Diplomarbeit

GENETIC ANALYSIS OF TWO CONSANGUINEOUS PAKISTANI FAMILIES WITH RARE HEREDITARY SKIN DISEASES

eingereicht von

Benjamin Tatrai

zur Erlangung des akademischen Grades

Doktor der gesamten Heilkunde

(Dr. med. univ.)

an der

Medizinischen Universität Graz

ausgeführt am

Diagnostik & Forschungsinstitut für Humangenetik

unter der Anleitung von

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

Dr.med.univ. Ingrid Lafer

Graz, am 20.01.2022

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 20.01.2022 *Benjamin Tatrai eh*

Danksagungen

Zuallererst möchte ich meinem Diplomarbeitsbetreuer, Assoz. Prof. Priv.-Doz. Mag. Dr.rer.nat. Christian Windpassinger, danken, ohne den es diese Diplomarbeit in dieser Form gar nicht geben würde. Er hat es mir ermöglicht, praktische und theoretische Erfahrungen am Institut für Humangenetik zu erlangen und so das Fundament für diese Arbeit gelegt. Er stand mir immer mit Rat zur Seite und half gerne mit seinem Wissen weiter.

Ein großer Dank geht an meine Familie. Nur dank ihrer Unterstützung konnte ich diesen Weg überhaupt einschlagen. Meine Mutter und Schwester haben stets an mich geglaubt und mich ermutigt, mein Bestes zu geben.

Ich möchte auch all meinen Freunden danken, die mir während meines Studiums auch in schwierigen Situationen immer Beistand leisteten.

Des Weiteren möchte ich danken: Prof. Muzammil A. Khan der Gomal Universität in D. I. Khan, Pakistan, für das Rekrutieren der Familien und die Kooperation; Dr. John B. Vincent des CAMH in Toronto, Kanada, für die Zurverfügungstellung der WES-Daten; Beatrice Brugger, MSc MSc, für das Vermitteln von praktischem und theoretischem Wissen, sowie die gründliche Einschulung in diverse Geräte und Programme, welche unentbehrlich für diese Arbeit waren; und Jasmin Blatterer, MSc, für ihre Mithilfe in der Suche nach Kandidatengenen.

Zuletzt möchte ich noch allen weiteren Mitarbeiterinnen und Mitarbeitern des Instituts für Humangenetik danken, die mir immer wieder bei der Überwindung kleiner, alltäglicher Probleme halfen.

Table of Contents

Glossary and Abbreviations

A	adenine
AAV	adeno-associated virus
AD	autosomal dominant (mode of inheritance)
AR	autosomal recessive (mode of inheritance)
bp	base pair
C	cytosine
Cas	CRISPR-associated genes
CNV	Copy Number Variation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	deoxyribonucleic acid
Dr.	doctor
DSB	double strand DNA break
EDTA	ethylenediaminetetraacetic acid
e.g.	for example (Latin: *exempli gratia*)
et al.	and others (Latin: *et alii/et aliae*)
etc.	and other similar things (Latin: *et cetera*)
g	gram
G	guanine
gRNA	guide RNA
i.e.	that is (Latin: *id est*)
indel	a mutation with either insertion or deletion of bases
kb	kilo base pairs (=1000 bp)
kDa	kilodalton
m^2	square metre
mA	milliampere
Mb	mega base pairs (=10^6 bp)
ml	millilitre
MSc	Master of Science
NADPH	the reduced form of nicotinamide adenine dinucleotide phosphate
ng	nanogram
NGS	Next Generation Sequencing
nm	nanometre

ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PPK	palmoplantar keratoderma
Prof.	professor
rcf	relative centrifugal force
RNA	ribonucleic acid
ROS	Reactive Oxygen Species
SH3	Src-Homology 3
SLR	synthetic long reads
SMRT	Single-Molecule Real-Time [sequencing]
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SR	spectrin repeat
T	thymine
TGS	Third Generation Sequencing
Tris	trishydroxymethylaminomethane
USA	United States of America
US-$	United States dollar
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
µl	microlitre
µM	micromolar (=10^{-6} mol/L or 10^{-3} mol/m^3)
°C	degrees Celsius
%	percentage

Table of Figures

Table of Tables

Zusammenfassung

Einleitung: Erbliche Hauterkrankungen können durch Mutationen in einer Vielzahl von Genen entstehen. Es sind sowohl dominante als auch rezessive Erbgänge möglich. Je nach der zugrundeliegenden Mutation können die Hautveränderungen unterschiedlich ausfallen und als Teil eines Syndroms oder isoliert auftreten. Für diese Diplomarbeit wurden zwei unterschiedliche, konsanguine, pakistanische Familien mit erblichen Hauterkrankungen für eine genetische Testung rekrutiert. In Familie 1 zeigten Betroffene Alopezie bzw. woolly hair, trockene Haut und Nagelveränderungen. In Familie 2 bestand das Hauptsymptom aus trockener Haut mit Bildung Abszess-ähnlicher Läsionen bei Sonneneinwirkung.

Material und Methoden: Es wurden bei beiden Familien DNA-Proben von betroffenen und gesunden Familienmitgliedern gewonnen. Es wurde ein Whole Exome Sequencing durchgeführt und basierend auf diesen Daten homozygote Bereiche ermittelt. Mithilfe der aus diesen Verfahren gewonnenen Daten wurden potenziell pathogene Varianten in Kandidatengenen bestimmt und deren Segregation mittels Sanger Sequenzierung überprüft.

Ergebnisse: Für Familie 1 wurde die bekannte Mutation c.1493C>T (p.Pro498Leu) im *DSP*-Gen als ursächliche Mutation für die Alopezie identifiziert. Für Familie 2 wurde eine neuartige Kandidatenvariante im *CARMIL2*-Gen gefunden: c.677A>T (p.Asp226Val). Des Weiteren wurde eine bekannte Art des G6PD-Mangels festgestellt.

Diskussion: Die Mutation c.1493C>T des *DSP*-Gens wurde erstmals durch Jan et al. 2015 beschrieben. Der Phänotyp, der in ihrem Paper beschrieben wurde, stimmt klinisch weitestgehend mit dem Phänotypen der Familie 1 dieser Diplomarbeit überein. Durch die weitere Zuordnung von Fällen konnte das klinische Spektrum dieser Mutation erweitert werden, aber es bedarf noch weiterer Forschungsarbeit, inklusive zellbiologischer Studien, um die pathomechanistischen Zusammenhänge dieser Mutation besser zu verstehen. Die Variante c.677A>T des *CARMIL2*-Gens wurde in der Literatur noch nicht beschrieben. Mit unseren Erkenntnissen kann kein kausaler Zusammenhang mit dem beobachteten Phänotypen hergestellt werden. Weitere Forschungsarbeit ist notwendig, um eine potenzielle Pathogenität dieser Variante zu evaluieren.

Abstract

Introduction: Hereditary skin diseases can be caused by mutations in a multitude of genes, with both dominant and recessive forms of inheritance. Depending on the mutation, the resulting skin changes can vary wildly and be syndromic or non-syndromic. For this diploma-thesis, two different, consanguineous Pakistani families with hereditary skin diseases were recruited for genetic testing. In family 1, affected individuals presented with alopecia/woolly hair, dry skin, and nail changes. In family 2, the main symptom was dry skin with formation of abscess-like lesions upon exposure to sunlight.

Materials and Methods: DNA samples were extracted from affected and healthy individuals in both families. Whole Exome Sequencing was performed, and this data was used to identify homozygous regions. The data acquired through these procedures was used to identify potentially pathogenic variants in candidate genes. Segregation of these variants was verified using Sanger sequencing.

Results: For family 1, the known mutation c.1493C>T (p.Pro498Leu) in the *DSP* gene was identified as a causative mutation for alopecia. For family 2, a novel candidate variant was found in the *CARMIL2* gene: c.677A>T (p.Asp226Val). In addition, a known version of G6PD deficiency was identified in this family.

Discussion: The mutation c.1493C>T of the *DSP* gene was first described by Jan et al. in 2015. The phenotype described in their paper is for the most part identical to the phenotype of family 1 of this thesis. The clinical spectrum of this mutation could be expanded through the correlation of the cases described here, but further research, including cell biological studies, is still required to better comprehend the pathomechanisms of this mutation. The variant c.677A>T of the *CARMIL2* gene has not been described in literature. No causal connection with the observed phenotype can be established with our findings. Further research is necessary to evaluate the potential pathogenicity of this variant.

Statement on Previous Publications

1. Abbas S, Brugger B, Zubair M, Gul S, Blatterer J, Wenninger J, Rehman K, Tatrai B, Khan MA, Windpassinger C. Exome sequencing of a Pakistani family with spastic paraplegia identified an 18 bp deletion in the cytochrome B5 domain of FA2H. Neurol Res. 2021 Feb;43(2):133-140. doi: 10.1080/01616412.2020.1831329. Epub 2020 Nov 27. PMID: 33246395.

1 Introduction

1.1 The Skin

The skin is the largest organ of the human body, measuring between 1,5m^2 and 2m^2, and forming the outer boundary between the individual and their surroundings. For this reason, the skin has to be capable of fulfilling various functions. It acts as a barrier against the surroundings, providing protection from mechanical stress, ultraviolet radiation, pathogens, and harmful substances. Furthermore, it helps regulate body temperature and hydration, serves as a sensory organ and produces vitamin D.[1,2] Finally, the skin also has a non-negligeable social function, as parts of it are visible at almost all times. This explains why diseases of the skin can have an immense impact on perceived quality of life.[3]

The skin is made up of three layers, the epidermis, the dermis and the subcutis, and furthermore contains appendages like hairs, nails, sebaceous glands and sweat glands.[1]

1.1.1 The Epidermis

The epidermis is a keratinized stratified squamous epithelium consisting of four layers: the stratum basale, stratum spinosum, stratum granulosum and stratum corneum. The so-called "thick skin", which can be found on the palms and soles, has an additional layer between stratum corneum and stratum granulosum, called stratum lucidum. The skin at these locations contains no hair follicles or sebaceous glands.[1,2]

Over 90% of all cells in the epidermis are keratinocytes. These keratinocytes are formed in the stratum basale and undergo changes as they progress outwards through the epidermis in a process called terminal epidermal differentiation. Keratin is already present in the keratinocytes of the stratum basale and solely undergoes biochemical changes in the following layers. It forms tonofilaments, which create a network and act as the cytoskeleton of the keratinocytes. These tonofilaments are also anchored to the desmosomes, which form intercellular junctions (see chapter 1.1.6). Other cell-to-cell connections include tight junctions, which act as a tight seal between individual cells, and gap junctions, which enable intercellular transfer of ions and small molecules via channels. The epidermis does not contain any blood vessels of its own but is nourished via diffusion from blood vessels found in the dermis instead.[1,2]

The stratum basale is the bottom-most layer of the epidermis. The keratinocytes in the stratum basale anchor the epidermis to the basement membrane via hemidesmosomes and

form new cells through mitosis. After mitosis, one daughter cell stays in the stratum basale as a stem cell for further division, while the other proceeds on to the suprabasal layers of the skin. Outside the stratum basale, no more mitosis of keratinocytes takes place and the keratinozytes become larger and flatter.[1,2] They are transformed into dead cells, so-called corneocytes, through cornification. This process includes the lysis of organelles and degradation of the nucleus, which leaves the corneocytes filled mostly with keratin. Below the plasma membrane, a cornified cell envelope is formed through cross-linking of keratin and other proteins. The cornified envelopes of multiple corneocytes are interconnected via corneodesmosomes. Lipids found in the intercorneocyte spaces and tight junctions between corneocytes help form the physical barrier of the skin. Eventually, the corneodesmosomes are dissolved and the outmost corneocytes are shed off to the environment in a process called desquamation.[4]

Apart from keratinocytes, the epidermis also contains melanocytes, Langerhans cells and Merkel cells.[1,2]

1.1.2 Dermoepidermal Junction Zone

The dermoepidermal junction corresponds to the basal membrane of the epidermis and has the function of linking the epidermis to the dermis. Its main parts are the lamina lucida and the lamina densa. The cells of the stratum basale of the epidermis connect to the lamina lucida by anchoring filaments attached to hemidesmosomes. The papillary layer of the dermis is connected to the lamina densa by anchoring fibrils. On a slightly larger, microscopical scale there are epidermal rete pegs reaching into the dermis, and dermal papillae that reach into the epidermis. This way the dermis and epidermis form a three-dimensional connective structure, strengthening their connection. Apart from the lamina lucida and the lamina densa, other components of the dermoepidermal junction zone are fibrils, collagen fibres and matrix.[1]

1.1.3 The Dermis

The dermis is the layer of skin beneath the epidermis. It extends to the subcutaneous fat tissue. The most common cells in the dermis are fibroblasts, which produce the dermal tissues and extracellular matrix. Their long appendages are interconnected to form a web. Other cell types found in the dermis include mast cells, macrophages, melanocytes, Langerhans cells and lymphocytes.[1]

The main type of dermal tissue are the collagen fibres, comprised of collagen fibrils. They are mostly made of type I collagen and form a network, which provides the mechanical stability of the dermis. Other collagen fibres found in the dermis are type VII collagen, which constitutes the anchoring fibrils, and type III collagen. Elastic fibres are the second most important dermal tissue and are often connected to the collagen fibres. Their function is providing firmness and elasticity to the dermis. All the previously described cells and fibres are embedded in a matrix comprised of proteoglycans like hyaluronic acid and other components.[1]

Histologically, the dermis comprises two parts: the stratum papillare and the stratum reticulare. The superficial stratum papillare reaches into the spaces between the rete pegs and includes mostly matrix, cells, and capillaries. The broad stratum reticulare consists mainly of collagen fibre bundles and elastic fibre, as well as hair follicles and sweat glands in its lower parts. Below the stratum reticulare lies the subcutis, comprised of fat tissue and connective tissue.[1]

1.1.4 Hair

In humans, just like the skin, hair carries an important social role due to aesthetic reasons. Furthermore, hair follicles play a role in wound repair and sensing touch.[5,6]

There are different types of hair. Lanugo hair is the foetal type of hair, which is replaced by vellus hair after birth. Stronger, visible hair is called terminal hair. Terminal hair covers the scalp and develops in the axillar and genital regions under the influence of hormones during puberty.[1]

Hair develops through growth of epidermis into the dermis during the foetal period. Each hair comprises the hair shaft and the hair follicle. The hair shaft represents the visible part of the hair, outside the epidermis. It consists of a cortex and a cuticle, as well as a medulla in terminal hair. The cortex comprises keratin-filled dead cells and melanin, which determines the pigmentation of the hair. The hair follicle is the part of hair that lies under the skin surface. At the base of the hair follicle lies the hair bulb, which contains hair matrix cells, responsible for cell division and growth of the hair, as well as melanin-producing melanocytes. The hair bulb encases the dermal papilla, which is a dermal structure containing connective tissue, a capillary loop and nerve fibres. The hair follicle outside the hair bulb comprises an inner and an outer root sheath. The inner root sheath produces keratin. The outer root sheath is surrounded by the basal membrane and merges with the epidermis at its end. The outer root sheath contains the attachment point of the

arrector pili muscle and the efferent duct of the sebaceous gland. Close to the attachment of the arrector pili muscle lies the bulge containing stem cells, which are responsible for hair follicle growth and help in wound repair. [5,6]

1.1.5 Alopecia and Genetic Hair Loss Disorders

Alopecia, or hair loss, is a common condition affecting scalp hair and sometimes body hair. Alopecia can have a big psychological impact and immensely hinder the well-being of affected individuals, potentially even leading to psychological conditions like depression. There are many possible causes of alopecia, including but not limited to: androgenetic alopecia, alopecia areata, illnesses, drugs, chemotherapy.[7] Androgenetic alopecia is caused by the effects of the androgen dihydrotestosterone on the hair follicle. In Caucasian males, its prevalence increases with age: it affects about 30% of males aged 30 and 80% of 70-year-old males.[8] Alopecia areata is an autoimmune, inflammatory form of hair loss which can affect small defined areas, or in worse cases lead to the entire loss of body and scalp hair. It has a lifetime incidence of about 2%.[9] While both androgenetic alopecia and alopecia areata very often have a genetic component,[8,9] they should not be confused with congenital hair loss disorders.

