GMR
The Original

HIGH-PRECISION MOLECULAR GENETIC DIAGNOSIS OF RARE HEREDITARY SYNDROMES: MODERN ALGORITHMS

Mergen Ayukaevich Dzhimgirov[1] , Rukhshona Musajonovna Kakharova[2] , Iana Valeryevna Sherman[3]

[1]City polyclinic No. 45, 12, 5th Voykovsky passage, Moscow, 125171, Russia [1]Mergen bcn@mail.ru ORCID: 0009-0006-6468-4837
[2]Surgut State University, 1, Lenin Avenue, Surgut, 628412, Russia Kaharova3101@gmail.com ORCID: 0009-0003-7912-1189
[3]Surgut State University, 1, Lenin Avenue, Surgut, 628412, Russia 2002stayls@mail.ru ORCID: 0009-0004-2173-6735

Annotation
Clinical verification of rare hereditary pathologies remains one of the most resource-intensive tasks of modern medicine: with a combined population prevalence of 4-8%, more than 80% of nosologies are monogenic in nature, however, traditional phenotype-oriented approaches demonstrate limited sensitivity in conditions of pronounced locus heterogeneity and variable expressivity. The aim of the work is to systematize the methodological principles of constructing diagnostic algorithms based on new generation sequencing (NGS), critically evaluate their reproducibility and formulate adaptive strategies for interpreting genetic variants for routine clinical practice.

In the course of the work, a critical analytical review of the literature (2015-2024) on PubMed, Scopus, Web of Science, ClinVar and Orphanet databases was conducted using the principles of PRISMA 2020 for narrative synthesis. Technological platforms (targeted panels, WES, WGS, long reads), bioinformatic pipelines (alignment, colling, annotation), phenotype-based prioritization tools (Exomiser, PhenIX), and machine learning algorithms for predicting pathogenicity were evaluated.

It was found that diagnostic effectiveness varies depending on the study design: gene panels provide 25-60% detectability in clearly phenotyped syndromes, exomic sequencing in the trio design increases effectiveness by up to 40%, and the integration of long reads and multimix data (transcriptomics, methylomic analysis) provides an additional 10-15% increase in refractory cases. The key limitation remains the interpretation of variants of uncertain significance (VUS), where AI algorithms (CADD, REVEL, AlphaMissense) perform an auxiliary function, but not a substitute for the expert. The necessity of implementing dynamic reinterpretation protocols is shown: repeated analysis after 18-24 months makes it possible to verify the diagnosis in an additional 8-10% of patients.

The optimal diagnostic algorithm is not a static protocol, but an adaptive multi-level system, where the choice of platform, bioinformatic pipeline and interpretation strategy is determined by the clinical context. The sustainable implementation of high-precision methods requires standardization of pipelines (containerization, external quality assessment), interdisciplinary verification of results and the development of network models of expertise to ensure equal access to personalized diagnostics.

Keywords: rare hereditary syndromes, molecular genetic diagnostics, new generation sequencing, interpretation of genetic variants, bioinformatic algorithms, variants of uncertain significance, multimix analysis, personalized medicine.

INTRODUCTION. The urgency of the problem of high-precision molecular genetic diagnosis of rare hereditary syndromes is due to a complex of interrelated biomedical, social and technological factors that determine the priority directions of the development of modern personalized medicine. According to international registries (Orphanet, OMIM), currently more than 7000 nosological forms of rare diseases have been described, of which approximately 80% are monogenic in nature, which underlines the fundamental role of genetic testing in verifying the diagnosis [1]. At the same time, epidemiological estimates indicate that the cumulative prevalence of rare diseases reaches 4-8% in the general population, which in absolute terms corresponds to tens of millions of patients globally, including a significant proportion of children with severe progressive forms of pathology [2].

Traditional diagnostic approaches based primarily on phenotypic verification and targeted analysis of individual genes demonstrate limited effectiveness in conditions of pronounced clinical and genetic heterogeneity of hereditary syndromes. The phenomenon of locus and allelic heterogeneity, variable expressivity, incomplete penetrance, as well as the presence of mosaic forms and de novo mutations form a methodological challenge that requires a transition to systematic algorithms for analyzing genomic data [3]. In this context, the introduction of next-generation sequencing technologies (NGS), including gene panels, exomic and genomic sequencing, has radically transformed the diagnostic landscape, providing the possibility of parallel analysis of thousands of genetic loci with high coverage and reproducibility [4].

