Medical University – Pleven

Faculty of Pharmacy

Department of Medical genetics

Dr. Slavena Enkova Nikolova

Preconception carrier screening in couples for common monogenic disorders

- study of prevalence, reproductive risk evaluation, and preventive genetic strategies

A B S T R A C T

for awarding an educational and scientific degree

„DOCTOR”

Scientific specialty: Medical genetics

Scientific supervisor: Prof. Katya Kovacheva, MD, PhD

Pleven, 2026

Medical University – Pleven

Faculty of Pharmacy

Department of Medical genetics

Dr. Slavena Enkova Nikolova

Preconception carrier screening in couples for common monogenic disorders: study of prevalence, reproductive risk evaluation, and preventive genetic strategies

A B S T R A C T

for awarding an educational and scientific degree

„DOCTOR”

Scientific specialty: Medical genetics

Scientific supervisor: Prof. Katya Kovacheva, MD, PhD

Examination Committee:

Prof. Savina Petrova Hadjidekova, MD, PhD

Prof. Alexey Slavkov Savov, DSc

Assoc. Prof. Borislav Nikolov Popov, MD, PhD

Assoc. Prof. Nikola Kalinov Popovski, MD, PhD

Assoc. Prof. Dobrinka Hristova Gincheva, MD, PhD

Pleven, 2026

With gratitude to Dr. Zornitsa Kamburova for her support and valuable guidance in the process of developing the present dissertation.

The dissertation is written in 177 pages and is illustrated with 64 tables and 14 figures. The bibliography contains 197 references, of which 195 are in Latin script and 2 in Cyrillic.

The present research is fully funded by the project BG05M2OP001-1.002-0010-C01 "CENTRE OF COMPETENCE IN PERSONALIZED MEDICINE, 3D AND TELEMEDICINE, ROBOTIC AND MINIMALLY INVASIVE SURGERY" under the Operational Programme “Science and Education for Smart Growth” and the European Regional Development Fund.

The author is a full-time PhD student at the Department of Medical Genetics, Faculty of Pharmacy, Medical University – Pleven.

The dissertation was reviewed and approved for public defense by the Department of Medical Genetics at the Medical University – Pleven at an extended departmental council meeting held on March 10, 2026.

Examination Committee:

Chairman:

Assoc. Prof. Nikola Kalinov Popovski, MD, PhD

Members:

Prof. Savina Petrova Hadjidekova, MD, PhD

Prof. Alexey Slavkov Savov, DSc

Assoc. Prof. Borislav Nikolov Popov, MD, PhD

Assoc. Prof. Dobrinka Hristova Gincheva, MD, PhD

Reserve members:

Prof. Ivanka Istalianova Dimova, DSc

Prof. Tihomir Pankov Totev, MD, PhD

The public defense will take place on May 12, 2026, at 13:00, in “Galen” Hall of the Telecommunication Endoscopic Center, Medical University – Pleven.

The defense materials are available at the Research Department of the Faculty of Pharmacy, Medical University – Pleven, as well as on the university website: www.mu-pleven.bg.

CONTENT

ABBREVIATIONS USED

ACMG/AMP – American College of Medical Genetics and Genomics and the Association for Molecular Pathology
AD – autosomal dominant
AR – autosomal recessive
ARC – autosomal recessive condition
ClinVar – Database of clinically significant genetic variants
CF – carrier frequency
CS – carrier screening
ECS – expanded carrier screening
GC – genetic counselling
gnomAD – The Genome Aggregation Database
LOF – loss of function
MD – monogenic disorder
NGS – Next-generation sequencing
PV – pathogenic variant
PGT-M – preimplantation genetic diagnosis for monogenic disorders
PND – prenatal diagnosis
RC – recessive condition
RR – reproductive risk
WHO – World Health Organization
XLR – X-linked recessive
XRC – X-linked recessive condition

INTRODUCTION

Monogenic disorders (MD), which account for approximately 40% of rare diseases, represent a significant medical, social, and economic challenge in contemporary healthcare. Despite their low individual prevalence, they affect a substantial proportion of the population and are associated with high morbidity, mortality, chronic disease course, and considerable burden on both affected families and healthcare systems. Of particular importance are recessive conditions (RC), including autosomal recessive (ARC) and X-linked recessive (XRC) conditions, in which the presence of asymptomatic carriers and the universal carrier burden in the human population result in a persistent risk of having affected offspring, even in couples without a family history.

Carrier screening (CS) has emerged as the most effective strategy for identifying at-risk couples and supporting informed reproductive decision-making. Historical experience with targeted, ethnicity-based screening programs has demonstrated high effectiveness in specific populations and diseases. However, it has also revealed important limitations, including reduced informativeness in pan-ethnic populations, the risk of missing a substantial proportion of carriers, challenges in accurately defining ethnic background, and the potential for stigmatization and discrimination.

Advances in high-throughput genomic technologies, particularly the introduction of massively parallel sequencing (next-generation sequencing, NGS), have enabled the transition toward expanded carrier screening (ECS) with pan-ethnic applicability. ECS offers significantly higher detection rates of carriers and at-risk couples, lower residual risk following negative results, and more equitable access to screening. The universal nature of this approach also contributes to reducing psychosocial risks, including stigmatization, by normalizing carrier status as a common phenomenon in the general population.

These considerations motivated us to initiate the present study, based on a prospectively conducted carrier screening program among a cohort of reproductive couples from the Bulgarian population.

I. AIM

To determine the frequency and carrier profile of pathogenic variants in genes associated with recessive monogenic conditions in a cohort of individuals from the Bulgarian population, to analyze the resulting reproductive risks, and to develop an approach for genetic counseling in the context of carrier screening results.

II. OBJECTIVES

1. To determine the frequency of carrier states and the average number for pathogenic variants (carrier burden) in genes associated with recessive monogenic conditions in the studied cohort of individuals from the Bulgarian population.
2. To characterize the profile of affected genes in the studied cohort.
3. To investigate the molecular characteristics of pathogenic variants identified in the most frequently affected genes.
4. To analyze potentially at-risk couples and to identify those with a confirmed reproductive risk.
5. To evaluate cases with identified pathogenic variants in genes associated with the carrier's personal health risk.
6. To develop a comprehensive approach to genetic counseling based on the results of expanded carrier screening and findings related to both reproductive risk in couples and the individual health risk of carriers.

III. MATERIALS AND METHODS

1. Characteristics of the Study Participants

The present dissertation study includes 150 reproductive couples (300 individuals) from the Bulgarian population. All participants met the following inclusion criteria: reproductive age, planned or ongoing reproduction, clinically healthy status, no personal or family history of monogenic diseases, and absence of consanguinity between partners.

Of the 150 included couples, 100 had no history of adverse reproductive outcomes. Among them, 70 couples had not yet achieved reproduction and had not attempted conception at the time of blood sampling; 9 couples had one healthy child; in 16 couples, the female partner was pregnant at the time of enrollment; and in 5 couples, there was one healthy child and an ongoing pregnancy. Couples with an ongoing pregnancy were referred to a genetic counseling (GC) unit for maternal serum screening, during which participation in the present study was offered. The remaining couples without adverse reproductive outcomes were enrolled following voluntary participation after public recruitment.

The remaining 50 reproductive couples had a history of adverse reproductive outcomes without successful reproduction. They were referred for genetic counseling by an obstetrician–gynecologist, and participation in the dissertation study was offered during the consultation.

Collection of personal and family history data, as well as sampling of biological material (venous blood), was performed by a medical geneticist at the Medical Genetics Counseling Unit of the "Dr. Georgi Stranski" – Pleven University Hospital. Samples were collected between January 2020 and May 2024.

All participants were enrolled voluntarily after signing informed consent approved by the Ethics Committee of the Medical University – Pleven.

2. Main Methods Used in the Study

2.1 Medical Interview - Personal medical history was obtained through a structured clinical interview conducted by a medical genetics specialist as part of a GC session.

2.2 Genealogical Method – A pedigree was constructed for all partners in the enrolled

couples who met the inclusion criteria, encompassing relatives across three to four generations for each individual.

2.3 Molecular Methods – Biological material (approximately 5 mL of venous blood) was collected from all study participants in EDTA-containing vacutainer tubes. Genomic analysis was performed by using massively parallel sequencing methods (Next-Generation Sequencing, NGS), comprising the following main steps:

2.3.1 DNA Extraction – Genomic DNA was isolated from peripheral venous blood using an automated method with the MagCore Genomic DNA Whole Blood Kit (Ref: MGB400-02, RBC Bioscience), according to the manufacturer's instructions. The reliability and reproducibility of the DNA extraction method were ensured by a quality certificate provided by the reagent supplier. Following extraction, the DNA samples were stored at −80 °C in the DNA biobank of the Center of Competence "Leonardo da Vinci" at the Medical University – Pleven.

2.3.2 Massively Parallel Sequencing (Next-Generation Sequencing, NGS) – Genetic analysis of the isolated genomic DNA was performed using next-generation sequencing (NGS), enabling simultaneous analysis of multiple genes across a large number of samples. Library preparation was carried out using the targeted panel TruSight One Expanded (Illumina©), which covers the exonic regions and exon–intron boundaries of 6,699 clinically relevant genes. Due to its scope and targeted design, this panel is considered a clinical exome and is appropriate for pilot carrier screening studies.

Library preparation was performed according to the manufacturer's protocol. The resulting libraries were sequenced on an Illumina NextSeq 550 platform with a paired-end read configuration of 2 × 150 bp. The applied method allows reliable detection of small genetic variants, including single nucleotide variants (SNVs) and small insertions and deletions (indels), but is not informative for the detection of large genomic rearrangements.

2.3.3 Analysis of Sequencing Data – The DNA fragments generated from the analyzed samples were aligned to the human reference genome (hg19). The resulting output files in gVCF format were processed and analyzed using the BaseSpace Variant Interpreter (Illumina©) software platform. To optimize variant annotation and interpretation, predefined

custom filters were applied, including a minimum coverage depth of 20× for each variant, as well as exclusion of variants classified as benign or likely benign, and those of uncertain clinical significance (VUS).

The identified genetic variants were classified according to the recommendations of the American College of Medical Genetics and Genomics (ACMG), based on the five-tier classification system: pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB), and benign (B). Classification was performed according to established pathogenicity criteria, as detailed in Table 1.

Table 1. ACMG/AMP criteria for variant pathogenicity

Criteria for pathogenicity	Category
Very strong	**PVS1** - variant leading to loss of protein function (nonsense, frameshift, canonical±1 or 2 splice sites, initiation codon, exon/exon deletion)
Strong	**PS1** - amino acid substitution, at a site where a pathogenic variant has been previously reported **PS2** - de novo variant, proven by segregation analysis **PS3** - presence of experimental evidence of pathogenicity of the variant **PS4** - variant incidence in affected individuals is higher than incidence in healthy controls
Moderate-strength	**PM1** - localized hotspot of pathogenic variants **PM2** - absent in healthy controls, or extremely low frequency **PM3** - detected in *trans* with another pathogenic variant (for recessive conditions) **PM4** - leading to protein length change (for in-frame mutations) or protein elongating variants **PM5** - a novel amino acid substitution at a site where a pathogenic variant has been reported **PM6** - assumed de novo variant, no segregation analysis conducted
Supporting	**PP1** - co-segregation with disease, with many affected relatives in one family **PP2** - missense variant in a gene with few benign missense variants, in a gene for which missense variants are known to be a pathogenic mechanism **PP3** - lots of computer-processed evidence of variant pathogenicity **PP4** - Patient's clinical picture is highly specific for a pathogenic variant in this gene **PP5** - Lots of evidence in the literature for pathogenicity of the variant, but the lab can't prove it independently

The automatically detected variants were verified using multiple databases, including ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), dbSNP (https://www.ncbi.nlm.nih.gov/snp/), and Ensembl (https://www.ensembl.org/index.html).

Given the focus on carrier status for RC, genes with autosomal dominant (AD) and other non-recessive modes of inheritance were excluded from the analysis. Only genes with autosomal recessive (AR), X-linked recessive (XLR), and mixed (AR/AD) inheritance patterns were included. Only variants classified as clinically significant (pathogenic and likely pathogenic) were entered into an electronic database containing their classification and relevant characteristics.

2.4 Genetic Counseling – During post-test counseling, all non-risk couples were informed about the implications of a negative result. It was explained that the absence of shared carrier status for ARC in both partners, or carrier status for XRC, does not completely eliminate genetic risk. It was emphasized that the applied analytical approach has certain limitations and that the test is not informative for some of the more common RC, particularly those caused by large deletions/duplications or variants located outside the analyzed coding and exon–intron boundary regions.

For at-risk couples, options for further management and reproductive choices were discussed.

In addition to reproductive risk (RR), participants were also counseled regarding findings with potential implications for their personal health risk, when such findings were identified. In these cases, the clinical significance of the findings was discussed and, when appropriate, recommendations were provided for further evaluation, follow-up, or referral to an appropriate clinical specialist.

2.5 Statistical Analysis – Statistical analysis of the obtained data was performed using the following methods:

- Frequency analysis of variables, including absolute frequencies expressed as percentages.
- Assessment of statistical association between two variables using the chi-square test for linear trend (Cochran–Armitage test).

For frequency data, 95% confidence intervals (95% CI) were calculated. A p-value was determined to assess statistical significance, with values of $p < 0.05$ considered statistically significant. Statistical analyses were performed using SPSS for Windows, version 29.0.2.0.

3. Methodological Limitations of the Applied Analytical Approach – The present dissertation is limited by the applied analytical approach for ECS, based solely on NGS using a clinical exome targeted panel. The absence of complementary methods for copy number analysis, such as multiplex ligation-dependent probe amplification (MLPA), does not allow reliable detection of large deletions and duplications, resulting in reduced sensitivity for identifying carrier status in certain clinically and epidemiologically significant recessive and X-linked monogenic disorders, including spinal muscular atrophy and Duchenne/Becker muscular dystrophy.

In addition, the limitations also arise from the fact that the applied panel does not cover regulatory and deep intronic regions, which further reduces the sensitivity of the method for detecting pathogenic variants characteristic of β-thalassemia—a condition with a relatively high carrier frequency in the Bulgarian population. As a result, the true carrier frequency in the studied cohort may be underestimated.

IV. RESULTS AND DISCUSSION

1. Quantitative Assessment of Carrier Burden – Frequency and Average Number of Carrier States for Pathogenic Variants in Recessive Monogenic Conditions in the Studied Bulgarian Cohort

The proportion of carriers of pathogenic/likely pathogenic variants (PV)* in genes associated with ARC and XRC identified in our study was 95.7% (n = 287/300) (Figure 1). Accordingly, the proportion of individuals in whom no carrier burden for RC was detected was 4.3% (n=13/300).