The hair follicle is the skin appendage responsible for hair growth. Many different genes are expressed in the hair follicle, some of which have been identified to be playing a role in congenital hair loss disorders. These disorders are much rarer than alopecia areata or androgenetic alopecia. Depending on whether additional (non-dermatological) symptoms are present or not, they can be classified into syndromic or non-syndromic variants. There are both autosomal dominant and recessive forms of hereditary alopecia and the underlying molecular changes can be very different.[10] The following genes have been chosen to provide some examples of the many different molecular causes and do not represent an extensive list of all congenital hair loss disorders.

The *hairless* gene (*HR*) is located on chromosome 8 and plays a vital role in controlling the hair cycle. Mutations in the *HR* gene have an autosomal recessive (AR) mode of inheritance and lead to atrichia with popular lesions through loss of function. *U2HR* lies in the 5'-untranslated region of *HR* and presents an inhibitory open reading frame (ORF). Mutations in *U2HR* have an autosomal dominant (AD) form of inheritance and lead to overexpression of the *HR* protein, which leads to Marie Unna hereditary hypotrichosis, a non-syndromic genetic hair loss disorder.[11,12]

The *lipase H* gene (*LIPH*) lies on chromosome 3 and encodes phospholipase A1. Through diacylation, phospholipase A1 changes phosphatidic acid into 2-acyl-lysophosphatidic acid, which promotes hair growth. Mutations in the *LIPH* gene can cause hypotrichosis or woolly hair and have an AR mode of inheritance.[10,12]

Desmosomes are intercellular junctions containing multiple proteins working together (see below). Mutations in the genes encoding these proteins can lead to different forms of genetic hair loss. For example, mutations in the *DSG4* (*desmoglein 4*) gene on chromosome 18 can lead to AR monilethrix, a condition with hypotrichosis and follicular papules on the scalp. Similar phenotypes can be caused by AR mutations in *DSC3* (*desmocollin 3*, also found on chromosome 18). Mutations in the *PKP1* (*plakophilin 1*) gene on chromosome 1 are a possible cause of ectodermal dysplasia/skin fragility syndrome with skin fragility, sparse hair and nail changes and possessing an AR mode of inheritance. Finally, Naxos disease and Carvajal syndrome are two similar diseases also caused by desmosome dysfunction. Naxos disease includes woolly hair, palmoplantar keratoderma (PPK), and arrhythmogenic right ventricular cardiomyopathy. It can be caused by mutations of the *JUP* (*junctional plakoglobin*) gene on chromosome 17 or of the *DSC2* (*desmocollin 2*) gene on chromosome 18 (both have an AR mode of inheritance). Carvajal syndrome includes woolly hair, PPK and left ventricular cardiomyopathy and can be caused by AR mutations of the *DSP* gene (see below).[10]

1.1.6 Desmosomes

Desmosomes are intercellular junctions that play an important role in cell-to-cell adhesion and in anchoring the keratin intermediate filaments of the cytoskeleton to the plasma membrane. This way, desmosomes increase the mechanical resilience of tissues. For this reason, they are common in tissues that are often exposed to mechanical stress, like heart and skin.[13,14] Desmosomes also have very important functions in the hair follicle, which is why defects in desmosomes can cause hair symptoms.[10]

The desmosome comprises the extracellular core region (which houses transmembrane glycoproteins), the outer dense plaque and the inner dense plaque. The transmembrane glycoproteins found in desmosomes are called desmogleins and desmocollins and belong to the cadherin family of adhesion molecules. The extracellular domains of these proteins attach to the cadherins of desmosomes of other cells, thereby forming intercellular adhesion, while the cytoplasmic domains bind to plakoglobin in the outer dense plaque. The outer dense plaque therefore includes the cytoplasmic domains of desmogleins and

desmocollins, plakoglobin, as well as plakophilins and desmoplakin. Plakoglobin and plakophilins are members of the armadillo family of linker proteins and bind to desmoplakin. In the inner dense plaque, desmoplakin attaches to keratin intermediate filaments (tonofilaments). This way, desmoplakin anchors the cytoskeleton to the desmosome and the plasma membrane.[15]

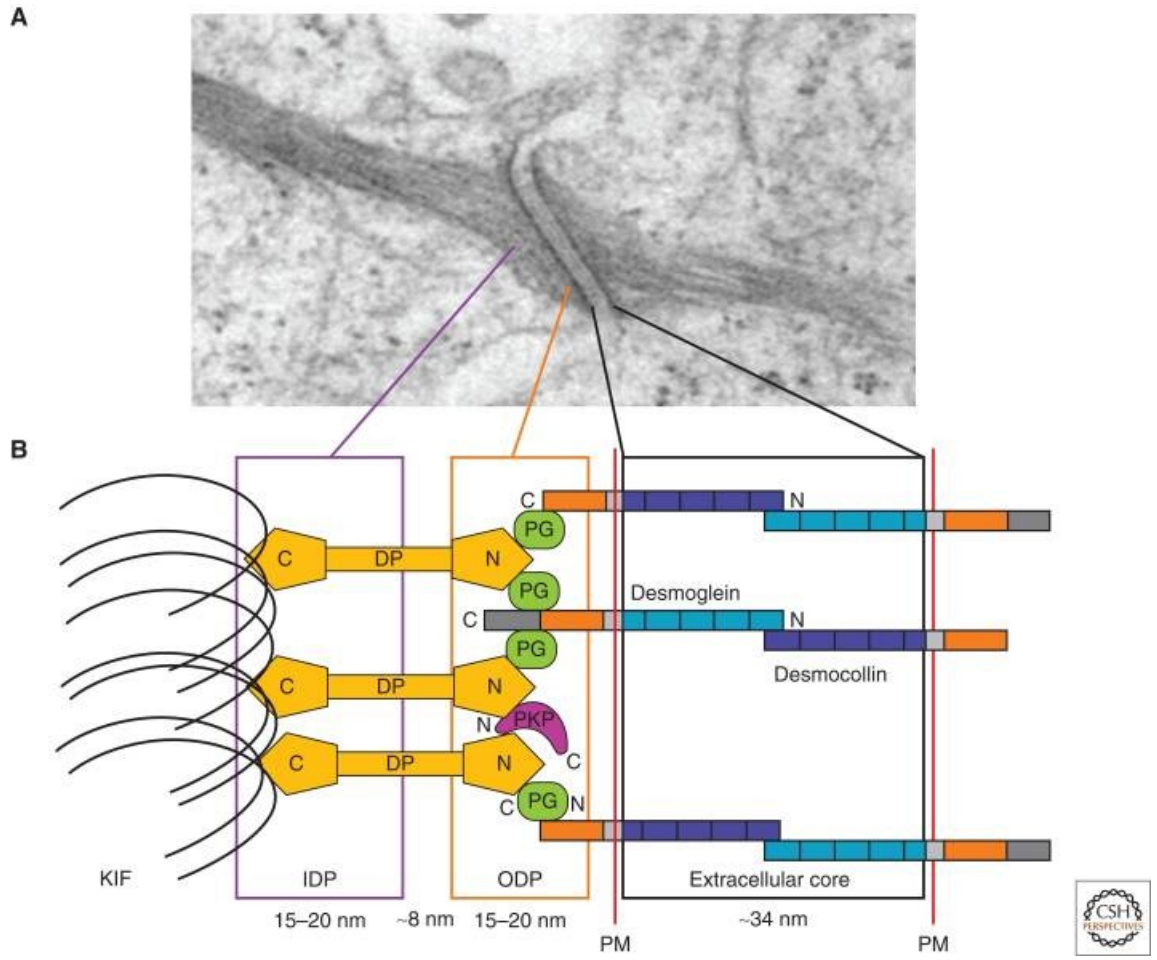

Figure 1 - Electron micrograph of a desmosome (A) and schematic diagram of a desmosome (B). KIF=keratin intermediate filaments, IDP=inner dense plaque, DP=desmoplakin, ODP=outer dense plaque, PG=plakoglobin, PKP=plakophilin. C refers to the C-terminal domain and N to the N-terminal domain of the respective protein. See text for an explanation of the desmosomal proteins.

Reprinted from Cold Spring Harbor Perspectives in Biology, Volume 1, Issue 2, Emmanuella Delva, Dana K Tucker, Andrew P Kowalczyk, The desmosome, Copyright 2009 Cold Spring Harbor Laboratory Press, with permission from Cold Spring Harbor Laboratory Press
Originally adapted from Journal of Cell Science, Volume 119, Issue 5, Margaret D Kottke, Emmanuella Delva, Andrew P Kowalczyk, The desmosome: cell science lessons from human diseases, 797-806, Copyright 2006 The Company of Biologists, with permission from The Company of Biologists

1.1.7 The Role of Desmoplakin in the Skin

For one of the two families examined in this paper, a mutation was found in the *DSP* (*desmoplakin*) gene, encoding the protein desmoplakin.

Desmoplakin is the most prevalent protein of the desmosome. It contains an amino-terminal domain (N-terminal domain), which attaches to plakoglobin and plakophilins, and a carboxy-terminal domain (C-terminal domain), which attaches to intermediate filaments. These two domains are connected by an alpha-helical coiled-coil rod domain.[14,15] Three different desmoplakin isoforms exist (DSPI, DSPII and DSPIa), which are encoded by the *DSP* gene in humans and created through alternative splicing.[16] The *DSP* gene lies on chromosome 6p24.3, comprises 24 exons and spans approximately 45 kDa of genomic DNA.[17] The three desmoplakin isoforms only differ in the length of the alpha-helical rod domain. Compared to DSPI, the alpha-helical rod domain is approximately half as long in DSPIa and only about one third as long in DSPII**.** While DSPI and DSPII are usually expressed together, DSPI is found more commonly in heart tissue and DSPII is found more commonly in complex epithelial tissues. DSPIa is also expressed in heart and epithelial tissues, albeit at much lower levels than DSPI and DSPII.[16]

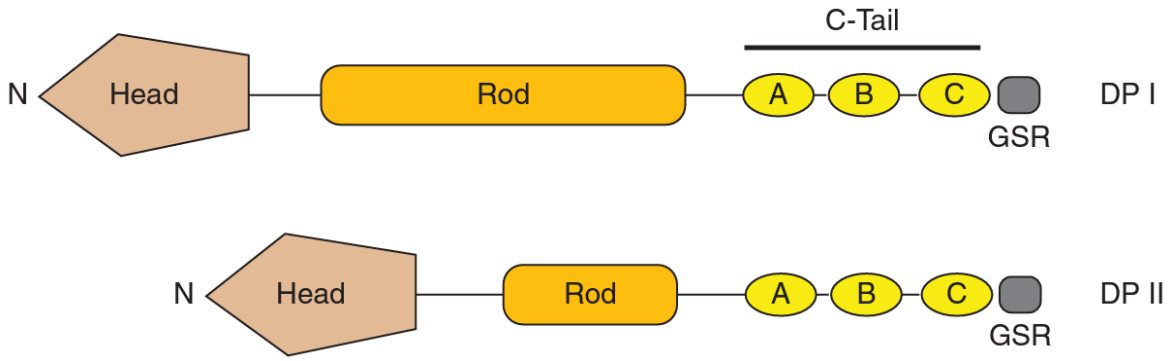

Figure 2 - Two isoforms of the desmoplakin protein, desmoplakin I (on top) and desmoplakin II (on bottom). The third known isoform, desmoplakin Ia, is not represented.
All isoforms share the same main domains, namely an amino-terminal domain (N-terminal domain), a carboxy-terminal domain (C-terminal domain) and an alpha-helical rod domain connecting the other two. The only difference in the isoforms lies in the varying length of the rod domain.

Reprinted from Cold Spring Harbor Perspectives in Biology, Volume 1, Issue 2, Emmanuella Delva, Dana K Tucker, Andrew P Kowalczyk, The desmosome, Copyright 2009 Cold Spring Harbor Laboratory Press, with permission from Cold Spring Harbor Laboratory Press

A great number of mutations in the *DSP* gene with several phenotypes and differing severity has been described in literature. Mutations with an autosomal dominant mode of inheritance can lead to arrhythmogenic right ventricular dysplasia and keratosis palmoplantaris striata II. On the other hand, the consequences of autosomal recessive mutations can range from skin fragility with woolly hair syndrome to lethal acantholytic epidermolysis bullosa.[18] The vital role of desmoplakin in correct desmosome function becomes even more apparent when looking at mouse experiments. Knockout mice with their *DSP* genes inactivated on both alleles form drastically fewer desmosomes than wildtype mice and die few days after implantation.[19]

1.2 Disease Gene Identification in Disorders with Presumed Autosomal Recessive Inheritance

It is estimated that over 5000 rare monogenic diseases exist, for half of which the causative gene has not yet been discovered. Furthermore, there is evidence that many diseases previously assumed to have complex multifactorial inheritance might in fact be caused by yet unknown monogenic mutations. Identifying such genes is a prerequisite for molecular diagnosis of patients and augments our understanding of the physiology of the underlying protein and the pathogenesis of the disease. Thus, the identification of a gene or mutation can serve as the first step towards developing therapeutic interventions.[20]

Before the wide availability of Next Generation Sequencing, Mendelian disease genes were identified through Sanger sequencing of previously selected candidate genes. Potential selection criteria for candidate genes include resemblance to previously discovered mutations with related pathologies or predicted role of the altered protein in the pathogenesis. However, the most successful approach for the identification of candidate genes has been the creation of a genetic positional mapping, as it does not require pre-existing knowledge of the gene or the protein it encodes. The following procedures have established themselves as the tools of choice for genetic positional mapping: karyotyping, homozygosity mapping, analysis based on copy number variation (CNV) or single-nucleotide polymorphism (SNP) and linkage analysis. However, each of these methods has its limitations, which is why a combined, sequential approach usually provides the best results.[20]

Next Generation Sequencing (NGS) has become increasingly affordable and therefore available to more researchers in the past decade.[21] With NGS, the amount of DNA that can be sequenced in a given time is magnitudes above that of Sanger sequencing. Nowadays, a whole genome can be sequenced in a short period of time. This change has eliminated the need for genetic positional mapping and the identification of candidate genes prior to sequencing. Instead, the challenge now lies in the interpretation of the millions of genomic variants identified per genome. Out of these millions of variants, the causative variant or variants need to be filtered and correctly identified.[20]

1.2.1 Consanguinity

The term consanguinity describes unions between individuals who share at least one common biological ancestor. In clinical genetics, the term usually refers to unions between second-degree cousins or closer. Nowadays, consanguinity is viewed negatively in most Western societies and is therefore often discouraged and legally restricted, or even illegal in certain states. However, in other parts of the world it is still common practice to perform consanguineous marriages, especially in the Middle East, North Africa and parts of Asia. While historically consanguinity has been common in ruling and upper classes in Western societies, today most consanguineous marriages are found in poor and rural communities. The reasons given for consanguineous unions are usually social and economic, as they are supposed to strengthen family ties and eliminate the necessity of paying dowries.[22]

It is estimated that worldwide about 10% of the population are offspring of a consanguineous union. The countries with the highest rates of consanguinity are Kuwait with 68%, Saudi Arabia with 56% and Pakistan with 55-59%. In certain Pakistani communities, consanguinity rates may even be as high as 76%.[23] According to the *2017-18 Pakistan Demographic and Health Survey*, 50% of the marriages in Pakistan are between first-degree cousins.[24]