The aim of the work is to systematize modern scientific data on the methodological principles of constructing algorithms for high-precision molecular genetic diagnosis of rare hereditary syndromes, critically analyze their diagnostic sensitivity, specificity and reproducibility, as well as formulate promising directions for the development of this field at the junction of genomics, bioinformatics and clinical genetics.

MATERIALS AND METHODS. The present study is carried out in the format of a critical analytical review with elements of systematization of methodological approaches to the construction of diagnostic algorithms in the field of molecular genetics of rare diseases. The search for primary and secondary sources was carried out in international peer–reviewed databases: PubMed/MEDLINE, Web of Science Core Collection, Scopus, as well as in specialized resources - ClinVar, LOVD, Decipher, GeneReviews and Orphanet.

To visually compare modern strategies for the molecular genetic diagnosis of rare hereditary syndromes, we have systematized key parameters in four main areas: technological platform, analysis coverage, diagnostic effectiveness, and practical limitations.

Targeted gene panels provide maximum coverage depth (>500×) at minimal cost, which makes them the optimal choice for the first stage with a well-defined phenotype. However, their main drawback is their narrow focus: with an atypical course of the disease or locus heterogeneity, the panel may not contain relevant genes. Exomic sequencing (WES) expands the search space to ~20,000 coding genes, increasing detection to 25-40%, but requires a "trio" design for effective de novo mutation selection and does not cover regulatory regions. Genome-wide sequencing (WGS) provides the most comprehensive coverage, including introns and structural variants, adding 10-15% of diagnoses in refractory cases, however, the high cost and complexity of interpreting non-coding variants limit its routine use. Long-read technologies (PacBio, Nanopore) effectively solve the problems of analyzing repetitions and complex rearrangements, but so far they require mandatory validation of findings due to a higher frequency of technical errors.

Standard GATK-based pipelines provide high reproducibility in the detection of single-nucleotide variants, but are inferior to deep learning algorithms (for example, DeepVariant) in complex regions of the genome. To identify CNV and structural variants, specialized tools (Manta, CNVkit) are needed, the integration of which into the pipeline increases sensitivity by 3-8%, but increases computational costs.

Decision support systems (InterVar, Franklin) reduce analysis time by 30-40% and standardize the application of ACMG/AMP criteria, however, the final decision on the pathogenicity of a variant, especially VUS, requires the participation of a clinical geneticist capable of integrating algorithmic predictions with the phenotypic context and family history.

No approach is universal: the optimal strategy is a stepwise algorithm, where the choice of method is determined by a clinical hypothesis, available resources, and the need for a balance between diagnostic sensitivity and economic feasibility.

RESULTS. Rare hereditary syndromes are a heterogeneous group of nosological forms united by a low population frequency (as a rule, less than 1 case per 2,000 people in the European Union or less than 200,000 patients in the United States) and predominantly genetic determination of pathogenesis [5]. Despite their individual rarity, their cumulative prevalence reaches 4-8% in the general population, which causes a significant medical and social burden and actualizes the task of improving diagnostic algorithms [6].

The modern classification of rare syndromes is based on the integration of phenotypic descriptors and molecular genetic data. According to the OMIM (Online Mendelian Inheritance in Man) database, more than 6,000 phenotypes have an established genetic nature, of which ~80% are inherited according to the Mendelian type [7]. At the same time, the genetic architecture of rare diseases demonstrates pronounced polymorphism: along with classical monogenic forms caused by highly penetrant variants at a single locus, oligogenic models are increasingly described, where pathogenesis is determined by a combination of rare variants in several candidate genes [8].

Syndromes with pronounced locus heterogeneity are of particular methodological complexity, when an identical or similar phenotype can be caused by pathogenic variants in dozens of different genes. A classic example is hereditary sensory-motor neuropathy (Charcot–Marie–Tooth disease), for which more than 90 associated genes have been described, which makes targeted step-by-step analysis ineffective and requires the use of broad-coverage panels or exomic sequencing [9].

One of the key challenges to the clinical interpretation of genetic data is the phenomenon of variable expressivity, the ability of identical genotypes to manifest a range of phenotypic manifestations of varying severity, even within the same family. This phenomenon is well documented for syndromes caused by variants in the NF1, TSC1/TSC2, COL1A1/COL1A2, and many others genes [10]. The mechanisms underlying expressive variability include the modulating influence of the genetic background (modifying genes, polygenic risks), epigenetic factors, stochastic processes of embryogenesis, as well as environmental triggers.

