Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):482-489

doi 10.18699/vjgb-26-52

Diagnostic efficiency of whole exome sequencing in the search for genetic causes of hereditary diseases in Yugra (West Siberia, Russia)

M.Yu. Donnikov [1] ✉, P.A. Suchko [2, 3], A.V. Morozkina [1], L.N. Kolbasin [1, 4], E.A. Popova [1, 4], S.I. Papanov [1, 4], Yu.S. Koshevaya [5], L.G. Danilov [3], Yu.A. Eismont [6], O.S. Glotov [2, 6], L.V. Kovalenko [1]

[1] Medical Institute of Surgut State University, KHMAO-Yugra, Surgut, Russia
[2] Research Institute of Obstetrics, Gynecology, and Reproductology named after D.O. Ott, St. Petersburg, Russia
[3] Saint Petersburg State University, St. Petersburg, Russia
[4] KHMAO-Yugra Surgut Regional Clinical Center for Maternity and Childhood Protection, Medical Genetic Counseling Service, Surgut, KHMAO-Yugra, Russia
[5] Saint-Petersburg State Medical Diagnostic Center (Genetic Medical Center), St. Petersburg, Russia
[6] Federal Scientific and Clinical Center of Infectious Diseases of the Federal Medical and Biological Agency, St. Petersburg, Russia
✉ donnikov@gmail.com

Abstract. Whole-exome sequencing (WES) has revolutionized the diagnostics of hereditary diseases, yet its efficacy varies across populations. Data on the genetic architecture of rare hereditary disorders in many Russian regions, including the ethnically diverse Khanty-Mansi Autonomous Okrug (Yugra) are scarce. The aim of this study was to evaluate the diagnostic yield of WES for identifying genetic variants associated with hereditary disorders in this ethnically heterogeneous population. The study involved 286 probands with suspected hereditary disorders observed by regional geneticists in the years 2021–2024. WES was performed on the DNBSEQ-G50 platform (MGI, China). Bioinformatic analysis included variant calling and annotation using population databases and pathogenicity prediction tools. Identified variants were classified according to ACMG/Russian Medical Genetics Society guidelines and correlated with clinical phenotypes. Molecular genetic diagnoses were categorized as definitive, partial, potential (based on variants of unknown significance), or unknown. The examined cohort was predominantly pediatric, the most common clinical indications were neurological, dysmorphic, and metabolic disorders. Definitive molecular diagnoses were established in 24.8 % of patients. Inclusion of potential diagnoses increased the total yield to 48.6 %. Diagnostic efficacy varied significantly among disease categories ranging from 58.3 % for renal disorders to 0 % for neurodevelopmental disorders. A total of 420 unique variants were analyzed, and missense changes were the most frequent among clinically significant findings. The most commonly implicated genes were *ATP7B*, *GJB2*, *ABCA4*, and *GALT*. The study results indicate that WES is an effective first-tier molecular tool for a wide range of suspected hereditary diseases in the Yugra population, with a diagnostic yield comparable to similar studies abroad. The findings support the utility of WES in diverse populations and highlight the potential for increasing yield through trio-WES and periodic data reanalysis.

Key words: whole exome sequencing; hereditary diseases; new generation sequencing; genetic counseling; diagnostic effectiveness; molecular genetic diagnosis

For citation: Donnikov M.Yu., Suchko P.A., Morozkina A.V., Kolbasin L.N., Popova E.A., Papanov S.I., Koshevaya Yu.S., Danilov L.G., Eismont Yu.A., Glotov O.S., Kovalenko L.V. Diagnostic efficiency of whole exome sequencing in the search for genetic causes of hereditary diseases in Yugra (West Siberia, Russia). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):482-489. doi 10.18699/vjgb-26-52

Funding. The study was supported by the Foundation for Scientific and Technological Development of Yugra, project 2023-574-05.

Диагностическая эффективность полноэкзомного секвенирования в поиске генетических причин наследственных заболеваний в ХМАО-Югре (Западная Сибирь, Россия)

М.Ю. Донников [1] ✉, П.А. Сучко [2, 3], А.В. Морозкина [1], Л.Н. Колбасин [1, 4], Е.А. Попова [1, 4], С.И. Папанов [1, 4], Ю.С. Кошевая [5], Л.Г. Данилов [3], Ю.А. Эйсмонт [6], О.С. Глотов [2, 6], Л.В. Коваленко [1]

[1] Медицинский институт, Сургутский государственный университет, Сургут, ХМАО-Югра, Россия
[2] Научно-исследовательский институт акушерства, гинекологии и репродуктологии им. Д.О. Отта, Санкт-Петербург, Россия
[3] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
[4] Сургутский окружной клинический центр охраны материнства и детства, Россия, Сургут, ХМАО-Югра, Россия
[5] Диагностический центр (медико-генетический), Санкт-Петербург, Россия
[6] Федеральный научно-клинический центр инфекционных болезней Федерального медико-биологического агентства, Санкт-Петербург, Россия
✉ donnikov@gmail.com

