, and without appropriate treatment, symptoms can persist throughout a patient's lifetime (Skokowa J. et al., 2017).

The causes of congenital neutropenia can be defects in the maturation and function of neutrophils, immune dysregulation syndromes (such as various hemophagocytic lymphohistiocytosis), some severe combined immunodeficiencies (such as reticular dysgenesis and PAC2 activation defect), as well as primary autoimmune neutropenia at different stages of neutrophil development. Typically, patients with severe congenital neutropenia require antimicrobial prophylaxis and treatment with granulocyte colony-stimulating factor. Radical cure is impossible without hematopoietic stem cell transplantation (Skokowa J. et al., 2017).

Currently, there are more than 30 congenital immune system errors (or primary immunodeficiencies) in which neutropenia is observed. Although each condition in isolation occurs rarely, the overall prevalence of these conditions in the population is significant, and a timely and accurate diagnosis is necessary to prescribe adequate therapy (McNulty S.N. et al., 2021).

About half of all cases of severe congenital neutropenia are caused by variants in the *ELANE* gene. A small percentage of all cases of this disease are associated with other related genes, including *TCIRG*1. The genetic basis of more than 30% of cases remains unknown.

Hennekam syndrome is an autosomal recessive disease and one of the rarest forms of primary immunodeficiency, which in its phenotype has malformations of the lymphatic system (Hilliard R.I. et al., 1990). The main manifestations of Hennekam syndrome are lymphedema-lymphangiectasia, affecting any part of the body with a predominance of the lower limbs, intestines, abdominal, and pleural cavities. In addition, patients with Hennekam syndrome often have flattened facial features, a wide nose, hypertelorism, epicantus, and other anomalies may be observed (Hennekam R.C. et al., 1989). Hennekam syndrome can be caused by mutations in the *CCBE1* gene (25% of cases), as well as in the *FAT4* and *ADAMTS3* genes, each of which affects the VEGF-C/VEGFR-3 signaling pathways in one way or another (Alders M. et al., 2013).

Thus, identifying the genetic causes of congenital neutropenia and Hennekam syndrome is crucial for proper diagnosis, management, and treatment of patients.

The aim of this study is to utilize bioinformatics analysis methods to determine the role of potentially pathogenic variants in causative genes in the pathogenesis of congenital immune disorders, specifically *RBCK1* deficiency, congenital neutropenia, and Hennekam syndrome.

The research objectives are:

1. To conduct a comparative analysis of gene expression in RBCK1 deficiency relative to healthy children and patients with CINCA/NOMID, Muckle-Wells syndrome, and mevalonate kinase deficiency.

2. To assess the pathogenicity of nonsynonymous single nucleotide variations in the *ELANE* and TCIRG1 genes in congenital neutropenia.

3. To identify potentially new candidate genes in the development of diseases related to congenital neutropenia.

4. To identify new variants of *CCBE1*, *ADAMTS3*, and FAT4 genes that lead to the development of Hennekam syndrome.

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

The differences in gene expression in peripheral blood mononuclear cells in individuals with RBCK1 deficiency compared to healthy individuals have been identified for the first time.

Newly identified pathogenic variants of TCIRG1 and *ELANE* genes, as well as previously published nsSNP variants, have been analyzed for their impact on the corresponding proteins.

New candidate genes for congenital neutropenia have been identified for the first time, which can be used in the future for disease diagnosis and the development of pathogenetically justified treatment methods.

Novel non-synonymous single nucleotide polymorphisms in causative genes of Hennekam syndrome (*CCBE1*, FAT4, and ADMATS3), along with previously published variants, have been shown to significantly affect the structure and function of these proteins.

Theoretical and practical significance of the study. Theoretical significance of the research is based on the acquisition of novel genetic-phenotypic associations that form the pathogenetic foundation of diseases linked with congenital immune deficiencies, such as RBCK1 deficiency, congenital neutropenia, and Hennekam syndrome, by employing a customized sequence of bioinformatics analysis techniques that involves molecular dynamics simulation. These findings will be beneficial in further studies for identifying therapeutic targets for these diseases.

The practical importance of the research lies in the utilization of predicted gene variants for differential diagnosis of primary immunodeficiency syndromes like RBCK1 deficiency, congenital neutropenia, and Hennekam syndrome. The developed program of sequential bioinformatics analysis techniques can be utilized to explore new candidate genes for congenital immune deficiency-related disorders.

The research methodology and methods involved conducting the study at the Immunochemistry Department of the Chemical Engineering Institute of the Federal State Educational Institution of Higher Professional Education "Ural Federal University named after the First President of Russia B.N. Yeltsin" and at the Laboratory of Inflammation Immunology of the Federal State Budgetary Scientific Institution, Institute of Immunology and Physiology, Ural Branch of Russian academy of science.

To achieve the study objectives, the researchers utilized four sets of data downloaded from the GEO database comprising 443 transcript samples. Additionally, single nucleotide polymorphism data for the *ELANE*, TCIRG1, *CCBE1*, FAT4, and *ADMATS3* genes were obtained from the dbSNP-NCBI and Ensmble databases, with a total of 222251 SNPs. Moreover, two sets of whole genome sequencing data of patients were included in the analysis.

Provisions made for the defense:

1. In case of RBCK1 deficiency, the expression of genes involved in immune response signaling pathways, inflammatory response, and protein phosphorylation is reduced.

2. In congenital neutropenia, a list of genes has been identified that expands the spectrum of known genes associated with this group of primary immunodeficiencies.

3. New nonsynonymous single nucleotide substitutions in TCIRG1 and *ELANE* genes are destabilizing for TCIRG1 and *ELANE* proteins.

4. The newly identified nonsynonymous single nucleotide substitutions in genes that lead to Hennekam syndrome destabilize the structure and function of *CCBE1*, *ADAMTS3*, and FAT4 proteins.

Degree of reliability of the results obtained. The scientific validity of the study's results was guaranteed by the accuracy of the initial theoretical assumptions, utilization of contemporary research techniques, adherence to proper data collection procedures, and sophisticated modern data analysis. The research design relied on a thorough examination of recent literature, and hypotheses were tested through several statistical and bioinformatics methods. The obtained data's reliability was substantiated by the expert committee of the Institute of Immunology and Physiology, Ural Branch of RAS, as evidenced by a primary documentation verification act dated 07.04.2023.

