Thesis

Evaluation of data filtering strategies for whole exome data using a combination of homozygosity mapping and software-based genome-wide variant prioritization

submitted by

Sophie Teres Spitzer

in partial fulfillment of the requirements for the degree of

Doktorin der gesamten Heilkunde

(Dr^{in}. med. univ.)

at the

Medical University of Graz

executed at the

Diagnostic and Research Institute of Human Genetics

under the supervision of

Christian Windpassinger, Assoz. Prof. Priv.-Doz. Mag. Dr.rer.nat.

Graz, 08.Januar 2026

Declaration of Academic Integrity

I hereby confirm that the present diploma thesis is the result of my own independent scholarly work. I also confirm that in all cases, where material from the work of others (in books, articles, essays, dissertations, and on the internet) is acknowledged, quotations and paraphrases are clearly indicated. No material other than that cited in the reference list has been used. I have read and understood the Medical University's regulations and procedures concerning plagiarism.

Furthermore, I hereby declare that if artificial intelligence (AI) tools were used for the generation and/or correction of certain text passages in the creation of this work, such employment was conducted in compliance with ethical principles, academic integrity, and the regulations of my university. Additionally, it was ensured that this usage was transparently disclosed and appropriately attributed.

Graz, 08. Januar 2026 Sophie Teres Spitzer , m.p.

Acknowledgements

First and foremost, I would like to thank my supervisor, Assoz.Prof.Priv.-Doz.Mag.Dr.rer.nat. Christian Windpassinger, for his continuous support and encouragement throughout the work on my thesis. His expertise, feedback and patience helped me navigate the challenges of academic research.

I also want to express my gratitude to the staff of the neurogenetic laboratory, in particular Amtsdirektor Ing. Johann Semmler-Bruckner, for his invaluable advice and support during this time. Furthermore, I want to thank Anita Harb and Irmgard Pölzl for their indispensable assistance in the laboratory. Special thanks to my fellow student Birgit Sommerauer, MSc, whose experience in the laboratory proved to be of immense value to me.

Finally, I am grateful to my family and my friends for their unconditional love, support, and understanding. Their belief in me has been a great source of strength.

To everyone who has contributed in any way to the completion of this thesis, thank you.

Zusammenfassung

Erbliche Hautkrankheiten werden durch eine Vielzahl unterschiedlicher Mutationen in verschiedenen Genen verursacht, sind häufig selten und daher auch schwieriger zu diagnostizieren. Abhängig von der spezifischen Mutation kann der Phänotyp stark variieren. Eine Diagnose ist für die Behandlung der Krankheit von entscheidender Bedeutung. Die Teilnehmer dieser Studie waren drei verschiedene pakistanische Familien mit konsanguinen Beziehungen und erblichen Hautkrankheiten.

DNA-Proben wurden aus Vollblut gewonnen, das von betroffenen und nicht betroffenen Personen in den Familien abgenommen wurde. Es wurde eine Whole-Exome-Sequenzierung durchgeführt, und die Daten wurden verwendet, um homozygote Regionen durch Homozygotie-Kartierung ausfindig zu machen und weiters mithilfe softwarebasierter Tools möglich Varianten zu filtern. Durch diesen Prozess konnten in allen drei Familien Kandidatenvarianten identifiziert werden. Diese wurden dann durch Segregation der Varianten mittels Sanger-Sequenzierung bestätigt.

In einer Familie wurde die Erkrankung als ektodermale Dysplasie identifiziert, mit einer Deletion von 4 bp (p.Lys240fs) in Exon 8 des EDAR-Gens. Diese Mutation führte wahrscheinlich zu einem verkürzten Protein, welches die Struktur der Haare und Zähne der betroffenen Personen beeinträchtigte. Die zweite Familie leidet an Porphyria variegata, einer seltenen genetischen Stoffwechselstörung, deren häufigstes Symptom eine Lichtempfindlichkeit der Haut ist. Eine Missense-Mutation (c.1387G>A) wurde in Exon 13 im PPOX-Gen gefunden und führte zu einer Veränderung der Aminosäuresequenz (p.Gly463Arg) und damit zu der Erkrankung.
In der dritten Familie wurde eine Missense-Mutation (c.400T>C) in Exon 3 des Gens TGM1 gefunden. Dieses Gen ist eine bekannte Mutationsstelle für Patienten mit autosomal-rezessiver kongenitaler Ichthyose (ARCI), welche auch dem Phänotyp der dritten Familie entspricht.

Die Entdeckung der ursächlichen Varianten ermöglicht eine definitive Diagnose für die Patienten in diesen drei Familien. Den Patienten können

Informationen und in einigen Fällen auch Behandlungen angeboten werden. Darüber hinaus treibt diese Arbeit die Forschung voran, indem sie die Datenbanken zu Krankheitsgenen erweitert und neue pathogene Varianten aufdeckt

Abstract

Hereditary skin diseases are caused by a multitude of different mutations in different genes. Many of these diseases are rare and therefore harder to diagnose. Depending on the specific mutation the phenotype can vary widely. A diagnosis is vital for the treatment of the disease. The participants of this study were three different Pakistani families with consanguineous relationships and hereditary skin diseases.

DNA samples were won from whole blood, given by both affected and unaffected individuals in the families. Whole exome sequencing was performed, and the data was used for finding homozygous regions by homozygosity mapping and using software-based tools for variant filtering. Through this process candidate variants could be identified in all three families. Sanger segregation confirmed that the variants were disease-causing.

In one family the condition was identified as ectodermal dysplasia, with a deletion of 4bp (p.Lys240fs) in exon 8 of the EDAR gene. This mutation likely resulted in a truncated protein, affecting the structure of hair and teeth in the affected individuals.
The second family suffers from variegate porphyria, a rare genetic metabolic disorder, whose most common symptom is skin photosensitivity. A missense mutation (c.1387G>A) was found in exon 13 in the PPOX gene, leading to a change in the amino acid sequence (p.Gly463Arg) and resulting in the condition.
In the third family a missense mutation (c.400T>C) was found in exon 3 in the gene TGM1. This gene is a known mutation site for patients with autosomal recessive congenital ichthyosis (ARCI), matching the phenotype exhibited by the third family.

The discovery of the causal variants offered a definitive diagnosis for the patients in these three families. Information and in some cases, treatment can be offered to the patients. Furthermore, such work advances research by expanding disease-gene databases and revealing novel pathogenic variants.

Table of contents

Abbreviations

°C	degree Celcius
µL	microliter
A	adenine
ala	alanine
ARCI	autosomal recessive congenital ichthyosis
arg	arginine
asp	aspartic acid
BAM	Binary Alignment Map
bp	base pairs
C	cytosine
chr	chromosome
CIE	Congenital Ichthyosiform Erythroderma
cys	cysteine
dest.	distilled
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
ED	ectodermal dysplasia
EDTA	Ethylenediaminetetraacetic Acid
EtOH	ethanol
FASTQ	"Fast" Quality scores
g	gravitational force
G	guanine
glu	glutamic acid
gly	glycine
H2O	water
his	histidine
indel	insertion or deletion
KCl	potassium chloride
lys	lysine
mg	milligram
MgCl	magnesium chloride
min	minute
mL	millilitre

mM	millimole
NaCl	sodium chloride
NFW	nuclease-free water
nm	nanometer
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
pH	potential of hydrogen
pmol	picomole
R. Temp	Ramp temperature
ROH	Region of Homozygosity
rpm	revolutions per minute
SDS	Sodium Dodecyl Sulfate
sec	second
SNP	Single Nucleotide Polymorphism
SNV	Single nucleotide variant
T	thymine
TBE	Tris-Borat-EDTA
TKM	Tris-KCl-MgCl2
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
tyr	tyrosine
U	unit
UCSC	University of California Sana Cruz Genome Browser
UV	Ultraviolet
VCF	variant call format
WES	Whole Exome Sequencing
WGS	whole genome sequencing

Table of figures

Table of tables

1. Introduction

1.1. The Skin and genetic diseases

The skin is the largest organ of the body and serves as an important barrier, protecting the body from injury, water loss, infection and many more. A variety of cells are needed to uphold this barrier. Between cells complex interactions facilitated by different molecules take place and mutations in one of those molecules can have far reaching consequences for an individual. Inherited skin diseases caused by genetic mutations, can severely compromise the skin's structural and functional integrity (1).

1.1.1. Skin Composition

To understand the consequences of the genetic variants presented in this work, it is first important to understand the structure of the skin.
The skin is structured into 3 layers, the epidermis, the dermis and the hypodermis, each of which contains different structures and performs different functions (2).

Epidermis

The deepest layer of the epidermis is the stratum basale. It consists of a single layer of cuboidal to columnar stem cells, which are attached to the basement membrane via hemidesmosomes. This layer also contains melanocytes. The next layer, the stratum spinosum, consists of 8-10 layers of polyhedral keratinocytes connected by desmosomes. The stratum granulosum contains 3-5 layers of diamond-shaped cells with keratohyalin and lamellar granules. The cells are attached to each other by glycolipids secreted by the lamellar granules.
The layer above is only present in thick skin such as the palms and soles of the feet. It is called the stratum lucidum a thin translucent layer that contains eleidin, a transformation product of keratohyalin. The final layer of the epidermis is the stratum corneum, which is composed of 20-30 layers of dead keratinocytes and functions as a protective barrier.
Important cells in the epidermis are the keratinocytes, producing keratin and lipids for the barrier function, and aiding in vitamin D activation through UVB absorption.

Melanocytes produce melanin, the pigment that gives the skin its colour. UVB light stimulates the production of melanin, which in turn protects the skin against UV radiation. Melanocytes transfer melanin to keratinocytes via cytocrine secretion. An important part of the human immune defence are Langerhans cells. They are immune dendritic cells in the stratum spinosum and are antigen-presenting cells. Lastly Merkel cells serve as mechanoreceptors for light touch and interact with sensory nerve endings. They are found in the stratum basale with the highest concentration recorded in the fingertips (2,3).

Dermis

Under the basement membrane the superficial dermis, the papillary layer, can be found. It is thinner, composed of loose connective tissue and forms dermal papillae. This layer contains fine collagen and elastin fibers, capillaries and sensory nerve endings.

Below the deep dermis, the reticular layer, is located. It consists of thicker, dense, irregular connective tissue and houses thick collagen bundles and elastin fibers, sweat glands, hair follicles, blood vessels and sensory neurons (2,3).

Hypodermis

The deepest layer of the skin is the hypodermis and the main components of this layer are adipose lobules, blood vessels, sensory neurons and sparse skin appendages (2).

1.1.2. Genodermatoses

Genodermatoses are inherited skin disorders that affect the skin structure and skin function and may also lead to increased morbidity and mortality. The most common genodermatoses are ichthyosis, epidermolysis bullosa and ectodermal dysplasia. Other notable conditions are albinism, cutis laxa, progeroid syndromes and precancerous conditions like xeroderma pigmentosa or dyskeratosis congenita (4).

1.1.3. Ectodermal Dysplasia

Ectodermal dysplasia (ED) is a term used to describe a group of genetic disorders affecting structures that are derived from the embryonic ectoderm, such as skin, sweat glands, hair, teeth and nails. Some affected individuals may also suffer from a cleft lip or cleft palate, immune system issues, as well as hearing and vision problems. There are over 180 different types of ED caused by genetic mutations, which can either be inherited or occur spontaneously (5).
Symptoms vary depending on the type and may also vary within families. Some examples include:

- Abnormal hair growth; thin, sparse hair
- Malformation or aplasia of some or all teeth
- Abnormal nails
- lack of the ability to sweat, causing overheating
- dry eyes, corneal erosions
- severe skin erosions, pigmentary abnormalities

The severity of these symptoms can range from mild to severe (6,7).

ED conditions may be grouped according to the molecular pathway affected by the genetic mutation. Three important pathways can be differentiated.
The EDA/NFKappaB pathway is associated with anhidrotic/hypohidrotic ectodermal dysplasia and mutations in this pathway occur in genes such as the X-linked EDA gene and autosomal genes like EDAR, EDARADD and TRAF6. EDA is expressed in interfollicular cells, while its receptor EDAR is expressed in follicular cells.
EDA encodes ectodysplasin, a type II transmembrane protein. The protein binds to the ectodysplasin A receptor, which is encoded by EDAR and EDARADD. This ectodysplasin-EDAR-EDARADD complex interacts with TRAF6 and ultimately results in NF-κB signalling for skin appendage development. Mutations in this pathway lead to defective formation of hair follicles, sweat glands and teeth (8,9).
Another pathway is the WNT pathway, which primarily results in dental abnormalities. Important genes affected in this pathway are WNT10A and PORCN (9).

The third pathway is the TP63 pathway. Its most important gene is TP63, which regulates gene activity by binding to specific DNA sequences. Individuals affected by this condition typically exhibit a combination of ectodermal dysplasia manifestations, including hypohidrosis, nail dysplasia, sparse hair, and tooth abnormalities. Additional features may include cleft lip/palate, split-hand/foot malformation/syndactyly, lacrimal duct obstruction, hypopigmentation, hypoplastic breasts and/or nipples, and hypospadias (9,10).

So far there exists no cure for ectodermal dysplasia. Treatment options depend on the specific type of ED and the affected ectodermal structures. Patients with anhidrosis/hypohidrosis have problems with their thermoregulation, therefore a cool environment and adequate hydration are important. Dental defects may require orthodontic treatment, including dental implants. Topical agents may help patients with xerosis or eczematous dermatitis. Patients who suffer from severe alopecia can try the use of topical minoxidil, which may improve hair growth (11,12).

1.1.4. Variegate Porphyria

Variegate porphyria is a rare genetic metabolic disorder caused by deficient protoporphyrinogen oxidase (PPOX) enzyme activity and falls under a group of disorders known as porphyrias. The disease develops due to variants in the PPOX gene and leads to the accumulation of porphyrins and porphyrin precursors, causing cutaneous and /or neurological symptoms.

Variegate porphyria has an autosomal dominant mode of inheritance, but incomplete penetrance, meaning not everyone who carries the variant will develop symptoms. Different variants in the gene PPOX have been identified in different families (13).

The clinical penetrance is low, and thus symptoms only manifest in about 35-40% of patients (14,15). The most common chronic symptom is skin photosensitivity. Sun-exposed skin is fragile and frequently exhibits symptoms such as blistering, milia, scarring, discolouration, and increased hair growth.

Systemic symptoms manifest as acute attacks and occur without cutaneous symptoms. During an acute attack common symptoms include gastrointestinal

manifestations, such as severe abdominal pain, vomiting and nausea, diarrhoea and constipation. Further neurological symptoms include insomnia, restlessness, muscle and limb weakness, low blood sodium levels, seizures and go as far as muscle paralysis and hallucinations. Other symptoms are reddish, dark urine and increased heart rate and blood pressure (13,16).

Variegate porphyria is caused by pathogenic variants in the PPOX gene, which encodes the enzyme protoporphyrinogen oxidase. This enzyme is crucial for heme production, a molecule that contains iron and is present in haemoglobin and other proteins. The defect primarily affects the heme synthesis in the liver and therefore important hemoproteins that are synthesized there. Protoporphyrinogen and coproporphyrinogen built up in the liver and become oxidized to protoporphyrin and coproporphyrin, which then spread to other tissues and cause photosensitivity and neurological symptoms.

Attacks are triggered by factors that increase liver heme synthesis, such as certain medications, hormones like progesterone, fasting or low carbohydrate intake, alcohol and stress or infections (13).