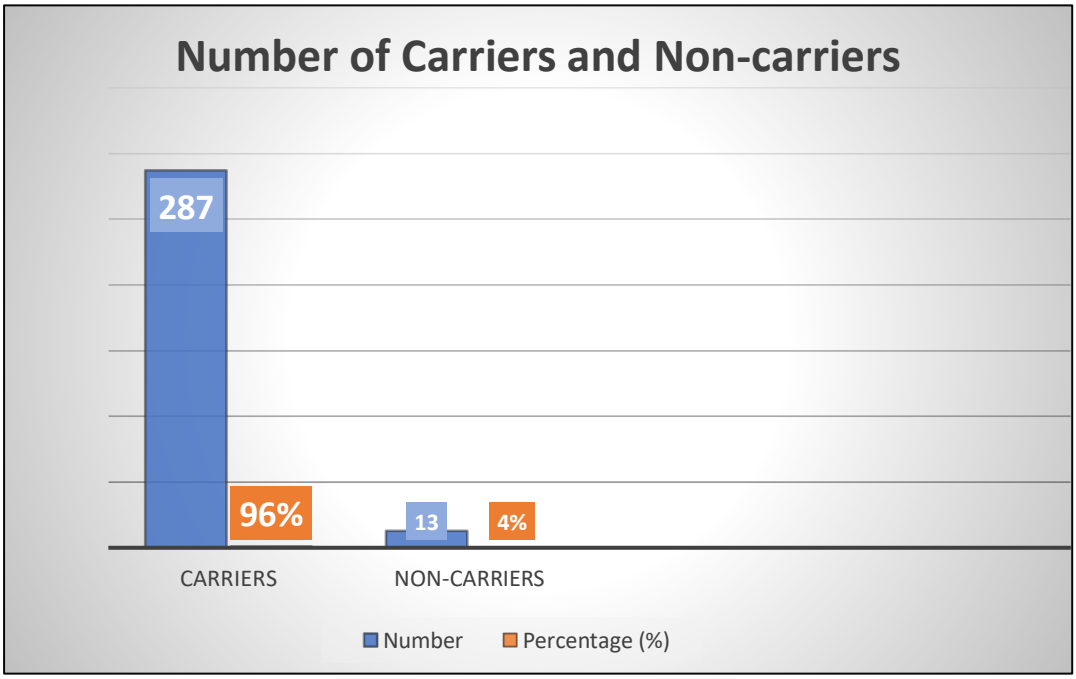

Figure 1. Distribution of the studied individuals according to the presence or absence of carrier status (absolute numbers and percentages).

Thus, within the studied cohort, 96 out of 100 individuals were carriers of at least one PV in a gene associated with RC, while 4 out of 100 individuals carried no PV.

*In the present study, the terms "pathogenic variant" and "likely pathogenic variant" are collectively referred to as "pathogenic variant." Although a formal distinction exists according to ACMG/AMP classification criteria, their clinical relevance in the context of this analysis is considered equivalent. Therefore, for the purposes of this study, the term "pathogenic variant" is used interchangeably with the term "mutation."

The total number of identified carrier states in the study was 1,051. Their distribution per individual is presented in Figure 2 and is as follows: carrier status for a PV in one gene was identified in 11% of carriers (n=33/287); in two genes – 18% (n=55/287); in three genes – 19% (n=56/287); in four genes – 22% (n=63/287); in five genes – 11% (n=32/287); in six genes – 8% (n=22/287); in seven genes – 6% (n = 18/287); in eight genes – 2% (n= 6/287); in ten genes – 0.3% (n=1/287); and in twelve genes – 0.3% (n=1/287).

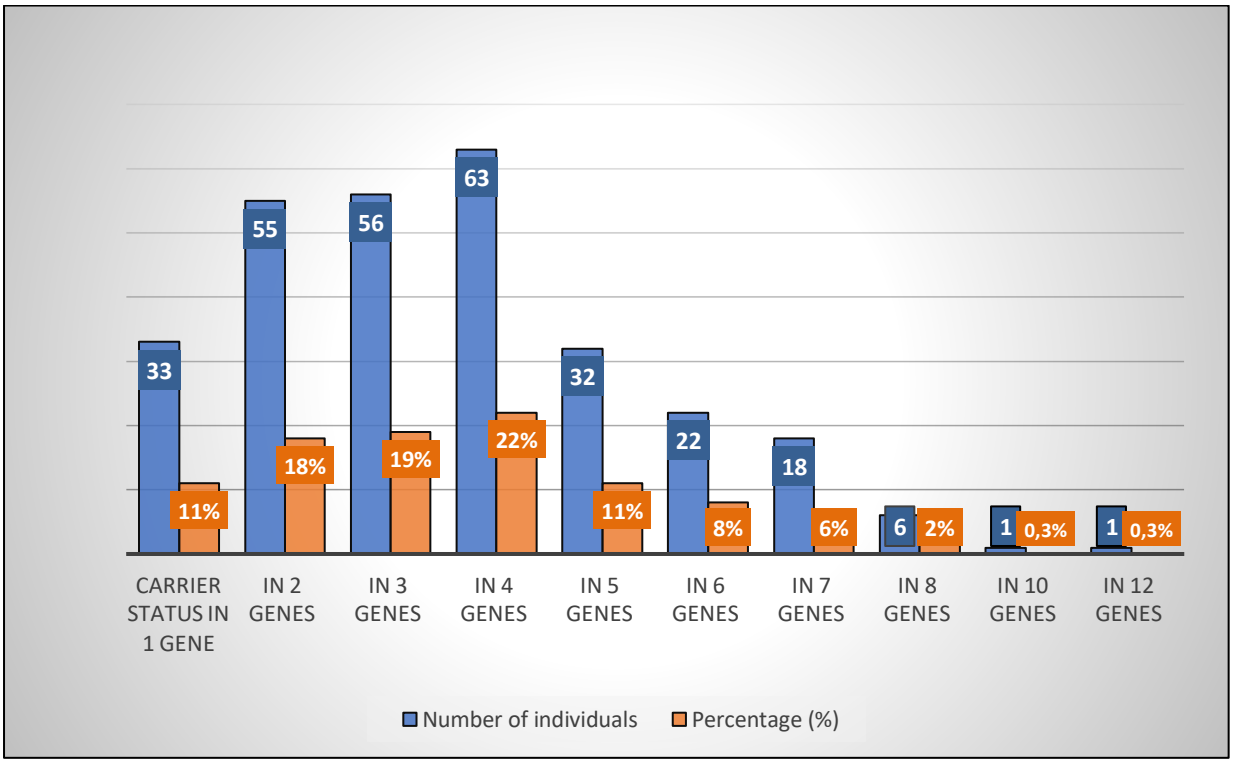

Figure 2. Distribution of carriers by number of pathogenic variants per individual

Based on the presented data, we found that the average number of carrier states for PV associated with RC per individual in the studied cohort was 3.5. The results of the study indicate that the mean carrier burden is 3.5 pathogenic or likely pathogenic variants per individual for RC.

It should be emphasized that direct numerical comparison with previously published data on carrier burden is methodologically limited due to substantial differences in analytical approaches, gene panel coverage, and variant inclusion criteria. In contrast to most published studies, which are based on predefined targeted panels, the present study employs an exome-based approach with broad gene coverage and without prior selection.

For decades, population genetics has established the concept that every clinically healthy individual is a carrier of several recessive PV, which manifest phenotypically only in the homozygous state. In this context, conceptual comparison with contemporary exome- and genome-based studies indicates that the carrier burden identified in the present study falls within the expected range for the human population. This finding further supports the applicability of ECS for precise characterization of individual carrier profiles.

2. General Characteristics of Genes with Identified Carrier Status and Detailed Analysis of the Most Frequently Affected Genes

2.1 General Characteristics of Affected Genes

The analysis identified carrier status for PV in 481 different genes, classified according to inheritance pattern as AR, mixed (AR/AD), and XLR. The distribution of carrier status across genes demonstrated a markedly uneven pattern (Figure 3).

For the majority of genes—62% (n=299/481)—carrier status was identified in a single individual. For 18% (n = 85/481) of genes, carrier status was observed in two individuals; for 7% (n=33/481), in three individuals; and for 6% (n=25/481), in four individuals. Higher carrier frequency (CF) was observed in a substantially smaller number of genes: in 2% (n=12/481), carrier status was identified in five individuals; in 2% (n = 8/481), in six individuals; and in 1% (n=3/481), in seven individuals.

At the upper end of the distribution, a limited number of genes exhibited notably higher carrier frequencies, including one gene (*C2*) identified in eight individuals; two genes (*CYP24A1* and *FUT8*) in nine individuals each; three genes (*CLCN1*, *SERPINA1*, and *G6PD*) in ten individuals each; one gene (*MPO*) in eleven individuals; and one gene (*BCHE*) in fifteen individuals. In addition, for two genes (*CFTR* and *NPHS2*), carrier status was identified in sixteen individuals each; for one gene (*GJB2*), in twenty individuals; for two genes (*BTD* and *CYP21A2*), in twenty-two individuals each; and for one gene (*MEFV*), in twenty-three individuals. The highest CF – 46 individuals in the studied cohort—was observed for a single gene (*ABCA4*).

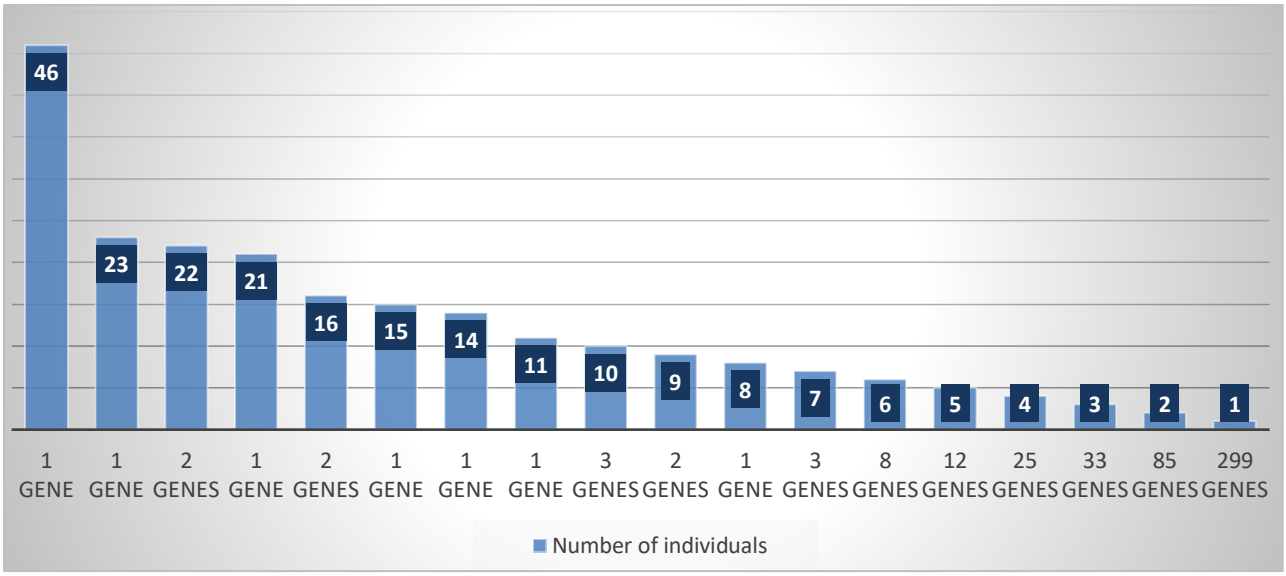

Figure 3. Distribution of affected genes (with carrier status) by number of individuals

The observed distribution pattern of carrier status across genes is consistent with the so-called "long-tail" distribution model. This concept reflects the presence of a limited number of more common conditions and genes with relatively higher carrier frequencies, contrasted by a large number of ultra-rare conditions represented by single or extremely rare findings in the population (Figure 4).

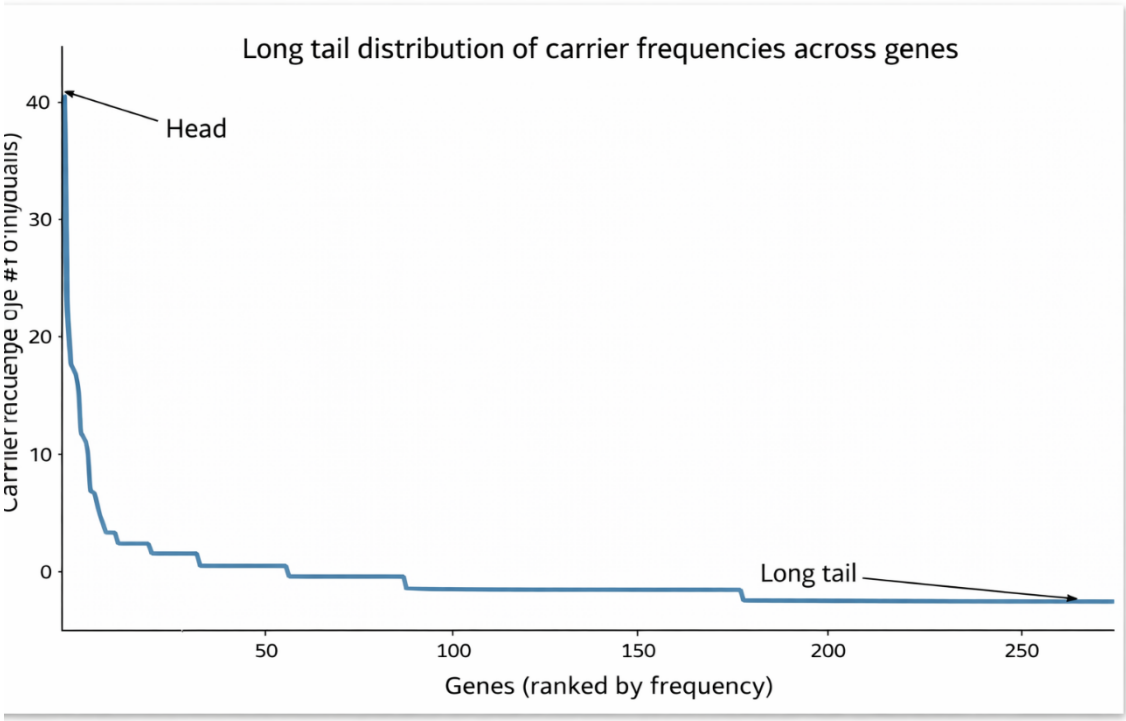

Figure 4. "Long-tail" model in the distribution of carrier states across genes

In the present study, of all 481 genes, 87% (n=419/481) were associated with AR inheritance, 12% (n=59/481) with mixed inheritance (AR and AD), and 1% (n=3/481) XLR inheritance (Figure 5).