Children resulting from a consanguineous union have a higher risk of genetic disorders, especially autosomal recessive disorders. On average, first-degree cousins share 1/8 or 12,5% of their genes. Therefore, the offspring of consanguineous unions between first-degree cousins are, on average, autozygous at 1/16 or 6.25% of all gene loci. The term autozygous describes that at specific regions in their genomes, these individuals have received identical gene copies from both parents (i.e., homozygous by descent). The percentage of autozygous regions can be estimated using the coefficient of inbreeding (F), with F values increasing the closer a couple is related. For example, the F value in first-degree cousins is 1/16, while for second-degree cousins it is 1/64. In addition to the increased risk for autosomal recessive disorders, the offspring of consanguineous couples is also presumed to be at a higher risk for disorders of multifactorial or complex inheritance. When estimating the risk for disorders in the offspring of a consanguineous union, the general population risk and family history must be considered in addition to the coefficient of inbreeding.[25] Compared to the general population, unions between first-degree cousins without any previously known genetic disorders in the family have a risk increase of 1.7%-2.8% for birth defects in their children. It has also been shown in multiple

studies that the risk of neonatal and infant deaths is 1.1% higher in the offspring of first-degree cousins than in the offspring of unrelated parents.[26]

In a more recent study, Fridman et al.[27] examined the rate of at-risk couples in a Dutch and an Estonian cohort (i.e., couples in which both partners are carriers for at least one identical autosomal recessive mutation). Their findings indicate that first-degree cousins have a 16-fold risk for an affected child with an autosomal recessive disease compared to unrelated parents.[27]

1.2.2 Homozygosity Mapping

Offspring of consanguineous parents will be autozygous at a certain percentage of gene loci and therefore have increased coefficient of inbreeding. The closer the parents are related, the higher the coefficient of inbreeding.[25] This leads to an increased probability of disease-causing genes in a homozygous region.[28] Therefore, in cases where consanguinity is suspected or confirmed and a rare recessively inherited disorder seems probable, it can be assumed that the pathogenic variant lies within a homozygous region.[20] Additionally, it has been shown that homozygous regions created through consanguinity usually span multiple megabases.[29] Due to these considerations it is possible to search only for homozygous variants of genes in homozygous regions and exclude variants outside these regions. Thus, the total number of variants detected by Whole Exome Sequencing can be drastically reduced, which facilitates the identification of the causative variant.[20] Homozygosity mapping can be carried out by SNP arrays or with data from WES.[30]

1.2.3 Sanger Sequencing

Sanger sequencing or, as Frederick Sanger named it in his in his 1977 publication, “DNA sequencing with chain-terminating inhibitors”,[31] is considered the first generation of sequencing.[32] As its name implies, this technique relies on chain-termination after incorporation of dideoxynucleotides in DNA-replication, instead of the traditional deoxynucleotides.[31] Sanger sequencing has presented the gold standard for determining nucleotide sequences for 30 years and was also used in sequencing the first human genome. However, the cost and effort associated with genome sequencing by Sanger sequencing have proven to be too exorbitant for clinical work. With the advent of Next Generation Seuqencing, this problem has been solved.[32]

While previously Sanger sequencing has been used for validating NGS results, this application is not advised anymore.[33] However, Sanger sequencing can still present a useful tool, e.g., in segregation analysis.

1.2.4 Next Generation Sequencing and Whole Exome Sequencing

In 2001, the Human Genome Project finished sequencing the first human genome using Sanger sequencing. This project took 13 years to complete, and the total cost is estimated to have been around 2.7 billion US-$.[34] Twenty years later, a whole genome can be sequenced using Next Generation Sequencing (NGS) in less than a day for under 1000 US-$.[21] The advances brought about by NGS have revolutionized the research of both rare and common genetic disorders. While in the past, specific regions had to be identified and chosen for sequencing, nowadays sequencing data for a whole genome can be generated in a short period of time.[34]

Next Generation Sequencing and massively parallel sequencing are synonymous expressions, describing various platforms that are capable of parallel sequencing of millions of small DNA fragments. The sequence of base pairs sequenced in such a fragment is referred to as a read. These fragments are then pieced together through comparison with the reference human genome. In this process, each base is sequenced multiple times, which helps to ensure precise data and mitigate artefacts. NGS systems can achieve outputs magnitudes above those of Sanger sequencing, with gigabases of nucleotide sequence achieved per run.[35] Nevertheless, they are highly accurate and reliable tools, so that routine Sanger validation of NGS results is not required and might even lead to contesting variants correctly identified by NGS.[33] However, NGS is limited by creating relatively short reads, which can lead to misassembly and gaps within longer repeated sequences, as well as impeding the reliable detection of larger structural variations.[36]

Whole Exome Sequencing (WES) describes the sequencing of only the exome, i.e., the parts of the DNA that are coding for proteins, as opposed to sequencing the whole DNA with Whole Genome Sequencing (WGS).[20] The exome only comprises about 30 Mb or 1% of the human genome, yet it contains about 85% of all disease-causing mutations.[34] Performing WES has the advantages of greatly reduced costs compared to WGS, as multiple exomes can be sequenced in a single NGS run, as well as a less overwhelming amount of data generated.[20]

With WES, usually 20,000 to 50,000 variants are identified in each sequenced exome**.** For this reason, multiple filters need to be applied before being left with a manageable number

of candidate mutations. The first step of filtering aims at excluding false-positive calls by applying quality criteria. These include the total number of reads as well as the percentage of reads including the variant (for heterozygous variants at least 20% of the reads must show the variant and for homozygous variants at least 80%). In the next step, synonymous variants, i.e. variants that do not change the encoded amino acid, as well as variants outside the coding regions and splice-sites are filtered out, as these are expected to have minimal effect on the protein.[20] Further filtering includes the exclusion of known variants, variants with a high population frequency, as well as filtering based upon the effect on the open reading frame and the protein (filtering steps taken from *Illumina VariantStudio* 3.0; Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA). This way the total number of variants can be reduced to usually less than 500 potentially pathogenic variants. The largest effect is achieved through the exclusion of known variants, which are typically taken from large databases or published studies. Through this step alone, a reduction of potential candidate mutations by 90-95% can be achieved.[20] It should be noted that the exact filtering steps can vary. If after filtering no plausible genes are left over, the filtering steps should be reconsidered and adjusted if necessary.

1.3 The Families

1.3.1 SK10

The family SK10 is a Pakistani family from the district Tank, in the province Khyber Pakhtunkhwa of Pakistan.

The pedigree of the family includes four generations with 3 affected individuals in the fourth generation alive. The phenotype of these affected individuals includes woolly hair/alopecia, dry skin, and nail irregularities. While the female patient has woolly hair on the scalp, the living male patients have basically no scalp hair at all. Eyelash and eyebrow growth is weak and in the male individuals, facial hair growth is greatly reduced (pictures of the phenotype in Figure 3). The hypotrichosis also includes body hair. Environmental factors and infections could be excluded as causative for the phenotype observed. The parents of the affected individuals did not show any of these signs, neither did the unaffected siblings. There were no abnormalities observed in teeth, eyes or hearing and the affected individuals did not report excessive or reduced sweating. Ultrasound examination of the abdomen showed no major pathologies of the gallbladder, liver, spleen, pancreas or the urinary tract and no cardiac abnormalities were present in the performed ECG.

Figure 3 - The phenotype as presented by the individuals SK10-1 (on the left) and SK10-3 (on the right)
Pictures kindly provided by Prof. Muzammil A. Khan

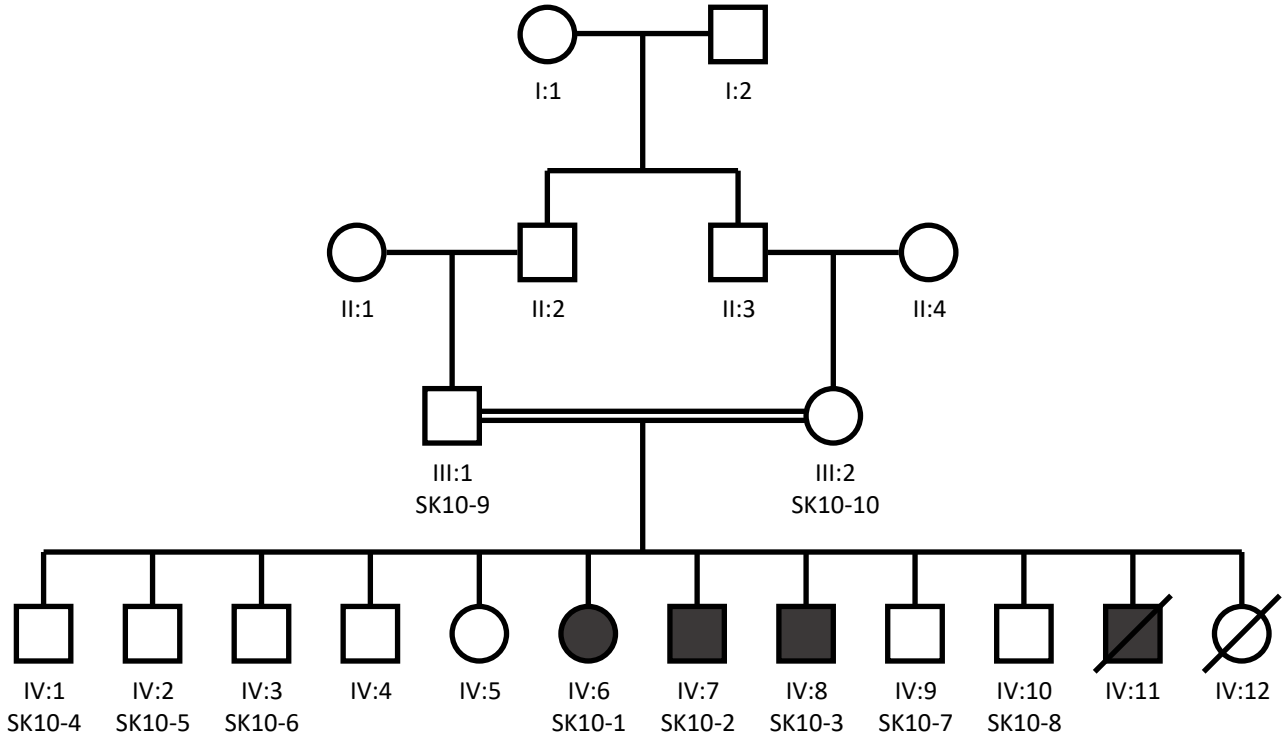

Figure 4 - Pedigree of the family SK10.
Squares symbolize male individuals and circles symbolize female individuals. Filled symbols represent people with the pathological phenotype, while people represented by empty symbols do not share that phenotype. Crossed out symbols stand for deceased individuals. A horizontal double line represents consanguinity.

As the pedigree demonstrates, the couple in third generation is of consanguineous union, as the two individuals are first degree cousins. Out of the twelve individuals in the fourth generation, 4 were affected by the abnormalities described. At the time of writing this thesis, an affected male individual has died at 2 months of age in an accident and an unaffected female individual has died at 5 months of age due to a measles infection.

Observing the pedigree, a traditional dominant form of inheritance can be excluded because the phenotype is not present in every generation, but only the fourth generation. Since there is an affected female individual (SK10-1), and her father (SK10-9) is unaffected, X-chromosomal recessive transfer of the disease seems highly unlikely and Y-chromosomal transfer can be excluded. Therefore, for further analysis an autosomal recessive mode of inheritance is assumed for the described condition. Due to the high inbreeding coefficient, a homozygous mutation in a homozygous region seems highly probable, however compound hemizygous variants should also be considered.

DNA material was isolated from blood samples taken from the 3 living affected individuals in the fourth generation (SK10-1, SK10-2 and SK10-3) and 5 of their unaffected siblings (SK10-4 through to SK 10-8), as well as their parents (SK10-9 and SK10-10).

1.3.2 SK6

The family SK6 is a Pakistani family from the district D. I. Khan, in the province Khyber Pakhtunkhwa of Pakistan.

The pedigree of the family includes four generations with 2 affected individuals in the fourth generation and 3 in the third. The affected individuals are affected by dry skin and skin scaling (as can be seen in Figure 5) and develop blisters on their skin upon exposure to sunlight, which develop pus. Of the two affected individuals in the fourth generation, the older brother (SK6-1) has died of spinal cancer after the initial examination. The younger affected brother (SK6-2) has developed thalassemia-like symptoms, for which the administration of blood transfusions has been necessary. SK6-2 also presents dust allergy, bad tolerance to high temperatures with hyperhidrosis, and photophobia with normal eyesight. The parents of the affected individuals do not show any of these signs.

Figure 5 - The phenotype as presented by the individual SK6-2
Pictures kindly provided by Prof. Muzammil A. Khan

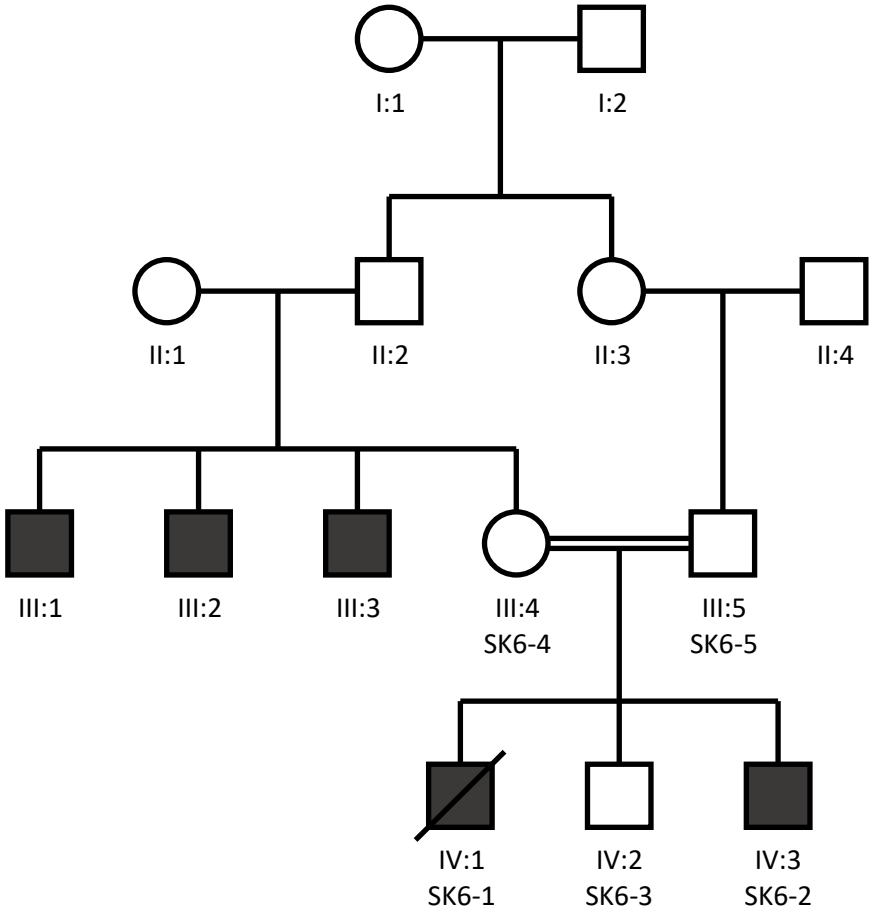

Figure 6 - Pedigree of family SK6.
The squares symbolize male individuals; the circles symbolize female individuals. Filled symbols represent people with the pathological phenotype, while people represented by empty symbols do not share that phenotype. The crossed out square stands for a deceased individual. A horizontal double line represents consanguinity.

