MEDICAL UNIVERSITY OF SOFIA

DEPARTMENT OF MEDICINAL CHEMISTRY AND BIOCHEMISTRY

****

MICHAELA BOYKOVA MLADENOVA

MOLECULAR BASIS OF RARE DISEASES S DISORDERS IN THE SKELETAL AND CONNECTIVE TISSUE

ABSTRACT

On a dissertation

for awarding an educational and scientific degree

"DOCTOR"

Field of higher education:

4. Natural sciences, mathematics and informatics

Professional direction: 4.3. Biological Sciences

PhD program: Molecular Biology

SCIENTIFIC SUPERVISORS:

Acad. Vanyo Ivanov Mitev, dm, dbn

Prof. Albena Parvanova Todorova – Georgieva, dbn

****

Sofia, 2024

The dissertation contains 115 pages, 55 figures and 20 tables.126 literary sources and 15 electronic sources are cited.

The dissertation work has been discussed and directed for defense by the Department of Medicinal Chemistry and Biochemistry, Medical University - Sofia.

The research included in the dissertation was carried out in:

• Department of Medicinal Chemistry and Biochemistry, Medical University - Sofia

• Genetic Medical-Diagnostic Laboratory "Genika", Sofia

The current research was partially financed by the Medical Science Council of the Medical University - Sofia under contracts: No. D - 117/2018, PROJECT ent. No. 7676/21.11.2017

The dissertation was presented at a meeting of the departmental council of the Department of "Medical Chemistry and Biochemistry", Medical University - Sofia, on 29.09.2023 and scheduled for defense on 30.01.2024.

The official defense of the dissertation work will take place on 30.01.2024 at the Department of Medical Chemistry and Biochemistry, Sofia Medical University, 2 Zdrave Street, Sofia, before a scientific jury composed of:

1. Doc. dr Lyubomir Mitev, dm (reviewer)
2. doc. dr Mariya Dragneva, dbn (reviewer)
3. Akad. prof. Rumen Pankov, dbn
4. Doc. dr Svetoslav Dimov, dbn
5. Prof. Aleksei Savov, dbn

The defense materials are available at the Department of "Medical Chemistry and Biochemistry", Medical University - Sofia and are published on the website of MU-Sofia - Academic Staff Development - http://career.mu-sofia.bg.

TABLE OF CONTENTS

Abbreviations used 3

Summary 4

Summary 6

Introduction 8

Aim and Objectives 9

Materials and methods 10

Results and Discussion 12

Conclusion 48

Conclusions 52

Contributions 53

References 55

Publications and scientific events 56

ABBREVIATIONS USED

FGFR1 - Fibroblast Growth Factor Receptor 1

FGFR2 - Fibroblast Growth Factor Receptor 2

COL1A1 - Collagen Type I Alpha 1 Chain

COL1A2 - Collagen Type I Alpha 2 Chain

COL11A1 - Collagen Type XI Alpha 1 Chain

COL2A1 - Collagen Type II Alpha 1 Chain

PCR/ПВР - Polymerase Chain Reaction (Полимеразна верижна реакция)

MLPA - Multiplex Ligation-dependent Probe Amplification

ARRAY - CGH-Array comparative genomic hybridization

NGS - Next-Generation Sequencing

WES - Whole Exsome Segurncing

DNA - Deoxyribonucleic acid

РЕЗЮМЕ

Скелетните дисплазии и заболяванията с нарушения в съединителната тъкан (колагенози) са две големи групи хетерогенни вродени състояния, които са социално-значими, тъй като са редки заболявания с тежко изявена клиника, която се отразява на цялостния психо-социален и обществено-икономически статус на индивида и неговото семейство.

Скелетните дисплазии са нарушения в хрущялната/скелетната тъкан, асоциирани с аномалии във формата и размера на скелета, диспропорция на дългите кости, гръбначния стълб и главата [Aravidis, Christos et al.,2014]. Известни са над 450 такива състояния и въпреки че всяко едно от тях е рядко, общо те се срещат с честота 1:5000 живородени деца [Barbosa et al,2009]. Причините за скелетни дисплазии са толкова разнообразни, колкото е и броят на различните заболявания, но има и чести мутации в няколко специфични гена (напр. *FGFR3, FGFR2*).

Заболяванията с нарушения на съединителната тъкан се срещат с честота 1:20000 живородени деца и са свързани основно с нарушения на хрущялната тъкан, ставите, мускулите и растежа на костите, но могат да засягат и други органи и системи - включително очите, сърцето, белите дробове, бъбреците, стомашно-чревния тракт, кръвоносните съдове и др. [Uttarilli et al.,2019]. Понастоящем, повече от 200 гена са асоциирани със заболявания на съединителната тъкан и са разпределени в различни групи, според тежестта на клиничната изява [Nicol et al.,2019]. Обект на настоящия дисертационен труд бяха гените *COL1A1, COL1A2, COL2A1* и *COL11A1*, които се асоциират с тясна група наследствени заболявания - Osteogenesis imperfecta, Marshall syndrome и Stickler syndrome.

В хода на изпълнението на настоящата работа бяха въведени и оптимизирани методи за изследване на гени, свързани със скелетни дисплазии и с колагенози. Научната разработка беше фокусирана върху търсене на молекулни дефекти по дължината на гените *FGFR3, FGFR2, COL1A1, COL1A2, COL2A1, COL11A1*. Разработеният молекулярно-генетичен подход беше приложен за изследване на 75 пациенти - при 66 от случаите анализите са проведени постнатално, а при 9 е извършена дородова диагностика. На базата на предварително проведените клинични изследвания пациентите бяха разпределени в 3 групи: 1) пациенти, насочени с диагноза скелетна дисплазия, 2) пациенти с нарушения в съединителната тъкан и 3) бременности, насочени за пренатална диагностика.

В първата група от 41 **пациенти с диагноза скелетна дисплазия** бяха извършени изследвания за мутации във *FGFR3* и *FGFR2*. Бяха открити точкови мутации в тези гени при 24 от изследваните (63.42%), които се разпределят в следните подгрупи: 14 с предполагаема диагноза ахондроплазия (34.15%), 8 - с хипохондроплазия (19.51%) и 2-ма - с Apert синдром (4.88%). Като допълнение, при един пациент се касаеше за хромозомно преустройство с интерстициални трипликации в дългото рамо на 15-та хромозома, водещо до тетразомия, а при друг пациент беше открит генетичен вариант в *MN1* гена, водещ до поява на стоп кодон и до преждевременно терминиране на белтъчния синтез.

В групата от 25 **пациенти, насочени с диагноза нарушения в съединителната тъкан** бяха открити патологични генетични варианти в гена *COL1A1* при 3-ма пациенти с диагноза Osteogenesis imperfecta тип I (12% от изследваните). При един пациент с диагноза синдром на Marshall-Stickler беше открит патологичен генетичен вариант в *COL11A1* гена (4% от изследваните). Като допълнение, при един пациент с предполагаема клинична диагноза Osteogenesis imperfecta се откри делеция на дългото рамо на X хромозомата с размер 870,743,253 кб, непубликувана в литературата.

Делецията включва гените *PLS3* (OMIM*300131) и *SLC6A14* (OMIM*300444), асоциирани със заболявания при човека.

При 31 от изследваните постнатално 66 пациентаи (47%) поставената клинична диагноза беше изяснена на молекулно ниво, като бяха открити както известни мутации, така и нови, непубликувани досега патогенни генетични варианти.

Пренатална диагностика беше проведена при 9 случая с УЗ данни за скелетни аномалии на плода или фамилна обремененост. При един от изследваните ембриони (12.5%), насочен по ултразвукови данни, беше открит молекулен дефект, водещ до Танатоформна дисплазия тип I.

Благодарение на въведената методика за диагностика на скелетните дисплазии и колагенози стана възможно провеждането на адекватна и навременна генетична консултация, профилактика и пренатална диагностика на засегнатите семейства

SUMMARY

The skeletal dysplasias and diseases with connective tissue disorders (collagenoses) are two large groups of heterogeneous congenital conditions that are socially significant, as they are rare diseases with severe clinical manifestations that affect the overall psycho-social and socio-economic status of the individual and his family.

Skeletal dysplasias are disorders of cartilage/skeletal tissue associated with abnormalities in the shape and size of the skeleton, disproportion of the long bones, spine and head [Aravidis, Christos et al.,2014]. There are over 450 known conditions, and although each one of them is rare, they collectively occur at a frequency of 1:5000 live births [Barbosa et al,2009]. The causes of skeletal dysplasias are as diverse as the number of different diseases, but there are known mutations in a few specific genes (eg, *FGFR3, FGFR2*).

Diseases with connective tissue disorders occur with a frequency of 1:20,000 live births and are mainly associated with disorders of cartilage tissue, joints, muscles and bones growth, but can also affect other organs and systems - including the eyes, heart, lungs, kidneys, gastrointestinal tract, blood vessels, etc [Uttarilli et al.,2019]. Currently, more than 200 genes are associated with connective tissue diseases and are divided into different groups, according to the severity of the clinical manifestation [Nicol et al.,2019]. The subject of this dissertation work were the genes *COL1A1, COL1A2, COL2A1* and *COL11A1*, which are associated with a narrow group of hereditary diseases - Osteogenesis imperfecta, Marshall syndrome and Stickler syndrome.

In the course of the implementation of the present work, optimized methods were also introduced for the research of genes, related to skeletal dysplasias and collagenoses. The scientific development was focused on searching for molecular defects along the length of the *FGFR3, FGFR2, COL1A1, COL1A2, COL2A1, COL11A1* genes. The developed molecular-genetic approach was applied for the study of 75 patients - in 66 of the cases, the analyses were performed postnatal, and in 9, prenatal diagnosis was performed. Based on the previously conducted clinical examinations, the patients were divided into 3 groups: 1) patients with a diagnosis of skeletal dysplasia, 2) patients with connective tissue disorders, and 3) pregnancies forwarded for prenatal diagnosis.

In the first group of 41 patients diagnosed with skeletal dysplasia, studies were performed for mutations in *FGFR3* and *FGFR2*. Point mutations in these genes were found in 24 of the patients (63.42%): in 14 with a suspected diagnosis of achondroplasia (34.15%), 8 with hypochondroplasia (19.51%), and in 2 with Apert syndrome (4.88%). In addition, one patient had a chromosomal rearrangement with interstitial triplications in the long arm of chromosome 15, leading to tetrasomy, and in another patient was found a genetic variant in the *MN1* gene, leading to a stop codon generation and to the premature termination of protein synthesis.

In the group of 25 patients with a diagnosis of connective tissue disorders, pathological genetic variants in the *COL1A1* gene were found in 3 patients with a diagnosis of Osteogenesis imperfecta type I (12% of the examined). In one patient diagnosed with Marshall-Stickler syndrome, a pathological genetic variant in the *COL11A1* gene was found (4% of the examined). In addition, in one patient with a suspected clinical diagnosis of Osteogenesis imperfecta a deletion of the long arm of the X chromosome, with the size of 870,743,253 kb, unpublished in the literature, was found. The deletion includes the *PLS3* (OMIM*300131) and *SLC6A14* (OMIM*300444) genes associated with human disease.

In 31 of the 66 patients examined postnatal (47%), the presumed clinical diagnosis was clarified at the molecular level, as there were detected both known mutations and new, previously unpublished pathogenic genetic variants.

The prenatal diagnosis was performed in 9 cases with ultrasound evidence for foetal skeletal anomalies or family burden. In one of the embryos studied (12.5%), forwarded for

analyses on the bases of an ultrasound data, a molecular defect leading to Thanatophoric dysplasia type I was found.

Thanks to the introduced methodology and systematic approach for the diagnosis of skeletal dysplasias and collagenoses, it became possible to conduct an adequate and timely genetic counselling prevention, and prenatal diagnosis of the affected families.

.

INTRODUCTION

Skeletal dysplasias and diseases with connective tissue disorders (collagenoses) are two large groups of heterogeneous congenital conditions that are socially significant, as they are rare diseases with severe clinical manifestations that affect the overall psycho-social and societal -economic status of the individual and his family.

Skeletal dysplasias are genetic diseases that are caused by mutations in specific genes. It is a heterogeneous group of diseases associated with abnormalities in the skeleton, more precisely with disorders of cartilage tissue and bone growth, which leads to pathological shape and size of the skeleton and disproportion of the long bones, spine and head. Although each of the diseases is rare, in total they occur with a frequency of 1:5000 live births [Orioli, 2017]. Diseases range in severity from premature arthropathies to severe dwarfism leading to perinatal death. Numerous orthopedic, neurological, auditory, visual, pulmonary, cardiac, renal and physiological complications can be observed as accompanying symptoms.

Connective tissue disorders are another large group of disorders involving the protein-rich tissue that supports various organs and parts of the body. Examples of connective tissue diseases are related to bone and cartilage disorders caused by genetic disorders, in total they occur with a frequency of 1:20,000 live births [Byers, 2012]. Currently, more than 200 genes are associated with connective tissue diseases and are divided into different groups according to the effect they exhibit. The subject of this dissertation work are the genes that are associated with a narrow group of hereditary diseases - Osteogenesis imperfecta, Marshall syndrome and Stickler syndrome.

Family history (including miscarriages or stillbirths) of skeletal dysplasias is of particular importance in assessing the nature and inheritance pattern of the disease. Parents, brothers, sisters and other relatives should be carefully examined for mild manifestations of the disease, due to the variable clinical penetrance and expressiveness of the diseases and due to the fact that by routine ultrasound examination, prenatally they cannot always be diagnosed.

AIM AND OBJECTIVES

AIM

The aim of this dissertation is to develop a molecular genetic approach for the analysis of the FGFR2, FGFR3, COL1A1, COL1A2, COL2A1 and COL11A1 genes for the diagnosis of diseases with disorders in the skeletal or connective tissue. Assessment of genotype-phenotype correlations.

OBJECTIVES

1. Introduction and optimization of a molecular genetic approach for the identification of pathological substitutions in the FGFR2, FGFR3, COL1A1, COL1A2, COL2A1 and COL11A1 genes.

2. Selection of a group of patients with a clinical diagnosis of skeletal dysplasias and diseases with disorders in the connective tissue for carrying out genetic studies.

3. Taking a family history.

4. Conducting molecular genetic analysis in patients with a clinical diagnosis potentially associated with genetic variants in the FGFR2, FGFR3, COL1A1, COL1A2, COL2A1 and COL11A1 genes.

5. Conducting prenatal molecular genetic analysis in case of family burden or ultrasound data for skeletal or connective tissue malformations, by means of targeted screening along the length of a specific gene.

6. Comparing the genetic variants found in the studied genes with those published in the world databases.

7. Searching for genotype-phenotype correlations in order to predict the course of disease development.

8. Summary of the obtained results.

MATERIALS AND METHODS

CLINICAL MATERIAL

Within the framework of the current dissertation work, molecular genetic studies were conducted on a total of 75 patients. Based on clinical diagnosis, patients were grouped as follows:

Group 1:

41 patients with a suspected diagnosis of skeletal dysplasia.

FGFR2 and FGFR3 gene studies were performed in this group of patients. The distribution of studied patients by suspected diagnosis and studied gene is presented in Table 1. The FGFR3 gene was studied in 35 patients, and the FGFR2 gene was studied in 6 patients.

Group 2:

25 patients with a suspected diagnosis involving connective tissue disorders.

The COL1A1, COL1A2, COL11A1 and COL2A1 genes were tested in the patients with collagenosis. The distribution of studied patients by presumptive diagnosis and studied gene is presented in Table 8. COL1A1 and COL1A2 genes were studied in 23 patients, and COL11A1 and COL2A1 genes were studied in 2 patients.

Group 3:

9 cases for prenatal diagnosis

Genetic testing after an invasive procedure was performed in 9 foetuses due to US findings of congenital skeletal anomalies or family history.

Table 1. Distribution of patients by presumptive diagnosis and studied genes

Clinical diagnosis	Number of the patients	Target gen
Achondroplasia *	23	*FGFR3*
Hypochondroplasia *	10	
Muenke syndrome	2	
Apert syndrome	6	*FGFR2*
Osteogenesis imperfecta	23	*COL1A1* и *COL1A2*
Marshall - Stickler syndrome	2	*COL11A1* и *COL2A1*
US evidence of skeletal anomalies or family history	9	*FGFR3*

BIOLOGICAL MATERIAL

For the purposes of molecular genetic studies, DNA was isolated from venous blood and from chorionic villi and amniotic fluid.

METHODS

✓ Pre-analytical processing of the material
o Isolation of high molecular weight DNA from venous blood by salt method

METHODS

- ✓ **Pre-analytical processing of the material**
  - o Isolation of high molecular weight DNA from venous blood by salt method
  - o Isolation of high molecular weight DNA from chorionic villi and amniotic cells using Chelex 100 (Chelex®100)
- ✓ **Analytical processing of the material**
- ✓ Polymerase chain reaction (pcr) of the following genes: fgfr2, fgfr3, col1a1, col1a2, col2a1 and col11a1. electrophoresis to assess the quality and quantity of the amplification product in an agarose gel
- ✓ sanger sequencing with fluorescently labeled dideoxy nucleotides
- ✓ multiplex ligation-dependent probe amplification (mlpa)
- ✓ karyotyping
- ✓ evaluation of the nucleotide sequence of the region around the site of occurrence of the detected pathological genetic variants
- ✓ conducting array-cgh analysis in selected patients in collaboration with Dijana Plaseska-Karanfilska, MD, academy of sciences and arts research center for genetic engineering and biotechnology "Georgi D. Efremov", Skopje Republic of Macedonia
- ✓ carrying out whole exome sequencing (wes) in selected patients in collaboration with prof. Borut Peterlin, center for mendelian genomics clinical institute of medical genetics, Ljubljana, Slovenia.