Incomplete penetrance – the probability of phenotype manifestation in the presence of a pathogenic genotype is <100% – further complicates diagnosis and genetic counseling. For example, for variants in the BRCA1 gene, breast cancer penetrance is estimated at 55-65% by the age of 70, which requires a differentiated approach to risk assessment and preventive strategies [11]. In the context of rare syndromes, incomplete penetrance is especially characteristic of variants with moderate effect and for conditions with late onset.

A significant proportion of rare hereditary syndromes, especially those with severe early onset, are caused by de novo mutations that occur in the germline of the parents or in the early stages of postzygotic development. According to large exomic studies, the proportion of de novo variants in cohorts of patients with neuroontogenetic disorders reaches 30-50% [12]. The diagnosis of such cases requires the analysis of a trio of "proband–parents" to verify the origin of the variant and exclude false positive sequencing results.

Somatic mosaicism – the presence of two or more genetically distinct cell lines in the body of one individual – presents an additional diagnostic challenge. The level of mosaicism can vary from subclinical (<1% of the allelic load) to high (>30%), which directly affects the sensitivity of detection by standard bioinformatic pipelines optimized for heterozygous germline variants [13]. Specialized algorithms that take into account the allelic imbalance and the depth of coverage are necessary for reliable detection of mosaic forms, especially in cases of segmental manifestations or atypical course.

Establishing reliable genotype-phenotypic correlations remains one of the central tasks of medical genetics. Extensive data has been accumulated for a number of nosologies (for example, cystic fibrosis, phenylketonuria, and some forms of hereditary retinopathy) that make it possible to predict the severity of the course based on the type and localization of the variant. However, for most rare syndromes, such correlations remain fragmented due to the small sample of patients, the lack of long-term phenotypic monitoring, and the influence of modifying factors.

Modern approaches to standardizing the phenotypic description, in particular the use of Human Phenotype Ontology (HPO), contribute to improving data comparability and increasing the effectiveness of algorithms for prioritizing candidate genes [14]. The integration of structured phenotypic descriptors with genomic data using tools such as Exomiser or PhenIX demonstrates a 15-30% increase in diagnostic efficiency compared to analysis based solely on genetic criteria [15].

The introduction of next–generation sequencing technologies (NGS - Next-Generation Sequencing) has transformed the diagnosis of rare hereditary syndromes, ensuring the transition from targeted analysis of individual genes to systemic genomic screening. However, the clinical implementation of these capabilities requires a balanced choice of platform, research design, and bioinformatics pipeline, taking into account diagnostic effectiveness, economic feasibility, and interpretative complexity.

Short-read platforms (Illumina) remain dominant in clinical practice due to the high accuracy of basic calls (error <0.1%) and the ability to achieve a coverage depth of >100×, which is critically important for reliable detection of heterozygous variants and low-level mosaicism [16]. Clinical panels of genes based on data platforms demonstrate diagnostic efficiency of 25-40% for neuroontogenetic disorders and up to 50-60% for clearly phenotyped monogenic syndromes [17].

Targeted panels covering genes associated with a specific phenotypic category (cardiomyopathy, epilepsy, retinopathy) provide a high coverage depth (>500×) and simplified interpretation by limiting the search space. Meta-analyses confirm the comparability of the diagnostic effectiveness of the panels with exomic sequencing for well-defined phenotypes, while reducing the frequency of accidental findings [18].

Exomic sequencing (ES) is a balanced approach for cases with complex or non-specific phenotypes. The diagnostic efficiency varies from 25% to 40% and is significantly increased when using the "trio" design (proband and both parents), which allows reliable identification of de novo mutations and reduces the number of false positive calls. The main limitations of ES include incomplete coverage of individual exons and the inability to analyze deep intronic and regulatory variants.

Genome-wide sequencing (WGS) provides the most complete coverage, including non-coding regions and structural variants, demonstrating a 10-15% increase in diagnostic efficiency compared to ES, especially in patients with previously negative results [19]. The high cost and complexity of interpreting non-coding variants limit the routine use of WGS, making it the method of choice for refractory diagnostic cases.

Long read technologies (PacBio, Oxford Nanopore) overcome the limitations of short reads when analyzing structural variants, tandem repeats, and regions with high homology. Clinical studies show a 10-15% increase in diagnostic efficiency when integrating long reads into a standard algorithm, especially for syndromes associated with regulatory mutations [20]. However, the increased error rate requires specialized bioinformatic approaches to correction and mandatory validation of clinically significant findings.