Author's personal contribution: The author's personal contribution consisted of direct participation at all stages of the dissertation research. The creation of the main idea, planning of the scientific work, including the formulation of the working hypothesis, goals, objectives, determination of the research methodology, interpretation, and analysis of the obtained results were carried out jointly by the author and scientific supervisors.

Implementation of research results into practice: The research findings have been integrated into the postgraduate curriculum at the Institute of Immunology and Physiology, Ural Branch of the Russian Academy of Sciences, and the Department of Immuno-chemistry at the Institute of Chemical Engineering Ural Federal University named after the first President of Russia B.N. Yeltsin. Additionally, they have been incorporated into the research procedures of the Inflammation Immunology Laboratory at the Institute of Immunology and Physiology, Ural Branch of the Russian Academy of Sciences. The obtained results have been put into practical use at the Clinical and Diagnostic Center for "Maternal and Child Health Protection" of the Healthcare Administration of the Sverdlovsk Region.

Approbation of the work. The key findings of the thesis were presented at several scientific conferences, including the International Euro-Asian Congress on Bioethics, Molecular and Personalized Medicine "Biomed-inn-2019" (Perm, 08.11.2019), the

International Conference "Current issues of organic chemistry and biotechnology" (Ekaterinburg, 21.11.2020), the Russian Conference with international participation "Experimental and Computer Biomedicine" dedicated to Corresponding Member of RAS Vladimir Semenovich Markhasin (Ekaterinburg, 28.05.2021), and the Second International Conference "Physician - Patient - Society: Immunology and Genetics 2022" (Ekaterinburg, 25.05.2022).

Publications. The applicant has published 13 papers based on the results of the dissertation, including 4 papers in journals recommended by the Higher Attestation Commission (HAC) with a K1 category, indexed in Scopus Q1-2 research articles, index Web Of Science Q2 – 1 article, and in Scopus and Web of Science (Q3-Q4) – 6 papers.

Size and structure of the dissertation. The dissertation consists of 230 pages of typewritten text, including an introduction, literature review, materials and methods, three chapters with the results of their own research, conclusions, practical recommendations, list of abbreviations, and bibliography (210 sources, including 12 domestic and 216 foreign). The work is illustrated with 20 tables, 79 figures, and 2 formulas.

MAIN CONTENT OF THE WORK

Materials and Methods: To study the pathogenesis of RBCK1 deficiency, a comparative analysis of gene expression was conducted by downloading 2 datasets from the NCBI Gene Expression Omnibus (GEO): GSE31064, which included data obtained from skin fibroblasts of patients - 2 with RBCK1 deficiency, 1 with MYD88 deficiency, 1 with NEMO syndrome, and 3 healthy individuals (as control); GSE40561, which included data obtained from whole blood collected from 2 patients with CINCA/NOMID disease, five patients with Muckle-Wells syndrome, 2 patients with mevalonate kinase deficiency, one patient with RBCK1 deficiency, and 41 healthy children (as control). To search for candidate genes in congenital neutropenia, an analysis of differential gene expression was conducted on two sets of data downloaded from NCBI GEO (https://www.ncbi.nlm.nih.gov). The GSE142347 dataset included the results of gene expression analysis of 93 female patients, 95 male patients, and 193 healthy controls; the GSE6322 dataset included the results of gene expression analysis of 2 parents and 2 children with neutropenia from one family.

Data on various genes and single nucleotide polymorphisms (SNPs) in congenital neutropenia and Hennekam syndrome were downloaded from the dbSNP-NCBI (https://www.ncbi.nlm.nih.gov/snp/) and Ensmble (https://www.ensembl.org/index.html) databases. The following SNPs were downloaded for the study of single nucleotide polymorphisms or substitutions (SNPs) in genes associated with congenital neutropenia: for the *ELANE* gene, 3646 SNPs, of which non-synonymous single nucleotide

polymorphisms (nsSNPs) were 301; for the TCIRG1 gene, a total of 5627 SNPs, of which 811 were nsSNPs. For the study of SNPs in genes associated with Hennekam syndrome, *CCBE1* had 73845 SNPs and 407 nsSNPs; FAT4 had 68257 SNPs and 3434 nsSNPs; *ADAMTS3* had 70876 SNPs and 911 nsSNPs. Other databases, including Swiss-Prot (http://expasy.org./) and OMIM (https://www.omim.org/), were used to evaluate gene variants.

The investigation of gene variants in patients with congenital neutropenia and Hennekam syndrome from the Sverdlovsk region was carried out using available sequencing results obtained at the Genome Center "Genomed", which were voluntarily provided by patients for research at the Institute of Immunology and Physiology of the Ural Branch of the Russian Academy of Sciences and further anonymized. To assess the harmfulness of non-synonymous single nucleotide substitutions on the protein structure and function, the following sequence of actions was followed. Initially, all identified nsSNPs in the databases were evaluated using the SIFT tool. Then, potentially deleterious variants were sorted and passed through the PolyPhen-2 program, and subsequently evaluated by other bioinformatics tools, including 18 programs and online services such as PROVEAN, FATHMM, LRT, M-CAP, META SVM, METALR, Mutation Assessor, Mutation Taster, FATHMM MKLCoding, CAAD, PHD-SNP, Panter, SNP&GO, PON-P2, DANN, and SNAP2. Some of these tools were available through VarCard and MutPred. The threshold values for the above tools were as follows: Mutation Taster: <0.5; CADD: >15; MetaLR: >0.5; M-Cap: >0.025; PANTHER: probably damaging (probably) substitution at time > 450my, possibly damaging (possibly, less likely) at 450my > time > 200my, probably benign at time < 200my; VEST3: >0.5; LRT: >0.001; PROVEAN: >-2.667; FATHMM-MKK: <0.5; PhDSNP: >0.5; SNP-GO: >0.5; SNAP2: a scale ranging from -100 (completely neutral) to +100 (exerts a strong effect); DANN: >0.5; Mutation Assessor: >0.65 (from -5.545 to 5.975, the higher the value, the more damaging); FATHMM: >0.453; PON-P2: >0.5.