RESULTS AND DISCUSSION

The group of patients that was the subject of molecular-genetic studies in the present scientific work was provided by the Genetic Medical-Diagnostic Laboratory "Genika", Sofia. The total number of patients examined was 75, and based on clinical data, they were divided into 3 groups: 1) patients referred with a suspected diagnosis of skeletal dysplasia, 2) patients with a suspected diagnosis of collagenosis, and 3) foetuses with ultrasound evidence of skeletal dysplasia and collagenoses or pregnancies with evidence of familial burden.

Figure 1 shows the distribution of prenatal and postnatal diagnostics performed - according to US data and preliminary clinical history, respectively.

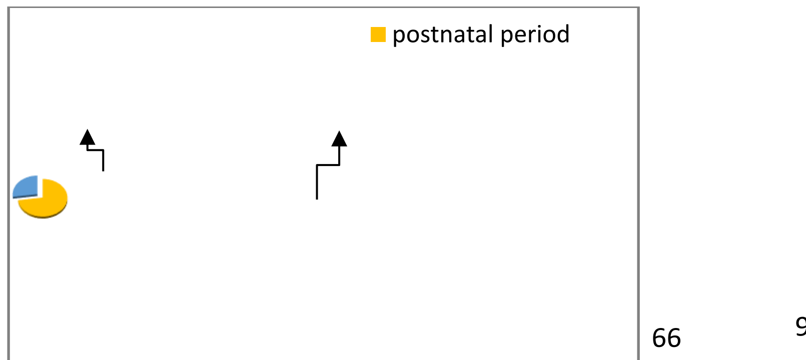


Figure 1. Distribution of patients, depending on the period in which they were examined.

DISEASES WITH DISORDERS IN THE CARTILAGE TISSUE

Based on the preliminary clinical data, the postnatally examined 41 patients were divided into 2 groups, depending on the submitted clinical history and the studied genes (table 2).

Table 2.
Distribution of patients by gene grouped by sex and number.

	Male	*Female*	**Total number**
FGFR3	19	16	35
FGFR2	2	4	6

RESULTS IN THE RESEARCHED GROUP OF PATIENTS S

FGFR3-RELATED SYNDROMES (FGFR3 - SKELETAL DYSPLASIAS):

Fibroblast growth factors (FGFs) are part of a large family of polypeptide growth factors that are involved in multiple processes in the cell, including mitogenesis and angiogenesis. FGF receptors such as FGRF3 include in their structure an extracellular domain with 2 or 3 immunoglobulin-like domains, a transmembrane domain and a cytoplasmic tyrosine kinase domain.

The *FGFR3* gene was studied in 35 patients with syndromes related to the *FGFR3* gene with a diagnosis of achondroplasia (23 individuals), hypochondroplasia (10 individuals), Muenke's syndrome (2 individuals). In the course of the conducted molecular genetic studies, point mutations were found in 22 patients (63% of those examined in this group).

Results are presented by diagnosis:

1. ACHONDROPLASIA

Achondroplasia is the most common non-lethal skeletal dysplasia with an incidence of 1:20,000 live births. The phenotypic presentation of the disease is characterised by extremely short stature and body disproportion, without intellectual deficit.

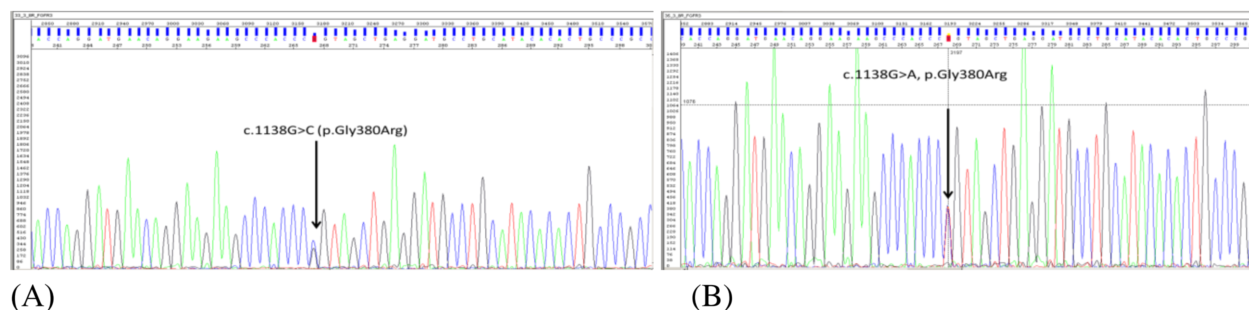
In the studied 23 patients with achondroplasia, 2 mutations were found in exon 8 of the *FGFR3* gene: c.1138G>A; p.Gly380Arg (in 12 of the casts) and c.1138G>C; p.Gly380Arg (in 2 of the studied). In total, in 14 patients (~61%), the diagnosis of achondroplasia was genetically verified (Table 10). In all proven cases, there was no evidence of achondroplasia in the family and the detected mutations arose de novo.

In 9 patients, the molecular defect was not clarified - 8 of them had an unclear clinical picture resembling skeletal dysplasia and in them only the exons of the *FGFR3* gene, which are mutation hotspots associated with skeletal dysplasias, were examined. Genetic studies in this gene in them were not continued, due to conflicting clinical symptoms, and they were redirected to analysis of other genes. One patient referred with a suspected clinical diagnosis of achondroplasia was negative for mutations in the *FGFR3* gene and underwent revision of the diagnosis because achondroplasia was ruled out.

Table 3. Presentation of the results obtained after analysis of the *FGFR3* gene

Пациент №	Ген	Екзон	Нуклеотидна Замяна	Аминокиселинна Замяна	Диагноза
1	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
2	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
3	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
4	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
5	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
6	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
7	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
8	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
9	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
10	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
11	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
12	*FGFR3*	8	c.1138G>A	p.Gly380Arg	Achondroplasia
13	*FGFR3*	8	c.1138G>C	p.Gly380Arg	Achondroplasia
14	*FGFR3*	8	c.1138G>C	p.Gly380Arg	Achondroplasia
15	*FGFR3*	1-17	-	-	Achondroplasia
16	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
17	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
18	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
19	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
20	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
21	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
22	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia
23	*FGFR3*	4,5,6,8-16	-	-	Skeletal dysplasia

Figure 2 shows the sequence profiles of the detected mutations c.1138G>A (p.Gly380Arg) and c.1138G>C (p.Gly380Arg) leading to achondroplasia.



(A) (B)

Figure 2.
A - Sequence profile of mutation c.1138G>A; p.Gly380Arg (sequencing with primer R).
B - Sequence profile of mutation c.1138G>C; p.Gly380Arg (sequencing with primer R)..

The two detected mutations are well known and described in the world literature. They are the most common pathological variants in the FGFR3 gene leading to the appearance of achondroplasia. The mutation c.1138G>A (p.Gly380Arg) was reported in over 98% of patients, while c.1138G>C (p.Gly380Arg) was found in only about 1% of patients [Rousseau et. et al., 1996].

Both mutations affect the same nucleotide from codon 380, which codes for the amino acid glycine falling within the transmembrane domain of the FGF receptor. In both mutations, the nucleotide substitutions result in the incorporation of arginine instead of glycine in the bulge. The nucleotide substitutions found are of two types: transition and transversion. The transition represents the replacement of one purine nucleotide with another purine nucleotide (c.1138G>A); the transversion is the replacement of a purine nucleotide with a pyrimidine nucleotide (p.1138G>C). The c.1138G>C transversion is of much lower frequency than the c.1138G>A transition. This is thought to be due to the fact that in transitions chemically identical bases are replaced, while in transversions there is a replacement of a purine with a pyrimidine base or vice versa. This, in turn, is reflected as a change in DNA conformation and is a prerequisite for easier recognition and removal of the "error" by the repair systems. When the altered nucleotide is missed by the repair systems and remains fixed in the genome as a pathological mutation, it is logical to ask whether regions of the surrounding sequence do not somehow influence the DNA polymerase so that it makes errors or is unable to repair those that have occurred. This issue is particularly important when dealing with the de novo occurrence of mutations affecting the same nucleotide in a given gene. To evaluate the role of the nucleotide sequence around the site of occurrence of the two pathogenic genetic variants in the *FGFR3* gene, the surrounding sequence at a distance of 30 nucleotides in the 5' and 3' direction from position c.1138 of the gene was examined. (Figure 3).

Of interest were nucleotide sequences such as palindromes, straight or inverted repeats, symmetry elements, and other sequence motifs that could be relevant to the course of DNA.

polymerase during replication.

c.1138G>A; p.Gly380Arg

CCAGGATGAAGAGGAAGAAGCCCACCCAGAAGCTGAGGATGCCTGCATA

Figure 3. Analysis of the surrounding sequence at the *FGFR3*:p.Gly380Arg mutation site.

After analysis of the surrounding sequence, a number of perfect straight repeats were identified, which could be involved in the formation of secondary structures that interfere with the normal functions of the polymerase. The presence of repeated motifs (AGGA), (GAAG) and, in addition, a CCSSCC-symmetric element near the mutation is striking. Repeated C-elements around the site of mutation could lead to the formation of secondary strand kinks, favouring DNA polymerase slippage, where a nucleotide identical to the previous one can be incorporated. The occurrence of mutations in the same position of the gene, as a result of independent mutational events, may be based entirely on features of the surrounding sequence [Todorova et. et al., 1996].

Figure 4 shows the spatial structure of the FGFR3 receptor and the p.Gly380Arg mutation, which falls within the alpha-helical region of its transmembrane domain.

Figure 4. Spatial structure of the FGFR3 gene protein using AlphaFold2, the location of the pathological mutation leading to Achondroplasia is marked with an arrow.

The wild-type amino acid is glycine, which has no side radical (one H). The replacement of the nonpolar amino acid glycine with the polar amino acid arginine, which has an additional

basic group and the ability to participate in the formation of ionic forces of attraction and repulsion, can lead to structural changes in the domain that affect the stability and activity of the receptor [Bochyńska et al ., 2018] The missense mutation affects a weakly conserved region in the FGFR3 domain. By the conservation of the amino acid among the species, the pathogenic influence of the mutation can be judged.

In addition, a mathematical processing software (PolyPhen-2 prediction of functional effects of human nsSNPs) was used to evaluate the pathogenicity of this substitution, which predicts the possible influence of the amino acid substitution on the structure and/or function of the protein (Figure 5). The software presents the result in the form of an index (score), the value of which varies from 0 to 1. Score 0 is interpreted as a neutral substitution, without impact on protein function, while score 1 is evaluated as a pathogenic substitution.

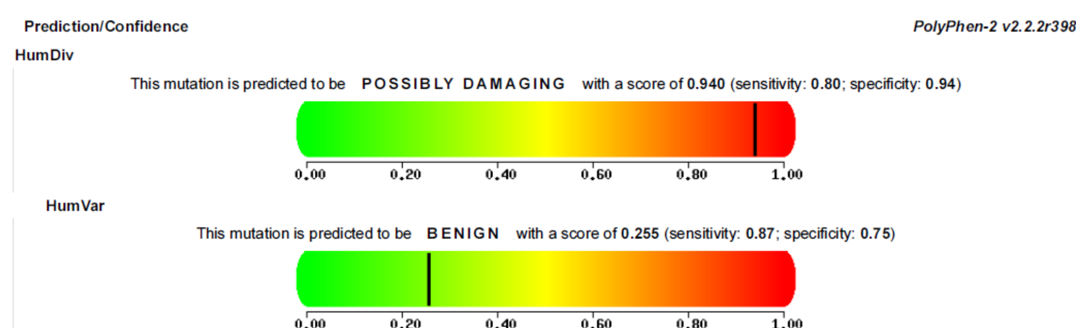


Figure 5. Result of the online software predictor Polyphen2 for estimating the pathogenicity of the FGFR3:p.Gly380Arg mutation.

The two mathematical models by Polyphen2 used to predict the possible impact of the substituted amino acid, in this case show a different index (0.940 and 0.255, respectively). The result of the first mathematical model shows that the probability of pathogenicity is close to 1, which is the maximum pathogenicity index. The second mathematical model gives an index of 0.255, which can be interpreted rather as preserving the normal properties of the FGFR3 protein and its functional specificity. This discrepancy in the mathematical interpretations may be due to the fact that the two models have different calculation algorithms that take into account different parameters. The influence of the fact that the amino acid sequence in the region is not entirely conserved can also be assumed. The amino acid glycine occurs normally at this position of the FGFR3 protein sequence, for example in Canis familiars, Mus musculus, Danio rerio, Chimpanzee, etc.

As already mentioned, the amino acid glycine at position 380 falls within the transmembrane domain (TM) region of the FGFR3 receptor (Figure 26). The TM domain of FGFR is an intracellular domain and functions by dimerization. The FGFR3 receptor has two kinase domains that are activated after phosphorylation. This process is mediated by the binding of FGF ligands, which stabilize the domain and help activate it [Shiang et al., 1994]. As a result of this phosphorylation, the process of ossification is activated - the conversion of cartilage tissue into bone.

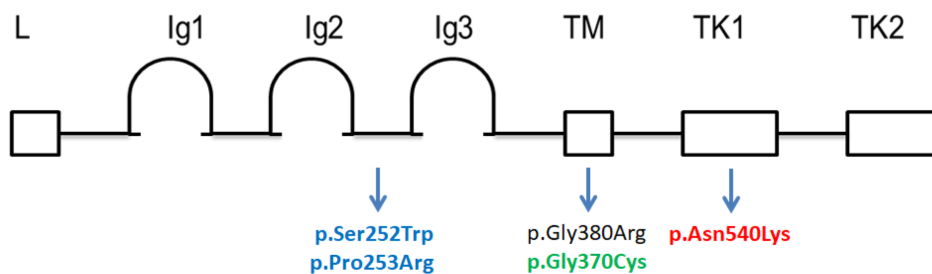


Figure 6. Localization of known pathological genetic variants in FGF receptors associated with various disorders in skeletal tissue.

Since the FGFR3:p.Gly380Arg mutation falls in the transmembrane domain, two hypotheses can be formulated regarding the increase in receptor activity:

1) The resulting mutation affects the dimerization or phosphorylation of the tyrosine domains in the activation loop of the receptor. The amino acid substitution likely stabilizes the receptor and thus increases its ligand-independent activation. Thus, the receptor is constantly active and this leads to premature ossification of cartilage tissue.

2) The mutation can be viewed as a Gain-of-function mutation that leads to overexpression of the product.

Regardless of the mechanism of action, the p.Gly380Arg mutation causes continuous activation of FGFR3, inhibiting chondrocyte proliferation and differentiation. When the signaling pathway is overactivated, chondrocytes cannot proliferate and differentiate normally, and as a result less bone tissue is formed [Jesse, 2012].

2. HYPOCHONDROPLASIA

Hypochondroplasia is a non-lethal skeletal dysplasia with a frequency of 1:40,000 live births. The phenotypic manifestation of the disease is a less pronounced disproportion of the body, short stature and delayed intellectual development.

The studied patients with clinical evidence of hypochondroplasia were a total of 10 and the results of the analyses performed are presented in Table 8. In 8 of the patients, the pathological variant c.1620C>A, p.Asn540Lys in exon 11 of the *FGFR3* gene was detected (Figure 7), as all patients lacked data on familial burden.

Two patients were left with an unexplained molecular defect after analysis of mutation hot regions of the *FGFR3* gene. In both cases, the clinical presentation was vague, resembling skeletal dysplasia, and therefore further screening for mutations in the *FGFR3* gene was discontinued. In them, the clinical diagnosis is subject to revision and referral to other genetic tests.

Table 4. Results of studies of patients for hypochondroplasia.

№	Gene	Exon	Nucleotide substitution	Amino acid substitution	Diagnosis
1	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
2*	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
3*	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
4	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
5	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
6	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
7	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
8	*FGFR3*	11	c.1620 C>A	p.Asn540Lys	Hypochondroplasia
9	*FGFR3*	11	-	-	Hypochondroplasia
10	*FGFR3*	4-6, 8-16	-	-	Skeletal dysplasia

***identical twins**

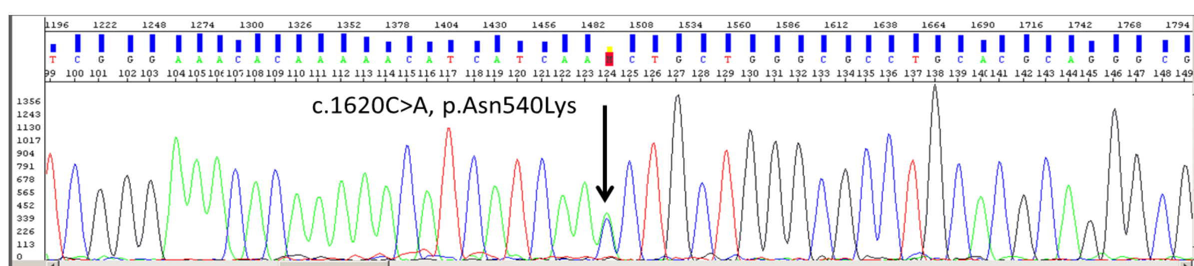


Figure 7. Sequence profile of the c.1620C>A mutation; p.Asn540Lys.

