APPLICATION OF NEXT GENERATION SEQUENCING IN THE DIAGNOSIS OF COMPLEX GENETIC DISORDERS

PhD. Thesis

Gábor Kovács

Department of Paediatrics

Faculty of Medicine

University of Szeged

SZEGED

2016

"The average patient with rare disease visits 8 physicians and receives 2 to 3 misdiagnosis before being correctly diagnosed. On average, this takes 7-6 years."

/Shire HGT: Rare Disease Impact Report-2013/

Table of contents

Publications related to the thesis

I. **Gábor Kovács**, Tibor Kalmár, Emőke Endreffy, Zoltán Ondrik, Béla Iványi, Csaba Rikker, Ibolya Haszon, Sándor Túri, Mária Sinkó, Csaba Bereczki, Zoltán Maróti. Efficient Targeted Next Generation Sequencing-based workflow for differential diagnosis of Alport-related disorders; *PlosOne*; 2016 DOI: 10.1371/journal.pone.0149241;
[IF: 3,2]

II. Endreffy Emőke; **Kovács Gábor**; Horváth Emese; Karg Eszter; László Aranka; Túri Sándor. Ectodermalis Dysplasias férfi molekuláris genetikai vizsgálata. *Gyermekgyógyászat*; 2008; 2008/59. p. 85-88.

List of abbreviations

AD	Autosomal dominant
AR	Autosomal recessive
AS	Alport syndrome
CES	Clinical exom sequencing
CKD	Chronic kidney disease
ECD	Ectodermal dysplasia
FBH	Familiar benign haematuria
NGS	New generation sequencing
RBC	Red blood cell
TBMN	Thin basement membrane nephropathia
XR	X-linked recessive

I.   Introduction

Rare diseases (RDs), also called orphan diseases are those which affect only a very small portion of the population and the majority of them (~80%) are genetic. According to the latest international agreement, the prevalence threshold for RD not more than 5 affected persons per 10,000 (European Parliament and of the Council of 16 December 1999). Up to date more than 6000 RDs are characterised by a broad diversity of symptoms. Many of them are characterized by mental retardation and affects 50-70% of children and even today around 30% of the affected patients will die under the age of five. The common symptoms can mask the basic rare diseases which can lead to misdiagnosis and delays treatment. In 1997 the Orphanet database was created which aim is to provide information about rare disease to clinicians, patients in order to contribute to the successful diagnosis and to better treatment. RDs were registered with one offered name and as many synonyms with an ORPHA number (www.orpha.net). As the result of this it makes even more difficult to even guess the exact number of rare diseases. Majority of RDs are monogenic, however due to the „umbrella terms" used for disease groups with overlapping phenotype there are numerous complex diseases that are caused by few to hundreds of genes (nephrosis, schizophrenia, ECD, COPD, etc.). Due to the technical limitations and the wide spectrum of the phenotypes the diagnosis of the rare disease cases are very complicated and the vast majority of them remains unsolved.

Up to recently the clinical genetic diagnosis of complex diseases were only partially solved. Sanger sequencing of hundreds of genes is not feasible in the diagnostics due to the enormous manual work and costs of the analysis. This challenge was partially handled by Orphanet where the most frequent genes or specific mutations causing various diseases were analysed in specific expert centres. Our genetic diagnostic laboratory is an expert centre of several RDs, including monogenic and complex diseases. Among others, we use diagnostic methods to analyse the most frequent X linked type (COL4A5 gene) Alport syndrome and X linked type ECD (EDA gene). In many cases the traditional (linkage analysis, high resolution melting analysis, Sanger sequencing based) methods investigating only the most frequent genes did not revealed the genetic cause of the disease.

Recently with the advancement of high-throughput massively parallel sequencing technologies (New Generation Sequencing, NGS) it became possible to sequence more genes and more samples at the same time. Since we had numerous unsolved cases of complex disease patients we wanted to establish and test the feasibility of NGS. In my thesis I chose to work with Alport syndrome and ectodermal dysplasia in order to improve their current molecular diagnostic approach in our laboratory, thus my thesis has two parts:

I.1. Alport related disorders

Familiar benign haematuria (FBH) and Alport syndrome (AS) are familial haematuria diseases which in case of AS regularly escalate to chronic kidney disease (CKD) stage 5 (formerly referred as end stage renal disease). AS patients usually have sensorineural high-tone deafness and ocular abnormalities affecting the lens and fundus [1,2]. Today more focus has been placed on treating patients early to prevent or delay future end stage kidney damage. Although, the pathogenesis of CKD is multifactorial, some of the suggested therapeutic interventions (anti-hypertensive therapy, glycaemic control, anti-proteinuric therapy, renoprotection, and life style management such as restricted protein intake, cessation of cigarette smoking and chronic analgesic-abuse) are encouraging. These preventive steps the more earlier are implemented the more efficient they are[3–5]. There has been an old and ongoing dispute to differentiate between AS and FBH based on the wide spectra of observed clinical symptoms, microscopic analysis of renal biopsy, immunological examination and family history [6]. The first observed clinical signs are postponed by the fact that in the intrauterine life our glomerular basal membrane does not contain *COL4A3-A4-A5*, but *COL4A1* and *COL4A2*. First symptom of AS is haematuria which appears usually 10-15% of cases in the first year of life, while the majority of patients show symptoms only around 5-7 years of age [7]. Because of this late manifestation and similarity between early symptoms of AS and FBH neither clinical nor histological methods are appropriate for early diagnosis. However, the correct genotyping can give an informed clue for the expected progression of the disease, aid the determination of mode of inheritance useful to the family members showing yet no symptoms, and most crucially is unavoidable for kidney donor selection [7,8]. In order to determine the mode of inheritance of AS and/or FBH a pathogenic mutation(s) in the *COL4A5* gene present on X chromosome or mutations in either the *COL4A3* or *COL4A4* genes on chromosome 2 should be

found [9]. The most common form of AS, with approximately 4 in every 5 cases, is inherited in an X-linked fashion. X linked carrier females usually show variable intermediate phenotype. Due to imbalances in random X inactivation the phenotype can vary even between family members. In case of FBH, the mode of inheritance is autosomal dominant, and this disease is caused by a single heterozygous mutation either in *COL4A3* or in *COL4A4* genes [10]. If there are two mutations either in *COL4A3* or *COL4A4* genes,—a more severe— form of AS develops. Because of this FBH can be viewed as the carrier state of AS. There are very few reports in the literature, the vast majority of which from the pre-next generation sequencing (NGS) era where autosomal dominant form of AS reported, and only one mutation was found in either the *COL4A3* or *COL4A4* gene. Up to now it is not clear whether this form of AS may only be the result of prior technical limitations or it is real, which can be solved only with sequence-based analysis of larger data set and with the introduction of a new technology [11].

In my thesis I focus on the improvement of genetic diagnosis of type IV collagenopathies, AS and FBH [12]. Sequential (one by one) genetic testing for mutations in *COL4A3-A4-A5* genes has become an integral part of the clinical evaluation. In our previous study on type IV collagenopathies we detected large number (more than a dozen) of non-synonymous variants in each individual in these 3 genes, therefore the distinction between causative mutations and benign variants is crucial [13]. Since all three genes are large (contains 52, 48 and 51 exons, respectively) the use of conventional Sanger sequencing is time-consuming, expensive and can suffer from some technical limitation (such as failing to detect insertion/deletion with certain sizes in a heterozygote subject). One way to overcome these problems is to sequence all three genes simultaneously using NGS-targeted sequencing [14]. We present a new, efficient, amplicon based NGS protocol for simultaneous analysis of the coding regions (all the exons and flanking intronic sequences) of the COL4A3, A4 and A5 genes since a previously published NGS-based approach failed to detect mutations in 45% of their cases [11]. Both mutations and polymorphisms in the 3 investigated COL4A genes are thought to be highly population-specific due to the lack of selection pressure in case of the polymorphisms and low selection pressure in case of TBMN. Thus, in order to further aid the classification of COL4A3-A4-A5 genetic variations, we present COL4A3-A4-A5 polymorphisms of 66 unrelated Hungarian non AS/TBMN patients' data obtained by NGS. In addition we set the task of identifying the genetic causes of 17 Hungarian AS/TBMN independent cases, analysing 17 AS/TBMN patients using our panel and Sanger validation of 55 additional family members.

I.2. Ectodermal dysplasia disorders

Ectodermal dysplasias (ECD) are inherited disorders characterized by abnormal morphogenesis of ectoderm [15]. Ectoderm is one of the three primary germ layers during the development of early embryo. Ectodermal dysplasias are hereditary diseases which represents a large group of conditions in which two or more ectodermally-derived tissues fail to develop. There are more than 150 types of Ectodermal dysplasias described so far, with an incidence of 7/10000 [45]. Based on the inheritance pattern within family one can differentiate between autosomal recessive (AR), autosomal dominant (AD), X-linked recessive (XR) and X-linked dominant (XD) inherited forms [16]. The number of the genes and the different phenotypes make difficult the diagnosis for the clinicians and for the geneticist. The most common form of ECD is the X-linked Hypohidrotic ectodermal dysplasia (XL-HED) characterized by hypotrichosis, hypohidrosis, hypodontia. The prevalence of XL-HED is 1/5000-10000 in newborns [17].
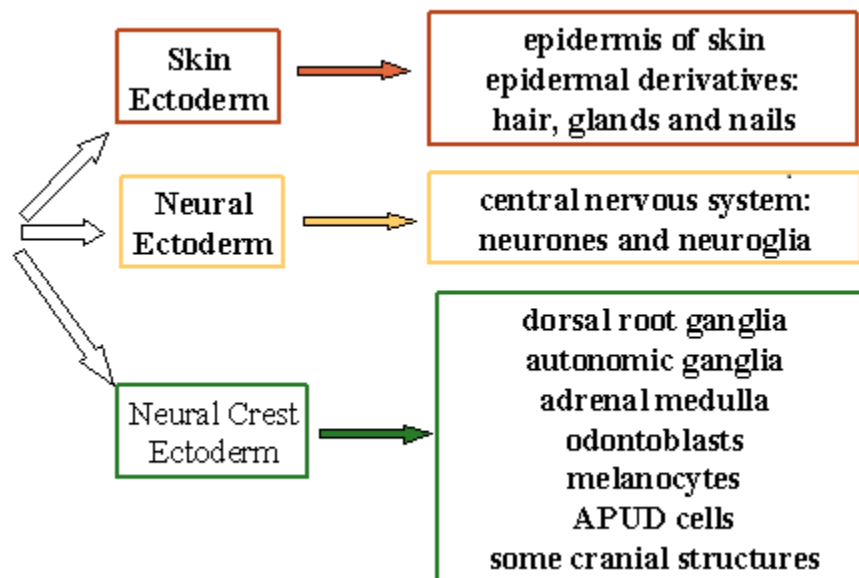

Figure 1. Derivatives of ectoderm

Based on the different clinical symptoms it is not easy to assign the exact type of ECD in the majority of the cases. The phenotypes are not always recognizable in newborn babies but usually identifiable in early childhood. The following tissues are affected in ECD:

Hair: the hair is absent on the scalp, sparse and light with abnormal texture. The hair may also be fragile and curly. The hair is usually dry because of the absence of the oil glands. The eyebrows and eyelashes or the other body hair may absent or sparse. Any of the hair defects could manifest at birth, while others are not until later age.

Nails: nail usually poorly developed on the fingers and on the toes. Usually thick, sharp, stands out, fragile and brittle. Additionally, the nails grow more slowly and become easier infected by fungus or yeast.

Teeth: may be missing or reduced in number. Teeth are widely spaced and tapered/peg shaped. If the teeth are missing than the jawbones in which were embedded do not develop well, that cause the „old age appearance" of the face**.**

Sweat glands: characterized by missing and scattered. Due to the functional problems the body is unable to maintain the normal body temperature. As a result the extreme high body temperature is usual that is why the patients unable to tolerate the high outside temperature. The recurrent fever periods are the first symptoms which help the diagnosis the ECD in early childhood (hypohidrotic and anhidrotic form of ECD).