The management of variegate porphyria depends on the particular symptoms exhibited by the patient, as there is no causal therapy so far. During an acute attack, hospitalization is often required for pain control, treatment of nausea and vomiting, as well as specific therapies depending on the symptoms. To prevent attacks, patients are advised to avoid triggers such as unsafe drugs, fasting, unsupervised diets and sun exposure. In cases where hormonal triggers are the issue, GnRH analogues can be prescribed. Cutaneous symptoms are mainly managed by reducing sun exposure through the use of protective clothing, opaque sunscreens and window tinting for homes and cars (17).

1.1.5. Ichthyosis

Ichthyosis is a group of skin disorders characterized by dry, scaly, thickened skin that may appear rough, red and itchy. The severity can vary from mild to severe, with some causes even affecting internal organs. There are over 30 different types of ichthyoses. Most of these types are inherited and are caused by genetic mutations. They are present at birth or develop in childhood. Less common is acquired ichthyosis, which is triggered by medical conditions or medications (18).

While most patients exhibit symptoms such as dry, scaly skin and pruritus, the clinical presentation can vary significantly depending on the specific type and severity of the disorder. There are eight major inherited forms.

- Ichthyosis vulgaris is the most common type and develops in early childhood, with patients appearing normal at birth. Patients present with fine white scales, sparse skin folds and exaggerated palmar creases. It is often associated with atopic dermatitis and allergies.
- X-linked recessive ichthyosis develops more commonly in males and typically manifests at about 3 to 6 months. Scaling usually presents as polygonal, greyish-brown scales and is primarily present at the neck, lower face, trunk and legs.
- Keratinopathic (epidermolytic) ichthyosis has its onset at birth with erythroderma. Newborns are born with fragile skin and blisters. The fragile epidermis leads to a high risk of secondary infections. As the disease progresses generalized thickened, hyperkeratotic scales develop and, especially in joint areas, the skin is ridged.
- Autosomal recessive congenital ichthyosis, such as lamellar ichthyosis and congenital ichthyosiform erythroderma (CIE), is present at birth. They can present as collodion baby, where a tight shiny membrane encases the newborn. After shedding this membrane large, dark, plate-like scales appear in lamellar ichthyosis and fine scaling with redness in CIE. Another possibility is the self-healing collodion baby. In such cases, the skin normalizes or only displays mild residual scaling.
- The most severe form is the Harlequin ichthyosis. At birth, patients present with thick, amor-like plates, which impair vital functions like breathing and can lead to ectropion (eyelid eversion), eclabium (lips pulled outward) and constrictive bands, which can cause digit amputation. Life-threatening complications include respiratory distress, feeding difficulties and infections. The mortality rate in neonates is as high as 50%.
- Other forms are erythrokeratodermia variabilis where skin can develop rough, thick or reddened areas, and progressive symmetric erythrokeratoderma, which is characterized by dry, red, scaly skin.

The cutaneous manifestations of acquired ichthyosis are similar to those of ichthyosis vulgaris, but only manifest in adults (18,19).
The focus of this thesis is autosomal recessive congenital ichthyosis (ARCI); consequently, the following paragraph will be focused on the genetic variants associated with such forms of ichthyosis.
The gene most commonly affected is TGM1, with 24%-34% of all ARCI patients showing variants in this gene. Patients with lamellar ichthyosis are especially affected. The TMG1 gene encodes an enzyme called transglutaminase 1, that is found in the epidermis. It is involved in the formation of the cornified cell envelope, a structure formed on the inner surface of the cell membrane. Transglutaminase 1 gives strength and stability to the epidermis by cross-linking proteins of the cornified cell envelope (20,21).
With a prevalence of 9%-21% the second most common gene in ARCI patients is ALOX12B. Over 75% of patient with a variant in this gene are born with a collodion membrane and later suffer from mild to moderate CIE. The protein encoded by ALOX12B is vital in the synthesis of corneocytes lipid envelope, which establishes a barrier and protects the skin from water loss (21,22).
Other genetic causes of ARCI include the genes ABCA12 (3%-12%), ALOXE3 (3%-10%), CYP4F22 (3%-10%) and NIPAL4 (5%-9%) (20).

There is currently no cure for inherited forms of ichthyosis and treatment focuses on the management of the symptoms. Severe forms, like collodion baby and harlequin ichthyosis, are life-threatening emergencies that require neonatal intensive care.
The general aim of treatment is to reduce the thickness of the scales, the redness of the skin and the itching. Patients should use non-soap cleansers and regularly apply moisturising creams containing keratolytics, such as urea or salicylic acid. Long baths can soften the scales, and oral or topical retinoids can reduce the scaling (18,19).

1.2. Mode of Inheritance

1.2.1. Autosomal dominant

Autosomal dominant inherited disorders are expressed in heterozygotes, meaning the affected individuals typically carry one mutant allele and one normal allele (Figure 1). The disease locus is located on an autosome, rather than on the X or Y chromosome. The condition can therefore affect both male and female individuals.

When one parent is affected, each child has a 50% chance of inheriting the disease. In rare cases, if both parents are affected, there is a possibility that a homozygous child may be born. In such cases the phenotype is usually more severe than that observed in heterozygous individuals (23).

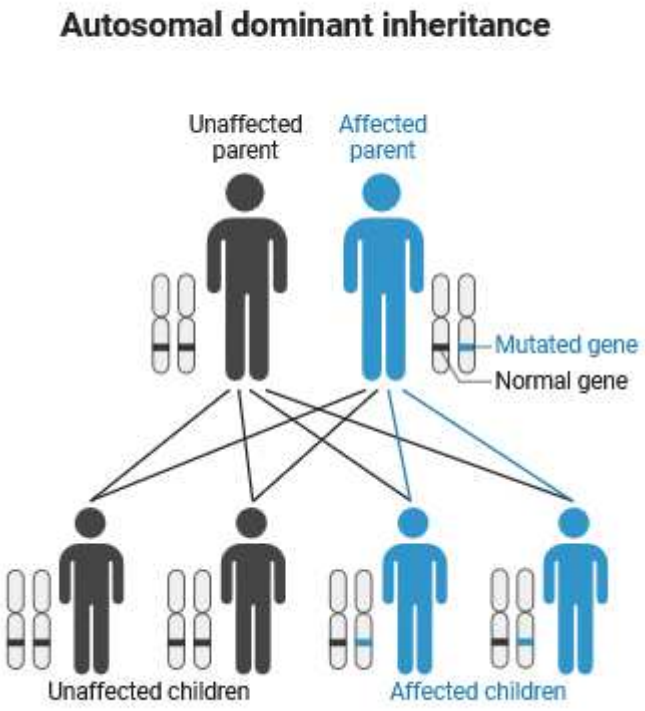

Figure 1: Autosomal dominant inheritance pattern.

created in https://BioRender.com

1.2.2. Autosomal recessive

An individual with an autosomal recessive disorder can also be of either sex and is typically born to two unaffected parents, who each carry a heterozygous variant (Figure 2). The affected individual inherits two mutant alleles, one from each parent, causing the disease. With two unaffected heterozygous parents, there is a 25% chance of each child being affected and a 50% chance of each child being an unaffected carrier of the mutant allele.

Every individual carries heterozygous mutant alleles at multiple loci associated with recessive phenotypes. For more common autosomal recessive disorders, carriers are frequent and affected children may inherit two different mutant alleles (compound heterozygotes) from parents who each carry different mutations. In the context of rare diseases, affected individuals often possess two identical mutant alleles due to their parents being close relatives (23).

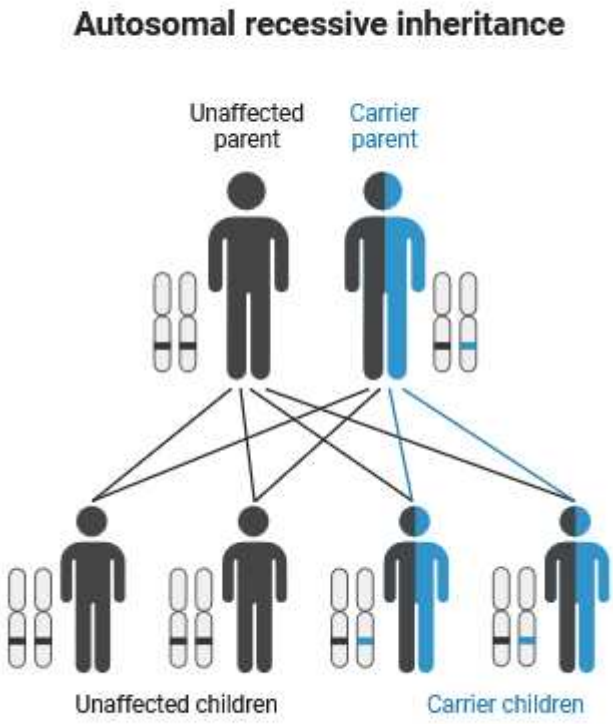

Figure 2: Autosomal recessive inheritance pattern.

created in https://BioRender.com

1.3. Consanguinity

Generally, a consanguineous union is defined as a marriage between second cousins or closer. For closely related individuals a coefficient of inbreeding can be calculated. In consanguineous marriages, this coefficient should be $F \geq 0.0156$ (24).

While marriages between closely related family members who share 50% of their DNA such as parents and children or siblings are legally prohibited and socially frowned upon, cousin marriages can be more frequent in some parts of the world, like in some communities from the Middle East and parts of the Indian subcontinent. Cousins share approximately $1/8^{th}$ of their genes, and their union can result in a high degree of homozygosity, with the increased chance of their children being affected by a recessive disorder (23).

1.4. The Families

1.4.1. SK11

The SK11 pedigree includes five generations, with only the latest generation displaying symptoms. Consanguinity occurs in the fourth generation, twice between first-degree cousins and once between second-degree cousins.

Seven affected individuals were identified in the fifth generation, among them were five men and two women. All of them have non-affected siblings.

Affected individuals are present in the fifth generation and no affected individuals are found in the first four generations, since the first consanguineous unions appeared in the fourth generation. Both male and female patients are affected, excluding a Y-chromosomal inheritance mode. X-chromosomal recessive inheritance can be ruled out as there are female patients, whose fathers are both unaffected. The presented pedigree highly suggests the mode of inheritance as autosomal recessive. In this mode of inheritance homozygous mutation in homozygous regions or compound heterozygotes variants can be expected.

Blood samples for Whole Exome Sequencing were taken from all affected individuals (SK11-1, SK11-2, SK11-3, SK11-4, SK11-5, SK11-6). However, the

Institute of Human Genetics at the Medical University of Graz only received five samples, which belonged to SK11-2, SK11-3, SK11-4, SK11-5 and SK11-6.
The pedigree of family SK11 (Figure 12) can be found under results in chapter 3.1.
In the physical examinations the patients exhibited symptoms consistent with ectodermal dysplasia of the hair and teeth type.
The teeth of the affected individuals are noticeably different. The patients have missing and widely spaced teeth. Their teeth are also smaller than average, and the enamel is thinner, which leads to increased rates of tooth decay.
The hair of these individuals is fine and sparse. The pictures (Figure 3,4) also suggest a fragile hair type as well as unruly hair.

Figure 3: Phenotype SK11
sparse and fine hair

Figure 4: Phenotype SK11
missing, widely spaced, pointed teeth

1.4.2. SK12

The SK12 family pedigree comprises four known generations. The first affected individual appears in the second generation, while both the third and fourth generations have multiple affected individuals, with four patients in the third and three in the fourth. In the third and fourth generation the ratio of affected to non-affected individuals can be assumed equal, as can the distribution between sexes. Consanguinity has also been observed in the third generation between first degree cousins. However, given the presence of the phenotype in the second and third generations, it appears that consanguinity is not the primary factor contributing to the development of the described disease.

Since female members of the family are also affected, it can be assumed that the mode of inheritance is not Y-chromosomal. For the same reason, as well as the presence of unaffected fathers and brothers, an X-chromosomal recessive transfer will also be excluded. While there is a possibility of an autosomal recessive transfer, it seems unlikely for there to be individuals in three consecutive generations. The pedigree indicates that the mode of inheritance of the phenotype is likely to be autosomal dominant.
The blood samples were taken from the affected individuals in the fourth generation (SK12-1, SK12-2, SK12-3) and their unaffected brother (SK12-5). The mother (SK12-4) of those individuals was also affected and provided a blood sample. Unfortunately, her affected brother and two affected sisters could not be recruited, nor could their mother, who is the affected person in the second generation. The pedigree of family SK11 can be found under results in chapter 3.2.

During the physical examination the following observations were made:

- fragile skin
- minor scratches leading to severe wound formations and bleeding
- a skin rash
- photosensitivity

Prof. Khan excluded ichthyosis as the cause of the symptom.
The photographs (Figure 5) showing the hands of the affected individuals show hyper- and hypopigmentation and scaring. In the picture an erosion on the back of the hand can be seen. The face is also affected by the condition. Skin lesions center around the lower part of the face especially around the mouth and cheeks.

Figure 5: Phenotype SK12
hyper- and hypopigmentation, erosion and scaring

1.4.3. SK13

In family SK13 the pedigree extends over of five generations with affected individuals being present in the fourth and fifth generations. A consanguineous union is present in the third generation between first degree cousins, resulting in one affected male individual (SK13-2) out of six children. However, another affected individual (SK13-3) can be found in the fourth generation who is not directly related to him. The affected individual in the fifth generation SK13-1 is related to both individuals (SK13-2 and SK13-3) and is their second cousin once removed.

All affected individuals are male, but a Y-chromosomal transfer can be excluded since none of their fathers are affected and SK13-2 and SK13-3 each have non-affected brothers. X-chromosomal transfer cannot fully be ruled out since there are no female affected individuals. An autosomal dominant mode of inheritance can be excluded in this case, since the phenotype is not present in every generation. The most likely mode of inheritance seems to be an autosomal recessive transfer.
Blood samples were taken from the three affected individuals, as well as from four non-affected individuals (SK13-4, SK13-5, SK13-6, SK13-7). The pedigree of family SK13 can be found under results in chapter 3.3.

As can be seen in the following pictures (Figure 6), the affected individuals show symptoms of ichthyosis with dry flaky skin. The scalp is particularly affected, exhibiting large plate-like scales, as well as alopecia.

Figure 6: Phenotype SK13
dry flaky skin, large plate-like scales on the scalp

1.5. Computational tools and approaches for disease gene identification

1.5.1. Nanoquant

The Nanoquant is a photometric procedure that can be used to evaluate the concentration of DNA samples.

Prior to using DNA samples in analytical techniques, it is necessary to ensure the quality and usability of these samples, through parameters such as DNA purity

and integrity. The ratio of absorbance at 260 nm and 280 nm is utilised to evaluate the purity of the DNA sample. At a ratio of about 1.8 the DNA is judged as pure. A lower ratio may suggest the presence of proteins, phenol, or other contaminants (25).

1.5.2. Whole exome sequencing

Whole Exome Sequencing (WES) is a method used to sequence all protein-coding regions, the so-called exons, of the genome. These make up about 1% of the human genome but account for about 85% of known disease-causing mutations (26).

The typical workflow of a WES analysis includes the following steps: quality control of the raw data, preprocessing, sequence alignment, post-alignment processing, variant calling, variant annotation, and variant filtration and prioritization. WES generates thousands of variant candidates and knowledge of the phenotype and family pedigree is used to add various filters to the data and identify disease causing variants (27).

1.5.3. Franklin by Genoox

Franklin by Genoox integrates data from publicly available population, disease and sequence databases and published literature.

It is a genomic analysis platform designed to help interpret genetic data efficiently. Franklin is particularly useful for diagnosing genetic disorders, cancer genomics, and personalized medicine. Advantages include the user-friendly interface, comprehensive variant databases, and collaboration features (28).