Аннотация. Полноэкзомное секвенирование произвело революцию в диагностике наследственных заболеваний, однако его эффективность варьирует в разных популяциях. Данные о генетической архитектуре редких наследственных заболеваний во многих регионах России, включая этнически разнообразный Ханты-Мансийский автономный округ – Югру, ограниченны. Настоящее исследование направлено на оценку диагностической ценности полноэкзомного секвенирования для выявления генетических вариантов, ассоциированных с наследственными заболеваниями в этой этнически неоднородной популяции. В исследование включили 286 пробандов с предполагаемыми наследственными заболеваниями, наблюдаемых генетиками региона в период с 2021 по 2024 г. Полноэкзомное секвенирование было выполнено на платформе DNBSEQ-G50 (MGI, Китай). Биоинформатический анализ включал выявление и аннотацию вариантов с использованием популяционных баз данных и инструментов прогнозирования патогенности. Найденные варианты были распределены по категориям в соответствии с рекомендациями ACMG/Российского общества медицинских генетиков и коррелировали с клиническими фенотипами. Молекулярно-генетические диагнозы были классифицированы как полные, частичные, потенциальные (на основании вариантов неизвестной значимости) или отсутствующие. Обследованная когорта преимущественно состояла из детей, наиболее частыми клиническими показаниями были неврологические, дисморфические и метаболические нарушения. Окончательный молекулярно-генетический диагноз был установлен у 24.8 % пациентов. С учетом потенциальных заключений общая значимость исследования увеличилась до 48.6 %. Диагностическая эффективность значительно варьировала в зависимости от категории заболеваний: от 58.3 % для заболеваний почек до 0 % для нарушений нейрогенеза. Проанализировано в общей сложности 420 уникальных вариантов генов, при этом миссенс-варианты были самыми встречающимися среди клинически значимых результатов. Наиболее часто выявлялись гены *ATP7B*, *GJB2*, *ABCA4*, *GALT*. Результаты исследования показали, что полноэкзомное секвенирование – эффективный молекулярный тест первой линии для широкого спектра предполагаемых наследственных заболеваний у жителей Югры с диагностической ценностью, сопоставимой с аналогичными исследованиями за рубежом. Полученные данные подтверждают результативность полноэкзомного секвенирования в гетерогенной популяции и подчеркивают потенциал повышения диагностической ценности за счет формата «трио» и периодического повторного анализа данных.

Ключевые слова: полноэкзомное секвенирование; наследственные заболевания; секвенирование нового поколения; генетическое консультирование; эффективность диагностики; молекулярная диагностика

Introduction

Due to the development of molecular genetic diagnostic methods, especially high-throughput sequencing, significant progress has been made in determining the molecular nature of hereditary diseases over the past 15 years. In particular, the introduction of whole-genome (WGS) and whole-exome (WES) sequencing has enabled both large-scale population projects to describe the frequencies of genetic variants and the analysis of complex clinical cases with unclear disease etiology.

Although technological advances allow for more precise results and cheaper research, the clinical interpretation of genomic data necessary for a specific patient has become a new challenge (Petersen et al., 2017). Given the enormous amount of data obtained from WGS, the main obstacles to implementing this method are the difficulty of interpretation in the context of a specific disease, the high cost, and the burden on laboratory infrastructure with the general diagnostic efficiency of about 40 % (Stranneheim et al., 2021).

In terms of diagnostic efficiency, WES allows focusing on the analysis of protein-coding regions of genes. It is hardly inferior to WGS, and its advantage is less laborious interpretation. However, its key disadvantages are the loss of information outside of exons, the uneven coverage of gene sequences by probes, and limitations in analyzing copy number variations (CNVs) (Wang et al., 2017; Ross et al., 2020).

Since the diagnostic efficiency of WES for hereditary diseases significantly exceeds that of targeted panels, its main niche includes cases with suspected rare genetic diseases, diseases with recently identified or extended genes, a suspected heterogeneous disease in a young child, and negative results from other diagnostic methods (Okuneva et al., 2020).

According to OMIM Morbid Map Scorecard (www.omim.org/statistics/geneMap, accessed September 25, 2025), there are currently 6,619 phenotypes with 4,661 involved genes corresponding to single gene disorders and traits with known molecular basis. Most common forms of monogenic diseases vary significantly not only among different countries but also within distinct regions of a country. Therefore, it is of utmost importance to study the spectrum of genetic variants and monogenic diseases in all regions of Russia (Zinchenko et al., 2019), taking into account the vast diversity of subpopulations that have been practically unexplored by large-scale studies (Barbitoff et al., 2024).

The aim of this study was to identify the genetic causes of the most common hereditary diseases in the West Siberian Russian Khanty-Mansi Autonomous Okrug (Yugra) and to assess the diagnostic efficiency of WES in this population.

Materials and methods

Patient enrollment. Patient selection was carried out from 2021 to 2024 inclusive, as they were admitted for consultation at the regional medical genetic service located in Surgut (Yugra). For the study, 286 probands of various ages and ethnicities were selected according to the following inclusion criteria: the suspected monogenic nature of the disease (early age of manifestation, indications in the family history, inherited nature of the disorder, rare and specific symptoms of multiple organ damage), as well as already diagnosed hereditary diseases with an unspecified molecular cause based on clinical picture and common biochemical tests. For molecular genetic testing, 5 mL of peripheral blood (with EDTA) were taken from patients and transferred for processing to the Yugra biobank laboratory, established at the Surgut State University.

All participants (or their official representatives) provided informed consent for participation in the study and personal data processing. The study was conducted in accordance with the Helsinki Declaration.

DNA extraction, library preparation, and sequencing. Genomic DNA was extracted from peripheral blood using a MagPure Blood DNA Kit (Magen, China). Whole-exome libraries were prepared using a KAPA HyperPlus Kit and KAPA HyperExome probes (Roche, United States). The libraries were converted with a MGIEasy Universal Library Conversion Kit (MGI, China) and sequenced on a DNBSEQ-G50 system (MGI) in the paired-end mode with the read length of 150 bp, following the manufacturer's recommendations.

Bioinformatics analysis. Samples with an average 70× coverage of target regions and at least 10× coverage width of 98 % were included in further analysis. Samples that did not pass quality control were sent for repeated library preparation and sequencing.

