SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHYLOSOPHY (PHD)

Mutation spectrum determination in monogenic diseases using next generation sequencing methods

by

László Madar

Supervisor: Prof. Dr. István Balogh

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2023

Mutation spectrum determination in monogenic diseases using next generation sequencing methods

By László Madar

Supervisor: Prof. Dr. István Balogh, PhD, DSc

Doctoral School of Molecular Cell and Immune Biology, University of Debrecen

Head of the **Defense Committee**:	Prof. Dr. Éva Csősz, PhD, DSc
Reviewers:	Dr. Nikoletta Nagy, PhD, DSc
	Dr. István Szatmári, PhD
Members of the Defense Committee:	Dr. Emese Beatrix Horváth, PhD
	Dr. Attila Borbély, PhD

The PhD Defense takes place at Department of Internal Medicine, Building A, University of Debrecen, at 13.00 p.m. on 15th of February 2024.

I. INTRODUCTION

In my studies I investigated two hereditary monogenic diseases, the genetic background of which has not been investigated in Hungary due to the size or number of the underlying genes, so only clinical incidence data were available without any genetic information on the disease background. Knowledge of the genetic background can greatly assist clinicians in assessing and preventing either potential treatments or expected complications of a given disease. Furthermore, patient data can be compared with the literature, providing a wealth of information that can be used to assess, for example, which variants are carried by patients in the home country. We can obtain information on the ethnic origin of certain variants or discover founder variants in the Hungarian population.

Marfan syndrome is a monogenic hereditary disease based on the development of abnormal connective tissue structure, which explains the involvement of various organ systems. Patients can present with a wide range of symptoms. The most prominent are skeletal abnormalities such as high stature or long fingers and limbs, and in the majority of patients, an abnormal position of the eye lens. The greatest risk to patients, however, is cardiovascular symptoms caused by lesions in the vascular walls. Marfan syndrome is most often caused by a single major gene defect, but there may also be abnormalities in other genes.

Familial hypercholesterolaemia is a type of high cholesterol that is caused by an abnormality in a gene that plays a key role in cholesterol metabolism. It is important to highlight that, to date, it is an under-diagnosed disease with very serious cardiological risks, with the majority of patients presenting at a young age due to cardiological complications. In most cases, this is when a possible family history is discovered, as cascade tests can often detect several family members who are still asymptomatic. It is important to note that in cases where a genetic abnormality is confirmed, medication is necessary from an early age to achieve the cholesterol target and to prevent cardiological complications.

For any monogenic disease, which may be caused by multiple, often large genes, genetic testing and identification of the variant responsible for the disease can be challenging.

Both Marfan syndrome and familial hypercholesterolaemia are monogenic diseases whose genetic background has not been investigated in our country due to the size and number of genes and the high costs, high labour requirements and long turnaround times. The appearance of a new generation of sequencing techniques in the early 2000s fundamentally changed molecular genetic testing. In contrast to Sanger sequencing, it became possible to test

multiple genes and multiple patients simultaneously with leaps compared to previously known methods, and at a much higher speed and cost-effectiveness. Thanks to rapid technological advances, the performance of these devices is doubling every few years, further reducing sequencing costs and increasing performance and reliability.

In my first study, my objectives included the genetic analysis of Marfan syndrome patients in Hungary and the development of a new method for routine use, including the potential of next-generation sequencing. So far, the size of the genes responsible for the disease has made it extremely time-consuming and costly for laboratories to screen such patients. The technological breakthrough has made it possible to test several patients at once in a significantly shorter time. After evaluating the results, we further investigated the detected variants to gain a picture of the variants in Hungary, focusing on genotype-phenotype correlations.

The second study aimed to analyse patients with familial hypercholesterolaemia in the country, where it is even more notable that only clinical data are available, without genetic information. Genetic testing can also be of great importance for the therapy of the disease, as different drug treatments are used when certain abnormalities are detected, and genetic results can also provide important information on the course of the disease. The variants detected were compared with the type and distribution of variants described in other countries and their location within genes was analysed. We also examined and reported the genotype-phenotype correlations in the patient. In the process of our work, we have also performed further detailed molecular and segregation analysis for some of the variants.

1.1 Sanger DNA sequencing

Sanger sequencing is the gold standard method used in both studies, developed by Frederich Sanger and his colleagues in 1977. The method is used to determine the exact nucleotide sequence of a given DNA fragment. The method consists of amplifying a DNA fragment while terminating replication with didezoxynucleotides. In the original method, the synthesis is carried out in four tubes, where different didezoxy nucleotides are added to the deoxyribonucleotides, so that the chain termination is stopped at a different nucleotide type in each case. Ultimately, fragments of different lengths are produced in each tube by ending up on the same nucleotide. Radioactively labelled dATP is added to the reaction. The contents of the four tubes are run on a polyacrylamide gel electrophoresis with sufficient resolution to detect a single nucleotide difference. The fragments were visualised by autoradiography. When the four samples are run side by side, fragments of different sizes are separated by length. Reads are taken from the furthest, shortest DNA fragment. This method is highly accurate but extremely

time-consuming, so an automated version of it was developed, also based on the Sanger method, with the modification that the four types of didezoxynucleotides are fluorescently labelled in different colours, so that they fluoresce in different colours when excited. This allowed readout in a single tube using a laser.

A detector identifies the nucleotides based on the wavelength of the light emitted and a program can be used to read and sequence the detected signals.

1.2 Next generation DNA sequencing

Next-generation DNA sequencing refers to high-throughput methods that are able to determine the nucleotide sequence of the entire human genome simultaneously and in parallel. The new methods differ significantly from Sanger sequencing in both the sequence synthesis and the detection method. They require highly advanced and precise automation and large computing capacity due to their speed and the amount of data generated. The most common of these methods is called Sequencing by synthesis (SBS), because the reads are taken when a complementary strand to the DNA strand under investigation is synthesised and detected in real time. Next-generation DNA sequencing methods have 3 main steps: library preparation, sequencing and alignment or assembly.

1.2.1 Pyrosequencing

Pyrosequencing is a next-generation sequencing method developed in the late 1990s and commercialised in the early 2000s. Pyrosequencing is also based on the principle of SBS, or synthesis-based sequencing, whereby a complementary strand of a single-stranded DNA molecule is synthesised enzymatically. The synthesis is monitored in real time using a chemiluminescent enzyme. The incorporation of a nucleotide produces pyrophosphate (PPi), the amount of which is measured by a coupled enzyme reaction. At the end of the reaction is a flash of light induced by the luciferase enzyme, which we are able to detect using an advanced sensitive camera system. In order to obtain a detectable signal from the light flash, it is necessary to amplify the signal, which can be achieved by parallel sequencing reactions. For this purpose, the so-called emulsion PCR reaction has been developed, whereby the DNA fragment to be tested is amplified by PCR on polystyrene beads. This process is called clonal amplification. Other reagents required for sequencing are added to the sequencing plate by the automated system. In these small wells, sequencing reactions take place independently and in parallel, and are continuously processed by the camera. In the process, it is important that only one type of nucleotide is present at a time, which is either incorporated into the newly synthesised DNA strand or not. Nucleotides that are not incorporated are removed from the system before the next nucleotide is added. If the nucleotide is incorporated, the signal size will

be one unit. If identical nucleotides, called homopolymers, are found in sequence on the DNA strand being tested, the signal intensity increases in proportion to the size of the homopolymers, for example, if 3 adenines are incorporated, the signal detected will be three times the size. A known problem with the pyrosequencing method is that this theoretically linear signal increase above a certain homopolymer size does not always give reliable results. This number is usually reliable up to 4 nucleotides but above this number it can give an uncertain measurement, depending on the type of nucleotide and the length of the homopolymer.