Franklin provides a wide range of filtering options, enabling users to analyse genetic variants through parameters like inheritance patterns, variant types, genomic regions, and clinical significance criteria. Hyperlinks to additional resources such as UCSC, gnomAD and OMIM are also displayed in the tool.

1.5.4. Homozygosity Mapper

Homozygosity mapping is a genetic technique used to identify disease-causing recessive mutations by detecting regions of the genome that are homozygous in affected individuals. Recessive mutations are more likely to be found in homozygous regions due to shared ancestry. SNP Microarrays are a common tool for homozygosity mapping, due to their cost-effectiveness and ability to genotype hundreds of thousands of single-nucleotide polymorphisms (SNPs) across the genome. WES is increasingly used alongside or instead of SNP arrays, especially for novel gene discovery. Overlapping runs of homozygosity (ROH) in affected individuals highlight possible disease-causing locations. Genes within these homozygous regions are therefore analysed for pathogenic mutations.

This approach only works for a recessive mode of inheritance, particularly in consanguineous families, as the disease-causing mutation likely comes from a common ancestor.

With homozygosity mapping, not the whole genome, but only a few candidate regions need to be searched for pathogenic variants. Homozygosity mapping has the potential to yield valuable evidence showing the pathogenicity of previously unidentified mutations, including those occurring in deep intronic or missense variants of uncertain significance (29).

1.5.5. Sanger Sequencing

Sanger sequencing is regarded as the most precise technique for detecting single nucleotide variants and small insertions or deletions. It is widely used in screening for known familial variants and for confirming results from next generation sequencing.

The method uses patient DNA as a template in a PCR reaction including normal bases and chain-terminating bases. Thereby DNA fragments of different lengths are generated and then separated by their size through capillary electrophoresis. Chain-terminating bases are fluorescently labelled and based on this data a chromatograph is generated representing the DNA sequence (30).

1.6. Aims

The objective of this thesis is to identify the underlying genetic variants associated with the skin diseases affecting the three Pakistani families. In this process, advanced data filtering techniques employed for whole-exome sequencing (WES) data will be systematically evaluated through the utilisation of homozygosity mapping and software-based variant calling.

This work aims to identify novel pathogenic variants or further annotate previously characterised disease genes with new potentially disease-causing variants. These findings will contribute to the broader understanding of rare, genetically heterogeneous diseases.

In any event, the spectrum of known mutations and/or gene loci will be expanded, thereby enhancing current genetic knowledge and accelerating the future molecular diagnosis and classification of genodermatoses.

2. Methods

2.1. Recruitment, family history

The patient cohort was recruited by Prof. Muzammil A. Khan at Gomal University, Pakistan. All patients originated from consanguineous families in rural regions of Pakistan. They were informed about blood sample collection, clinical examination procedures, and the use of photographs taken of affected individuals. Written informed consent was obtained from all participants. All these procedures were managed and carried out by Prof. Khan and his associates. Ethical approval for the genetic study was obtained from the Ethical Review Board of Gomal University (32/DQA/GU 24. January 2023). Blood samples were sent to the Institute of Human Genetics at the Medical University of Graz for further analysis.

2.2. DNA extraction and measurement

Table 1: Working solutions for DNA extraction

Buffer name	**Buffer contents**
TKM 1	10 mM Tris-HCl pH=8,8 10 mM KCl 10 mM MgCl2 x 6H2O 2 mM EDTA Triton X-100 (=1%) dest. H2O
TKM 2	0,4 M NaCl 10 mM Tris-HCl pH=8,8 10 mM KCl 10 mM MgCl2 x 6H2O 2 mM EDTA dest. H2O
10 mg/mL Proteinase K	100 mg Proteinase K in 8 mL dest. H2O lösen mit dest. H2O auf 10 mL auffüllen

10%iges SDS	5 g SDS in 40 mL dest. H2O lösen mit dest. H2O auf 50 mL auffüllen

Genomic DNA was extracted from whole blood samples using phenol–chloroform extraction. 1 mL EDTA-anticoagulated blood was lysed in 5 mL of TKM1 buffer (Table 1). The sample was centrifuged at 2000 g for 10 min at room temperature, and the supernatant discarded. The cell pellet was washed in 5 mL TKM1 buffer until the pellet appeared white and the supernatant clear (~7 times washed). The remaining supernatant was carefully removed by pipette.
Each pellet was resuspended in 160 µL TKM2 buffer (Table 1), 20 µL 10 mg/mL Proteinase K (Table 1), and 20 µL 10 % SDS (Table 1). A mock control tube containing 100 µL TKM1 buffer was processed in parallel. All tubes were then incubated at 48°C for 18h at 950 rpm to ensure lysis and protein digestion.

After the incubation, 500 µL Phenol:Chloroform:Isoamylalcohol (25:24:1) was added to the cell lysate. The mixture was vortexed for 20 seconds and centrifuged for 5 minutes at 13,000 rpm at room temperature. 500 µL of Chloroform was added to the supernatant, again vortexed for 20 seconds and centrifuged for 5 minutes at 13,000 rpm at room temperature. The supernatant was transferred into a 15 mL centrifuge tube, and the volume was estimated (~800 µL). The 6-fold volume of 100% EtOH (-20°C), ~4,8 mL, was added. The DNA was precipitated by gently pivoting the tube. Since no DNA was visible at this point, the samples were transferred into four tubes and incubated overnight at -20°C.

Following precipitation, samples were centrifuged for 10 min at 13,000 rpm at room temperature, and the supernatant was discarded. The DNA pellet was washed with 500 µL of 70% ethanol and centrifuged for 1 min at 13,000 rpm. This washing step was repeated once with 70% ethanol and once with 100% ethanol. The resulting DNA pellet was incubated at room temperature for 50 min. The DNA pellet was dissolved in 50 µL of 10 mM TE buffer pH=8,8 at 37°C for 10 minutes at 950 rpm, resulting in 200µL DNA solution per sample.
DNA concentration and purity were determined spectrophotometrically at 260/280 nm using a NanoQuant instrument (Tecan, Switzerland).

2.3. Whole Exome Sequencing

The most promising samples with the highest DNA concentration and purity, as determined by NanoQuant measurements, were selected for whole-exome sequencing (WES). For family SK12, both samples SK12-3 and SK12-4 were included. WES was performed using Twist Library Preparation Kit (Twist Bioscience) on a Nova Seq 6000 (Illumina) at the Institute of Human Genetics of the Medical University of Graz.

WES data analysis

For WES data analysis, we performed quality control of the raw data, preprocessing, sequence alignment, post-alignment processing, variant calling, annotation, and subsequent variant filtration and prioritization.

Raw sequencing data were exported in FASTQ format. Data preprocessing included quality control, removal of technical contaminants, filtering of low-quality reads, and elimination of redundant sequences. Sequence alignment to the reference genome was then performed, thereby determining the genomic location of each exome fragment.

The new format in the second step is referred to as a BAM file. In the post-alignment processing low-quality sequence fragments are identified and filtered out. Variant calling is performed through specialized variant calling software, with single nucleotide polymorphisms (SNPs), insertion-deletion mutations (Indels), and other genomic variations being identified. It is important to distinguish between somatic and germline variants. Following the variant calling process, the file format is now a VCF file.

The third step is the variant prioritization.

For the families in this study WES was already performed in Pakistan by Prof. Khan and VCF files for the sample SK11-1, SK12-2, SK13-2 and SK13-3 were sent. These files were used for further analysis together with the new VCF files for SK12-3 and SK12-4 that were created in Graz.

2.4. Data filtering process

2.4.1. Franklin by Genoox

The data from all three families was uploaded to Franklin in the form of a VCF file. The system then analysed variants, filtered them based on clinical relevance, and provided annotations with supporting literature. Automated classifications were reviewed, and further filtering steps were taken. Based on the inheritance patterns observed in the family pedigrees heterozygous or homozygous variants were excluded. Further filtering steps were taken to detect possible variants. The exact filtering steps are presented in chapter 4.1, 4.2 and 4.3. Franklin's ability to customize filters while maintaining broad variant coverage proved useful in identifying potential pathogenic mutations without overlooking less obvious candidates.

2.4.2. Homozygosity Mapper

In this thesis the tool Homozygosity Mapper was used for families SK11 and SK13 as they showed an autosomal recessive mode of inheritance. The same VCF files as in Franklin were used, SK11-1 for family SK11, and SK13-2 and SK13-3 for family SK13.

As there was only one sample homozygosity mapping was more difficult for family SK11, and the primary focus was therefore put on family SK13.

The results were exported into an Excel file and further analysed using the web-based tools Galaxyuse (Version 23.2) and Genevenn (Pirooznia, Nagarajan, & Deng, 2007).

2.5. Primer Design

For the Primer design the online tool primer3 (Version 2.0.1) was used. The gene in which the suspected variant was located was found using the UCSC genome browser. The sequence was then entered into the primer3 tool, and the appropriate product size range was selected. A forward and reverse primer was generated. The primers were verified using the UCSC tool in silico PCR. The predicted sequence was then checked for common snips in the UCSC genome

browser. No common snips were found in all the primers and therefore they were ordered. Primers used in this study are displayed in Table 2.

Table 2: Primers for generating specific DNA fragments

Primer name	**Sequence (5'-3')**
SK11 primers	
EDAR_E8f	GCTCTGGGTCATTCATGCCT
EDAR_E8r	CCCCACGGTAAGCACAGTAT
SK12 primers	
PPOX_E13f	ACCCCCAGCTAAAACATTCCT
PPOX_E13r	CTTTGCTGCCCCGCATTCTT
SK13 primers	
TGM1_E3f	GCTAATGATAAGGGCCTGGG
TGM1_E3r	GACCCTAATGCCAGCCTCTC

2.6. PCR-Optimization and PCR-Reactions for Sanger-Control-Sequencing

Table 3: PCR reaction master mix

18,550 µL	NFW
0,125 µL	100 pmol/µL forward Primer
0,125 µL	100 pmol/µL reverse-Primer
0,100 µL	25 mM je Base dNTP-Mix
5,000 µL	5xPuffer
0,100 µL	5 U/µL Go Taq Polymerase

Following whole-exome sequencing, PCR reactions were optimised to evaluate the designed primers and ensure the quality of subsequent Sanger sequencing. Genomic DNA samples were thawed and vortexed prior to use. A master mix was prepared as outlined in Table 3, 24 µL were dispensed into PCR 8-strip tubes and supplemented with 1 µL of DNA (100 ng/µL). Negative controls were prepared by adding 1 µL nuclease-free water, while positive controls received 1

µL of a reference DNA sample. All preparation steps were carried out on ice to prevent premature primer–DNA interactions.

PCR amplification was performed in a thermal cycler using a gradient program to determine the optimal annealing temperature. Amplification products were analysed by agarose gel electrophoresis (1,5%). For each reaction, 3 µL of PCR product was mixed with 3 µL of TBE loading buffer and loaded onto the gel. Electrophoresis was conducted at 100 V for 20 minutes. DNA fragments were visualised under UV light using a photo documentation system. The agarose gel electrophoresis process separated DNA molecules in the samples, so the purity of the DNA fragments could be visualized (31).

Based on the electrophoretic profiles, the optimal annealing temperature for each primer pair was determined (Table 4). PCR amplification was subsequently repeated using customised cycling conditions reflecting the optimal annealing temperatures.

Table 4: Programme for standard PCR
for x the temperature was selected according to the annealing temperature of each primer (EDAR: 62,8° C; PPOX: 62,8°C; TGM1: 60°C)

Programme:	**96°C**	**96°C**	**x°C**	**72°C**	**72°C**
	5 min	10 sec +2°C/sec R.Temp	10 sec	10 sec +2°C/sec R.Temp	5 min
		35 cycles			

2.7. Sanger Sequencing

Prior to Sanger sequencing, PCR products were purified using the ExoSAP-IT PCR Product Cleanup Kit (Thermo Fisher Scientific, USA), according to the manufacturer's instructions. 1 µL ExoSAP-IT reagent was added to 2,5 µL PCR product, incubated at 37 °C for 15 min to degrade primers and nucleotides, and then incubated at 80 °C for 15 min for enzyme inactivation.

The PCR product was now ready for the subsequent steps (32).

For sequencing, 1 µL of either forward or reverse primer (diluted 1:10) was added to the sequencing master mix (Table 5). Subsequently, 1 µL of purified PCR

product was added, and cycle sequencing was performed in a thermal cycler using the program in Table 6.

Table 5: Master mix for sequencing reaction

BigDye Terminator v1.1 Cycle Sequencing Kit	1,0 µL
5x sequencing buffer	2,0 µL
Millipore H_2O	5,0 µL

Table 6: Thermal cycler programme for the sequencing reaction

Programme:	**96°**	**50°**	**60°**
	20sec	20sec	4 min
	35x		

Sequencing reaction products were purified using Sephadex G-50 Superfine (Thermo Fisher Scientific). 395 µL Sephadex was added to each well of 0,2 mL MicroAmp Optical 8-tube strips, centrifuged at 800 g for 1 min, and the excess water was removed. This step was repeated, and the gel-containing tubes were then transferred to new 0,2 mL MicroAmp Optical 8-tube strips. Next, 10 µL of each sequencing product, diluted with 10 µL nuclease-free water, was loaded onto the Sephadex columns and centrifuged at 800 g for 2 min. The purified sequencing products were collected in the new MicroAmp tubes.
Capillary electrophoresis was performed on a SeqStudio Genetic Analyzer (Thermo Fisher Scientific). Electropherograms were analysed using Sequencing Analysis Software v7.0 (Thermo Fisher Scientific).

3. Results

This chapter presents the results of homozygosity mapping and the Franklin analysis. These two analyses were conducted on three different families. For family SK11 and SK13, both analyses were performed and for family SK12, only the Franklin analysis was performed. The same set of procedures was applied in all three cases. This included whole exome sequencing, data analysis and data filtering, followed by primer design and confirmation by Sanger sequencing.

3.1. SK11

The family SK11 showed an autosomal recessive mode of inheritance and the generation in which the variants were first observed highly suggested consanguinity to be the deciding factor in the appearance of the mutations.
The five samples received by the Institute of Human Genetics at the Medical University of Graz (SK11-2, SK11-3, SK11-4, SK11-5 and SK11-6) first underwent DNA extraction and measurement. Unfortunately, all five samples were highly contaminated and thus could not be used for WES and the subsequent software-based analysis. WES had already been performed on the missing sample SK11-1 in Pakistan and the Institute of Human Genetics received a VCF file. This file was used for the next steps in the analysis. Using only one patient's DNA complicated the search.
HomozygosityMapper was first employed in the search for homozygous regions. Probably due to low quality of the data the search did not work as intended.
Consequently, the tool Franklin by Genoox was used for filtering the variants.
Without any filters activated, a total of 36,066 variants in 11,713 genes were found. As the pedigree suggests an autosomal recessive inheritance mode, all heterozygous variants were excluded in a first filtering step, and the mode of inheritance was selected as autosomal recessive. Following this step, 3,033 variants in 1,300 genes remained. Next synonymous variants were excluded, further narrowing down the findings to 1,998 variants in 1,069 genes. The Franklin Classification filter was used and the filters 'pathogenic', 'likely pathogenic' und 'uncertain' ('leaning pathogenic', 'uncertain significance') were used. Only 9 variants in 9 genes remained. Two of these variants were classified as pathogenic, while the remainder were all classified as variants of uncertain

significance. The pathogenic variants EDAR:c.719_722delAAGA and CD36:c.1240_1243dupCTTT were further examined using the phenotype of the affected individuals. This indicated a high likelihood of EDAR:c.719_722delAAGA, with the associated condition of ectodermal dysplasia, being the underlying cause for the skin disease affecting the SK11 family.