Further symptoms: light skin contains red and brown pigments. The skin on the fingers and on the toes is particularly thick. The abnormal development of the ear can lead to hearing loss/hearing impairment. The presence of cleft palate/lip is usual. The lack of the secretion in glands (in the mouth and nose) can result throat infections which can make breathing relatively difficult.

We distinguish four main symptoms of ECD based on their clinical features [18].

ECD1: Trichodysplasia (hair dysplasia)
ECD2: Dental dysplasia (teeth dysplasia)
ECD3: Onychodysplasia (nail dysplasia)
ECD4: Dyshidrosis (sweat gland dysplasia).

Based on the combination of the previously mentioned 4 groups of symptoms the following clinical syndromes can occur:

- Acro-dermato-ungual-lacrimal-tooth syndrome (ADULT syndrome)
- Ankyloblepharon-ectodermal defects-cleft lip and palate syndrome (AEC syndrome)
- Cartilage-hair hypoplasia syndrome
- Cleft lip/palate-ectodermal dysplasia syndrome
- Clouston syndrome
- Congenital hypotrichosis with juvenile macular dystrophy
- EEC Syndrome
- Ectodermal dysplasia, pure hair and nail type
- Ectodermal dysplasia, skin fragility syndrome
- Ectodermal dysplasia, with ectrodactyly and macular dystrophy
- Ectrodactyly-ectodermal dysplasia-cleft lip/palate syndrome (EEC syndrome)
- Focal dermal hypoplasia syndrome
- Hypohidrotic ectodermal dysplasia
- Hypohidrotic ectodermal dysplasia, with immune deficiency
- Hypohidrotic ectodermal dysplasia, with immune deficiency, osteopetrosis and lymphoedema
- Incontinentia pigmenti
- Limb-mammary syndrome
- Naegeli syndrome
- Rapp-Hodgkin syndrome
- Rothmund-Thomson syndrome

Around 141 gene were linked to ECD. Due to the number of the genes the traditional molecular genetic methods when analysis carried out gene by gene is not feasible. Because of this I used a different approach called clinical exom sequencing (CES) which allows simultaneous investigation of all known 141 ECD genes. The exome sequencing is a targeted sequencing approach which is restricted to the protein-coding regions of the genome (exom).

I.3. Aims

In my thesis addressed the following problems:

- Identifying the causative genetic alterations in 14 nephropathy families and 3 nephropathy individuals
- Try to give a better description to the controversial inheritance (AR, AD) mode of Alport and TBMN.
- Identifying the causative genetic alterations in three families with different ECD syndromes.

II.   Patients

II.1. Alport related disorders

Our Hungarian cohort consisted of 3 individuals (I1-I3, where family members were not available for either clinical or genetic analysis) and 14 COL IV nephropathy families (F1-F14) where clinical data and inheritance of clinical symptoms supported Alport/FBH diagnosis while other possible clinical causes were excluded. All volunteer family members underwent urinalysis and renal function evaluation. Kidney biopsy was not performed on all proband. All available clinical and pathological information is summarized in Table 1. After obtaining written informed consent peripheral blood samples were collected from both the affected and unaffected members of the families. We used anonymised DNA samples from our licensed biobank for the 66 Hungarian non AS/FBH control samples. Criteria for FBH: mild persistent haematuria (25-30 RBC/HPF), mild (0.5-1.0 g/day) or no proteinuria, first appearance after 10 years of age. Criteria for Alport: persistent haematuria (>30, 50-100 RBC/HPF), persistent proteinuria (0.5-5.0 g/day).

Family	Individuals	Age	Hematuria	Proteinuria	Renal function	Biopsy, histology	hearing
F1	I/2	36 yrs	microhematuria	none	normal	none	normal
	II/2	3 yrs	persistent hematuria	none	normal	At age 3 could not discriminate between AS or TBMN	normal
F2	I/1	58 yrs	none	none	normal	not performed	normal
	II/2	31 yrs	microhematuria	yes	normal	At age of 31 TBMN was suggested	normal
F3	I/2	53 yrs	hematuria	none	normal	At age 7: suggested TBMN	ND
	II/1	32 yrs	hematuria	yes	on dialysis	At age 17: suggested TBMN	ND
	II/2	30 yrs	hematuria	0.5 g/day	eGFR 90 (CKD Stage 1), had transplantation and now on dialysis	not performed	ND
F4	I/1	38 yrs	microhematuria	no	normal level of se-creatinine	not performed	mild right-side and moderate left-side hipacusis
	II/1	15 yrs	ND	ND	on dialysis since age of 15	not performed	bilateral mild hipacusis
	II/2	13 yrs	hematuria	5 g/day at age of 13		At age of 18 diffusely thin GBM, basket weaving in several loops, α3 chain was entirely absent, α5 chain exhibited very focal expression	ND

	II/3	16 yrs	microhematuria	ND	normal	ND	normal at age of 29
F5	I/2	86 yrs	hematuria	ND	ND	ND	ND
	II/3	61 yrs	hematuria	ND	ND	ND	ND
	II/4		hematuria	ND	ND	ND	ND
	III/2	30 yrs	ND	ND	kidney transplantation at age of 31	ND	ND
	IV/1	17 yrs	hematuria	yes	no kidney transplantation at age of 25	ND	normal at age of 25
	IV/2 and IV/3 twins	16 yrs	ND	ND	both had kidney transplantation at age of 16	ND	ND
F6	II/1 died at age 13	13 yrs	ND	ND	ND	ND	ND
	II/5	21 yrs	hematuria	yes	ND	At age 11; not conclusive	ND
F7	I/1	80 yrs	none	none	se-creatinine normal	not performed	ND
	I/2	49 yrs	10-12 dysmorphic RBC/hpf	yes	se-creatinine normal	At age of 49 diffusely thin GBM, glomerular scars	moderate bilateral hearing loss
	II/2	60 yrs	10-15 dysmorphic RBC/hpf	yes	se-creatinine normal	not performed	ND
	II/4	62 yrs	10-15 dysmorphic RBC/hpf	none	ND	not performed	ND
	III/1	46 yrs	8-10 dysmorphic RBC/hpf	none	se-creatinine normal	At age of 46 diffusely thin GBM	bilateral hearing loss
	III/3	42 yrs	hematuria	yes	ND	not performed	ND
	III/4	40 yrs	hematuria	ND	ND	ND	ND
	III/6	47 yrs	30-35 dysmorphic RBC/hpf	1.5 g/day	se-creatinine normal	At age of 47 diffusely thin GBM, doubled in some loops	bilateral hearing loss
	III/8	44 yrs	hematuria	ND	ND	not performed	ND

	III/9	42 yrs	25-30 dysmorphic RBC/hpf	0.8 g/day	se-creatinine normal	At age of 42 diffusely thin GBM	moderate bilateral hearing loss
	IV/1	17 yrs	8-10 dysmorphic RBC/hpf	3 g/day	se-creatinine normal	At age of 17 diffusely thin GBM, with splitting in some loops	moderate bilateral hearing loss
	IV/2	24 yrs	8-10 dysmorphic RBC/hpf	3 g/day	se-creatinine normal	At age of 24 diffusely thin GBM, with splitting in some loops	bilateral hearing loss
	IV/3	36 yrs	hematuria	ND	ND	not performed	normal
	IV/7	15 yrs	50-60 dysmorphic RBC/hpf	0.8 g/day	se-creatinine normal	At age of 15 diffusely thin GBM	moderate unilateral hearing loss
	V/1	4 yrs	persistent hematuria	ND	ND	not performed	normal at age of 4
F8	I/1	38 yrs	hematuria	ND	ND	not performed	normal
	II/1	5 yrs	hematuria	ND	ND	not performed	normal
	II/2	12 yrs	100 RBC/hpf	ND	ND	not performed	normal
F9	I/1	71 yrs	hematuria	ND	ND	not performed	ND
	II/2	45 yrs	hematuria	ND	ND	At age of 32 could not confirm either AS or TBMN	ND
	III/1	16 yrs	hematuria	ND	ND	At age of 3 TBMN was suggested	ND
F10	II/1	65 yrs	hematuria	ND	ND	not performed	ND
	II/2	67 yrs	hematuria	ND	ND	not performed	ND
	III/2	44 yrs	hematuria	ND	ND	not performed	ND
	IV/1	12 yrs	hematuria	ND	ND	not performed	mild hipacusis

	Individuals	Age	Hematuria	Proteinuria	Renal function	Biopsy, histology	hearing
F11	I/1	50 yrs	30 RBC/hpf	ND	ND	not performed	ND
	II/1	24 yrs	30 RBC/hpf	yes	ND	At age of 24 secondary FSGS, likely TBMN but could not exclude AS	normal
	II/2	15 yrs	30 RBC/hpf	yes	decreasing eGFR	ND	ND
F12	II/1	69 yrs	hematuria	none	ND	not performed	normal
	II/3	71 yrs	hematuria	none	ND	not performed	normal
	II/5	68 yrs	hematuria	none	ND	not performed	normal
	III/3	37 yrs	hematuria	none	ND	At age of 42 TBMN was suggested	normal
	III/6	42 yrs	hematuria	none	ND	not performed	normal
	III/7	35 yrs	hematuria	none	ND	not performed	normal
	III/8	39 yrs	hematuria	none	ND	not performed	normal
F13	I/1	38 yrs	hematuria	no	ND	ND	normal
	II/2	6 yrs	hematuria	no	ND	ND	normal
F14	I/2	35 yrs	microhematuria	no	ND	not performed	ND
	II/1	2 yrs	microhematuria at age of 3	no	ND	not performed	ND
	Individuals	**Age**	**Hematuria**	**Proteinuria**	**Renal function**	**Biopsy, histology**	**hearing**
	I1	35 yrs	hematuria	yes	ND	not performed	ND
	I2	38 yrs	microhematuria	0.8 g/day	ND	not performed	ND
	I3	1 yr	hematuria	none	ND	not performed	ND

Table 1. Clinical and pathological details of the investigated Hungarian AS/FBH families.

RBC – red blood cell ; eGFR – estimated glomerular filter rate; CKD – chronic kidney disease; ND – no data; hpf – (microscopic) high power field; FSGS – focal segmental glomerulosclerosis, GBM – Glomerular basement membrane, AS – Alport syndrome, FBH –Familiar Benign Hematuria. There was no data on eye sight problems in the patients.

II.2. Ectodermal dysplasia patients

II.2.1. Family I.