As can be seen in the pedigree, there is evident consanguinity between the individuals SK6-4 and SK6-5. The 3 affected persons in the third generation are the brothers of SK6-4. Unfortunately, these three individuals could not be recruited for analysis. It is remarkable that only male individuals are affected and therefore X-chromosomal transfer of the disease cannot be excluded. Y-chromosomal transfer, however, can be eliminated for further analysis due to one brother in the fourth generation and the father not showing the phenotype (SK6-3 and SK6-5, respectively). Therefore, the most likely mode of inheritance is X-chromosomal recessive, but autosomal recessive mutations must be considered too.

DNA material was isolated from blood samples taken from the 2 affected individuals in the fourth generation (SK6-1 and SK6-2), their unaffected brother (SK6-3) and their parents (SK6-4 and SK6-5).

2 Materials and Methods

2.1 Recruitment and Examination

For the purpose of this thesis, two unrelated, consanguineous Pakistani families with different forms of skin diseases were recruited by Professor Muzammil A. Khan of Gomal University, Dera Ismail Khan (D. I. Khan), Pakistan. The persons involved were informed about the procedures and gave their written consent for genetic testing and use of their photographs. Ethical approval was given by the ethical review board of Gomal University. The family history was taken and physical examination of the individuals was performed by Prof. Khan. DNA material was extracted from Blood samples taken in D. I. Khan and then sent to the Institute of Human Genetics of the Medical University of Graz, Austria.

2.2 Whole Exome Sequencing

After arrival, the DNA samples were quantified using *Qubit 2.0 Fluorometer* (ThermoFisher Scientific Inc., 168 Third Avenue, Waltham, MA 02451, USA). For each family, out of the affected individuals, the samples with the most promising results were selected for whole exome sequencing (WES). For the family SK10, SK10-1 was identified as the best sample and WES was performed on a *MiSeq* Sequencer (Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA) at the Institute of Human Genetics of the Medical University of Graz. Due to low quality in the WES results for SK10, another run was performed. In this second run, the number of reads remained low and the general quality suboptimal. Nevertheless, it was decided that analysis of this data should be attempted. For the family SK6, the samples of both affected individuals (SK6-1 and SK6-2) were sent to the Molecular Neuropsychiatry and Development Laboratory of the CAMH (Centre for Addiction and Mental Health) in Toronto, Canada, where WES was performed on an *Illumina* (Illumina Inc., see above) NGS platform.

The data obtained from the WES was analysed using *Illumina VariantStudio 3.0* (Illumina Inc., see above) and the *Integrative Genomics Viewer* (IGV).[37] For SK6, further analysis was performed using *Golden Helix* (Golden Helix Inc., 1487 North 14th Avenue, Bozeman, MT 59715, USA) with the help of Doctorate student Jasmin Blatterer, MSc from the Institute of Human Genetics of the Medical University of Graz. The reference genome build used in every step was GRCh37/hg19.

2.3 Homozygosity Mapping

After WES, each DNA sample was examined for homozygous regions by uploading the VCF file to *HomozygosityMapper*.[38] Unfortunately, for SK10 no homozygous regions could be identified. This was due to the low quality of the SK10-1 WES results, caused by low read counts and parts of the genome not getting covered sufficiently. For SK6, homozygosity mapping was performed for SK6-1 and SK6-2 individually and then the homozygous regions shared by both were identified. These homozygous regions were considered in the search for candidate genes.

2.4 Sanger Sequencing

Sanger sequencing was used to evaluate the segregation of the identified variants. The required primers were created using the *UCSC genome browser*[39,40] (available at http://genome.ucsc.edu) together with its *In-Silico PCR* tool, as well as *Primer3*.[41] In *Primer3*, a product size range of 150 – 550bp was specified. The primers were ordered from the Swiss company *Microsynth* (Microsynth AG, Schützenstrasse 15, 9436 Balgach, Switzerland). After arrival, the primers were diluted to a concentration of 10µM using nuclease-free water.

For the DNA amplification the following reaction mix was used:

- 0.5µl DNA (~50ng/µl)
- 0.5µl forward primer (10µM)
- 0.5µl reverse primer (10µM)
- 6.0µl *HotStarTaq® Master Mix* (QIAGEN NV, Hulsterweg 82, 5912 PL Venlo, The Netherlands)
- 4.5µl *LiChrosolv* water (Merck KGaA, Frankfurter Straße 250, 64293 Darmstadt, Germany)

For primers with a GC-content of 50% or more, 1µl of water per sample was replaced with 1µl of dimethyl sulfoxide.

In the next step, the PCR for amplification was performed on a *9800 Fast Thermal Cycler* by *Applied Biosystems* (ThermoFisher Scientific Inc., 168 Third Avenue, Waltham, MA 02451, USA). The temperatures and times of the PCR program used were as follows:

Temperature	**Time**	**Cycles**
94°C	15 minutes	1
94°C 57°C 72°C	30 seconds 30 seconds 45 seconds	34
72°C	7 minutes	1
4°C	∞	

Table 1 - PCR program used for amplification

Following this, correct amplification was confirmed through gel electrophoresis with a 1% agarose gel. For this, 1g of LE Agarose (Biozym Scientific GmbH, Steinbrinksweg 27, 31840 Hessisch Oldendorf, Germany) was dissolved in 100ml of 1x TBE (Tris/Borate/EDTA) buffer under heat. Then the mixture was cooled down and 10µl of *GelRed*™ (Biotium, Inc., 46117 Landing Parkway, Fremont, CA 94538, USA) were added before solidification of the gel. Per sample, 5µl of PCR product were mixed with 1µl of loading dye. The samples and a standard of 1kb loading dye were applied to the gel and the electrophoresis was run at 130 volts and 400mA for 30 minutes.

In the next step, the following reaction mix was prepared on ice, with an individual reaction mix prepared both for forward and reverse primers:

- 1.0 µl PCR product
- 0.4 µl forward respectively reverse primer (10µM)
- 0.5 µl *BigDye® Terminator v3.1* (Applied Biosystems, see above)
- 1.4 µl 5x Sequencing Buffer (Applied Biosystems, see above)
- 7.2 µl *LiChrosolv* water (Merck, see above)

The PCR was performed on a *9800 Fast Thermal Cycler* by *Applied Biosystems* (see above). The PCR program used had the following times and temperatures:

Temperature	**Time**	**Cycles**
96°C 50°C 60°C	30 seconds 15 seconds 4 minutes	25
6°C	∞	

Table 2 - PCR program before Sanger Sequencing

Following this second PCR, the product was purified using *Sephadex™ G-50* (GE Healthcare, Chicago, Illinois, USA) columns. For this, *Sephadex G-50* was pipetted into *CentriStep™* columns (Princeton Separations, 100 Commerce Ave, East Freehold, NJ 07728, USA), centrifugated at 750rcf for 15 seconds and then again for 2 minutes, with any excess water being discarded after every centrifugation. Next, the PCR product was diluted 1:1 with *LiChrosolv* water (Merck, see above) and applied to the columns with new collection tubes. Subsequently, the columns were centrifugated again at 750rcf for 2 minutes.

In the final step, the purified PCR product samples were applied to a 96-well sequencing plate, centrifugated and then sequenced on a *3130 DNA Analyzer* by *Applied Biosystems* (see above).

Analysis of the Sanger sequencing was performed with *Chromas* v.2.6.6 (available at www.technelysium.com.au; Technelysium Pty Ltd, 8 Cordelia St, South Brisbane QLD 4101, Australia). The individual sequences were copied from *Chromas* in the FASTA format and compared with the reference genome build GRCh37/hg19 using the *BLAT* tool[42] of the *UCSC Genome Browser*.[39,40]

3 Results

3.1 SK10

The pedigree of the family SK10 comprises four generations with four affected individuals in the fourth generation, 3 being alive (see Figure 4). Affected individuals in this family are affected by hypotrichosis/woolly hair syndrome as well as dry skin and nail irregularities. The parents do not show any signs of hypotrichosis or skin changes. For Whole Exome Sequencing, the DNA sample of SK10-1 was chosen.

3.1.1 Search for Candidate Genes

WES was performed on a *MiSeq* Sequencer (Illumina Inc., see above) at the Institute of Human Genetics of the Medical University of Graz. Due to bad quality of the isolated DNA and poor results in the first run, a second WES run was performed on the same machine. Unfortunately, the result of this second run was only marginally better. A search for homozygous regions using the VCF-file from WES was attempted with *HomozygosityMapper*[38] but proved futile because of the poor WES results.

Next, the WES-data was analysed using the program *Illumina VariantStudio 3.0* (Illumina Inc., see above) for filtering. The following table shows the filtering steps that were applied together with the number of leftover variants after each step. The low number of variants before filtering are an indicator of suboptimal WES results:

Before filtering	152,847 variants
Population frequency <0.1%	58,948 variants
Homo- or hemizygous	4,398 variants
Consequence: missense, frameshift, stop gained, stop lost, initiator codon, in-frame insertion, in-frame deletion or splice	190 variants

Table 3 - SK10-1 filters applied and leftover variants after each step

The respective genes of the leftover variants were examined individually with the *UCSC genome browser*[39,40] and the *OMIM genes* plugin.[43] The reference genome used in every program was GRCh37/hg19. Genes were considered as candidate genes if a phenotype similar to that of SK10 had previously been described in *OMIM*. The gene *DSP* on

chromosome 6 was identified as a candidate gene with the following variant: NM_004415.2:c.1493C>T (GRCh37/hg19). This variant lies in exon 12 (out of 24) of the gene *DSP* and is in a well conserved position. In this variant, the normally present cytosine on base position 7,569,492 of chromosome 6 has been replaced with a thymine, leading to the substitution of proline with leucine (p.Pro498Leu) and therefore constituting a missense variant. In *gnomAD*,[44] the allele frequency of this variant was specified as $1.591*10^{-5}$ and the online prediction tool *MutationTaster*[45] described it as disease causing. The reads were verified using IGV2.10.3.[37] In the WES results, on position 7,569,492 of chromosome 6 all 12 reads showed thymine.

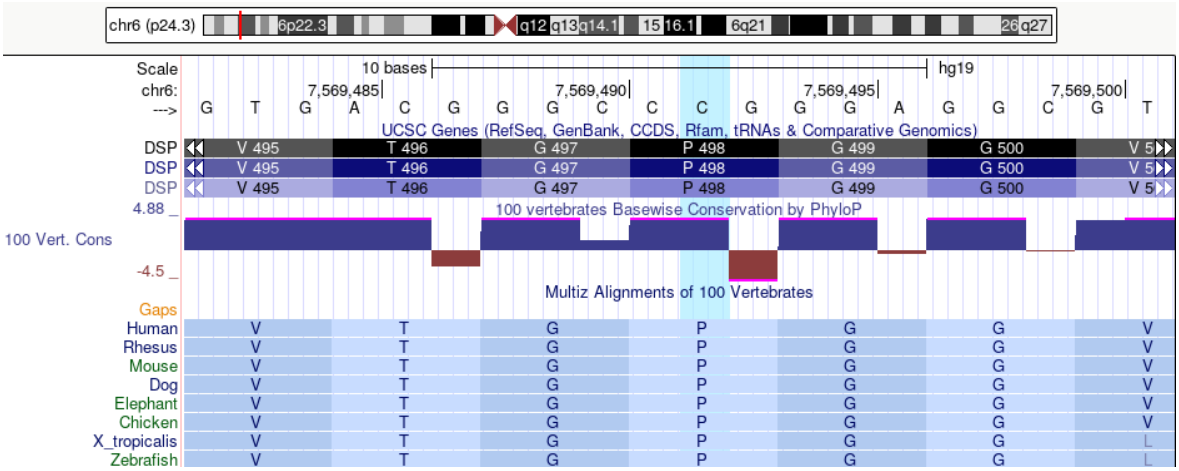

Figure 7 - Conservation of cytosine at position c.1493 respectively proline at p.498 of the *DSP* gene
Screenshot taken in the UCSC genome browser (GRCh37/hg19), available at http://genome.ucsc.edu

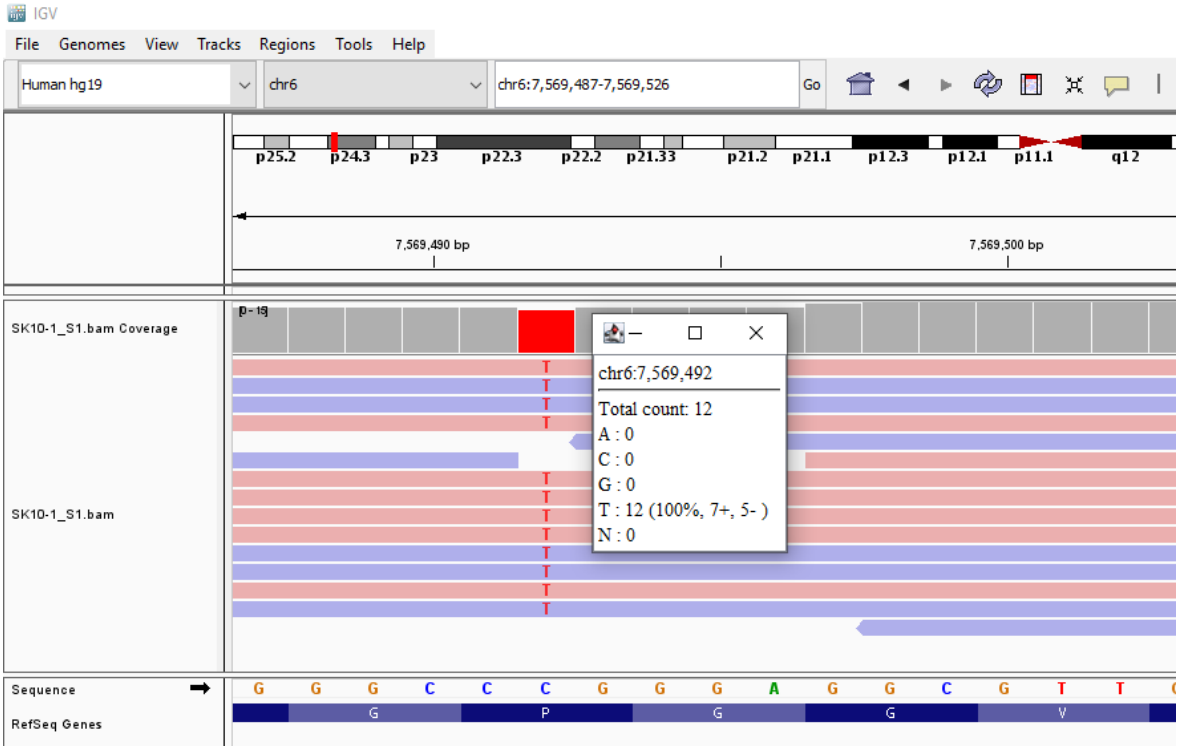

Figure 8 - WES reads at chr6:7,569,492 (GRCh37/hg19) of SK10-1
Screenshot taken in IGV2.10.3

3.1.2 Sanger Sequencing

Following the identification of the candidate gene and the variant, corresponding primers were designed using *Primer3*.[41] The product size was 398bp and there were no common SNPs in the primer sequences. Table 4 shows the forward and reverse primers.

	sequence	**GC-%**
Forward primer	TGAGGGGAAAAACGTAAAAGC	42.86%
Reverse primer	GAAGGTGTTTATTATGGCAAGG	40.91%

Table 4 - Forward and reverse primers for *DSP* exon 12

Figure 9 shows the results of the Sanger sequencing.