The study utilized various bioinformatics analysis programs to assess the impact of single nucleotide substitutions on protein structure and stability. The I-Mutant 3.0, iStable 2.0, and MU-PRO programs were utilized for this purpose. Additionally, protein-protein interactions were evaluated using the STRING and Cytoscape software packages, while functional enrichment was analyzed with the KEGG database. Gene co-expression was assessed using the CemiTool program. To construct 3D models of wild-type and mutant protein structures and evaluate the effect of mutations on protein function, the HHPred, AlphaFold 2, Phyre2, I-Tasser, Chimera UCSF Chimera, and PyMOL software programs were utilized. Molecular dynamics simulations were carried out using the Maestro and Gromacs 4.5.3 software packages. Whole-genome sequencing and SNP identification were

performed on a supercomputer provided by the Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences - Ural Branch of the Russian Academy of Sciences "Supercomputer Center" (Yekaterinburg). Informed consent was obtained from the parents of the patients for the use of de-identified research results. The statistical analysis, bioinformatics, and computational biology analysis 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.

RESULTS AND DISCUSSION

The aim of this study was to evaluate gene expression differences and investigate key signaling pathways in patients with RBCK1 deficiency. To achieve this goal, a comparative analysis of gene expression was conducted between transcription data of patients with RBCK1 deficiency, CINCA/NOMID syndrome, Muckle-Wells syndrome, mevalonate kinase deficiency (MVK), and transcription data of healthy children (Figure 1). Genes with differential expression were annotated and functionally enriched to obtain information on their role in organism functioning, the signaling pathways they are involved in, and under what conditions they were expressed according to data obtained by other researchers earlier.

From the dataset GSE40561, which includes 48,803 genes in different individuals, 380 differentially expressed genes (DEGs) were detected: 229 genes had increased expression, and 151 genes had decreased expression. Comparative analysis of transcription samples from healthy individuals and patients with RBCK1 deficiency showed the largest number of DEGs - 119 genes with significantly reduced expression. Additionally, when comparing RBCK1 and MWS samples, a significant difference was found in the relatively high expression of 142 genes in RBCK1 deficiency (*figure 1*).
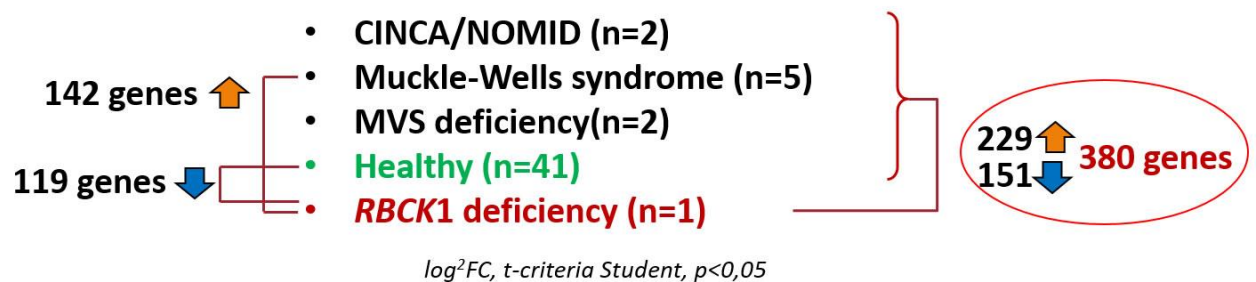

Figure 1 - Significant differences in gene expression in RBCK1 deficiency compared to healthy children and patients with other autoinflammatory syndromes

The next step involved determining the involvement of differentially expressed genes in RBCK1 deficiency in key signaling pathways and evaluating their impact on biological functions. The gene annotation results in the biological processes category

indicated the involvement of immune response signaling pathways, inflammatory response, and protein phosphorylation.

The functional enrichment of differentially expressed genes revealed that in RBCK1 deficiency (relative to healthy individuals), several signaling pathways involved in leishmaniasis development, susceptibility to staphylococcal infection, cholera, NK cell cytotoxicity, and other pathways affecting immune response are activated. It should be noted that this does not necessarily indicate an increased likelihood of developing the corresponding pathology in RBCK1 deficiency. However, it provides insights into how the deficiency of one protein can affect various processes that, in one way or another, impact the immune system and anti-infective defense.

In addition, alterations in the activity of the mTOR, PI3K/AKT, Rho, and Nf-kB signaling pathways can have a direct or indirect impact on the expression of genes involved in the immune system. A significant reduction in the expression of the CISD2 gene was observed in patients with RBCK1 deficiency in this study. Previous research has shown that defects in CISD2 result in endoplasmic reticulum stress and apoptosis (Shen Z. Q. et al., 2021). Given the close functional relationship of CISD2 with cellular stress processes and apoptosis, it can be postulated that the decrease in CISD2 expression in RBCK1 deficiency has an adverse effect on the stability of peripheral blood mononuclear cells to apoptosis and cell death. These findings provide insights into the molecular interactions underlying the pathogenesis of RBCK1 deficiency.

The identified differences in gene expression shed light on the molecular interactions underlying the pathogenesis of RBCK1 deficiency.

To investigate the impact of nonsynonymous single-nucleotide substitutions on the structure and function of *ELANE* and TCIRG1 proteins, which are causally related to severe congenital neutropenia, a study was conducted.

Using the SIFT and PolyPhen-2 tools, the malignancy of nonsynonymous single-nucleotide substitutions was determined. Out of the 301 nsSNPs obtained from the NCBI database in the *ELANE* gene, eight nsSNPs were identified as highly deleterious/damaging as they show to destabilize the structure and function of protien. In contrast, the combined analysis of SIFT and PolyPhen-2 programs allowed the selection of only 34 nsSNPs out of 811 nsSNPs in the TCIRG1 gene.

Additional *in silico* tools were employed to corroborate the deleteriousness of the selected polymorphisms identified by SIFT and PolyPhen-2. Eighteen computational tools were utilized to ascertain the number of potentially deleterious substitutions in the *ELANE* gene and to select the 15 most damaging nonsynonymous single-nucleotide substitutions in the TCIRG1 gene.

Assessment of the impact of high-risk pathogenicity nsSNPs on *ELANE* protein stability and function using I-Mutant 3.0 indicated that amino acid substitutions V101L and A166V enhance *ELANE* protein stability, whereas R34W, C71R, V101M, P139L, R143C, C151Y, A166T, T175M, R182H, V190M, R193W, G203S, L206F, N209K, G210R, G214R, F218L, P262S, and P262L R50C reduce its calculated stability.