The mutation c.1620C>A, p.Asn540Lys is located in the ATP-binding segment of the tyrosine kinase domain of the FGF receptor (see Figure 6 - highlighted in red). According to literature data, this pathological variant is one of the two most common in patients with hypochondroplasia. The detected mutation is of the transversion type - replacement of a pyrimidine with a purine nucleotide. By analogy with the frequent mutation in achondroplasia, the region around the c.1620C>A substitution was analyzed, noting that the triplet in which the mutation occurs, as well as the nucleotide sequence before it, are -AAC-rich direct repeats (Figure 8). Immediately before the mutation there is a symmetrical element -ACAAAAACA-, again flanked on both sides by the straight repeat -AAC-, and in the second -AAC- triplet the nucleotide substitution occurs. Furthermore, the nucleotide sequence after the substitution is rich in CTG repeats, which can additionally affect the synthesis and repair activity of the

polymerase. It can be assumed that these motifs affect the increased mutability of the specific nucleotide and its formation as a hot mutation point, which also explains the high frequency of occurrence of this mutation. According to literature data, the p.Asn540Lys variant occurs with a frequency of about 70% in patients with clinical evidence of hypochondroplasia. The remaining 30% of patients remain with an unclear genetic defect, demonstrating the genetic heterogeneity of the disease [Song et al.,2012]

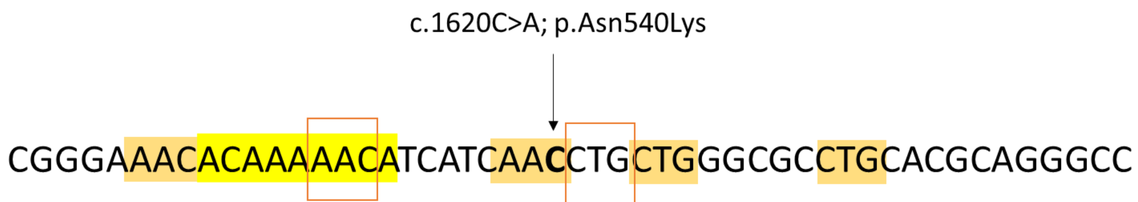
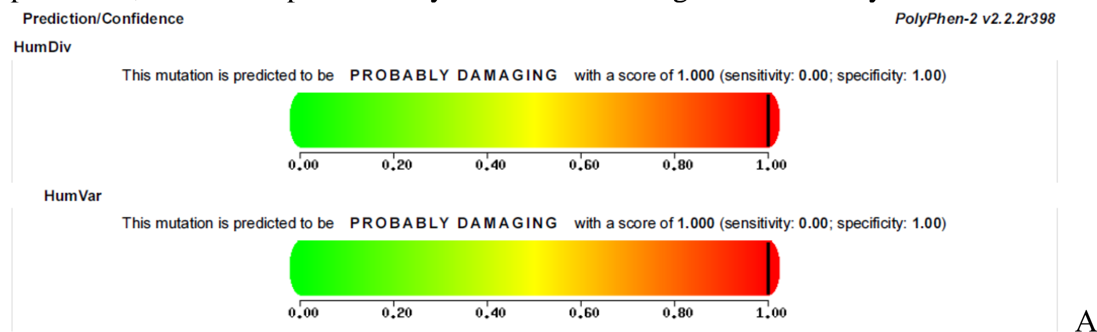
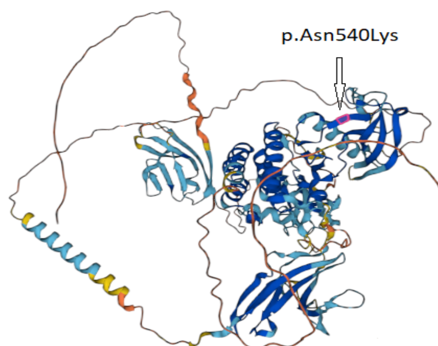


Figure 8. Analysis of the region of the FGFR3 gene where the mutation c.1620C>A, p.Asn540Lys falls.

The PolyPhen2 predictor was used to assess the pathogenicity of the p.Asn540Lys substitution (Figure 9A). In the same figure (Figure 9B). a spatial model of the FGFR3 receptor is also presented, with the p.Asn540Lys mutation falling within the tyrosine kinase domain 1.



A



B

Figure 9. A-Result of the online software predictor Polyphen2 for estimating the pathogenicity of the FGFR3:p.Asn540Lys mutation. B- Spatial structure of the FGFR3 receptor modeled

using AlphaFold2. The p.Asn540Lys mutation leading to hypochondroplasia is marked with an arrow.

Both mathematical models for evaluating the pathogenicity of the amino acid substitution show an index with a maximum value, which means a high pathogenicity of the mutation. The amino acid asparagine is located in a highly conserved region of the tyrosine kinase domain of the FGFR3 receptor. The replacement of the amino acid asparagine (with an amide group in the side radical) by lysine (with an amino group in the side radical) results in the generation of an additional positive charge in this conserved region, enabling the formation of new ionic interactions and conformations, which disrupts normal activity of the kinase domain of the receptor.

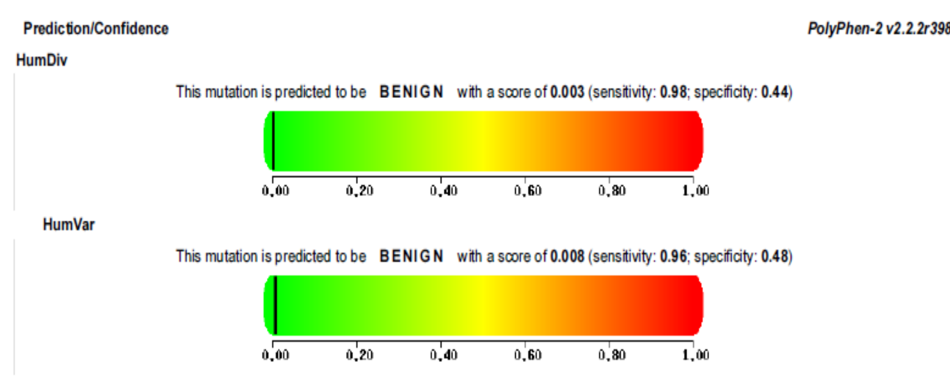
MUENKE SYNDROME

Table 5. Results of patients tested for Muenke syndrome.

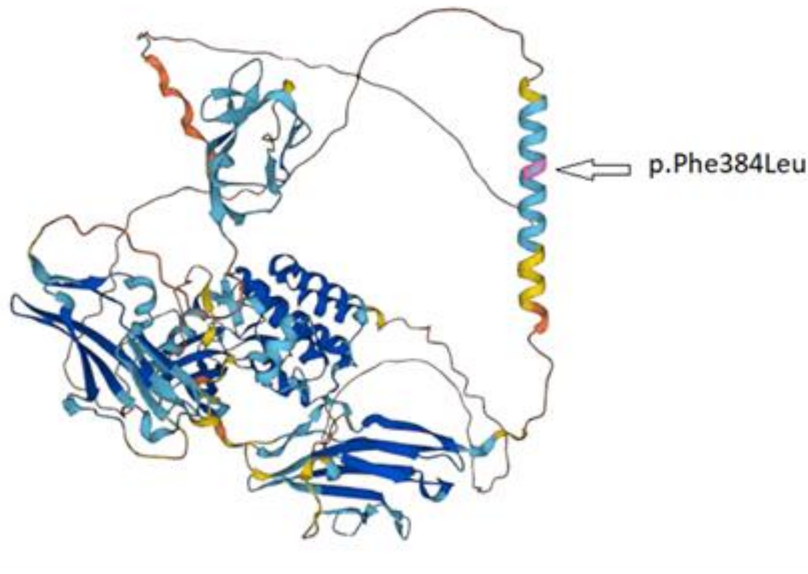
№	Gene	Exon	Nucleotide substitution	Aminoacid substitution	Diagnosis
1	*FGFR3*	1-17	c.1150C>T	p.Phe384Leu	Muenke syndrome
2	*FGFR3*	1-17	-	-	Muenke syndrome

Two patients from our sample were referred for genetic testing with a clinical diagnosis of Muenke syndrome. There is evidence that this syndrome is caused by the mutation c.749C>G, p.Pro250Arg in exon 6 of the *FGFR3* gene. The patients were initially screened for this mutation, but the result was negative. Whole gene sequencing was subsequently performed and a nucleotide substitution c.1150C>T, p.Phe384Leu in exon 8 of the *FGFR3* gene was found in one patient. The detected nucleotide substitution changes the TTC codon encoding the amino acid phenylalanine (containing an aromatic ring) to CTC encoding the amino acid leucine (a nonpolar, aliphatic amino acid). Amino acid p.Phe384 is located in a region of the transmembrane domain of the FGFR3 receptor that is not characterized by high conservation. Furthermore, p.Phe384 is located only three amino acid residues away from p.Gly380, whose change to arginine leads to achondroplasia. Therefore, we expected that the detected molecular variant p.Phe384Leu would also prove to be pathogenic.

The software predictor PolyPhen2 (Figure 10A) was used to assess the pathogenicity of this substitution, which unambiguously assessed the detected substitution as non-pathogenic (indices of both assessment methods – 0.003 and 0.008, respectively).



A



B

Figure 10. A-Result of the online software predictor Polyphen2 for assessing the pathogenicity of the mutation FGFR3:p.Phe384Leu; B- Spatial structure of the FGFR3 receptor modelled using AlphaFold2 The p.Phe384Leu substitution is marked with an arrow.

The literature search for p.Phe384Leu confirmed that the substitution of phenylalanine with leucine at position 384 of the protein has no pathogenic effect and is not related to the observed clinical symptoms in the patient. It has been suggested in the literature that this variant may play a role as a modulator of the phenotypic expression in patients with skeletal dysplasia. This hypothesis was later rejected because the results did not prove such a modulating effect of the p.Phe384Leu substitution [Thatiane Yoshie Kanazawa, 2014].

From a molecular point of view, the pathologically damaged FGFR3 gene is characterized by several features:

The detected genetic variants c.1138G>A, c.1138G>C and c.1620C>A in the FGFR3 gene affect different exons and fall into different domains of the protein – respectively the transmembrane domain and the tyrosine kinase domains (figure 6). However, many similarities are observed between the p.Asn540Lys and p.Gly380Arg mutations causing Hypochondroplasia and Achondroplasia, respectively, which might be expected to be associated with increased activity of the encoded receptor. As a result, the ends of growing bones ossify prematurely, fail to grow, and remain short, a symptom seen in FGFR3-linked skeletal dysplasias.

The FGFR3 gene is mainly affected by emerging mutations due to repeated independent mutational events in certain regions called "mutational hotspots". Approximately 80% of patients with achondroplasia and ~80% of patients with hypochondroplasia have parents of normal height and the mutation arises de novo. Newly occurring mutations are thought to be predominantly associated with advanced paternal age (> 35 years) [Penrose, 1955; Stoll, 1982]. De novo mutations causing achondroplasia more often affect the paternal allele, resulting in germline mosaicism during spermatogenesis [Wilkin, 1998].

RESULTS IN THE RESEARCHED GROUP OF PATIENTS

FGFR2-RELATED SYNDROMES (FGFR2 SKELETAL DYSPLASIA)

FGFR2-linked skeletal dysplasias are a group of 8 disorders with an autosomal dominant pattern of inheritance. These include Pfeiffer syndrome, Apert syndrome, Crouzon syndrome, Beare-Stevenson syndrome, FGFR2-related isolated coronary synostosis, Jackson-Weiss syndrome. The most characteristic clinical features for each of the diseases are shown as follows in Table 6.

Table 6. Clinical features in FGFR2-related skeletal dysplasias

Syndrome	Craniosynostosis	Airway features	Hands/Feet	Neurological complications
Crouzon syndrome	+	Midface retrusion, +/- airway obstruction	Normal	Hydrocephalies
Crn syndrome-acanthosis nigricans	+	Midface retrusion, +/- airway obstruction	Normal	Hydrocephalies
Apert syndrome	+	Midface retrusion Cleft palate Dental anomalies airway obstruction	Soft tissue & bony syndactyly	Agenesis of corpus callosum
Pfeiffer syndrome	+	Midface retrusion,proptosis +/- airway obstruction	Brachydactyly Bony nodules on phalanges & metacarpals	Hydrocephalies
Jackson-Weiss syndrome	+	Variable midface retrusion & proptosis airway obstruction	Broad medially deviated great toes, syndactyly,	Normal ID
Beare-Stevenson syndrome	+	Severe midface retrusion w/proptosis Cleft palate airway obstruction	Normal	Hydrocephalies
FGFR2-coronary synostosis	+	Most w/ moderate to severe midface retrusion & proptosis Cleft palate airway obstruction	Normal	Normal

A total of 6 patients were referred for genetic testing in connection with clinical data on *FGFR2*-related syndromes. Table 7 shows the results obtained after the molecular genetic studies.

Table 7. Results of patients tested for syndromes related to the *FGFR2* gene.

№	Exon	Nucleotide substitution	Aminoacid substitution	Diagnosis
1	7	c.755 C>G	p.Ser252Trp	Apert syndrome
2	7	c.755 C>G	p.Ser252Trp	Apert syndrome
3	1-18	-	-	Skeletal dysplasia
4	1-18	-	-	Skeletal dysplasia
5	1-18	-	-	Skeletal dysplasia
6	1-18	-	-	Skeletal dysplasia

In two of the studied patients with a clinical diagnosis of Apert syndrome, the pathological variant c.755C>G, p.Ser 252Trp in exon 7 of the *FGFR2* gene was detected (Figure 31).

In 4 patients with a suspected clinical diagnosis of skeletal dysplasia, the study included sequencing of all exons of the *FGFR2* gene, which did not lead to the detection of a molecular defect. The diagnosis in these 4 patients was subject to revision as Apert syndrome was ruled out.

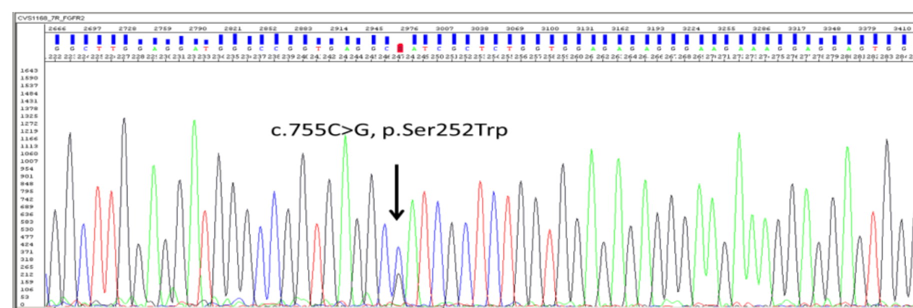


Figure 11 Sequence profile of the mutation c.755C>G, p.Ser 252Trp in exon 7 of the *FGFR2* gene in a patient with Apert syndrome

Apert syndrome is a rare autosomal dominant disorder characterized by craniosynostosis, limb deformities, congenital heart disease, and other systemic malformations, including intellectual disability. The nucleotide substitution c.755C>G, p.Ser 252Trp in exon 7 of the *FGFR2* gene found in our patients is one of the described mutations associated with Apert syndrome. In this exon, according to literature data, two mutations have been described - c.755C>G, p.Ser252Trp, which was found in our patients, and c.758C>G, p.Pro253Arg. Both mutations are of the transversion type and seem to affect two adjacent codons. Analogous to the mutations in the FGFR3 gene, an analysis of the surrounding nucleotide sequence in the region of the two mutations was performed (Figure 12). The presence of the CCGG palindrome and multiple straight and inverted repeats that are S/T rich is striking. These repeats enclose symmetrical elements, with the participation of which the generation of mutational

transformations can be explained. The analyzed repeats could be related to the formation of secondary structures interfering with the normal functions of the polymerase, which would explain the independent occurrence of the mutational event at the same position of the *FGFR2* genetic sequence.

c.755C>G ,p.Ser252Trp c.758C>G, p.Pro253Arg

CTTCCCTCTCTCCACCAGAGCGATCGCCTCACCGGCCCATCCTCCAAGCG

Figure 12. Analysis of the region of the FGFR2 gene in which the mutations c.755C>G, p.Ser252Trp and c.758C>G, p.Pro253Arg fall.

Both reported missense mutations in *FGFR2* are located in the linker region between the 2nd and 3rd activation loop of the immunoglobulin-like domains see Figure 6 (blue colored).