1.2.2. Illumina sequencing technology

The most widely used sequencing method worldwide is currently the Illumina synthesis-based sequencing technology. Originally designed for short and fast reads, it is now available for hundreds of nucleotide reads. To prepare the sample, DNA is fragmented into short fragments of around 100-200 bp either mechanically or enzymatically. The ends of the double-stranded DNA strands are corrected and an adenine is added to the 3' ends. An adapter piece of DNA with an overhanging thymine is ligated to this. This adapter contains the so-called index sequences, which serve as a kind of molecular barcode to identify the samples, and short sequences at the ends of the adapters that are identical at each amplicon end (designated as adapters P5 and P7). These allow the denatured single-stranded DNA library to bind to the sequencing chips, such that complementary primers are anchored to the chip surface by these sequences. Detection here also requires clonal amplification, which in this technology means a so-called bridge amplification. In this step, the molecules bound to the chip surface are amplified in situ and after amplification, the DNA molecules form so-called clusters. Each cluster represents thousands of copies of a single copy of DNA, located very close together. Since with this technology, each of the four types of nucleotides is labelled with a different fluorescent dye, identification is done one at a time, unlike pyrosequencing.

2. Theoretical background

2.1 Marfan syndrome

Marfan Syndrome (MFS, MIM#154700) is an autosomal dominant inherited rare connective tissue disorder. In the past, MFS was also known as arachnodactyly (spider fingers), as it is a symptom of Marfan syndrome, which is characterised by abnormally long, slender or spider-like fingers and toes. Approximately half of cases have a positive family history. It mainly affects the cardiovascular, ocular and skeletal systems, with a prevalence of 1 to 5 per 10,000 and a prevalence unrelated to ethnicity or gender. Aortic valve and aortic arch diameters are significantly greater in patients with a positive family history than in those without such a

history, and life expectancy has been shown to be shorter. Clinical manifestations of Marfan syndrome become more apparent with advancing age. One of the most common symptoms in patients with Marfan syndrome is myopia and 60% of patients develop an abnormal position of the lens (ectopia lentis). Patients are also at increased risk of retinal detachment, glaucoma and early cataracts. Other common symptoms include involvement of the skeletal and connective tissue system, such as joint laxity, dolichocephaly, pectus excavatum, pectus carinatum and scoliosis. The diagnosis of Marfan syndrome is based primarily on clinical and revised Ghent nosology. Most commonly, mutations in the FBN1 (fibrillin 1, 15q21.1, MIM 134797) gene lead to Marfan syndrome or other connective tissue diseases called type I fibrillinopathies, e.g. Shprintzen-Goldberg syndrome, Weill-Marchesani syndrome and ectopia lentis syndrome. In patients with a phenotype reminiscent of Marfan syndrome, abnormalities in the TGFBR1 and TGFBR2 genes are rare but occur. The clinical features of individuals with Marfan syndrome with such mutations can often overlap with other severe connective tissue diseases, such as Loeys-Dietz syndrome (LDS) or familial thoracic aortic aneurysm (FTAA) syndrome.

The FBN1 gene itself was described in 1991. The protein encoded by the gene is a cysteine-rich monomeric glycoprotein with a molecular mass of 350 kDa. The major component of the 10-12 nm microfibrils of the extracellular matrix is fibrillin1. The protein is divided into 5 structural domains (A-E). The FBN1 gene is large, spanning approximately 200 kb of genomic DNA, organised into 65 coding exons and with a transcript size of 10 kb. To date, more than 3000 pathogenic or putative pathogenic mutations have been described, and the majority of these mutations are unique variants. Early mortality and morbidity are closely associated with cardiovascular symptoms. Progressive dilatation of the aortic root and ascending aorta predisposes to aortic rupture. Although mutation detection rates are higher in patients who meet all of the Ghent criteria, testing for FBN1 mutations may be useful in a number of clinical situations, including testing patients who do not meet the Ghent criteria or in sporadic cases at a young age to determine whether aortic dilatation requires follow-up or preventive treatment.

2.2 Familial hypercholesterolaemia

Familial hypercholesterolaemia (FH, MIM#143890) is one of the most common autosomal dominant genetic diseases. It is a disorder of cholesterol metabolism, predisposes to premature atherosclerosis and cardiovascular complications. Its prevalence is estimated to be 1:200-250, while the estimated prevalence in the Hungarian population is 1:340. Early detection and appropriate treatment of the disease greatly improves clinical outcome. The diagnosis of FH is

primarily based on the clinical presentation of the patient. The disease is still underdiagnosed and under-treated.

To aid diagnosis, a FH scoring system has been established based on the recommendations of the European Atherosclerosis Society. This scoring system, known as the Dutch Lipid Clinic Network Score (DLCNS), has been used mainly in the Netherlands and has been adopted in clinical practice in several countries. The scores are calculated by clinicians on the basis of the patient's family history, clinical history and physical examination, which can be used to decide whether the diagnosis of FH is unlikely, possible, probable or definite.

The LDL molecule binds to the LDL receptor, which is located on the cell surface and then attaches to a membrane-bound vesicle. LDL is degraded by lysosomal enzymes and the released cholesterol is then released into the cell plasma. The LDL receptor returns to the surface of the cell so that it can reintroduce cholesterol into the cell by binding more LDL. The binding between LDL and its receptor can also be impaired if there is a mismatch in the apoB100 carrier protein. In this case, even if the receptor is present, if apoB100 does not bind strongly enough, cholesterol uptake is slowed.

Most cases of FH are caused by mutations in the *LDLR* (Low density lipoprotein receptor, 19p13.2, MIM#143890) and *APOB* (Apolipoprotein B, 2p24.1, MIM 144010) genes. Much less frequently, but mutations in other genes causing FH have also been described, such as the *PCSK9* gene or, even more rarely, the *LDLRAP1* and *STAP1* genes. Mutations in these genes may have a negative effect on the proper elimination of LDL lipoproteins (low-density lipoproteins) from the blood. This may result in abnormal increases in LDL cholesterol (LDL-C) and total cholesterol (TC) levels.

Patients with heterozygous mutations have LDL-C levels higher than 5 mmol/L, while patients with homozygous or complex heterozygous genotypes usually have LDL-C levels higher than 13 mmol/L. Clinical symptoms can be more severe in homozygous or complex heterozygous patients. In these patients, phenotypic symptoms may appear that raise the suspicion of homozygous familial hypercholesterolaemia. Such symptoms include cutaneous and tuberous xanthomas, extracellular deposition of fats in the cornea (arcus corneae), and atherosclerotic cardiovascular disease (ACVD) which present young, usually before the age of 20 years. To date, more than 2100 *LDLR* and 250 *APOB* mutations have been described in the literature. No mutation hotspots have been identified in the *LDLR* gene, whereas the *APOB* gene has a long-standing mutation hotspot in exon 26. In the European population, a recurrent variant was known as early as the late 1980s and was reported in the literature as R3500Q for a long time.

2.2.1 Structure and function of *APOB*

The *APOB* gene is located on chromosome 2, consists of 29 exons and encodes 4563 amino acids. Apoproteins play a key role in binding lipoproteins to receptors. A major function of lipoproteins is to supply cells with lipids and to remove excess lipids. There are 2 forms of ApoB, one is ApoB100 and the other is ApoB48. Both protein forms are products of the *APOB* gene, but APOB48 is a truncated form of APOB100. APOB100 is produced in the liver, while APOB48 is produced in the proximal small intestine.

2.2.2 Structure and function of the *LDLR* gene

The low-density lipoprotein receptor is encoded by the 44.5 kb *LDLR* gene, located on the short arm of chromosome 19. The 18 exon gene encodes a type I transmembrane protein of 860 amino acids with 5 functional domains. LDLR mediates the clearance of cholesterol and cholesterol ester-containing low-density lipoprotein (LDL) particles from the blood. LDLR plays a key role in cholesterol homeostasis. This is demonstrated by the fact that more than 2100 *LDLR* mutations have been described in patients with familial hypercholesterolaemia. These mutations may affect LDLR activity to varying levels. Based on the phenotypic effects of the mutations, *LDLR* mutations can be classified into 5 classes. Class I: null allele, Class II: allele with a transport defect, Class III: allele with a binding defect, Class IV: allele with an internalisation defect, Class V: allele with a recycling defect.

3. Objectives

1) Our aim is to investigate the applicability and efficiency of different high-throughput, next-generation DNA sequencing methods for monogenic diseases where the size or number of pathogenic genes would have made diagnostic testing with conventional methods expensive and time-consuming.

2) On the basis of molecular genetic analysis of unrelated Hungarian patients with suspected Marfan syndrome, we aimed to determine the mutational spectrum of the Hungarian patient population, to summarize its phenotypic features and to investigate the pathogenicity of a recurrent silent mutation.