The following table contains the clinical symptoms of the investigated family I. members.

| Family member	Code	Clinical symptoms
Proband	ECD2	frequent periods of high fever hypohidrosis peg-shaped teeth lack of eyebrows/lashes/hair
Mother	ECD7	fever periods hypodontia body hair in various part of the body
Father	ECD6	unaffected
Grandmother	ECD4	fever periods hypodontia body hair in various part of the body
Sister	ECD8	unaffected
Brother	ECD3	unaffected
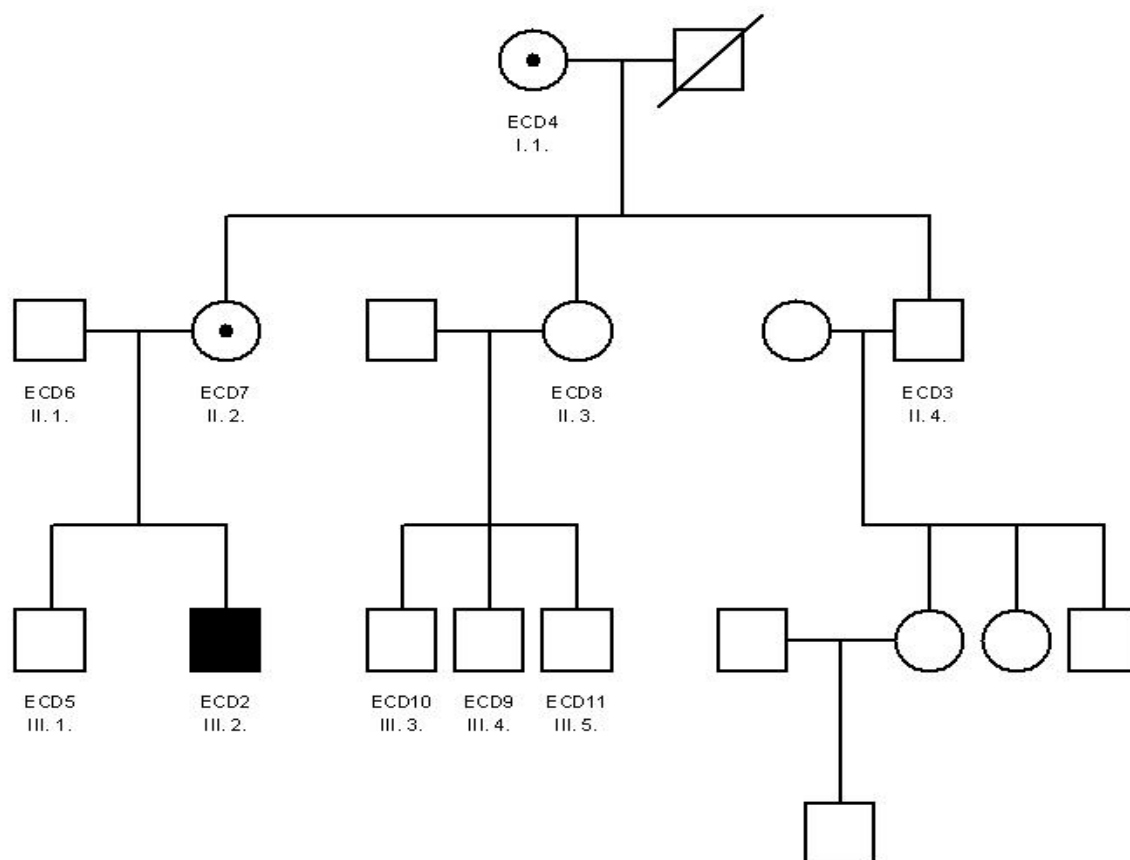
Figure 2. Family tree of the Family I.

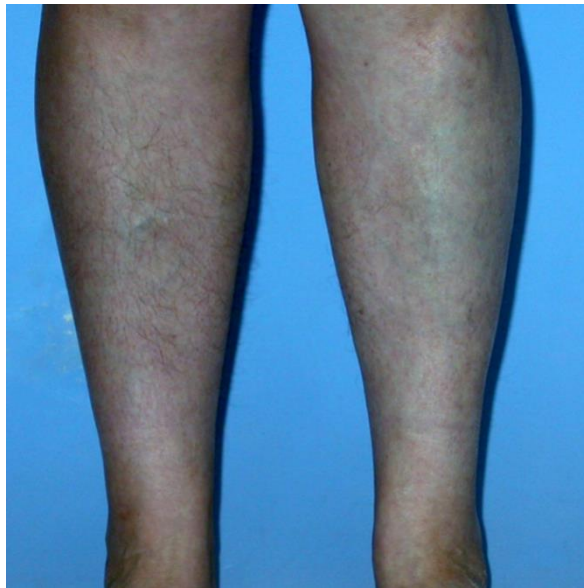**Figure 3**. The affected man's (ECD2) head

Figure 4. Leg of the mother (ECD7). The right leg is hairy but the left is hairless.

II.2.2. Family II.

The following table contains the clinical symptoms of the investigated family II. members.

| Family member	Code	Clinical symptoms
Proband	ECD14	Hipohydrosis Sparse hair malformed-, peg-shaped teeth syndactyly double nipples
Mother	ECD19	tooth agenesis, sharp tooth sparse hair reduced sweating ability
Father	ECD20	unaffected
Sister	ECD21	Tooth agenesis, sharp tooth sparse hair reduced sweating ability
Nephew	ECD22	Tooth agenesis

■ Hypohidrosis, sparse hair ▨ Hypodontia
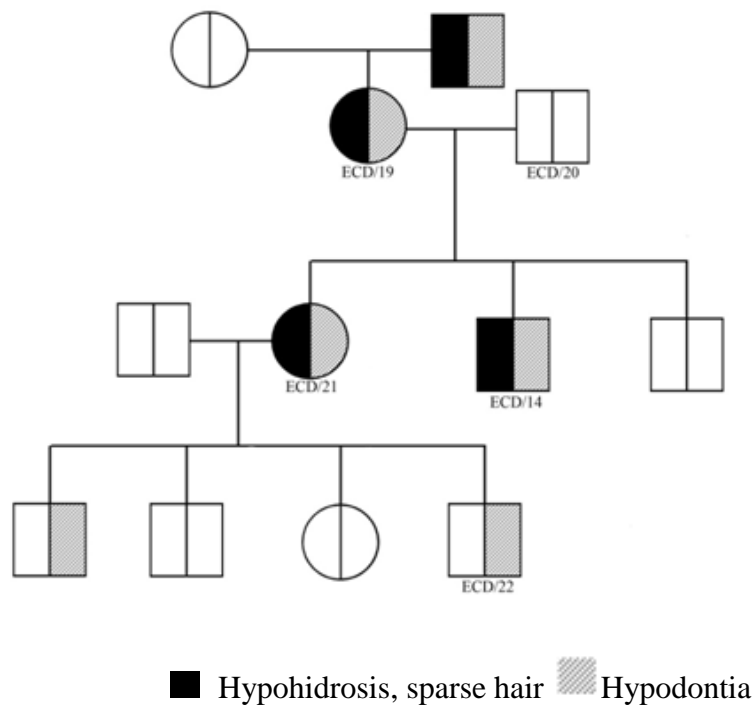
Figure 5. Family tree of the Family II.

Figure 6. The symptoms of the affected boy (ECD14). **a, b**: parse hair; **c**: lack of eyelashes and eyebrows; **d**: syndactily; **e**: hypodontia-only two permanent teeth; **f**: double nipple.

II.2.3. Family III.

The following table contains the clinical symptoms of the investigated family II. members.

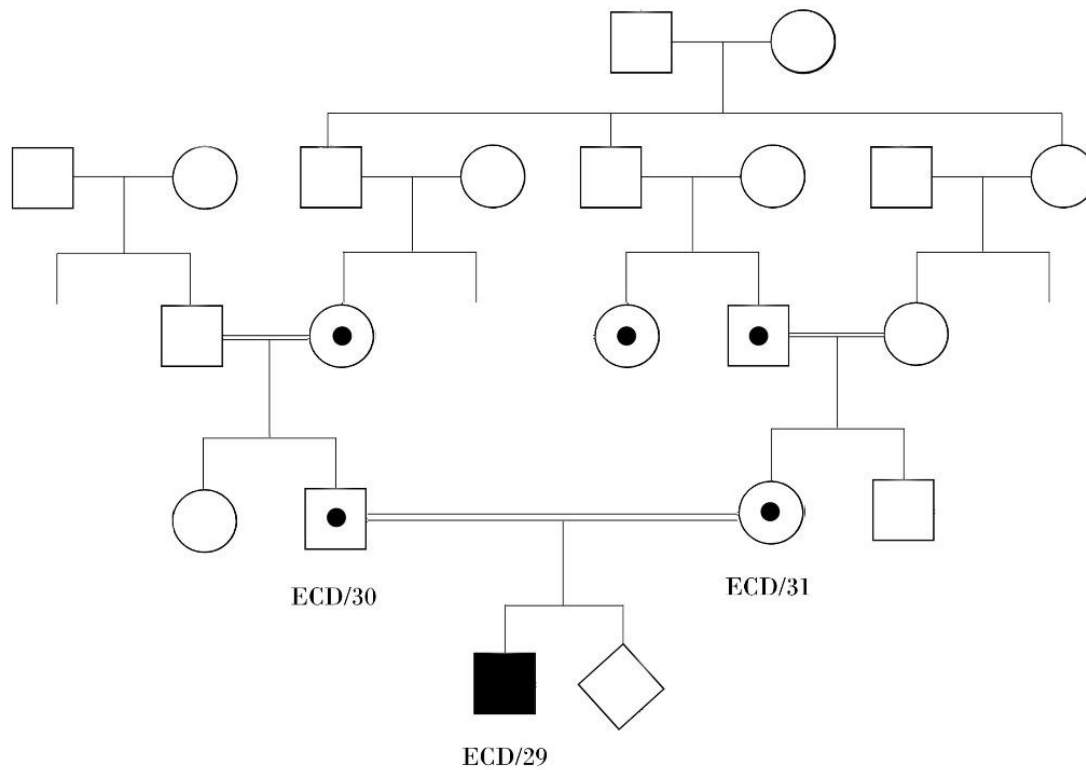| Family member	Code	Clinical symptoms
Proband	ECD29	absent hair, eyebrows, eyelashes, severely dystrophic nails syndactyly
Mother	ECD31	unaffected
Father	ECD30	unaffected

Figure 7: Family tree of the affected boy (ECD29)

Figure 8. The affected boy (ECD29)

Based on the clinical symptoms we had chosen one affected member (proband) from every family to be first investigated with NGS.

III. Material and Methods

In accordance with Hungarian regulations written informed consent was obtained from each patient in the framework of a genetic counselling session. The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by the regional ethical committee (Ethical board of University of Szeged).

The genomic DNA from peripheral blood leukocytes was isolated by salting out procedure.

III.1. Targeted resequencing of Alport related genes

Our AS/FBH panel consisting of all the three *COL4A3-4-5* genes which was designed in-house using the Ion AmpliSeq™ Designer version 4.2.1 from Life Technologies. Gene enrichment was performed by the AmpliSeq amplicon based system. Sample enrichment was performed by the Ion AmpliSeq(TM) Library Kit, Life Technologies. The samples were barcoded and sequenced on IonTorrent 316 chip. We aimed for at least 25 times coverage (more than 90% of the target were covered at least 100 times and more than 95% of the target were covered at least 50 times). Bioinformatical analysis was performed by CLC Bio, Torrent Suite 4.2 and Illumina VariantStudio softwares.

III.2. Clinical exom sequencing of ectodermal dysplasia related genes

To analyse the samples we used the Illumina TruSight One Sequencing Panel. The TruSight One sequencing panel provides comprehensive coverage of 4813 clinically relevant genes selected from the Online Mendelian Inheritance in Man (OMIM), Human Gene Mutation Database Professional (HGMD) and GeneTests databases. We followed the user manual of TruSight One kit for the library preparation and sequencing. We used 50ng gDNA. The cumulative target region: 12Mb; the minimum coverage of the target region was >20x (no. target exons: 62000; probe size: 80mer; 150bp pair-end reads). The samples were sequenced on Illumina MiSeq platform.

Alignment and variant calling was performed by BWA + GATK HaplotypeCaller. We used in house algorithms for detection of large structural rearrangements and copy number variations (CNV). We used VariantStudio for variant annotation, filtration and classification.

III.3. Variant confirmation workflow

To confirm and evaluate the putative disease causing genetic variations we designed primers by Primer3Plus software and executed PCR amplification for all members of our cohort [19]. PCR products were purified using High Pure PCR Product Purification Kit, (Roche) and sequenced on Applied Biosystems HITACHI 3500 Genetic Analyser. For visualization of the Sanger sequencing we used FinchTV (Geospiza Inc.).