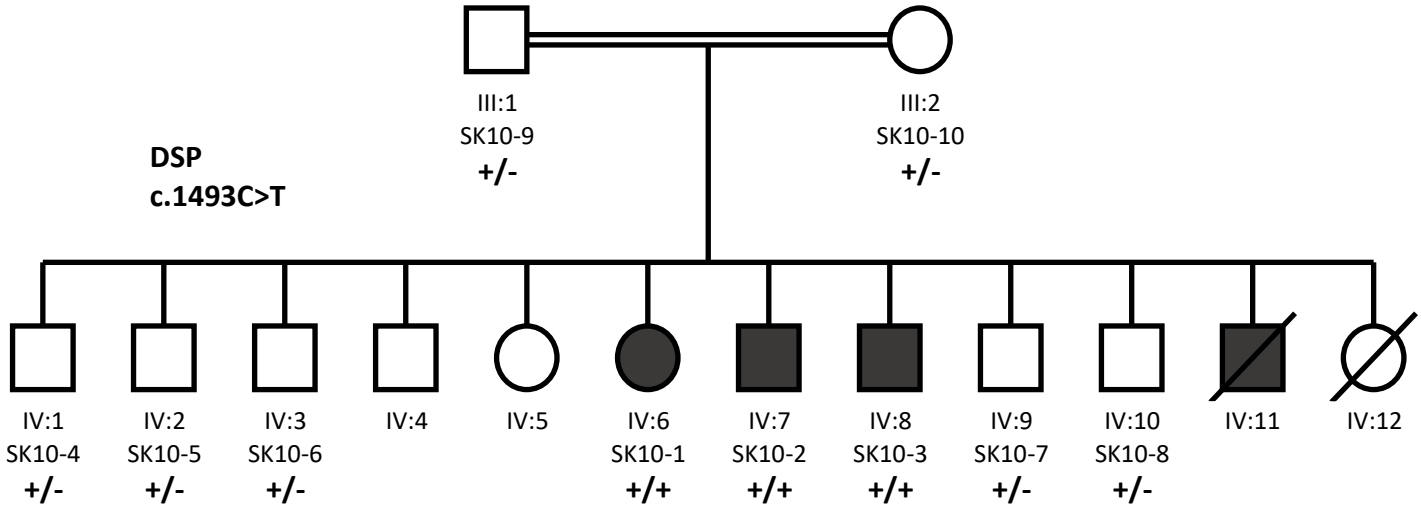

Figure 9 - Pedigree of the family SK10 with the results of Sanger sequencing

The pedigree combined with the Sanger Sequencing results reveals that the variant segregates as expected of an autosomal recessive disorder.

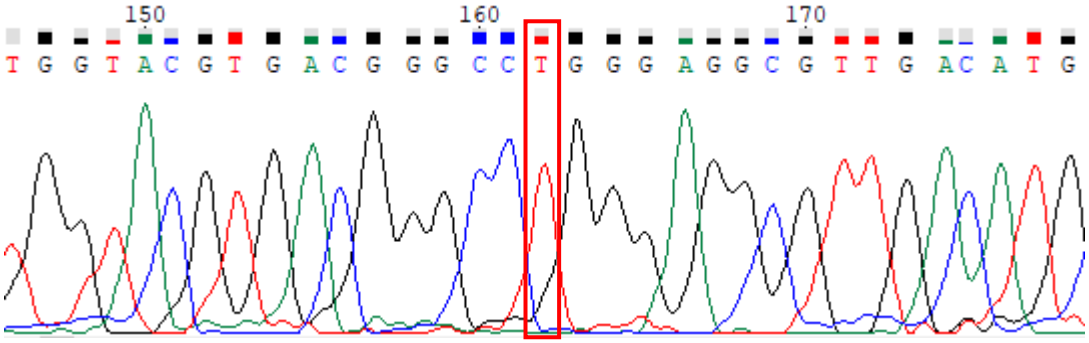

Figure 11 - Electropherogram of SK10-2 (homozygous for the variant)
Screenshot taken in Chromas v2.6.6

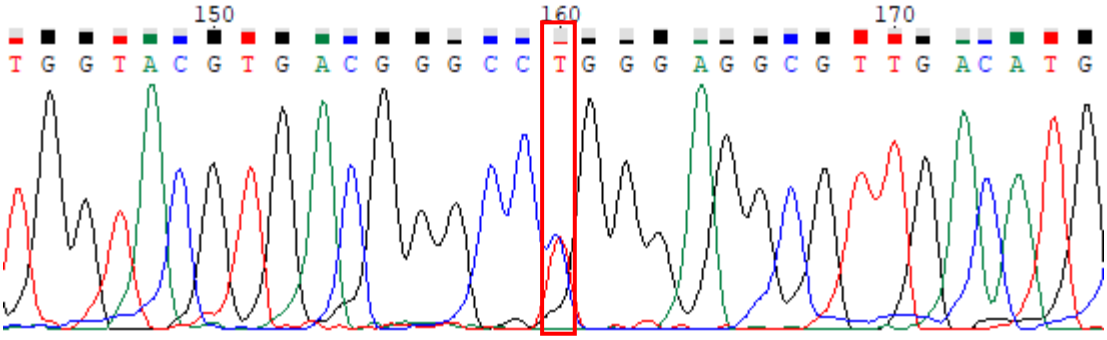

Figure 10 - Electropherogram of SK10-8 (heterozygous for the variant)
Screenshot taken in Chromas v2.6.6

Figures 10 and 11 show the electropherogram results of SK10-2 and SK10-8 as examples for a homozygous and a heterozygous genotype, respectively. In SK10-2, the cytosine which is normally present on base position 7,569,492 of chromosome 6 has been replaced with a thymine (c.1493C>T). In SK10-8, at the same position, cytosine and thymine have been sequenced roughly equally as often, which indicates a heterozygous genotype.

This mutation has previously been described in literature and constitutes a known cause of alopecia (see Discussion).

3.2 SK6

The pedigree of the family SK6 comprises four generations with three affected individuals in the third generation and two affected individuals in the fourth (see Figure 6). Affected individuals are affected by dry skin, skin scaling and hyperhidrosis and develop blisters on their skin upon exposure to sunlight. None of the parents of affected individuals show any of the symptoms. The DNA samples of SK6-1 and SK6-2 were used for Whole Exome Sequencing.

3.2.1 Search for Candidate Genes

WES performed on an *Illumina* NGS platform (Illumina Inc., see above) at the Molecular Neuropsychiatry and Development Laboratory of the CAMH in Toronto, Canada.

In all the following steps, the reference genome build used in every program was GRCh37/hg19. First, the WES data was used to create a homozygosity mapping of the homozygous regions shared between SK6-1 and SK6-2, using the online tool *HomozygosityMapper*.[38] Because this technique achieves a lower precision compared to a SNP array, the identified homozygous regions were extended by about 0.5Mb in each direction and rounded, which resulted in the following regions:

Chromosome 5	128,000,000 – 137,000,000
Chromosome 5	167,000,000 – 171,000,000
Chromosome 7	15,000,000 – 35,000,000
Chromosome 10	18,000,000 – 28,000,000
Chromosome 12	25,000,000 – 41,000,000
Chromosome 15	84,000,000 – 89,000,000
Chromosome 16	58,000,000 – 75,000,000
Chromosome 19	16,000,000 – 39,000,000

Table 5 - Homozygous regions shared by the affected individuals (SK6-1 and SK6-2)

The *UCSC genome browser*[39,40] was used together with the *OMIM genes*[43] plugin to search the homozygous regions for genes which could cause similar symptoms to SK6. These genes were compared with the variants identified by *Illumina VariantStudio 3.0* (Illumina Inc., see above). This method did not provide any plausible candidate genes.

In the next step, *Illumina VariantStudio 3.0* was used for filtering all the homo- and hemizygous variants of individual SK6-1. The filters that were applied and the corresponding leftover variants are shown in the following table.

Before filtering	1,300,016 variants
Population frequency <0.1%	175,732 variants
Homo- or hemizygous	108,084 variants
Consequence: missense, frameshift, stop gained, stop lost, initiator codon, in-frame insertion, in-frame deletion or splice	97 variants

Table 6 - SK6-1 filters applied for homo- & hemizygous variants and leftover variants after each step

The respective genes of the leftover variants were again examined in the *UCSC genome browser* with the *OMIM genes* plugin. The only relevant variant found this way was NM_005514.7:c.354_355del (GRCh37/hg19) of the gene *HLA-B* on chromosome 6. Mutations in this gene can lead to an increased susceptibility for Stevens-Johnson syndrome and toxic epidermal necrolysis. This variant was dropped because of the high allele frequency described in *gnomAD* (0.08686), because *MutationTaster* predicted the variant to be a polymorphism and because it fell outside of the homozygous regions identified before. The filtering was repeated looking for compound heterozygous variants. Table 7 shows the filters applied and the corresponding leftover variants:

Before filtering	1,300,016 variants
Population frequency <0.1%	175,732 variants
Heterozygous	67,648 variants
Consequence: missense, frameshift, stop gained, stop lost, initiator codon, in-frame insertion, in-frame deletion or splice	463 variants

Table 7 - SK6-1 filters applied for heterozygous variants and leftover variants after each step

Each gene that had two or more heterozygous variants was examined in the *UCSC genome browser* with the *OMIN genes* plugin. The only relevant combination of variants found were NM_001042440.4:c.1914+25_1914+28del (GRCh37/hg19) and NM_001042440.4:c.1914+26_1914+28del (GRCh37/hg19) of the gene *CAST* on chromosome 5. Mutations in this gene can cause peeling skin with leukonychia and acral punctate keratoses. This combination was also dropped because the variants had a population frequency of 0.01950 and 0.1816 respectively in *gnomAD* and because it fell outside of the homozygous regions identified.

In the next step, WES data analysis was performed in *Golden Helix* (Golden Helix Inc., 1487 North 14th Avenue, Bozeman, MT 59715, USA) with the help of Doctorate student Jasmin Blatterer, MSc from the Institute of Human Genetics of the Medical University of Graz. Table 8 shows the two filter strategies employed in *Golden Helix*.

Filter-Strategy 1
1) Quality&Artefacts
Filter=PASS
Genotype Quality >20
Exclusion of common artefacts (custom annotation from internal datasets)
Exclusion of intergenic regions
2) ACMG auto classification is not benign/likely benign
3) Zygosity = homozygous
4) 0x homozygous in GnomAD Exomes
Filter-Strategy 2
1) Quality&Artefacts
Filter=PASS
Genotype Quality >20
Read Depth >20
Exclusion of common artefacts (custom annotation from internal datasets)
Variant Allele Frequency >0.1
2) Population Frequency
AF GnomAD <0.01
3) ClinVar classification is not benign/likely benign
4) ACMG auto classification is not benign/likely benign
5) Variants of interest (variant fits at least one of the following criteria)
GeneRank >0,8 (HPO-Term-association)
Zygosity = Homozygous
Effect = Loss of Function
>2 variants/gene (count alleles by gene for potential compound heterozygous variants)
X-linked
Inheritance = autosomal dominant and variant is rare (GnomAD AF <0.001 and Allele count <5)
rare variant (0x in GnomAD)

Table 8 - Filter strategies used in *Golden Helix*

For this second analysis of WES data, the program *Golden Helix* was used because it allows for more precise filtering commands and the inclusion of HPO (Human Phenotype Ontology)[46] terms. For SK6, the following HPO terms were used: "HP:0011121 - abnormality of skin morphology" and "HP:0000613 – photophobia". Through the WES analysis in *Golden Helix*, two candidate genes were identified.

The first candidate gene identified was the gene *CARMIL2* with the following variant: NM_001013838.2:c.677A>T (GRCh37/hg19). This variant lies on exon 9 (out of 38) of the gene *CARMIL2*. In this variant, the normally present adenine on base position 67,681,084 of chromosome 16 has been replaced with a thymine, which constitutes a missense variant and leads to aspartic acid being replaced with valine (p.Asp226Val). Three different isoforms are known for this gene and this position is part of all three. Only one allele carrying this variant has been described in *gnomAD*, leading to an allele frequency of $4.02*10^{-6}$. The online prediction tool *MutationTaster* described this variant as disease causing. Additionally, the variant lies in one of the homozygous regions identified before and the position is well conserved. The reads were verified using IGV2.10.3.[37] Mutations in the gene *CARMIL2* have previously been linked with immunodeficiency leading to early-onset skin lesions (amongst other symptoms).

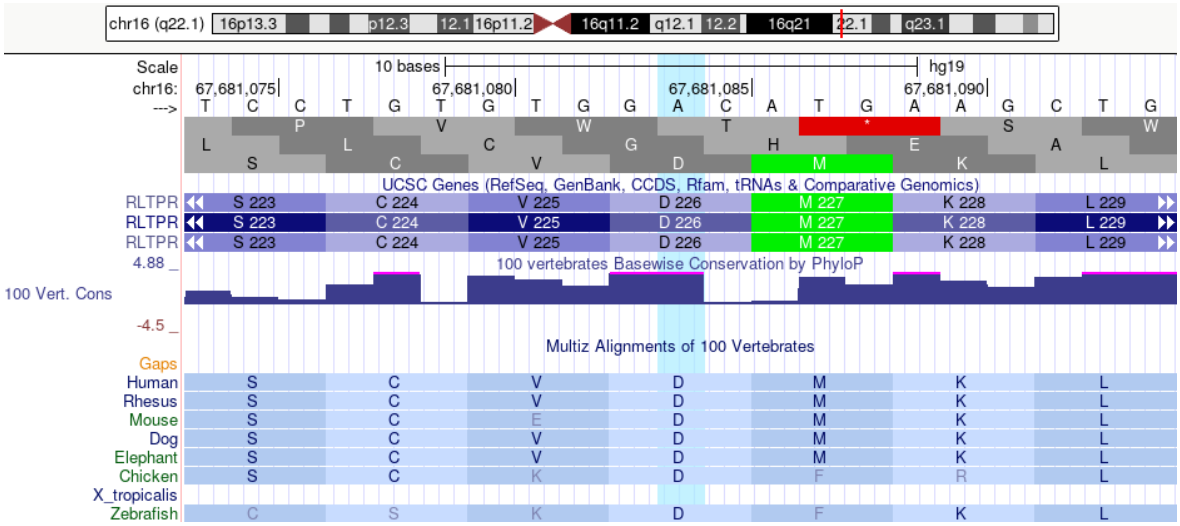

Figure 12 - Conservation of adenine at c.677 respectively aspartic acid at p.226 of the *CARMIL2* gene Screenshot taken in the UCSC genome browser (GRCh37/hg19), available at http://genome.ucsc.edu