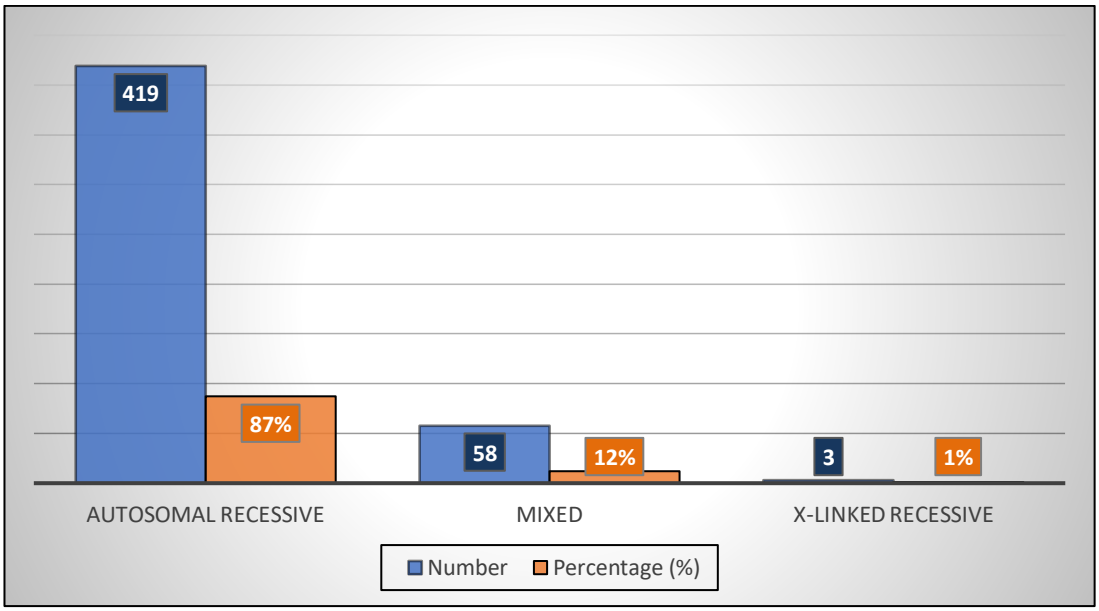

Figure 5. Distribution of detected conditions according to mode of inheritance

Adhering to ACMG standards and applying the empirical framework proposed by Lazarin et al., each condition for which carrier status was identified was classified into one of the following categories: profound, severe, moderate, or mild. The classification was based on clinical criteria, including age of onset, impact on survival, degree of physical and/or intellectual impairment, need for medical intervention, and overall impact on quality of life.

In cases where multiple phenotypes were associated with a single gene, classification of disease severity was based on the most severe clinically relevant phenotype observed in individuals with biallelic PV (homozygotes or compound heterozygotes). Among the conditions associated with carrier status identified in this study, 19% (n=92/481) were classified as profound, 40% (n=192/481) as severe, 35% (n=168/481) as moderate, and 6% (n=29/481) as mild (Figure 6).

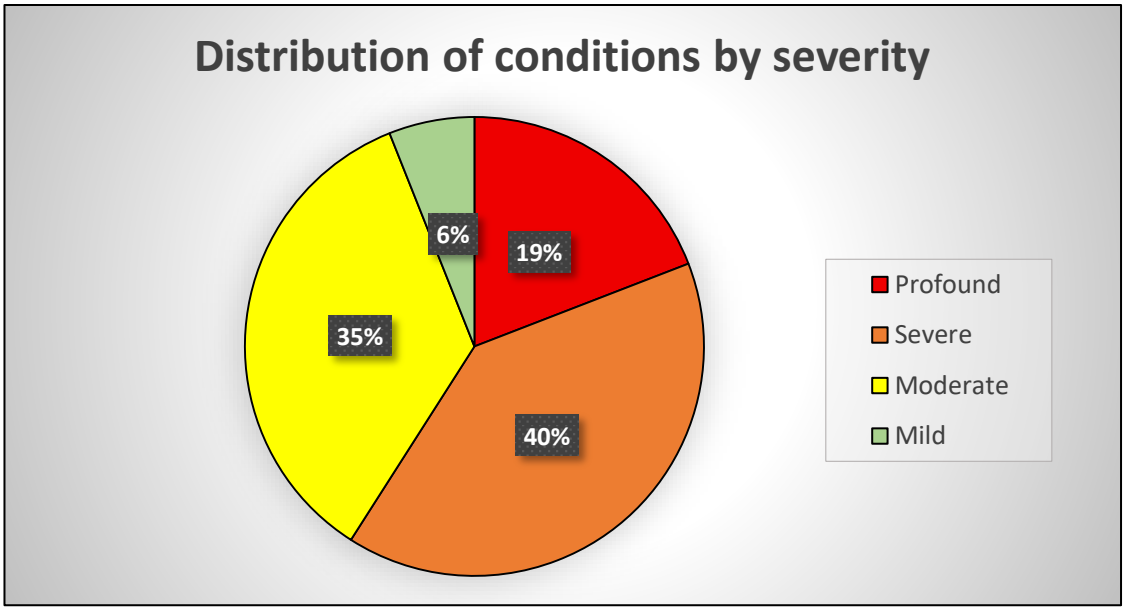

Figure 6. Proportions of conditions according to severity

2.2 Definition and Analysis of the Most Frequently Affected Genes

Genes were defined as most frequently affected if carrier status was identified in ≥5 individuals (≥5/300), corresponding to a minimum observed frequency of ≥1:60 in the studied cohort. The clinical conditions associated with such carrier status may be considered relatively common disorders.

This frequency is comparable to the carrier frequencies of some of the most prevalent ARC, such as spinal muscular atrophy (~1:50), congenital adrenal hyperplasia due to 21-hydroxylase deficiency (~1:60), and β-thalassemia (~1:50 in Bulgaria).

In the present study, of the total 481 affected genes in which carrier status was identified, 7.9% (n=38/481) met this definition (Table 2). This threshold was selected to increase the robustness of frequency estimates and to limit the influence of single observations in the analysis of a large number of genes.

Table 2. Profile of the most frequently affected genes

Gene	Phenotype	Mode of Inheritance	Severity	Total Number of Individuals	Carrier Frequency (%)	95% CI (%)
ABCA4	Stargardt disease 1	AR	Moderate	46	15,3	11,7–19,8
MEFV	Familial Mediterranean Fever	Mixed	Moderate	23	7,7	5,2–11,2
BTD	Biotinidase deficiency	AR	Moderate	22	7,3	4,9–10,9
CYP21A2	Congenital adrenal hyperplasia	AR	Severe	22	7,3	4,9–10,9
GJB2	Autosomal recessive nonsyndromic hearing loss 1A	AR	Moderate	21	7	4,6–10,5
NPHS2	Nephrotic syndrome, type 2	AR	Severe	16	5,3	3,3–8,5
CFTR	Cystic fibrosis	AR	Severe	16	5,3	3,3–8,5
BCHE	Butyrylcholinesterase deficiency	AR	Mild	15	5	3,1–8,1
MPO	Myeloperoxidase deficiency	AR	Mild	11	3,7	2,1–6,4
CLCN1	Congenital myotonia	Mixed	Moderate	10	3,3	1,8–6,0
G6PD	Glucose-6-phosphate dehydrogenase deficiency	XLR	Moderate	10	3,3	1,8–6,0
SERPINA1	Alpha-1 antitrypsin deficiency	AR	Moderate	10	3,3	1,8–6,0
CYP24A1	Infantile hypercalcemia type 1	AR	Moderate	9	3	1,6–5,6
FUT8	Congenital disorder of glycosylation with defective fucosylation type 1	AR	Profound	9	3	1,6–5,6
C2	Complement component 2 deficiency	AR	Moderate	8	2,7	1,4–5,2
RBM8A	Thrombocytopenia-absent radius syndrome	AR	Severe	7	2,3	1,1–4,7
DNAH11	Primary ciliary dyskinesia, type 7, with or without situs inversus	AR	Moderate	7	2,3	1,1–4,7
WNT10A	Ectodermal dysplasia 16	Mixed	Moderate	7	2,3	1,1–4,7

GBA1	Gaucher disease	AR	Severe	6	2	0,9–4,3
ABCC6	Pseudoxanthoma elasticum	Mixed	Moderate	6	2	0,9–4,3
C8B	Complement component 8 deficiency, type II	AR	Severe	6	2	0,9–4,3
DHCR7	Smith-Lemli-Opitz syndrome	AR	Profound	6	2	0,9–4,3
ECM1	Urbach-Wiethe disease	AR	Moderate	6	2	0,9–4,3
PAH	Phenylketonuria	AR	Severe	6	2	0,9–4,3
REL	Immunodeficiency 92	AR	Severe	6	2	0,9–4,3
USH2A	Usher syndrome, type 2A	AR	Moderate	6	2	0,9–4,3
VWF	von Willebrand disease	AR	Moderate	6	2	0,9–4,3
ACY1	Aminoacylase 1 deficiency	AR	Mild	6	2	0,9–4,3
CD36	Platelet glycoprotein IV deficiency	AR	Mild	5	1,7	0,7–3,8
GNRHR	Hypogonadotropic hypogonadism 7 without anosmia	AR	Mild	5	1,7	0,7–3,8
HFE	Hereditary hemochromatosis	AR	Moderate	5	1,7	0,7–3,8
MMP20	Amelogenesis imperfecta, type IIA2	AR	Mild	5	1,7	0,7–3,8
PEX6	Peroxisome biogenesis disorder 4A	Mixed	Profound	5	1,7	0,7–3,8
POLG	Mitochondrial DNA depletion syndrome 4B	Mixed	Profound	5	1,7	0,7–3,8
PROP1	Combined pituitary hormone deficiency 2	AR	Moderate	5	1,7	0,7–3,8
RAD50	Nijmegen breakage syndrome-like disorder	Mixed	Severe	5	1,7	0,7–3,8
TRPM6	Hypomagnesemia 1, intestinal	AR	Moderate	5	1,7	0,7–3,8
UGT1A1	Синдром на Gilbert	AR	Mild	5	1,7	0,7–3,8

Among the group of common conditions, 11% (n=4/38) were classified as profound, 24% (n=9/38) as severe, 47% (n=18/38) as moderate, and 18% (n=7/38) as mild (Figure 7). Notably, in contrast to the overall group of conditions associated with carrier status, in which severe conditions predominated (40%), the group of common conditions was characterized by a predominance of moderate conditions (47%).

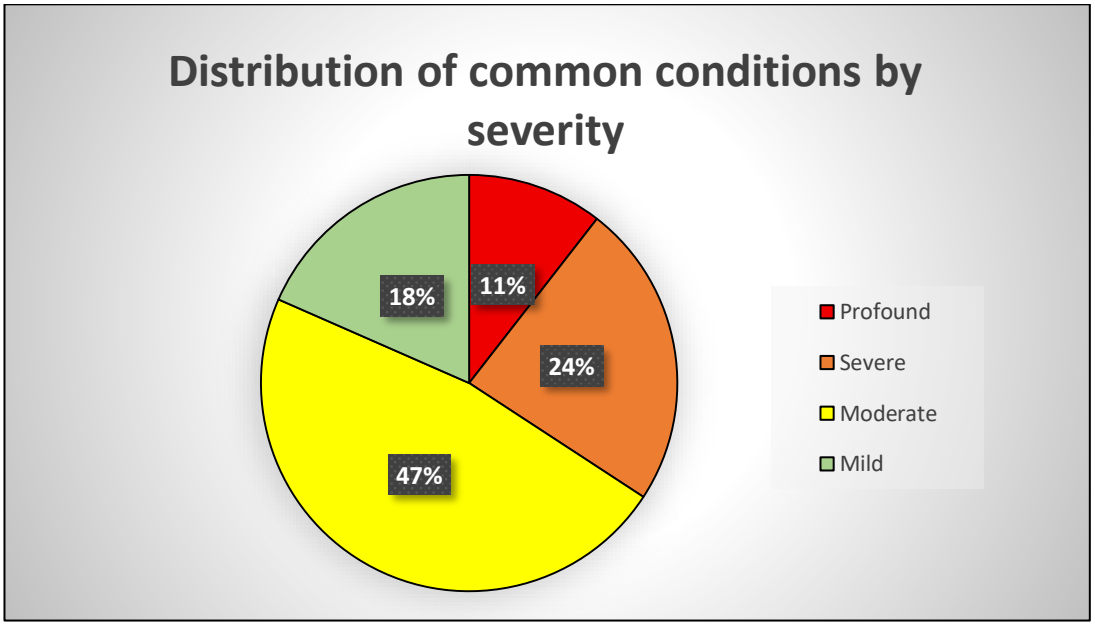

Figure 7. Distribution of common conditions in the studied cohort by severity

A comparative analysis of the association between CF and clinical severity was performed using the chi-square test for linear trend (Cochran–Armitage test). A statistically significant difference in severity distribution was observed between common and rare genes, as well as a significant negative association between CF and disease severity.

Table 3. Distribution of common and rare genes according to the severity of associated conditions

Group of conditions	**Common genes**	**Rare genes**
Mild + Moderate	65%	39%
Severe + Profound	35%	61%

Approximately two-thirds of the most frequently affected genes in the studied cohort were associated with moderate and mild conditions, whereas approximately two-thirds of rare genes were associated with profound and severe clinical phenotypes (Table 3). This observation supports the presence of negative selection against highly severe genetic conditions at the population level.

For conditions with well-defined reference population carrier frequencies, a formal statistical comparison with published data was performed using a one-sample Z-test for proportions. This test allows assessment of whether the observed CF in the studied Bulgarian cohort differs significantly from a predefined reference frequency, taking into account the sample size and expected statistical variability. The results of these statistical comparisons are summarized in Table 4.

Table 4. Comparison of carrier frequency in 12 of the most frequently affected genes in the studied cohort with reference frequencies

Gene/Phenotype	BG carriers (x/300)	BG frequency (%)	Ref. frequency (%)	Z	p-value	Conclusion
ABCA4/Stargardt disease 1	46/300	15,3	5,0	+8,21	<0,0001	BG > ref.
BTD/Biotinidase deficiency	22/300	7,3	0,83	+12,38	<0,0001	BG > ref.
CYP21A2/Congenital adrenal hyperplasia	22/300	7,3	1,82	+7,15	<0,0001	BG > ref.
NPHS2/ Nephrotic syndrome, type 2	16/300	5,3	7,69	−1,53	0,13	n.s.
CFTR/Cystic fibrosis	16/300	5,3	4,0	+1,18	0,24	n.s.
CLCN1/Congenital myotonia	10/300	3,3	0,89	+4,49	<0,0001	BG > ref.
SERPINA1/A1AT deficiency	10/300	3,3	10,0	−3,85	0,0001	BG < ref.
CYP24A1/Infantile hypercalcemia type 1	9/300	3,0	1,0	+3,48	0,0005	BG > ref.
DHCR7/Smith–Lemli–Opitz syndrome	6/300	2,0	1,85	+0,19	0,85	n.s.
PAH/Phenylketonuria	6/300	2,0	2,0	0,00	1,00	n.s.
USH2A/Usher syndrome, type 2A	6/300	2,0	1,43	+0,83	0,40	n.s.
GBA1/Gaucher disease	5/300	1,7	0,5	+2,87	0,004	BG > ref.

Among the conditions for which formal statistical comparison was possible, approximately 42% (n=5/12) showed concordance between the observed CF in the Bulgarian cohort and reference values, whereas 67% (n=8/12) demonstrated a statistically significant difference. In most cases of discrepancy, the CF was higher in the Bulgarian sample, suggesting possible population-specific enrichment of PV. This applied to carrier status in the following genes:

ABCA4, BTD, CYP21A2, CLCN1, CYP24A1, and *GBA1*. Only one gene – *SERPINA1* – showed a statistically significantly lower frequency compared to published data.

These findings highlight both the overall comparability with European data for a subset of conditions and the presence of population-specific deviations with potential clinical relevance.

3. Molecular Profiling of Pathogenic Variants in the Most Frequently Affected Genes

In the present study, among all analyzed individuals (n=300), a total of 690 distinct variants classified as pathogenic/likely pathogenic according to ACMG/AMP criteria were identified. Of these, 68% (n=470/690) were previously reported and documented in the medical literature. The remaining approximately one-third, 32% (n=220/690), represent novel variants that had not been reported in population or clinical databases at the time of analysis (Figure 8).