To visualize the results of NGS we used IGV 2.3 and Illumina Variant Studio 2.2 versions. We compared the results with the following public databases: HGMD [20], LOVD [21], 1000Genomes.org [22], Alport database (ARUP) [23]. All the previously unpublished mutation were submitted in LOVD database (http://grenada.lumc.nl/) and can be found in ClinVar database as well (http://www.ncbi.nlm.nih.gov/clinvar/).

See our variant classification criteria system in supplement 1.

Gene	Variant	MAF	Consequence	Published in	Family (F) or Individual (I)
COL4A3	c.127G>C	0.153	p.Gly43Arg	[24]	F1 (M),F5 (M),F6 (M),F10 (M),F13 (M)
	c.1721C>T	0.121	p.Pro574Leu	[25]	F1 (M),F5 (M),F2 (F),F4 (F),F8 (F), F10 (M),F11 (F),F12 (F),F13 (M),F14 (F),I1(M), I2 (M), I4,F1 (M), F5 (M)
	c.3325C>T	0.024	p.Pro1109Ser	[26]	F4 (F)
	c.547-9A>C	0.024	IV	rs55667591	F3 (F), F4(F)
	c.3807C>A	0.040	p.Asp1269Glu	[25]	F2 (F), F6 (M)
	c.976G>T	0.226	p.Asp326Tyr	[27]	F3 (F), F6 (M), I3 (M), F7 (M), F13 (M), F11 (F), F8 (F)
	c. 2384-5T>C	0.113	IV	[28]	F10 (M), F7 (M)
	c.485A>G	0.798	p.Glu162Gly	[29]	F3 (F), I1 (F), F6 (M), F7 (M), F11 (F), F9 (F), F1 (M), F2 (F), F10 (M), F4 (F), I2 (M), F5 (M), F12 (F), F12 (F), I3 (M), F8 (F), F13 (M), F13 (F)
	c. 144+12C>A	0.339	IV	[29]	F8 (F) F3 (F), F10 (M), I1 (F), I2 (M), F6 (M), I3 (M), F13 (M), F11 (F), F9 (F)
	c.422T>C	0.798	p.Leu141Pro	[26]	F1 (M), F2 (F), F10 (M), F4 (F), I2 (M), F5 (M), F12 (F), F12 (F), I3 (M), F8 (F), F13 (M), F13 (F), F3 (F), I1 (F), F6 (M), F7 (M), F11 (F), F9 (F)
	c. 1352A>G	0.105	p.His451Arg	[25]	I2 (M), F9 (F)
COL4A4	c.2717-5A>T	0.040	IV	[30]	F11(F)
	c.2501A>G	0.008	p.Lys834Arg	[31]	F11 (F)
	c. 4207T>C	0.355	p.Ser1403Pro	[32]	F2 (F), F3 (F), I1 (F), I2 (M), F5 (M), F12 (F), F6 (M)F10 (M), F4 (F), F12 (F), F7 (M)
	c.3817+9G>C	0.363	IV	[33]	F1 (M), F10 (M), F4 (F), I3 (M), F7 (M), F13 (F)
	c.3979G>A	0.363	p.Val1327Met	[11]	F1 (M), F10 (M), F4 (F), F12 (F), I3 (M), F7 (M), F13 (F)
	c 1444C>T	0.435	p.Pro482Ser	[24]	F2 (F), F10 (M), I1 (F), I2 (M), F12 (F), F7 (M) F1 (M), F3 (F), F5 (M), F12 (F), F6 (M), F13 (F), F11 (F)
	c.3011C>T	0.444	p.Pro1004Leu	[24]	F2 (F), F3 (F), F10 (M), I1 (F), I2 (M), F5 (M), F12 (F), F6 (M), F7 (M), F1 (M), F4 (F), F12 (F), I3 (M), F13 (M), F13 (F)
	c.2996G>A*	0.008	p.Gly999Glu	[34]	F4 (F)
	c.4394G>A**		p.Gly1465Asp	[11]	F3(F)
COL4A5	c.2768-11A>G	0,065	IV	[35]	F3 (F), I1 (F), F10 (M), I2 (M), I3 (M)

Table 2. List of the identified polymorphisms in the investigated non AS/FBH cohort.

minor allele frequency;; IV: intronic variant; F: female; M: male)

* variant effect remained uncertain since we have found it in a control person (age 13 with two negative urinal samples, but we could not rule out later manifestation of FBH).

** this is a rare genetic variant with uncertain consequences. It was published both as polymorphism and as pathogenic mutation. In Family 3 this variant is co-segregating with a known pathogenic mutation (c.2320G>C), so we can not elaborate on the consequence of this variant when it is occurring separately.
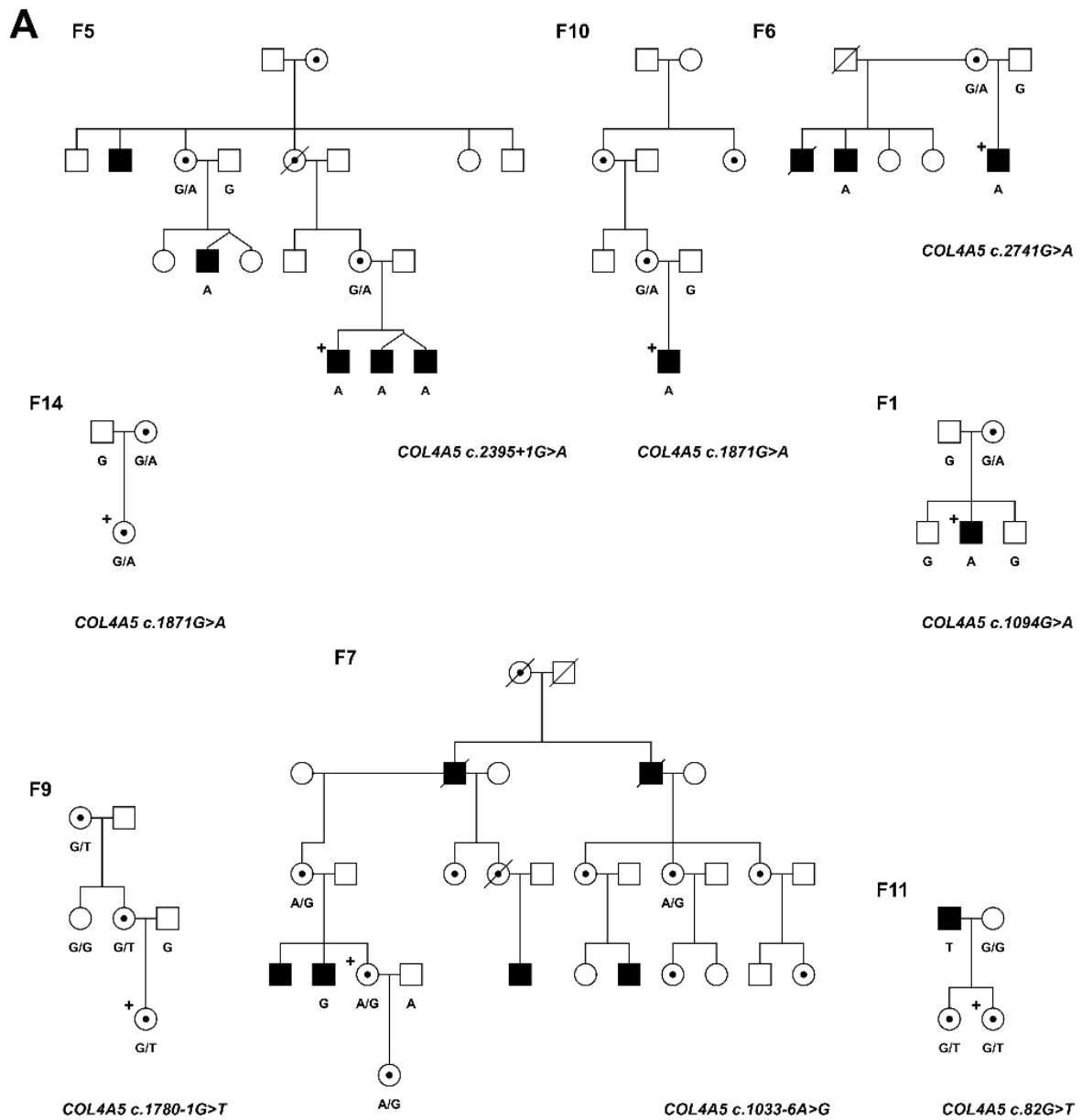
The available medical records and the predictive clinical and pathological results are often incomplete or not conclusive enough to distinguish between TBMN and the different types of Alport (ARAS, XR), especially in not so-large families (see Table 1 and Figure 9.) Therefore, we decided to do sequencing-based genetic analysis on the Hungarian AS/TBMN cohort using our AmpliSeq panel. Selection criteria for individuals for the initial NGS analysis were as detailed in Methods. Then, as results of NGS analysis all the identified unique 37 genetic alterations were compared to data of publicly available databases such as LOVD, dbSNP, HGMD and Alport database (University of Utah). We identified the 21 published, previously mentioned polymorphisms (Table 2) and 4 published mutations (2 in COL4A4 and 2 in COL4A5 genes). All the four published mutations were validated in the appropriate AS/TBMN cases, where we confirmed that only affected individuals carried them. The 14 remaining alterations were investigated among AS/TBMN family members. We have classified all of these alterations as new mutations (5 in COL4A3 gene; 3 in COL4A4 and 6 in COL4A5 genes) as their exclusive presence was confirmed in the affected family members by Sanger sequencing (see individual family trees in Supplements). Six out of the thirteen SNP mutations were amino acid change (missense mutations), while the remaining seven SNP mutations were found in the splice recognition sites (splice variants). We also identified a COL4A4 mutation affecting part of exon 20 and 2 bases in the splice donor site that is a 52 bp deletion shortening the exon and altering splicing (Table 3).

Gene	Variant	Consequence	Family/Individual	Published in
COL4A3	c.687G>A	splice region	F4(F)	this paper
	c.3490G>T	p.Gly1164Cys	F4(F)	this paper
	c.765+2T>C	splice donor	F12(F)	this paper
	c.3410G>A	p.Gly1137Asp	F13(M)	this paper
	c.765G>T	splice region	F2(F)	this paper
COL4A4	c.2320G>C	p.Gly774Arg	F3(F)	[36]
	c.2986G>A	p.Gly996Arg	I3(F)	this paper
	c.1100-2A>C	splice acceptor	F8(F)	this paper
	c.1320_1369+2del	deletion, splice donor	F3 (F)	this paper
COL4A5	c.1780-1G>T	splice acceptor	F9(F)	this paper
	c.1871G>A	p.Gly624Asp	F10(M), F14(F), I2(M)	[36]
	c.1033-6A>G	presumed splice variant	F7(M)	this paper
	c.2395+1G>A	splice donor	F5(M)	this paper
	c.1094G>A	p.Gly365Glu	F1(M)	[37]
	c.2741G>A	p.Gly914Asp	F6(M)	this paper
	c.82G>T	p.Ala28Ser/ splice acceptor	F11(F)	this paper
	c.1010G>T	p.Gly337Val	I1(M)	this paper

Table 3. List of the identified mutations in our Hungarian AS/FBH cohort

(F: female; M: male)