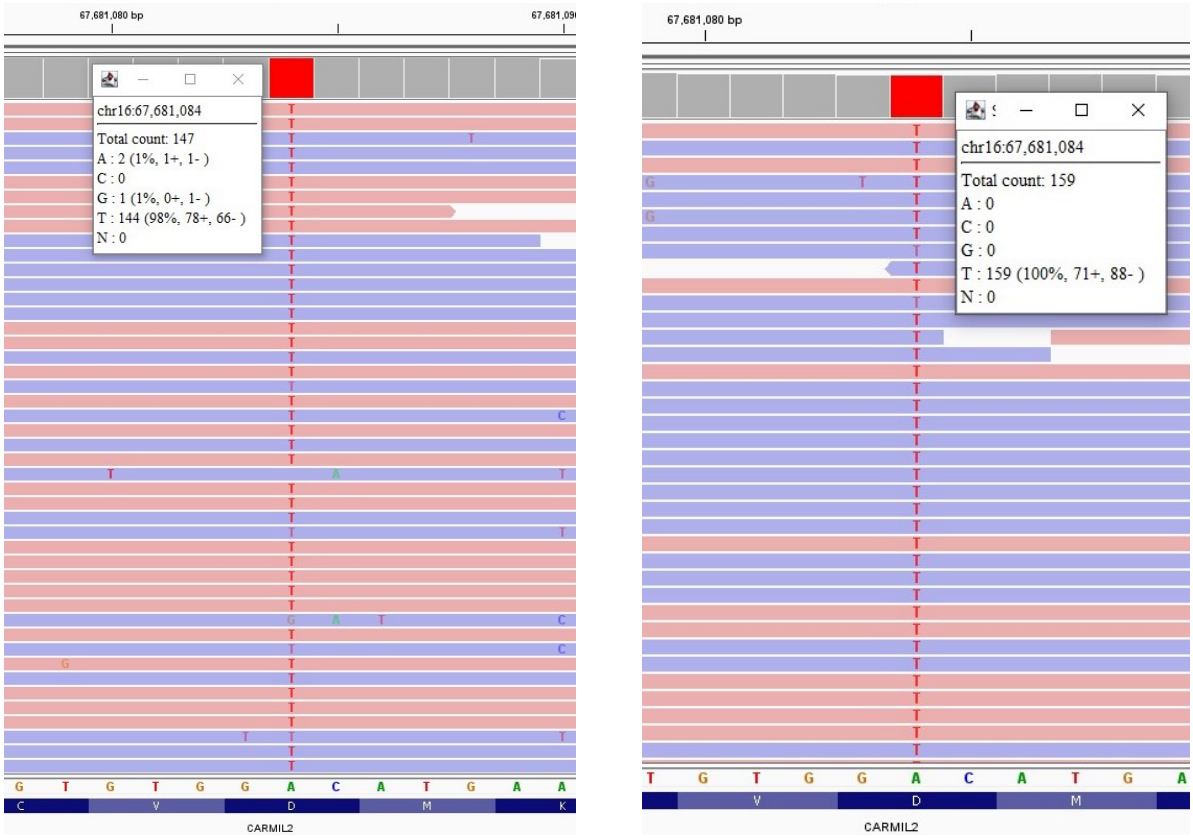

Figure 13 - WES reads at chr16:67,681,084 (GRCh37/hg19) of SK6-1 (right) and SK6-2 (left) Screenshot taken in IGV2.10.3

The second candidate gene was the gene *G6PD* with the following variant: NM_001042351.2:c.563C>T (GRCh37/hg19). This variant lies in the exon 6 (out of 13) of the gene *G6PD*. It is a missense variant in which the normally present guanine on base position 153,762,634 of the X-chromosome has been replaced with an adenine, which in turn causes the amino acid serine to be substituted by phenylalanine (p.Ser188Phe respectively p.Ser218Phe). Two different isoforms have been described for this gene and this position is part of both. *gnomAD* described the allele frequency of this variant with $2.301*10^{-3}$ and the online prediction tool *MutationTaster* predicted it to be a polymorphism. Nevertheless, it was decided that further analysis of this variant should be performed in addition to the *CARMIL2* variant described above. The reads were also verified for this variant in IGV2.10.3.[37] Mutations in the gene *G6PD* have been linked with haemolytic anaemia.

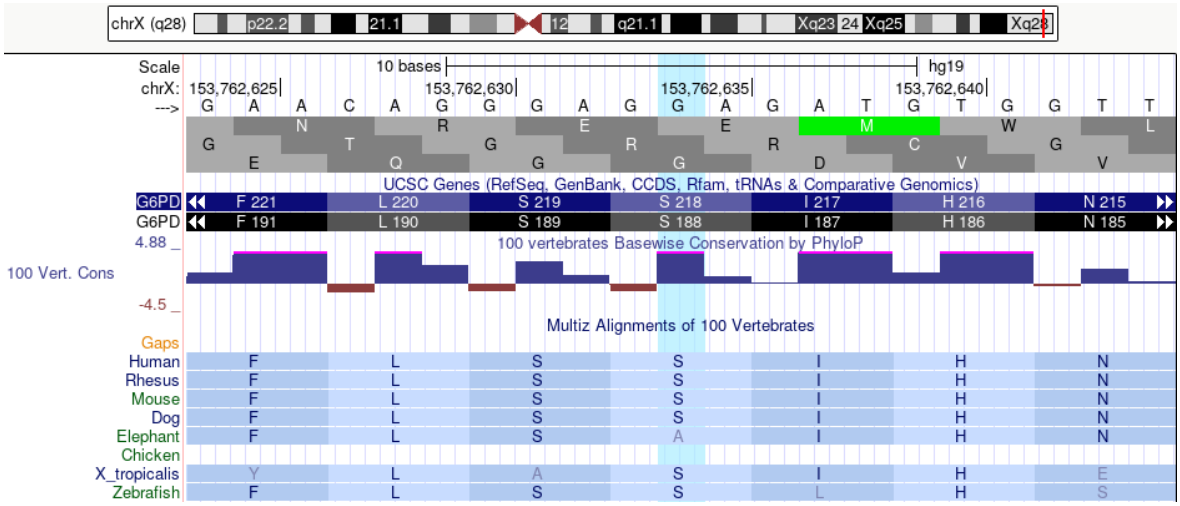

Figure 15 - Conservation of guanine at position 153,762,634 of the X-chromosome respectively serine at p.188/p.218 of the *G6PD* gene
Screenshot taken in the UCSC genome browser (GRCh37/hg19), available at http://genome.ucsc.edu

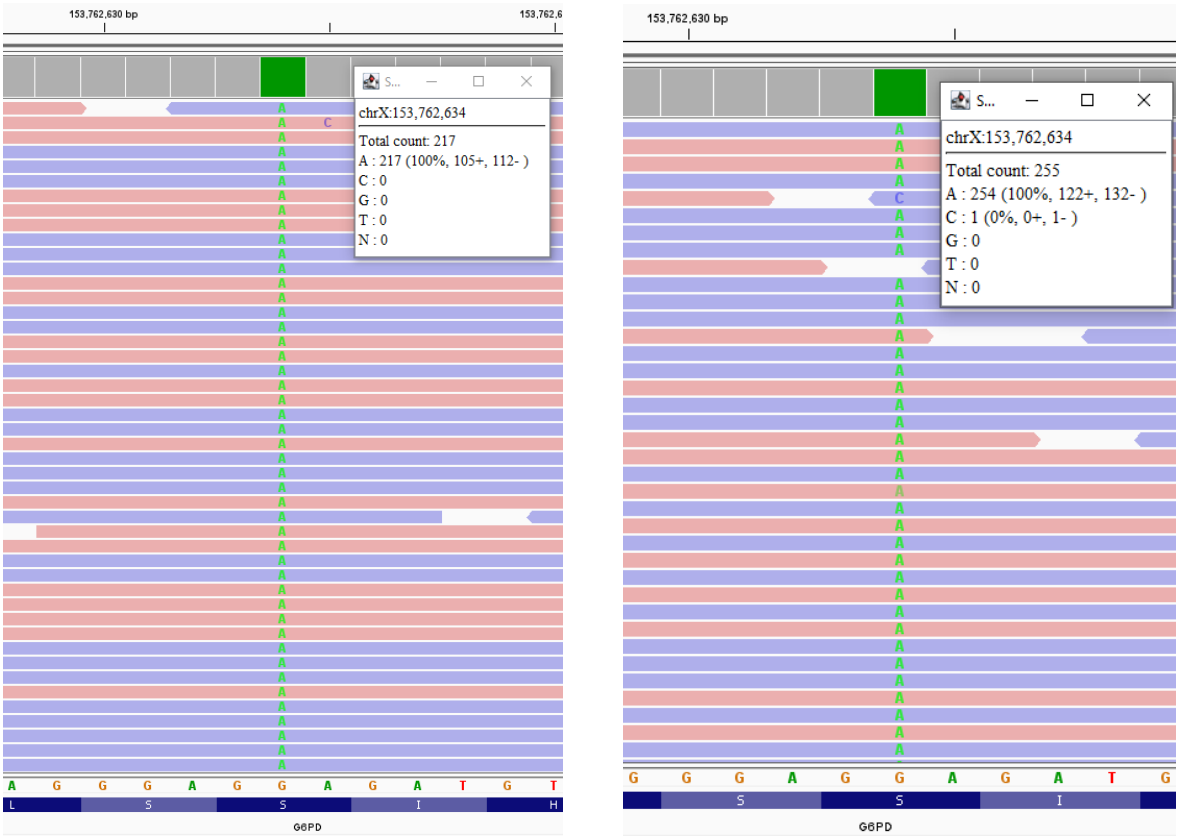

Figure 14 - WES reads at chrX:153,762,634 (GRCh37/hg19) of SK6-1 (right) and SK6-2 (left)
Screenshot taken in IGV2.10.3

3.2.2 Sanger Sequencing

Following the identification of the two candidate genes and the variants, corresponding primers were designed using *Primer3*.[41] No common SNPs have been identified in the primer sequences. The following tables show the forward and reverse primers and their respective product sizes.

CARMIL2 exon 9	**sequence**	**GC-%**	**product size**
Forward primer	CTGAGCCTCTGCCTCTGTTCC	60,00%	227bp
Reverse primer	CCTCAAGGCTCTGAGAAGCA	55,00%	227bp

Table 9 - Forward and reverse primers for *CARMIL2* exon 9

G6PD exon 6	**sequence**	**GC-%**	**product size**
Forward primer	CCAGGTGAGGCTCCTGAGTA	60,00%	538bp
Reverse primer	ATGGCTTGGATGCCTCCTC	57,89%	538bp

Table 10 - Forward and reverse primers for *G6PD* exon 6

Figure 16 shows the results of the Sanger sequencing. The variant in *CARMIL2* segregates as expected, however the variant in *G6PD* does not. The mother is homozygous for the variant and SK6-3 is hemizygous for the variant, despite both of them having stayed asymptomatic so far.

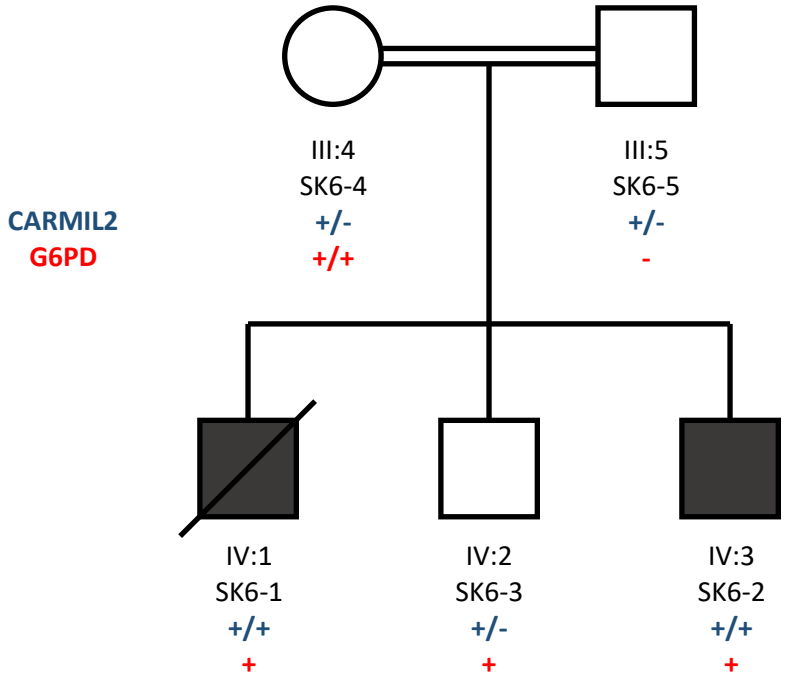

Figure 16 - Pedigree of the individuals examined together with the results of the Sanger sequencing of *CARMIL2* exon 9 and *G6PD* exon6; a + represents a mutated allele and a - represents a wildtype allele

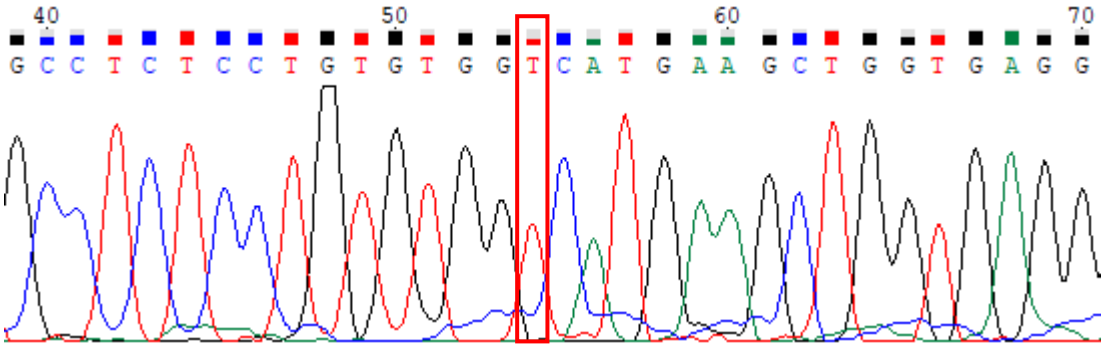

Figure 17 - Electropherogram of *CARMIL2* of SK6-1 (homozygous for the variant) Screenshot taken in Chromas v2.6.6

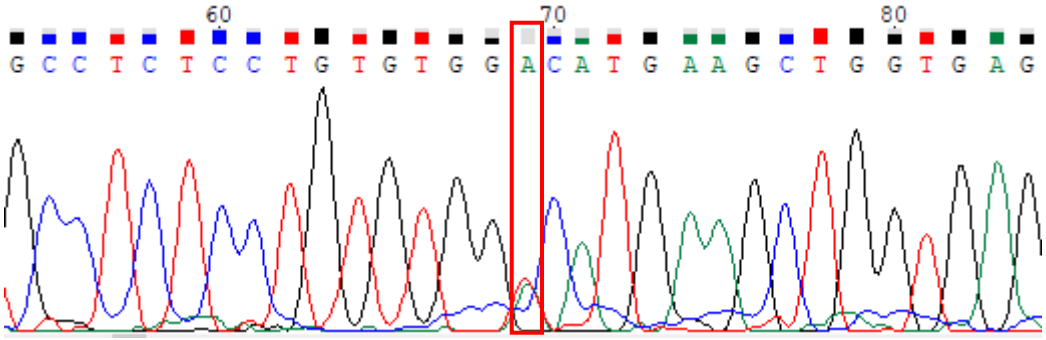

Figure 18 - Electropherogram of *CARMIL2* of SK6-4 (heterozygous for the variant) Screenshot taken in Chromas v2.6.6

Figures 17 and 18 show the electropherogram results for *CARMIL2* of SK6-1 and SK6-4 as examples for a homozygous and a heterozygous genotype, respectively. In SK6-1, the adenine which is normally present on base position 67,681,084 of chromosome 16 has been replaced with a thymine (c.677A>T). In SK6-4, at the same position, adenine and thymine have been sequenced roughly equally as often, which indicates a heterozygous genotype.