Regarding TCIRG1, the results of the evaluation of selected nsSNPs in I-Mutant 3.0 showed that amino acid substitutions G405R, S474W, and A778V enhance protein stability, whereas P572L, M546V, I730N, F610S, A732T, F51S, A717D, E722K, R57H, R109W, R191W, S532C, G192S, F529L, H804Q, G458S, R444L, R56P, G379S, R757C, N730S, V375M, T314M, D517N, R92W, T368M, A417T, R363C, R56W and R50C decrease its calculated stability.

Moreover, the Mu-pro algorithm revealed that all nsSNPs with a high pathogenicity score detected by previous tests reduce the stability of the tested proteins.

For both proteins, the effect of amino acid substitutions on the highly conserved regions of the proteins was analyzed using the ConSurf service. The findings indicated that several putative pathogenic substitutions affect highly conserved regions of the protein, including functional amino acid residues.

The prediction of *ELANE* structures with the selected amino acid substitutions was performed using Phyre2 and 3D modeling with I-Tasser. The resulting 3D models were then evaluated using the Zhanggroup online service, which generated comparison metrics such as TM-score and RMSD of C-alpha atoms between the wild-type and mutant proteins (*table 1*).

Four nsSNPs with the highest RMSD values (C71Y, G214R, R34W, and F218L) were chosen and subjected to remodeling and structure comparison with the wild type using I-Tasser. The validation results for both the wild-type and mutant versions of the 3D models were satisfactory. These selected mutant *ELANE* types were then used in an *in silico* experiment for molecular docking screening.

The docking process was visually assessed using Discovery Studio and PyMol. Docking site interactions were evaluated to determine binding strengths, which are important for stabilizing the formation of the receptor-ligand complex. Analysis of the docking results, which included calculating hydrogen bonds, Van der Waals, and hydrophobic contacts, revealed that the mutations G214R and R34W had interactions that were similar to the wild-type. However, the mutations C71Y and F218L had fewer hydrogen bonds, suggesting that they may impact the stability and energy of the *ELANE* protein.

Table 1 - TM-score and standard deviation values (RMSD) for selected nsSNPs in *ELANE* protein

nsSNP	A.A.S	TM-Score	RMSD
rs28931611	C71Y	0.85993	2.05
rs201163886	R34W	0.86482	1.98
rs200384291	F218L	0.87828	1.96
rs137854451	G214R	0.96114	1.12
rs201723157	R193W	0.95176	0.96
rs201139487	G203S	0.99524	0.49
rs137854448	P139L	0.99994	0.04

Fifteen of the most damaging non-synonymous single nucleotide substitutions (and corresponding amino acid substitutions) for the TCIRG1 protein were selected based on the predictions of the algorithms used in the previous research stage, as they were expected to have the most significant effects on the protein's structure and function. These mutations included rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489, rs116675104, and rs118141250. One of these substitutions (rs118141250, Val52Leu) had previously been identified in a patient from the Sverdlovsk region through whole-genome sequencing. All of the mutations were incorporated into a single three-dimensional structure of the TCIRG1 protein, as they were located in different regions. As a result, differences in the 3D structure were observed for each region where a substitution could occur when comparing the models. Finally, molecular dynamics simulations of the wild-type and mutant TCIRG1 variants' 3D models in AlphaFold2 were conducted.

The RMSD evaluation results indicate that the mutant protein has a higher RMSD during the 100 ns modeling period (*figure 2*).

The results of the assessment of the impact of amino acid substitutions on the molecular dynamics, taking into account the secondary structures of proteins, indicate that the stability of the TCIRG1 molecule is reduced in the mutant type relative to the wild type protein. Therefore, non-synonymous single nucleotide substitutions in the TCIRG1 (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489, and rs116675104) and *ELANE* (rs200384291, rs201163886, rs193141883, rs201139487, and rs201723157) genes destabilize TCIRG1 and *ELANE* proteins, suggesting their influence on the development of severe congenital neutropenia syndrome.
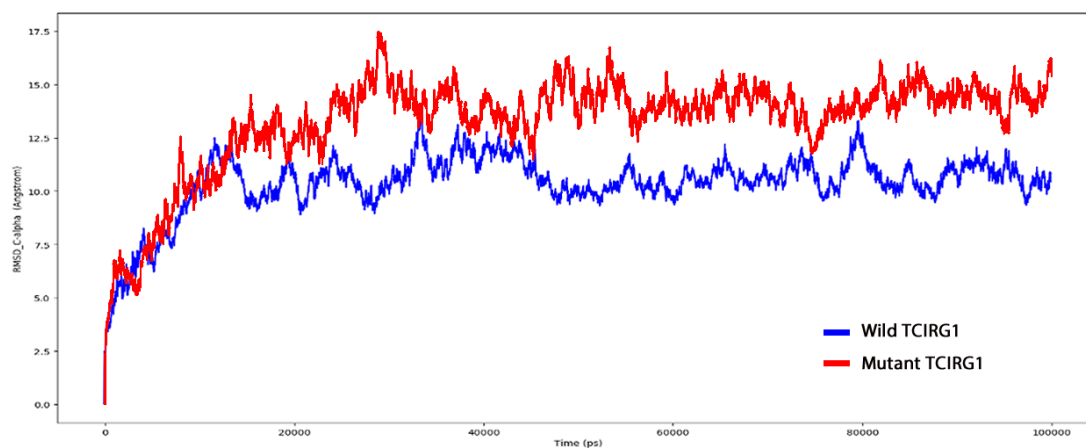
Figure 2 - RMSD of wild-type and mutant Cα atoms by HHpred data over time (100ns)

Note: Abscissa axis is time in ps, ordinate axis is RMSD (Å).

To search for and predict new candidate genes for severe congenital neutropenia, the first steps involved a review analysis of published information, an analysis of information in genome and inherited disease databases, as well as a review of genetic studies related to the phenotype.

Using the STRING database, information on protein-protein interactions was extracted for all known genes of primary immunodeficiencies. The results showed a strong interaction between genes associated with congenital neutropenia.