Mapping of obtained reads to the human reference genome (hg19 in 2021–2022 and hg38 in 2023–2024) was performed using BWA (0.7.16) (Li, 2011). Post-processing steps of alignments, variant calling, and filtering were carried out using the Genome Analysis Toolkit (Van der Auwera, O'Connor, 2020).

Variant annotation for all known transcripts of each gene from the RefSeq database was performed using snpEff (v.5.1) (Cingolani et al., 2012), with population frequencies of identified variants from The 1000 Genomes Project and gnomAD samples added to the annotation. Pathogenicity prediction of sequence variants was performed using DANN (Quang et al., 2015), GERP (Davydov et al., 2010), REVEL (Ioannidis et al., 2016), SIFT (Kumar et al., 2009), PolyPhen2 (Adzhubei et al., 2010), PrimateAI (Sundaram et al., 2018). Algorithms AdaBoost (Pashaei et al., 2016) and SpliceAI (Strauch et al., 2022) were also used for assessing the impact of variants on the splice site function.

Clinical data interpretation. To assess the clinical relevance of identified sequence variants the following resources were used: OMIM database, disease-specific databases (if available), and scientific literature data. Reports included only variants that had a possible relationship to the patient's clinical manifestations or met other criteria specified in the physician's referral. Polymorphisms classified as benign or likely benign were excluded from the report.

Identified variants were categorized according to criteria from ACMG guidelines (Richards et al., 2015) and the Russian Society of Medical Genetics (Ryzhkova et al., 2019) as pathogenic (P), likely pathogenic (LP), variants of unknown (clinical) significance (VUS), and asymptomatic carriers (AC). The last category included heterozygous pathogenic and likely pathogenic variants not causing the disease in the proband but strongly associated with other monogenic diseases.

Classification of molecular diagnoses. Based on the number of identified variants in each gene and inheritance type, molecular testing results were categorized as follows:

– complete molecular genetic diagnosis (MGD): at least one heterozygous or hemizygous P/LP variant with a dominant or X-linked type of inheritance, as well as one homozygous P/LP variant or two (potentially) compound heterozygous P/LP variants with a recessive type of inheritance;
– partial MGD: one heterozygous P/LP variant in a gene associated with an autosomal recessive disease;
– potential MGD: one VUS clearly associated with the phenotype with autosomal dominant or X-linked inheritance, or two VUSes clearly associated with the phenotype with autosomal recessive inheritance;
– no MGD: absence of identified variants, one P/LP variant or VUS in a gene with autosomal recessive inheritance, incomplete relation of VUS associated with autosomal dominant, or X-linked inheritance to clinical presentation;
– incidental findings: only AC variants.

Statistical analysis and graph plotting were performed using the RStudio programming environment. Patient data and genetic variants were imported from a proprietary database (Glotov et al., 2025). They are available upon request.

Results

Characteristics of the examined cohort of patients

From 2021 to 2024 inclusive, 286 probands with suspected genetic diseases were referred from the regional medical genetics counseling service for whole-exome sequencing. The gender, age, and ethnic compositions of the subjects are presented in Table 1. The majority of patients (86.7 %) belong to the child-adolescent age group, the mean age of the cohort being 10.9 years. The most represented ethnic groups are Russians (65.4 %) and Tatars (11.5 %).

Regarding the structure of symptom categories (Table 2), neurological disorders occupy the leading position in Yugra (41.3 %). They are followed by dysmorphic syndromes (12.2 %) and metabolic disorders (11.2 %). It should be noted that psychiatric disorders are the only category not represented in the study group.

Clinical and molecular characteristics of identified genetic variants related to the phenotype

A total of 420 genetic variants (SNVs, InDels) were identified in patients. These variants were associated either with the observed phenotype or with the asymptomatic heterozygous carrier status (AC) of various monogenic diseases. Although, as shown in Figure 1, most of the identified variants are not of confirmed clinical significance, their potential contribution to disorder manifestation cannot be ignored.

The clinically significant variants detected (194 out of 420) (Fig. 2) belong to the following classes: missense, 43.3 %; stop-gained, 21.6 %; frameshift deletions, 20.1 %; splice donor/acceptor variants, 8.2 %; frameshift insertions, 5.2 %; inframe deletions, 1.5 %. It should be noted that the class of repeat length variation is not represented in our study, as the bioinformatics algorithm used has not undergone strict validation for the analysis of this variation type.

The inheritance patterns of identified clinically significant variants are presented in Figure 3. Gene variants with the autosomal recessive inheritance pattern have a significant

Table 1. Demographic characteristics of the studied cohort (*N* = 286)

Characteristic	Patients, no. (%)	Characteristic	Patients, no. (%)
Gender		Ethnic structure	
Male	159 (55.6)	Russians	187 (65.4)
Female	127 (44.4)	Tatars	33 (11.5)
Onset age		Azerbaijanis	11 (3.8)
Neonatal	38 (13.3)	Tajiks	10 (3.5)
Infancy	85 (29.7)	Ukrainians	8 (2.8)
Childhood	128 (44.8)	Chechens	8 (2.8)
Adolescent	18 (6.3)	Dagestanis	7 (2.4)
Adult	17 (5.9)	Uzbeks	6 (2.1)
Median onset age [Q1; Q3]	2 [0.5; 4.5]	Kumyks	3 (1.1)
Age at sampling		Armenians	3 (1.1)
Neonatal	1 (0.4)	Belarusians	2 (0.7)
Infancy	4 (1.4)	Germans	2 (0.7)
Childhood	180 (62.9)	Kazakhs	2 (0.7)
Adolescent	64 (22.4)	Moldovans	2 (0.7)
Adult	37 (12.9)	Bashkirs	1 (0.4)
Median age at sample accession [Q1; Q3]	8.4 [3.8; 14.9]	Lezgians	1 (0.4)