The p.Ser252Trp mutation we discovered is a replacement of the amino acid serine, which is polar uncharged with a hydroxyl group, with the amino acid tryptophan - a hydrophobic, aromatic amino acid with an indole ring. [Mundhofer, 2013]. Both types of amino acid substitutions described in *FGFR2* - p.Ser252Trp and p.Pro253Ar - are associated with the incorporation of an amino acid different in chemical nature and structure compared to the wild type amino acid. This, in turn, could be considered a prerequisite for a drastic change in the orientation of the ligand-binding sites of the receptor, since each of these changes results in increased FGFR2 ligand-binding affinity and decreased specificity.

According to literature data, the p.Ser252Trp mutation we discovered is the more common pathological replacement [Fanganiello, 2007]. The estimated pathogenicity coefficient for the p.Ser252Trp mutation from the PolyPhen2 predictor was 0.999, which corresponds to a high probability of pathogenicity (Figure 13) Clinical symptoms such as cleft palate, midface hypoplasia, underdeveloped upper jaw, ocular abnormalities such as bulging eyes, hypertelorism and palpebral fissures, which were also observed in our patients (see Table 6). Intellectual deficit and syndactyly were more common in patients carrying the other described mutation p.Pro253Arg [Faradz et . al., 2013]. These results suggest strict genotype/phenotype correlations and therefore disease course can be predicted based on the detected mutation

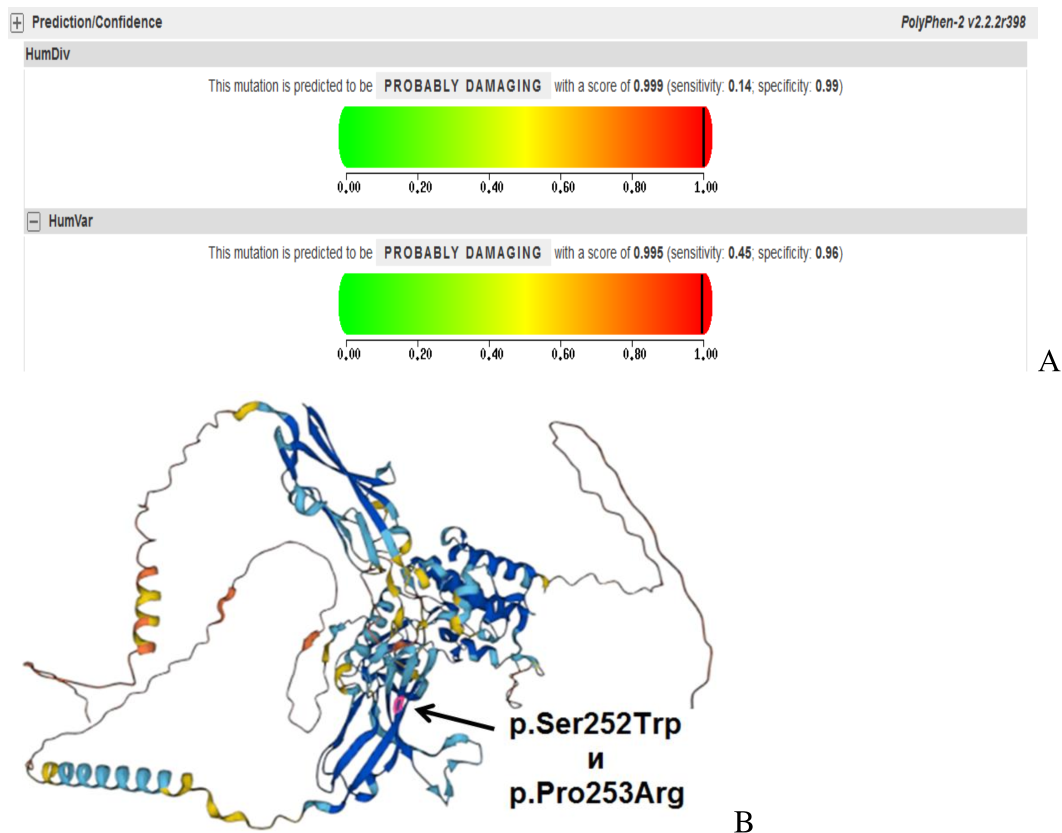
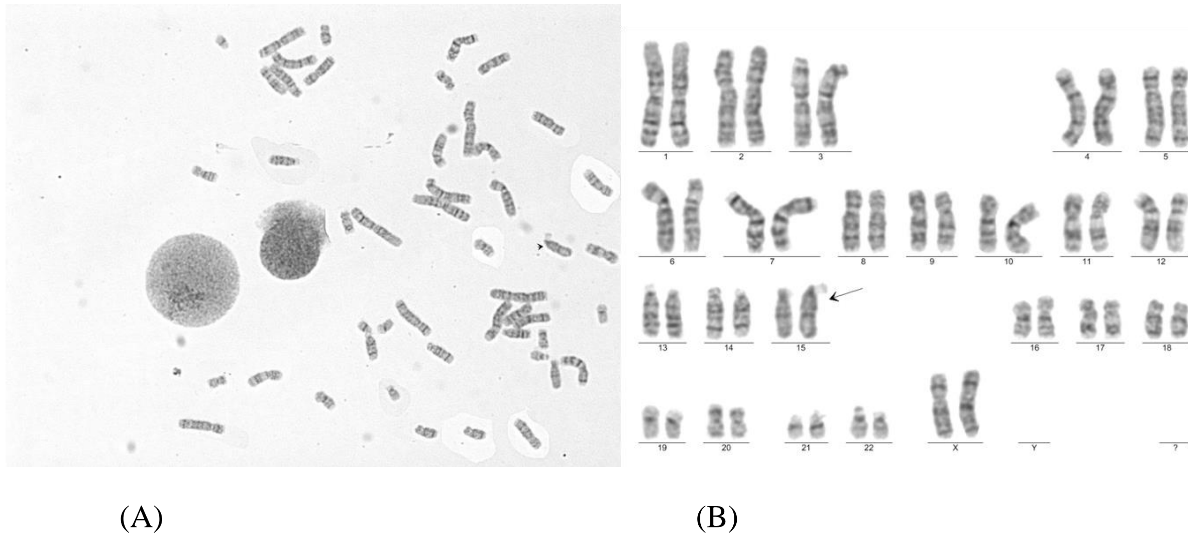


Figure 13. A-Result of the online software predictor Polyphen2 for assessing the pathogenicity of the mutation *FGFR2*:p.Ser 252Trp; B Spatial structure of the *FGFR2* receptor modelled using AlphaFold2 Arrow marks the two substitutions p.Ser252Trp and p.Pro253Arg located next to each other

PATIENTS WITH SKELETAL DYSPLASIA AND AN UNCLEAR MOLECULAR DEFECT

After conducting molecular genetic studies on the FGFR3 and FGFR2 genes in 15 of the patients, no pathological genetic variants associated with skeletal dysplasia were found. In two of them, performing additional genetic tests to clarify the clinical diagnosis led to the discovery of interesting genetic variants:

The first patient was initially referred with a suspected clinical diagnosis of achondroplasia, but the diagnosis was subsequently revised. Short stature, unsteady gait, delayed neuropsychomotor development, dysmorphic features, microcephaly, and the appearance of unilateral hypopigmentation were described. During the diagnostic workup of the patient, new diagnoses began to be discussed because of the combination of dysmorphic features and skin lesions, one of which was hypomelanosis of Ito (OMIM*300337). Based on this information, a cytogenetic study (karyotyping) was performed, where an interesting and previously unpublished finding was found (Figure 14).



(A) (B)

Figure 14.
(A) Photomicrograph of the metaphase plate showing dup(15)(q11.2q13.3);
(B) Karyotype of the patient, arrow indicates the detected dup (15)(q11.2q13.3)

The result obtained showed an abnormal karyotype with additional material on the long arm of chromosome 15, (15) (q11q13) bands. To confirm the presence of this additional material on the long arm of chromosome 15, our next step was to perform MLPA analysis. A copy number bias was detected in him with respect to probes located in the target region 15q11.2 (SNRPN and UBE3A genes).

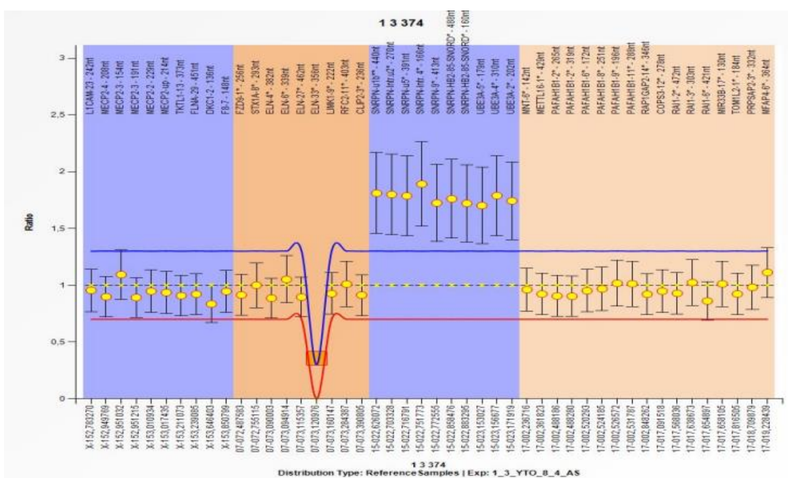


Figure 15. MLPA profile indicating additional material in the long arm of chromosome 15 in the 15q11.2 region.

To clarify the boundaries of the duplication, aCGH analysis was performed (Figure 16), which revealed the following finding: arr [GRCh37] 15q11.2q13.3 (22,558,697-30,366,124) x4, 15q13.2q13.3 (30,652,489-32,462,701) x3.

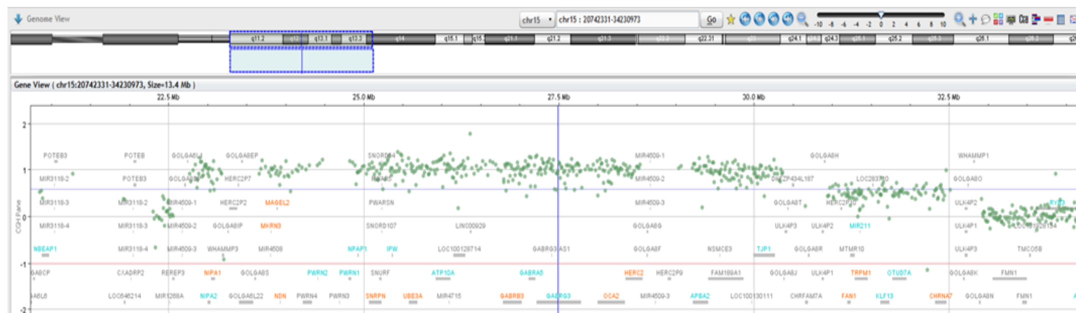


Figure 16.
aCGH profile indicating the boundaries of the duplication (15)(q11.2q13.3) found in the index patient: arr[GRCh37] 15q11.2q13.3 (22,558,697-30,366,124)x4, 15q13.2q13.3 (30,652,489-32,462,701)x3 .

As a final step, family segregation analysis by QF-PCR of polymorphic loci was performed to identify the origin of the chromosomal rearrangement. The following 6 markers located on chromosome 15 were selected for analysis: D15S1513, D15S657, D15S643, D15S659, D15S128 and D15S822. The first four proved uninformative, while the last two proved a maternal origin of the additional material on chromosome 15q (Figure 17).

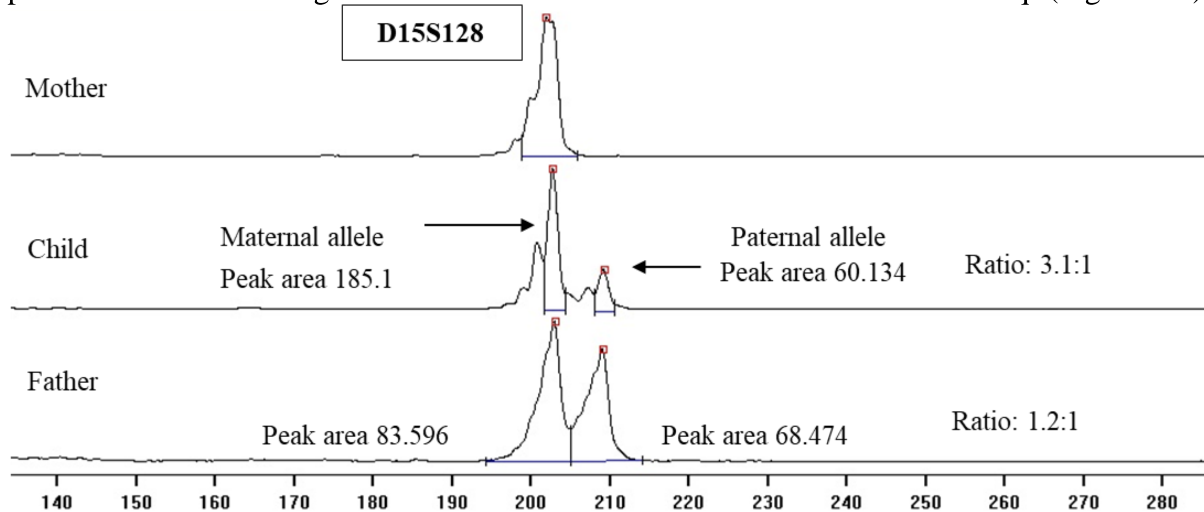


Figure 17. Segregation analysis by QF-PCR of the polymorphic marker D15S128 in a family with hypomelanosis of Ito. Segregation analysis indicated maternal origin of the additional material on chromosome 15q, maternal allele: paternal allele area ratio 3:1.

After a literature review, it appears that the proximal end of the long arm of chromosome 15 is prone to cytogenetic abnormalities such as deletions, duplications, triplications, translocations and inversions, as well as inverted duplications (15) [Castronovo et. al., 2014]. According to literature data, these chromosomal transformations are exclusively of maternal origin [Pettigrew et.al., 1987; Clayton-Smith et. al., 1993].

Our hypothesis is that the detected chromosomal rearrangement is an interstitial triplication of the 15q11.2q13.3 region leading to tetrasomy (see Figure 17, allele ratio 3:1). Tetrasomy on chromosome 15 (15q11.2–15q13.3 region) is probably a very rare event. To our knowledge, only a few cases of Ito hypomelanosis have been reported in the literature, but they are associated with other molecular pathologies that do not involve interstitial triplications of the 15q11–q13 region [Holowinsky et. al., 1993, Schinzel et. al., 1994, Crawford et al. al., 1995, Cassidyet. Al., 1996, Chadwicket. al., 1996, Reddy et al. al., 2000, Ungaro et.al., 2001, Vialard et.al., 2003, Crolla et.al., 1995, Huang et.al., 1997].

On the other hand, the literature review showed that cases with hypomelanosis of Ito and hypopigmentation of the skin are exclusively associated with deletions in this region (15) (q11.2q13.3) [Hogart et.al., 2010, Turleau et. al., 1986, Pellegrino et. et al., 1995]. Triplications of about 6.8Mb located between repeats BP2 and BP4 on chromosome 15 result in tetrasomy of the affected region [Locke et. al., 2004]. At the proximal end of the long arm of chromosome 15, a group of low copy repeats (LCR) located at breakpoints BP1 - BP5 is located. These repetitive motifs mediate various deletions and duplications through allelic non-homologous recombination [Pujana et.al., 2002]. The duplication detected in the present study is located within the BP2 - BP3 critical region associated with Prader-Willi / Angelman syndromes, both caused by deletions or uniparental disomy in the target region.

To our knowledge, this is the first genetically proven case of hypomelanosis of Ito caused by a de novo interstitial 15q11.2q13.3 triplication. The case was published in the Journal of Clinical & Medical Genomics [Mladenova M et al., 2019].

The second patient, in whom no mutations were detected in the *FGFR3* and *FGFR2* genes, but with interesting genetic variants found after further genetic testing, was referred with a presumptive clinical diagnosis of skeletal dysplasia. The patient is a six-year-old girl with congenital anomalies, dysmorphic features, craniofacial synostosis and retardation in psychomotor development. The previously performed cytogenetic examination showed no deviations from the normal profile: 46, XX. In order to discover the genetic cause of the established clinical diagnosis, it was decided to conduct a subsequent extended study using the whole exome sequencing (WES) method. This method enables multiple genes associated with skeletal dysplasias to be analyzed simultaneously in a short period of time.

After performing whole exome sequencing and analysis of the obtained data, a genetic variant presented in Table 7 was discovered.

Table 7. Results of the performed exome sequencing.

Gene	Variant (UCSC,hg19)	Nucleotide substitution	Aminoacif substitution	Codon	Zygosity
MN1	Chr.22:g.28192789 C>T	NM_002430.3: c.3743G>A	p.Trp1248Ter	Stop codon	Heterozygous (De novo)

The detected genetic variant c.3743G>A, p.Trp1248Ter in the *MN1* gene was of the nonsense mutation type and was not reported in the databases at the time of analysis. As a result of the nucleotide substitution, a stop codon is generated in exon 1 of the gene at position 1248 in the amino acid sequence, which probably leads to premature termination of protein synthesis.