3) For Hungarian clinically suspected FH patients, we aimed to perform genetic testing and to determine the mutation spectrum and compare it with literature data. To describe the suspected

pathological abnormalities and genotype-phenotype correlations detected in unrelated patients and to determine the exact breakpoint in the case of a single exon deletion.

4. Patients and Methods

4.1. Patients

4.1.1 Patients with suspected Marfan syndrome

In twenty-six patients who presented with suspected Marfan syndrome or related connective tissue disorders, the phenotypic features and available clinical data, together with the detected abnormalities, are discussed in detail in the results section. The clinical diagnosis, when available, was made by a clinical geneticist, paediatrician or internist using the Ghent criteria. The mean age of the patients was 19 years. The youngest patient was 1 month old and the oldest 52 years old. Seventeen patients were younger than 20 years and 2 patients were under 1 year old. For all patients, we had a genetic informed consent form signed by the patient or legal representative.

4.1.2 Patients with suspected familial hypercholesterolaemia

Patients with a clinical diagnosis of familial hypercholesterolaemia (FH) from all over the country were studied. For patients where a genetic abnormality was detected, genetic testing of available blood family members was also performed. For each patient, we had a genetic consent form signed by the patient or legal representative. Relevant clinical information and laboratory results were available for most patients. Clinical diagnosis was made by a clinical geneticist or an internist, where possible, based on laboratory findings and possible phenotypic features. Patients with pathogenic variants included 31 females and 13 males. The mean age of the patients was 48 years. The youngest was 14 years old and the oldest 78 years old at the time of clinical diagnosis. The mean LDL cholesterol level was 7.2 mmol/L, while the mean total cholesterol level was 10.4 mmol/L among patients carrying at least one pathogenic variant.

4.2 Genomic DNA isolation

DNA isolation was performed from peripheral blood samples coagulated with K3-EDTA (ethylenediamine tetraacetic acid) or Na-citrate using the QIAamp Blood Mini kit according to the manufacturer's instructions. Concentrations and purity of extracted DNA samples were

determined using a NanoDrop 2000/2000c UV-Vis spectrophotometer and a Qubit 4 Fluorometer, respectively.

4.3 Application of pyrosequencing and Sanger sequencing to analyze the FBN1 gene

In fifteen patients, the FBN1 gene coding exons and exon-intron boundaries were amplified in 65 individual PCR reactions using PCR primers from a previous publication. Twenty-seven PCR primers were redesigned using the Primer3 design software to meet our requirement that all PCR products should be of similar length, as PCR products that are too short would be over-represented in clonal amplification. Amplicon analysis was performed by bidirectional pyrosequencing with a coverage criterion of 40x (coverage is defined as the number of times a nucleotide was read) on a Roche GS Junior 454 pyrosequencer. Because of the well-known homopolymer problem of pyrosequencing methodology, 49 exons were analyzed by pyrosequencing, while 16 exons containing homopolymers were analyzed by Sanger sequencing.

4.4 Next-generation sequencing on the Illumina platform for patients with Marfan syndrome

In eight patients with suspected Marfan syndrome, a commercially available kit called Marfan MASTR was used to test the FBN1 gene. This kit is a multiplex PCR based kit that tested the coding exons and exon-intron boundaries of the FBN1 gene. For 3 additional patients, we used a self-designed gene panel (Qiagen) to test not only FBN1, but also TGFBR1 and TGFBR2 genes. Unlike the previous kit, this gene panel was a so-called "capture" library kit. The prepared samples were run on an Illumina Miseq. The raw data were analyzed using NextGene analysis software.

4.5 FBN1 gene mRNA analysis

Skin biopsies were taken from the axilla of patient 26 for culturing and analysis of fibroblast cells. Total RNA was isolated from skin fibroblasts after 3-5 passages using TRIzol reagent. Thereafter, cDNA was obtained from the sample by reverse transcription using High-Capacity cDNA Reverse Transcription kit. After reverse transcription, exon 26 and the surrounding exon-exon boundaries were amplified and analyzed by bidirectional Sanger sequencing. Primers were designed using Primer3 software. The ratio of mutant to wild-type alleles was also determined

from both cDNA and genomic DNA. Pyrosequencing was used to determine the ratio and the raw data were evaluated using Amplicon Variant Analyzer software.

4.6 Sequencing the genes responsible for familial hypercholesterolaemia

To test the genes most commonly associated with FH, the Devyser FH V2 kit was used according to the manufacturer's instructions. This amplicon-based library construction kit is able to screen gene coding regions and exon/intron boundaries at a minimum distance of +/- 10bp. The kit tests the following genes *APOB*: MIM#144010, *LDLR*: MIM#143890, *PCSK9*: MIM#603776, *LDLRAP1*: MIM#603813, *STAP1*. Sequencing was performed on an Illumina MiSeq with 300 cycle reads according to the manufacturer's instructions. Raw data were evaluated using NextGENe software with a coverage criterion of 40x. Samples were also analysed for copy number variation (CNV) based on coverage data. The detected pathogenic variants were also confirmed by Sanger sequencing.

4.7 MLPA analysis of the *LDLR* gene

In three cases (patients 12, 33, 38), MPLA was used to confirm copy number variations detected by next-generation sequencing using the SALSA MLPA Probemix P062 MIX1 kit according to the manufacturer's instructions.

4.8 Variant filtering and classification when examining genes associated with FH

All detected variants with a MAF (minor allele frequency) >0.01, that is, greater than 1%, were filtered out in the GnomAD population database as non-pathogenic variants. ACMG (American College of Medical Genetics and Genomics) recommended classification of the remaining variants was performed. ACMG uses a very widely accepted and used five-category system, which are: pathogenic, probably pathogenic, of uncertain clinical significance, probably benign, benign. In addition, the HGMD (The Human Gene Mutation Database) professional and ClinVar mutation databases were used to classify variants.

4.9 Breakpoint determination in the *LDLR* gene

In one patient, a heterozygous deletion of exon 7 in the *LDLR* gene was detected by next-generation sequencing. To determine the exact breakpoints, primers were designed around the putative deletion in introns 6 and 8. Following PCR amplification, bidirectional Sanger sequencing was performed.

4.10 Mutation databases and prediction tools

During my work I used several databases to check and describe the detected variants, which were Human Gene Mutation Database (HGMD) Professional, ClinVar, The Genome Aggregation Database (gnomAD), 1000 Genomes project, Mutation Taster, SIFT, PolyPhen-2, Franklin, Human Splicing Finder.

5. Results

5.1 Variants detected in Hungarian patients with Marfan syndrome

5.1.1 Analysis of the FBN1 gene mutation

In this study, genomic DNA and, in one case, total RNA samples were analysed from a total of 26 patients with Marfan syndrome or other related fibrillinopathies with a primary diagnosis. A total of 23 unique pathological or potentially pathological FBN1 abnormalities were detected in 23 patients. More than 30.4% (7/23) of the abnormalities were new variants, while 69.6% (16/23) were recurrent variants. A family history was available in 23 patients, which revealed that 7 cases were sporadic and 16 cases were familial. We detected missense mutations in 69.6% (16/23), small deletions/duplications in 13% (3/23) and splice site variants in 17.4% (4/23). In addition, we detected one recurrent silent mutation in three unrelated patients, which we were able to classify as a harmless variant after mRNA analysis.

5.1.2 Cysteine residue-affecting missense FBN1 variants

Of 16 missense variants, 11 (68.8%) affected conserved cysteine amino acids in the fibrillin 1 protein. According to the revised Ghent criteria, *FBN1* missense mutations affecting cysteine amino acids are considered pathogenic aberrations. Three variants (patients 4, 9, 11) have not been previously described.

5.1.3 Non-cysteine residue-affecting missense FBN1 variants

Almost a third of the missense mutations we detected (5/16, 31.2%) did not involve a cysteine amino acid, and in these cases we carefully studied the available literature and databases. Their classifications were performed according to the revised Ghent criteria and the American College of Medical Genetics and Genomics' (ACMG) recent recommendation for variant interpretation.

Several cases of c.640G>A, p.(Gly214Ser) variant have been described by different working groups. In patients with Marfan syndrome, ectopia lentis syndrome (ELS) and sporadic aortic dissection. The abnormality was co-segregated with phenotypic symptoms and was absent in healthy individuals, in a large control population, and was not included in either the gnomAD or 1000 Genomes Project population databases. Glycine at position 214 is conserved among species and is also likely to be classified as a pathogenic variant by in silico prediction software. Our 10-year-old patient 12 also had ectopia lentis and systemic involvement without aortic root dilatation (diagnosis: possible MFS).