Table 2. Categories of clinical symptoms (*N* = 286)

Symptom category	Patients, no. (%)	Symptom category	Patients, no. (%)
Neurological disorders	118 (4.3)	Autoimmune or rheumatological disorders	8 (2.8)
Dysmorphic and congenital abnormality syndromes	35 (12.2)	Cardiovascular disorders	7 (2.4)
Metabolic disorders	32 (11.2)	Gastrointestinal and hepatic disorders	7 (2.4)
Hearing and ear disorders	16 (5.6)	Endocrine disorders	4 (1.4)
Skeletal disorders	15 (5.2)	Ophthalmological disorders	3 (1.1)
Renal and urinary tract disorders	12 (4.2)	Respiratory disorders	2 (0.7)
Neurodevelopmental disorders	9 (3.1)	Growth disorders	1 (0.4)
Hematological and immunological disorders	8 (2.8)	Tumor syndromes	1 (0.4)
Dermatological disorders	8 (2.8)		

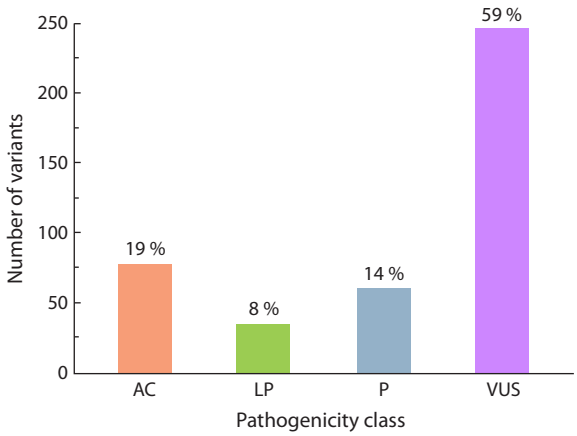

Fig. 1. Pathogenicity classes of the identified variants.

AC – asymptomatic carrier; LP – likely pathogenic; P – pathogenic; VUS – variant of unknown significance.

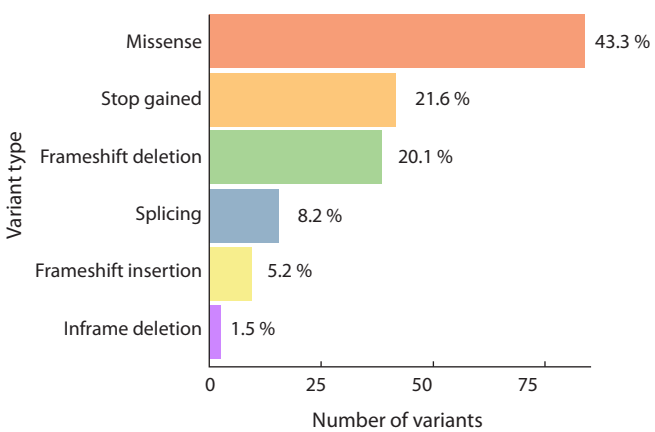

Fig. 2. Molecular classes of clinically significant variants.

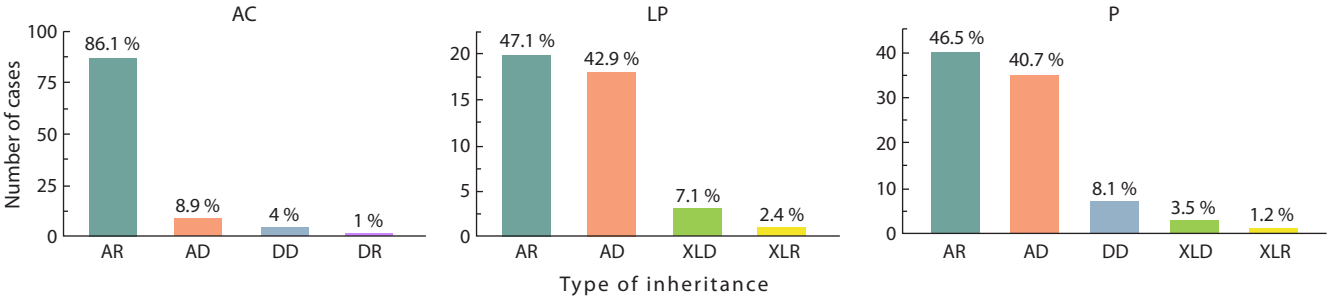

Fig. 3. Distribution of pathogenicity classes by inheritance types.

AR – autosomal recessive; AD – autosomal dominant; DD – digenic dominant; DR – digenic recessive; XLD – X-linked dominant; XLR – X-linked recessive.

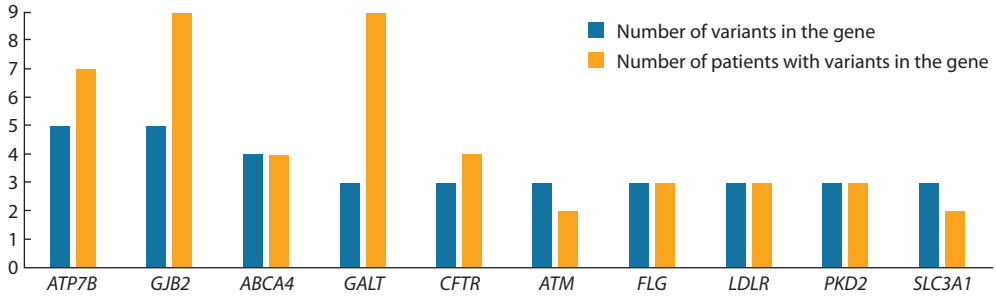

Fig. 4. Human genes with the largest numbers of variants in the studied cohort.