In order to clarify the origin of the mutation, a segregation analysis was performed in the family of the affected child. The obtained molecular genetic results showed that the variant c.3743G>A, p.Trp1248Ter in the *MN1* gene was not inherited from the parents, but arose de novo.

The *MN1* gene encodes a transcriptional cofactor that was first identified as an oncogene in meningioma [Lekanne et al., 1995]. It interacts with other transcription factors

that attach to specific regions of DNA and influence the expression and activity of certain genes. The transcriptional cofactor *MN1* is believed to act as a regulator of the activity of a number of genes required for skull and brain development. Because *MN1* does not contain a DNA-binding domain and *MN1* itself has transactivating activity, *MN1* likely activates transcription by binding to other DNA-binding transcription factors that can bind enhancer and/or promoter regions. MN1 is involved in transcriptional regulation by interacting with the transcription factors *PBX1, PKNOX1* and *ZBTB24*. Mutant *MN1* impairs binding to ZBTB24 and *RING1*, which is an E3-ubiquitin ligase. Based on literature studies, the model that the C-terminal deletion interferes with the interacting molecules of *MN1* related to the ubiquitin-mediated proteasome pathway, including *RING1*, and increases the amount of the mutant protein; this increase leads to dysregulation of *MN1* target genes by inhibiting rapid turnover of MN1 protein [Miyake et al., 2021] (Figure 18).

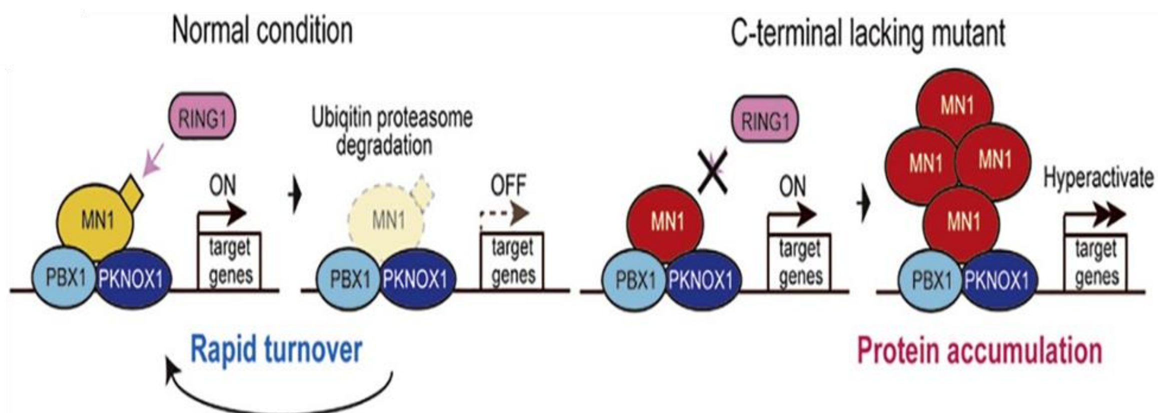


Figure 18. Pathomechanism resulting from C-terminally defective *MN1* [Noriko Miyake et.al.,2021]

In our case, the stop codon occurs at position 1248, which is located in the - 3′ end region of exon 1 of the MN1 gene (Figure 19).

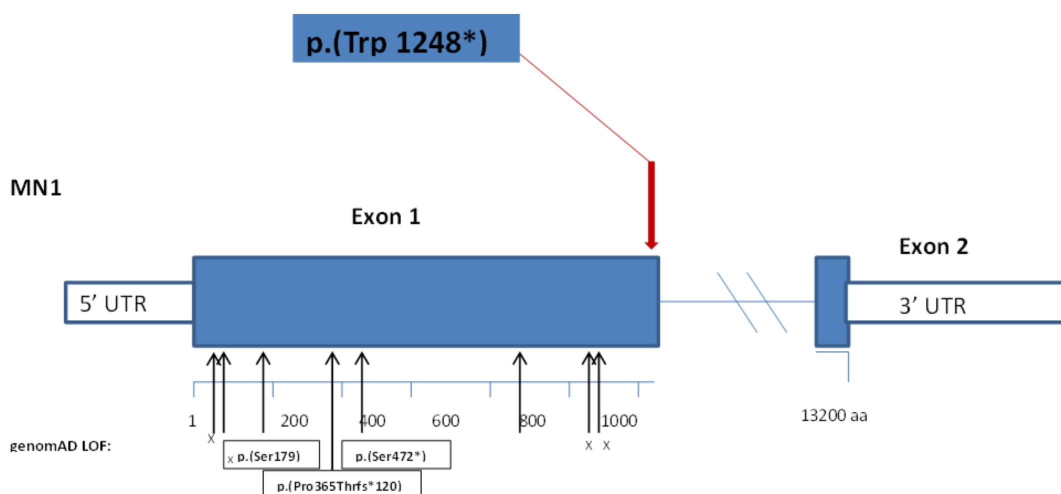


Figure 19. Localization Distribution of the C-terminal mutation c.3743G>A, p.Trp1248* identified in the MN1 gene in the proband (red arrow). The positions of loss-of-function variants (p.Ser179*, p.Pro365Thrfs*120 and p.Ser472*) described in the gnomAD database are shown below the gene scheme with black arrows.

All variants leading to the generation of a premature stop codon in the last 37 nucleotides of exon 1 or the final exon 2 (Figure 19) are expected to result in MN1 transcripts that avoid nonsense-mediated mRNA decay. This suggests that genetic variants of *MN1* affecting this region may lead to the expression of a C-terminally truncated protein with a pathogenic effect. The unique phenotypes associated with C-terminal truncation mutations that avoid nonsense-mediated mRNA decay have also been reported for several other genes [White, 2016]. All reported cases of C-terminal truncation syndrome are due to nonsense mutations in the region where our mutation falls. It can be assumed that one of the reasons for the occurrence of mutations in this region of the *MN1* gene (at the extreme 3′ end of exon 1 or in exon 2) is the nucleotide composition of this region. The p.Trp1248Ter mutation is generated by a G>A transition at a CpG dinucleotide. This mutational profile is overrepresented among human germline mutations [Acuna-Hidalgo, 2016] and there are no other CpGs regions with the potential to generate a stop codon by a C>T substitution in the last 55 nucleotides of exon 1 or in exon 2 of *MN1*. Pathogenic heterozygous variants in the MN1 gene have been shown to cause C-terminal truncation syndrome (MCTT) (RMID:31834374)

MCTT syndrome occurs in individuals with the following clinical findings:

1) Intellectual disorders with a pronounced speech delay

2) Hypotonia

3) Delay in motor development

4) Hearing loss (conductive or sensorineural)

5) Distinctive craniofacial dysmorphism

The disease has an autosomal dominant pattern of inheritance. Most of the heterozygous variants associated with MCTT syndrome are de novo mutations. An exception is a report of a family in which two affected brothers inherited the mutation from a mildly affected father who carried the mutation in mosaic form. Deletions including the entire MN1 gene sequence are associated with a variable clinical phenotype, including neurological abnormalities, facial dysmorphism, congenital heart defects, etc. [Said, 2011; Davidson, 2012; Mak, 2020].

In support of the assumptions about the pathogenic nature of the variant are:

- ✓ Previous reports on the pathogenic nature of stop codon variants in the MN1 gene (PMID:31834374)
- ✓ Stop codon variants in this region of the gene have been reported as pathogenic in the LOVD database (eg: p,Glu1249fsTer, p.Glu1260Ter, p.Gln1273Ter, etc.)
- ✓ De novo variant origin
- ✓ The absence of the variant from all control populations of the gnomAD project

The genetic variant c.3743G>A in the *MN1* gene found in our patient has not been reported in the world database so far.

1.6. PRENATAL DIAGNOSIS

9 cases were referred for genetic testing of the *FGFR3* gene for the purpose of prenatal diagnosis. 6 of them had ultrasound evidence of foetal bone growth retardation without a family history; in the remaining 3 cases there was a family burden.

Table 8 presents the clinical data and the reason for conducting antenatal invasive procedure, as well as the results of the genetic screening for mutations in the *FGFR3* gene.

Table 8. Clinical data of the patients examined in the prenatal period and genetic findings.

№	**Gestation week***	**Clinical data**	**Results– *FGFR3* gene**
1	**23^{+1}**	The mother has a genetically confirmed diagnosis of hypochondroplasia (mutation FGFR3:c.1620C>A; p.Asn540Lys)	Normal
2	**20**	Family history, child with genetically verified achondroplasia (mutation FGFR3:c.1138G>A;p.Gly380Arg)	Normal
3	**20**	The father and sister have a genetically verified diagnosis of hypochondroplasia (mutation FGFR3:c.1620C>A;p.Asn540Lys)	Normal
4	**12^{+5}**	US findings of a narrow chest and severe micrognathia	c.1108G>T
5	**20^{+6}**	US data on a discrepancy in the development of the bones of the fetus compared to the gestational week	Normal
6	**26**	US data showing shorter limbs relative to gestational week	Normal
7	**15**	US data showing shorter limbs relative to gestational week	Normal
8	**17**	US data for narrow chest, enlarged head	Normal
9	**21^{+1}**	US data on macrocephaly	Normal

* The gestational week in which the invasive manipulation (amniocentesis or chorionic biopsy) was performed to extract DNA material from the foetus.

In the **first** case of prenatal diagnosis performed in a family with a family history, the risk of an affected embryo was 50% because the mother carried the FGFR3:c.1620C>A mutation; p.Asn540Lys and herself diagnosed with hypochondroplasia. The analyses

performed showed that the foetus was a healthy non-carrier of the maternal mutation (Figure 20).

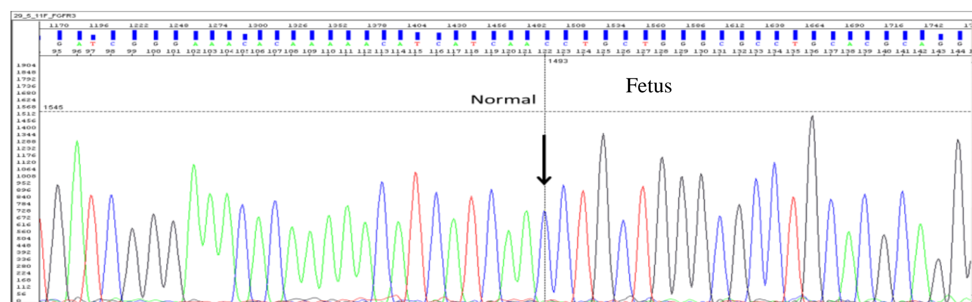


Figure 20. Sequential profile of prenatal diagnosis in a family with hypochondroplasia. The foetus did not inherit the familial mutation FGFR3:c.1620C>A, p.Asn540Lys**.**

At prenatal diagnosis of the **second** familial case with genetically verified achondroplasia in the family (FGFR3 mutation: c.1138G>A; p.Gly380Arg), no carrier of the familial mutation was detected in the foetus. The risk of recurrence of the mutation in a subsequent pregnancy in these cases was estimated at about 10-15%, due to the phenomenon of germline mosaicism, affecting mostly spermatogenesis.

The **third** case of prenatal diagnosis with a family history of hypochondroplasia was similar. Genetic testing ruled out carrier of the familial mutation FGFR3:c.1620C>A;p.Asn540Lys in the foetus.

In one of the cases, prenatal diagnosis was performed on the occasion of ultrasound findings of limbs with dimensions inconsistent with the gestational week, a narrow chest and severe micrognathia. In the study of the *FGFR3* gene, a pathological genetic variant c.1108G>T, p.Gly370Cys associated with Thanatophoric dysplasia type I was found. The results of the sequencing analysis are shown in Figure 21.

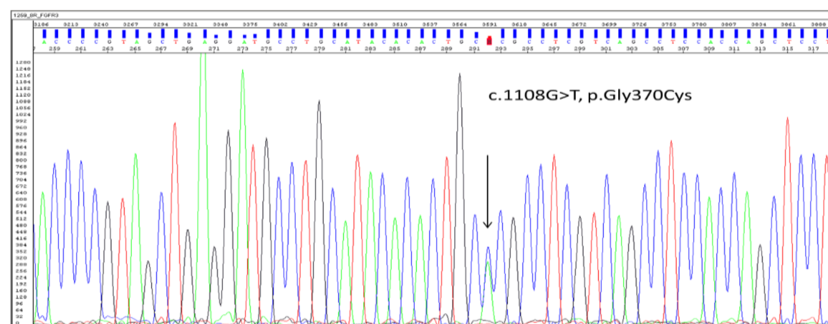


Figure 20. Sequence profile of the pathogenic genetic variant FGFR3:c.1108G>T, p.Gly370Cys from US-guided prenatal diagnosis of skeletal abnormalities.

Thanatophoric dysplasia is a severe form of skeletal dysplasia, characterized by multiple skeletal malformations, such as severely shortened limbs and the presence of additional skin folds on the arms and legs. Thanatophoric dysplasia type I is characterized by a normal skull and curved femurs and is associated with pathological genetic variants in the *FGFR3* gene. The detected variant c.1108G>T, p.Gly370Cys was located in exon 10 of the gene and was reported in the databases. The altered codon 370 falls within the region of the intracellular transmembrane kinase domain (TM) of the FGF receptor (see Figure 6 - in green), which functions by dimerization and post-phosphorylation.

The mutation c.1108G>T, p.Gly370Cys in the *FGFR3* gene is a variant that often arises de novo, therefore we performed analysis of the surrounding nucleotide sequence (Figure 42). Numerous G/T rich semi-perfect straight repeats as well as perfect straight repeats are found in the analyzed region. What is interesting in this case is that the mutant nucleotide falls in the center of a symmetrical element. This could explain the mechanism of occurrence of the mutation at this exact position. The ability of palindromes to initiate genetic recombination lies in their ability to form secondary structures in DNA that can cause replication delays and double-strand breaks. Given their recombinogenic nature, it is not surprising that palindromes in the human genome are involved in genetic rearrangements, translocations, and deletions associated with certain human syndromes (Figure 28). These repeated sequences could be involved in strand mispairing, polymerase slippage, and generation of mutational conversions. The analyzed repeats could be related to the formation of secondary structures hindering the normal functions of the polymerase, which could explain the occurrence of the mutation in a specific nucleotide position. The estimated pathogenicity coefficient for the c.1108G>T, p.Gly370Cys mutation from the PolyPhen2 predictor was 0.95, corresponding to a high probability of pathogenicity (Figure 29)