The previously undescribed missense variant c.3038G>T, p.(Gly1013Val) was detected in patient 13. It affects an amino acid position where another missense variant c.3037G>A, p.(Gly1013Arg) has been previously described. Evolutionarily conserved amino acid position 1013. We also tested for the c.3038G>T variant in 50 control samples representing 100 alleles, which was not detected in any allele, and the variant was not present in either the gnomAD or the 1000 Genomes Project population database, and thus is predicted to be a potentially pathogenic mutation. Our patient presented with a severe atypical form of MFS with neonatal onset of symptoms, similar to cases previously described in the literature. Based on the available data, we classified the variant as probably pathogenic. In the absence of samples, we were unable to perform a parental analysis, but if the origin of the abnormality is proven de novo, the evidence suggests that it can be reclassified as a pathogenic abnormality.

The c.3838 G>A, p.(Asp1280Asn) variant was not present in either the gnomAD or the 1000 Genomes Project population database. It was reported once in the ClinVar database with ClinVar ID:200025 as a probable pathogenic variant. It was also found in the locus specific database (http://www.umd.be/FBN1/), where the variant was reported in a patient with Marfan syndrome. Aspartic acid 1280 is evolutionarily conserved between species and is part of a conserved cbEGF consensus sequence (X-D-X-(N/D)-E-C-X(6)-C-X(4)-C-X-N*-X(2)-G-X-(Y/F)-X-C-X-C-X(2)-G-X(9)-C).

The aspartic acid (D) concerned is highlighted in bold in the sequence. Based on the criteria for pathogenic mutations, this variant was judged to be a pathogenic missense mutation.

The c.4727T>C, p.(Met1576Thr) variant has been previously described in association with Marfan syndrome (aortic dilatation was described in both cases) and was absent from 400 control chromosomes. The variant was also described in three patients with adolescent idiopathic scoliosis. In the latter study, Western blot analysis of paraspinal muscle samples from three patients showed elevated levels of phosphorylated SMAD2 (pSMAD2), providing

indirect evidence of increased TGF-β signaling in MFS. The variant occurred with very low frequency in the GnomAD database (34/282694) and was not present in the 1000 Genomes Project database. Although mentioned as an MFS associated mutation in the literature, it is predicted as a benign variant by prediction software (PolyPhen2, SIFT, Mutation Taster). Based on the ambiguous data, the variant was classified as a variant of uncertain significance (VUS).

The c.5582 G>A, p.(Ser1861Asn) missense variant was included in the ClinVar database (ClinVar ID:406289) as a variant of uncertain clinical significance (VUS) associated with MFS. It was present in the GnomAD database with a very low frequency (4/245638).
The serine amino acid at position 1861 is not fully conserved between species. Considering the available data, the variant was classified as a variant of uncertain significance (VUS).

5.1.4 Small-scale deletions/duplications in the FBN1 gene

Small-scale deletions/duplications in the FBN1 gene were detected in 3 of our 23 patients (13%). All of these variants lead to reading frame shifts and likely truncated protein or mRNA degradation by nonsense mutation-mediated mRNA degradation (NMD), and thus all of these variants were classified as pathogenic. Two variants were novel, whereas one was previously described in an 18-year-old patient with Marfan syndrome who had ascending aortic dilatation and mild systemic involvement. Our patient 19 (44 years old) had similar ascending aortic dilatation with <7 systemic involvement.

5.1.5 Slice site variants in the *FBN1* gene

A total of 4 splice site variants (4/23, 17.4%) were detected in the gene. Two splice variants were the previously described variants c.1468+5G>A and c.4337-2A>G. The c.1468+5G>A mutation was first described in 1997. The authors demonstrated by RT-PCR that this variant caused exon skipping in total RNA samples isolated from fibroblasts. The c.4337-2A>>G variant was described in a female patient who died following postpartum aortic dissection. The c.1837+4C>T variant has been detected in one patient and is included in the ClinVar database (ClinVar ID:406295) and occurs with a very low frequency (9/246104) in the gnomAD database. It is a pathogenic variant based on predictions (Human Splicing Finder, Mutation Taster) as it potentially interferes with the splicing process. Based on available data, the variant was classified as a variant of uncertain significance (VUS).

The previously undescribed variant c.3082+1G>A was detected in a 10-month-old boy with ectopia lentis and mild systemic involvement. Based on the predictions (Human Splicing Finder, Mutation Taster), it is a pathological abnormality. The variant is located in intron 24

and affects the canonical +1 donor splice site, which mostly disrupts normal splicing (possibly causing exon skipping). For this reason, the variant was defined as a pathogenic variant.

5.1.6 Recurrent silent variant in three unrelated patients

The same silent variant c.3294C>T, p.(Asp1098=) was detected in three unrelated patients with only mild systemic phenotypic symptoms. The variant is listed in the HGMD database as well as in the Universal Mutation Database (http://www.umd.be/FBN1/). Two personal communications (Ads, Holman, 2000; Boileau, 2013) and two publications are available in relation to the phenotype. Sheikhzadeh et al. summarized that there is no comprehensive evidence to judge the pathogenicity of the variant, although a splicing defect cannot be excluded based on software predictions. The minor allele frequency (MAF) of the c.3294C>T, p.(Asp1098=) variant is less than 1% in the GnomAD and 1000 Genomes Project databases. Human Splicing Finder (HSF) prediction software predicted the potential change in the splicing site.

Transcript analysis was performed on RNA samples isolated from skin fibroblasts from patient 26. We also performed bidirectional Sanger sequencing of exon 26 and exon-exon boundaries, which showed that the mutant allele was present in approximately 50% of the patient's cDNA sample and that the exon-exon boundaries were intact. For patient 26, an accurate allele ratio determination was performed by bidirectional pyrosequencing with high coverage on both genomic DNA and cDNA samples. The latter analysis showed that wild-type and mutant alleles were equally distributed in the patient's cDNA sample, similar to the genomic DNA sample. In conclusion, neither NMD nor exon skipping was observed. The variant was classified as a benign variant on the basis of the following: the allele frequency of the variant is higher than would be expected from the disease, no abnormal splicing was detected, and the suspected disease (MFS, other fibrillinopathies) is known to have genetic (locus) heterogeneity, i.e. the symptoms may be caused by a variant in other genes.

5.2 Mutation spectrum of familial hypercholesterolaemia patients in Hungary

5.2.1 Detected variants

A total of 47 mutations associated with familial hypercholesterolaemia were detected in 44 patients, of which 3 patients carried two variants, meaning they were of complex heterozygous genotype. The vast majority of the mutations detected (40 mutations) were located in the *LDLR* gene, and in 7 cases we detected a mismatch in the *APOB* gene. No pathogenic mutations associated with FH were detected in the *PCSK9*, *LDLRAP1*, *STAP1* genes.

In terms of the distribution of mutations within the *LDLR* gene, 1 mutation was found in the promoter of the *LDLR* gene, 17 mutations in the ligand binding domain, 18 in the epidermal growth factor (EGF) precursor homology domain, 1 in the O-linked sugar domain, and 3 mutations in the cytoplasmic domain.

5.2.2 Variants detected in the *APOB* gene

During the study, *APOB* mutations were identified in 7 patients, of which the most common *APOB* mutation causing c.10580G>A p.(Arg3527Gln) FH was identified in 4 patients.
We detected two missense mutations (c.8213T>A, p.(Ile2738Lys), c.10438A>G, p.(Lys3480Glu)) in 2 probands whose pathogenicity is not clear, but neither of these variants is included in the population databases and therefore not present in the normal population, and no other variants in the genes tested were detected in any of the patients. The mutations are classified as variants of uncertain clinical significance (VUS) according to ACMG, so further studies are needed to assess the pathogenicity of the mutations. In one additional patient, a mutation causing reading frame shift c.13242delG in exon 29 of the *APOB* gene was detected in heterozygous form. The deletion results in the insertion of a premature stop codon at amino acid position 4415, which, in turn, results in a possibly truncated protein. Since the mutation is located in the last exon, it is unlikely to result in nonsense-mediated mRNA degradation, but rather it is likely that the truncated protein will lack 148 amino acids at the C-terminal. The mutation is not listed in population databases, but is mentioned once in the ClinVar database as a variant of uncertain significance detected in a clinical trial (ClinVar ID:927203). Overall, the classification by prediction software and the fact that a pathogenic truncating mutation was described 3' away from the detected variant, which also supports the pathogenicity of the mutation, led us to classify the variant as VUS/probably pathogenic.