Table 3. The most common clinically significant variants in the studied cohort

cDNA variant	Patho-genicity	Inheritance	Position (hg19)	gnomAD AF, %	Mutation type	No. of patients	Relationship with diseases
ATP7B c.3207C>A	P, AC	AR	chr13:52518281G>T	0.1524	Missense	6	Wilson disease
GJB2 c.109G>A	P, AC	AR, DD	chr13:20763612C>T	0.1392	Missense	4	GJB2-related deafness
GJB2 c.35delG	P, AC	AR, DD	chr13:20763685AC>A	0.9578	Frameshift deletion	4	GJB2-related deafness
PAH c.1222C>T	P, AC	AR	chr12:103234271G>A	0.1719	Missense	4	Phenylketonuria
NEB c.23989C>T	AC	AR	chr2:152357937G>A	0.05838	Stop gained	3	Nemaline myopathy 2
C2 c.841_ 868del	AC	AR	chr6:31902065ATGGTGGACAG GGTCAGGAATCAGGAGTC>A	0.7159	Frameshift deletion	2	C2-complement deficiency
CFTR c.274G>A	AC	AR	chr7:117170953G>A	NA	Missense	2	Cystic fibrosis
KIAA0586 c.392del	AC	AR	chr14:58899156AG>A	0.4773	Frameshift deletion	2	Joubert syndrome
OTOA c.2359G>T	LP	AR	chr16:21747639G>T	0.03104	Stop gained	2	Autosomal recessive deafness 22

weight in the cohort studied (47.6 % for LP and 46.5 % for P). Autosomal dominant inheritance is observed for a substantial number of P/LP variants (40.7 and 42.9 %, respectively). X-linked inheritance has a relatively low prevalence in the cohort, accounting for no more than 10 % of all likely pathogenic variants and no more than 5 % of all pathogenic variants.

Among the genes with the greatest number of detected variants, as shown in Figure 4, the most prominent are *ATP7B*, *GJB2*, *ABCA4*, and *GALT*, which are associated with Wilson's disease, GJB2-related deafness, Stargardt disease, and galactosemia, respectively.

The most frequent clinically significant variant, as presented in Table 3, is c.3207C>A in the *ATP7B* gene, associated with Wilson's disease. However, the majority of the most frequent variants were detected in the AC (asymptomatic carrier) state. All most frequent variants detected are described in the literature as pathogenic, including the *CFTR* c.274G>A (p.Glu92Lys) variant (Chuvash mutation), which is absent from the gnomAD database.

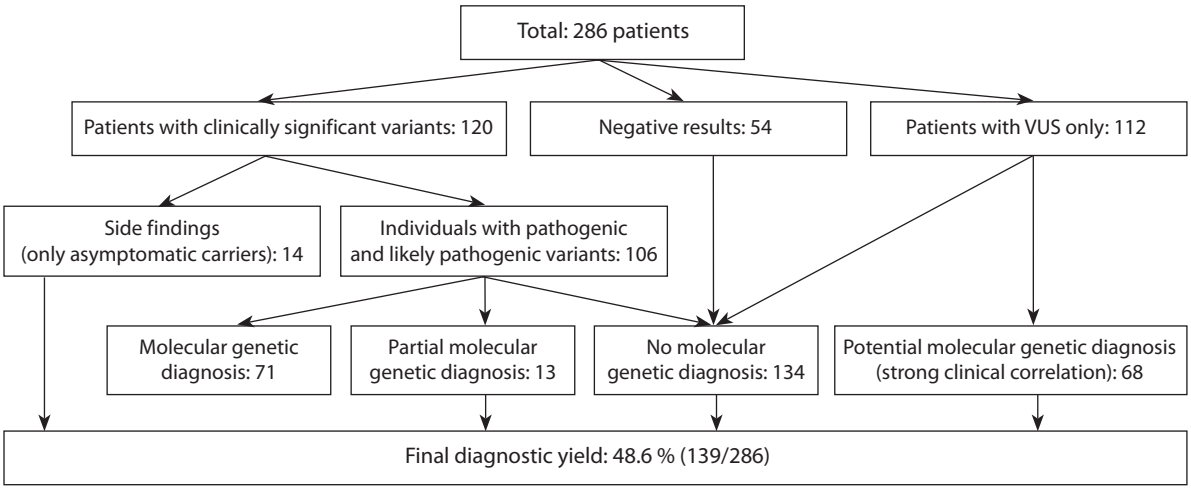

Fig. 5. WES testing summary.