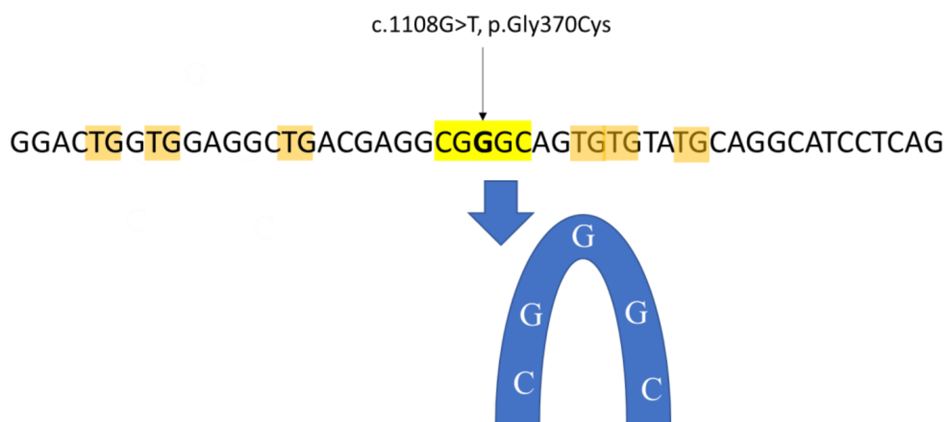
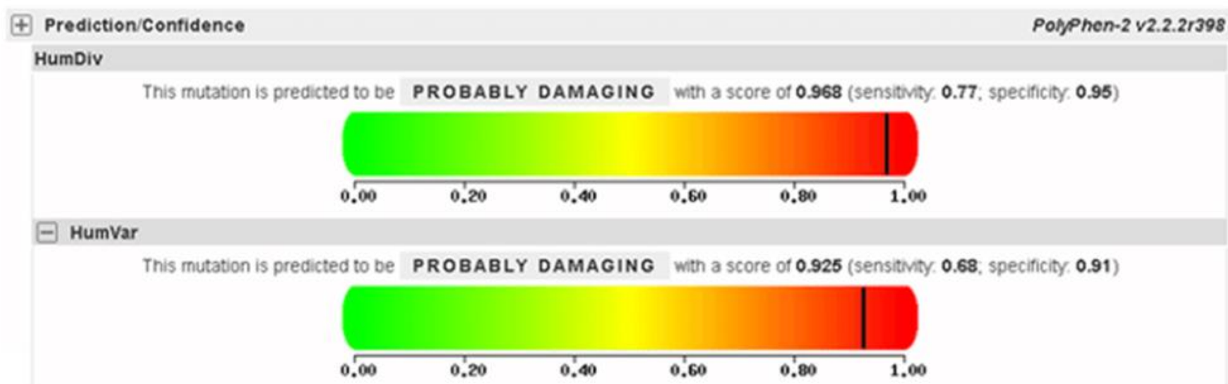
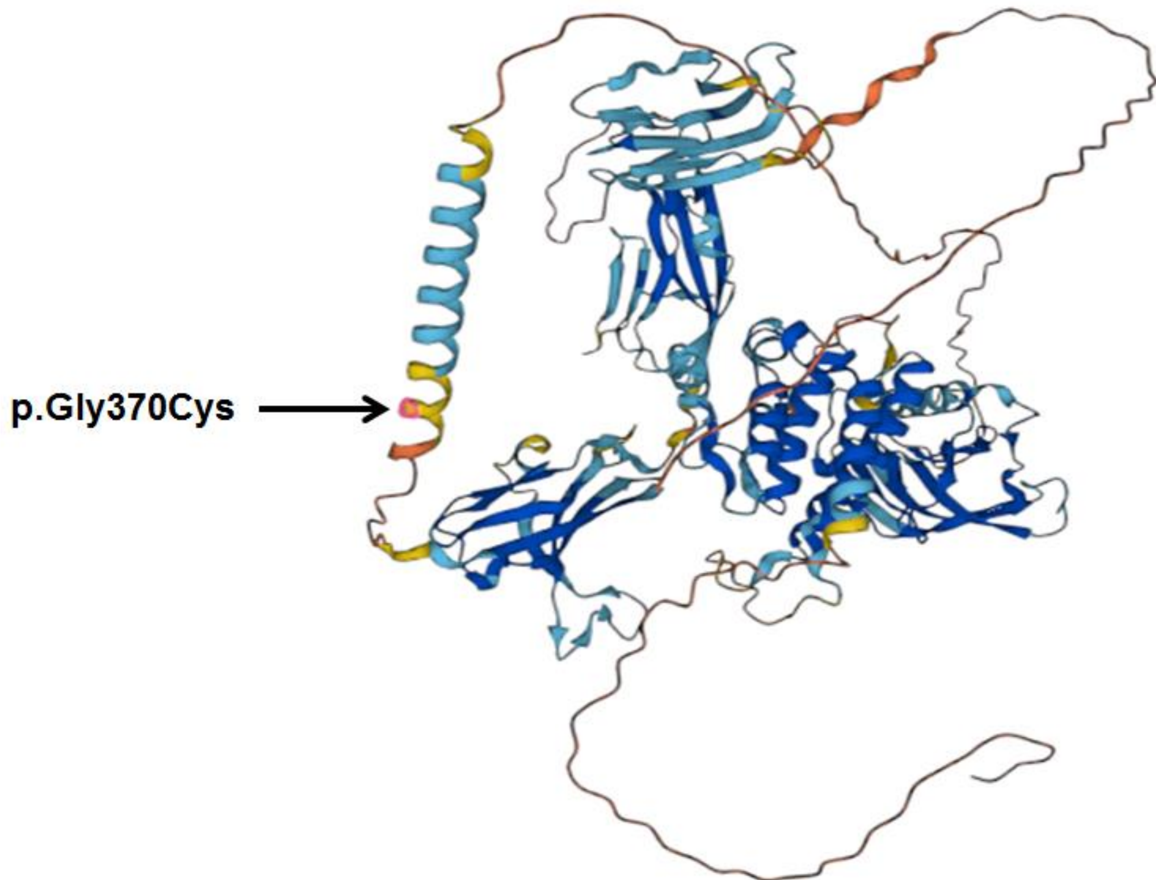


Figure 28. Analysis of the region of the FGFR3 gene in which the mutation c.1108G>T, p.Gly370Cys falls and a schematic representation of the palindromic sequence encompassing the triplet of the pathological mutation



A



B

Figure 29. A-Result of the online software predictor Polyphen2 for assessing the pathogenicity of the FGFR3 mutation: c.1108G>T, p.Gly370Cys; B- Spatial structure of the FGFR3 receptor modeled using AlphaFold2. The p.Gly370Cys substitution leading to thanatophoric dysplasia is marked with an arrow.

In the amino acid sequence of the receptor at position 370 is located the amino acid glycine, which has no side radical (one H atom) and is nonpolar. As a result of the substitution, the polar amino acid cysteine is incorporated, which has a side radical containing a sulfur atom as part of the sulfhydryl group (-CH2-SH), which is reactive and participates in the construction of intra- and intermolecular disulfide bridges. This puts cysteine in a special position when it is incorporated at the wrong place in the amino acid sequence, since disulfide bridges are essential for the formation of tertiary protein structure. This is an additional hypothesis as to why the amino acid substitution p.Gly370Cys has such a pathological effect on protein structure and/or function and manifests itself with the observed severe clinical picture.

In addition, no mutations in the *FGFR3* gene were detected in the remaining five examined foetuses with evidence of skeletal dysplasia in the US.

DISEASES WITH DISORDERS IN THE CONNECTIVE TISSUE

Based on the preliminary clinical data, the 25 postnatally examined patients were divided into gene groups corresponding to the given anamnesis and clinical diagnosis.

Table 9 presents the groups of patients studied respectively for the genes *COL1A1, COL1A2, COL2A1, COL11A1*, grouped by gender.

Table 9. Distribution of patients by genes, by gender and number.

	Male	Female	Total number
COL1A1 / COL1A2	12	11	23
COL2A1 / COL11A1	1	1	2

SYNDROMES ASSOCIATED WITH COL1A1 / COL1A2 GENES (OSTEOGENESIS IMPERFECTA)

Patients with a clinical diagnosis falling into the group of syndromes associated with the *COL1A1 and COL1A2* genes were a total of 23 in number (Table 10).
After conducting molecular genetic analyses concerning the *COL1A1* gene, 3 mutations associated with Osteogenesis imperfect were found (table 11). 12 patients remained with an unclear molecular defect, and one of them had an unclear clinical picture resembling Osteogenesis imperfecta. This patient was tested for all exons of the *COL1A1* gene and is subject to further genetic studies in the future.

Table 11. Presentation of the results obtained after analysis of the COL1A1 gene.

Пациент №	Ген	Екзон	Нуклеотидна Замяна	Аминокиселинна Замяна	Диагноза
1*	*COL1A1*	1-51	c.2424delC	p.Gly809AlafsTer299	Osteogenesis imperfecta
2*	*COL1A1*	36	c.2424delC	p.Gly809AlafsTer299	Osteogenesis imperfecta
3*	*COL1A1*	36	c.2424delC	p.Gly809AlafsTer299	Osteogenesis imperfecta
4*	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
5**	*COL1A1*	1-51	c.2784delT	p.Gly929AlafsTer179	Osteogenesis imperfecta
6**	*COL1A1*	40	c.2784delT	p.Gly929AlafsTer179	Osteogenesis imperfecta
7**	*COL1A1*	40	c.2784delT	p.Gly929AlafsTer179	Osteogenesis imperfecta
8**	*COL1A1*	40	c.2784delT	p.Gly929AlafsTer179	Osteogenesis imperfecta
9***	*COL1A1*	7	-	-	Osteogenesis imperfecta
10***	*COL1A1*	7	c.579delT	p.Gly194ValfsTer71	Osteogenesis imperfecta
11***	*COL1A1*	7	-	-	Osteogenesis imperfecta
12	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
13	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
14	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
15	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
16	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
17	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
18	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
19	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
20	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
21	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
22	*COL1A1*	1-51	-	-	Osteogenesis imperfecta
23	*COL1A1*	1-51	-	-	Osteogenesis imperfecta

*** Family №1 / ** Family №2 / *** Family №3**

Family #1

The family was referred for genetic counselling due to the birth of a child with a clinical diagnosis of Osteogenesis imperfecta. The mother of the index patient and her father also had clinical evidence of this condition. Genetic testing of the COL1A1 gene was performed and the index patient was found to be a heterozygous carrier of the mutation p.Gly809AlafsTer299 in exon 36 of the COL1A1 gene. The genetic finding is of the type of mutations with

displacement of the reading frame (so-called frameshift mutation) and deletion of one or several nucleotides. This change can result in a protein with an altered amino acid sequence or a shorter protein resulting from the occurrence of a premature stop codon due to a reading frame shift. The newly discovered variant p.Gly809AlafsTer299 in the index patient represents a deletion of one C nucleotide at position 809 in the protein, leading to a shift in the reading frame and premature termination of protein synthesis after codon 299, due to the generation of a new stop codon.

In order to clarify the origin of the newly discovered mutation, a segregation analysis was performed in the family of the affected child. Molecular genetic analysis was performed targeting exon 36 only, in order to search for this molecular defect. Heterozygous carrier of the pathological genetic variant COL1A1:p.Gly809AlafsTer299 was demonstrated in the child's mother and maternal grandfather, who reported frequent fractures. This is consistent with the expected autosomal dominant pattern of inheritance (Figure 19)

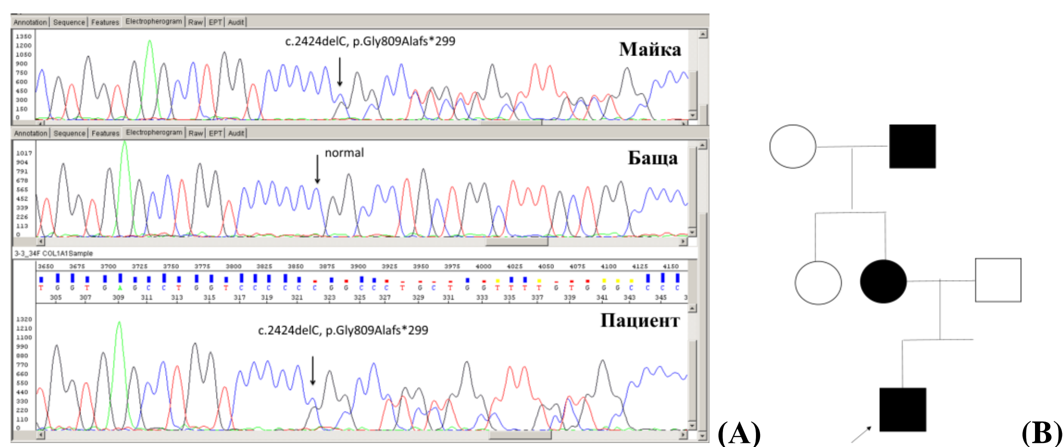


Figure 19.

(A) Sequence profile of the familial mutation COL1A1:p.Gly809AlafsTer299;

(B) Pedigree tree of the family, arrow shows the proband.

Family #2

The proband of the family was a 10-year-old child referred with a clinical diagnosis of Osteogenesis imperfecta type I due to the presence of multiple long bone fractures occurring spontaneously or with minor trauma. The family history is positive with multiple fractures reported in the mother, uncle and their father.

Molecular genetic testing of the *COL1A1* gene was performed in the index patient and a heterozygous c.2784delT, p.Gly929AlafsTer179 mutation in exon 40 of the gene was detected. The genetic variant detected is a single T nucleotide deletion at position 2784 of the gene, resulting in a frameshift and generation of a premature stop codon 179 codons further in the protein sequence. The detected pathogenic variant in the *COL1A1* gene has been reported in databases and is associated with the clinical diagnosis of Osteogenesis imperfecta type I.

In order to clarify the origin of the newly discovered mutation, a segregation analysis was performed in the family of the affected child. The child's mother, maternal uncle, and maternal grandfather were found to be carriers of the same genetic variant, which matched the initially presented clinical data of frequent fractures among relatives. The sequence profile of the mutation and the family tree are shown in Figure 20

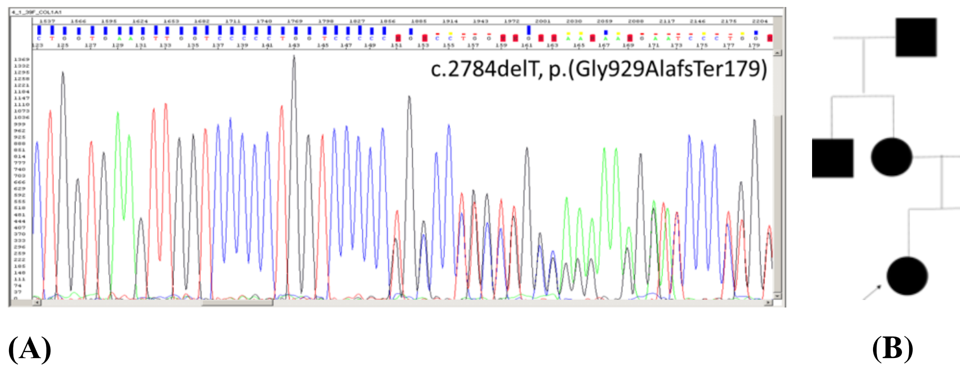


Figure 20.

(A) Sequence profile of the familial COL1A1 mutation: p.Gly929AlafsTer179;

(B) Pedigree tree of the family, index patient indicated by arrow

Family #3

The index patient from family #3 was a 1-year-old child referred with a clinical diagnosis of Osteogenesis imperfecta type I because of the presence of multiple long bone fractures. The molecular genetic study of the *COL1A1* gene showed a heterozygous carrier of the genetic variant p.Gly194ValfsTer71 in exon 7 of the gene.

The detected variant p.Gly194ValfsTer71 represents a deletion of one T nucleotide at position 579 of the gene, leading to a shift in the reading frame and generation of a premature stop codon 71 codons upstream. The detected pathogenic variant in the *COL1A1* gene was reported in the ClinVar database and was associated with the clinical diagnosis of Osteogenesis imperfecta type I.

In order to clarify the origin of the detected mutation, a segregation analysis targeting only exon 7 was performed in the family of the affected child. The mother and father were found not to carry this pathogenic variant and therefore the detected genetic **defect arose de novo in the child (Figure 21).**

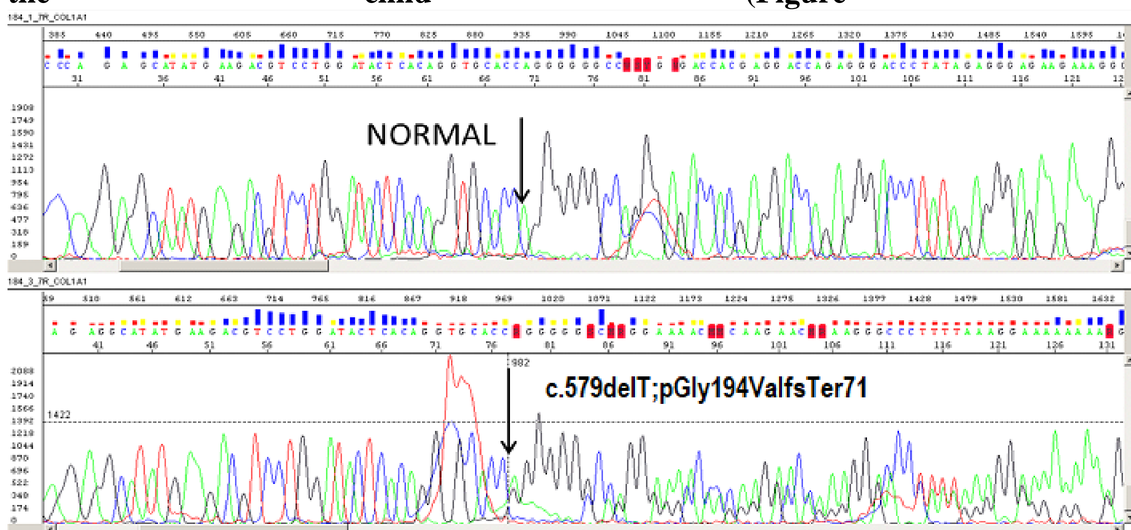


Figure 21. . Sequence profile of the de novo COL1A1:c.579delT mutation in COL1A1 in the studied family

The variants found in the *COL1A1* gene in the three studied families were of the same type of frameshift mutations, resulting in the deletion of one nucleotide and the generation of a premature stop codon in the new reading frame. The analysis of the environment of the pathological sequence variant COL1A1:p.Gly809AlafsTer299 found in family #1 showed repetitive perfect straight repeats (TGG and TT), a semi-symmetrical element in the mutation region -GTCCCCCCGGCCCTG-. From the schematic representation of the semiperfect repeat it is seen that the site of the doublet nucleotide is at the top of the secondary structure formed and the mechanism by which this hot spot for mutation occurs is most likely the purpose of the DNA polymerase to create absolute symmetry in the region. (Figure 22 A). In addition, the six-stranded sequence from the C-rich region is present, with one of these C nucleotides missing as a result of the deletion.

The COL1A1:p.Gly929AlafsTer179 mutation found in family #2 showed the presence of repeated perfect straight GGT repeats (Figure 22B) flanking the mutation site. In addition, the pentaplex C-rich sequence flanking the position of the deleted T nucleotide is present. Notably, the entire region surrounding the mutation site is rich in repetitive elements, including a perfectly symmetrical CCCCCTGGTCCCCC element flanked by straight repeats, which could further complicate the work of DNA polymerase.