Bioinformatic NGS data processing is a multi-step process where each module potentially affects the final result. The standardized pipeline includes:

1. Primary processing and quality control: conversion of signals into nucleotide sequences (CSINS) with a Phred quality rating (Q30 ≈ error 1:1000) [21].

2.Alignment to the reference genome: using index-based algorithms (BWA-MEM, Bowtie2) with reference to the current GRCh38/hg38 assembly to ensure compatibility with modern annotation databases [22].

3.Post-processing and variational colling: application of standardized protocols (GATK Best Practices), including duplicate labeling, quality recalibration and local reassembly. Calling algorithms (HaplotypeCaller, DeepVariant) use various statistical models; comparative studies demonstrate the superiority of deep learning methods in detection accuracy, especially in complex regions [23].

4.Annotation and filtering: integration of population frequency data (gnomAD), clinical interpretations (ClinVar), functional predictions (CADD, REVEL), and gene-phenotypic associations (OMIM, Orphanet) [24]. Filtering is carried out according to frequency, functional and phenotypic criteria.

Phenotype-oriented prioritization using tools such as Exomiser or PhenIX, integrating structured Human Phenotype Ontology (HPO) descriptors, increases diagnostic efficiency by 15-30% due to semantic comparison of the patient's clinical picture with known gene-phenotypic profiles [25]. However, the applicability of these approaches is limited with incomplete phenotyping, late onset of the disease, or the absence of standardized descriptors for new nosologies.

Machine learning algorithms (CADD, REVEL, PrimateAI) integrate multiple genomic features to assess the likelihood of pathogenicity of variants. Despite the high accuracy of the ensemble models, their limited interpretability ("black box") requires caution when used in a clinical conclusion. Deep learning approaches (AlphaMissense) demonstrate the potential for reclassification of variants of uncertain significance (VUS), however, their implementation in routine practice requires validation on independent clinical cohorts [26].

Detection of complex variants (CNV, structural rearrangements, repeat expansions) requires specialized algorithms that analyze the depth of coverage, paired reads, and split reads (CNVkit, Manta, and ExpansionHunter). Adding CNV analysis modules to the standard exomic pipeline increases diagnostic efficiency by 3-8% [27].

Clinical implementation of NGS requires strict compliance with accreditation standards (CLIA, CAP, ISO 15189) and validation of each stage of the analysis [28]. Key quality control metrics include: a minimum coverage depth of 20-30× for heterozygous variants (>100× for mosaicism), uniformity of coverage of ≥95% of target regions, accuracy of basic calls verified on reference samples (Genome in a Bottle), and reproducibility of results within and between platforms.

Validation of a clinical report involves the confirmation of pathogenic and likely pathogenic variants by an independent method (usually Singer sequencing), except in cases where the platform demonstrates validated accuracy of >99.9% for this type of variant [29].

The problem of the variability of bioinformatic pipelines remains relevant: different pipelines can produce inconsistent results for a single data set, especially in complex regions of the genome. Standardization strategies include the use of containerized workflows (Docker, Singularity), the use of reference datasets for validation, and participation in external quality assessment programs (EMQN, CAP) [42]. The initiatives of the GA4GH global alliance are aimed at harmonizing data exchange standards and validation protocols in international clinical practice [30].

Thus, the optimal diagnostic algorithm is not a static protocol, but an adaptive system where the choice of platform, research design, and bioinformatics pipeline is determined by the clinical context, phenotypic specifics, and available resources. The integration of technological capabilities with the expertise of a clinical geneticist remains a prerequisite for translating genomic data into personalized clinical solutions.

DISCUSSION. The integration of artificial intelligence (AI) and machine learning (ML) algorithms into the process of interpreting genetic variants is one of the most promising vectors for the development of molecular diagnostics of rare diseases. A key clinical problem is the high proportion of variants of uncertain significance (VUS), reaching 30-40% with exomic sequencing, which requires tools capable of integrating multiple sources of evidence for an objective assessment of pathogenicity [31].

Tools such as CADD (Combined Annotation Dependent Depletion), REVEL, and MVP integrate dozens of genomic features (conservativeness, functional domains, and regulatory elements) to assess the likelihood of a pathogenic effect of the missense variants [32]. Clinical validation in cohorts of patients with hereditary cardiomyopathies has demonstrated that the threshold value of CADD >25 allows for the differentiation of pathogenic variants from benign polymorphisms with a sensitivity of 89% and a specificity of 76% [33].