To demonstrate the ease and usefulness of our methodology, we categorized all of the 14 families as follows: F1, F5-7, F9-11 and F14 have X-linked recessive AS (XR); F3 and F4 have autosomal recessive AS (ARAS); F2, F8, F12 and F13 have TBMN (see Figure 1A-C). We had no information about additional family members in the individual (I1-I3) cases, but using our genetic diagnostic approach we could categorize them as X-linked recessive AS (I1 and I2) and TBMN (I3).
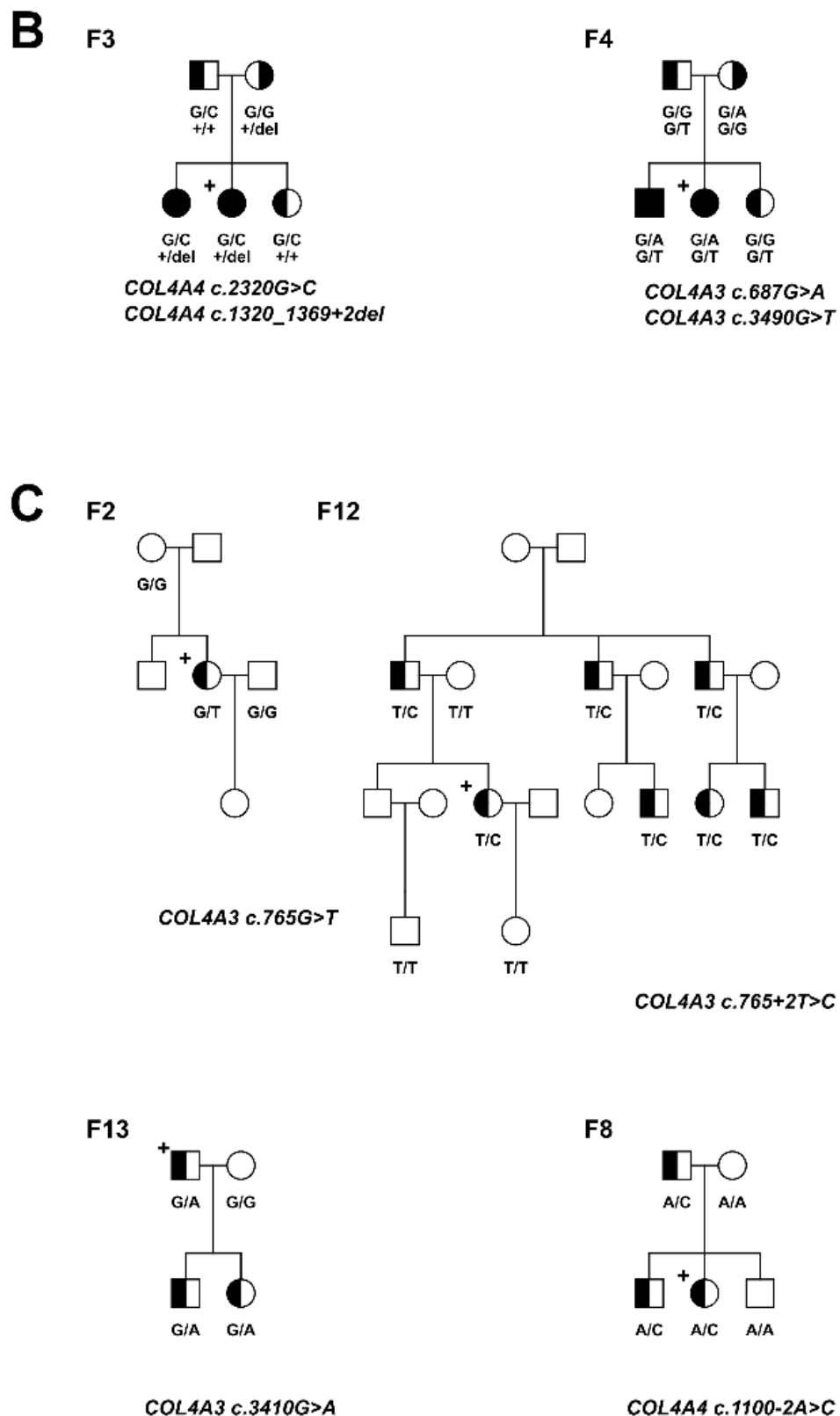
A

F5

F10   F6

COL4A5 c.2741G>A

F14

COL4A5 c.2395+1G>A   COL4A5 c.1871G>A

F1

COL4A5 c.1871G>A

COL4A5 c.1094G>A

F7

F9   F11

COL4A5 c.1780-1G>T   COL4A5 c.1033-6A>G   COL4A5 c.82G>T

Figure 9. Pedigree of investigated AS/TBMN Hungarian families. A, Families with X-linked recessive AS (XR). B, Families with autosome recessive AS (ARAS). C, Families with TBMN.

Open squares indicate males and open circles indicate females. An oblique bar indicates a deceased individual. White symbols indicate individuals without clinical sings of the AS/TBMN disease. Filled black symbols denote individuals with hematuria and/or proteinuria, hypoacusis or renal failure. The segment shaded circles and squares indicate carriers for a Chromosome 2 (COL4A3 or COL4A4) mutation. A circle that has a dot inside indicates a COL4A5 mutation carrier. The plus signs indicate the index patients whose DNA was analysed by NGS. The identified mutation(s) for each family is shown.

IV.2. Clinical exom sequencing of ECD associated genes.

Family I.

One missense variant in FGF10 gene (*c.620A>C; p. His207Pro*) was identified in the sample of the affected boy by NGS (Figure 10).

Figure 10. The found c.620A>C mutation in FGF10 gene in the sample of ECD2 patient.

Since, this is not a published variant we performed the segregation analysis of both the affected and the unaffected family members (Table 4.)

Individual	Gene	Variant c.620A> C	Consequence	Status (unaffected, carrier, ECD)
ECD2	**FGF10**	c.[=]+[620A> C]	p. [=]+[His207Pro]	LADD
ECD3	**FGF10**	c.[=]+[=]	p. [=]+[=]	normal
ECD4	**FGF10**	c.[=]+[620A> C]	p. [=]+[His207Pro]	LADD
ECD5	**FGF10**	c.[=]+[=]	p. [=]+[=]	normal
ECD7	**FGF10**	c.[=]+[620A> C]	p. [=]+[His207Pro]	LADD
ECD8	**FGF10**	c.[=]+[=]	p. [=]+[=]	normal

Table 4: Segregation analysis of Family I.

Family II.:

In the affected boy's sample after the CES analysis we found an alteration in both the WNT10A *(c.321C> A; p. Cys107Ter)* (Fig. 11.) and in the EDA (c.*752T> G; p. Val251Gly*) genes (Fig. 12.).

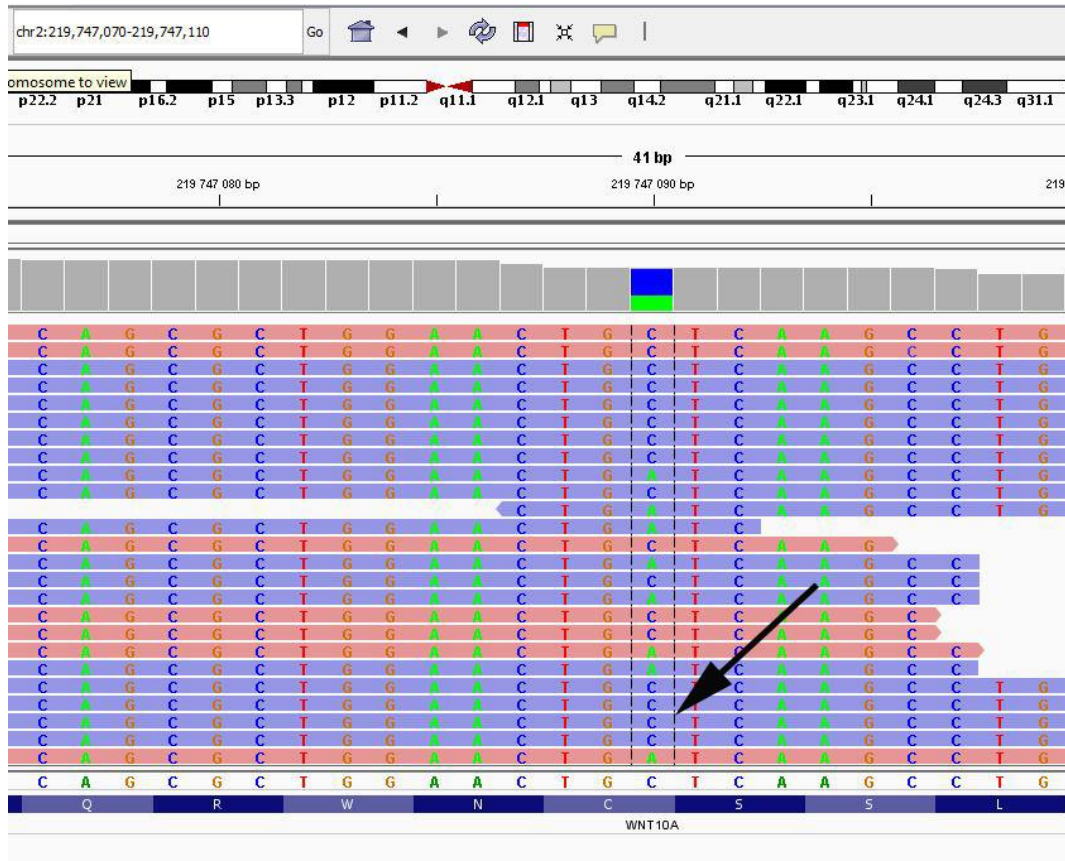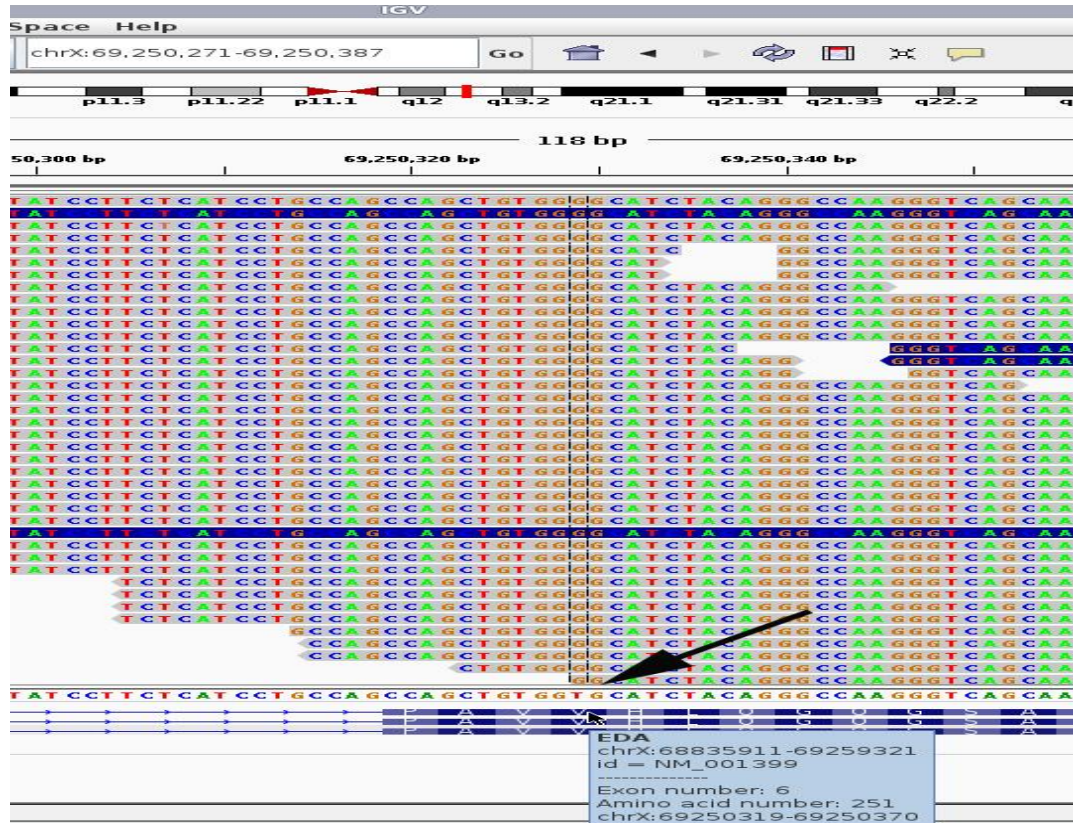**Figure 11.** The identified *c.321C*>A mutation in the WNT10A gene.