According to the ACMG guidelines, this variant was classified as pathogenic. The effect on the protein is very strong. A change to the same amino acid results in a known pathogenic variant, and the variant lies in a gene where loss of function is a known mechanism of disease.

To further evaluate the possible variant gene, the UCSC genome browser was used. The reference genome used by the UCSC genome browser was GRCh38/hg38. The identified variant is NM_022336.4:c.719_722delCTCTT>C (GRCh38/hg38). The variant is located on chromosome 2q.13 in exon 8 out of 12 on the EDAR gene (Figure 7). dbSNP reports common variants in the region. Those variants are synonymous and thus do not affect the resulting protein. In this variant after the present cytosine on base position 108,910,783 the next four bases in the positions 108,910,784 to 108,910,787 were deleted. The normal order of nucleotide bases is illustrated in figure 8. In this case, AAGA was deleted. This resulted in a frameshift, affecting glutamic acid and lysine (p.Lys240fs) and therefore the whole protein.

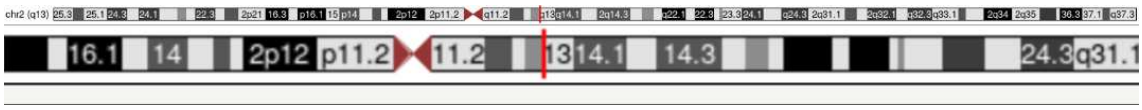

Figure 7: Location of the EDAR gene on chr2q13
modified illustration from UCSC

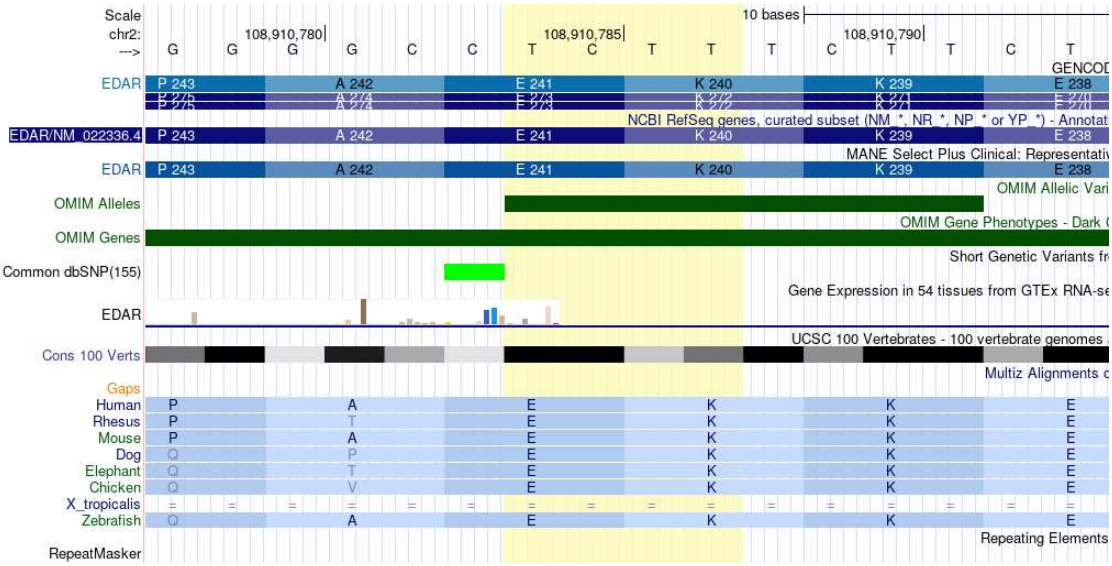

Figure 8: Screenshot from UCSC showing the EDAR gene the highlighted section shows the deleted bases in NM_022336.4:c.719_722delCTCTT>C(p.Lys240fs)

Primers for EDAR were ordered, to confirm the variant through Sanger sequencing. The primers were assessed in a gradient PCR assay, with the annealing temperature set from 58° C to 70° C. The PCR fragments were then analysed by agarose gel electrophoresis. Based on the results an annealing temperature of 62,8°C was selected.

Another PCR assay was performed using the SK11 samples (SK11-2, SK11-3, SK11-4, SK11-5, SK11-6), the positive control and a blank as a negative control. Agarose gel electrophoresis was used again to determine the accuracy of the PCR (Figure 9).

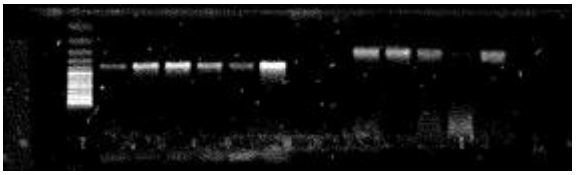

Figure 9: Agarose gel electrophoresis of SK11 samples with EDAR primer on the left-hand side

Sanger segregation analysis was then performed using all the samples mentioned above and a positive control. In this family only samples from affected

family members were available, therefore the positive control was used as the wildtype and helped in comparing the electropherograms.

While the results align with the pedigree of family SK11, the bad quality of the DNA can also be seen in the electropherograms. In SK11-2 the reverse strand could not be read and in SK11-5 the forward strands quality was not satisfactory. Unfortunately, there was also a problem with the forward strand of the wildtype, and it broke off after 250 bases. The deletion was illustrated in samples SK11-2, SK11-3, SK11-4 and SK11-6 without the comparison to a wildtype, as seen in figure 10. The reverse strand of samples SK11-3, SK11-4, SK11-5 and SK11-6 and in comparison, the wildtype is shown below (Figure 11). In each strand, 4 bases were deleted.

The expected deletion affected bases AAGA and segregation could clearly be shown in all samples in the forward and reverse strand. A comparison to the pedigree (Figure 12) shows a clear correlation between the mutation and the affected individuals in the family.

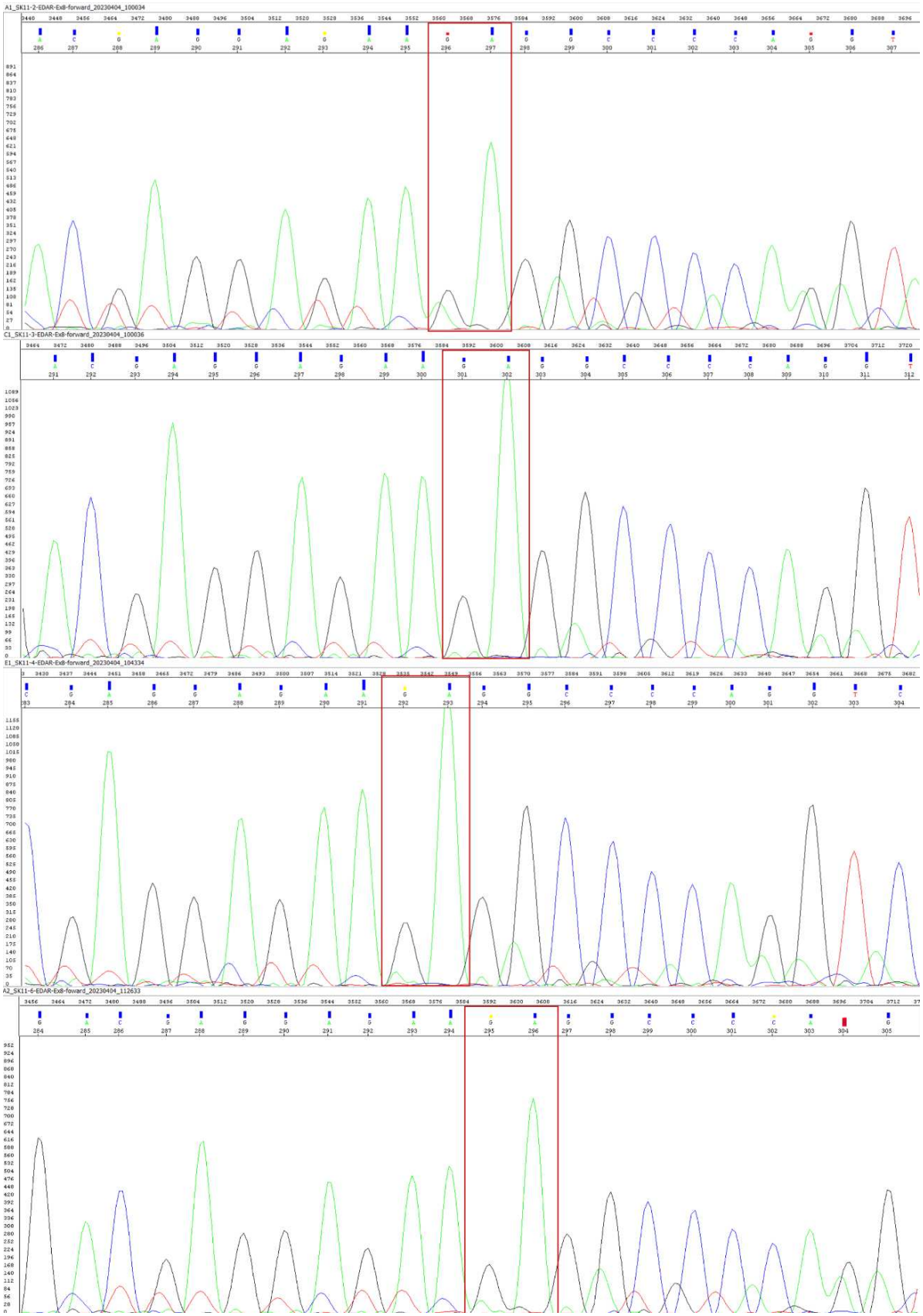

Figure 10: Electropherogram of SK11-2, SK11-3, SK11-4 and SK11-6; forward strand red highlighted section shows the deletion of four bases

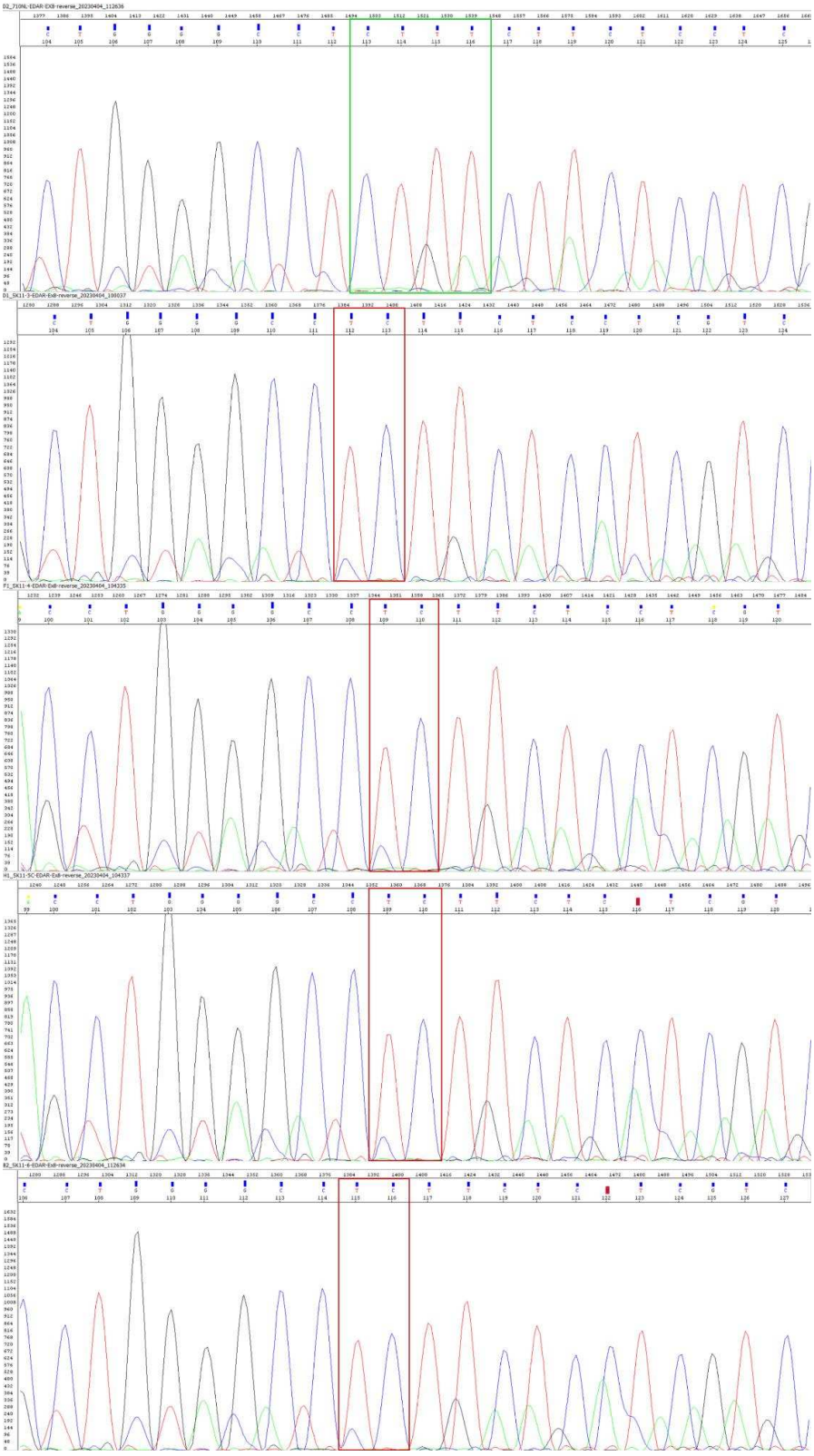

Figure 11: Electropherogram of SK11-3, SK11-4, SK11-5C, SK11-6 and the wildtype; reverse strand
red highlighted section shows the deletion of four bases and green highlighted section shows the wildtype

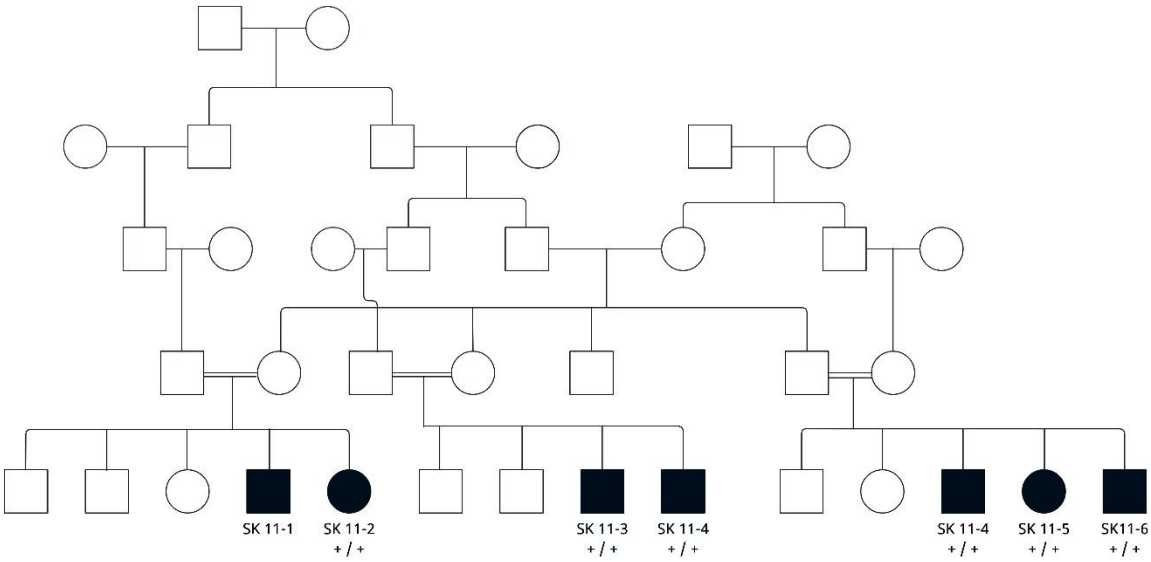

Figure 12: Pedigree SK11
Squares symbolize male individuals and circles symbolize female individuals. Filled symbols represent people with the pathological phenotype; empty symbols represent people that do not share the phenotype. The autosomal recessive inheritance pattern for the variant *EDAR:c.719_722delAAGA* is shown using the symbols + / + or + / -, which indicate the presence of the variant on both (*homozygous) or one (heterozygous)* of the two alleles, respectively. *A horizontal double line represents consanguinity.*

3.2. SK12

The apparent mode of inheritance in family SK12 was autosomal dominant, and consanguinity did not seem to be the cause of the mutation. The Institute of Human Genetics at the Medical University of Graz received 11 samples from 4 patients. One patient's sample (SK12-1) was missing. The samples were used for DNA extraction and measurement, resulting in 5 samples of good quality, that could be used for further analysis. Fortunately, at least 1 sample for each patient was of good quality. For patient SK12-2 a VCF file was already obtained from WES in Pakistan. Therefore, WES was performed using sample SK12-3 and sample SK12-4(3), as these samples showed the most promising results in the DNA extraction and measurement process.