Based on Pearson correlation analysis and protein-protein interactions (Cheng F. et al., 2018), we obtained 4,613 specific gene interactions that are functionally related to congenital neutropenia and identified 177 candidate genes. We conducted functional enrichment analysis of known genes of congenital neutropenia using KEGG data to link the genes from the list with their biological functions. Our KEGG pathway analysis revealed five statistically significant signaling pathways ($p<0.05$), including cytokine-cytokine receptor interaction and chemokine signaling pathways. We used this information to identify specific candidate genes that are functionally similar to known congenital neutropenia genes and enriched in at least one of the five KEGG pathways. This led to the discovery of 15 new candidate genes for congenital neutropenia, including *STAT1, STAT2, STAT3, STAT5B, LYN, FGR, SRC, PIK3CG, ITK, VAV1, CDC42, PTK2, CRKL, PLCG1, and ARRB2*.

Further functional enrichment analysis of congenital neutropenia genes, including the 15 candidate genes, revealed a total of 15 statistically significant signaling pathways described in the KEGG database, such as Epstein-Barr virus infection, cytokine-cytokine receptor interaction, and B-cell receptor signaling pathway (*figure 3*).

To analyze the network of the 15 candidate genes for congenital neutropenia and 31 k associated with the disease, we used the STRING database. The results demonstrated a close relationship between these genes, indicating their involvement in the pathogenesis of congenital neutropenia (*figure 3*).

Additionally, each candidate gene for congenital neutropenia was merged with known genes for congenital neutropenia, and the biological distance of the merged gene was re-evaluated. The merged genes then underwent phylogenetic analysis using FGA to determine the biological relationship between the congenital neutropenia genome and the candidate gene for congenital neutropenia. The results showed that the 15 candidate genes for congenital neutropenia were evenly distributed throughout the range of known genes, suggesting a close association with the known genes for congenital neutropenia.
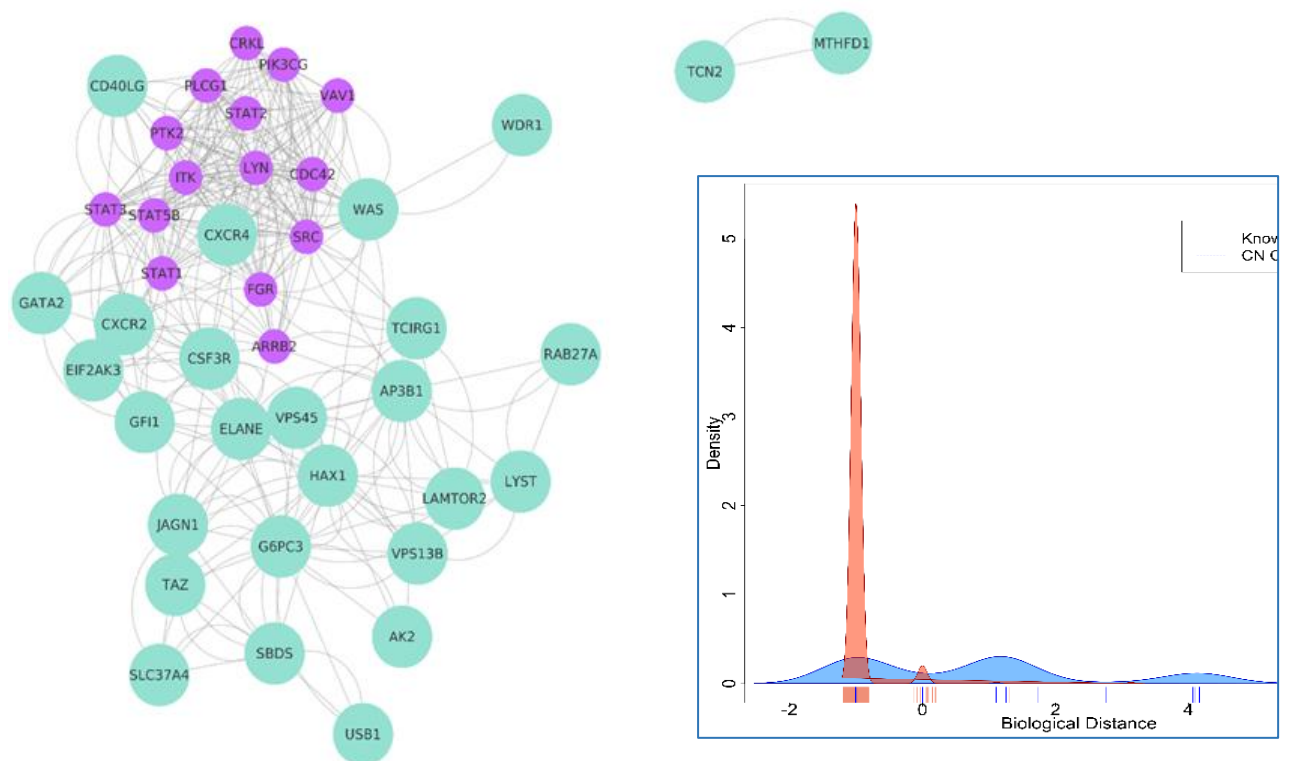

Figure 3 - Network of interprotein interactions of known and candidate genes for congenital neutropenia (Cytoscape). In the inset - comparison of the biological distance between known genes and candidate genes for congenital neutropenia
Note. In the main figure: green color represents known genes for congenital neutropenia, and purple represents candidate genes. In the inset: blue represents network density and biological distance of known genes for congenital neutropenia, while red represents candidate genes.

The evaluation of differences in gene expression in peripheral blood neutrophils of patients in the analyzed datasets confirmed the preliminary results using another approach. Moreover, when comparing genes with differences in expression in peripheral blood neutrophils and B-cells of patients with congenital neutropenia, common genes with

reduced and elevated expression relative to control samples were identified. This finding indicated the identification of common transcriptional properties of neutrophil and B-cell transcriptomes in patients with congenital neutropenia.

Having identified new candidate genes, we conducted a quality assessment study by reviewing some relevant literature. In particular, ten candidate genes for congenital neutropenia (*STAT1, STAT2, STAT3, STAT5B, LYN, FGR, SRC, PIK3CG, ITK, VAV1, CDC42*), which were not listed in ESID but were predicted by us, were identified by other researchers in clinical cases of congenital neutropenia. This confirms the significance of the candidate genes for congenital neutropenia that we identified.