5.2.3 Variants detected in the *LDLR* gene

A total of 40 *LDLR* variants were detected in 37 patients. In three patients, a complex heterozygous genotype was detected, which explained the patients' more severe clinical symptoms. A complex heterozygous genotype causes clinical symptoms as severe as a homozygous genotype, in both cases resulting in loss of function of both alleles.

5.2.3.1 Patients carrying a missense *LDLR* variant

We detected several missense mutations in the *LDLR* gene that were present in more than one patient. The most frequently occurring variant was c.1618G>A, p.(Ala540Thr) in four families, while c.1130G>A, p.(Cys377Tyr) was detected in three families and c.2054C>T, p.(Pro685Leu) was also detected in three families. Variants c.662A>G, p.(Asp221Gly), c.1775G>A, p.(Gly592Glu), c.2000G>A, p.(Cys667Tyr) and c.1865A>G, p.(Asp622Gly) were all found in two families. The other missense mutations were unique.

5.2.3.2 Patients carrying a nonsense *LDLR* variant

The nonsense mutations c.82G>T, p.(Glu28*) and c.337G>T, p.(Glu113*) were detected in two families, while the c.1048C>T, p.(Arg350*) variant was detected in one family in heterozygous form. In patient #13, the c.337G>T, p.(Glu113*) nonsense variant was detected in a compound heterozygous form together with another pathogenic missense mutation.

5.2.3.3 Patients with *LDLR* variant affecting splice site

Splicing defect causing variant was detected in a total of 8 patients. Two of these were present in more than one proband. The c.940_940+14del variant was also detected in three patients (patients 8,11,23). This mutation is likely to affect the splicing process and therefore impair protein function. The c.694+2 T>C variant was found in two families. In one patient, a previously unreported splicing variant c.1706-2A>G variant in the *LDLR* gene was also detected.

5.2.3.4 Small deletions/duplications in the *LDLR* gene detected in patients with FH

The c.2416dupG, p.(Val806Glyfs*11) variant was detected in two families in heterozygous form. This variant is a 1 bp insertion in exon 17 of the *LDLR* gene, a homopolymeric region of 5 guanines. The insertion results in 6 instead of 5 guanines in the homopolymer stretch, which shifts the reading frame and results in premature stop codon incorporation. This mutation belongs to class III of *LDLR* mutations, i.e. it results in a defective or deficient LDL receptor protein. In patient 28, we detected the c.2167delG, p.(Glu723Argfs*7) variant in heterozygous form. This small deletion in exon 15 also disrupts the normal reading frame and leads to premature stop codon incorporation. In this patient, we also identified an additional heterozygous missense mutation (c.862G>A, p.(Glu288Lys)), and this patient is discussed in detail in the section on patients carrying compound heterozygous *LDLR* mutations. In only one

of the patients studied, a deletion not affecting the reading frame was detected (c.654_656delTGG, p.(Gly219del)).

5.2.3.5 Patients with compound heterozygous *LDLR* mutations

In this study, there were 3 patients (P12, P13, P28) who had a compound heterozygous genotype. In all three cases, the genotype was associated with a severe clinical picture.

In patient 12, in addition to a heterozygous missense mutation (c.1618G>A, p.(Ala540Thr)), a deletion of exon 7 of the *LDLR* gene (c.941-190_c.1061-270del) was detected in heterozygous form. A targeted cascade analysis was performed, which demonstrated trans inheritance of the two mutations. We also successfully determined the breakpoints of the exon 7 deletion in the *LDLR* gene. A homologous DNA sequence of 32bp in length was found both upstream and downstream of exon 7 of the wild type *LDLR* gene. In the case of the patient, this sequence was detected only once after sequencing the PCR product, and exon 7 of the gene was also deleted. Our study suggests that deletion of this short intronic sequence may explain the deletion of the entire exon 7.

In patient 13 (P13), we detected one nonsense (c.337G>T, p.(Glu113*)) and one missense mutation (c.1618G>A, p.(Ala540Thr)) in heterozygous form. Cascade analysis showed that the nonsense mutation was inherited from his mother. A paternal DNA sample was not available as the patient's father had previously died due to cardiovascular complications.

In patient 28, we detected one missense (c.862G>A, p.(Glu288Lys)) and one frameshift mutation (c.2167delG, p.(Glu723Argfs*7)) in heterozygous form.

The c.2167delG, p.(Glu723Argfs*7) variant has been previously described in a patient with elevated total cholesterol.

In our patient, elevated LDL-C and TC were also measured. Although xanthoma was not described due to the young age of the patient (13 years), it may develop with advancing age. DNA samples from the family members of patient 28 were not available so cascade analysis could not be performed and thus trans inheritance could not be confirmed.

5.2.3.6 Patients with *LDLR* copy number variation

In three patients, copy number variation in the *LDLR* gene was detected in heterozygous form. In one patient described above (patient 12) a compound heterozygous form was found. In patient 38, a duplication of exons 4-8 as and in patient 33 a deletion of exons 3-8 was detected in heterozygous form based on next-generation sequencing coverage data. The detected variants were also confirmed by MLPA.

5.2.4 Family members of patients with *LDLR* variants

For all patients where a pathogenic or likely pathogenic variant was detected, a targeted analysis of the available family member sample was performed. A total of 15 family members were found to have a variant, of which 1 case involved the *APOB* gene and 14 cases involved the *LDLR* gene.

6. Discussion

6.1 Summary of variants in Marfan syndrome patients

In this study, we identified 23 FBN1 variants in patients with Marfan syndrome or other connective tissue disorders. The Ghent nosology shows excellent specificity for identifying MFS patients with FBN1 mutations and defines criteria for pathogenic FBN1 mutations, which greatly facilitates the assessment of the pathogenicity of new genetic variants found in approximately 25-30% of MFS cases.

In patients who meet the Ghent criteria, mutation detection rates of up to 97% can be achieved. However, the clinical diagnosis of MFS is not always straightforward due to: i) the large number of diseases that can present with overlapping cardiovascular, skeletal and ocular symptoms with Marfan syndrome, ii) the age-dependent onset of some clinical manifestations, iii) intra- and inter- and inter-observer variability in terms of age at onset, tissue involvement and severity of symptoms. FBN1 mutations are most commonly associated with Marfan syndrome, but variants have also been described in a number of Marfan-like phenotypes, although mutation detection rates are much lower in other mild fibrillinopathies. For the reasons mentioned above, testing of the FBN1 gene was warranted for a number of clinical situations that were representative of our patient population. In total, we detected 23 pathogenic or potentially pathogenic variants in the FBN1 gene, including three variants of uncertain significance, of which approximately 30% were previously undetected novel variants. Two-thirds of the mutations were missense variants, the majority of which were located in one of the cbEGF patterns and about 70% of which involved a conserved cysteine residue, in accordance with the literature. The cysteine amino acids in the FBN1 gene are highly conserved between species, showing their essential role. It is also known that mutations affecting cysteine are associated with a higher chance of ectopic lentis compared to missense mutations affecting other amino acids. We also found a much higher proportion (11/9) of ectopia lentis in cases where a cysteine-associated missense mutation was detected, compared to other abnormalities (12/2).

In five cases we identified non-cysteine missense variants. This group of variants illustrates the difficulties in interpreting genetic results. The c.4727T>C, p.(Met1576Thr) variant has been previously described in association with Marfan syndrome and aortic root dilatation, but has also been reported in adolescent idiopathic scoliosis. The proband (14 years old) and his sister (16 years old) inherited the variant from their father (no clinical data were available for the father), both of whom had scoliosis and mild systemic involvement (flat feet, thumb symptom, joint laxity). Obviously, regular cardiological follow-up is recommended, but it is questionable how the outcome will affect the lifestyle of the sister who is playing contact sports.