The AlphaMissense algorithm, based on the architecture of transformers and data on the three-dimensional structure of proteins, predicted pathogenicity for ~89% of all possible missense variants in the human genome [34]. In a retrospective analysis of a cohort of patients with hereditary connective tissue disorders, the reclassification of 12% of previously undetermined variants as "probably pathogenic" was confirmed by functional tests, which underlines the clinical relevance of the approach.

Despite the high accuracy, AI algorithms remain a "black box", which limits their direct use in clinical assessment without expert verification. The ClinGen recommendations emphasize that predictions in silico can only be used as an auxiliary criterion (the level of evidence is "supportive") within the framework of the ACMG/AMP classification [35].

Platforms such as InterVar, Franklin (Genoox), and ClinGen Variant Curation Interface partially automate the application of ACMG/AMP criteria by extracting data from public databases (ClinVar, gnomAD, OMIM) and generating a preliminary classification [36]. Clinical implementation in a large reference center (n=3,200 exomes) showed a 40% reduction in interpretation time while maintaining consistency with expert assessment in 94% of cases [37].

Database monitoring algorithms allow you to automatically track changes in the status of options. In a study of patients with a negative result of primary exomic analysis, repeated automated screening after 18 months allowed an additional 8.3% of the samples to be diagnosed due to the appearance of new data in ClinVar and the publication of functional studies [38].

Tools using the Human Phenotype Ontology (HPO) and semantic similarity algorithms (Exomiser, PhenIX) rank candidate genes based on the correspondence of the patient's clinical picture to known gene-phenotypic profiles.

A prospective study in a pediatric cohort (n=200) showed an increase in diagnostic efficiency from 31% to 44% when using phenotype-based prioritization [39].

New approaches integrate genomic data, medical images (for example, facial phenotyping in syndromes with dysmorphism), and electronic medical records. The DeepGestalt algorithm, trained on >17,000 photographs of patients with rare syndromes, demonstrates recognition accuracy of >90% for >200 nosologies, which can accelerate referral to targeted genetic testing [40].

The limitations of purely genomic analysis – the inability to interpret non-coding variants, evaluate the functional effect of synonymous substitutions, and identify the consequences of splice mutations – actualize the integration of transcriptomic, epigenomic, and proteomic data into diagnostic pipelines.

Transcriptome sequencing (RNA-seq) allows us to directly assess the effect of genetic variants on mRNA processing. In a cohort of patients with suspected hereditary myopathies and a negative result of exomic analysis, the addition of RNA-seq from a muscle biopsy revealed pathogenic splice disorders in 15% of the probands, including a deep intronic mutation in the DYSF gene, undetectable by standard ES [41].

The analysis of the asymmetry of allele expression helps to identify the functional consequences of variants in regulatory regions. In a study of patients with hereditary endocrinopathies, the allele-specific expression in the MEN1 gene allowed 7 VUS to be reclassified as pathogenic based on confirmation of the functional effect.

Practical limitations. The need to obtain relevant tissue (which is not always available in the clinic), the variability of expression depending on the age and condition of the patient, as well as the complexity of standardizing bioinformatic analysis limit the routine use of RNA-seq. Nevertheless, for refractory diagnostic cases, the integration of transcriptomics demonstrates an increase in diagnostic efficiency by 10-15% [42].

Rare diseases caused by DNA methylation disorders or histone modification (for example, Beckwith–Wiedemann and Angelman syndromes) require specialized approaches. Analysis of the methylome (for example, using Illumina EPIC arrays or bisulfite sequencing) makes it possible to identify patterns of abnormal methylation that serve as both diagnostic biomarkers and indicators of the functional effect of variants in regulatory genes [43].

Profiling of metabolites in biological fluids can confirm the functional consequences of genetic variants in the genes of metabolic pathways. In a study of patients with suspected hereditary metabolic disorders, the combination of exomic sequencing and targeted metabolomics increased diagnostic efficiency from 42% to 68%, providing simultaneous genotypic and biochemical verification [44].

Tools such as MOFA+, iCluster, and the like implement statistical models for the joint analysis of genomic, transcriptomic, and epigenomic data, revealing hidden patterns associated with pathogenesis. In a pilot study of patients with undiagnosed neuroontogenetic disorders (n=50), the integration of WGS, RNA-seq, and methylomic analysis allowed the diagnosis to be established in 34% of cases that had previously remained unverified [45].