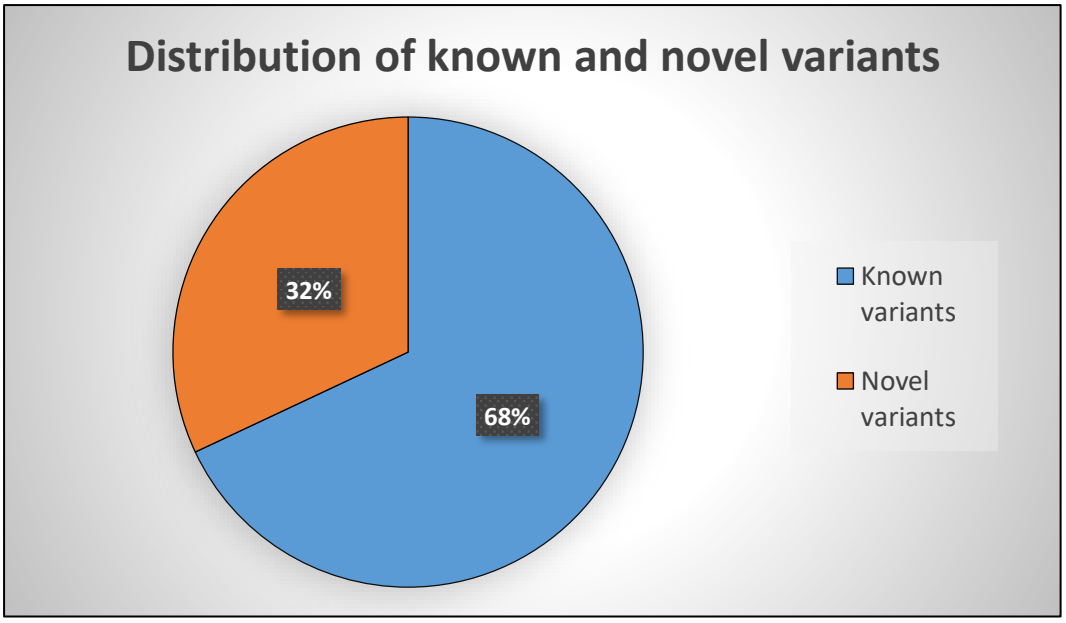

Figure 8. Distribution of known and novel pathogenic variants identified in the present study

NGS technologies are characterized by high detection capacity, enabling the identification of a broad spectrum of genetic variation. This includes not only rare variants but also novel, previously unreported variants, as demonstrated in large-scale exome and genome studies. In this context, the observed proportion of 32% novel PV in the present study can be interpreted as a logical consequence of the applied analytical approach – clinical exome sequencing without prior gene selection – as well as the limited representation of the Bulgarian population in

existing clinical and population reference databases.

As noted above, the most frequently affected genes were defined as those in which carrier status was identified in ≥5 individuals (≥5/300), corresponding to a minimum observed frequency of ≥1:60 in the studied cohort. A detailed analysis of the molecular characteristics of the identified pathogenic variants was performed for all such genes. This analysis was conducted in accordance with ACMG/AMP pathogenicity criteria, as presented in Table 1 in the Materials and Methods section.

Across the 38 genes meeting this definition, a total of 119 PV were identified and classified according to their clinical significance based on available literature data. The largest proportion of variants – 48% (n=57/119), was associated with severe clinical effects, identified in 172 individuals. Approximately 34% of variants (n=41/119) were associated with mild effects or milder phenotypes and were identified in 176 individuals. A considerably smaller proportion, 8% (n=9/119), corresponded to variants with intermediate effects, identified in 13 individuals. It should be noted that classification of intermediate effect required in-depth expert interpretation of available clinical and literature data and was applicable to a limited number of genes. For 10% of variants (n=12/119), the severity of the phenotypic effect could not be definitively determined; these variants were identified in 20 individuals.

These findings indicate that variants associated with severe clinical effects predominated in the studied group, followed by those associated with milder phenotypes (Figure 9). Nevertheless, the distribution of severe and mild variants among the most frequently affected genes was approximately comparable.

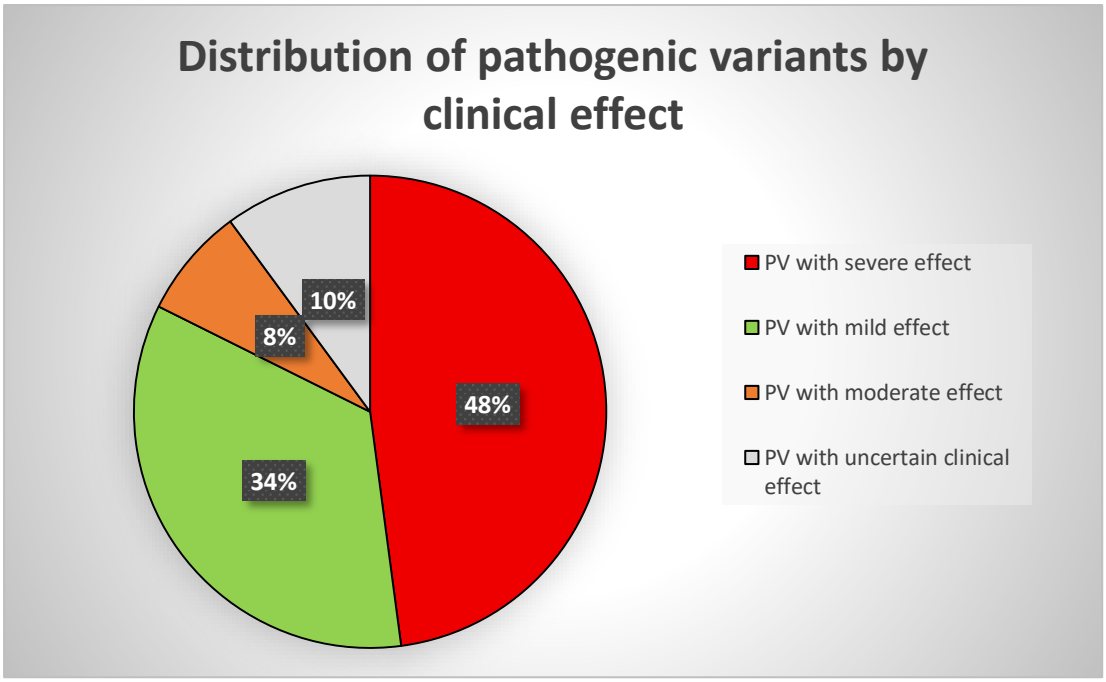

Figure 9. Distribution of pathogenic variants identified in the group of most frequently affected genes

Of the 119 PV identified in the 38 most frequently affected genes, 6% (n=7/119) were novel and had not been previously reported at the time of analysis (Table 5).

Table 5. Novel pathogenic variants identified in the group of most frequently affected genes

Gene/Phenotype	Variant	Transcript	Variant type
GJB2/Autosomal recessive nonsyndromic hearing loss 1A	c.252del (p.Ser85ProfsTer6)	NM_004004.6	Frameshift
FUT8/Congenital disorder of glycosylation, type If	c.286C>T (p.Gln96Ter)	NM_004480.4	Stop gained
ECM1/Urbach–Wiethe disease	c.135_136insG (p.Pro46AlafsTer16)	NM_004425.5	Frameshift
ECM1/Urbach–Wiethe disease	c.386-1G>A	NM_004425.5	Splice acceptor
CD36/Platelet glycoprotein IV deficiency	c.535_536del (p.Leu179MetfsTer4)	NM_001001548.3	Frameshift
CD36/Platelet glycoprotein IV deficiency	c.962dup (p.Asn321LysfsTer32)	NM_001001548.3	Frameshift
TRPM6/Hypomagnesemia 1, intestinal	c.955del (p.Thr319GlnfsTer25)	NM_017662.6	Frameshift

The identified variants were classified as likely pathogenic according to ACMG/AMP criteria, based on similar lines of evidence, namely their rarity in the population, as reflected by their absence from population databases such as gnomAD (PM2), as well as concordance of their molecular characteristics (loss of function, LOF) with the established pathogenic mechanism of the respective disease (PVS1).

4. Assessment of Couples with Potential Reproductive Risk and Identification of At-Risk Couples

4.1 Definition of At-Risk Couples in the Present Study

An at-risk couple is defined as a reproductive couple with an increased RR for a genetic disorder in the offspring, including ARC when both partners are carriers of PV in the same gene, and XRC when the female partner is a carrier of a PV.

In standard screening models, panel content is typically limited to genetic disorders with severe clinical presentation. In contrast, in the present study, RR—and accordingly at-risk couples – is defined more broadly, as the clinical exome approach includes conditions with mild and moderate phenotypes, which are subsequently subject to detailed molecular and clinical interpretation in terms of their clinical significance and the magnitude of the estimated RR.

A total of 11.3% (n=17/150) of the studied couples were considered potentially at risk. This group included couples with shared carrier status for PV in the same gene associated with ARC, as well as couples in which the female partner was a carrier of a PV in a gene associated with XRC.

4.2 Reproductive Couple No. 1

Table 6. Reproductive risk–associated genetic variants identified in Couple No. 1

Gene/Phenotype	Partner	Variant	Type	Classification
CYP21A2/ Congenital adrenal hyperplasia	Female	c.1174G>A (p.Ala392Thr)	Missense	Likely pathogenic
	Male	c.1174G>A (p.Ala392Thr)	Missense	Likely pathogenic

The variant identified in both partners of reproductive couple No. 1 was classified as likely pathogenic according to ACMG/AMP criteria. *CYP21A2* is a gene in which missense variants represent an established disease mechanism, and the rate of benign missense variation is low

(PP2). The identified variant is rare in the general population and below the thresholds for ARC (PM2). Functional studies demonstrating reduced 21-hydroxylase activity support a deleterious effect of the variant (PS3). In the literature and databases, the p.Ala392Thr variant has been reported in affected individuals in trans with another PV, consistent with a recessive mode of inheritance (PM3).

It should be emphasized that the *CYP21A2* gene and the variants identified within it are challenging to analyze due to the presence of a highly homologous pseudogene, *CYP21A1P*. Therefore, the interpretation remains conditional and requires clear documentation that the functional data refer to a variant unambiguously localized within the active *CYP21A2* gene (i.e., excluding pseudogene co-amplification or gene conversion). To refine the risk assessment for this specific couple, it is necessary to verify whether the variant identified in each partner originates from the functional *CYP21A2* gene or from the pseudogene.

In summary, the risk for the offspring of the studied couple to be affected by an ARC associated with *CYP21A2* can be considered increased (25%) only if the c.1174G>A (p.Ala392Thr) variant identified in both partners is confirmed to be localized in the functional *CYP21A2* gene and not derived from the pseudogene *CYP21A1P*.

4.3 Reproductive Couple No. 2

Table 7. Reproductive risk–associated genetic variants identified in Couple No. 2

Gene/Phenotype	Partner	Variant	Type	Classification
NEB/ Nemaline myopathy	Female	c.24094C>T (p.Arg8032Ter)	Stop gained	Pathogenic
	Male	c.25441C>T (p.Arg8481Ter)	Stop gained	Pathogenic

The variant c.24094C>T (p.Arg8032Ter), identified in the female partner of the couple, is classified as pathogenic according to ACMG/AMP criteria. It is a nonsense variant leading to a premature stop codon and an expected LOF of the protein, which is an established pathogenic mechanism for the *NEB* gene (PVS1). It is rare or absent in the general population, consistent with an AR mode of inheritance (PM2). In addition, the variant type is in agreement with the well-described spectrum of pathogenic truncating variants in *NEB* (PP2).

The variant c.25441C>T (p.Arg8481Ter), identified in the male partner, is analogous to the previously described variant in terms of variant type (nonsense), expected molecular effect

(LOF due to a premature stop codon), and fulfillment of the same ACMG/AMP pathogenicity criteria; therefore, it is also classified as pathogenic.

Based on these findings, there is a 25% risk in each pregnancy that both pathogenic variants will be co-inherited by the offspring in a compound heterozygous state. The combination of two nonsense alleles in the *NEB* gene results in biallelic LOF, which represents the established molecular mechanism underlying NEB-related nemaline myopathy. Given the molecular characteristics of the variants, the expected phenotype is of moderate to severe clinical severity. Moderate and severe forms are characterized by early childhood muscle weakness, hypotonia, delayed motor development, and frequently respiratory involvement, typically with preserved intellectual function.

4.4 Reproductive Couple No. 3

Table 8. Reproductive risk–associated genetic variants identified in Couple No. 3

Gene/Phenotype	Partner	Variant	Type	Classification
ABCA4/ Stargardt disease 1	Female	c.514G>A p.(Gly172Ser)	Missense	Likely pathogenic
	Male	c.5603A>T p.(Asn1868Ile)	Missense	Pathogenic

The variant c.514G>A (p.Gly172Ser) in the *ABCA4* gene, identified in the female partner of reproductive couple No. 3, is classified as likely pathogenic due to its effect on an evolutionarily conserved amino acid within a functionally important region of the protein, supported by deleterious in silico predictions (PP3), low population frequency in large reference databases (PM2), and reports in affected individuals in combination with other pathogenic variants (PP4).

The variant c.5603A>T (p.Asn1868Ile), identified in the male partner, is not a classical PV but is classified as a hypomorphic PV with incomplete penetrance (<5%) and variable clinical expression, typically manifesting in combination with severe PV in *ABCA4*. It is associated with later onset and a milder form of ABCA4-associated retinopathy. The penetrance of the p.Asn1868Ile variant, when present in trans with a severe PV, is estimated to be approximately 5% or lower at the population level. However, higher penetrance estimates, reaching up to approximately 65%, have been reported in family-based studies involving affected individuals. In the present case, no family history or clinical manifestations of ABCA4-associated disease are reported. Therefore, population-based penetrance estimates should be applied in the risk

assessment for the offspring.

Given a 25% probability of inheritance of both variants in a compound heterozygous state, and assuming a penetrance of <5% for p.Asn1868Ile in combination with a likely pathogenic variant in *ABCA4*, the actual risk of developing a clinical phenotype in the offspring is estimated to be approximately 1%.

Considering the low estimated magnitude of phenotypic risk for the offspring, the reproductive couple was excluded from the group of at-risk couples.

* Hypomorphic variant – a genetic variant that results in partial reduction of gene or protein function. For some hypomorphic variants, clinical manifestation is observed primarily when present in trans with a classical pathogenic allele.

** Incomplete penetrance – a condition in which not all carriers of a given variant express the associated clinical phenotype. In the context of ARC, incomplete penetrance means that not all individuals with biallelic variants in a given gene develop a clinical phenotype, with expression depending on the residual function of the variants and the genetic context.

4.5 Reproductive Couple No. 4

Table 9. Reproductive risk–associated genetic variants identified in Couple No. 4

Gene/Phenotype	Partner	Variant	Type	Classification
MEFV/ Familial Mediterranean Fever	Female	c.2230G>T (p.Ala744Ser)	Missense	Likely pathogenic
	Male	c.2084A>G (p.Lys695Arg)	Missense	Context-dependent pathogenicity

The variant c.2230G>T (p.Ala744Ser) in the *MEFV* gene, identified in the female partner, is classified as likely pathogenic according to ACMG/AMP criteria. It has been repeatedly reported in patients with familial Mediterranean fever, including in homozygous and compound heterozygous states (PS4). The variant affects a functionally important domain of pyrin, in which pathogenic missense variants are enriched (PM1). It has a low population frequency, inconsistent with a benign polymorphism (PM2), and in silico analyses support a deleterious functional effect (PP3). It has also been reported in patients with a phenotype associated with *MEFV* (PP4), although clinical expression is characterized by variable expressivity and incomplete penetrance.