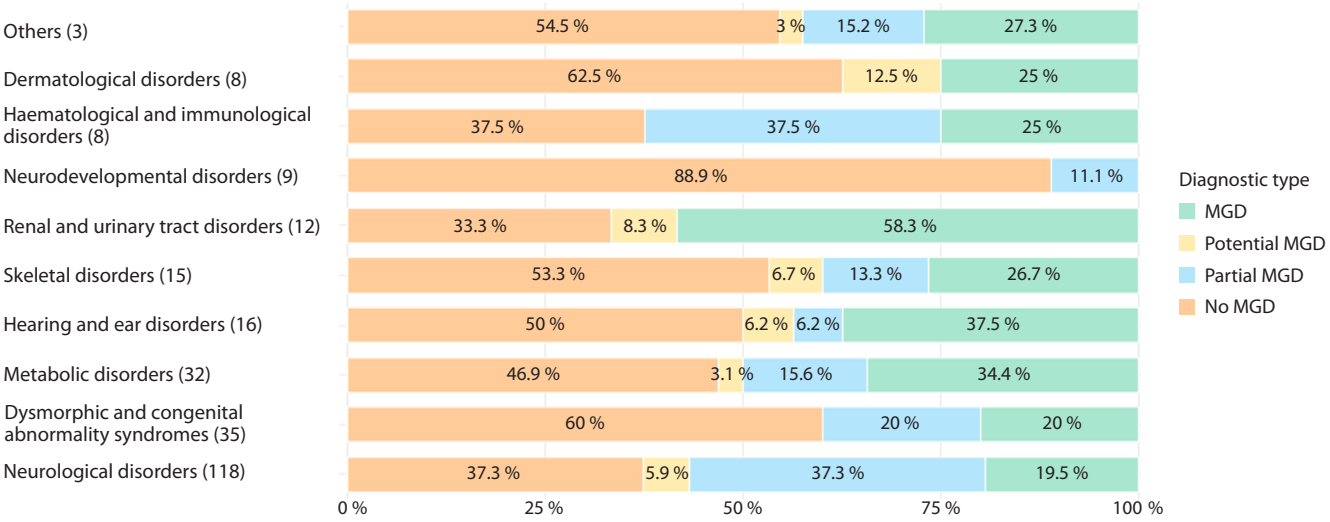

Fig. 6. Types of molecular genetic diagnosis (MGD) in the most common disorder groups.

Clinical effectiveness of WES in the examined cohort

As shown in Figure 5, WES results produced no findings in 54 out of 286 examined patients. Variants of P/LP and AC classes were found in 120 patients, with 14 of them having only AC as a secondary finding. Among 106 patients with phenotype-relevant variants, 71 received a molecular genetic diagnosis, and 13 had only a partial diagnosis. Among the 112 studied samples with only VUS, 68 patients had variants that allowed for a potential molecular genetic diagnosis. Thus, the diagnostic effectiveness of WES in the examined cohort was 48.6 % (139/286).

The distribution of molecular genetic diagnosis types varies significantly across different symptom categories (Fig. 6). The highest percentages of patients with confirmed molecular genetic diagnosis belong to the following categories: *Renal and urinary tract disorders*, 58.3 % (7/12 patients); *Hearing and ear disorders*, 37.5 % (6/16 patients), and *Metabolic disorders*, 34.4 % (11/32 patients). The lowest diagnostic effectiveness among the most common categories is observed in the *Neurodevelopmental disorders* group, where 88.9 % of patients (8/9) received no molecular genetic diagnosis.

The most ambiguous molecular diagnoses are observed in *Haematological and immunological disorders* – 37.5 % with potential MD (3/8 patients) and *Neurological disorders* – 37.3 % with potential MD (44/118 patients). The presence of a large number of variants of uncertain clinical significance (VUS) explains the ambiguity in these groups.

Discussion

We performed whole-exome sequencing for 286 probands with suspected hereditary diseases to clarify the molecular genetic diagnosis. The majority of patients (86.7 %) belonged to the child-adolescent age group, which is associated with the early manifestation of hereditary diseases. Male subjects were more numerous in the group studied (55.6 %), owing to the presence of X-linked recessive disorders.

The diagnostic effectiveness of the method was 24.8 % for cases with complete molecular genetic diagnosis and increased up to 48.6 % when taking into account variants of uncertain significance that have a close relationship with the observed phenotype. These variants may be reclassified using updated information in the foreseeable future, as with similar findings from studies in other countries (Han et al., 2025).

As long as WES testing was conducted only for probands, the experience of other researchers brought us to the assumption that diagnostic effectiveness could be increased significantly through the implementation of duo- and trio-WES (Lai et al., 2024), mainly due to more reliable identification of compound heterozygous variants and *de novo* variants (Tan et al., 2019). This is particularly relevant for the studied population, where the proportion of autosomal recessive pathogenic or likely pathogenic variants is close to 50 %

Additionally, diagnostic effectiveness could be improved through re-analysis of exome data due to the constant emergence of new data about the relationships between variants and phenotypes (Arteche-López et al., 2022). Finally, since approximately 13 % of genetic variability is associated with copy number variations (CNVs) (Stankiewicz, Lupski, 2010), we conceive that it would be desirable for patients with negative WES results to undergo whole-genome sequencing (WGS) or low-coverage genome sequencing in combination with re-analysis of WES (Moey et al., 2025).

We found that the most common symptom categories of hereditary diseases in Yugra were *Neurological disorders* (118 out of 286 cases), *Dysmorphic and congenital abnormality syndromes* (35/286); and *Metabolic disorders* (32/286).

Our study shows significant differences in WES diagnostic effectiveness among different symptom categories: from zero yield for percentage of confirmed molecular genetic diagnosis cases in the *Neurodevelopmental disorders* group to 58.3 % in the *Renal and urinary tract disorders* group. Thus, the high occurrence of neurological disorders in our cohort presumes that just nervous system disorders pose the greatest challenge to productive WES analysis. Such differences may be related to yet uncharacterized molecular causes of these diseases (Adams, Eng, 2018), and generally they are consistent with results obtained from the analysis of 3,040 WES samples, where a wide range from 4 to 55 % of positive reports (Retterer et al., 2016) was also observed in different disease groups.