FBN1 mutations have been described in the entire coding region of the gene. Mutations in exons 24-32 have been associated with the most severe and rapid phenotypic progression. In two of our patients under 1 year of age, we identified one missense (c.3038G>T, p.(Gly1013Val), exon 24) and one slice site (c.3082+1G>A, intron 24) mutation in the FBN1 gene (patients 13 and 22). In patient 22, the missense was described as a de novo missense, as no familial variant was detected when the parents were tested. In patient 13, the father had a marfanoid phenotype but did not agree with the genetic test. The p.(Gly1013Arg) variant affecting the same amino acid position is a relatively common recurrent variant described in association with atypical severe cardiovascular, skeletal manifestations and ectopia lentis. The amino acid glycine 1013 is highly conserved and appears to be a key residue at this position, located in the interdomain binding regions of TGFβ1bp #03 and cbEGF #11 motifs, indicating a key role for this region. The glycine 1013 position could also be considered as a mutational hotspot in patients with early onset and severe phenotypes.

The c.3082+1G>>A splicing affecting variant in intron 24 affects the donor splicing site. Another substitution (c.3082+1G>T) has been previously described at this position in a patient with classical Marfan syndrome. A likely consequence of this mutation is the skipping of exon 24 during transcription. Both of our patients were at the severe end of the MFS symptom spectrum, with early onset of symptoms consistent with the literature. Splice site variants can occasionally result in truncated protein and are therefore similar in effect to nonsense or frameshift mutations, whereas other splice site variants cause deletions without frameshift. In a study of 197 patients, it was found that almost all patients who had an early aortic event had a splicing or truncating mutation. This finding has important implications for patients who carry these types of abnormalities (in our patient group, the overall proportion of these patients was 30.4%). It is known that exonic silent mutations in the FBN1 gene can cause exon skipping. In our study, we investigated the effect of a recurrent silent mutation (c.3294C>T) on mRNA

splicing. FBN1 transcript analysis was performed from total RNA isolated from a skin biopsy of a corresponding cell type (fibroblast).

The c.3294C>T variant has been previously described in the literature in patients with symptoms suggestive of connective tissue disease but not fulfilling the criteria for Marfan syndrome. For the c.3294C>T variant, exon skipping could not be confirmed and exon-exon boundaries were intact in mRNA studies. The ratio of mutant to wild-type alleles was close to 1:1 in both the patient's genomic and cDNA samples, which were analyzed by pyrosequencing. This recurrent variant was also detected in 3 unrelated patients. The patients had systemic phenotypic symptoms as described in the literature, without ectopia lentis and aortic root dilatation. The mild, previously unknown effect of this variant may be consistent with this mild phenotype.

In conclusion, we have described clinical and molecular data from 26 unrelated patients with Marfan syndrome or other connective tissue diseases. We identified 7 previously undescribed and 16 recurrent pathogenic or probable pathogenic variants in 23 patients. Most of the FBN1 variants were associated with Marfan syndrome, two of them with an early and severe form of the disease.

6.2 Summary of variants found in FH patients

Clinically diagnosed patients with familial hypercholesterolaemia were tested using a next-generation sequencing method that can detect not only small-scale mutations, but also large-scale duplications and deletions, with appropriate bioinformatic analysis.

Except for one variant, 39 *LDLR* mutations have been previously described in the literature. Although the deletion of exon 7 has been previously described, we also determined the breakpoints, which has not been done before.

A total of 59 individuals were found to have alterations associated with familial hypercholesterolaemia during the study of patients. Forty-four probands and fifteen family members carried the ACMG criteria pathogenic/probably pathogenic, while in two cases we were able to classify the detected abnormality as a variant of uncertain significance. In terms of the distribution of *LDLR* mutations, most of the variants were located in the ligand-binding domain and the EGF-like domain.

To the best of our knowledge, no previous study has investigated the genetic background of familial hypercholesterolemic patients in Hungary. Regarding the mutation spectrum, Hungarian data showed a large overlap with data from neighbouring countries such as Poland

and the Czech Republic. The most common *LDLR* variants in the Polish population were also found in our patient population.

The 1618G>A, p.(Ala540Thr) missense mutation was detected in 4 unrelated individuals (4/37, 10.8%) and in 2 additional family members during cascade testing. This pathogenic variant occurred in several European and Latin American populations and was a common variant among the patients we studied. The phenotype reported in the literature and in our own patients was very similar. This variant is located in the epidermal growth factor (EGF) precursor homologous domain. Its pathogenicity is supported by the fact that it has been previously described in many cases in the literature and that patients carrying the homozygous form of the variant had severe clinical symptoms such as premature coronary artery disease, tendon xanthomas and high total cholesterol.

The c.1130G>A, p.(Cys377Tyr) missense mutation was detected in 3 families. The first study in Sweden described the variant. Pathogenicity is supported by the fact that since then, further European population studies have described this variant in patients with FH.

Another missense mutation c.2054C>T, p.(Pro685Leu) variant was also detected in 3 families. According to a comprehensive study describing more than 2000 FH-causing mutations, this variant is considered pathogenic. It should be noted that in this study, 7 mutations were found that have been described in several countries on all continents, and one of them is the c.2054C>T, p.(Pro685Leu) missense mutation. This fact suggests that the variant may be a very old founder mutation or a frequently recurring variant.

In two patients we detected the c.1775G>A, p.(Gly592Glu) missense variant in heterozygous form. This mutation has been previously described in several populations, and a previous study reported it as a founder mutation in a Greek region, where it was present in almost one third of the patients studied, and was detected in several cases in homozygous form. Based on literature data, this class V LDLR variant does not cause a drastic increase in cholesterol levels in the homozygous form, and it can be concluded that the variant only slightly reduces LDLR activity, which may have contributed to the spread of the mutation.

The nonsense mutations we detected have been previously described as a pathogenic FH-causing variant. These mutations all result in a truncated, non-functional form of the LDLR protein. The uptake of LDL particles is reduced in the liver and other tissues.

The c.-153C>T promoter mutation has been previously detected in a study of Czech FH patients. This variant involves a conserved sequence called sterol-dependent regulatory element 2 (SRE2) in the promoter region of the *LDLR* gene. Using a luciferase-based functional assay, they concluded that the mismatch could significantly reduce the activity of the *LDLR* promoter.

The c.940_940+14del mismatch was previously described in a study of a Scandinavian population, where it was co-segregated with disease based on family history. In that study, both the clinical picture and family history confirmed pathogenicity, with early cardiovascular disease described in patients carrying the mutation. LDL-C levels were also very high in our patients before treatment (7.9-10.5 mmol/L).

The c.694+2T>C variant was previously described as an Icelandic founder mutation, with origins traced back to the 18th century. The variant affects a conserved spiking locus and its pathogenicity has been demonstrated by segregation studies.

The pathogenicity of the previously undescribed c.1706-2A>G variant detected by us is supported by the fact that other nucleotide substitutions have been described at this position (c.1706-2A>C; c.1706-2A>T). c.1706-2 A>T nucleotide substitution has been detected in 2 Arab families and further studies have shown that the adenine-thymine substitution at this nucleotide position is pathogenic, as greatly reduced *LDLR* mRNA levels were measured in patients and the mismatch involves a conserved splicing site.

The c.2416dupG, p.(Val806fs*11) frameshift mutation was described in a Swedish population study and has been found in Pakistani and other population studies. The proband patient and 8 other family members had elevated TC values (7.8-12.9 mmol/L) typical of heterozygous FH patients, while the proband's 5-year-old child had elevated TC values (23.2 mmol/L) typical of homozygotes and typical xanthomas. Patient 18 also had markedly elevated TC (8.1 mmol/L), coronary artery disease and xanthamas, while patient 44 had elevated TC (10 mmol/L).

The c.654_656delTGG, p.(Gly219del) variant was previously identified in an Ashkenazi Jewish population as a founder mutation causing FH. This variant represents 21% of the pathogenic mutations detected in patients studied there. The mutation results in the deletion of a glycine amino acid at position 219, located in a highly conserved ligand-binding domain.

The c.337G>T, p.(Glu113*) variant is located in exon 4 of the *LDLR* gene and is classified as a class 1 pathogenic variant resulting in a null allele. It was first described by Hobbs and colleagues in 1992.