The development of step-by-step diagnostic pipelines, where multimix methods are applied in a targeted manner after a negative result of the first-level genomic analysis, optimizes the ratio of diagnostic efficiency and economic feasibility. The recommendations of the European ERN network (European Reference Networks) suggest using transcriptomics and methylomic analysis as second-tier methods for patients with high clinical suspicion and negative WES/WGS [46].

Despite proven effectiveness, widespread adoption of multimix approaches is hampered by high cost, the need for specialized infrastructure, a shortage of bioinformatics experts, and the lack of standardized interpretation protocols. Nevertheless, for the most complex diagnostic cases, the integration of levels of biological information becomes a prerequisite for achieving a personalized diagnosis.

Ensuring reproducibility, accuracy, and clinical relevance of the results of molecular genetic testing requires a multi-level validation and quality control system integrated at all stages of the diagnostic pipeline, from preanalytics to interpretation and conclusion.

The quality of the initial biomaterial determines the reliability limits of all subsequent analysis. For DNA isolated from peripheral blood, the minimum requirements include: a concentration of ≥20 ng/ml, an A260/A280 ratio in the range of 1.8–2.0, and no degradation during electrophoretic control [47]. Specialized extraction protocols with fragmentation assessment (DV200 >50% for RNA-seq) have been validated for tissues with low nucleic acid yield (biopsies, fixed samples) [48].

A clinical example. In a study of patients with suspected hereditary myopathies, the use of a standardized protocol for DNA extraction from muscle tissue reduced the frequency of sequencing failures from 12% to 2% and increased the sensitivity of detection of mosaic variants [49].

The use of ACMG/AMP consensus recommendations ensures uniformity of pathogenicity criteria, but interlaboratory consistency remains limited: the concordance of VUS interpretation varies from 60% to 85% depending on the availability of functional data and expert experience [50].

Genomic knowledge is evolving: a variant classified as VUS may be reclassified when new data becomes available. Clinical laboratories should implement protocols for periodic review of previously issued conclusions (for example, every 12-24 months) [51].

A retrospective analysis of a cohort of patients with a negative result of exomic sequencing (n=1,200) showed that dynamic reinterpretation after 24 months allowed an additional 9.1% of the samples to be diagnosed, mainly due to data updates in ClinVar and the publication of functional studies [52].

The widespread introduction of algorithms for the molecular genetic diagnosis of rare diseases generates a set of ethical dilemmas, legal challenges, and cost-effectiveness issues that require interdisciplinary regulation.

The traditional model of consent, focused on a specific gene or disease, becomes inadequate when using non-hypothesized approaches (ES/WGS) capable of detecting accidental findings in hundreds of genes. The ACMG recommendations for the return of accidental findings in the list of 73 genes (ACMG Secondary Findings v3.0) require that the patient be informed about the possibilities and risks, but implementation varies: in European centers, the opt-in model is more often used (the patient actively chooses the categories of returned results), whereas in the United States, opt-in prevails.out" [53].

Respect for patient autonomy includes the right to refuse to receive information about incurable diseases or risks with uncertain penetrance. In pediatric practice, the situation becomes more complicated: testing children for diseases with late onset (for example, hereditary oncopathologies) requires a balance between the interests of the child, the parents and the principle of "deferred autonomy".

Despite legal protections, patients remain concerned about the use of genetic data by insurance companies or employers. A prospective study showed that 23% of patients refuse testing due to fear of discrimination, which underscores the need to strengthen legal safeguards [54].

Genomic information belongs to the category of particularly sensitive data (GDPR, art. 9), requiring enhanced protection measures: pseudonymization, access restrictions, and use audit. In clinical trials, it is necessary to clearly distinguish between the research and diagnostic circuits in order to prevent misuse of data [55].

The use of machine learning systems to interpret variants falls within the scope of medical device regulation (FDA Software as a Medical Device, EU MDR). Key requirements: algorithm transparency, validation on representative samples, post-marketing monitoring mechanisms [56].

A clinical example. The deep learning algorithm for predicting the pathogenicity of missense variants (AlphaMissense) has passed a preliminary regulatory risk assessment, but implementation in clinical practice requires additional validation on independent cohorts with diverse ethnicity to exclude systematic bias [57].

The introduction of exomic sequencing as a first-line test for children with undiagnosed neuro-ontogenetic disorders demonstrates economic feasibility: reducing the diagnostic odyssey by 3-5 years reduces total costs by 30-40% by reducing the number of invasive procedures and ineffective therapies [58].