HomozygosityMapper could not be used in this instance as there seemed to be an autosomal dominant mode of inheritance, suggesting heterozygosity.

In the Franklin analysis the VCF file (SK12-2) from Pakistan was used, as well as the two new VCF files (SK12-3, SK12-4) from Graz. First an individual analysis

was conducted for each file. Due to the autosomal dominant mode of inheritance, the filtering steps differed from those used for SK11.

Sample SK12-2 showed 80,116 variants in 15,318 genes without the application of any filters. In the first filtering step, an autosomal dominant mode of inheritance and heterozygosity were selected, reducing the results to 7,244 variants in 1,444 genes. Synonymous variants were excluded again, leaving 6,108 variants in 1,391 genes. The same filter for the Franklin Classification was used, including 'pathogenic', 'likely pathogenic' and 'uncertain' ('leaning pathogenic', 'uncertain significance') variants. Now only 152 variants in 81 genes were found.

Unfiltered 42,181 variants in 12,767 genes were found for sample SK12-3. Gene properties and zygosity were filtered in the same way, leaving 3,283 variants in 1,069 genes. Synonymous variants were excluded, and the Franklin Classification was applied in the same way as for sample SK12-2. The remaining variants numbered 110 and were located in 52 genes.

The same steps were taken for sample SK12-4, narrowing down 43,319 variants in 12,897 genes to just 99 variants in 53 genes.

The 152 variants of sample SK12-2, 110 variants of sample SK12-3 and 99 variants of sample SK12-4 were then compared to find variants present in all three affected individuals. This search resulted in the discovery of the variant PPOX:c.1387G>A, and numerous different HLA-DRB1 variants. However, these HLA variants were excluded, as they were judged to not be the underlying cause of the skin disease affecting family SK12.

The variant PPOX:c.1387G>A was promising as the associated condition was variegate porphyria, which coincided with the described phenotype of family SK12, namely fragile skin and atopic dermatitis.

The ACMG guidelines classified this variant as a variant of uncertain significance. In-silico predictions support a deleterious effect on the gene for a missense or a splicing region variant. Functional data also indicates that this gene has a low rate of benign missense mutations, despite missense mutations being a common mechanism of disease.

UCSC genome browser was used for further analysis. The reference genome used by the UCSC genome browser was GRCh38/hg38, and the identified variant is NM_001122764.3:c.1387G>A. The variant is located in chromosome 1q23.3 in the gene PPOX in exon 13 out of 13 (Figure 13,14). At position 1387 of the

cDNA a G>A substitution occurred, resulting in the amino acid glycine being switched to arginine and thus affecting the resulting protein. dbSNP reports no common variants in the region. At the mutation site position 161,171,127, nucleotide base conservation is indicated as high, as shown by the dark value (Figure 14). This confirms that any change in the nucleotide base pair has far reaching consequences for the resulting protein and, consequently, for the affected individual.

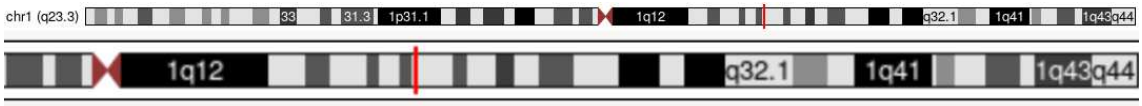

Figure 13: Location of the PPOX gene on chr1q23.3
modified illustration from UCSC

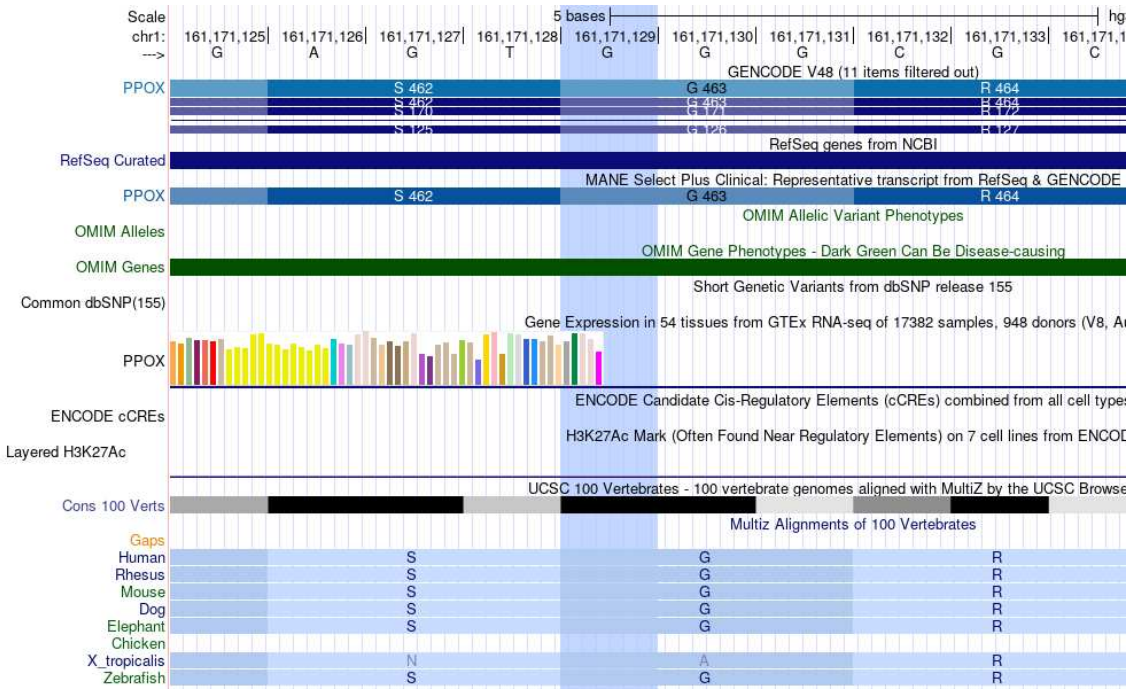

Figure 14: Screenshot from UCSC showing the PPOX gene
the highlighted section shows the base being substituted in
NM_001122764.3:c.1387G>A(p.Gly463Arg)

The associated primers were ordered for further sequencing and confirmation of the variant. First a gradient PCR assay was performed using the primers. The annealing temperature was set between 58°C and 70°C. Based on the results of

the agarose gel electrophoreses, the annealing temperature was set at 62,8°C. Subsequently, a PCR assay was performed with the samples SK12-2(3), SK12-3, SK12-4(3) and SK12-5, as well as a positive control and a blank as a negative control. Agarose gel electrophoresis was employed, controlling that all PCR results for the SK12 samples were positive (Figure 15).

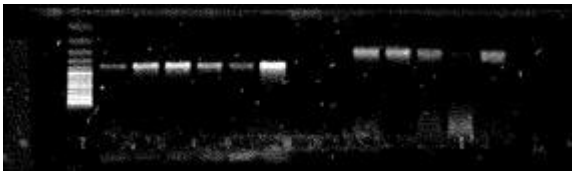

Figure 15: Agarose gel electrophoresis of SK12 samples with PPOX primer on the right-hand side

The next step was to conduct Sanger sequencing, using the aforementioned samples. The reverse strand for SK12-5, the only sample from an unaffected individual was contaminated and could not be used. Thus, the wildtype was once again used for comparison. Figure 16 shows the normal base sequence in the wildtype and the heterozygous C>T mutation in the affected individual's samples SK12-2, SK12-3 and SK12-4. As is evident when compared to the pedigree of family SK12 (Figure 17), the affected individuals in the family show segregation for this variant. Functional studies are needed to further confirm the pathogenicity of this variant.

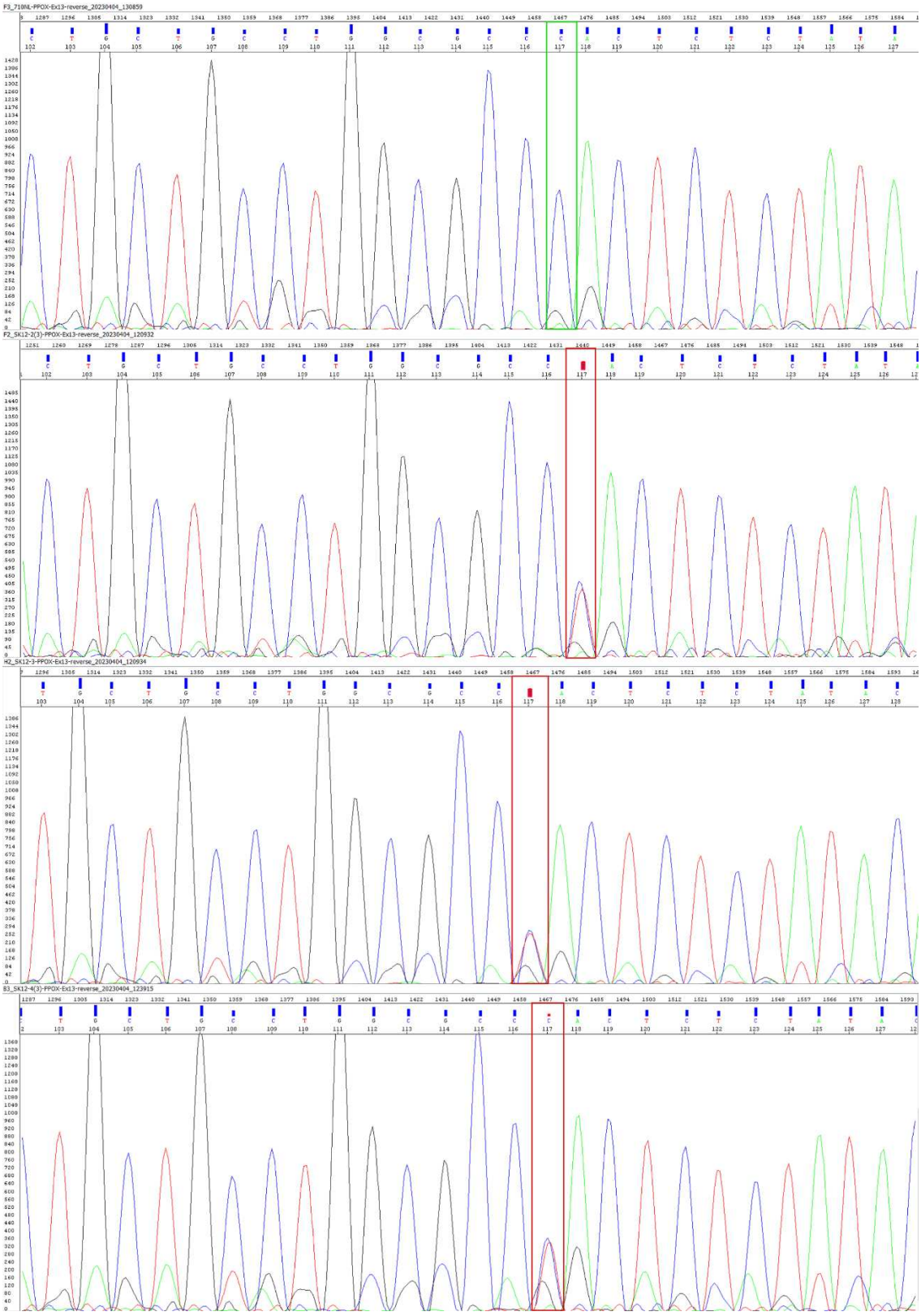

Figure 14: Electropherogram of SK12-2(3), SK12-3 and SK12-4(3) and the wildtype; reverse strand
red highlighted section shows heterozygote genotype of affected individuals and green highlighted section shows the wildtype

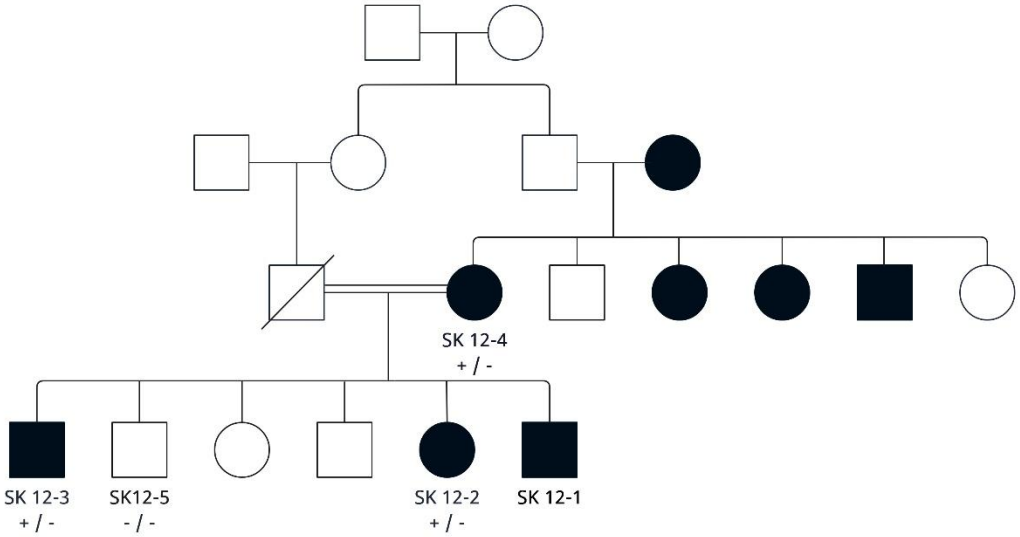

Figure 15: Pedigree SK12
Squares symbolize male individuals and circles symbolize female individuals. Filled symbols represent people with the pathological phenotype; empty symbols represent people that do not share the phenotype. The autosomal dominant inheritance pattern for the variant PPOX:c.1387G>A is shown using the symbol + / - indicating the presence of the heterozygous variant or - / - indicating its absence. A horizontal double line represents consanguinity. A crossed-out symbol represents a deceased individual.

3.3. SK13

The pedigree of family SK13 suggested an autosomal recessive transfer. Consanguinity could neither be confirmed, nor excluded as the cause of the mutation. 9 samples from 6 patients were received from which 6 samples, one from each patient, were found to be of good quality following DNA extraction and measurement. Like in the other two families, the sample of an affected individual, SK13-1, was missing. WES had already been performed in Pakistan using the samples SK13-2 and SK13-3. Due to the good quality of the results, no further WES was performed.