To assess the potential pathogenicity of new nonsynonymous single nucleotide polymorphisms (nsSNPs) in Hennekam syndrome genes (*CCBE1*, FAT4, and *ADAMTS3*), we used modern *in silico* tools. Specifically, we investigated the most pathogenic nsSNPs of *CCBE1*, *FAT4* , and *ADAMTS3*. Our findings indicate that seven non-synonymous single nucleotide polymorphisms (nsSNPs) in the *CCBE1* gene (rs1159828792479, rs149792489, rs374941368, rs121908254, rs149531418, rs121908251 and rs372499913) are likely to have pathogenic effects. Among them, four (G330E, C102S, C174R, and G107D) are highly probable to be pathogenic, with two (G330E and G107D) never reported in Hennekam syndrome. Additionally, we evaluated two significant substitutions (rs374941368 and rs200149541) in the *CCBE1* gene, which may affect post-translational modifications by impacting a potential phosphorylation site.

In the analysis of *ADAMTS3* gene variants obtained from the dbSNP database, a total of 919 nonsynonymous single nucleotide polymorphisms (nsSNPs) were pre-sorted. Out of these, five substitutions (G298R, C567Y, A370T, C567R, and G374S) were predicted to have the highest potential for being pathogenic and could be associated with diseases. The modeling of *ADAMTS3* protein demonstrated that it can be segmented into parts 1, 2, and 3, connected by short loops. Using various molecular dynamics simulation tools, it was found that some substitutions destabilize the protein structure, causing disruptions in secondary structures, especially in segment 2 (as shown in Figure 4). The pathogenic effects of mutations in segment 1 may not be due to destabilization, but rather to other factors such as changes in phosphorylation, as suggested by studies of post-translational modification.

Our study is the first to examine *ADAMTS3* gene polymorphisms using multiple tools, including molecular dynamics simulations. Some of the predicted substitutions in

the *ADAMTS3* protein have not yet been reported in the PubMed library. We hope that our findings will be useful for diagnostic purposes and in the search for therapies.
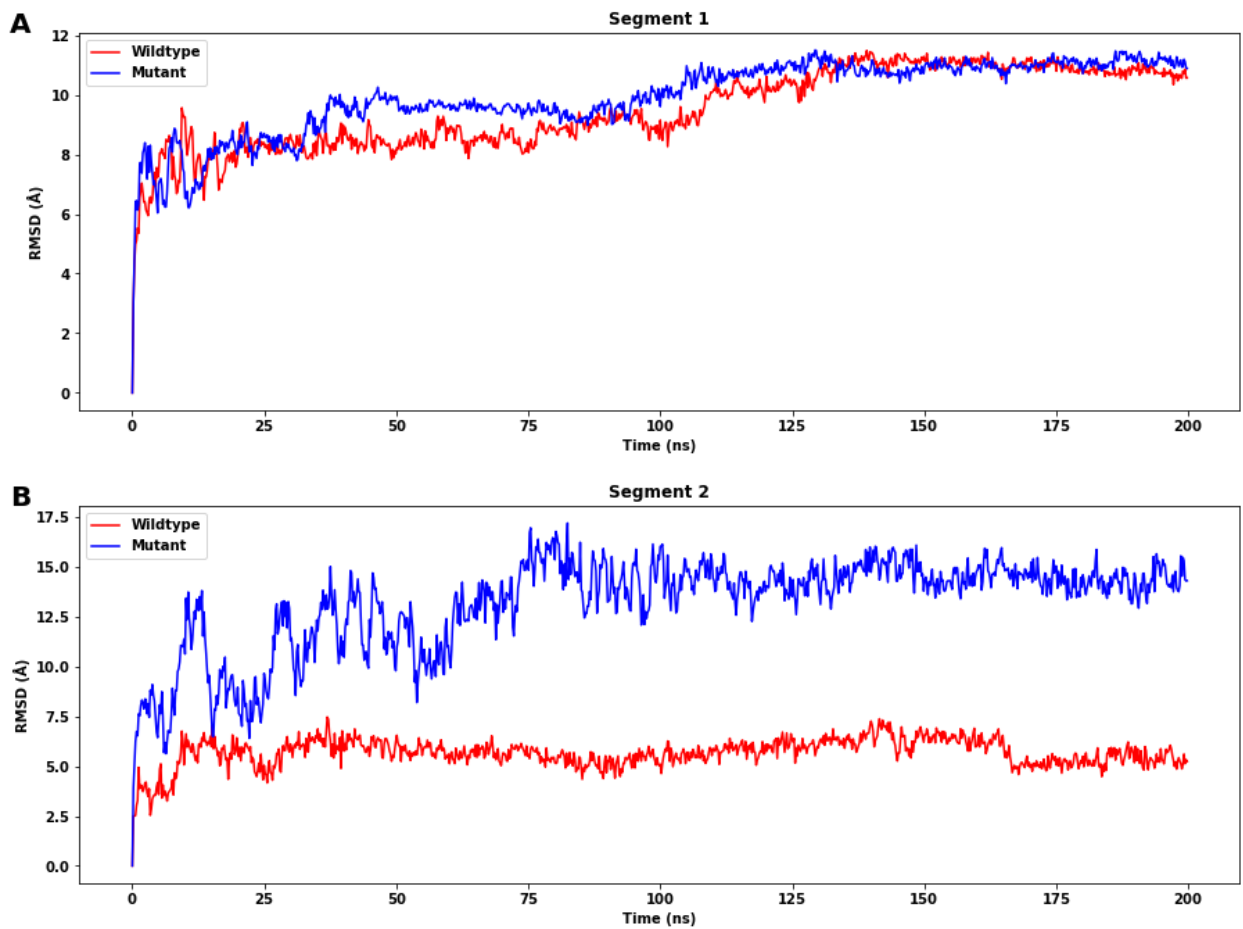

Figure 4 - Root Mean Square Deviation (RMSD) of the Cα atoms of wild-type (red) and mutant (blue) *ADAMTS3* protein segments 1 (A) and 2 (B) over time (in nanoseconds)

Note. that the difference in equilibrium RMSD is insignificant for segment 1 (A), while there is a significant difference in RMSD for segment 2 (B). All mean values and standard deviations (SD) were calculated using values after 170 ns.