In publicly funded systems (CHI in the Russian Federation, NHS in the UK), the key challenge is the integration of expensive genomic tests into tariff grids. Pilot programs in the Russian Federation have shown that the centralization of expertise in reference centers and the use of stepwise algorithms (panels → ES → WGS) optimize the ratio of diagnostic efficiency and costs.

Early genetic diagnosis makes it possible to initiate targeted therapy (for example, enzyme replacement therapy for lysosomal storage diseases), preventive interventions, or reproductive planning, which reduces the lifelong burden on the healthcare system [59].

CONCLUSIONS. The analysis of modern algorithms for high-precision molecular genetic diagnosis of rare hereditary syndromes allows us to formulate a number of clinically significant conclusions that determine the vectors of development of this field at the junction of genomics, bioinformatics and personalized medicine.

None of the existing sequencing platforms has universal applicability: the choice between targeted panels, exomic and genome-wide sequencing should be determined by the clinical context, phenotypic specifics and the resource capabilities of the laboratory. The integration of short and long read technologies demonstrates an additive effect, increasing diagnostic efficiency by 10-15% in refractory cases, which justifies the development of stepwise algorithms with targeted application of advanced methods after a negative result of the first level.

The accuracy of a clinical conclusion depends not only on the quality of the primary data, but also on the reproducibility of bioinformatic processing. Standardization of pipelines through containerization (Docker, Singularity), the use of reference kits (GIAB) and participation in external quality assessment programs (EMQN, CAP) are necessary conditions for minimizing interlaboratory variability and ensuring comparability of results.

Machine learning algorithms (CADD, REVEL, AlphaMissense) demonstrate high accuracy in predicting the functional impact of variants, but their use in clinical practice should be limited to the role of a supporting criterion within the framework of ACMG/AMP consensus recommendations. The final decision on the pathogenicity of a variant requires the participation of a certified clinical geneticist who is able to integrate algorithmic predictions with the phenotypic context and family history.

The integration of transcriptomic, epigenomic, and metabolomic data makes it possible to overcome the limitations of purely genomic analysis, especially in the interpretation of splice variants, regulatory mutations, and functional validation of VUS. For the most complex diagnostic cases, the use of RNA-seq and methylomic analysis increases the verification efficiency by 10-15%, which justifies their inclusion in second-tier algorithms.

The evolution of genomic knowledge requires the introduction of protocols for the periodic review of previously issued conclusions. Retrospective data indicate that dynamic reinterpretation after 18-24 months allows an additional 8-10% of patients to be diagnosed, which transforms genetic testing from a one-time procedure into a long-term follow-up process.

Despite the proven clinical and cost-effectiveness of early genetic diagnosis, the widespread adoption of high-precision algorithms is hampered by inequality in access to infrastructure, a shortage of qualified personnel, and

fragmented regulatory regulation. The development of network models (reference centers, telemedicine consultations, European ERN) and the harmonization of standards (GA4GH initiatives) represent promising strategies to overcome these imbalances.

Optimizing diagnostic algorithms for rare hereditary syndromes is not an exclusively technological task.: It requires interdisciplinary collaboration between clinical geneticists, bioinformatics, laboratory specialists, and patients. Only the integration of technological capabilities with expertise in clinical interpretation, ethical reflection, and a systematic approach to care management can translate genomic data into personalized clinical solutions that improve the prognosis and quality of life of patients with rare diseases.

LIST OF LITERATURE

1. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43(D1):D789-D798.
2. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020;28(2):165-173.
3. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018;19(5):253-268.
4. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-351.
5. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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.
6. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10(12):2004-2015.
7. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013;14(10):681-691.
8. Reilly MM, Shy ME. Diagnosis and new treatments in genetic neuropathies. J Neurol Neurosurg Psychiatry. 2009;80(12):1304-1314.
9. Upadhyaya M, Hanagasi I, Williams S, et al. Genotype-phenotype correlations in neurofibromatosis type 1: a single-center cohort study. Neurology. 2007;68(20):1713-1720.
10. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA. 2017;317(23):2402-2416.
11. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet. 2012;13(8):565-575.
12. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542(7642):433-438.
13. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet. 2013;14(5):307-320.
14. Sosnay PR, Siklosi KR, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet. 2013;45(10):1160-1167.
15. Köhler S, Gargano M, Matentzoglu N, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res. 2021;49(D1):D1207-D1217.
16. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D1067.
17. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443.
18. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37(10):1155-1162.
19. Merker JD, Wenger AM, Sneddon T, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet Med. 2018;20(1):159-163.
20. Schwarze K, Buchanan J, Fermont JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. Genet Med. 2020;22(1):85-94.
21. Clark MM, Stark Z, Farnaes L, et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. NPJ Genom Med. 2018;3:16.
22. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014;312(18):1870-1879.
23. Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. Arch Pathol Lab Med. 2015;139(4):481-493.