HomozygosityMapper was used in a first analysis of the data. The VCF files for SK13-2 and SK13-3 were each uploaded and the genotype was analysed. Next the results were exported into Excel and formatted, before being converted into bed files. Using the online tool ‘Galaxyuse’ the bed files of both samples were then searched for intersecting intervals. The overlapping ROHs are illustrated in table 7.

Table 7: Overlapping ROHs of SK13-2 and SK13-3

chr	from (bp)	to (bp)	mb
chr14	24473478	25076702	0, 60
chr16	88808918	90141477	1, 33

The results were again exported into Excel, after which variants in these regions were selected for formatting. Lastly the tool ‘GeneVenn’ was used to illustrate the results in a Venn diagram (Figure 18).

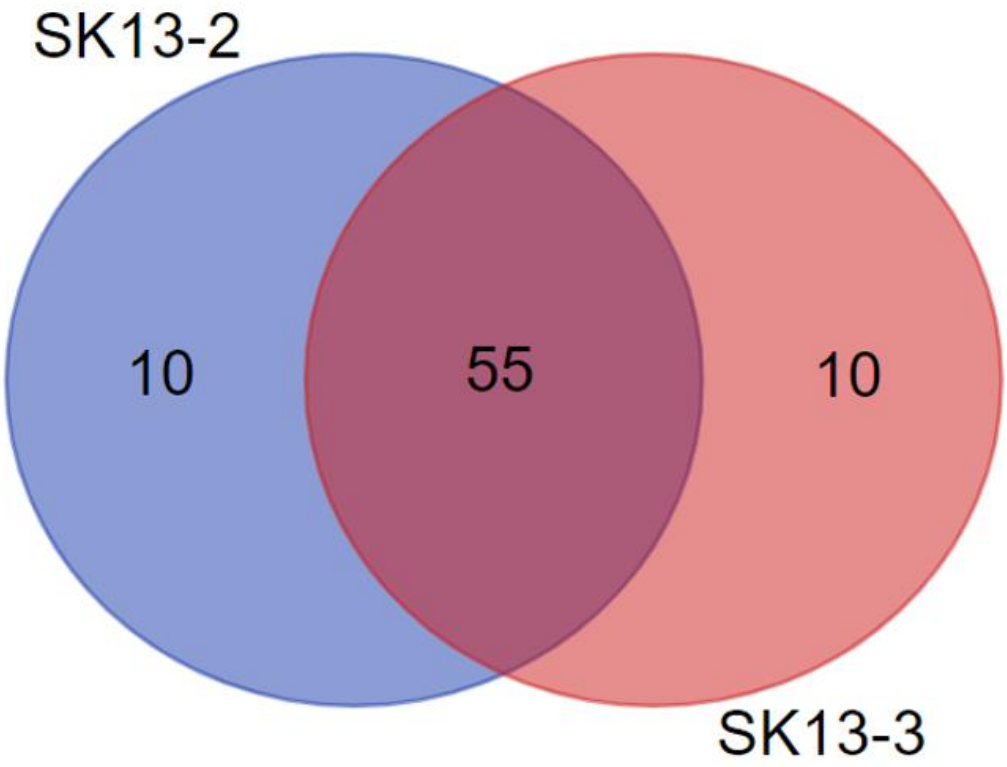

Figure 16: Venn diagram of variants from samples SK13-2 and SK13-3

The variants that were not shared between SK13-2 and SK13-3 could be excluded from the search, which left 55 variants that had to be manually searched and compared with the reported phenotype affecting family SK13. Following analysis of the variants, a missense variant in the TMG1 gene (NM_000359.3:c.400T>C) on chromosome 14 was identified as a candidate variant.

In comparison, the Franklin analysis was performed using the same VCF files. SK13-2 showed 75,400 variants in 15,024 genes, without the application of any filters. Homozygosity and autosomal recessive mode of inheritance were applied, and 7,884 variants in 1,926 genes remained. Using the same filters for the Franklin classification (‘pathogenic’; ‘likely pathogenic’; ‘uncertain’ (‘leaning pathogenic’; ‘uncertain significance’)) and excluding a synonymous effect, 11 variants in 11 genes were found. The phenotype was then investigated for the 11 variants.

In this case, the underlying variant TGM1 was displayed first in the list without any filtering options applied.

Without any filters 79,658 variants in 15,166 genes were found in sample SK13-3. The same filtering methods left only 2 variants in 2 genes, which also appeared in sample SK13-2. TRPM1:c.2T>C is a variant of uncertain significance and associated with night blindness. Therefore, it was phenotypically excluded from further analysis. The other variant, TGM1:c.400T>C, was classified as ‘likely pathogenic’ and associated with congenital ichthyosis, so primers were ordered for this variant.

The classification of ‘likely pathogenic’ was suggested using aggregated data from public databases using ACMG Guidelines. The effect on the protein was judged as ‘likely pathogenic’ because a different amino acid change occurred than in a known pathogenic variant. In-silico predictions and reputable source data also support this evaluation.

UCSC genome browser with the reference genome GRCh38/hg38 was used once again. The identified variant was NM_000359.3:c.400T>C in chromosome 14q12 in exon 3 out of 15 of the TGM1 gene (Figure 19).

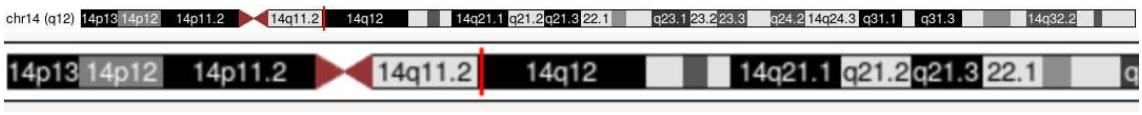

Figure 17: Location of the TMG1 gene on chr14q12
modified illustration from UCSC

At position 400 of the cDNA, a T>C substitution occurred, affecting the third base pair of tyrosine and changing it to histidine (p.Tyr134His). No common variants were reported in this region by dbSNP. Figure 20 shows the darkest value in the track ‘100 Vert. Cons’, indicating a high degree of conservation at the variant site at position 24,731,009.

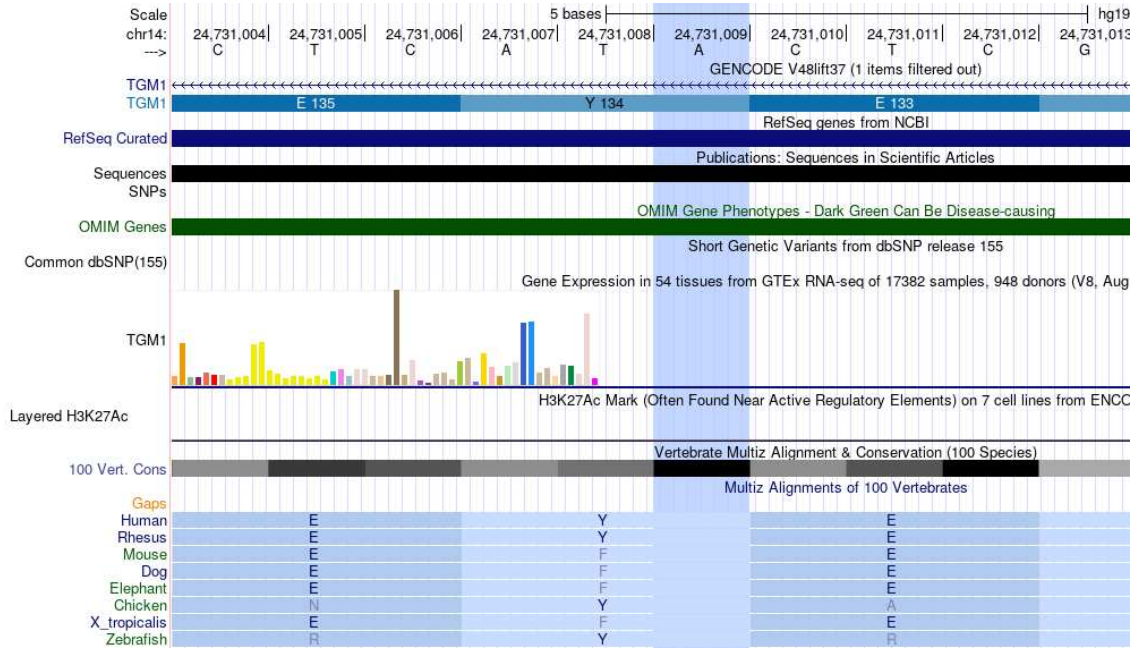

Figure 18: Screenshot from UCSC showing the TGM1 gene the highlighted section shows the base being substituted in NM_000359.3:c.400T>C(p.Tyr134His)

Thereafter primers were ordered for the TGM1 gene. A gradient PCR assay was performed to find the optimal annealing temperature for the primers. The annealing temperature was set from 55°C to 65°C. After agarose gel electrophoreses a temperature of 60°C was chosen for the next step. For this step, all available samples were used, in varying dilutions. The dilution was necessary to ensure the successful PCR amplification, despite the poor quality of the DNA samples. The diluted samples are shown in table 8.

Table 8: Dilutions used in SK13 samples for PCR amplification

Sample name	Dilution
SK13-2	3x (0,5 µl DNA + 1 µl H2O)
SK 13-2B	50x (0,5 µl DNA + 24,5 µl H2O) 100x (0,5 µl DNA + 49,5 µl H2O)
SK13-3	25x (0,5 µl DNA + 12 µl H2O)
SK13-4	10x (0,5 µl DNA + 4,5 µl H2O)
SK13-4B	50x (0,5 µl DNA + 24,5 µl H2O) 100x (0,5 µl DNA + 49,5 µl H2O)
SK13-5	10x (0,5 µl DNA + 4,5 µl H2O)
SK13-6	10x (0,5 µl DNA + 4,5 µl H2O)
SK13-6B	50x (0,5 µl DNA + 24,5 µl H2O) 100x (0,5 µl DNA + 49,5 µl H2O)
SK13-7	100x (0,5 µl DNA + 49,5 µl H2O)

Agarose gel electrophoresis was used to select the best samples for Sanger sequencing (Figure 21). The selected samples are presented in table 9.

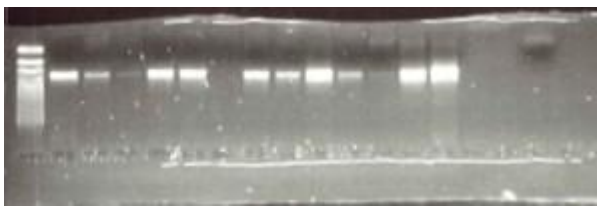

Figure 19: Agarose gel electrophoresis of SK13 samples with TGM1 primer

Table 9: SK13 samples chosen after successful PCR amplification

Sample name	Dilution
SK13-2	3x (0,5 µl DNA + 1 µl H2O)
SK 13-2B	50x (0,5 µl DNA + 24,5 µl H2O)
SK13-3	25x (0,5 µl DNA + 12 µl H2O)
SK13-4	10x (0,5 µl DNA + 4,5 µl H2O)
SK13-4B	50x (0,5 µl DNA + 24,5 µl H2O)
SK13-5	10x (0,5 µl DNA + 4,5 µl H2O)
SK13-6	10x (0,5 µl DNA + 4,5 µl H2O)
SK13-6B	100x (0,5 µl DNA + 49,5 µl H2O)
SK13-7	100x (0,5 µl DNA + 49,5 µl H2O)

The results of Sanger sequencing were presented in the form of electropherograms. In the affected individual's samples (SK13-2 and SK13-3), the substitution of T>C could be shown in the forward strands of the DNA, and the corresponding substitution of A>G could be seen in the reverse strands. Homozygosity was therefore illustrated (Figure 22,23). As indicated by the pedigree, the samples from the unaffected family members exhibited a heterozygote pattern. In comparison, the wildtype shows the normal base at this position. The comparison can be seen in figure 24 and 25.

Comparing the results to the pedigree of family SK13 (Figure 26), the affected patients clearly showed segregation.

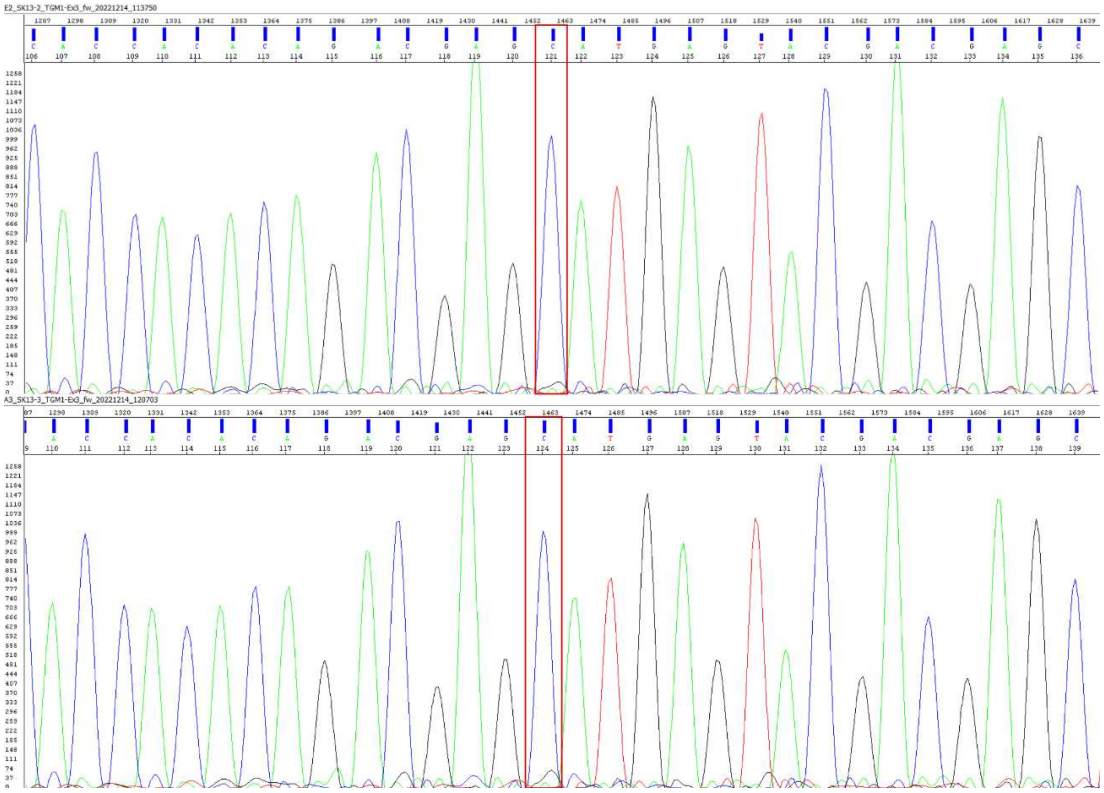

Figure 20: Electropherogram of SK13-2 and *SK13-3; forward strand*
the red highlighted section shows the substitution of T>C