Figure 12: The identified c.752T>G mutation in the EDA gene.

The WNT10A mutation is a known pathogenic variant which causes the abnormal development of the teeth [38].

The EDA variant has not been published, yet. So we investigated its segregation in the family.

Individual	Gene	Variants in the family	Consequence	STATUS	
		EDA 752T> G, WNT10A c. 321 C>T			
ECD14	EDA	c.[752 G]	p.[Val251Gly]	XL-HED	
ECD19	EDA	c.[=]+[752 T> G]	p. [=]+[Val251Gly]	XL-HED	
ECD20	EDA	c.[752 T]	p. [=]	normal genotype	
ECD21	EDA	c.[=]+[752 T> G]	p. [=]+[Val251Gly]	XL-HED	

ECD22	EDA	c.[752 T]	p. [=]	normal genotype	
ECD14	WNT10A	c.[=]+[321 C> A]	p. [=] +[Cys107Ter]	OODD	
ECD19	WNT10A	c.[=]+[321 C> A]	p. [=] +[Cys107Ter]	OODD	
ECD20	WNT10A	c.[=]+[=]	p. [=]+[=]	normal genotype	
ECD21	WNT10A	c.[=]+[321 C> A]	p. [=] +[Cys107Ter]	OODD	
ECD22	WNT10A	c.[=]+[321 C> A]	p. [=] +[Cys107Ter]	OODD	

Table 5: Segregation analysis of Family II.

Family III.

After the evaluation the affected boy's CES result we found a not-published, homozygous, presumably pathogenic missense variant (*c.173G>A; p.Arg58His*) in KRT 74 gene (Fig.13.).
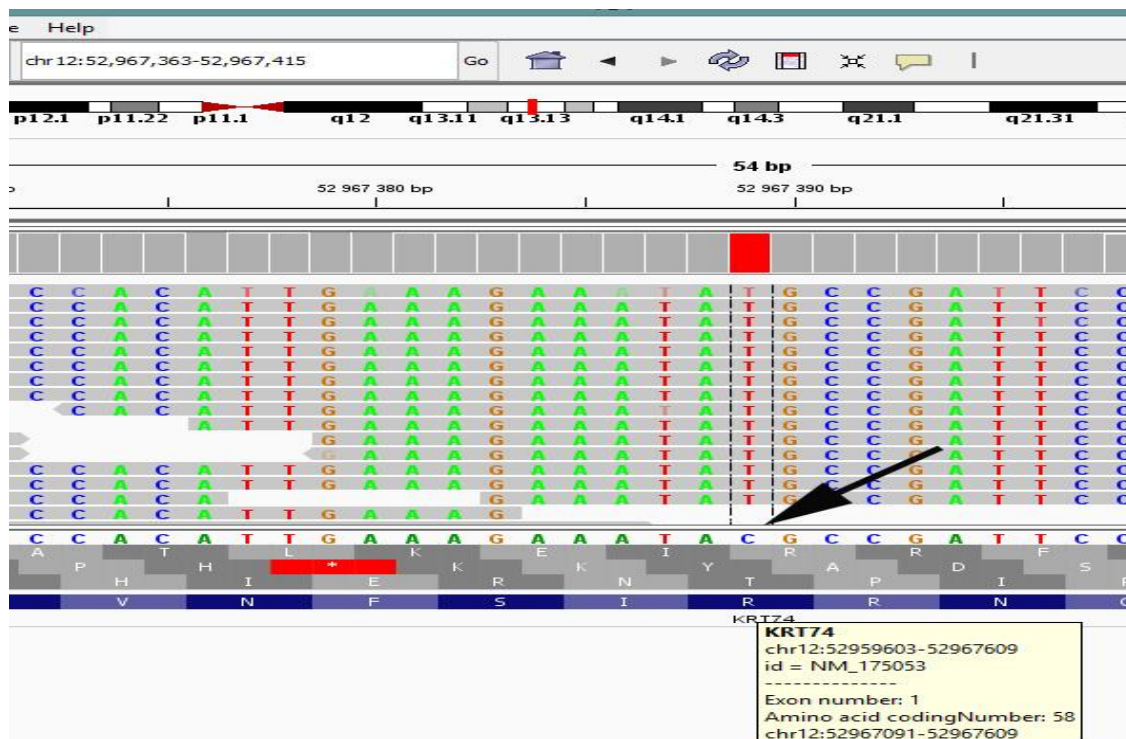

Figure 13: KRT74 mutation in the sample of the affected boy.

By the segregation analysis (Table 6) we proved that this 2 mutation were inherited from both parents.

Individual	Gene	Variant c. 173G>A	Consequence	STATUS
ECD/29	KRT74	c. [173 G>A]+[173 G>A]	p.[Arg58His] + [Arg58His]	Hair/nail type ED
ECD/30	KRT74	c.[=]+ [173 G>A]	p.[=]+[Arg58His]	carrier
ECD/31	KRT74	c.[=]+ [173 G>A]	p.[=]+[Arg58His]	carrier

Table 6. Segregation analysis of Family III.

V.   Discussion

V.1. Alport related disorders

AS and FBH are inherited renal diseases, with an incidence rate of 1:50,000 in case of AS and around 1:200 in case of FBH [36,39]. Type IV collagen is an important structural protein in basement membranes, has a triple helical structure, which consists of a Col4A3, a Col4A4 and a Col4A5 protein [40]. Since *COL4A3* and *COL4A4* genes are coded in an autosomal chromosome, a single mutation in any of these two genes only leads to FBH in both sexes, as half of the triple helix has normal structure. However, a single mutation in X chromosome coded Col4A5 protein will lead to AS in males (disrupting all forming triple helix) and varying, intermediate phenotype in females. Genetically FBH is a special carrier form of AS since it also causes phenotype (not like in classical recessive metabolic diseases such as PKU) and traditionally treated as independent disease. Successful identification of genetic mutation in any of these 3 *COL4A* genes - even before the clinical signs can be observed - could provide an early differential diagnosis for FBH/Alport patients.

We established a new, amplicon based NGS method which has three advantages: 1, it is reliable, cheaper (less than 500 euro per sample) and faster than the traditionally used Sanger capillary sequencing, 2, can provide DNA sequence information from all three *COL4A* genes simultaneously and 3, has the potential to detect single nucleotide polymorphisms (SNPs) and various sizes of insertions/deletions as well. Using our NGS panel and the approach outlined in this thesis we successfully identified causative mutations in all investigated AS/FBH instances.

In our study 37 genetic variations were identified from which there are 14 unpublished and 3 published mutations, thus causative mutations were successfully identified in all investigated cases. Out of our 17 cases 10 (as was expected) were X-linked AS (caused by 8 different *COL4A5* mutations), we had 5 FBH cases (3 with three different *COL4A3* and 2 with two different *COL4A4* mutations). In case of Family 4 we have identified two *COL4A3* mutations, while in Family 3 we have identified two *COL4A4* mutations causing autosomal AS in both cases. In our cohort we not only detected non-synonymous (amino acid change) SNPs and splice mutations, nine and seven respectively, there was also a 52 bp deletion revealed. Eight amino acid changes caused Glycine substitution in collagenous domains, what is a well-

known pathogenic alteration in Col4A proteins. Interestingly, a missense mutation (c.82G>T; p. Ala28Ser mutation in family 11, which also alters a conserved base of the splice acceptor site) was detected in a non-collagenous domain of Col4a5, which segregation was coherent with the observed clinical symptoms in the family. In two other cases the splice regions within exons, in two cases splice donor sites, while in 3 cases splice acceptor sites were affected. There was a *COL4A5* splice site mutation variant which is outside the usual +/- 1-2 splice donor/acceptor site (c.1033-6G>A in Family 7), that although fulfilled all of our mutation classification criteria (detailed in Supplement 1.) without additional functional test we only presumed that it is pathogenic.

Up to now there are less than 1000 classified causative variants in public databases (HGMD, dbSNP, ClinVar) regarding the three *COL4A* genes causing AS/FBH which are around 4% of the whole exonic target region. The vast majority of the published mutations coming from X-linked *COL4A5* data, even if either of the two autosomal *COL4A* genes is longer and their mutations can cause not only the more frequent FBH but also AS as well. One of the possible explanations could be that lots of FBH cases (caused by a single mutation either of *COL4A3* or *COL4A4*) remain undiagnosed. With the introduction of simultaneous sequencing of all three *COL4A* genes by NGS, we should expect more sequence data from *COL4A3* and *A4* in the near future. Both mutations and polymorphisms in the 3 investigated *COL4A* genes are thought to be highly population-specific (due to non-existing selection pressure in case of the polymorphisms and only low selection pressure in case of FBH). But, to our surprise, none of the 21 identified genetic variations in our control group proved to be unpublished and Hungarian population specific. In scientific literature an autosomal dominant (AD) form of AS is also mentioned. Most of the autosomal dominant AS data comes from an era where the technical limit (error rates) of sequencing was comparable to the claimed frequency (around 1%) of this form of AS. However, a recent paper using NGS claims that there was an astonishing portion of autosomal dominant AS (31%) of their solved cases [10]. In the light of these data we should expect to find around 5 AD AS among our 17 cases, but we have not found any.

In our opinion there is a widespread confusion in the clinical and pathological terminologies used to describe FBH, TBMN and AS. FBH is used by nephrologist as a descriptive term for cases when only haematuria (usually non progressive, without proteinuria or impaired hearing or vision) could be detected throughout the family, which is only a subset (although the largest) of all the TBMN cases. TBMN on the other hand is a term used by pathologists describing the thin basement membrane phenomenome of the kidney, which in

some cases result slowly progressive haematuria with proteinuria and as an age-related process can reach ESRD in elderly patients. In their paper Fallerini C. et al presented 7 AD AS families in detail with a single identified autosomal *COL4A3* or *A4* mutation [11]. Their conclusion could be biased by the fact that 45% of their AS/FBH cases went without mutation identification. It is also worth to note that in all of their published AD AS cases ESRD presented only in older age (between 52-74 with an average of 63 years), without a single case with eye problem and with only 4 elderly cases with hearing loss. According to WHO data (updated in 2015 March) approximately one-third of people over 65 years of age are affected by disabling hearing loss which is fully agrees with their reported rate.

Further on there is no known explanation for the biological foundation of AD AS. In cases where collagen triple helix formed by one (as in case of Type VII collagen, 3 identical chains coded by *COL7A1* gene) or two different collagen chains (as in case of Type I collagen, consisting of two chains coded by *COL1A1* and one coded by *COL1A2*) the dominant negative effect of a single autosomal mutation is known [41,42]. In these cases, due to the formation of homodimers/homotrimers, a single mutation in an autosomal gene could lead to stronger phenotype than a null allele. The mutated protein will bind to the normal protein and result only around 25% of functional dimer (and even less in case of trimers) while autosomal null alleles lead to ~50% functional protein in heterozygotes. This is not the case in the forming Col4a3-4-5 triple helix, where all three chains are different and coded by different genes, so it does not comply with the current biological model of negative dominant effect. FBH biologically is the result of reduced amount of proper Col4a3-4-5 triplex, while AS is the result of the lack of the functional protein.

Based on our own observation on the genetic analysis presented here; confusing interpretation on age related clinical symptoms and the lacking of biological foundation we think that AD AS, as genetic term doesn't exist and shouldn't be used. We think that the alleged autosomal Alport cases genetically are either a single severe autosomal mutation leading to phenotype in between FBH/AS only in higher age (so it is a slowly evolving TBMN) or it could also be the result of unsuccessful genotyping of autosomal AS due to technical (for example the use of ineffective insertion/deletion detection algorithm) or differences in their genetic variation classification approach that heavily relies on prediction softwares.