It is also important to consider the relevance of applying special molecular genetic methods to specific inherited diseases, such as search for duplications on chromosome 17 in cases of hereditary motor-sensory neuropathies (Shchagina et al., 2020), which was not conducted within the scope of our study. Other studies note the effectiveness of duo- and trio-WES for patients with intellectual disability, developmental delays, and epilepsy due to the high number (up to 80 %) of *de novo* variants in these cases (Lai et al., 2024).

It is particularly important to emphasize the significance of providing sufficient data on clinical picture by referring physicians (mostly by geneticists), as incomplete understanding of the symptoms by the interpreter may lead to false elimination of many genes from the search criteria and produce false-negative results. Additionally, for multiethnic populations determining the patient's ethnicity (at least based on a questionnaire) is an important aspect, which was considered in our study, involving 16 ethnic groups.

Furthermore, the importance of secondary findings classified in our study as the asymptomatic carrier state (about 20 % of all identified variants) should not be underestimated, as they provide insight into the genetic burden of the population and should be appropriately considered during genetic counseling of families, particularly for planning future pregnancies.

It can be noted that 8.9 % of asymptomatic carrier cases among the examined patients are referred to genes with an autosomal dominant inheritance pattern. However, for these cases no significant effect on the phenotype was observed, which may be due to the different inheritance patterns of the variant depending on its belonging to a particular protein domain, as in the case of *GJB2* (Xiang et al., 2023). Another consideration is the risk of mutation manifestation at an older age, as in the case of mutations in the *BRCA1* and *BRCA2* genes (Azzollini et al., 2016).

Conclusions

In our study of 286 residents of the KHMAO-Yugra region, we accurately established molecular genetic diagnoses in 24.8 % of patients by using the whole-exome sequencing approach and identified the possible genetic nature of inherited diseases in 48.6 % of cases within the ethnically heterogeneous group. We characterized 420 gene variants, excluding benign and likely benign ones. Thus, whole exome sequencing should be considered a sufficiently effective first-tier diagnostic method for unveiling the genetic causes of a broad range of presumably hereditary diseases. However, some specific cases may demand duo- or trio-WES, whole-genome sequencing, or additional specific methods well proven in revealing the molecular background of certain classes of molecular disorders.

References

Adams D.R., Eng C.M. Next-generation sequencing to diagnose suspected genetic disorders. *N Engl J Med*. 2018;379(14):1353-1362. doi 10.1056/NEJMra1711801

Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S., Sunyaev S.R. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4): 248-249. doi 10.1038/nmeth0410-248

Arteche-López A., Ávila-Fernández A., Riveiro Álvarez R., Almoguera B., Bustamante Aragonés A., Martin-Merida I., López Martínez M.A., … Blanco-Kelly F., Tahsin Swafiri S., Lorda Sánchez I., Trujillo Tiebas M.J., Ayuso C. Five years' experience of the clinical exome sequencing in a Spanish single center. *Sci Rep*. 2022;12(1): 19209. doi 10.1038/s41598-022-23786-6

Azzollini J., Scuvera G., Bruno E., Pasanisi P., Zaffaroni D., Calvello M., Pasini B., Ripamonti C.B., Colombo M., Pensotti V., Radice P., Peissel B., Manoukian S. Mutation detection rates associated with specific selection criteria for BRCA1/2 testing in 1854 high-risk families: a monocentric Italian study. *Eur J Intern Med*. 2016;32: 65-71. doi 10.1016/j.ejim.2016.03.010

Barbitoff Y.A., Khmelkova D.N., Pomerantseva E.A., Slepchenkov A.V., Zubashenko N.A., Mironova I.V., Kaimonov V.S., … Aseev M.V., Shcherbak S.G., Glotov O.S., Isaev A.A., Predeus A.V. Expanding the Russian allele frequency reference via cross-laboratory data integration: insights from 7452 exome samples. *Natl Sci Rev*. 2024; 11(10):nwae326. doi 10.1093/nsr/nwae326

Cingolani P., Platts A., Wang le L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., Ruden D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3. *Fly* (*Austin*). 2012;6(2):80-92. doi 10.4161/fly.19695

Davydov E.V., Goode D.L., Sirota M., Cooper G.M., Sidow A., Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* 2010; 6(12):e1001025. doi 10.1371/journal.pcbi.1001025

Glotov O.S., Donnikov M.Yu., Glotov A.S., Popova E.A., Suchko P.A., Kolbasin L.N., Belotserkovtseva L.D., Kovalenko L.V. Genetic variants identified in patients with rare and multifactorial diseases in the Khanty-Mansi Autonomous Okrug – Yugra based on exome sequencing: registered database No. 2025622641. 2025. Available: https://new.fips.ru/publication-web/publications/document?type=doc&tab=PrEVM&id=60EDD49E-F82F-481E-BFB7-474B01087E34, accessed on 28.09.2025 (in Russian)

Han H., Seo G.H., Hyun S.I., Kwon K., Ryu S.W., Khang R., Lee E., … Yang S., Lee S., Jang S., Lee J., Lee H. Exome sequencing of 18,994 ethnically diverse patients with suspected rare Mendelian disorders. *NPJ Genom Med*. 2025;10(1):6. doi 10.1038/s41525-024-00455-3

Ioannidis N.M., Rothstein J.H., Pejaver V., Middha S., McDonnell S.K., Baheti S., Musolf A., … Bailey-Wilson J.E., Radivojac P., Thibodeau S.N., Whittemore A.S., Sieh W. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877-885. doi 10.1016/j.ajhg.2016.08.016

Kumar P., Henikoff S., Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081. doi 10.1038/nprot.2009.86

Lai G., Gu Q., Lai Z., Chen H., Chen J., Huang J. The application of whole-exome sequencing in the early diagnosis of rare genetic diseases in children: a study from Southeastern China. *Front Pediatr*. 2024;12:1448895. doi 10.3389/fped.2024.1448895