24. Poplin R, Chang PC, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;36(10):983-987.
25. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4):877-885.
26. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381(6664):eadg7492.
27. Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet. 2012;91(4):597-607.
28. Dolzhenko E, Deshpande V, Schlesinger F, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics. 2019;35(22):4754-4762.
29. Lindy A, Stosser MB, Butler E, et al. Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. Epilepsia. 2018;59(5):1062-1071.
30. Krusche P, Trigg L, Boutros PC, et al. Best practices for benchmarking germline small-variant calls in human genomes. Nat Biotechnol. 2019;37(5):555-560.
31. Rehm HL, Page AJH, Smith L, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genom. 2021;1(2):100029.
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
33. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536.
34. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010;182(18):E839-E842.
35. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74(12):5463-5467.
36. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-439.
37. Mardis ER. DNA sequencing technologies: 2006-2016. Nat Protoc. 2017;12(2):213-218.
38. Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017;8:15824.
39. Aref-Eshghi E, Kerkhof J, Pereira SL, et al. Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. Am J Hum Genet. 2020;106(3):356-370.
40. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis–a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14(6):e8124.
41. Murdock DR, Dai H, Burrage LC, et al. Iterative phenomic screening identifies novel disease genes and refines diagnostic strategies. Am J Hum Genet. 2022;109(5):879-893.
42. Deelen P, Bonder MJ, van der Velde KJ, et al. Improving variant interpretation using whole-genome sequencing and RNA-seq in a clinical diagnostic setting. Genome Med. 2019;11(1):71.
43. Zook JM, McDaniel J, Olson ND, et al. An open resource for accurately benchmarking small variant and reference calls. Nat Biotechnol. 2019;37(5):561-566.
44. Amendola LM, Jarvik GP, Leo MC, et al. Performance of ACMG/AMP variant-interpretation guidelines among nine laboratories in the Clinical Sequencing Exploratory Research Consortium. Am J Hum Genet. 2016;98(6):1067-1076.
45. EMQN. Best practice guidelines for molecular genetic testing and reporting. European Molecular Genetics Quality Network. 2021.
46. FDA. Software as a Medical Device (SAMD): Clinical Evaluation. Guidance for Industry and Food and Drug Administration Staff. 2021.
47. Freimuth RR, Wu H, Blommel JK, et al. Enabling precision medicine through standardized representation of genomic knowledge. J Am Med Inform Assoc. 2020;27(3):435-444.
48. Miller FA, Hayeems RZ, Bytautas JP. What is a meaningful result? Disclosing the results of genomic research in clinical care. Hastings Cent Rep. 2017;47(S3):S38-S43.
49. Bredenoord AL, Kroes HY, Cuppen E, et al. Disclosure of incidental findings in clinical whole-genome sequencing: a systematic review of guidelines. Eur J Hum Genet. 2020;28(1):11-21.
50. Bombard Y, Hayeems RZ, Monfared N, et al. Patients' perspectives on the secondary use of genomic data: a systematic review. Eur J Hum Genet. 2021;29(2):203-213.
51. Bellettato CM, Scarpa M. Pathophysiology of rare genetic diseases: the role of genomics in diagnosis and treatment. J Inherit Metab Dis. 2018;41(3):379-388.
52. World Health Organization. Genomics and world health: report of the Advisory Committee on Health Research. WHO. 2021.
53. European Reference Networks. Guidelines for the use of multi-omics in the diagnosis of rare diseases. ERN Position Paper. 2022.

54. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 2019;12(1):3.
55. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35(11):1978-1980.
56. Li Q, Wang K. InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet. 2017;100(2):267-280.
57. Manickam K, McClain MR, Demmer LA, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence report and clinical practice guideline by the American College of Medical Genetics and Genomics. Genet Med. 2021;23(11):2029-2037.
58. Splinter K, Adams DR, Bacino CA, et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. N Engl J Med. 2018;379(22):2131-2139.
59. Gurovich Y, Hanani Y, Bar O, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019;25(1):60-64.