As far as we see Alport syndrome and its related diseases and their observed clinical/pathological heterogeneity (heamaturia, ESRD, hypoacusis, TBMN) is such a biological problem, which has quantitative inheritance elements. In individuals without

mutations in the three investigated *COL4A* genes has the full amount of fully functional triple-helix. In the individuals (independently from gender) with a mutation in either *COL4A3* or *COL4A4* gene there is about half of the theoretical maximum amount of functional triple-helix. If there is a *COL4A5* mutation in a male, than there isn't any fully functional triple-helix formed. In females, since random inactivation of X chromosome occurs, approximately 50% of fully functional triple-helix will form, but this could be altered (positively or negatively) by the imbalances of X inactivation. This also could explain the variable phenotype even in a family. The situation in the case of extracellular basal membrane is even more complex, since the observed clinical phenotype does not depend on a single cell genotype rather from the genotype ratio within Type IV collagen producing cell population. The other sources of the heterogeneity naturally are the mutations causing different level of malfunction in the forming triple-helix. This phenotypic variability leads to introduce the term semi-dominant inheritance, which has an intermediate phenotype different from both dominant (since the phenotype of a heterozygote is not the same as a homozygote mutant) or recessive (since the phenotype of a heterozygote is not the same as a homozygote normal). As a result of this we can view Alport and its related diseases as a spectrum disorder, where the biological base of the different phenotypes is the amount of fully functional (or impaired/null) Type IV collagen formed. The two extremities of this spectrum are FBH with only mild, non-progressing haematuria, while the other end is Alport syndrome with serious kidney, eye and hearing phenotype.

Because the prognosis of FBH and AS is entirely different while the early symptoms could be similar, both the genetic classification and the above mentioned distinction in our view is extremely important. By providing correct genetic data and by establishing the correct inheritance pattern within family, it could help both clinicians and genetic counsellors 1) to communicate and make an early, informed decision about the possible speed and direction of the progression in the family, 2) to establish the circle of the family members who need regular checking, 3) to select a possible kidney donor within the family and 4) support positive family planning. The genetic test could lead to the earliest diagnosis even before the manifestation of clinical symptoms that would be a big benefit for patients, for their families and for the health systems as well. For patients with some of the earlier clinical symptoms or even before, genetic diagnosis could mean preventive treatment such as using ACE inhibitors which could delay the onset of end-stage renal failure and improve life expectancy, even when begun before the onset of proteinuria [5,6,30]. For family members the biggest benefit could be the possibility for positive family planning, and the awareness of possible kidney problems. And for the health system the

cheap genetic diagnostic procedure eliminates the need for long and uncertain clinical staying, and lots of unnecessary, less conclusive and much more expensive tests and examinations. The best tool for this job is the DNA-based molecular genetic analysis of the proband and his/her relatives. Using peripheral blood as DNA source for genetic analysis causes much less trauma for patient (especially children) compared to the kidney biopsy, which carries certain level risk of severe complications[31]. The result of the DNA sequencing is independent of the age of the proband, from the stage of the disease and in most cases provides quick and accurate confirmation of the diagnosis. It can be also performed on a single person independently when the family history is not available. Furthermore, in cases where no detailed clinical and pathological data is available genetic analysis could be the first choice.

V.2. Ectodermal dysplasia related disorders

There are more than 150 types of Ectodermal dysplasia described so far, with a combined incidence of 7/10000. Every organs (skin, skin appendages, hair, teeth, eccrine-glands) and tissues which are developing from the embryonic ectoderm can be affected. More than 150 types of ECD were described coded by more than 140 genes. Due to the large number of the affected genes it was nearly impossible to confirm the clinical diagnoses by genetic analysis in every cases. To diagnose ECD we choose the CES sequencing approach, where we can investigate all the known pathogenic ECD genes at the same time. In my thesis I show the genetic analysis of three ECD patients and their families by CES and Sanger sequencing.

In the first family we found a FGF10 (fibroblast growth factor 10) gene (c.620A> C) mutation. The encoded protein is part of a family of proteins which take part in important processes during the early embryonal development (such as cell division, regulation of cell growth/maturation, wound healing, the lacrimal system, the ears, the salivary glands (saliva glands in mouth, teeth etc.). We validated the presence in the proband by Sanger sequencing. This mutations was not published in the literature for this we performed the segregation analysis in the family. The segregation analysis (Table 4) shows that only the affected family members have the mutant variant, the unaffected family members carry the wild type variant. The clinical phenotype agree the symptoms which were described in OMIM caused by mutations in FGF10 gene. Mutations in FGF10 gene known to cause the lacrimo-auriculo-dento-digital (LADD) syndrome [43] with autosomal dominant inheritance therefore based on the detailed clinical -, genetic analysis and our variant assertion criteria we confirmed the LADD type of ED in this family.

In Family II. we found two missense mutations in two different genes. One is in the WNT10A gene (c.321C> A; p. Cys107Ter) and another one in the EDA gene (c.752T> G; p.*Val251Gly*). The identified WNT10 mutation is a known pathogenic, which was published by Axel Bohrin et. al[38]

Mutations in WNT10A gene know to be responsible for tooth malformations and plays role in another ED type, Odontoonychodermal dysplasia (OODD). Fadhil et al. first described phenotype related WNT10A mutation and the first demonstrated, that altered WNT signalling pathway could contribute to ectodermal dysplasia (OODD)[44]. However in randomly investigated samples with tooth agenesis only 40% had at least one mutation in the WNT10A gene [45,46] Phenotype of family II. agrees with this findings. The affected boy's nephew whose got only WNT10A mutation has only Odontoonychodermal dysplasia. Based on the segregation analysis and the phenotype we confirmed the OODD type of ED in ECD14, ECD19, ECD20, ECD22. The identified novel EDA mutation was not described in the literature so we had to confirm the pathogen status. EDA gene encode the Ectodysplasin-A transmembrane protein which consist of 391 amino acids coded by 8 exons. The 6th exon codes 18 amino-acids (247-265) which is part of the extracellular domain of the protein UniProt database). This mutation is likely to affect that part of the extracellular EDA which can interact with EDA-receptor (EADR). This interaction is responsible for initialisation of the proper differentiation of the sweat glands.

In view of this and the segregation analysis (only the affected members have the mutant gene) we concluded that the identified novel EDA mutation and the published causative WNT10A mutation was responsible for the mixed phenotypes in this family. The affected boy, his mother, his grandmother and his sister have both Hypohidrotic ectodermal dysplasia and OODD. The proband's sister has 3 children. Two of them have only tooth agenesis and have normal sweating abilities. We confirmed his clinical status by genetic analysis. He only inherited the WNT10A mutation which responsible for OODD.

In Family III. we found a KRT74 gene variant both in homozygous and in heterozygous (carrier) forms. KRT74 belongs to the family of type II keratins that are specifically expressed in the inner root sheath of hair follicles and in nail matrix and its mutations known to cause ectodermal dysplasia type 7 [47–49] . The identified mutation in KRT74 is in the first exon at c. 173G>A which causing Arg to His amino-acid change at the position of 58. The investigated boy is carrying this mutation in homozygous form, while both of his parents are carrying this mutation in heterozygote fashion which is an agreement with the inheritance mode of this

disease (autosomal recessive). To date, according to our knowledge, the medical relevance of this variant is not described elsewhere.

Up to now there has not been any cure for ECD. However, recently researchers developed a new therapy for X-linked Hypohidrotic form of ED (XLHED) called EDI200 [50]. Treatment of pregnant mother with EDI200 or of the newborn baby shortly after birth leads to the normal phenotype (normal working hair follicules and sweat glands)[50]. To start the therapy in time the early genetic diagnosis which confirms is crucial. To exclude other genes which cause likely XLHED phenotype (EDAR, EDARADD, etc.) not enough to investigate only the EDA gene. We need to investigate all ECD related genes. To meet these requirements, we think that CES could be a reliable and economical choice. In addition the simultaneous investigation of all ECD related genes open the door to the further development gene specific therapies[50]. Besides the mentioned three cases we investigated other 5 families affected with different types of ECD. To diagnose the complex diseases is a big challenge to both the clinicians and to the diagnostic laboratories. Earlier, we made genetic tests to confirm the co-segregation pattern of haematuria with STR markers and LOD score analysis and we performed mutation screening with High Resolution Melting (HRM) analysis and Sanger sequencing of a different set of complex diseases. However this approach was not successful to identify the causative mutations in many cases. In cases where more than a single gene is responsible for the disease, we need diagnostic tools that can simultaneously investigate all the associated genes. In this study I first presented cases with Alport syndrome and its related diseases where we developed a custom targeted NGS panel capable to detect genetic variations in all three COL4A genes simultaneously. In the second part of my thesis I presented three cases with clinical diagnoses of ECD, where we used a commercially available CES kit (TruSight One, Illumina) in order to identify causative mutations in any of the 141 known ECD-causing genes.

The combined use of NGS data and targeted validation and segregation analysis of the mutations in the analysed cohort proved to be a suitable strategy for the assessment of correct genetic diagnosis in complex diseases

Acknowledgements

There are number of people without whom this thesis might not have been written, and to whom I am greatly indebted:

To my supervisor **Dr. Emőke Endreffy** who supervised my scientific work when I was an undergraduate student and let me continue my work as a PhD student.

Let me express my gratitude to **Dr. Zoltán Maróti** for his leaderships and professional advice which I have had the good fortune to benefit from while finalizing my thesis and ever since. I thank to **Dr. Tibor Kalmár** for his leaderships during my PhD studies.
I thank you both again for your trust and friendship which allows me to stand here in front of you this day.

I should also like to thank *Dr. Zsolt Rázga* for his kind help during my PhD study.

I am also very grateful for my dear colleagues: *Dr.Ibolya Haszon, Dr. Eszter Karg, Dr. Mária Sinkó and Dr. Ákos Baráth* whose suggestions helped to improve the quality of my work,

I would like to say a special thanks to *Mrs Erzsebet Borzási, Mrs Ildikó Csipő, Mrs Mária Mustoha and Mrs Tünde Vida-Szűcs* for the technical assistance.

To the ex-chair of the Department of Pediatrics *Prof. Dr. Sándor Túri* and to the present chairman of Department of Pediatrics *Dr. Csaba Bereczki* for making my research possible and being so supportive during my undergraduate and doctoral studies.

I would like to thank my family *Mrs. Teréz Kovács* and *Mihály Kovács*, my *brother Zsolt Kovács* for their unconditional support, both financially and emotionally throughout my studies

Supplement 1.

Genetic Diagnostic Laboratory

Assertion Criteria

For variant description we follow the recommendations of the Human Genome Variation Society (Nomenclature for the description of sequence variants **). For variant classification we** follow the guidelines of the ACMG Laboratory Practice Committee Working Group (Richards et al 2008. Gen in Med. 10: 294-300).

Sequencing variants are classified in one of five categories. Below is a summary of the guidelines used by our laboratory (as of 1/1/2016)

For clinical whole exome or targeted (re)sequencing, variants classified as pathogenic variants will be confirmed by Sanger sequencing.