When analyzing various variants of the FAT4 gene from the 3,343 nsSNPs available in the NCBI database using various pathogenicity prediction tools, 11 substitutions in the FAT4 protein (D2978G, V986D, Y1912C, R4799C, D1022G, G4786R, D2439E, E2426Q, R4643C, N1309I, and Y2909H) were predicted to be potentially pathogenic. In addition, three substitutions in the FAT4 gene (rs12650153, rs1567047, and rs1039808) were previously detected in a patient with Hennekam syndrome from the Sverdlovsk region using candidate variant filtering with whole-genome sequencing, and *in silico* analysis of these mutations demonstrated that they could significantly destabilize the structure and function of the protein.

Using molecular dynamics simulation, we focused on 19 mutations in the FAT4 gene, including 11 predicted in our *in silico* study, three non-synonymous single nucleotide polymorphisms (nsSNPs) detected in the patients, and five nsSNPs already published as probable causes of Hennekam and Van Maldergem syndromes, which have phenotypic differences from Hennekam syndrome.

The results of the molecular dynamics simulation method confirmed lower stability of the mutant protein compared to the wild-type. The genetic variants identified in this study cohort were not previously reported as causes of Hennekam syndrome. It should be noted that due to the limitations of supercomputer resources and software, we were only able to simulate fragments of the large *FAT4* protein, which consists of 4981 amino acids. In our experiment, the *FAT4* molecule was analyzed using five models, each containing a segment with the analyzed substitution less than 1000 amino acids long. Nevertheless, we hope that these results can contribute to a better understanding of the predisposition to diseases associated with FAT4 protein activation and may help in developing effective approaches for diagnosing and treating diseases related to this gene.

Prospects for further development of the topic. Identification of specific genetic changes and determination of the molecular basis of immunopathology will enable the investigation of pathogenic mechanisms, differentiation of nosological forms from the extensive heterogeneous group of inborn errors of the immune system, and bring the creation of specific targeted therapies, including gene editing and antisense oligonucleotides, closer. This will help to address issues of radical cure for patients. Additionally, even the simple acceleration of the diagnostic process will enable timely diagnosis, prescription of substitutional and pathogenetic therapy, improvement of prognosis, and enhancement of the quality of life of patients.

The process of verifying primary immunodeficiency genes can be improved by developing software for predicting candidate genes of various immunopathologies. The inclusion of methods for predicting the impact of genetic changes on the protein *in silico* provides the possibility of its effective application in clinical research.

Furthermore, studying rare cases of human pathology allows for solving general pathological questions regarding the formation of diseases, enriching science with knowledge about the functioning of the immune system and the human organism as a whole. By delving into the molecular level of pathology, researchers gain access to objective justifications for the development and application of targeted therapeutic tactics, opening up the prospect of creating new targeted drugs.

Findings

1. New genetic disorders have been identified in three types of primary immunodeficiency disorders: RBCK1 deficiency, congenital neutropenia, and Hennekam syndrome.

2. Significant differences in gene expression have been identified in RBCK1 deficiency compared to healthy children and patients with CINCA/NOMID syndrome, Muckle-Wells syndrome, and mevalonate kinase deficiency.

3. Nonsynonymous single nucleotide substitutions in the *TCIRG1* gene (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489, rs116675104) and *ELANE* gene (rs200384291, rs201163886, rs193141883, rs201139487, rs201723157) destabilize the TCIRG1 and *ELANE* proteins in neutrophils.

4. *CDC42, CRKL, FGR, CRC, NYK, PLCG1, ARRB2, PIK3CG, PTK2, STAT1, STAT2, STAT3, STAT5B, VAV1*, and *ITK* genes are new candidate genes for the development of congenital neutropenia.

5. Nonsynonymous single nucleotide substitutions in the *CCBE1* (rs115982879, rs149792489, rs374941368, rs121908254, rs149531418, rs121908251, and rs372499913), FAT4 (rs147663284, rs192514171, rs138137489, rs199895179, rs372060616, rs138173652, rs142184187, rs147633644, rs181607904, rs184971791, rs148655455), and *ADAMTS3* (rs61757480, rs61741624, rs140806973, rs140595148, rs140914273, rs142268705, rs142781084, rs143059623, rs146979323, rs372067284, rs370857003, rs375983592, rs367831484, rs202031187, and rs150012152) genes lead to destabilization of the CCBE1, FAT4 , and ADAMTS3 protein structures and may be the cause of Hennekam syndrome development.

6. The developed program for the sequential use of bioinformatics methods is effective in finding genes that influence the pathogenesis of diseases associated with primary immunodeficiency disorders.

PRACTICAL RECOMMENDATIONS

1. In diagnosing primary immunodeficiency disorders (congenital immune deficiencies), it is necessary to determine the gene expression profile by analyzing differential gene expression, signaling pathways, and genetic ontologies, as well as to determine biomarkers of pathology in patients. This will reduce the cost of treatment and prevent the development of side effects.

2. In studies aimed at predicting new candidate genes for congenital neutropenia, factors of co-expression, protein-protein interactions, and signaling pathways should be included in the analysis.

3. For the differential diagnosis of congenital neutropenia, in addition to the genes listed on the ESID website and in the IUIS classification, additional genes (*CDC42, CRKL, FGR, CRC, NYK, PLCG1, ARRB2, PIK3CG, PTK2, STAT1, STAT2, STAT3, STAT5B, VAV1,* and *ITK*) identified as candidate genes in our study should be included.

4. For the differential diagnosis of Hennekam syndrome and congenital neutropenia, in addition to the listed missense mutations in the genes *ADAMTS3*, *FAT4* , *CCBE1*, *ELANE*, and *TCIRG1*, the presence of nsSNP missense mutations identified in our study for the following genes should be evaluated:

- *CCBE1* (rs115982879, rs149792489, rs374941368, rs121908254, rs149531418, rs121908251 и rs372499913),

- *ELANE* (rs200384291, rs201163886, rs193141883, rs201139487 и rs201723157),

- TCIRG1 (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489 и rs116675104),

- FAT4 (rs147663284, rs192514171, rs138137489, rs199895179, rs372060616, rs138173652, rs142184187, rs147633644, rs181607904, rs184971791, rs148655455) и

- *ADAMTS3* (rs61757480, rs61741624, rs140806973, rs140595148, rs140914273, rs142268705, rs142781084, rs143059623, rs146979323, rs372067284, rs370857003, rs375983592, rs367831484, rs202031187, и rs150012152).