The variant c.2084A>G (p.Lys695Arg) is a missense variant with conflicting interpretation in clinical databases. It is most commonly classified as a VUS or as likely pathogenic with incomplete penetrance. It has been reported both in patients with mild or atypical forms of familial Mediterranean fever and in clinically unaffected carriers, including in heterozygous and compound heterozygous states. The variant is associated with variable expressivity, low predictive clinical significance, and lack of association with a severe phenotype; therefore, it rarely results in clinically significant disease on its own and is generally interpreted with caution in the context of the second allele and the clinical presentation.

There is a 25% risk of a compound heterozygous genotype in the offspring of this couple, comprising both variants. Given the characteristics of the identified variants, a mild to moderate clinical phenotype with good therapeutic responsiveness is expected.

4.6 Reproductive Couple No. 5

Table 10. Reproductive risk–associated genetic variants identified in Couple No. 5

Gene/Phenotype	Partner	Variant	Type	Classification
BTD/ Biotinidase deficiency	Female	c.1336G>C (p.Asp446His)	Missense	Pathogenic
	Male	c.1336G>C (p.Asp446His)	Missense	Pathogenic

The variant c.1336G>C (p.Asp446His) in the *BTD* gene was identified in both partners of the couple. It is classified as pathogenic, as it has been repeatedly reported in patients with biotinidase deficiency, including in homozygous and compound heterozygous states (PS4), is associated with reduced biotinidase activity demonstrated by functional/enzymatic studies (PS3), affects a functionally critical region of the protein (PM1), and is consistent with the established disease mechanism—loss of enzymatic function (PP4).

A 25% risk of a homozygous genotype in the offspring is therefore considered. However, available clinical and functional data indicate that p.Asp446His is a hypomorphic variant associated with partial biotinidase deficiency, which in the homozygous state is typically asymptomatic or associated with minimal clinical manifestations. The expected phenotype does not result in clinically significant health impairment and is fully preventable and treatable with timely biotin supplementation.

Accordingly, despite the presence of a theoretical genetic risk, Couple No. 5 was not classified as an at-risk couple.

4.7 Reproductive Couple No. 6

Table 11. Reproductive risk–associated genetic variants identified in Couple No. 5

Gene/Phenotype	Partner	Variant	Type	Classification
SERPINA1/ Alpha-1 antitrypsin deficiency	Female	c.1177C>T (p.Pro393Ser)	Missense	Pathogenic
	Male	c.1096G>A (p.Glu366Lys)	Missense	Pathogenic

The variant c.1177C>T (p.Pro393Ser, M Würzburg) in the *SERPINA1* gene, identified in the female partner, is classified as pathogenic. Functional data demonstrate impaired stability and secretion of α1-antitrypsin, resulting in reduced serum levels consistent with partial α1-

antitrypsin deficiency (PS3). The variant has been repeatedly reported in individuals with biochemically confirmed α1-antitrypsin deficiency and is enriched in affected individuals compared to the general population (PS4). It is reported at low frequency in the general population (PM2) and is associated with a SERPINA1-specific phenotype characterized by reduced serum α1-antitrypsin levels, typically with mild to moderate clinical significance (PP4). Limited familial data supporting segregation with disease are also available (PP1).

The variant c.1096G>A (p.Glu366Lys, Pi*Z), identified in the male partner, is classified as pathogenic. It is associated with markedly reduced serum α1-antitrypsin levels due to impaired protein folding and intracellular retention (PS3). It has been extensively reported in individuals with clinically and biochemically confirmed α1-antitrypsin deficiency (PS4) and is associated with a SERPINA1-specific phenotype characterized by moderately reduced serum α1-antitrypsin levels, with a variable but generally low to moderate risk of pulmonary involvement and a low risk of liver disease (PP4).

A 25% risk of transmission of both variants in a compound heterozygous genotype in the offspring is therefore considered. In such a genotype, intermediate to low serum α1-antitrypsin levels are expected, typically lower than those observed in PiMZ and within a range similar to PiSZ, which is associated with a moderate to increased risk of pulmonary disease (COPD/emphysema). This risk is strongly influenced by environmental factors, particularly smoking and other inhalational exposures. The risk of liver involvement in this genotype is considered low to moderate—substantially lower than in the classical Pi*ZZ genotype—but cannot be excluded, given the known secretory defect of the affected alleles and the potential for cumulative cellular stress in hepatocytes.

Incomplete penetrance and variable clinical expression are observed; therefore, the actual clinical phenotype cannot be predicted solely on the basis of genotype. Clinical assessment and follow-up should be based on both serum α1-antitrypsin levels and the individual’s clinical status and exposure history.

4.8 Reproductive Couple No. 7 and 8

Table 12. Reproductive risk–associated genetic variants identified in Couple No. 7 and 8

Gene/Phenotype	Partner	Variant	Type	Classification
NPHS2/ Nephrotic syndrome, type 2	Female	c.686G>A (p.Arg229Gln)	Missense	Likely pathogenic
	Male	c.686G>A (p.Arg229Gln)	Missense	Likely pathogenic

An analogous scenario was observed in two of the studied couples, in which each partner was a carrier of the c.686G>A (p.Arg229Gln) variant in the *NPHS2* gene. This variant is hypomorphic and exhibits well-defined context-dependent pathogenicity, which manifests in compound heterozygosity when present in trans with a severe PV in *NPHS2*. Available functional data demonstrate partially impaired podocin function, with reduced interaction with nephrin, without complete loss of protein activity (PS3). Clinical association with disease is observed predominantly in compound heterozygous states with another PV in *NPHS2*, in which case PM3 is applicable, along with supporting evidence for segregation and enrichment in affected individuals (PP1, PS4).

Existing clinical and population data clearly indicate that homozygous carrier state of p.Arg229Gln does not result in a disease phenotype. Therefore, a reproductive couple in which both partners are heterozygous carriers of this variant is not considered to be at increased risk of having affected offspring. On this basis, reproductive couples No. 7 and 8 were excluded from the group of at-risk couples.

4.9 Reproductive Couple No. 9

Table 13. Reproductive risk–associated genetic variants identified in Couple No. 7 and 8

Gene/Phenotype	Partner	Variant	Type	Classification
MARS/ Interstitial lung and liver disease	Female	c.2114dup (p.Leu705PhefsTer19)	Frameshift	Likely pathogenic
	Male	c.2114dup (p.Leu705PhefsTer19)	Frameshift	Likely pathogenic

At the time of analysis, the identified variant c.2114dup (p.Leu705PhefsTer19) was novel and not reported in any database. It represents a duplication leading to a frameshift and the creation of a premature stop codon. The expected effect is an absent or truncated protein product, i.e.,

LOF (PVS1). LOF is an established disease mechanism for MARS, associated with AR MARS-related encephalopathies and interstitial lung disease. The variant is absent from population databases (PM2). Based on these data, it is classified as likely pathogenic according to ACMG/AMP criteria.

In the case of a homozygous genotype in the offspring of the couple (with a 25% risk), p.Leu705PhefsTer19 is expected to result in LOF of the *MARS* gene. Given the established LOF disease mechanism for *MARS*, clinical manifestation is highly likely. However, due to the lack of published data for this specific variant and the known phenotypic variability, the severity and clinical spectrum of the condition cannot be reliably predicted.

4.10 Reproductive Couple No. 10,11,12,13,14 and 15

Table 14. Reproductive risk–associated genetic variants identified in Couple No. 10,11,12,13,14 and 15

Gene/Phenotype	Partner	Variant	Type	Classification
G6PD/ G6PD deficiency	Female	c.653C>T (p.Ser218Phe)	Missense	Pathogenic

An identical scenario was observed in six at-risk couples, involving heterozygous carrier state of the c.653C>T (p.Ser218Phe) variant in the *G6PD* gene in the female partner. Given the XLR mode of inheritance of glucose-6-phosphate dehydrogenase deficiency, there is a 50% risk for each male offspring in these couples.

The p.Ser218Phe variant is classified as pathogenic, as it has been repeatedly associated with clinically manifest G6PD deficiency (PS4), leads to significantly reduced enzymatic activity demonstrated by functional and biochemical studies (PS3), and shows a clear genotype–phenotype correlation with acute hemolytic anemia under oxidative stress, typically corresponding to WHO class II (occasionally II–III), which is characteristic of G6PD-related disease (PP4). In the medical literature, it is commonly referred to as the “Mediterranean” variant and is associated with severe deficiency of glucose-6-phosphate dehydrogenase activity (WHO class II).

Clinically, this variant is characterized by an increased susceptibility to and severity of hemolysis following exposure to oxidative stress (including infections, certain medications, and ingestion of fava beans). It manifests as acute hemolytic anemia, hemoglobinuria, and may

require hospitalization, with a higher risk of severe episodes compared to milder *G6PD* variants.

4.11 Reproductive Couple No. 16

Table 15. Reproductive risk–associated genetic variants identified in Couple No. 16

Gene/Phenotype	Partner	Variant	Type	Classification
G6PD/ G6PD deficiency	Female	c.934G>C (p.Asp312His)	Missense	Pathogenic

In the present couple, there is again a 50% risk for each male offspring, due to carrier status of a pathogenic variant in the *G6PD* gene in the female partner. The identified variant c.934G>C (p.Asp312His) is classified as pathogenic. It has been repeatedly reported in patients with clinically significant G6PD deficiency and a characteristic phenotype (PS4), leads to demonstrably reduced enzymatic activity (PS3), and is associated with a specific phenotype—acute hemolysis under oxidative stress—corresponding to WHO class III (PP4).

The difference between this and the previously described G6PD-related cases lies in the degree of clinical severity. The p.Asp312His variant is associated with a mild phenotype, typically manifesting as hemolysis only under significant oxidative stress, and is linked to a minimal risk of severe or chronic complications.

4.12 Reproductive Couple No. 17

Table 16. Reproductive risk–associated genetic variants identified in Couple No. 17

Gene/Phenotype	Partner	Variant	Type	Classification
DMD/ Duchenne/Becker muscular dystrophy	Female	c.960+2T>C	Splice donor	Likely pathogenic

In Couple No. 17, a novel variant not previously reported at the time of analysis was identified in the *DMD* gene in the female partner. The c.960+2T>C variant affects the canonical donor splice site (+2 position) of an intron in the *DMD* gene and is highly likely to result in aberrant splicing (e.g., exon skipping, intron retention, or activation of a cryptic splice site). As a secondary consequence, disruption of the reading frame is expected, leading to the generation of a premature stop codon. The predicted outcome is loss LOF of the dystrophin protein, either

due to nonsense-mediated mRNA decay or, less likely, production of a truncated, non-functional protein. LOF represents an established and predominant disease mechanism for the *DMD* gene, underlying Duchenne/Becker muscular dystrophy (PVS1). The variant is novel and absent from population databases (PM2). Based on these data, c.960+2T>C is classified as likely pathogenic.

The condition follows an XLR mode of inheritance, resulting in a 50% risk for each male offspring. Given the high likelihood of a frameshift and consequent loss of dystrophin function, c.960+2T>C is expected to be associated with a Duchenne muscular dystrophy phenotype. However, due to the possibility of alternative splicing mechanisms, a Becker phenotype cannot be entirely excluded.

4.13 Characteristics of the Group of At-Risk Reproductive Couples

Following the analysis of the 17 reproductive couples, we determined that classification as at-risk was not justified for 4 of them. According to the criteria applied in the present study, the proportion of truly at-risk couples was 9% (n=13/150) (Figure 9).

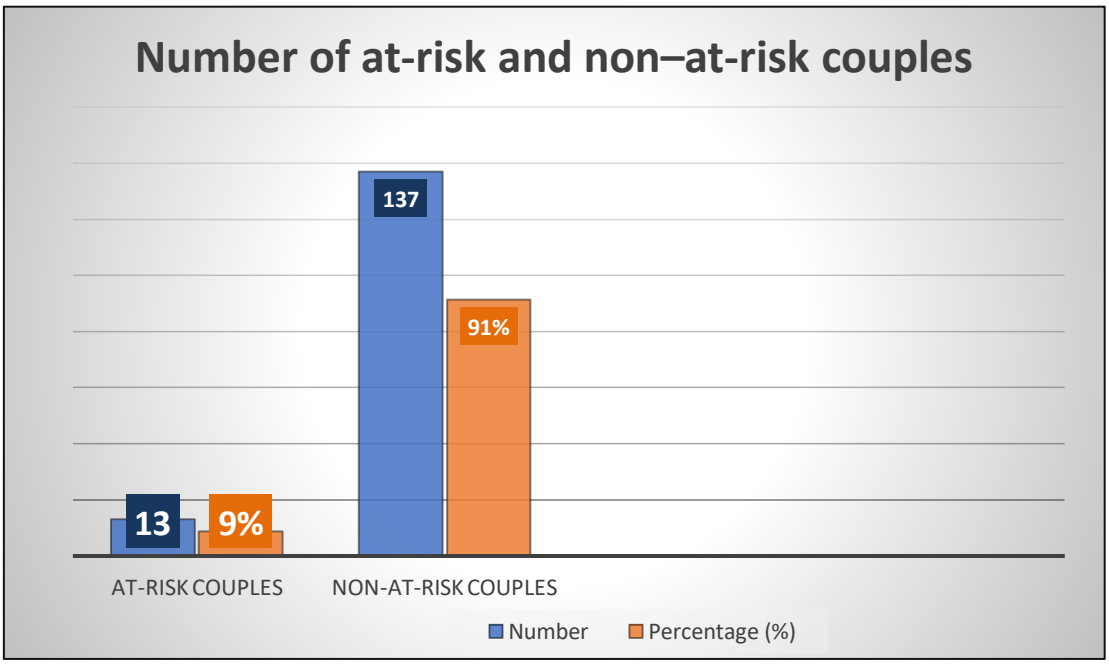

Figure 9. Distribution of couples according to the presence or absence of reproductive risk (number and percentage)

A total of eight genes associated with RC were identified among at-risk couples as contributing to an increased risk for the offspring. Of these, four were associated with ARC, two with XLR, and two with mixed inheritance (AR and AD) (Figure 10).