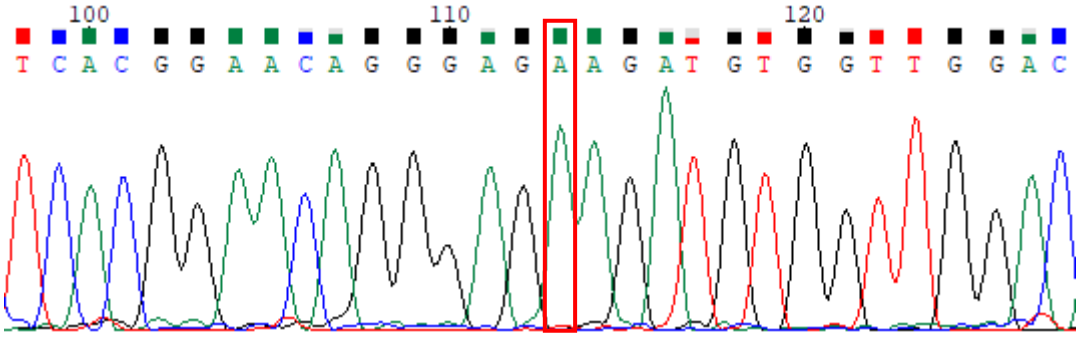

Figure 19 - Electropherogram of *G6PD* of SK6-2 (hemizygous for the variant)
Screenshot taken in Chromas v2.6.6

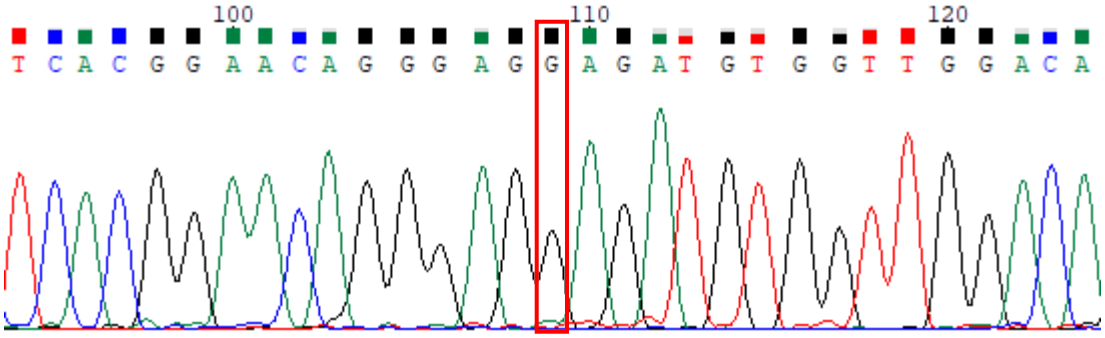

Figure 20 - Electropherogram of *G6PD* of SK6-5 (hemizygous for the wildtype)
Screenshot taken in Chromas v2.6.6

Figures 19 and 20 show the electropherogram results for *G6PD* of SK6-2 and SK6-5 as examples for an allele with the described variant and a wildtype allele. Since both are male individuals, they each only carry one X-chromosome. At position 153,762,634 of the X-chromosome of the wildtype allele a guanine can be found. In SK6-2, an adenine was sequenced at this position instead, meaning that SK6-2 is hemizygous for the variant (chr.X:153,762,634G>A). In SK6-5, at the same position a guanine can be found, meaning hemizygous for the wildtype.

4 Discussion

The two families presented in this thesis were chosen because they both presented phenotypes affecting mainly the skin. This choice was made purely for thematic reasons. During my laboratory work at the Institute of Human Genetics of the Medical University of Graz, I have worked on many cases with consanguinity involved. Phenotypes of the families examined outside the scope of this thesis included ophthalmologic disorders, microcephaly, intellectual disability, and muscular dystrophy. For each of those, a similar approach to the one described here was used. In the majority of cases, a combination of WES, homozygosity mapping and Sanger sequencing allowed for the identification of a causative gene. SNP array or Array-comparative genomic hybridisation could be added to identify larger deletions. Throughout all this work, WES has proven itself as a fast and cost-efficient way of identifying disease-causing mutations. In cases where the first candidate gene turned out to be wrong or no candidate gene was found, the data from WES could be consulted again, filtering steps could be reconsidered and slightly changed. Since the whole exome had been sequenced, all the data necessary for re-evaluation was already available and just needed to be viewed from a different angle. This is one of the many advantages WES possesses over the older forms of identifying candidate genes described in the introduction.

But in cases where no causative mutations can be identified with WES, massively parallel sequencing also allows for whole genome sequencing (WGS). 85% of all disease-causing mutations lie inside the exome,[34] which still leaves 15% in the non-coding genome. For the most part, mutations of the non-coding genome have not been as widely explored as their genomic counterparts, but newer technologies have brought increased attention towards them. Mutations in the non-coding part of the DNA can affect functional elements like promoters, enhancers, or silencers. Such changes can cause gene dysregulation and therefore also lead to pathological phenotypes.[47]

But regardless of whether WES or WGS is used, massively parallel sequencing continues to become more affordable and readily available for even smaller research teams, thanks to steadily decreasing costs associated with it.[21] This will, without a doubt, lead to many more ground-breaking discoveries in the field of genetics.

4.1 SK10

The mutation c.1493C>T of the *DSP* gene was first described in 2015 by Jan A. et al.[48] In their paper, they described a family also originating from the province Khyber Pakhtunkhwa in Pakistan, however to our knowledge relatedness between this family and SK10 could not be established. This family comprised 5 generations with 5 affected individuals, where all the affected individuals, including females, were affected by alopecia totalis. Two twin sisters had some hair growth on the scalp after birth, however they lost that hair after a few months and no more hair growth was observed later. In all individuals, other body hair was also absent and in the male individuals, facial hair growth was greatly reduced. In addition to the alopecia, hyperhidrosis was also observed in this family. This publication is highly interesting, as the main phenotype presented by the family is very similar to that of SK10, however there are some minor differences. In the family described by Jan A. et al., all the individuals were affected by the alopecia regardless of sex, but in SK10 the only female affected individual presents woolly hair instead. Other differences include hyperhidrosis described in the family and no skin or nail changes. It can be speculated that these differences are due to other genes also playing a role in the individual phenotypes. Further research into mutations of the *DSP* gene and this mutation in particular needs to be conducted to be able to answer all questions of varying resulting phenotypes. Cell biological approaches could be used to evaluate the effects this mutation has on desmosome structure in cell culture, compared to the wildtype.

The desmoplakin protein comprises three parts: an amino-terminal domain (N-terminal domain), a carboxy-terminal domain (C-terminal domain) and an alpha-helical coiled-coil rod domain, which is connecting the other two.[14,15] The N-terminal domain of desmoplakin contains a plakin domain, which comprises spectrin-like repeats and an Src-Homology 3 (SH3) domain. A spectrin repeat (SR) is made of three α helices, named A, B, and C. These three helices are arranged in a way to create an antiparallel triple helical structure. SRs are connected by α-helical linkers. The plakin domain of desmoplakin contains the spectrin repeats SR3 through to SR6 as well as SR8 and SR9; SR1, SR2 and SR7 are absent. The SH3 domain is located in between helices B and C of SR5 and is made up of 5 β-strands. While in other proteins the SH3 domain facilitates protein to protein interactions by constituting a binding domain, no binding partners for the SH3 domain of desmoplakin have been identified and its function has not been completely clarified. Hydrophobic parts of the desmoplakin SH3 domain interact with hydrophobic parts of SR4. This interaction

leads to increased stability in the region between SR4 and SR5. This area presents the highest rigidity in the plakin domain. This leads to the conclusion that the SH3 domain contributes to the structural stability of the plakin family.[49]

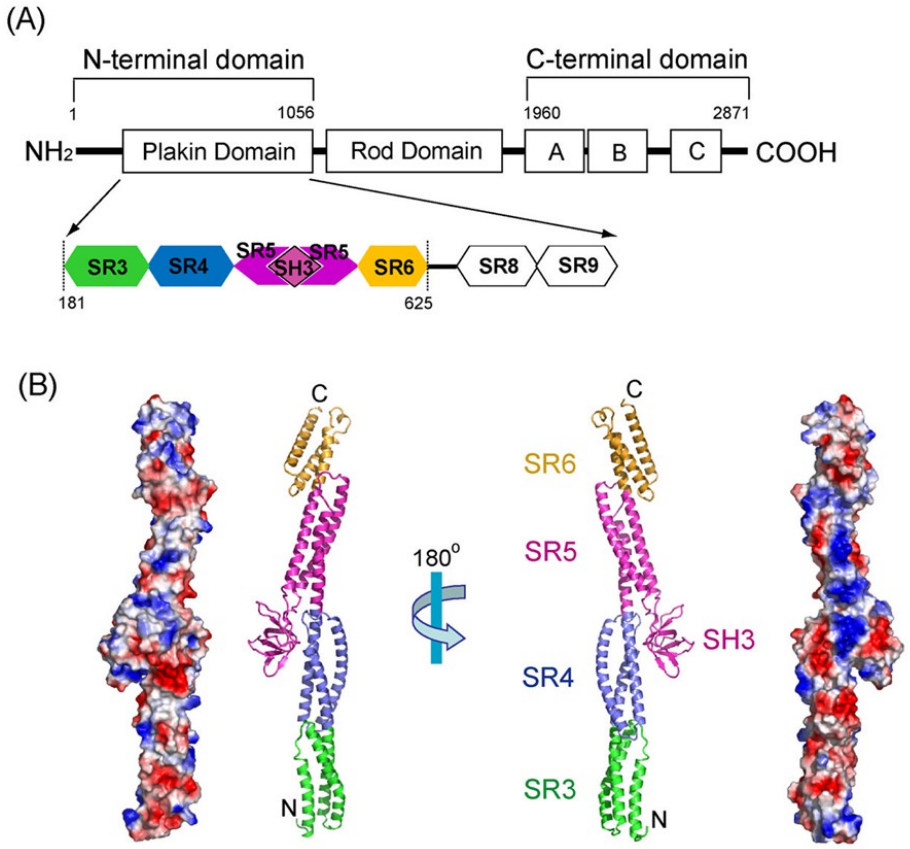

Figure 21 - Structure of desmoplakin protein residues 175–630.
(A) Desmoplakin primary structure. Desmoplakin comprises three main domains. The plakin domain is further enlarged, thus showing the spectrin repeats and the Src-Homology 3 domain.
(B) Ribbon representation and an electrostatic surface (contoured at ±5 kT/e) of desmoplakin (175–630), corresponding to the region between SR3 and SR6.

Reprinted from Journal of Molecular Biology, Volume 409, Issue 5, Hee-Jung Choi, William I. Weis, Crystal structure of a rigid four spectrin repeat fragment of the human desmoplakin plakin domain, 800-812, Copyright 2011 Elsevier, with permission from Elsevier

The mutation c.1493C>T respectively p.Pro498Leu lies in a highly conserved region and falls into this SH3 domain. This provides evidence that a missense mutation in the SH3 domain of desmoplakin might cause structural instabilities and therefore lead to pathologies in tissues where DSP is expressed.

4.2 SK6

The protein CARMIL2 plays a vital role in CD28 costimulation. In the absence of functional CARMIL2 proteins, the differentiation of CD4+ T cells to Th1 and Th17 cells is impaired.[50] The functions of Th17 cells include the repelling of fungal and bacterial infections. They are also associated with inflammatory and autoimmune disorders.[51] The variant c.677A>T of the *CARMIL2* gene found in SK6 has not yet been described in literature. Previously discovered mutations of the *CARMIL2* gene (also called *RLTPR* gene) have caused combined immunodeficiency with very heterogenous phenotypes associated. Combined immunodeficiency describes immunodeficiencies with defective T cell development and/or function with inhibited B cell function, thus affecting both the cellular and humoral immune response. The symptoms of patients affected by mutations in the *CARMIL2* gene differ immensely from case to case. Symptoms described in the literature include mucocutaneous candidiasis, recurrent infections of the upper and lower respiratory tract including pneumonia, asthma, esophagitis, inflammatory bowel disease, failure to thrive, susceptibility to EBV-associated smooth muscle tumours and skin manifestations. These skin manifestations include subcutaneous abscesses (often caused by S. aureus), persistent and often eczematous dermatitis, psoriasis-like lesions, and skin warts.[52-58] Photodermatitis has been described in two publications: Sorte et al.[56] described three patients with UVA-sensitive dermatitis, who suffered from solar urticaria appearing after very brief exposure to sunlight. Shayegan et al.[55] described one patient affected by photodermatitis with erythematous, lichenified skin with superficial fissures and excoriation, which improved upon topical steroids and photoprotection.

The variant c.563C>T of the gene *G6PD* present in SK6 has previously been described in literature. It is the cause of a form of G6PD deficiency, the so-called G6PD Mediterranean. On the DNA-level, this variant was first described by De Vita et al.[59] in 1989. The gene *G6PD* encodes the enzyme glucose-6-phosphate dehydrogenase (G6PD). This enzyme performs a significant function in the pentose phosphate pathway, creating NADPH, which in turn helps in the reduction of reactive oxygen species (ROS). Erythrocytes, which have no nucleus and are incapable of protein synthesis, are completely dependent on these enzymes to combat ROS. It is estimated that worldwide 500 million people are affected by some form of G6PD deficiency. The prevalence varies between practically 0% and over 20%, with high prevalence rates in countries affected by malaria. Because of the high allele frequencies of the variants causing G6PD deficiency, about 50% of these variants are classified as polymorphisms.[60] Looking at 110 patients with falciparum malaria and 100

healthy individuals, Moiz et al.[61] estimated the allele frequency of G6PD Mediterranean to 2.85% in Southern Pakistan. However, all these variants and polymorphisms can only lead to reduced G6PD activity and not absence of G6PD, as this would lead to death in early embryonic stages. Usually, people with G6PD deficiency are completely asymptomatic until they are exposed to an exogenous trigger, in which case they develop acute haemolytic anaemia (AHA). The triggers are infections, certain drugs, and the consumption of fava beans. Possible symptoms associated with AHA are jaundice, paleness, abdominal pain, splenomegaly and dark urine. In new-born, G6PD deficiency can cause neonatal jaundice, which can lead to kernicterus and can necessitate exchange blood transfusion. Since the gene *G6PD* lies on the X-chromosome, usually males are affected by this disease, however heterozygous female individuals can also be affected to varying extents.[60]

One of the main problems regarding G6PD deficiency arises in drug induced AHA, which can be triggered by drugs like the antimalarials Dapsone or Primaquine. This is problematic because G6PD deficiency is especially widespread in areas where malaria is present since these polymorphisms of the *G6PD* gene lead to higher resilience against malaria. However, Primaquine is the only medication currently available that can kill gametocytes and hypnozoites, which might necessitate the administration of Primaquine despite existent G6PD deficiency in certain cases. In these cases, additional caution should be taken.[62]

For the family SK6, this combination of two variants might explain some aspects of the phenotype. The polymorphism in *G6PD* might be a possible reason for the newly acquired thalassemia-like symptoms of SK6-2. The fact that most people with a *G6PD* variant stay asymptomatic explains why SK6-3 (hemizygous for G6PD Mediterranean) and SK6-4 (homozygous for G6PD Mediterranean) have so far never shown any signs of anaemia. Individuals with G6PD Mediterranean should get the possibility of a consultation, informing them about the potential danger of certain medications and the consumption of fava beans. They should also be made aware of the increased risk of neonatal jaundice in children with the polymorphism. A genetic counselling for all the individuals explaining the mode of inheritance and the consequences of this genetic variant can also prove beneficial.

On the other hand, the variant in *CARMIL2* has not been described in literature. Nevertheless, photosensitivity, abscesses and allergies have all been described in conjunction with *CARMIL2* mutations, even if these mutations have usually led to more

severe phenotypes. For this reason, it cannot be ruled out that the variant c.677A>T is the cause of at least some of the symptoms present in SK6. Further research needs to be conducted on this variant, including cell biological studies on the effects this variant has on the resulting protein, compared to the wildtype.

In the end, no definitive answers which could fully explain the observed phenotype of SK6 have been found yet. Photophobia, intolerance to high temperatures and hyperhidrosis specifically have not been described in conjunction with *CARMIL2* mutations and probably have a different cause. For further analysis, the recruitment of the affected brothers of SK6-4 would be the logical next step. Unfortunately, so far, these three individuals have refused to participate in this study and all the information available about them is that they present a similar phenotype to that of SK6-2. Ideally, they could be recruited in the future, undergo thorough physical examination, and participate in the genetic investigation. In this case the number of homozygous regions shared by all affected individuals could most likely be narrowed down to one or two through a SNP array of all individuals or WES of the newly recruited brothers of SK6-4. Additionally, if the variant in *CARMIL2* segregated in these individuals, it would strengthen the hypothesis of the potential pathogenic effect of this variant.