1. **Pathogenic** For a variant to be classified as a pathogenic variant it must be located in a gene known to be responsible for disease and the phenotype of the patient being tested is consistent with that disease.
   a. Variants predicted to result in the loss of protein function (may or may not have been previously reported in patients with disease)
      1. frameshift (an insertion or deletion that is not a multiple of 3 nucleotides)
      2. nonsense (introduction of a premature stop codon)
      3. splice junction (at positions +1,+2, -1 and -2 in an intron)
      4. change in an initiation codon
      5. change in the termination codon
   b. Variants predicted to result in an amino acid replacement (missense) with one of the following conditions met:

1. variant demonstrated to result in reduced protein function (loss of function), or aberrant protein function (gain of function) in an appropriate functional assay

2. common disease causing pathogenic variant in a specific population based on evidence in the literature

3. variant reported in multiple affected individuals and demonstrated to segregate with disease in at least a single family

c. Variants demonstrated to result in aberrant splicing in an appropriate functional assay (eg. intronic or silent)

2. **Likely Pathogenic**

   a. Recessive conditions: (all the following conditions must be met)

      1. diagnosis has been confirmed by biochemical testing or patient phenotype is specific for disease

      2. variant located on opposite chromosome from a known disease causing pathogenic or likely pathogenic variant

      3. variant occurs at an evolutionarily conserved nucleotide and/or amino acid

      4. variant not present in dbSNP, EVS (Exome Variant Server) or other publicly available database at a frequency consistent with being a benign variant

   b. Dominant conditions: (all of the following conditions must be met)

      1. variant segregates with phenotype in the family being tested **or**

      2. testing parental samples demonstrates that the variant occurred *de novo*

      3. variant not present in dbSNP, EVS (Exome Variant Server) or other publicly available database at a frequency consistent with being a benign variant

3. **Benign variant** (one of the following conditions must be met)*

    a. Variant reported in dbSNP, EVS, locus specific databases or EGL database at a population frequency higher than expected given the prevalence of the disease and mode of inheritance.

    b. Variant reported in a control population at a frequency inconsistent with being causative of disease

    c. Other evidence from published literature that indicates the variant has no effect on function

*Benign variants are interpreted as described above and not reported in clinical reports. A list of these variants is available upon request.

4. **Likely benign variant** (one of the following conditions must be met)

    a. Variant found heterozygous (for dominant) or homozygous (for recessive) in multiple unaffected individuals.

    b. Variant found in *cis* with a pathogenic variant in multiple unrelated individuals

    c. Variant found in an unaffected family member (for dominant)

5. **Variant of unknown clinical significance (VOUS)**

If a variant is in a gene that is not consistent with the patient's phenotype, it will be reported as a variant of unknown significance and additional clinical information may be needed. Variants that are identified in genes not known to cause disease will be also classified as variants of unknown significance.

 (if unable to classify the variant in one of the four categories above, it will be classified as a VOUS)

    a. Variant not reported in HGMD, locus specific databases, published literature, dbSNP or EVS

    b. Variant reported in dbSNP or EVS, but at an allele frequency insufficient to rule out clinical significance based on mode of inheritance and severity of disorder

   c. Variant reported in a single individual with insufficient segregation and/or functional data

   d. Variant reported in a single individual with inadequate clinical information

Note: In the presence of conflicting data the term variant of *uncertain* significance may be used.

References

1. Alport, A. C. HEREDITARY FAMILIAL CONGENITAL HAEMORRHAGIC NEPHRITIS. *Br. Med. J.* **1,** 504–506 (1927).

2. REYERSBACH, G. C. & BUTLER, A. M. Congenital hereditary hematuria. *N. Engl. J. Med.* **251,** 377–380 (1954).

3. Mogensen, C. E. Preventing end-stage renal disease. *Diabet. Med.* **15 Suppl 4,** S51–6 (1998).

4. Savige, J. *et al.* Expert guidelines for the management of Alport syndrome and thin basement membrane nephropathy. *J. Am. Soc. Nephrol.* **24,** 364–375 (2013).

5. Gross, O. *et al.* Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy. *Kidney Int.* **81,** 494–501 (2012).

6. Kashtan, C. E. *et al.* Clinical practice recommendations for the treatment of Alport syndrome: a statement of the Alport Syndrome Research Collaborative. *Pediatr. Nephrol.* **28,** 5–11 (2013).

7. Kashtan, C. E. Alport syndromes: phenotypic heterogeneity of progressive hereditary nephritis. *Pediatr. Nephrol.* **14,** 502–512 (2000).

8. Soininen, R., Huotari, M., Hostikka, S. L., Prockop, D. J. & Tryggvason, K. The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. *J. Biol. Chem.* **263,** 17217–17220 (1988).

9. Mencarelli, M. A. *et al.* Evidence of digenic inheritance in Alport syndrome. *J. Med. Genet.* **52,** 163–174 (2015).

10. Buzza, M., Wilson, D. & Savige, J. Segregation of hematuria in thin basement membrane disease with haplotypes at the loci for Alport syndrome. *Kidney Int.* **59,** 1670–1676 (2001).

11. Fallerini, C. *et al.* Unbiased next generation sequencing analysis confirms the existence of autosomal dominant Alport syndrome in a relevant fraction of cases. *Clin. Genet.* **86,** 252–257 (2014).

12. Torra, R., Tazon-Vega, B., Ars, E. & Ballarin, J. Collagen type IV (alpha3-alpha4) nephropathy: from isolated haematuria to renal failure. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* **19,** 2429–2432 (2004).

13. Endreffy, E. *et al.* Collagen type IV nephropathy: genetic heterogeneity examinations in affected Hungarian families. *Mol. Cell. Probes* **25,** 28–34 (2011).

14. Chatterjee, R. *et al.* Targeted exome sequencing integrated with clinicopathological information reveals novel and rare mutations in atypical, suspected and unknown cases of Alport syndrome or proteinuria. *PLoS One* **8,** e76360 (2013).

15. Deshmukh, S. & Prashanth, S. Ectodermal dysplasia: a genetic review. *Int. J. Clin. Pediatr. Dent.* **5,** 197–202 (2012).

16. Ghosh, S., Ghosh, E. & Dayal, S. Autosomal recessive anhidrotic ectodermal dysplasia: a rare entity. *Indian J. Dermatol.* **59,** 422 (2014).

17. Wright, J. T., Grange, D. K. & Richter, M. K. in (eds. Pagon, R. A. et al.) (1993).

18. Chikkannaiah, P., Nagaraju, S., Kangle, R. & Gosavi, M. Perinatal Autopsy Findings in a Case of De Novo Hypohidrotic Ectodermal Dysplasia. *J. Lab. Physicians* **7,** 131–133 (2015).

19. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35,** W71–4 (2007).

20. Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* **133,** 1–9 (2014).

21. Fokkema, I. F. A. C. *et al.* LOVD v.2.0: the next generation in gene variant databases. *Hum. Mutat.* **32,** 557–563 (2011).

22. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

23. Crockett, D. K. *et al.* The Alport syndrome COL4A5 variant database. *Hum. Mutat.* **31,** E1652–7 (2010).

24. Longo, I. *et al.* COL4A3/COL4A4 mutations: from familial hematuria to autosomal-dominant or recessive Alport syndrome. *Kidney Int.* **61,** 1947–1956 (2002).

25. Heidet, L. *et al.* Structure of the human type IV collagen gene COL4A3 and mutations in autosomal Alport syndrome. *J. Am. Soc. Nephrol.* **12,** 97–106 (2001).

26. Wang, Y. Y. *et al.* COL4A3 mutations and their clinical consequences in thin basement membrane nephropathy (TBMN). *Kidney Int.* **65,** 786–790 (2004).

27. Hoefele, J. *et al.* Novel heterozygous COL4A3 mutation in a family with late-onset ESRD. *Pediatr. Nephrol.* **25,** 1539–1542 (2010).

28. Papazachariou, L. *et al.* Frequency of COL4A3/COL4A4 mutations amongst families segregating glomerular microscopic hematuria and evidence for activation of the unfolded protein response. Focal and segmental glomerulosclerosis is a frequent development during ageing. *PLoS One* **9,** e115015 (2014).

29. Badenas, C. *et al.* Mutations in theCOL4A4 and COL4A3 genes cause familial benign hematuria. *J. Am. Soc. Nephrol.* **13,** 1248–1254 (2002).

30. Frasca, G. M. *et al.* Thin glomerular basement membrane disease: clinical significance of a morphological diagnosis--a collaborative study of the Italian Renal Immunopathology Group. *Nephrol. Dial. Transplant* **20,** 545–551 (2005).

31. Voskarides, K. *et al.* COL4A3/COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy. *J. Am. Soc. Nephrol.* **18,** 3004–3016 (2007).

32. Ciccarese, M. *et al.* Identification of a new mutation in the alpha4(IV) collagen gene in a family with autosomal dominant Alport syndrome and hypercholesterolaemia. *Nephrol. Dial. Transplant* **16,** 2008–2012 (2001).

33. Baek, J.-I. *et al.* Identification of novel variants in the COL4A4 gene in Korean patients with thin basement membrane nephropathy. *Indian J. Med. Res.* **129,** 525–533 (2009).

34. Buzza, M. *et al.* Mutations in the COL4A4 gene in thin basement membrane disease. *Kidney Int.* **63,** 447–453 (2003).

35.   Ma, J. *et al.* Twenty-one novel mutations identified in the COL4A5 gene in Chinese patients with X-linked Alport's syndrome confirmed by skin biopsy. *Nephrol. Dial. Transplant* **26,** 4003–4010 (2011).

36.   Martin, P. *et al.* High mutation detection rate in the COL4A5 collagen gene in suspected Alport syndrome using PCR and direct DNA sequencing. *J. Am. Soc. Nephrol.* **9,** 2291–2301 (1998).

37.   Knebelmann, B. *et al.* Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. *Am. J. Hum. Genet.* **59,** 1221–1232 (1996).

38.   Bohring, A. *et al.* WNT10A mutations are a frequent cause of a broad spectrum of ectodermal dysplasias with sex-biased manifestation pattern in heterozygotes. *Am. J. Hum. Genet.* **85,** 97–105 (2009).

39.   Levy, M. & Feingold, J. Estimating prevalence in single-gene kidney diseases progressing to renal failure. *Kidney Int.* **58,** 925–943 (2000).

40.   Sundaramoorthy, M., Meiyappan, M., Todd, P. & Hudson, B. G. Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. *J. Biol. Chem.* **277,** 31142–31153 (2002).

41.   Fritsch, A. *et al.* Dominant-negative effects of COL7A1 mutations can be rescued by controlled overexpression of normal collagen VII. *J. Biol. Chem.* **284,** 30248–30256 (2009).

42.   Gajko-Galicka, A. Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans. *Acta Biochim. Pol.* **49,** 433–441 (2002).

43.   Milunsky, J. M., Zhao, G., Maher, T. A., Colby, R. & Everman, D. B. LADD syndrome is caused by FGF10 mutations. *Clin. Genet.* **69,** 349–354 (2006).

44.   Fadhil, M., Ghabra, T. A., Deeb, M. & Der Kaloustian, V. M. Odontoonychodermal dysplasia: a previously apparently undescribed ectodermal dysplasia. *Am. J. Med. Genet.* **14,** 335–346 (1983).

45.   Mues, G. *et al.* The WNT10A gene in ectodermal dysplasias and selective tooth agenesis. *Am. J. Med. Genet. A* **164A,** 2455–2460 (2014).

46.   Haara, O. *et al.* Ectodysplasin and Wnt pathways are required for salivary gland branching morphogenesis. *Development* **138,** 2681–2691 (2011).

47.   Rasool, M. *et al.* Autosomal recessive pure hair and nail ectodermal dysplasia linked to chromosome 12p11.1-q14.3 without KRTHB5 gene mutation. *Eur. J. Dermatol.* **20,** 443–446 (2010).

48.   Langbein, L., Rogers, M. A., Praetzel, S., Winter, H. & Schweizer, J. K6irs1, K6irs2, K6irs3, and K6irs4 represent the inner-root-sheath-specific type II epithelial keratins of the human hair follicle. *J. Invest. Dermatol.* **120,** 512–522 (2003).

49.   Shimomura, Y., Wajid, M., Petukhova, L., Kurban, M. & Christiano, A. M. Autosomal-dominant woolly hair resulting from disruption of keratin 74 (KRT74), a potential determinant of human hair texture. *Am. J. Hum. Genet.* **86,** 632–638 (2010).

50.   Huttner, K. Future developments in XLHED treatment approaches. *Am. J. Med. Genet. A* **164A,** 2433–2436 (2014).

Appendix

I.

II.