The 1618G>A, p.(Ala540Thr) variant is also a well known pathogenic variant that has been detected in several studies in different populations such as Spanish, German, Italian or Greek.

The c.862G>A, p.(Glu288Lys) missense mutation was first described in a German patient in 1999. In Italian patients, significantly elevated LDL-C values and xanthomas in severe cases were observed as a phenotypic consequence of the mutation.

The most frequent variant in the *APOB* gene in the Hungarian population was also the c.10580G>A, p.(Arg3527Gln) mutation, which is in agreement with the literature data. By

testing the entire *APOB* gene, we were able to detect 3 additional variants, two of which are missense variants not previously described. Both missense variants were classified as variants of uncertain clinical significance. The third was a mutation with a reading frame shift (c.13242delG, p.(Leu4415*)), which was previously described once in the ClinVar database as an FH-associated variant. This result suggests that sequencing the entire *APOB* gene rather than targeted *APOB* testing can greatly improve diagnostic efficiency.

Our study revealed a new mutation in the *LDLR* gene, while most of the mutations in the Hungarian population have been described previously. The proportion of mutation types and their localisation within the gene is also broadly similar to other population studies. The importance of establishing a genetic diagnosis is indicated by the fact that the timing of initiation of lipid-lowering therapy can greatly influence the development and severity of expected complications.

A large study that followed children for 20 years concluded that early treatment greatly delayed the risk of adult cardiovascular complications and the severity of expected symptoms. It is important to know whether the patient has a heterozygous, compound heterozygous or homozygous genotype before appropriate treatment can be started. In addition, cascade testing can identify affected but asymptomatic family members, which is important for prevention and appropriate, timely treatment.

7. Summary

In the studies presented here, we investigated the applicability and efficiency of different next-generation sequencing methods in two monogenic diseases. For both diseases, molecular genetic testing methods have undergone a complete transformation in the last 10-15 years. Previously, genetic testing for mutations in known variants involved targeted testing, such as restriction digestion, whereas for unknown nucleotide-level variants, the only option was the costly and time-consuming Sanger DNA sequencing. This problem has been solved by the advent of a next-generation sequencing technique that has burst onto the molecular diagnostics market. The size of a gene was no longer a reason for exclusion, nor was the need to test several genes in the background of a disease in search of a pathogenic variant.

In the first study, we described genotype-phenotype correlations in 26 unrelated individuals with a guideline diagnosis of Marfan syndrome or related connective tissue disease. Two different next-generation sequencing platforms and three DNA library preparation methods were used in the study. In total, 23 patients were identified as having a pathogenic or potentially pathogenic variant in the FBN1 gene, of which 16 were recurrent variants and 7 variants had

not been previously described. We compared our detected variants with literature data. In the majority of cases, FBN1 variants were associated with Marfan syndrome, and in 2 cases with a severe early-onset (neonatal) form of the syndrome. In Marfan syndrome, the identification of pathogenic variants may be of great importance, as it is crucial to identify the exact genotype-phenotype correlation because of the expected consequences. With knowledge of the pathogenic variant, the literature can provide a good guide to prevent or delay the expected complications. In many cases, close cardiology, ophthalmology and orthopaedic follow-up of patients is recommended. Genetic diagnosis also has an important role to play in potential surgical interventions because of severe aortic involvement.

In the second study we tried to establish the genetic background of familial hypercholesterolaemia patients in Hungary. Since pathogenic variants in several genes are involved in the disease, a kit was used to analyse the samples for genes associated with the disease. In total, 44 patients were found to have pathogenic or potentially pathogenic variants. Seven patients carried an alteration in the *APOB* gene and 37 patients in the *LDLR* gene. Three patients had a complex heterozygous genotype, which explained the more severe clinical symptoms. The main clinical parameters of the patients and the genotype-phenotype correlation are described. We detected two or two *LDLR* and *APOB* variants that have not been previously described. Described and undescribed variants in the background of familial hypercholesterolemia were compared with literature data. We also determined the exact breakpoints of a 1 exon deletion.

Early diagnosis of familial hypercholesterolaemia is extremely important. The disease is still very under-diagnosed and is detected late or not at all due to fatal cardiovascular complications. Timely initiation of lipid-lowering treatment of FH detected early in life (especially in children), possibly targeted drug therapy, diet and close follow-up can all contribute to patients living a good quality of life with a reduced cardiovascular risk. It is also necessary to emphasise the importance of genetic counselling, as education and screening of affected family members is of paramount importance because of the potential complications.

8. New findings of the thesis

1) Using different types of next-generation DNA sequencing platforms, we determined the genetic background of Marfan syndrome or other related connective tissue disorders in 23 patients. In doing so, we described 7 cases of FBN1 abnormalities not previously reported in the literature. Experimental work was performed to investigate the pathogenicity of a recurrent silent variant using genomic DNA and RNA isolated from fibroblasts in cDNA samples, which

led to the final classification of the variant as a benign variant. Due to the size and number of genes underlying Marfan syndrome, the genetic background of Hungarian patients had not been previously mapped and thus no information on the variants and their expected consequences was available. The detected variants were compared with literature data.

2) We investigated the genetic background of patients with familial hypercholesterolaemia in Hungary. Previously, no data were available on the genetic background of the Hungarian FH population, so this study provided important results comparable with literature data and other population studies. The identification of the genetic background of the disease will also help clinical decision making regarding potential therapies and expected phenotypic symptoms. New variants that may be responsible for familial hypercholesterolaemia have been described. In the *APOB* gene, two missense variants (c.8213T>A, p.(Ile2738Lys)) and (c.10438A>G, p.(Lys3480Glu)) were identified, while in the *LDLR* gene, a novel splicing variant (c.1706-2A>G) and a 1 exonic deletion (c.941-190_c.1061-270del) were identified.

3) In both studies, it became clear that the former traditional Sanger sequencing method will be replaced by various next-generation sequencing devices. The high throughput and cost-effective operation has allowed routine investigation of the genetic background of previously unstudied disorders that were previously too costly and extremely time-consuming to investigate due to the size and/or number of genes.

VIII. Publications

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/93/2023.PL
Subject: PhD Publication List

Candidate: László Madar
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology
MTMT ID: 10060418

List of publications related to the dissertation

1. **Madar, L.**, Juhász, L., Szűcs, Z., Kerkovits, L., Harangi, M., Balogh, I.: Establishing the Mutational Spectrum of Hungarian Patients with Familial Hypercholesterolemia.
Genes. 13 (1), 1-13, 2022.
DOI: http://dx.doi.org/10.3390/genes13010153
IF: 4.141 (2021)

2. **Madar, L.**, Szakszon, K., Pfliegler, G., P. Szabó, G., Brúgós, B., Ronen, N., Papp, J., Zahuczky, K., Szakos, E., Fekete, G., Oláh, É., Koczok, K., Balogh, I.: FBN1 gene mutations in 26 Hungarian patients with suspected Marfan syndrome or related fibrillinopathies.
J. Biotechnol. 301, 105-111, 2019.
DOI: http://dx.doi.org/10.1016/j.jbiotec.2019.05.012
IF: 3.503

List of other publications

3. Deák, A., Koczok, K., Bessenyei, B., Szűcs, Z., **Madar, L.**, Csorba, G. É., Orosz, O., Laki, I., Halász, A., Marsal, G., Balogh, I.: A magyar Cystás Fibrosis Regiszter genetikai revíziója.
Orv. hetil. 163 (51), 2052-2059, 2022.
DOI: http://dx.doi.org/10.1556/650.2022.32655
IF: 0.707 (2021)

4. Szabó, T. M., Balogh, I., Ujfalusi, A., Szűcs, Z., **Madar, L.**, Koczok, K., Bessenyei, B., Csürke, I., Szakszon, K.: Helsmoortel-Van der Aa Syndrome: Cardiothoracic and Ectodermal Manifestations in Two Patients as Further Support of a Previous Observation on Phenotypic Overlap with RASopathies.
Genes. 13 (12), 1-13, 2022.
DOI: http://dx.doi.org/10.3390/genes13122367
IF: 4.141 (2021)