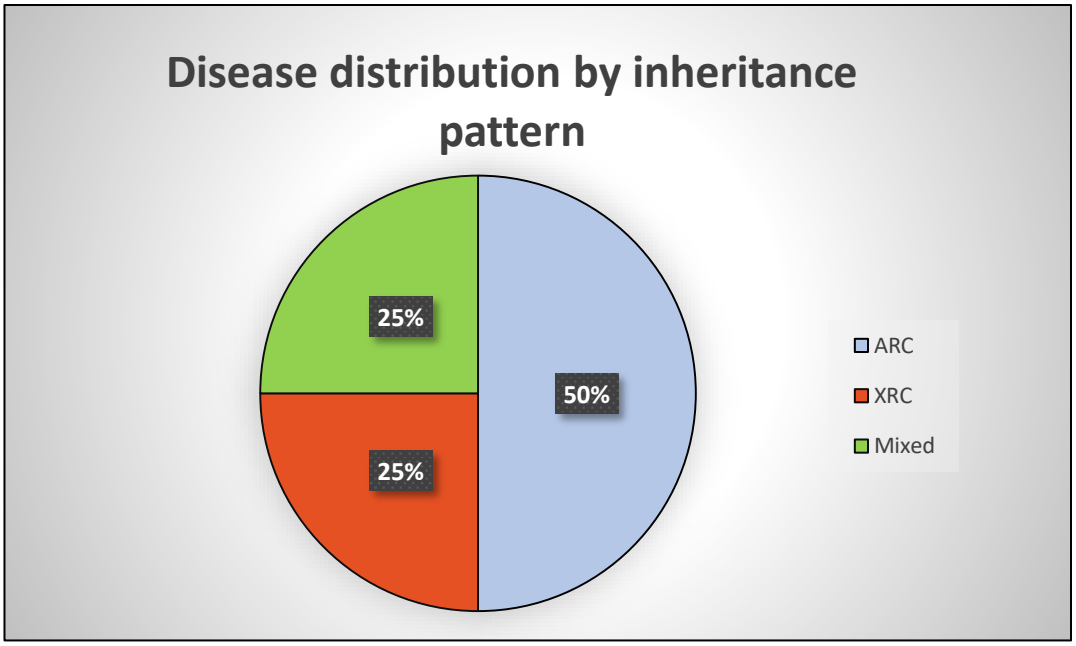

Figure 10. Distribution of conditions in at-risk couples by mode of inheritance

In approximately two-thirds, or 62%, of the at-risk couples (n=8/13), an increased risk for XRC was identified (Table 17).

Table 17. Conditions in at-risk couples with X-linked recessive inheritance

Gene	Condition	Severity	Number of at-risk couples
G6PD	G6PD deficiency	Moderate	7
DMD	Duchenne/Becker muscular dystrophy	Severe	1

In 31% of the at-risk couples (n=4/13), the condition associated with increased risk was AR in its inheritance (Table 18).

Table 18. Conditions in at-risk couples with autosomal recessive inheritance

Gene	Condition	Severity	Number of at-risk couples
CYP21A2	Congenital adrenal hyperplasia	Severe	1
NEB	Nemaline myopathy	Severe	1
ABCA4	Stargardt disease 1	Moderate	1
SERPINA1	Alpha-1 antitrypsin deficiency	Moderate	1

In 15% of the at-risk couples (n=2/13), an increased risk was identified for a condition with mixed inheritance (Table 19).

Table 19. Conditions in at-risk couples with mixed inheritance

Gene	Condition	Severity	Number of at-risk couples
MEFV	Familial Mediterranean Fever	Moderate	1
MARS	Interstitial lung and liver disease	Severe	1

Based on the presented data, it is evident that in the group of at-risk couples, the proportions of severe and moderately severe diseases are equal (Figure 11).

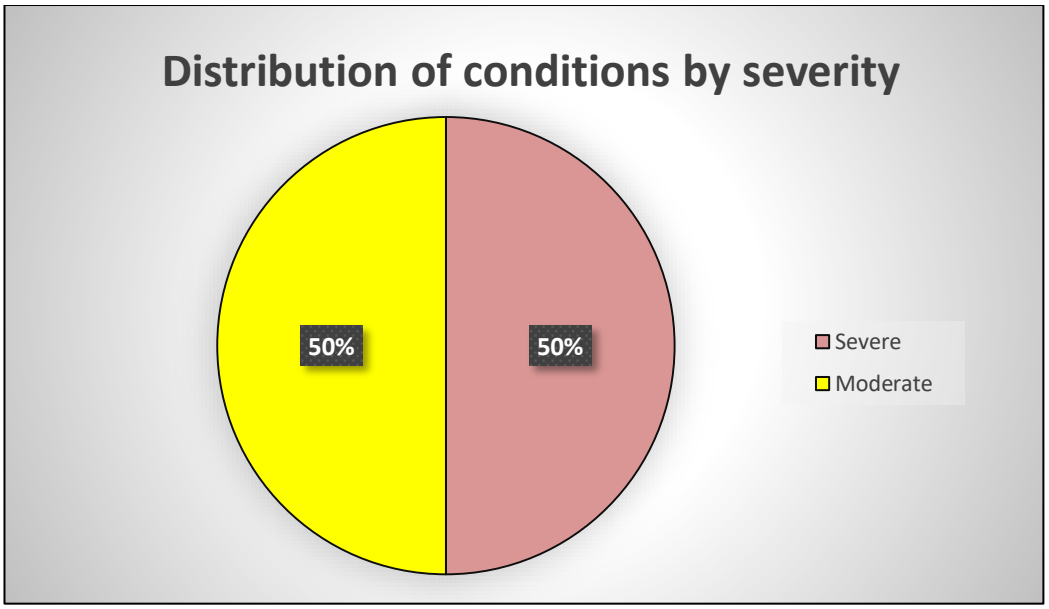

Figure 11. Distribution of conditions identified in the at-risk couples according to disease severity

When the analysis is restricted to severe and very severe conditions, as is the case in commonly used carrier screening panels, the proportion of at-risk couples in our cohort would decrease from 9% to 3%. This represents a threefold reduction in the estimated RR. As demonstrated, limiting the analysis exclusively to conditions with high clinical severity results in a substantially lower proportion of identified at-risk couples. In contrast, the exome-based, comprehensive approach enables a more complete assessment of RR and supports improved individualized genetic counseling and informed reproductive decision-making.

Of the total number of conditions identified among at-risk couples, 62% (n=5/8) belong to the group of the most frequent conditions in the studied cohort, while 38% (n=3/8), or approximately one-third, are not included in this group (Figure 12).

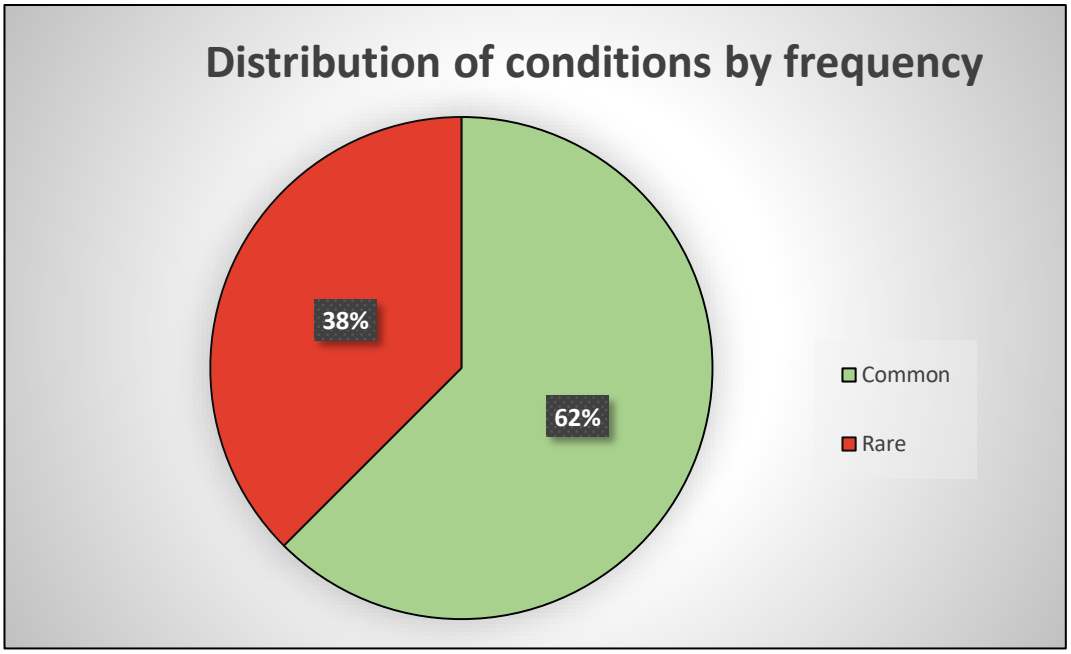

Figure 12. Distribution of conditions identified in the at-risk couples according to their frequency within the studied cohort

The finding that most conditions associated with at-risk couples are linked to the most frequently affected genes in the cohort is expected and reflects the disproportionate contribution of common carrier states to the combined RR. At the same time, the presence of at-risk couples associated with rarer conditions demonstrates that clinically significant RR is not limited to a restricted set of common disorders. This observation supports the utility of the exome-based approach in identifying risks beyond the scope of standard panels, particularly in the context of individualized GC.

5. Evaluation of Cases with Pathogenic Variants Relevant to the Personal Health Risk of Carriers

5.1 General Characteristics

In the present study, variants with potential relevance to personal health risk were identified in 13% (n=38/300) of the participants. These findings were categorized according to the nature and degree of individual risk as follows: confirmed personal health risk (13%; n=5/38); potential personal health risk requiring additional genetic testing (3%; n=1/38); conditional personal health risk, including genetic predisposition to oncological diseases (66%; n=25/38); and conditional risk for clinical manifestation of MD with mixed inheritance patterns (18%; n=7/38) (Figure 13).

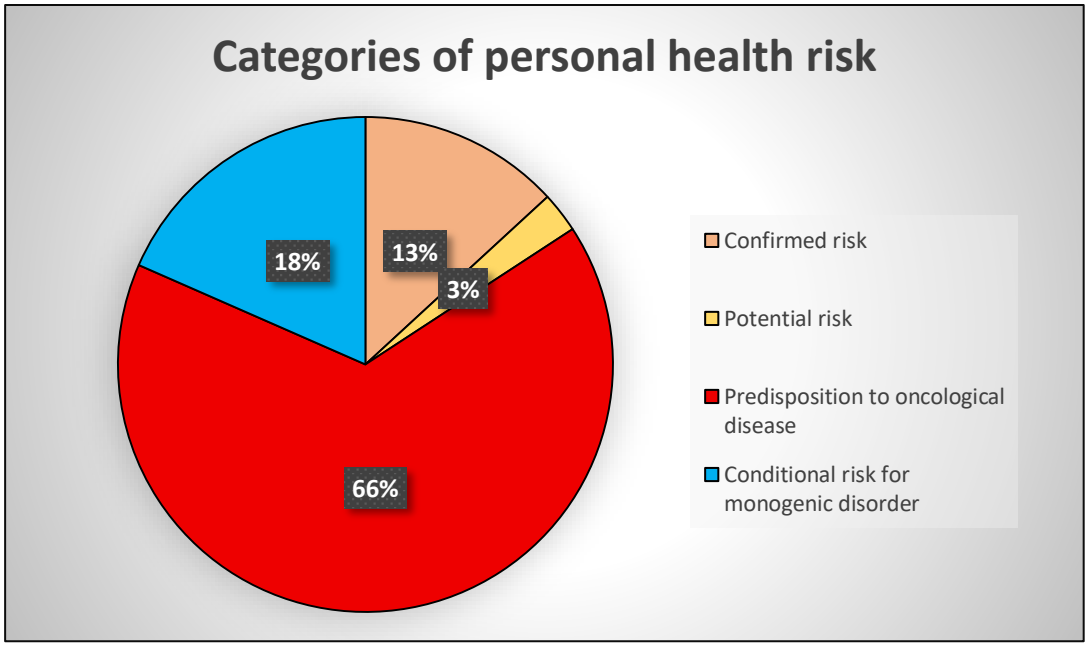

Figure 13. Distribution of different categories of personal health risk

5.2 Cases with Confirmed Personal Health Risk

The first category included four male individuals in whom the same PV in the *G6PD* gene – c.653C>T (p.Ser218Phe) – was identified. This gene is associated with glucose-6-phosphate dehydrogenase deficiency, a pharmacogenetic defect with an XLR mode of inheritance. This confers a real risk of clinical manifestations related to hemolytic episodes upon exposure to oxidative stress, including certain medications, infections, and specific foods. Recommendations for these individuals include avoidance of drugs with oxidative potential (e.g., sulfonamides, antimalarial agents, certain antibiotics, and analgesics), avoidance of fava beans, and increased clinical vigilance during infections. No direct risk is considered for male offspring of hemizygous men. However, it should be noted that all daughters will be heterozygous carriers, conferring a 50% risk of clinical manifestation in their sons (i.e., the male grandchildren of the affected individuals).

One additional case with confirmed personal health risk involved a male individual in whom a clearly pathogenic variant in the *ABCD1* gene—c.593C>T (p.Thr198Met)—was identified. This variant has been reported in individuals with clinical and/or biochemical manifestations of X-linked adrenoleukodystrophy (PS4), with evidence of segregation with disease in affected families (PP1). It is absent from population databases, including gnomAD (PM2). The amino acid residue p.Thr198 represents a clinically significant and functionally important region of the ABCD1 protein, with other PV affecting the same residue described in the literature (PM1).

In silico and structural-functional modeling further support a deleterious effect of the variant on protein function (PP3).

Based on these data, the p.Thr198Met variant can be classified as pathogenic. Given the XLR pattern of adrenoleukodystrophy, hemizygous males are considered to be at real risk for clinical manifestation. At the time of analysis, the individual was asymptomatic. However, the disease is characterized by variable age of onset, including the possibility of late manifestation, which precludes exclusion of risk and necessitates clinical follow-up.

Recommendations for the individual include regular clinical and biochemical monitoring due to the possibility of late disease onset, as well as consideration of cascade testing in maternal relatives. Due to the X-linked mode of inheritance, all daughters of the affected male will be obligate heterozygous carriers, which confers an indirect RR for clinical manifestation in their future sons.

5.3 Case with Potential Personal Health Risk

The second category included a single case of a female individual in whom two pathogenic variants in the *BCHE* gene were identified—c.1253G>T (p.Gly418Val) and c.293A>G (p.Asp98Gly). This gene is associated with the pharmacogenetic defect butyrylcholinesterase deficiency and follows an AR mode of inheritance. Given that clinical manifestations occur only under specific circumstances—most commonly upon exposure to ester-type muscle relaxants or other cholinesterase-metabolized drugs during general anesthesia—and in the absence of such clinical history, it cannot be determined whether the two variants are in cis or trans configuration. To clarify the clinical significance of these variants, molecular phasing is recommended to determine their allelic configuration.

5.4 Cases with Conditional Personal Health Risk

PV in genes with mixed inheritance patterns (AR and AD) were identified in 49% (n=147/300) of all analyzed individuals. Of the total 59 genes associated with both recessive and dominant phenotypes, 22% (n=13/59) are linked to cancer predisposition, while the remaining 78% (n=46/59) are associated with MD. These findings defined the following two categories of cases with variants relevant to personal health risk, presented in sections 5.4.1 and 5.4.2.

5.4.1 Cases with Predisposition to Oncological Disease

The group of cases with PV in cancer predisposition genes (identified in 13 genes) includes 25 individuals (Table 20).