Li H. Exploring single-sample SNP and INDEL calling with whole-genome *de novo* assembly. *Bioinformatics*. 2011;28(14):1838-1844. doi 10.1093/bioinformatics/bts280

Moey L.H., Seo G.H., Cheah B.E., Keng W.T., Lee H., Ch'ng G.S. A first large study of whole-exome sequencing (WES) in 489 patients with suspected rare genetic disorders at a tertiary centre in Malaysia. *Rare*. 2025;3:100102. doi 10.1016/j.rare.2025.100102

Okuneva E.G., Kozina A.A., Baryshnikova N.V., Krasnenko A.Yu., Klimchuk O.I., Stetsenko I.F., Plotnikov N.A., Surkova E.I., Ilinsky V.V. The utility of exome sequencing in diagnosis of hereditary diseases. *Med Genet*. 2020;19(12):18-24. doi 10.25557/2073-7998.2020.12.18-24 (in Russian)

Pashaei E., Yilmaz A., Ozen M., Aydin N. Prediction of splice site using AdaBoost with a new sequence encoding approach. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary. IEEE, 2016;003853-003858. doi 10.1109/SMC.2016.7844835

Petersen B.S., Fredrich B., Hoeppner M.P., Ellinghaus D., Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18(1):14. doi 10.1186/s12863-017-0479-5

Quang D., Chen Y., Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015; 31(5):761-763. doi 10.1093/bioinformatics/btu703

Retterer K., Juusola J., Cho M.T., Vitazka P., Millan F., Gibellini F., Vertino-Bell A., … Richard G., Brandt T., Haverfield E., Chung W.K., Bale S. Clinical application of whole-exome sequencing across clinical indications. *Genet Med*. 2016;18(7):696-704. doi 10.1038/gim.2015.148

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., Grody W.W., Hegde M., Lyon E., Spector E., Voelkerding K., Rehm H.L.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi 10.1038/gim.2015.30

Ross J.P., Dion P.A., Rouleau G.A. Exome sequencing in genetic disease: recent advances and considerations. *F1000Res*. 2020;9(F1000 Faculty Rev):336. doi 10.12688/f1000research.19444.1

Ryzhkova O.P., Kardymon O.L., Prohorchuk E.B., Konovalov F.A., Maslennikov A.B., Stepanov V.A., Afanas'ev A.A., … Kostareva A.A., Pavlov A.E., Golubenko M.V., Polyakov A.V., Kucev S.I. Guidelines for the interpretation of massive parallel sequencing variants (update 2018, v2). *Med Genet.* 2019;18(2):3-24. doi 10.25557/2073-7998.2019.02.3-23 (in Russian)

Shchagina O.A., Ryzhkova O.P., Chuhrova A.L., Milovidova T.B., Gundorova P., Mironovich O.L., Orlova A.A., Orlova M.D., Polyakov A.V. Diagnostic utility of exome sequencing for inherited peripheral neuropathies. *Neuromuscular Disord*. 2020;10(4):12-26. doi 10.17650/2222-8721-2020-10-4-12-26 (in Russian)

Stankiewicz P., Lupski J.R. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437-455. doi 10.1146/annurev-med-100708-204735

Stranneheim H., Lagerstedt-Robinson K., Magnusson M., Kvarnung M., Nilsson D., Lesko N., Engvall M., … Soller M.J., Nordgren A., Wirta V., Lindstrand A., Wedell A. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med*. 2021;13(1):40. doi 10.1186/s13073-021-00855-5

Strauch Y., Lord J., Niranjan M., Baralle D. CI-SpliceAI-improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. *PLoS One.* 2022;17(6):e0269159. doi 10.1371/journal.pone.0269159

Sundaram L., Gao H., Padigepati S.R., McRae J.F., Li Y., Kosmicki J.A., Fritzilas N., Hakenberg J., Dutta A., Shon J., Xu J., Batzoglou S., Li X., Farh K.K. Predicting the clinical impact of human mutation with deep neural networks. *Nat Genet*. 2018;50(8):1161-1170. doi 10.1038/s41588-018-0167-z

Tan T.Y., Lunke S., Chong B., Phelan D., Fanjul-Fernandez M., Marum J.E., Kumar V.S., … Martyn M., Goranitis I., Thorne N., Gaff C.L., White S.M. A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *Eur J Hum Genet.* 2019;27(12):1791-1799. doi 10.1038/s41431-019-0471-9

Van der Auwera G.A., O'Connor B.D. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media, 2020

Wang Q., Shashikant C.S., Jensen M., Altman N.S., Girirajan S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep*. 2017;7(1):885. doi 10.1038/s41598-017-01005-x

Xiang J., Sun X., Song N., Ramaswamy S., Abou Tayoun A.N., Peng Z. Comprehensive interpretation of single-nucleotide substitutions in *GJB2* reveals the genetic and phenotypic landscape of *GJB2*-related hearing loss. *Hum Genet*. 2023;142(1):33-43. doi 10.1007/s00439-022-02479-0

Zinchenko R.A., Kutsev S.I., Aleksandrova O.Yu., Ginter E.K. Main methodological approaches to the identification and diagnosis of monogenic hereditary diseases and problems in the organization of medical care and unified preventive programs. *Probl Social Hyg Public Health Hist Med.* 2019;27(5):865-877. doi 10.32687/0869-866X-2019-27-5-865-877 (in Russian)

Conflict of interest. The authors declare no conflict of interest.
Received September 29, 2025. Revised December 24, 2025. Accepted February 9, 2026.