List of publications on the topic of the dissertation

Articles from the list of editions recommended by the Higher Attestation Commission (K-1):

1. Ретроспективный анализ случаев первичных иммунодефицитов у детей с врожденными пороками сердца / С.С. Дерябина, Д.А. Черемохин, И.А. Тузанкина, Х. Шинвари // Российский иммунологический журнал. 2020. Т. 23, № 4. С. 505-514 (RSCI, ИФ РИНЦ – 0.165, ВАК К1).

2. Классификация врожденных ошибок иммунитета человека, обновленная экспертами комитета международного союза иммунологических обществ в 2019 году / М.А. Болков, И.А. Тузанкина, Х. Шинвари, Д.А. Черемохин // Российский иммунологический журнал. 2021. Т. 24 (1). С. 7-68 (RSCI, ИФ РИНЦ – 0.165, ВАК К1).

3. Роль врожденных ошибок иммунитета в группе детей с летальными исходами на первом году жизни / Д.А. Черемохин, И.А. Тузанкина, В.А. Черешнев, М.А. Болков, Х. Шинвари // Российский иммунологический журнал. 2022. Т. 25 (4). С. 555-560 (RSCI, ИФ РИНЦ – 0.165, ВАК К1).

4. Analysis of the TREC and KREC Levels in the Dried Blood Spots of Healthy Newborns with Different Gestational Ages and Weights / D.A. Cheremokhin, K. Shinwari,

S.S. Deryabina, M.A. Bolkov, I.A. Tuzankina, D.A. Kudlay // Acta naturae. 2022. V. 14 (1). P. 101–108 (RSCI, ИФ РИНЦ – 2.121, ВАК К1).

Articles indexed in Scopus and Web of Science (Q1-Q4):

5. Is Up-Regulation Gene Expression of the Certain Genes During the Viral Respiratory Tract Infection Would Have Any Influence in Pathogenesis of the SAR-CoV-2 Infection? / K. Shinwari, G. Liu, M. Bolkov, I. Ahmad, M. Daud, I. Tuzankina, V. Chereshnev // Acta Med. Iran. 2020. Vol. 58 (5). P. 246-248 (Scopus Q4).

6. Identifying Main Genes and Pathways by Using Gene Expression Profiling in Primary Immunodeficiency HOIL-1/RBCK1 Disorder Patients / K. Shinwari, G. Liu, M.A. Bolkov, M. Ullah, I. Tuzankina // Acta Med. Iran. 2021. V. 59 (5). P.265-279 (Scopus Q4).

7. Predicting the Most Deleterious Missense Nonsynonymous Single-Nucleotide Polymorphisms of Hennekam Syndrome-Causing CCBE1 Gene, In Silico Analysis / K. Shinwari, L. Guojun, S.S. Deryabina, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshnev // The Scientific World Journal. 2021. Vol. 2021. P. 6642626. (Scopus Q3).

8. Additional Pathogenic Pathways in RBCK1 Deficiency / E.I. Demicheva, K. Shinwari, K.S. Ushenin, M.A. Bolkov // Mathematical Biology and Bioinformatics. 2022. V. 17 (2). P. 174-187 (Scopus Q4, RSCI, ИФ РИНЦ – 0.527).

9. Checking gene expression profile associated with IRF7 and UNC93B deficient patient peripheral blood mononuclear cells infected with pH1N1 influenza virus / K. Shinwari, G. Liu, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshnev // AIP Conference Proceedings. 2022. Vol. 2390. P. 030089 (Scopus).

10. Gene expression and pathway analysis in patients with inborn error of TLRs and IL-IRs signaling using microarray data / K. Shinwari, G. Liu, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshnev // AIP Conference Proceedings. 2022. Vol. 2390. P.030088 (Scopus).

11. In Silico Analysis Revealed Five Novel High-Risk Single-Nucleotide Polymorphisms (rs200384291, rs201163886, rs193141883, rs201139487, and rs201723157) in ELANE Gene Causing Autosomal Dominant Severe Congenital Neutropenia 1 and Cyclic Hematopoiesis / K. Shinwari, M.A. Bolkov, M. Yasir Akbar, L. Guojun, S.S. Deryabina, I.A. Tuzankina, V.A. Chereshnev // Scientific World Journal. 2022. Vol. 2022. P.3356835 (Scopus Q1)

12. Novel Disease-Associated Missense Single-Nucleotide Polymorphisms Variants Predication by Algorithms Tools and Molecular Dynamics Simulation of Human TCIRG1 Gene Causing Congenital Neutropenia and Osteopetrosis / K. Shinwari, H.M. Rehman, G. Liu, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshnev // Front. Mol. Biosci. 2022. Vol.9. P. 879875. (WoS Q2, Scopus Q1).

13. Novel high-risk missense mutations identification in FAT4 gene causing Hennekam syndrome and Van Maldergem syndrome 2 through molecular dynamics simulation / K. Shinwari, H.M. Rehman, N. Xiao, L. Guojun, M.A. Khan, M.A. Bolkov,

I.A. Tuzankina, V.A. Chereshnev // Informatics in Medicine Unlocked. 2023. Vol. 37. P. 101160 (Scopus Q2).

LIST OF ABBREVIATIONS

IEI	Inborn errors of immunity
PID	Primary immunodeficiencies
CINCA/NOMID	Chronic infantile neurological skin and joint disease/Multisystemic inflammatory disease with neonatal onset
ELANE	Elastase expressed by neutrophils
FEL	Free energy landscape
LUBAC	Linear ubiquitin chain assembly complex
NF-kB	Nuclear factor kB
nsSNP	Non-synonymous single nucleotide substitution/polymorphism
PAC2	Chaperone for assembling proteasomes 2
PCA	Principal component analysis
RBCK1 (HOIL-1)	(HOIL-1) RanBP-type and C3HC4-type zinc finger-containing protein 1 (Hem-oxidized ubiquitin ligase-1 IRP2)
RMSD	Root Mean Square deviation
RMSF	Root Mean-Square Fluctuation
SASA	Solvent-accessible surface area
VEGF-C / VEGFR-3	Vascular endothelial growth factor C / Receptor to vascular endothelial growth factor 3

Shinwari Khyber

NOVEL GENE VARIANTS IN THE EVALUATION OF INBORN ERRORS OF IMMUNITY: RBCK1 DEFICIENCY, CONGENITAL NEUTROPENIA, HENNEKAM SYNDROME

3.2.7 - Immunology

ABSTRACT

Of dissertation

for the science degree

of candidate of biological sciences