Analysis of the flanking sequence of the COL1A1:p.Gly194ValfsTer71 mutation found in family #3 (Figure 22B) showed that the mutation site was again preceded by a hex-stranded sequence rich in nucleotide C and surrounded by the triplet TGG - analogously to the frameshift mutations found in the previous two families (Figure 22)

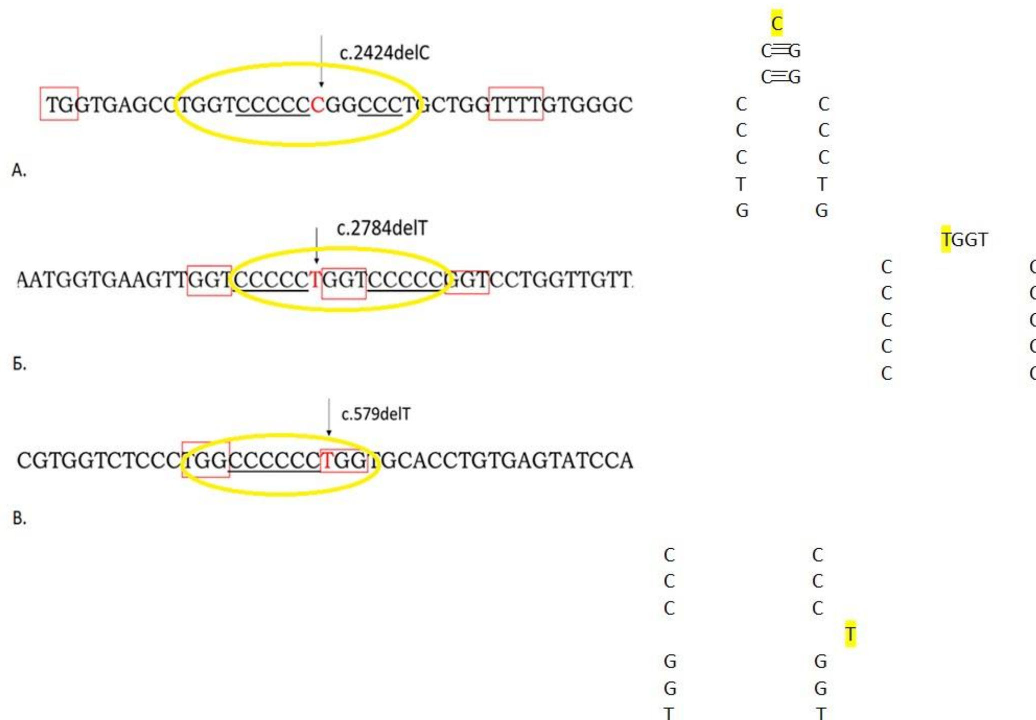


Figure 22.
Analysis of the regions of the COL1A1 gene in which the three mutations found in Bulgarian patients with Osteogenesis imperfecta fall.

A. Analysis of the surrounding sequence around the site of the COL1A1:p.Gly809AlafsTer299 mutation found in family #1. And schematic representation of the symmetrical element in which the site of mutation also falls.
B. Analysis of the surrounding sequence around the COL1A1:p.Gly929AlafsTer179 mutation site found in family #2. And schematically presented the symmetrical element, which also includes the place of occurrence of mutation.

C. Analysis of the surrounding sequence surrounding the site of the COL1A1 mutation: p.Gly194ValfsTer71 found in family #3. And schematically presented the symmetrical element, which also includes the place of occurrence of mutation.3.

In all three families, analysis of the sequence surrounding the mutation showed that the site of the mutation was preceded by a C-rich motif and that repeated TG-rich motifs and perfect or semi-perfect symmetry elements abounded (Figure 47). These motifs most likely favour the formation of secondary structures, hindering the work of DNA polymerase and leading to its slipping and skipping of one nucleotide. These secondary structures probably make the omission of a single nucleotide difficult to detect by DNA-repair systems, which is probably relevant to the fixation of the mutation in the genome.

The most common type of COL1 pathogenic variant in Osteogenesis imperfecta affects the triple helical structure of the COL1A1/2 chains. Type I collagen is a heterotrimer consisting of two alpha-1 chains and one alpha-2 chain. Triple helical domains are composed of continuous repeats of the tripeptide Gly-X-Y, where X and Y can be any amino acid. X is often proline (Pro) and Y is often lysine (Lys). Triple helix formation requires a glycine residue in every third position of the chains, as glycine is the only residue small enough to fit into the confined space on the inside of the helix [Zhytnik et.al., 2019]. Any mutation that disrupts the Gly-X-Y sequence of the collagen triple helix will result in impaired collagen function. Such mutations disrupt triple helix formation, its stability, and its interactions with the extracellular matrix.

As a general rule, phenotypic severity depends on the affected alpha chain, the position of the mutation, the amino acid substitution, or a combination of these three variables. Two main types of mutations in type I collagen are associated with Osteogenesis imperfecta. The first type of mutation results in haploinsufficiency, usually associated with a mild form of Osteogenesis imperfecta type I. Haploinsufficiency is usually the consequence of a nonsense mutation or a frameshift mutation that introduces a premature termination codon into the coding sequence of one allele of the COL1A1 gene, as was the case with our patients. This initiates nonsense-mediated mRNA decay [Marini et. al., 2017]. The result of these mutations is a quantitative defect in type I collagen, as only about half the normal amount of protein is produced. Despite the reduced amount, the type I collagen produced has a normal structure.

The second type of mutations are abnormalities in the nucleotide sequence that lead to structural changes in collagen type I. The most common mutations leading to Osteogenesis imperfecta represent the replacement of the most common amino acid in the collagen chain, glycine, with another amino acid. Any mutation that disrupts the Gly-X-Y sequence of the collagen triple helix will result in impaired collagen structure and/or function [Rauch et. al., 2010]. Very few mutant collagen monomers are required to disrupt the structure of the various multitrimeric assemblies. Thus, mutations affecting glycine in type I collagen exert a so-called dominant negative effect.

The phenotype-genotype correlation in Osteogenesis imperfecta type I can be summarized in the following points [Rauch et. al., 2010]:

1) The COL1 pathogenic variant generally affects triple-stranded helical organization;

2) Genetic variants can cause quantitative changes in the protein product associated with milder clinical symptoms;

3) Genetic variants can cause structural changes in the collagen chain, associated with more severe and even lethal clinical symptoms;

4) The severity of the phenotypic manifestation depends on the affected alpha chain;

5) The severity of the clinical symptomatology depends on the substituted amino acid and on the type of mutation that occurred;

6) The severity of the phenotypic expression is determined by whether one or both copies of the gene are affected.

In patients with a suspected clinical diagnosis of Osteogenesis imperfecta who were included in our studies and tested negative for pathogenic genetic variants in the *COL1A1* gene, additional testing of the *COL1A2* gene was performed. The results here were also negative, no pathogenic variants were found in the *COL1A2* gene.

One patient in our sample was referred with a presumptive clinical diagnosis of Osteogenesis imperfecta, due to multiple fractures. It was initially screened for the *COL1A1* gene, but as no pathogenic variants were found, we continued with the *COL1A2* gene, which was also negative. Subsequently, whole exome sequencing (WES) was performed, analyzing a panel of genes associated with the occurrence of collagenopathies. After next-generation sequencing, again no pathogenic variants were detected, but in the course of the data analysis, it was found that certain regions of the X chromosome lacked coverage (Figure 48) Such a problem in the data could be due to a misalignment or other error during the bioinformatic analysis of the data, as well as being an indication of the presence of a deletion in the X chromosome region, especially in cases of examination of male patients (such as is also our patient). In order to rule out or confirm this suspected hemizygous microdeletion of the long arm of the X chromosome in region 23 (Xq23), array CGH analysis was performed. The results of the performed comparative genomic hybridization showed the presence of a deletion on the X-chromosome Table 12. The deletion is 870,743,253 kb in size and includes two genes associated with human diseases PLS3 (OMIM*300131) and SLC6A14 (OMIM*300444).

Table 12. The results of the conducted array-CGH analysis.

Gene	Variant (UCSC,hg19)	Zygosity/ pathogenic
PLS3, SLC6A14	Arr[GRCh37]Xq23(114,799,463-115,670,182)x0	Chromosomal microdeletion / Class 4: probably pathogenic variant

Evidence regarding the pathogenic nature of such deletions in the Xq23 region is limited. According to literature data and mutation databases, no patients with the same deletion of the Xq23 region have been reported. The proven deletion involves the PLS3 gene, in which loss-of-function variants are associated with X-linked dominant osteoporosis and susceptibility to

osteoporotic fractures with high penetrance in males and variable penetrance in females (OMIM:300910) [Van Dijk et. al., 2013]. In addition to point variants, hemizygous Xq23 deletions involving the entire PLS3 gene have been reported in two men with bone fractures [Manny et. al., 2017, Kampe et al., 2017]. In support of the pathogenic nature of the variant and its association with the patient's clinical presentation, we can present the following evidence:

1) The complete absence of such deletions in the control groups of healthy people

2) The expected loss of function of genes falling within the hemizygous deletion of this chromosomal region

3) The correspondence between the effect of the variant with the established mechanism of PLS3-related diseases

4) Correspondence of the variant with the clinical presentation of the patient

5) The presence of the variant in a hemizygous state consistent with the inheritance pattern of PLS3-related diseases.

SYNDROMES ASSOCIATED WITH COL11A1 / COL2A1 GENES

There were a 2 patients with a clinical diagnosis falling into the group of syndromes associated with *COL11A1* and *COL2A1* genes - a father and a daughter.

The index patient was a 2-year-old girl with hypertelorism, midface hypoplasia, a small saddle nose, and craniofacial dysplasia. At the time of examination, the father was 38 years old, with the same clinical characteristics as the proband, except that he had hearing loss and moderate mental retardation. Both the father and the proband were clinically diagnosed with Marshall-Stickler syndrome.

Since Marshall-Stickler syndrome is most commonly associated with mutations in the *COL2A1* gene, Sanger sequencing of this gene in the family was performed as a first step in the analyses, but no pathogenic variants were detected. Sequencing of the second candidate gene *COL11A1* was undertaken. In the index patient, the c.3474+1G>A mutation in intron 44 of the *COL11A1* gene, not published to date in the databases, was detected. Segregation analysis performed showed that the father was also a carrier of the aforementioned variant and therefore the variant segregated with the disease in the family.

The COL11A1:c.3474+1G>A variant is a splice-site mutation affecting the donor consensus splicing sequence. Mutations that occur at the splice sites disrupt the signal that is required for the correct excision of the non-coding intronic sequences in the gene. Splice-site mutations can occur in the GT-sequence, which as a rule defines the 5'-donor splice site, or in the AG-sequence, which defines the 3'-acceptor site. In addition to these positions, mutations can also occur in sequences close to the donor or acceptor site, which again have an effect on splicing. If one of these signals is modified as a result of a mutation, then the introns cannot be cut out correctly and some of them remain in the mature mRNA (insertions) or an entire exon is cut out (deletions). This type of defect can lead to the breakdown of the damaged transcript or to the production of a protein product with a disturbed amino acid sequence.

The c.3474+1G>A variant affects a donor splice site of the triple-helical domain, which is a known "hot spot" for mutations in the *COL11A1* gene [Herrmann et. et al., 1975]. Analysis of the surrounding sequence indicated the presence of repetitive perfect straight repeats (AAGG and TT) and in addition a trinucleotide palindromic sequence AAA/TTT located close to the position of the substitution (Figure 23). The repetitive elements and the palindromic sequence could be related to structures favouring the replacement of the purine guanyl nucleotide (G) with the purine adenyl nucleotide (A) and its fixation in the gene sequence. At this position, the work of the DNA polymerase is most likely impaired by the formation of secondary structural elements, making chemically similar substitutions difficult to recognize.

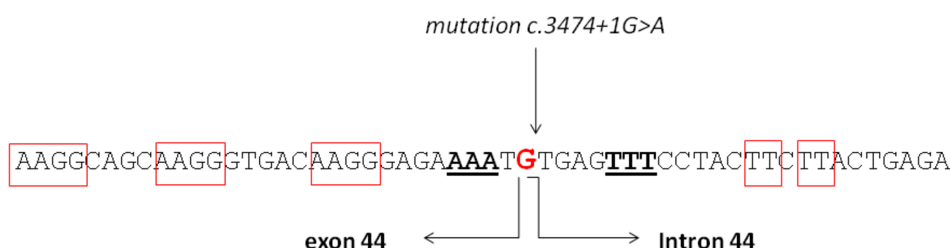


Figure 23. Analysis of the region of the *COL11A1* gene in which the mutation COL11A1: c.3474+1G>A occurs in patients with Marshall-Stickler syndrome.

In conclusion, it can be summarized that the unpublished pathological splice variant c.3474+1G>A affecting the triple-helical domain in the *COL11A1* gene is the cause of the first genetically verified case of Marshall-Stickler syndrome in Bulgaria: a familial case affecting at least two generations [Mladenova et al., 2021].

CONCLUSION

During the implementation of the present scientific work, methods were also introduced and optimized for the diagnosis of genes related to skeletal dysplasias and collagenoses (*FGFR3, FGFR2, COL1A1, COL1A2, COL11A1* and *COL2A1*). The developed molecular genetic approach was applied to study 75 patients, 66 of whom were screened postnatally and 9 prenatally. Protocols for amplification of target genes and subsequent analysis by sequencing were optimized.

The optimized protocols were applied to the diagnosis of the group of 66 patients in order to clarify the genetic causes of the observed clinical symptoms. Thirty-five patients were screened for mutations in the FGFR3 gene, six for mutations in the FGFR2 gene, twelve unrelated patients for *COL1A1/COL1A2* genes, and two related patients for *COL11A1/COL2A1*.

In 31 of these 66 patients (47%), the clinical diagnosis was confirmed at the molecular level, with both mutations published in the worldwide literature and mutations not published worldwide. In 27 patients (87%) known mutations were found:

► *FGFR3*: c.1138G>A, p.Gly380Arg; c.1138G>C, p.Gly380Arg; c.1620C>A, p.Asn540Lys;

► *FGFR2*: c.755C>G , p.Ser252Trp;

► *COL1A1*: p.Gly809AlafsTer299; p.Gly929AlafsTer179; p.Gly194ValfsTer71.

Previously unpublished mutations were found in 4 patients (13%):► 15q11.2q13.3 *de novo* interstitial triplication;

► *MN1*: c.3743G>A, p.Trp1248Ter;

► Hemizygous microdeletion of the long arm of the X chromosome (Xq23);

► *COL11A1*: c.3474+1G>A.

From the target group of patients referred with diagnoses related to cartilage tissue disorders, a total of 35 patients were examined. Detected mutations by gene are presented in Figure 25.

A

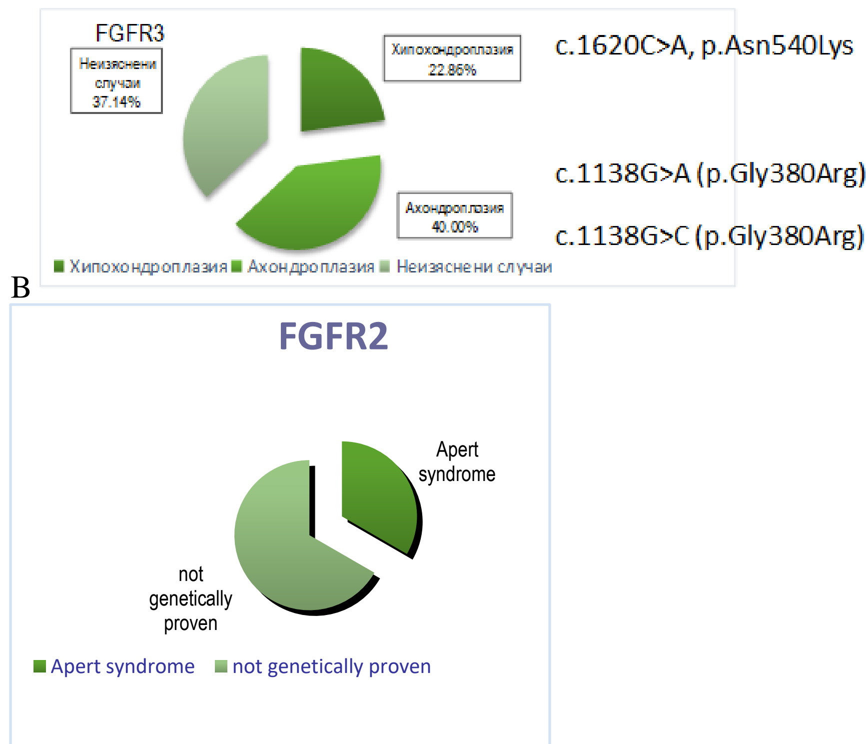


Figure 25. *Molecularly confirmed diagnoses versus unclear cases by genes in patients with an initial clinical diagnosis of skeletal dysplasia:*
A - Patients with FGFR3 gene analyses,
B - Patients with FGFR2 gene analyses;

Diagram "A" presents the percentage of proven cases of patients with a suspected clinical diagnosis of achondro/hypochondroplasia (associated with mutations in the FGFR3 gene) as well as the detected mutations. A total of 35 patients were examined in this group, and in 14 (40.00%) patients the diagnosis of Achondroplasia was proved at the molecular level (mutations FGFR3:c.1138G>A;p.Gly380Arg; FGFR3:c.1138G>C;p.Gly380Arg) , in 8 (22.86%) the diagnosis of Hypochondroplasia (mutations FGFR3:c.1620C>A;p.Asn540Lys) was proven at the molecular level. The remaining 13 (37.14%) cases presented with a clinical picture resembling skeletal dysplasia remained with an unclear molecular defect.