Table 20. Pathogenic variants associated with cancer predisposition

Gene	Trancscript	Variant	Number of individuals
RAD50	NM_005732.4	c.326_329del (p.Thr109AsnfsTer20)	5
ERCC2	NM_000400.5	c.1703_1704del (p.Phe568TyrfsTer2)	3
	NM_000400.5	c.1418dup p.(His474ProfsTer27)	1
FANCL	NM_018062.5	c.2T>Cp.?	3
ATM	NM_000051.4	c.1564_1565del p.(Glu522IlefsTer43	1
	NM_000057.4	c.217C>T (p.Arg73Cys)	1
BLM	NM_020937.6	c.1642C>T (p.Gln548Ter)	2
FANCM	NM_006231.3	c.5048_5052del p.(Lys1683ArgfsTer3)	1
	NM_032043.4	c.2953del p.(Glu985ArgfsTer3)	1
POLE	NM_005213.4	c.6433C>T p.(Arg2145Ter)	1
	NM_000124.4	c.5484del p.(Ser1829ProfsTer16)	1
BRIP1	NM_001018113.3	c.2010dup (p.Glu671Ter)	1 (detected in a carrier of a PV in the *BLM* gene)
ERCC4	NM_178031.3	c.2395C>T (p.Arg799Trp)	1
ERCC6	NM_032444.4	c.1397+8147_1397+8150del	1
FANCD2	NM_000400.5	c.3707G>A (p.Arg1236His)	1
FANCE	NM_018062.5	c.1239dup (p.Pro414SerfsTer54)	1
SLX4	NM_000051.4	c.2808_2809del (p.Ala938ThrfsTer7)	1

Assessment of personal health risk was performed based on published data on penetrance and relative risk for the respective gene, as well as taking into account the variant type and mode of inheritance. The interpretation also considered individual factors such as age, family history, and available clinical data.

5.4.2 Conditional Risk for Manifestation of a Dominant Phenotype in Carriers of Pathogenic Variants in Genes with Mixed Inheritance

PV were identified in 121 individuals in a total of 46 genes associated with MD exhibiting both dominant and recessive phenotypes. Assessment of personal health risk was performed by integrating data on age of onset, disease severity, available literature for the specific variant, and family history. For novel variants, interpretation was based on their molecular characteristics and concordance with the pathogenic mechanism associated with the dominant phenotype.

In none of the cases was there evidence of familial occurrence in the context of the respective condition. In 60 individuals, no increased risk for clinical manifestation was identified due to age exceeding the typical period of phenotypic expression. In 4 cases, the dominant phenotype corresponded to an asymptomatic condition. In 22 cases, the identified PV had not been reported in association with a dominant clinical manifestation. In 15 cases, although the variant had been reported, data regarding its association with a dominant phenotype were limited or lacking. In these cases, the risk was excluded due to discordance between the molecular characteristics of the variant and the pathogenic mechanism underlying the dominant phenotype.

In 13 cases, the identified variants were novel and had not been previously reported; their molecular characteristics did not correspond to the pathogenic mechanism of the dominant phenotype. In 7 individuals, a conditional risk for clinical manifestation of monogenic disorders with mixed inheritance patterns was established (Table 21). To refine the assessment of personal health risk in each of these seven individuals, segregation analysis was recommended to clarify the clinical significance of the identified variants.

Table 21. Pathogenic variants associated with conditional risk for a dominant phenotype

Gene	Dominant phenotype	Transcript	Variant	Variant type	Description
ALG8	Polycystic liver disease 3 with or without renal cysts	NM_024081.4	c.1090C>T (p.Arg364Ter)	Stop gained	The variant has been reported in the heterozygous state in patients with a dominant phenotype, and its molecular characteristics are consistent with the established pathogenic mechanism – LOF."
		NM_024081.4	c.174+1G>T	Splice donor	A novel, previously unreported variant whose molecular characteristics are consistent with the pathogenic mechanism of the dominant phenotype – LOF.
DSG2	Arrhythmogenic right ventricular dysplasia 10	NM_001943.5	c.3G>A	Initiator codon (ATG) loss	The variant has been reported in the heterozygous state in patients with a dominant phenotype.
MME	Charcot–Marie–Tooth disease type 2T (late-onset axonal neuropathy)	NM_007289.4	c.1408G>T (p.Glu470Ter)	Stop gained	A novel, previously unreported variant with molecular characteristics consistent with the pathogenic mechanism of the dominant phenotype
MYH11	Familial thoracic aortic aneurysm 4	NM_002474.3	c.3757AAG[3] (p.Lys1256del)	Inframe deletion	Reported in the heterozygous state in patients with a dominant phenotype. The molecular characteristics are consistent with a dominant-negative pathogenic mechanism.
PLG	Hereditary angioedema type 4	NM_000301.5	c.988A>G (p.Lys330Glu)	Missense	The variant has been reported in the heterozygous state in patients with a dominant phenotype
POGLUT1	Dowling-Degos disease 4	NM_015560.3	c.394C>T (p.Arg132Ter)	Stop gained	The variant's molecular characteristics are consistent with the established pathogenic mechanism. – LOF.

6. Development of a Genetic Counseling Approach in Expanded Carrier Screening, Focused on Reproductive Risk Assessment in Couples and Interpretation of Findings Relevant to Personal Health Risk

6.1 Genetic Counseling in Carrier Screening

CS, as a genetic test, should be accompanied by GC in accordance with the recommendations of international professional organizations (ACOG, ESHG, ACMG). Given the complexity of genetic information and its potential implications, GC in the context of CS should be provided by a qualified and appropriately trained specialist. In line with established international practice, CS is widely accepted to be accompanied by counseling both before (pre-test) and after (post-test) the genetic testing process.

Pre-test counseling

Pre-test genetic counseling in CS includes comprehensive information provided to individuals regarding the nature, purpose, and scope of the test, as well as possible results, limitations of the genetic methods used, and the presence of residual risk in the case of a negative result. During counseling, the concept of carrier status should be explained as a common phenomenon in the general population, typically not associated with clinical manifestation in the carrier.

It should be emphasized that carrier status becomes clinically relevant in terms of RR when both partners carry PV in the same gene associated with ARC, or when the female partner is a carrier of a PV in a gene associated XRC. The possibility of identifying findings related to the individual's personal health risk should also be discussed.

An essential component of counseling is the collection of personal, family, and reproductive history.

Genetic testing

Exome sequencing and clinical exome approaches provide a substantially broader diagnostic scope compared to targeted panels, enabling simultaneous analysis of a large number of genes associated with RC and identification of carrier status for rare conditions that would otherwise remain undetected with limited panels. This enhances the clinical utility of CS and improves identification of at-risk couples.

Simultaneous testing of both partners further optimizes the process by enabling faster and more precise determination of actual RR.

Post-test counseling

Post-test genetic counseling includes interpretation of the genetic test results, discussion of reproductive options, and explanation of the significance of PV relevant to personal health risk, when such findings are identified.

For couples in whom no increased risk for RC in the offspring is identified, recommendations are limited to general population preventive measures before and during pregnancy.

6.2 Genetic Counseling Approach in Expanded Carrier Screening Focused on Reproductive Risk Assessment in Couples

When carrier status for clinically significant variants in the same gene is identified in both partners, a key step in genetic counseling is the detailed evaluation of the molecular characteristics of the variants, as concordance of carrier status alone is not sufficient to define a couple as at risk. The actual RR depends on the type and functional significance of the variants, their genotype–phenotype correlation, and the expected clinical manifestation in a potential homozygous or compound heterozygous genotype.

The presence of hypomorphic variants or variants associated with milder phenotypes may substantially modify risk assessment. According to ACMG guidelines, only pathogenic and likely pathogenic variants should be considered when assessing RR, whereas VUS should not be used for clinical or reproductive decision-making.

In confirmed at-risk couples, genetic counseling should be a structured and non-directive process, tailored to the clinical severity of the condition and the timing of risk identification. The optimal period for identifying at-risk status is the preconception stage, when the full range of reproductive options is available (Table 22). Reproductive history is also of importance, particularly in cases with previous adverse reproductive outcomes or when planning assisted reproductive techniques.

Table 22. Comparison between preconception and prenatal identification of reproductive risk

Aspect	**Preconception risk identification**	**Risk identification during pregnancy**
Time pressure	None or minimal (except advanced parental age)	Significant, constrained by timelines for diagnosis and decision-making
Reproductive options	**Full range of options:** natural conception with prenatal diagnosis (PND), reimplantation Genetic Testing for Monogenic Disorders (PGT-M), donor gametes, or opting out of biological reproduction	Limited options, primarily related to PND and subsequent decision-making
Approach to genetic counseling	Comprehensive, planned, with the possibility of multiple sessions	Time-constrained, requiring prompt and clear communication
Psychological aspects	Lower level of emotional stress	Increased emotional burden requiring support
Decision-making autonomy	Maximally supported	Limited by time and emotional factors
Organizational capacity	Opportunity for referral to specialized centers and strategic planning	Limited, often concurrent with pregnancy
Role of genetic counseling	Optimal management of RR	Management of existing risk and support for time-sensitive decision-making

The GC approach to RR assessment for ARC in the studied couple, based on ECS results, is presented in Figure 14

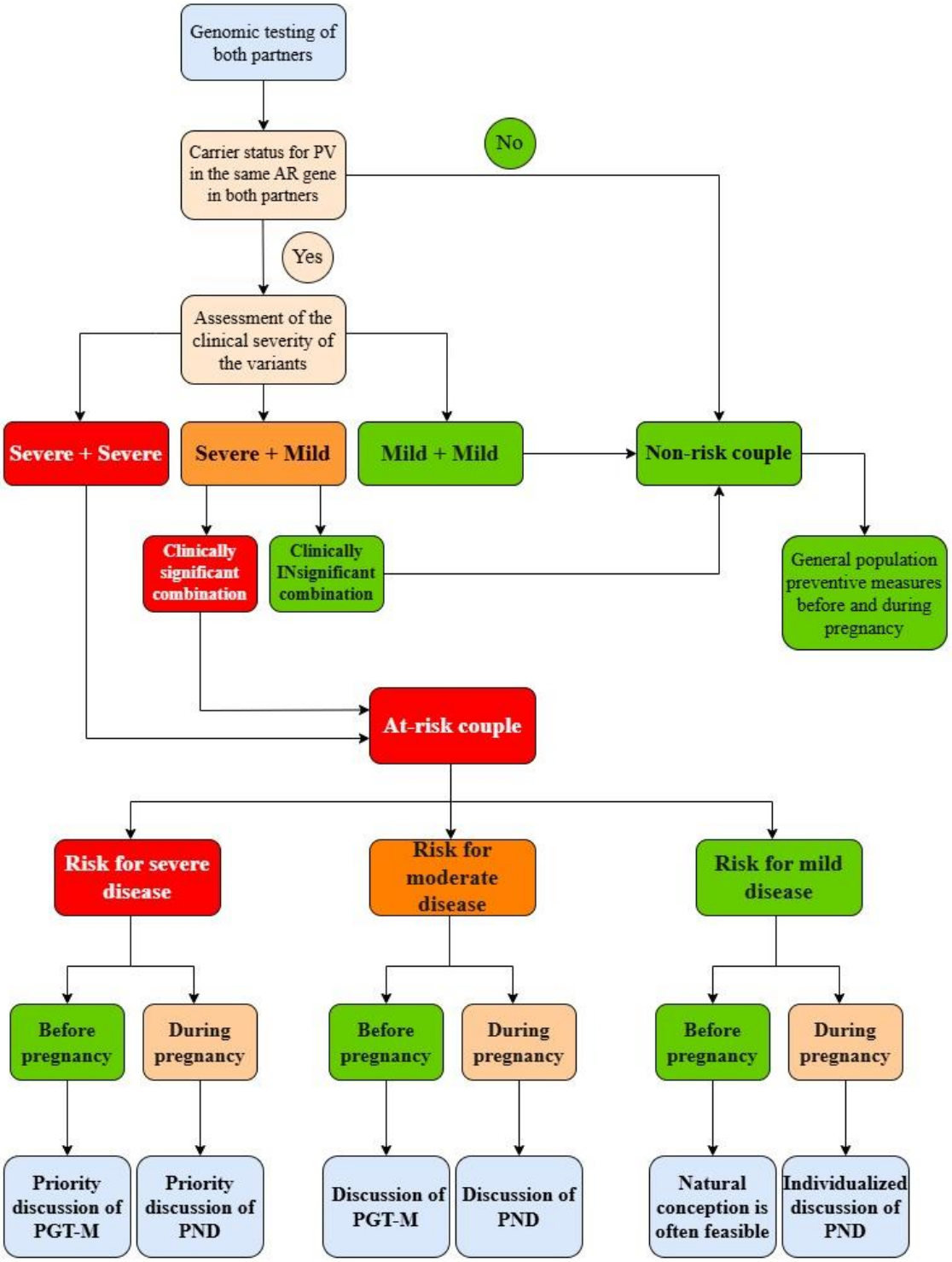

Figure 14. Approach to genetic counseling based on expanded carrier screening results with respect to the couple's reproductive risk

6.3 Approach to Genetic Counseling for Findings Related to the Individual's Personal Health Risk

When genetic findings are identified that extend beyond RR and have potential implications for the individual's personal health status, the approach to such results requires careful interpretation, clear differentiation of risk magnitude, and individualized GC. Depending on the nature and clinical significance of the finding, we propose that these cases be provisionally grouped into several main categories.

It should be emphasized that the disclosure of results with implications for personal health risk is performed only upon explicit request by the individual.

1. Established Personal Health Risk

This category includes findings in which there is a clear and well-established association between the identified PV/genotype and an increased risk of clinical manifestation in the carrier. In such cases, genetic counseling is focused on providing information regarding the natural history of the condition, available options for prevention, surveillance, or treatment, as well as referral for appropriate clinical follow-up and consultation with relevant specialists. The approach follows established clinical guidelines and recommendations for risk management.

2. Potential Personal Health Risk Requiring Further Genetic Characterization

This group includes findings for which there is evidence suggesting a possible association with increased health risk, but the available information is insufficient for a definitive conclusion, necessitating additional genetic analyses. In these cases, genetic counseling aims to clarify the degree of uncertainty, the need for further testing, and the expected timeframe for risk re-evaluation.

3. Conditional Personal Health Risk, Including Genetic Predisposition to Oncological Diseases

This category includes findings associated with a predisposition to cancer, in which the risk of clinical manifestation is conditional and depends on multiple factors, including gene penetrance, the age of the individual, family history, and environmental influences. Genetic counseling is again focused on interpreting risk in an individualized context, discussing recommendations for surveillance and prevention, and avoiding overinterpretation or

unnecessary anxiety.

4. Conditional Risk of Clinical Manifestation in Monogenic Disorders with Mixed Inheritance (AR/AD)

The most specific and complex group includes findings in genes associated with monogenic disorders for which both AR and AD forms have been described. In such cases, identification of a PV in the context of carrier screening requires a precise assessment of the risk of clinical manifestation in the carrier, for which we propose a stepwise analytical approach.

The initial and essential step in interpretation is the assessment of the presence or absence of family history, including evidence of clinical manifestations in first-degree and more distant relatives.

In the presence of a family history consistent with the dominant form of the disease, the risk of clinical manifestation in the carrier is considered increased. In this context, segregation analysis plays a key role in refining individual risk. This involves targeted genetic testing for the specific variant (identified in the counseled individual) in affected relatives with a phenotype consistent with the condition under consideration. The aim is to confirm or refute the association between the variant and the observed phenotype. In parallel, an active consultative and follow-up approach is recommended, including extended clinical evaluation and referral to an appropriate specialist.