5. Gaál, Z., Szűcs, Z., Kántor, I., Luczay, A., Tóth, H. P., Benn, O., Felszeghy, E. N., Karádi, Z., **Madar, L.**, Balogh, I.: A Comprehensive Analysis of Hungarian MODY Patients-Part I: Gene Panel Sequencing Reveals Pathogenic Mutations in HNF1A, HNF1B, HNF4A, ABCC8 and INS Genes.
Life (Basel). 11 (755), 1-13, 2021.
DOI: http://dx.doi.org/10.3390/life11080755
IF: 3.251

6. Gaál, Z., Szűcs, Z., Kántor, I., Luczay, A., Tóth, H. P., Benn, O., Felszeghy, E. N., Karádi, Z., **Madar, L.**, Balogh, I.: A Comprehensive Analysis of Hungarian MODY Patients-Part II: glucokinase MODY Is the Most Prevalent Subtype Responsible for about 70% of Confirmed Cases.
Life (Basel). 11 (8), 1-20, 2021.
DOI: http://dx.doi.org/10.3390/life11080771
IF: 3.251

7. Mokánszki, A., Chang Chien, Y. C., Mótyán, J. A., Juhász, P., Bádon, E. S., **Madar, L.**, Szegedi, I., Kiss, C., Méhes, G.: Novel RB1 and MET Gene Mutations in a Case with Bilateral Retinoblastoma Followed by Multiple Metastatic Osteosarcoma.
Diagnostics. 11 (1), 1-11, 2021.
DOI: http://dx.doi.org/10.3390/diagnostics11010028
IF: 3.992

8. Juhász, L., Balogh, I., **Madar, L.**, Kovács, B., Harangi, M.: A Rare Double Heterozygous Mutation in Low-Density Lipoprotein Receptor and Apolipoprotein B-100 Genes in a Severely Affected Familial Hypercholesterolaemia Patient.
Cureus. 12 (12), 1-5, 2020.
DOI: http://dx.doi.org/10.7759/cureus.12184

9. Mezei, Z. A., Tornai, D., Földesi, R., **Madar, L.**, Sümegi, A., Papp, M., Antal-Szalmás, P.: A DNA pool of FLT3-ITD positive DNA samples can be used efficiently for analytical evaluation of NGS-based FLT3-ITD quantitation: testing several different ITD sequences and rates, simultaneously.
J. Biotechnol. 303, 25-29, 2019.
DOI: https://doi.org/10.1016/j.jbiotec.2019.06.305
IF: 3.503

10. Ivády, G., **Madar, L.**, Dzsudzsák, E., Koczok, K., Kappelmayer, J., Krulisova, V., Macek, J. M., Horváth, A., Balogh, I.: Analytical parameters and validation of homopolymer detection in a pyrosequencing-based next generation sequencing system.
BMC Genomics. 19, 1-8, 2018.
DOI: http://dx.doi.org/10.1186/s12864-018-4544-x
IF: 3.501

11. Koczok, K., Merő, G., P. Szabó, G., **Madar, L.**, Gombos, É., Ajzner, É., Mótyán, J. A., Hortobágyi, T., Balogh, I.: A novel point mutation affecting Asn76 of dystrophin protein leads to dystrophinopathy.
Neuromusc. Disord. 28 (2), 129-136, 2018.
IF: 2.612

12. Szabó, T., Orosz, P., Balogh, E., Jávorszky, E., Máttyus, I., Bereczki, C., Maróti, Z., Kalmár, T., Szabó, A., Reusz, G., Várkonyi, I., Marián, E., Gombos, É., Orosz, O., **Madar, L.**, Balla, G., Kappelmayer, J., Tory, K., Balogh, I.: Comprehensive genetic testing in children with a clinical diagnosis of ARPKD identifies phenocopies.
Pediatr. Nephrol. 33 (10), 1713-1721, 2018.
DOI: http://dx.doi.org/10.1007/s00467-018-3992-5
IF: 2.816

13. Koczok, K., Gombos, É., **Madar, L.**, Török, O., Balogh, I.: Interfering effect of maternal cell contamination on invasive prenatal molecular genetic testing.
Prenat. Diagn. 38 (9), 713-719, 2018.
DOI: http://dx.doi.org/10.1002/pd.5319
IF: 2.434

14. Molnár, Z., Balogh, L., Kappelmayer, J., **Madar, L.**, Gombos, É., Balogh, I.: Congenital hyperinsulinism caused by a de novo mutation in the ABCC8 gene: a case report.
eJIFCC. 8 (28), 85-91, 2017.

15. Zádori, D., Szpisjak, L., **Madar, L.**, Varga, V. E., Csányi, B., Bencsik, K., Balogh, I., Harangi, M., Kereszty, É., Vécsei, L., Klivényi, P.: Different phenotypes in identical twins with cerebrotendinous xanthomatosis: case series.
Neurol. Sci. 38 (3), 481-483, 2017.
DOI: http://dx.doi.org/10.1007/s10072-016-2776-6
IF: 2.285

16. László, B., Ferenczi, A., **Madar, L.**, Gyöngyösi, E., Szalmás, A., Szakács, L., Veress, G., Kónya, J.: CpG methylation in human papillomavirus (HPV) type 31 long control region (LCR) in cervical infections associated with cytological abnormalities.
Virus Genes. 52 (4), 552-555, 2016.
DOI: http://dx.doi.org/10.1007/s11262-016-1338-6
IF: 1.431

17. Ivády, G., Koczok, K., **Madar, L.**, Gombos, É., Tóth, I., Győri, K., Balogh, I.: Molecular Analysis of Cystic Fibrosis Patients in Hungary - an Update to the Mutational Spectrum.
J. Med. Biochem. 34, 1-6, 2015.
IF: 0.742

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

18. Varga, V. E., Katkó, M., Harangi, J., Balogh, I., Kapás, I., **Madar, L.**, Seres, I., Molnár, M. J., Paragh, G., Kovács, G. G., Harangi, M.: Egy ritka, veleszületett neurodegeneratív betegség: a cerebrotendinosus xanthomatosis laboratóriumi diagnosztikája.
Orv. Hetil. 155 (21), 811-816, 2014.
DOI: http://dx.doi.org/10.1556/OH.2014.29887

19. Ivády, G., **Madar, L.**, Nagy, B., Gönczi, F., Ajzner, É., Dzsudzsák, E., Dvorakova, L., Gombos, É., Kappelmayer, J., Macek, J. M., Balogh, I.: Distribution of CFTR mutations in Eastern Hungarians: relevance to genetic testing and to the introduction of newborn screening for cystic fibrosis?
J. Cyst. Fibros. 10 (3), 217-220, 2011.
DOI: http://dx.doi.org/10.1016/j.jcf.2010.12.009
IF: 3.19

Total IF of journals (all publications): 45,5
Total IF of journals (publications related to the dissertation): 7,644

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

31 March, 2023

9. Keywords

Sanger DNA sequencing, next generation DNA sequencing, pyrosequencing, MLPA, Illumina MiSeq, synthesis-based sequencing, Marfan syndrome, *FBN1*, familial hypercholesterolemia, *LDLR*, *APOB*.

10. Acknowledgements

I would like to thank all those who helped me to write my PhD thesis.

First of all, I would like to thank my supervisor, Professor István Balogh, Head of the Department of Clinical Genetics and the Head of the Department of Human Genetics, thanks to whom I chose genetics. As a leader, I am always encouraged to think independently, work and and continuous development. Thank you for supporting me in my professional career at all times.

I would like to thank Dr. Katalin Koczok, from whom I have received a lot of knowledge and a lot of help and advice during our more than 10 years of professional relationship.

Thanks to the former and current staff of the Molecular Genetics Department, Éva Kálmáncheyné Gombos, Gabriella Csorba, Zsuzsanna Szűcs, Orsolya Nagy, PhD.

I am grateful to Professor János Kappelmayer, Director of the Institute of Laboratory Medicine, for the opportunity to do my PhD in this institute. With his continuous support, he provided the necessary financial and material conditions for my scientific work.

I would like to thank my co-authors, especially Professor Mariann Harangi, for their valuable work and support in the preparation of the publications on which this thesis is based.

I thank my wife, my children, my parents, my grandparents and my friends for always supporting me and for allowing me to come this far.

The research work was supported by the GINOP-2.3.2-15-2016-00039 grant "Research on the pathogenesis of rare diseases, development of new diagnostic and therapeutic procedures". I would like to take this opportunity to thank all those who made my PhD thesis possible and supported me in my work.