Chart "B" presents the percentage of proven cases of patients with a suspected clinical diagnosis of skeletal dysplasia. All of them were molecular tested for the exons of the FGFR2 gene, which are hotspots mutational associated with skeletal dysplasias. A total of 6 patients were researched in this group. In 2 (33%) the FGFR2:c.755C>G;p.Ser252Trp mutation leading to Apert syndrome was detected ,but 4 patients (67%) remained with an unclear molecular defect.

In addition, two of the patients with a suspected clinical diagnosis of skeletal dysplasia were genetically verified to have newly discovered pathogenic molecular defects. The first case concerns a patient referred with a suspected clinical diagnosis of Achondroplasia, after performing molecular genetic studies of the entire *FGFR3* gene, followed by sequencing of the entire *FGFR2* gene, in order to clarify the diagnosis, no pathological findings were found in either gene. After refining the clinical diagnosis and conducting additional molecular genetic studies, a de novo interstitial 15q11.2q13.3 triplication was found, leading to the manifestation of the disease Hypomelanosis of Ito.

The second case involved a patient with a suspected clinical diagnosis of Skeletal dysplasia, and similarly to the first case, no pathological variants were detected in the *FGFR3* and *FGFR2* genes. After refining the clinical diagnosis and conducting additional molecular genetic studies, a heterozygous carrier of the nonsense variant c.3743G>A, p.Trp1248Ter in the *MN1* gene was found, leading to a very rare C-terminal truncation syndrome with an autosomal dominant inheritance pattern.

The percentage ratio of detected pathologies in the studied group of 41 patients with diseases with disorders in the cartilage tissue is presented in Figure 26.

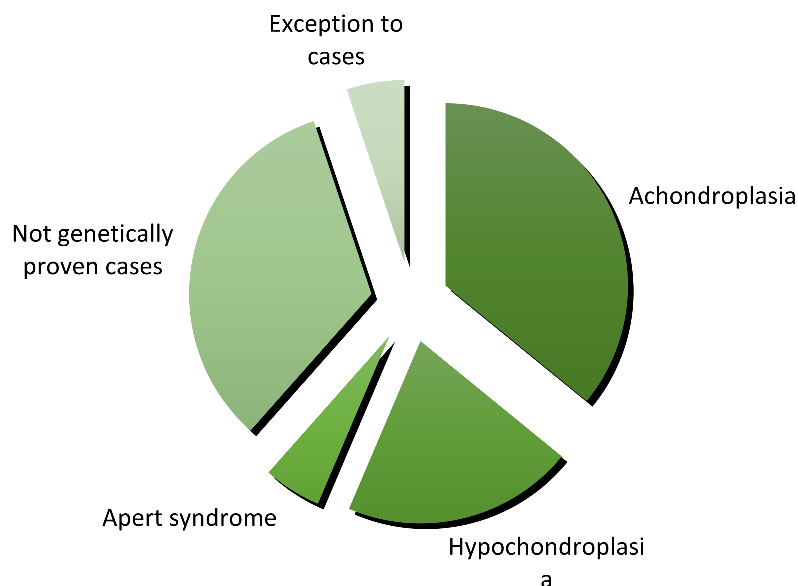


Figure 26. *Percentage ratio of molecularly clarified and unexplained pathologies involving cartilage tissue.*

For the purpose of **prenatal diagnosis**, 9 foetuses were referred for genetic testing of the FGFR3 gene Six of the cases were investigated on ultrasound (US) evidence of bone growth retardation, in the another three cases was a family history. In one embryo (12.5%), referred by ultrasonography for shorter limbs inconsistent with gestational week, narrow chest and severe micrognathia, the molecular defect FGFR3:c.1108G>T;p.Gly370Cys leading to Thanatophoric dysplasia type I was detected .

Thanks to the introduced methodology for the diagnosis of skeletal dysplasias, it became possible to conduct adequate and timely genetic counselling, prevention and prenatal diagnosis of the affected families.

From the target group of patients referred with suspected diagnoses involving connective tissue disorders, a total of 25 patients were examined, 16 of whom were unrelated. 3 pathologies with a diagnosis Osteogenesis imperfecta type I were proven at the molecular level

COL1A1:c.242delC;p.Gly809AlafsTer299;

*COL1A1:*c.2784delT;p.Gly929AlafsTer179;

COL1A1:c.579delT;p.Gly194ValfsTer71

Unpublished pathologic mutation COL11A1:c.3474+1G>A was found in one patient diagnosed with Marshall-Stickler syndrome. Additionally, one patient with an unclear clinical diagnosis of Osteogenesis imperfecta was found to have a hemizygous microdeletion of the long arm of the X chromosome in region 23 (Xq23), hitherto unpublished worldwide.

The percentage ratio of detected pathologies in the studied group of 13 unrelated patients is presented in Figure 27.

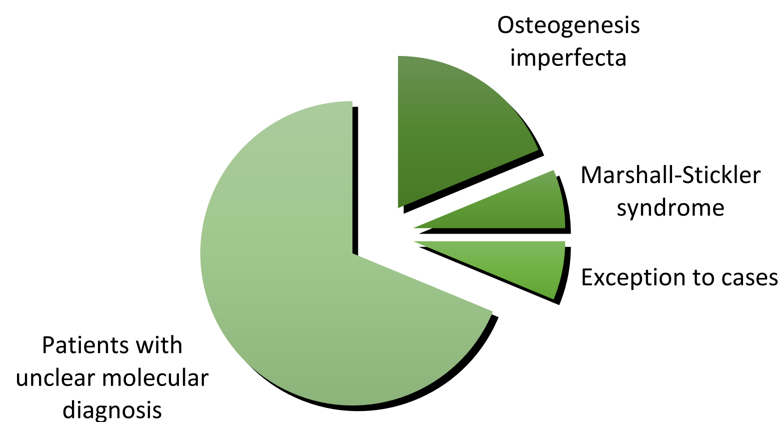


Figure 27. Percentage ratio of explained and unexplained pathologies at the molecular level among the group of collagenoses.

Thanks to the introduced methods for diagnosing diseases related to collagen disorders, it became possible to conduct adequate and timely genetic counselling, prevention and prenatal diagnosis in the affected families.

CONCLUSION

1. The selected panel of genes (*FGFR3, FGFR2, COL1A1, COL1A1, COL11A1* and *COL2A1*) for the analysis of patients suffering from cartilage and connective tissue disorders led to clarification of the molecular diagnosis in 47% of cases.
2. Introduced and optimized screening methods were successfully used to detect genetic variants in target genes.

3. Genetically verified diagnoses with disorders in cartilage tissue are as follows:
   A total of 35 patients were examined in this group, as
   - ➢ In 14 (40.00%) patients the diagnosis of Achondroplasia was proved at the molecular level (mutations FGFR3:c.1138G>A;p.Gly380Arg; FGFR3:c.1138G>C;p.Gly380Arg),
   - ➢ In 8 (22.86%) the diagnosis of Hypochondroplasia (mutations FGFR3:c.1620C>A;p.Asn540Lys) was proved at the molecular level. The remaining 16 (37.14%) cases presented with a clinical picture resembling skeletal dysplasia remained with an unclear molecular defect.
   - ➢ In 2 (33%) the FGFR2:c.755C>G;p.Ser252Trp mutation leading to Apert syndrome was found.
   - ➢ The obtained results are in accordance with the literature data, as the most common non-lethal skeletal dysplasia is achondroplasia caused by the mutation p.Gly380Arg in the FGFR3 gene.
4. The most common mutation causing hypochondroplasia in our sample was p.Asn540Lys in the FGFR3 gene, which is consistent with published data.
5. The p.Ser252Trp mutation in the FGFR2 gene is associated with Apert syndrome in Bulgarian patients as well, in agreement with published data.
6. The first case of Ito's hypomelanosis associated with interstitial 15q11.2q13.3 triplication in Bulgaria was discovered and published, which is also the first proven case of triplication worldwide.
7. The first case of de novo heterozygous variant c.3743G>A, p.Trp1248fsTer in the MN1 gene associated with C-terminal truncation syndrome in Bulgaria was discovered and published.
8. . The genetically verified diagnoses with disorders in the connective tissue are as follows: In this group, a total of 25 patients were examined, 16 of whom were unrelated.

Three pathologies (18.75%) with a diagnosis of Osteogenesis imperfecta type I were proven at the molecular level

- ➢ *COL1A1*:c.242delC;p.Gly809AlafsTer299;
- ➢ *COL1A1:*c.2784delT;p.Gly929AlafsTer179;
- ➢ *COL1A1*:c.579delT;p.Gly194ValfsTer71

9. Three pathological frameshift mutations in the COL1A1 gene associated with Osteogenesis imperfecta were found in Bulgarian patients.
10. In all frameshift mutations in the COL1A1 gene, the site of mutation is preceded by the presence of perfect or semiperfect symmetrical repeats.
11. The first family case of Marshall-Stickler syndrome, genetically verified in Bulgaria, associated with a mutation in the COL11A1 gene, was discovered and published.
12. A previously unpublished X-chromosomal deletion encompassing the PLS3 and SLC6A14 genes was also found to be associated with X-linked dominant osteoporosis and susceptibility to high-penetrance osteoporotic fractures in males.

13. On the basis of the obtained results of the molecular genetic studies, it is possible to carry out adequate genetic counselling and prenatal diagnosis in the affected families.

CONTRIBUTIONS:

METHODOLOGICAL:

With the present work, proprietary methods were developed for molecular genetic diagnosis of diseases with disorders in the skeletal and connective tissue.

SCIENTIFIC THEORETICAL:

1. Studies have been conducted on the mechanisms of mutagenesis in human

genome and the involvement of the surrounding sequence in the genomic architecture in provoking the mutational events.

2. Mutagenesis patterns were interpreted involving straight, inverted repeats and symmetric elements provoking the formation of the double-helical secondary structures.

APPLIED-DIAGNOSTIC CONTRIBUTIONS:

1. Phenotype-genotype correlation of target diseases was achieved.
2. A total of 47% of affected families were genetically verified, genetically counselled and prepared for prenatal diagnosis.

SCIENTIFIC CONTRIBUTIONS TO WORLD DATABASES:

1. The first case of hypomelanosis of Ito associated with interstitial 15q11.2q13.3 triplication was found in Bulgaria, which is also the first proven case of triplication worldwide.
2. The first case of de novo heterozygous variant c.3743G>A, p.Trp1248fsTer in the MN1 gene associated with C-terminal truncation syndrome was found in Bulgaria.
3. The first familial case of Marshall-Stickler in Bulgaria associated with COL11A1:c.3474+1G>A mutation was discovered

BIBLIOGRAPHY

Foreign literature

1. Acuna-Hidalgo R et al (2016) Genome biology17.1.
2. Albena Todorova et al (1997) Hum Mut. 9:537-547
3. Aravidis C et al (2014) J Matern Fetal Neonatal 27:1502–6.
4. Barbosa M et al (2009) Am J Med Genet. 2009;149A:260.
5. Bochyńska, A. et al (2018) Cells 7, 17.
6. Byers PH et al (2012) Annu Rev Genet.46:475-97.
7. Cassidy S et al (1996) Am J Med Genet Part A 62:206(A4).
8. Castronovo C et al (2014) Am J Med Genet Part A 9999:1–10.
9. Chadwick D et al (1996) Am J Hum Genet 59A114(627).
10. Clayton-Smith J et al (1993) J Med Genet 30:529–531.
11. Crawford EC et al (1995) Am J Hum Genet 57A111(618).
12. Crolla JA et al (1995) Hum Genet 95:161–170.
13. Davidson L et al (2012) World psychiatry 11.2 (2012): 123-128.
14. Daubeney PE et al (1993) J. Pediatr. 1993; 152:715–6.
15. Fanganiello R et al (2007) Molecular Medicine. Vol. 13. No. 7. BioMed Central, 2007.
16. Farmaditya EP et al (2013) Report of Indonesian patients. 2013; 54(3): e72-e75
17. Fonseca R et al (2008) Am J Med Genet A146:658–60.
18. Griebel V et al (1998) Neuropediatrics 20:234–7.
19. Herrmann J et al (1975) Birth Defects 11:76–103
20. Hogart A et al (2010) Neurobiol Dis.38:181–91.
21. Holowinsky S et al (1993), Am JHum Genet 53A125.
22. Huang B et al (1997) Hum Genet99:11–17.
23. Jesse P et al (2012) FRET
24. Lekanne D et al (1995) Journal of Neuropathology & Experimental Neurology 54.2: 224-235.
25. Locke D et al (2004) J Med Genet 41:175–182s
26. Mak C et al (2020) Journal of neurology vol. 143,1 (2020): 55-68.
27. Marini J et al (2017) Nat. Rev. Dis. Prim. 3, 17052. 10.1038/nrdp.2017.52
28. Mladenova et al (2019) J Clin Med Genomics
29. Michiko H et al (1996) A neurocutaneous syndrome. Brain and Development 9:141–53.
30. Mundhofir F et al (2013) Singapore Med J 54.3 (2013): e72-e75.
31. Nehal K et al (1996) Arch Dermatol 132:1167-70.
32. Nicol L et al (2019) Bone 120:70-74.
33. Orlova E et al (2017) The Journal of Clinical Endocrinology & Metabolism 102.9 3546-3556.
34. Pascual-Castroviejo I et al (2008) Springer Vienna page 363–85.
35. Pellegrino Jet al (1995) Hum. Genet. 96: 485-489
36. Penrose L et al (1955) Parental age and mutation 269:312–3.
37. Pettigrew A et al (1987) Am JMed Genet Part A 28A:791–802.
38. Planté-Bordeneuve V et al (2011) The Lancet Neurology 10.12 1086-1097.
39. Pujana M et al (2002) Eur J Hum Genet10:26–35
40. Rauch F et al (2010) Eur. J. Hum. Genet. 18, 642–647
41. Reddy K et al (2000) Clin Genet 58:134–141
42. Ruggieri M et al (2000) Journal of Child Neurology 15:635–44

43. Rousseau Fet al (1996), Hormone research vol. 45,1-2
44. Ruggieri M et al (2011) Medlink Corporation.
45. Ruggieri M et al (1996) Neurology 46:485–92
46. Rutland BM et al (2006) Pediatric Neurology 34:51–4
47. Shiang R et al (1994) Cell. 1994; 78: 335-342
48. Schinzel A et al (1994) JMedGenet 31:798–803
49. Song SH et al (2012) Am J Med Genet A.158A:2456
50. Stoll et al. (1989) Clinical genetics 35.2 88-92.
51. Turleau, C et al (1986) Hum. Genet. 74: 185-187.
52. Ungaro P et al (2001) J Med Genet 38:26–34.
53. Uttarilli A et al (2019) Bone 120:204-211.
54. Van Dijk et al (2013) Am J Med Genet Part A
55. Vialard Fet al (2003) Am J Med Genet Part A 118A:229–234.
56. White et al (2016) Routledge.
57. Zhytnik et al (2019) Frontiers in genetics vol. 10 722.

Electronic sources

1. ENSEMBL. http://www.ensembl.org/index.html
2. Genetics home reference. https://ghr.nlm.nih.gov/
3. Human Genome Variation Society. http://www.hgvs.org/mutnomen/
4. Leiden Open Variation Database. http://www.lovd.nl
5. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/
6. National Institute of Health. https://www.nih.gov/
7. PolyPhen-2: prediction of functional effects of human nsSNP2. http://genetics.bwh.harvard.edu/pph2/
8. The Human Gene Mutation Database. http://www.hgmd.org/
9. University of California Santa Cruz. In-Silico PCR. https://genome.ucsc.edu/cgi-bin/hgPcr

PUBLICATIONS AND SCIENTIFIC DISCUSSIONS IN CONNECTION WITH THE DISSERTATION

Iv. Turturikov, Sl. Atemin, **M. Mladenova**, Z. Pavlova, T. Todorov, P. Dimova, N. Chilingirova, V. Mitev, Alb. Todorova (2021) Application of Next Generation Sequencing in Pediatrics: Rare Diseases in Early Childhood Screening. Practical Pediatrics issue 8/2021

Mladenova, M., Todorov, T., Grozdanova, L., Mitev, V., & Todorova, A. (2021). Novel Mutation in the COL11A1 Gene Causing Marshall-Stickler Syndrome in Three Generations of a Bulgarian Family. Balkan journal of medical genetics : BJMG, 24(1), 95–98. Импакт фактор:0.88

Mladenova M, Koleva M, Rodopska E, Alexandrova I, Bojinova V, et al. (2019) Hypomelanosis of Ito and De novo Interstitial 15q11.2q13.3Triplication in Bulgarian Family. J Clin Med Genomics 7: 155. Импакт фактор: 0.20

TOTAL IMPACT FACTOR: 1.08