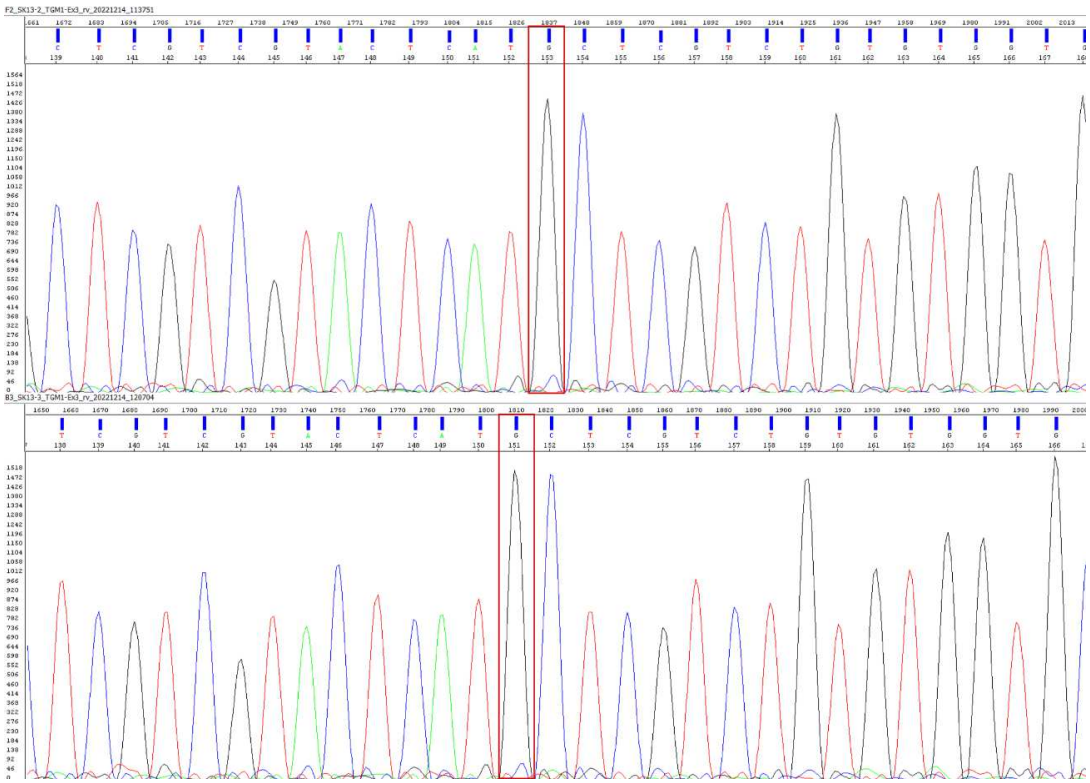

Figure 21: Electropherogram of SK13-2 and *SK13-3; reverse strand*
red highlighted section shows the substitution of A>G

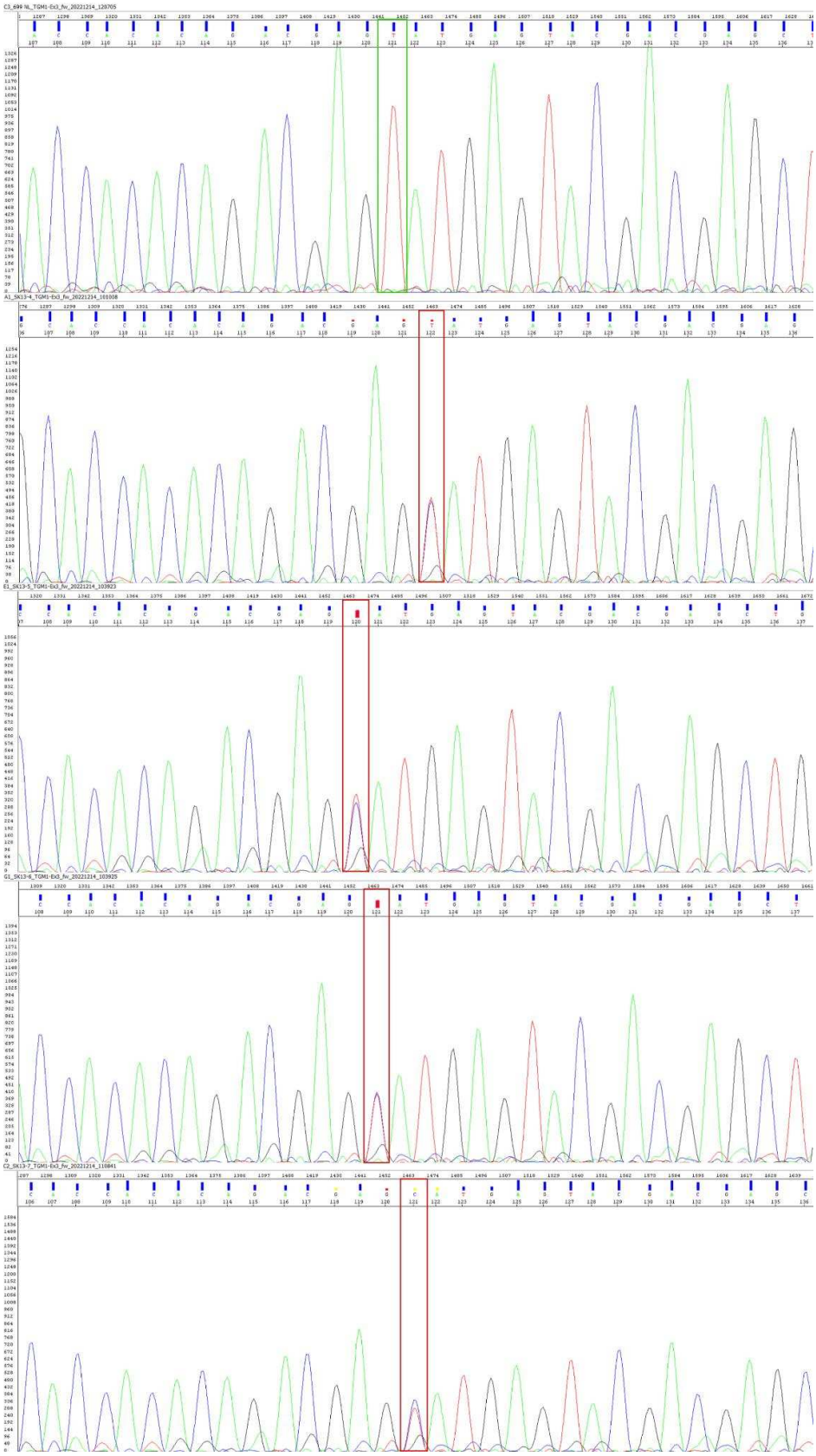

Figure 22: Electropherogram of SK13-4, SK13-5, SK13-6, SK13-7 and the wildtype; forward strand
red highlighted section shows heterozygote genotype of unaffected individuals and green highlighted section shows the wildtype

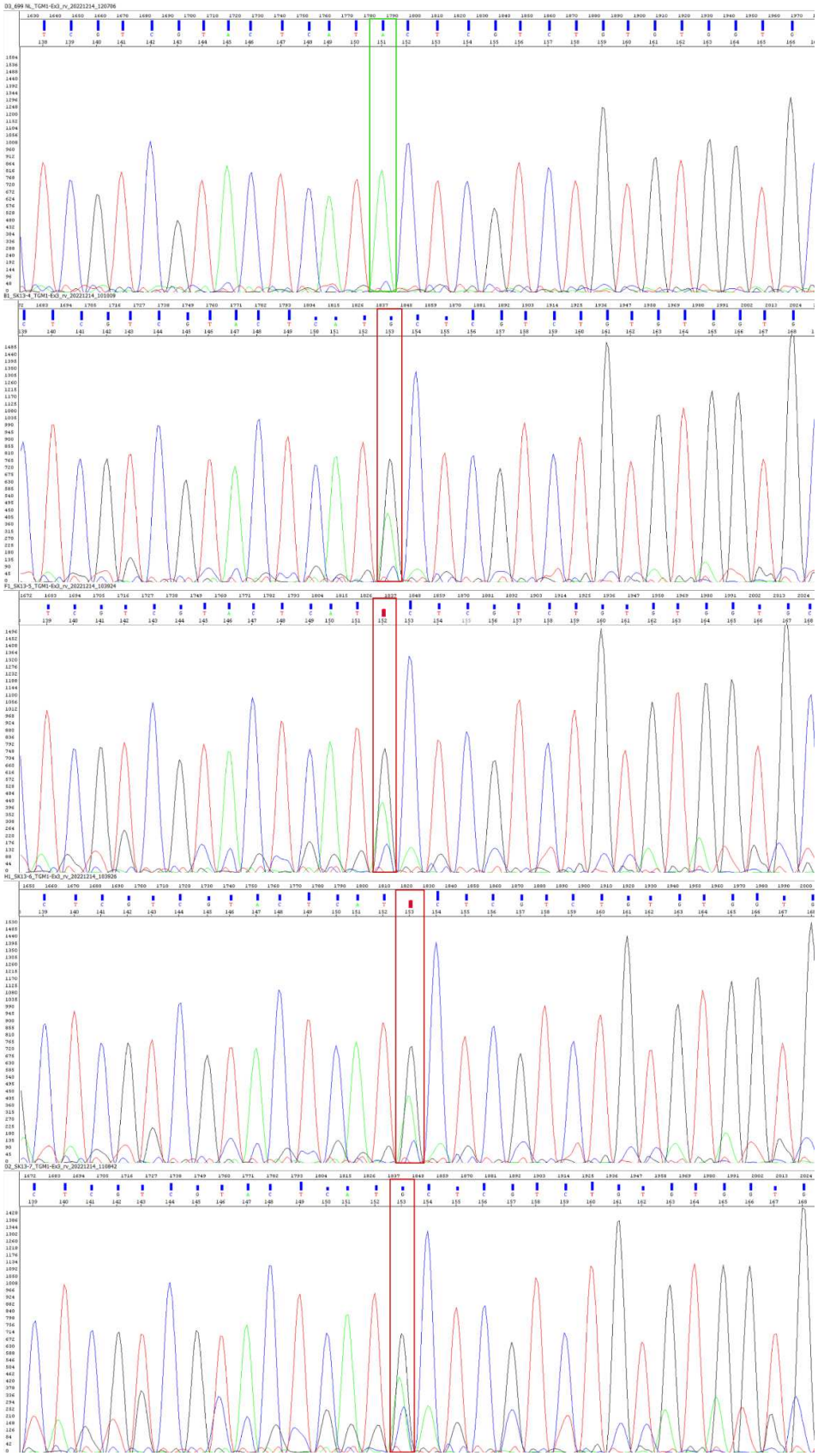

Figure 23: Electropherogram of SK13-4, SK13-5, SK13-6, SK13-7 and the wildtype; reverse strand
red highlighted section shows heterozygote genotype of unaffected individuals and green highlighted section shows the wildtype

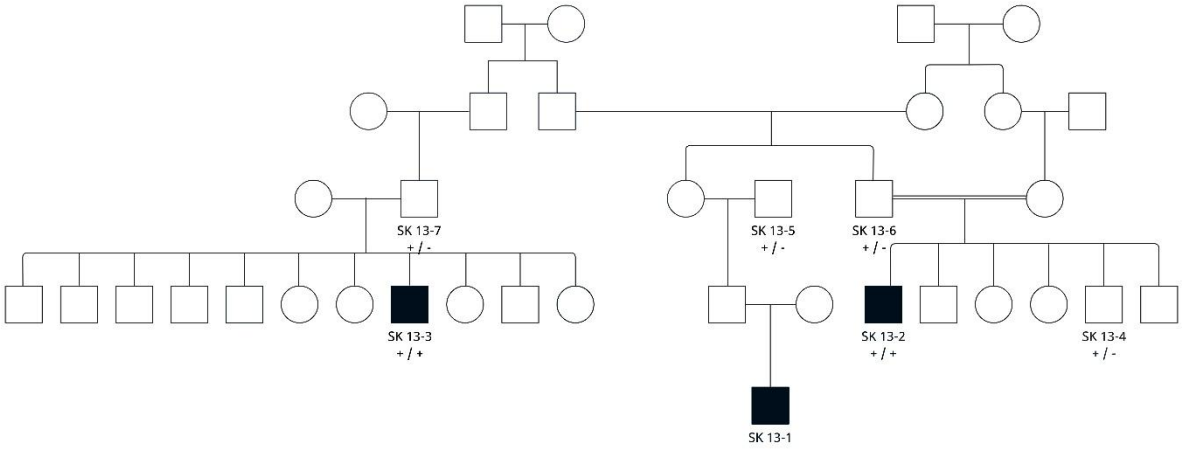

Figure 24: Pedigree SK13
Squares symbolize male individuals and circles symbolize female individuals. Filled symbols represent people with the pathological phenotype; empty symbols represent people that do not share the phenotype. The autosomal recessive inheritance pattern for the variant TGM1:c.400T>C is shown using the symbols + / + or + / -, which indicate the presence of the variant on both (homozygous) or one (heterozygous) of the two alleles, respectively. A horizontal double line represents consanguinity.

4. Discussion

This thesis focuses on data analysis of whole exome sequencing (WES) data using homozygosity mapping and software-based data filtering strategies. The data was obtained by Prof. Khan from Gomal University in Pakistan. The aim of this thesis was to identify genetic variants in consanguineous families from Pakistan, that were all affected by some form of skin disease. The discovery of novel variants is an important step toward improving disease prevention, diagnostics and therapy. For patients, knowledge of their specific genetic diagnosis is also of significant value.

In this study, three families were included, each with multiple members affected by a skin disorder. Genetic variants corresponding to the described phenotype were identified in all three families. Other families had initially been considered for inclusion; however, due to limitations such as time constraints, insufficient data quality, or missing information, they could not be analysed further and were therefore excluded from this study.

The strategy for this thesis initially included homozygosity mapping and Franklin analysis. However, homozygosity mapping ultimately proved to be unfeasible, both because one family exhibited an autosomal dominant mode of inheritance and because the overall data quality was insufficient. In contrast, the use of Franklin analysis was possible and proved to be highly valuable.

4.1. SK11

In family SK11 the variant NM_022336.4:c.719_722delCTCTT>C was identified using the Franklin analysis tool. The deletion is located in the gene EDAR on chromosome 2q13. Through Sanger sequencing electropherograms could be generated to verify the deletion. A deletion could be verified in the affected patients affecting the amino acids glutamic acid and lysine and resulting in a frameshift completely changing the amino acids following the mutation site. As the mutation occurs mid-gene, the impact is likely severe. Possible outcomes for the protein include the induction of a premature stop codon downstream in exon 8 or later, resulting in a truncated, non-functional protein. This protein may also degrade at the mRNA level due to its instability (23).

EDAR stands for ectodysplasin A receptor, and the gene encodes a member of the tumour necrosis factor receptor family. The encoded protein is a receptor for the ligand ectodysplasin A and plays a crucial role in cell death pathways. It is essential for the development of hair, teeth, and other ectodermal products. Mutations in this gene result in autosomal dominant and recessive forms of hypohidrotic ectodermal dysplasia (33).
In this family only samples from affected patients were collected. Due to the absence of a sample from an unaffected individual in this family, while a deletion and frameshift could be shown in the affected individuals, it was not possible to definitively confirm the absence of this frameshift in the DNA of unaffected family members.
The variant identified by Franklin had the rs number (rs797044436) and was listed in gnomAD, The variant p.Lys240fs is classified in ClinVar as pathogenic/likely pathogenic. There is a high probability that the frameshift may result in a premature stop gain and caused by the altered base sequence, the resulting protein would be entirely different from the wildtype.
There are two publications previously describing this variant. The 2005 publication by Naeem, M et al., entitled ‘Novel mutations in the EDAR gene in two Pakistani consanguineous families with autosomal recessive hypohidrotic ectodermal dysplasia’, describes the exact same variant in a Pakistani family. Their study found that this variant results in a frameshift and a premature termination codon 18 bp downstream in the same exon (34).
In the publication by Bashyam, M D et al. “A founder ectodysplasin A receptor (EDAR) mutation results in a high frequency of the autosomal recessive form of hypohidrotic ectodermal dysplasia in India.” the same mutation is described (35).