In the absence of family history, or if segregation analysis does not demonstrate a correlation between the variant and the phenotype in affected relatives, a similar approach is applied. The risk of clinical manifestation may be excluded in the following situations:

- In individuals whose age exceeds the typical age of onset of the dominant phenotype;
- When the identified PV is associated exclusively with a recessive phenotype;
- When the variant has been reported, but available evidence linking it to a dominant phenotype is limited or lacking, its molecular characteristics should be evaluated. If these are not consistent with the pathogenic mechanism of the dominant form of the disease, the risk is excluded. For example, a variant with a LOF effect (e.g., frameshift, nonsense, or splice variant) identified in a gene in which the dominant phenotype is caused by a gain-of-function or dominant-negative mechanism. Conversely, a gain-of-function variant (e.g., missense) identified in a gene in which the dominant phenotype

is due to haploinsufficiency would not support an increased risk;

- In the case of a novel, previously unreported variant whose molecular characteristics do not match the pathogenic mechanism described for the dominant phenotype.

In these situations, the risk of clinical manifestation in the carrier is considered low or negligible, and genetic counseling is focused on clarifying the result, preventing misinterpretation, and ensuring the possibility of future re-evaluation as new data become available.

An increased risk may be considered when the dominant phenotype has adult onset and:

- the identified variant is associated with a dominant phenotype;
- the identified variant (novel or with limited evidence in the literature) has molecular characteristics consistent with the pathogenic mechanism of the dominant phenotype.

It should be emphasized that conditional risk is not equivalent to a diagnosis and does not imply inevitable clinical manifestation, but rather represents a framework for follow-up and re-evaluation. Management in such cases includes referral for appropriate clinical follow-up, recommendations for periodic re-assessment of the genetic finding as new scientific and clinical data accumulate, and consideration of additional investigations or specialist consultations where appropriate. Genetic counseling in these cases aims to prevent both underestimation and overinterpretation of risk, while maintaining the principles of non-directiveness and informed choice.

The approach to GC in the presence of a PV in a gene associated with a monogenic disorder with mixed inheritance is illustrated in Figure 15.

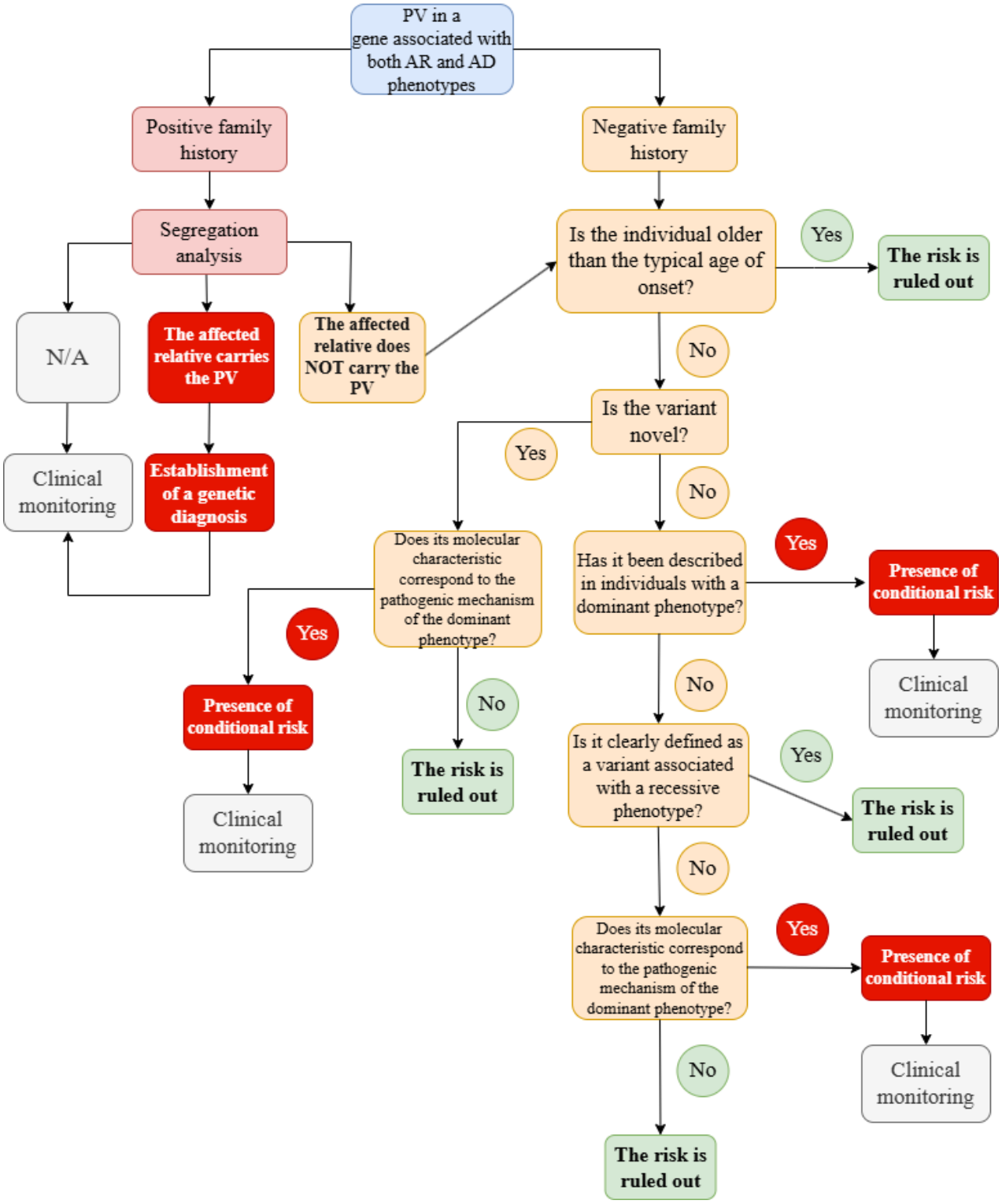

Figure 15. Approach to genetic counseling in the presence of a pathogenic variant in a gene associated with both recessive and dominant phenotypes, and assessment of the carrier's personal health risk

SUMMARY OF THE PRESENT STUDY

In the modern era, RC represent a significant medical and social challenge due to their population frequency, early onset, and potentially severe clinical course. Although individual conditions are rare, their cumulative impact is substantial, as carrier status for PV in genes associated with RC is widespread. This underlies the key role of ECS as a tool for reducing adverse reproductive outcomes by identifying couples at increased risk and supporting reproductive autonomy through informed decision-making.

In the present study, we found that approximately 96% of the analyzed individuals (n=287/300) were carriers of PV in at least one gene associated with RC. The high carrier rate is expected given the broad detection capacity of the applied methodology – NGS using a clinical exome panel. At the same time, this finding supports the concept that carrier status should be considered the norm, particularly for certain RC in specific populations.

Within the cohort, a total of 690 distinct PV were identified across 481 genes associated with RC. Approximately one-third, 32% (n=220/694), of the detected variants were novel and had not been previously reported in clinical or population databases at the time of analysis. This finding highlights the high degree of genetic heterogeneity in the Bulgarian population, as well as its limited representation in international reference databases. At the same time, it underscores the diagnostic potential of exome-based approaches for identifying rare variants.

Two major interpretative challenges emerged from the present study. First, the identification of at-risk couples for ARC cannot rely solely on shared carrier status for PV in the same gene. Instead, it requires detailed evaluation of the molecular characteristics of the variants, including their functional effect, penetrance, and associated clinical severity. This is clearly illustrated by our findings, where 11% (n=17/150) of reproductive couples were initially classified as potentially at risk based on shared carrier status. Following detailed assessment of the clinical relevance and severity of the identified variants, the proportion of truly at-risk couples decreased to 9% (n=13/150).

Second, exome-based CS inevitably leads to the identification of genetic findings with potential implications for the individual health risk of carriers, extending beyond the primary reproductive aim of the study and raising additional clinical and ethical challenges. In support

of this, findings relevant to personal health risk were identified in 13% (n=38/300) of the studied individuals. The largest proportion of these findings was related to predisposition to oncological diseases (66%, n=25/38). In 18% (n=7/38) of individuals, a conditional risk for dominant clinical manifestation of a MD associated with genes exhibiting mixed inheritance patterns (AR/AD) was identified, while in 13% (n=5/38) a confirmed personal health risk was established. In a single case (3%, n=1/38), a finding with potential relevance to personal health risk was identified, requiring additional genetic testing to clarify its clinical significance. These results demonstrate that the exome-based approach substantially expands the scope of genetic information beyond RR and necessitates an integrated approach to interpretation and GC.

V. CONCLUSIONS

1. The carrier frequency of PV in genes associated with RC identified in a cohort of individuals from the Bulgarian population was 96%, indicating that nearly every individual carries a variant in at least one gene associated with RC, with an average carrier burden of 3.5 PV per individual. The high carrier detection rate confirms the advantages of the applied methodology—NGS using a clinical exome panel—for comprehensive variant detection in CS.

2. The majority (92%) of the affected genes (associated with RC) were rare, while 8% met the criteria for commonly affected genes (carrier frequency in ≥5 individuals and frequency ≥1:60). The clinical conditions associated with the latter group are considered relatively common, with approximately two-thirds characterized by mild or moderate clinical severity. This supports the hypothesis that many severe genetic conditions are subject to negative selection, resulting in lower population prevalence.

3. Analysis of the molecular characteristics of PV in the most frequently affected genes revealed significant heterogeneity, including variants with clear pathogenic effects as well as hypomorphic and context-dependent variants. Although variants associated with severe clinical impact predominated (48%), their carrier frequency in the studied cohort was comparable to that of milder variants.

4. Approximately one-third of the identified PV were novel and had not been previously reported at the time of analysis. This highlights the high diagnostic and scientific value of NGS for identifying rare and previously unreported variants with implications for both reproductive and personal health risk.

5. Among all included couples, 11% were initially classified as potentially at-risk, based on the combination of identified genetic variants in both partners suggesting an increased RR. Following additional analysis incorporating detailed molecular and clinical data, the proportion of couples with a confirmed RR was reduced to 9%.

6. Among couples with confirmed RR, 62% were associated with conditions of moderate clinical severity. In 77% of these cases, the affected gene (and corresponding phenotype) belonged to the group of rarely affected genes, indicating that such conditions could be missed by standard carrier screening approaches not based on exome analysis. This

underscores the importance of an exome-based, clinically oriented approach for precise RR assessment.

7. Findings relevant to personal health risk were identified in 13% of the studied individuals, predominantly (66% of these cases) related to predisposition to oncological diseases, and to a lesser extent (18%) associated with a conditional risk of dominant manifestation of monogenic disorders involving genes with mixed inheritance patterns (AR/AD).

8. Exome sequencing (clinical exome) with simultaneous testing of both partners emerges as an optimal strategy for CS, given the pronounced genetic heterogeneity of the Bulgarian population.

9. Genetic counseling is an integral component of carrier screening, both in the pre-test and post-test setting. In 21% of the studied individuals, CS identified findings with implications for reproductive or personal health risk, and this clinically relevant information was provided within the framework of specialized GC.

10. A comprehensive approach to genetic counseling in the context of ECS has been developed, in which the role of the genetic counselor is central at all stages of the process. This approach begins with the initial contact with the couple (including information about the test and pedigree assessment), continues through expert evaluation of test results (molecular and clinical interpretation, assessment of reproductive and personal health risk), and concludes with risk communication, recommendations, and support for informed and autonomous reproductive decision-making.

VI. CONTRIBUTIONS

Scientific and Original Contributions

1. For the first time in Bulgaria, a comprehensive study based on prospectively conducted carrier screening in couples from the Bulgarian population has been performed. The study was carried out using a unified and standardized protocol and was based on an expanded gene panel of 6,699 genes utilizing NGS technology, enabling highly sensitive assessment of carrier status for pathogenic variants in genes associated with recessive conditions.
2. Novel data are presented regarding carrier frequency, the profile of affected genes, and the spectrum of pathogenic variants in a studied cohort from the Bulgarian population, thereby complementing the currently limited data available for Bulgaria.
3. A large number of novel variants were identified, which have not yet been reported and will contribute to the enrichment of both national and international genetic databases.
4. The role of carrier screening using NGS technology with a clinical exome approach is confirmed for the identification of novel or rare pathogenic variants with potential implications for both reproductive and personal health risk.

Practical (Applied) Contributions

1. It is demonstrated that, given the genetic heterogeneity of the Bulgarian population, the application of expanded carrier screening using a clinical exome represents an optimal strategy for identifying carrier status of pathogenic variants in genes associated with recessive conditions.
2. An approach for assessing reproductive risk in recessive conditions based on carrier screening results has been developed, along with a framework for genetic counseling in such couples.
3. An approach has been established for the evaluation of findings/pathogenic variants relevant to the individual health risk of carriers, as well as for providing genetic counseling in these cases.
4. A comprehensive model for genetic counseling in the context of expanded carrier screening has been developed. This model begins with the initial contact with the couple (including test-related information and pedigree assessment), continues through expert evaluation of test results (molecular and clinical interpretation, assessment of actual reproductive and personal health risk), and concludes with risk communication, recommendations, and support for informed and autonomous reproductive decision-making.

VII. PUBLICATIONS AND CONFERENCE PRESENTATIONS RELATED TO THE DISSERTATION

Publications:

- Kovacheva KS, **Nikolova SE**, Kamburova ZB. Carrier screening for single-gene disorders – A brief review. Journal of Biomedical and Clinical Research, 2021, 14(2):105–116; DOI: 10.2478/jbcr-2021-0015.
- **Nikolova SE**, Kamburova ZB, Vasilev PP, Kovacheva KS, Yordanova IA. Autosomal recessive type of dystrophic epidermolysis bullosa with a novel variant in the COL7A1 gene. Biomedical Reports, 2024, 21(5):167; DOI: 10.3892/br.2024.1855; PMID: 39301563; PMCID: PMC11411400; Web of Science (ESCI) / Scopus.
- Kamburova ZB, Dimitrova PD, Dimitrova DS, Kovacheva KS, Popovska SL, **Nikolova SE**. Lynch-like syndrome with germline WRN mutation in Bulgarian patient with synchronous endometrial and ovarian cancer. Hereditary Cancer in Clinical Practice, 2023, 21(1):13; DOI: 10.1186/s13053-023-00257-1; PMID: 37452354; PMCID: PMC10349469; Web of Science / Scopus.

Participation in scientific conferences:

- **Nikolova S,** Kovacheva K, Kamburova Z**.** Carrier screening: a pilot study amongst a Bulgarian population. *Jubilee Scientific Conference with International Participation "50 Years of Medical Education and Science in Pleven"*, 01–03 November 2024.
- **Nikolova S,** Kovacheva K, Kamburova Z, Ivanova M, Stankov I. Variant c.895_904del (p.Val301SerfsTer8) in a newborn girl with isolated postaxial polydactyly – a case report. *European Human Genetics Conference 2023*, Glasgow, Scotland, UK, 10–13 June 2023.
- **Николова С,** Ковачева К, Камбурова З, Иванова М, Станков И. Роля на геномните изследвания при пренатално установена бъбречна поликистоза на плода. *Трета научна конференция „Генетика в клиничната практика"*, хотел Белвил, Дюни, 25–28 май 2023.