In this family, there is a high likelihood that consanguinity played a role in the development of the mutation. The pedigree indicates that the occurrence of skin defects in this family coincided with the introduction of cousin marriage. It is possible that a heterozygote mutation is present in members of this family who are not affected. This could result in children from consanguineous unions being born with a homozygous variant and in turn would result in the phenotypic manifestation of the disease.

In this case, the bad quality of the samples limited several research steps. Samples of better quality are necessary, in order to carry out all necessary research steps. A sample from an unaffected individual should also be sequenced so that the genome can be compared to the affected individuals.

4.2. SK12

The candidate variant identified for family SK12 is NM_001122764.3:c.1387G>A. The single base exchange of G to A in this position results in a change of the amino acid from glycine to arginine.
PPOX pathogenic variants typically cause a severe loss of function, with most mutations leading to the complete inactivation or substantial reduction of the enzyme's activity. If there is some residual enzyme activity in affected individuals, it usually stems solely from the remaining wild-type allele. Therefore, different pathogenic variants in PPOX do not lead to variations in disease severity, as all loss-of-function mutations similarly affect enzymatic function (36).

The candidate variant could not be found in gnomAD or any other database and Franklin classified it as a variant of uncertain significance (VUS) /leaning pathogenic. Two other missense variants in the region were also classified as VUS [NM_001122764.3(PPOX):c.1390C>T (p.Arg464Cys) and NM_001122764.3(PPOX):c.1391G>A (p.Arg464His)]. Other missense variants in the gene were classified as pathogenic [NM_001122764.3(PPOX):c.1043A>G (p.Tyr348Cys) and NM_001122764.3(PPOX):c.1046A>C (p.Asp349Ala)]. Both pathogenic variants are strongly associated with variegate porphyria.

For further predictions, the platform mutationtaster.org was used, which predicts the mutation to be disease-causing, with a changed amino acid sequence, affected protein features or changes in the splice site.
The Grantham score, a method used to predict the distance between two amino acids from an evolutionary perspective, was calculated as 125 out of 215. Which indicates a greater evolutionary distance and therefore a high probability of the mutation being damaging.
In this family, consanguinity is not a contributing factor to the occurrence of the disease. The first affected individual appears in the second generation and seems

to not be a child of consanguineous parents. The autosomal dominant mode of inheritance also does not match the typical inheritance pattern in consanguineous families, which is autosomal recessive and thus homozygous. This family shows a clear autosomal dominant pattern, and the consanguineous relationship in the third generation has no effect on the inheritance of the candidate variant.

The received samples were of a higher quality. Even so, the electropherogram for the reverse strand of sample SK12-5 could not be read, despite the good PCR results before sequencing. Reasons for this might be sequencing reaction problems, capillary electrophoresis errors, data artefacts or human as well as software error. Unfortunately, the exact problem could not be identified.

The affected individuals whose DNA samples were available all showed the expected heterozygous variant, and the phenotypes matched the typical phenotypes of variegate porphyria. Though neurological symptoms were not reported, that does not mean an absence of those symptoms, since the clinical assessment in Pakistan was limited.

Although many factors suggest that the variant may be the underlying cause of the disease affecting the SK12 family, further functional studies are required to confirm its pathogenicity.

4.3. SK13

The candidate variant NM_000359.3:c.400T>C is located on chromosome 14q12 in exon 3 out of 15 of the TGM1 gene. A missense mutation with a base exchange from T>C could be proven in this family. At this position, the amino acid tyrosine was substituted by histidine (p.Tyr134His). The rs number for this variant is rs1230140208, and the variant is also listed in gnomAD. According to the gnomAD database, the allele frequency is 0.000003977, making this variant extremely rare.

There are already publications regarding this variant that detail the genotype phenotype correlation. A publication by Youssefian, Leila et al. classifies the variant as pathogenic (37).

TGM1 is the gene that encodes the protein Transglutaminase 1, which is one of three transglutaminases found in the human epidermis.

This protein is a membrane-bound enzyme that functions in the formation of the cornified cell envelope, which is vital for protecting the epidermis against water loss and infectious agents (38).
Mutations in this gene are associated with autosomal recessive congenital ichthyosis (ARCI). The most common cause of ARCI are mutations in the TGM1 gene, but six other genes have also been identified. Due to both genetic and clinical heterogeneity, the genetic forms cannot be clearly identified by the phenotype (39).

While consanguinity was first posited as a possible reason for the occurrence of the mutation, the pedigree showed no evidence for this being the case, excluding any potential unknown relations between the parents of affected individuals. The affected individuals in the fourth generation represent the first documented occurrence of the TGM1 variant as a disease-associated mutation within this family.
Sanger sequencing clearly identified a homozygous variant in the affected individuals and a heterozygous variant in the unaffected individuals in the third and fourth generations. The mutation may be a regional variant, given that all family members are from the same village. This could explain the occurrence of heterozygous mutations in non-related individuals in the third generation.
In this family, both HomozygosityMapper and Franklin by Genoox could be used for the analysis of the whole exome sequencing data.
Homozygosity mapping is a great tool for finding runs of homozygosity (ROHs) and in this case, the possible variants could be narrowed down to 55 using HomozygosityMapper and additional tools such as ‘Galaxyuse’ and ‘GeneVenn’. However, the analysis using Franklin proved superior, as only 2 possible variants for both samples remained after applying the filters described in part 3.3.
Both contained the variant NM_000359.3:c.400T>C, but using Franklin the time for finding the variant could be significantly reduced.

5. Conclusion

In conclusion, this thesis provides new insights into the genetic basis of familial skin disorders and identifies potential molecular causes of disease in three affected families. Two known variants in established disease-causing genes were confirmed in two families, strengthening existing genotype–phenotype correlations. In family SK12 a possible new variant in a known disease-causing gene was found, but further research is needed to establish whether the identified variant is truly linked to the disease or is a non-disease-causing variant. For definitive proof functional studies are the gold standard and should be performed for all variants identified in this thesis.

These findings contribute to a more precise understanding of the underlying mechanisms of hereditary skin disorders and offer a definitive diagnosis for patients with previously undifferentiated symptoms. The identification of causal variants not only enables more accurate distinction between phenotypically similar diseases but also supports clinical counselling by informing patients about risks, potential treatment options, and the implications of consanguinity in the case of recessive disorders.

Furthermore, such work advances research by expanding disease-gene databases and revealing novel pathogenic variants.

From a methodological perspective, this work highlights the central role of WES in the genetic investigation of rare familial disorders. By enabling the simultaneous analysis of a wide range of genes with high accuracy in detecting single-nucleotide variants and small insertions or deletions, WES represents an efficient and cost-effective approach compared with WGS (40).

The study also demonstrates the complementary use of different analytical tools. Franklin proved particularly valuable as an intuitive and versatile platform for variant prioritisation across both homozygous and heterozygous mutations, with broad applicability in diagnostics, carrier screening, and research. In contrast, homozygosity mapping remains especially useful for identifying recessive mutations in consanguineous pedigrees, despite its limitations regarding variant

classification and data quality requirements. Taken together, the combined use of both approaches offers a comprehensive strategy for variant detection and interpretation of hereditary diseases.

Despite these advances, several limitations of this study must be acknowledged. The overall quality of some samples restricted the reliability, and the small cohort size and missing healthy family members limited the statistical power. Furthermore, current analytical tools remain biased towards well-characterised genes and coding variants, leaving non-coding or structural alterations underexplored.
In this thesis, Franklin proved to be the more useful tool, however homozygosity mapping remains important for classifying new variants, particularly in autosomal recessive disorders in consanguineous families. In conclusion, both tools should be used in combination for optimal results.

To strengthen the validity of the findings, further segregation analyses are required to exclude false positives, while functional studies will be essential to establish the pathogenic impact of the identified variants. Future work should also prioritise improved sequencing strategies, higher-quality datasets, and the inclusion of larger family cohorts. Such efforts will not only confirm the results of this thesis but also expand the collective understanding of hereditary skin disorders, contributing to more accurate diagnosis and patient care.

Bibliography

1. DeStefano GM, Christiano AM. The Genetics of Human Skin Disease. Cold Spring Harb Perspect Med. 2014 Oct 1;4(10):a015172–a015172.

2. Yousef H, Alhajj M, Fakoya AO, Sharma S. Anatomy, Skin (Integument), Epidermis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 May 24]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK470464/

3. Communications E. Duale Reihe Dermatologie [Internet]. [cited 2025 May 28]. Available from: https://eref-1thieme-1de-10012a8kt153e.han.medunigraz.at/ebooks/1278819#/ebook_1278819_SL54808686

4. Babu NA, Rajesh E, Krupaa J, Gnananandar G. Genodermatoses. J Pharm Bioallied Sci. 2015 Apr;7(Suppl 1):S203–6.

5. Ectodermal Dysplasia | National Institute of Dental and Craniofacial Research [Internet]. [cited 2024 Jul 3]. Available from: https://www.nidcr.nih.gov/health-info/ectodermal-dysplasia

6. Ectodermal Dysplasias - Symptoms, Causes, Treatment | NORD [Internet]. [cited 2025 May 26]. Available from: https://rarediseases.org/rare-diseases/ectodermal-dysplasias/

7. What is Ectodermal Dysplasia? - The ED Society [Internet]. 2018 [cited 2025 May 26]. Available from: https://edsociety.co.uk/what-is-ed/

8. Cui CY, Schlessinger D. EDA Signaling and Skin Appendage Development. Cell Cycle Georget Tex. 2006 Nov 1;5(21):2477–83.

9. Timothy W, Fete M, Schneider H, Zinser M, Koster MI, Clarke AJ, et al. Ectodermal Dysplasias: Classification and Organization by Phenotype, Genotype and Molecular Pathway. Am J Med Genet A. 2019 Mar;179(3):442–7.

10. Sutton VR, van Bokhoven H. TP63-Related Disorders. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cited 2025 May 26]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK43797/

11. Ectodermal Dysplasia Treatment & Management: Medical Care, Surgical Care, Consultations. 2024 Feb 6 [cited 2025 May 26]; Available from: https://emedicine.medscape.com/article/1110595-treatment#d5

12. Cleveland Clinic [Internet]. [cited 2025 May 26]. Ectodermal Dysplasia: Learning to Live with a Rare Genetic Disorder. Available from: https://my.clevelandclinic.org/health/diseases/ectodermal-dysplasia

13. Variegate Porphyria - Symptoms, Causes, Treatment | NORD [Internet]. [cited 2025 May 27]. Available from: https://rarediseases.org/rare-diseases/variegate-porphyria/

14. Variegate porphyria - an overview | ScienceDirect Topics [Internet]. [cited 2025 May 27]. Available from: https://www.sciencedirect.com/topics/medicine-and-dentistry/variegate-porphyria

15. Baumann K, Kauppinen R. Penetrance and predictive value of genetic screening in acute porphyria. Mol Genet Metab. 2020 May;130(1):87–99.

16. DermNet® [Internet]. 2023 [cited 2025 May 27]. Variegate porphyria. Available from: https://dermnetnz.org/topics/variegate-porphyria

17. American Porphyria Foundation [Internet]. [cited 2025 May 27]. Variegate Porphyria (VP). Available from: https://porphyriafoundation.org/for-patients/types-of-porphyria/vp/

18. Branch NSC and O. National Institute of Arthritis and Musculoskeletal and Skin Diseases. NIAMS; 2012 [cited 2024 Jul 4]. Ichthyosis. Available from: https://www.niams.nih.gov/health-topics/ichthyosis

19. Ichtyoses - Therapeutics in Dermatology [Internet]. [cited 2025 May 27]. Available from: https://www.therapeutique-dermatologique.org/spip.php?article1683#paragraphe-3

20. TGM1 gene: MedlinePlus Genetics [Internet]. [cited 2025 May 27]. Available from: https://medlineplus.gov/genetics/gene/tgm1/

21. Richard G. Autosomal Recessive Congenital Ichthyosis. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cited 2025 May 27]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1420/

22. Zheng Y, Yin H, Boeglin WE, Elias PM, Crumrine D, Beier DR, et al. Lipoxygenases mediate the effect of essential fatty acid in skin barrier formation: a proposed role in releasing omega-hydroxyceramide for construction of the corneocyte lipid envelope. J Biol Chem. 2011 Jul 8;286(27):24046–56.

23. Lucassen A. Genetics and Genomics in Medicine. London: CRC Press; 2022.

24. Bittles A. Consanguinity and its relevance to clinical genetics. Clin Genet. 2001 Aug;60(2):89–98.

25. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R. DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreservation Biobanking. 2016 Aug 1;14(4):264.

26. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci. 2009 Nov 10;106(45):19096.

27. Bioinformatics Workflow of Whole Exome Sequencing - CD Genomics [Internet]. [cited 2025 May 13]. Available from: https://www.cd-genomics.com/bioinformatics-workflow-of-whole-exome-sequencing.html

28. About [Internet]. [cited 2025 May 14]. Available from: https://www.genoox.com/about/

29. Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. HomozygosityMapper—an interactive approach to homozygosity mapping. Nucleic Acids Res. 2009 Jul 1;37(Web Server issue):W593–9.

30. Campen DAF Dr Julia van. GeNotes. [cited 2025 Sep 28]. Sanger sequencing — Knowledge Hub. Available from: https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/sanger-sequencing/

31. How to Interpret Agarose Gel Data: The basics [Internet]. [cited 2024 Jul 6]. Available from: https://www.labxchange.org/library/items/lb:LabXchange:a03c81b4:html:1

32. 78200b.pdf [Internet]. [cited 2024 Jul 6]. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/78200b.pdf

33. EDAR ectodysplasin A receptor - NIH Genetic Testing Registry (GTR) - NCBI [Internet]. [cited 2025 May 22]. Available from: https://www.ncbi.nlm.nih.gov/gtr/genes/10913/

34. Naeem M, Muhammad D, Ahmad W. Novel mutations in the EDAR gene in two Pakistani consanguineous families with autosomal recessive hypohidrotic ectodermal dysplasia. Br J Dermatol. 2005 Jul;153(1):46–50.

35. Bashyam MD, Chaudhary AK, Reddy EC, Reddy V, Acharya V, Nagarajaram HA, et al. A founder ectodysplasin A receptor (EDAR) mutation results in a high frequency of the autosomal recessive form of hypohidrotic ectodermal dysplasia in India. Br J Dermatol. 2012 Apr 1;166(4):819–29.

36. Singal AK, Anderson KE. Variegate Porphyria. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cited 2025 May 22]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK121283/

37. Youssefian L, Vahidnezhad H, Saeidian AH, Touati A, Sotoudeh S, Mahmoudi H, et al. Autosomal recessive congenital ichthyosis: Genomic landscape and phenotypic spectrum in a cohort of 125 consanguineous families. Hum Mutat. 2019;40(3):288–98.

38. Farasat S, Wei MH, Herman M, Liewehr DJ, Steinberg SM, Bale SJ, et al. Novel transglutaminase-1 mutations and genotype-phenotype investigations of 104 patients with autosomal recessive congenital ichthyosis in the USA. J Med Genet. 2008 Nov 3;46(2):103–11.

39. Fischer J. Autosomal Recessive Congenital Ichthyosis. J Invest Dermatol. 2009 Jun 1;129(6):1319–21.

40. Campen DAF Dr Julia van. GeNotes. [cited 2025 May 23]. Whole exome sequencing — Knowledge Hub. Available from: https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/whole-exome-sequencing/

Appendix

The following tool was used to optimize the language of the text:

- Deepl Write, Version 25.8.2
- Deepl GmbH
- 12. October 2025
- https://www.deepl.com/de/write