4.3 The current state of NGS and the future of Sequencing technologies

As pointed out above, massively parallel sequencing/NGS has proven itself a massive boon for the field of genetics. However, despite the impressive technology behind NGS, it is still far from perfect. The main inherent problem associated with NGS lies in the relatively short reads it produces, which are usually <300bp long. Many individual short reads are assembled into an exome or genome. While this works very well for SNVs and small indels within the exome, it can lead to misassembly and gaps within longer repeated sequences (which can often be longer than individual reads). For the same reason, the detection of larger structural variants and the correct characterization of transcript isoforms also pose a challenge with NGS. Another limitation lies in the reliance on PCR, as PCR is inefficient at amplifying regions with very low or very high GC%, which leads to a bias in the short reads produced. Long-Read Sequencing/Third Generation Sequencing (TGS) promises to resolve many, if not all, of the problems NGS is facing.[36]

There are two main Long-Read Sequencing/TGS technologies in use. The first is called single-molecule real-time (SMRT) sequencing. In SMRT, hairpin adaptors are ligated to both ends of a double strand DNA molecule to create a circular single strand DNA. The sequencing of this circular single strand DNA is performed using the sequencing by synthesis technique. Fluorescently labelled nucleotides emit a light signal upon integration, which is recorded by a camera in real time. The second technology is called nanopore sequencing. In nanopore sequencing, a nanopore is integrated into a membrane, which separates two compartments filled with an electrolyte solution. When voltage is applied across the membrane, the movement of electrolytes through the nanopore creates a current flow. At the same time a single DNA strand is squeezed through the nanopore. As the DNA strand moves through the nanopore channel, it causes changes in the current, which can be assigned to individual nucleobases and therefore used for sequencing. The advantages these two methods hold compared to traditional NGS include longer read lengths (upwards of 10kb; >1Mb in nanopore sequencing), and the identification of epigenetic modifications like methylation. Both technologies are also independent of PCR. The main disadvantage of TGS lies in its high error rate, even though this is constantly declining as the technologies are improving and error correction techniques evolving.[36,63]

The combination of short reads from NGS and long reads from TGS can be used to

leverage the strengths of both technologies and thereby achieve more precise and reliable results.[63]

Another technology for obtaining long reads is called synthetic long reads (SLR), which is based on traditional NGS and relies on PCR. In this technology, genomic DNA is first split into long fragments of 10kb-100kb length and then further split into the small fragments used for NGS. However, each of these small fragments contains a "barcode", such that it can be retraced to the initial long fragment. Unlike SMRT and nanopore sequencing, SLR is not considered a TGS technology and cannot be used for the identification of epigenetic modifications.[36]

For the sequencing of exomes, NGS is a well refined and reliable tool that has revolutionized the field of genetics and remains highly valuable. However, it comes with its share of inherent limitations, most of which result from the short read lengths and the dependency on PCR technology. TGS is better equipped for resolving these issues, but it also has very relevant disadvan-tages like high error rates. In the past years, constant improvement of TGS technologies has helped bring these error rates down, and further reduction can be achieved through the incorporation of error correction techniques. In the long run, TGS technologies will probably replace NGS completely, but currently they are not perfected yet. As a matter of fact, currently the best results can be achieved by using a combination of TGS and NGS. For this reason, NGS is still a long way from becoming obsolete. Until TGS technologies don't manage to achieve a similar accuracy as NGS, the two technologies should be seen as complementary, rather than competing.

4.4 Gene therapy

In gene therapy, genetic material (DNA or RNA) is inserted into cells of patients with the intent of treating diseases. This can be achieved by one of two main ways: gene addition or genome editing. In gene addition, a vector is used to transfer a functioning copy of a specific gene into a cell with a defective copy of that gene. The functioning gene copy can take over the function of the defective gene. Usually, integrating vectors are used when transferring genes to stem cells and nonintegrating vectors are used for postmitotic or slowly dividing cells.[64] A modified version of this technique aims at transferring small inhibitory RNAs into the cell instead of genes. These small inhibitory RNAs inhibit the expression of the defective gene through RNA interference. In genome editing, the defective gene is removed or repaired. In addition to the different types of gene therapy,

there are also different ways of delivering it. *In vivo* gene delivery relies on local or systemic administration of a vector-gene construct. The most commonly used vectors for *in vivo* gene therapy are adeno-associated viruses (AAV). Unlike retrovirus vectors, AAV vectors are non-integrating. In *ex vivo* gene delivery, patient cells are removed from the patient's body, genetically modified (usually using retroviruses as vectors), cultured, and then readministered as autologous transplants. Retroviral vectors are integrating, meaning that the genetic information of the vector is integrated into the host's genome. If the site of integration is not specified in the vector, it carries the risk of insertional mutagenesis if the integration takes place within a functional element of the DNA.[64,65]

While initially gene therapy was envisioned to be used for monogenic inherited diseases, scientists quickly noticed its massive potential, and its application was extended to acquired diseases. Up until 2020, over 20 different gene therapies had been officially approved by regulating authorities for clinical use. These include treatments for melanoma, severe combined immunodeficiency, spinal muscular atrophy, large B-cell lymphoma, or retinal dystrophy, among others. For most of these, gene therapy provides significantly better results than conventional therapies, or there even is no alternative treatment to gene therapy (e.g., spinal muscular atrophy). One of the main issues with the gene therapies currently available lies in the high prices, as a single treatment can cost hundreds of thousands of US-$.[65]

The skin presents a particularly attractive organ for gene therapy due to its easy accessibility. For *ex vivo* skin therapy, a patient's skin stem cells can easily be attained through simple skin biopsy. These skin stem cells are modified in a laboratory, where scientists can verify whether the genetic modification was successful, and then reapplied to the patient.[66] Gene therapy for malignant melanomas has been approved for clinical use in the European Union and the USA.[65] Multiple studies have used gene therapy to great effect in epidermolysis bullosa, a monogenic disease with AR or AD mode of inheritance, causing painful blistering of the skin and mucous membranes. While these studies used the more conservative gene addition approach, in the future genome editing procedures using nucleases might be the tool of choice for gene therapy. Unlike traditional gene therapies, which only add a functioning copy of a gene to the cell, these procedures can correct the defective gene copy inside cells. Laboratory studies using these procedures have shown promising results for monogenic skin disorders like epidermolysis bullosa and epidermolytic ichthyosis. Currently, CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR-associated genes) technology is the most promising for

genome editing.[66] This technology is based on the Cas9 nuclease of prokaryotic organisms, where it plays a role in adaptive immunity by cutting the nucleic acid sequences of invading bacteriophage viruses, thereby destroying them.[67] CRISPR-Cas9 uses a guide RNA (gRNA) to find the complementary DNA sequence in a genome. Once in position, the nuclease performs a double strand DNA break (DSB) at specific target sites. Inside the cell, the repair of DSBs can occur in two ways. Non-homologous end-joining leads to short, random indels at the DSB site, which cause gene disruption. In contrast, homology-directed repair uses a homologous DNA template to insert precise insertions at the DSB site, which can be used for correcting defective genes.[67,68]

In conclusion, unfortunately no causal treatments exist currently for the mutations of the families described in this paper. Gene therapy is still a relatively new approach of treating diseases and only a very limited number of medications have been officially approved for clinical use. Nevertheless, there is an enormous amount of ongoing research into gene therapy. Monogenic skin diseases specifically are very interesting to researchers, as stem cells of the skin can be obtained relatively easily, and the application of modified cells is also easier than in other organ systems. There remains hope that in the future treatments for the mutations of these two families will become available and affordable.

But despite the rapid advancements made in the field of gene therapy, many ethical aspects are still being discussed. This is especially true for germline therapy, where gametes or preimplantation embryos are genetically modified. Germline therapy allows for the elimination of any specific genetic disease not just from the DNA of an individual, but also all the descendants of that individual.[65] In a theoretical future, in which germline gene therapy has been perfected, it could be used to achieve a society in which genetic diseases can be identified and treated long before the birth of an individual. While such prospects sound appealing, misuse of this technology might become dangerously easy, too. Most people would probably agree that skin diseases like epidermolysis bullosa present serious medical conditions and that gene therapy should be embraced to treat them. Similar arguments could be brought up for the treatment of monogenic hair loss disorders. While hereditary hair loss disorders usually don’t cause immediate medical problems, they can impact the well-being of the individual, lead to social marginalization and thus negatively affect mental health. But is the same true of androgenetic alopecia, which also has genetic components? Should every genetic trait that can potentially lead to diseases or unwanted phenotypical traits be removed from society? In this theoretical future, the same technology that is used in germline therapy could also be used by parents to modify the

phenotype of their children, e.g., choose their hair colour, skin pigmentation, etc. Babies would be designed the same way as houses. Couples wishing to have children would experience social pressure to utilize germline therapy, regardless of whether they approve of it or not. Authoritarian regimes could identify certain “desirable” traits for their citizens und use genome editing to create a society like in Huxley’s *Brave New World*. Gene therapy’s potential for curing diseases is seemingly endless, but the potential for misuse must not be ignored. The scenarios painted here might never come to happen and even if, it will take decades or centuries. Nevertheless, we have to be aware of future consequences that current scientific advances might have and develop laws to ensure the positive potential can be realised and the negative warded off. For this reason, some form of regulation for gene therapy seems indispensable. It is up to society to define where we draw the line.

5 Literature

1. Moll I. Duale Reihe Dermatologie. 8th edition. Stuttgart: Thieme; 2016

2. Yousef H, Alhajj M, Sharma S. Anatomy, Skin (Integument), Epidermis. 2021 Jul 26. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 29262154.

3. Goebler M, Henning H. Basiswissen Dermatologie. Berlin: Springer; 2017

4. Eckhart L, Lippens S, Tschachler E, Declercq W. Cell death by cornification. Biochim Biophys Acta. 2013 Dec;1833(12):3471-3480. doi: 10.1016/j.bbamcr.2013.06.010. Epub 2013 Jun 20. PMID: 23792051.

5. Brown TM, Krishnamurthy K. Histology, Hair and Follicle. 2021 May 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 30422524.

6. Hoover E, Alhajj M, Flores JL. Physiology, Hair. 2021 Jul 26. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 29763123.

7. Nalluri R, Harries M. Alopecia in general medicine. Clin Med (Lond). 2016 Feb;16(1):74-8. doi: 10.7861/clinmedicine.16-1-74. PMID: 26833522; PMCID: PMC4954340.

8. Lolli F, Pallotti F, Rossi A, Fortuna MC, Caro G, Lenzi A, Sansone A, Lombardo F. Androgenetic alopecia: a review. Endocrine. 2017 Jul;57(1):9-17. doi: 10.1007/s12020-017-1280-y. Epub 2017 Mar 28. PMID: 28349362.

9. Strazzulla LC, Wang EHC, Avila L, Lo Sicco K, Brinster N, Christiano AM, Shapiro J. Alopecia areata: Disease characteristics, clinical evaluation, and new perspectives on pathogenesis. J Am Acad Dermatol. 2018 Jan;78(1):1-12. doi: 10.1016/j.jaad.2017.04.1141. PMID: 29241771.

10. Shimomura Y. Congenital hair loss disorders: rare, but not too rare. J Dermatol. 2012 Jan;39(1):3-10. doi: 10.1111/j.1346-8138.2011.01395.x. Epub 2011 Nov 2. PMID: 22044263.

11. Basit S, Khan S, Ahmad W. Genetics of human isolated hereditary hair loss disorders. Clin Genet. 2015 Sep;88(3):203-12. doi: 10.1111/cge.12531. Epub 2014 Nov 22. PMID: 25350920.

12. Hayashi R, Shimomura Y. Update of recent findings in genetic hair disorders. J Dermatol. 2021 Oct 21. doi: 10.1111/1346-8138.16204. Epub ahead of print. PMID: 34676598.

13. Choi HJ, Weis WI. Purification and Structural Analysis of Desmoplakin. Methods Enzymol. 2016;569:197-213. doi: 10.1016/bs.mie.2015.05.006. Epub 2015 May 30. PMID: 26778560.

14. Kowalczyk AP, Green KJ. Structure, function, and regulation of desmosomes. Prog Mol Biol Transl Sci. 2013;116:95-118. doi: 10.1016/B978-0-12-394311-8.00005-4. PMID: 23481192; PMCID: PMC4336551.

15. Delva E, Tucker DK, Kowalczyk AP. The desmosome. Cold Spring Harb Perspect Biol. 2009 Aug;1(2):a002543. doi: 10.1101/cshperspect.a002543. PMID: 20066089; PMCID: PMC2742091.

16. Cabral RM, Wan H, Cole CL, Abrams DJ, Kelsell DP, South AP. Identification and characterization of DSPIa, a novel isoform of human desmoplakin. Cell Tissue Res. 2010 Jul;341(1):121-9. doi: 10.1007/s00441-010-0989-1. Epub 2010 Jun 4. PMID: 20524011; PMCID: PMC2896628.

17. Lai Cheong JE, Wessagowit V, McGrath JA. Molecular abnormalities of the desmosomal protein desmoplakin in human disease. Clin Exp Dermatol. 2005 May;30(3):261-6. doi: 10.1111/j.1365-2230.2005.01736.x. PMID: 15807686.

18. Al-Owain M, Wakil S, Shareef F, Al-Fatani A, Hamadah E, Haider M, Al-Hindi H, Awaji A, Khalifa O, Baz B, Ramadhan R, Meyer B. Novel homozygous mutation in DSP causing skin fragility-woolly hair syndrome: report of a large family and review of the desmoplakin-related phenotypes. Clin Genet. 2011 Jul;80(1):50-8. doi: 10.1111/j.1399-0004.2010.01518.x. Epub 2010 Jul 22. PMID: 20738328.

19. Gallicano GI, Kouklis P, Bauer C, Yin M, Vasioukhin V, Degenstein L, Fuchs E. Desmoplakin is required early in development for assembly of desmosomes and cytoskeletal linkage. J Cell Biol. 1998 Dec 28;143(7):2009-22. doi: 10.1083/jcb.143.7.2009. PMID: 9864371; PMCID: PMC2175222.

20. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012 May;20(5):490-7. doi: 10.1038/ejhg.2011.258. Epub 2012 Jan 18. PMID: 22258526; PMCID: PMC3330229.

21. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata. Accessed [15.10.2021].

22. Bittles A. Consanguinity and its relevance to clinical genetics. Clin Genet. 2001 Aug;60(2):89-98. doi: 10.1034/j.1399-0004.2001.600201.x. PMID: 11553039.

23. Oniya O, Neves K, Ahmed B, Konje JC. A review of the reproductive consequences of consanguinity. Eur J Obstet Gynecol Reprod Biol. 2019 Jan;232:87-96. doi: 10.1016/j.ejogrb.2018.10.042. Epub 2018 Nov 1. PMID: 30502592.

24. National Institute of Population Studies (NIPS) [Pakistan] and ICF. 2019. 2017-18 Pakistan Demographic and Health Survey Key Findings. Islamabad, Pakistan, and Rockville, Maryland, USA: NIPS and ICF

25. Bennett RL, Motulsky AG, Bittles A, Hudgins L, Uhrich S, Doyle DL, Silvey K, Scott CR, Cheng E, McGillivray B, Steiner RD, Olson D. Genetic Counseling and Screening of Consanguineous Couples and Their Offspring: Recommendations of the National Society of Genetic Counselors. J Genet Couns. 2002 Apr;11(2):97-119. doi: 10.1023/A:1014593404915. PMID: 26141656.

26. Bennett RL, Malleda NR, Byers PH, Steiner RD, Barr KM. Genetic counseling and screening of consanguineous couples and their offspring practice resource: Focused Revision. J Genet Couns. 2021 Jul 26. doi: 10.1002/jgc4.1477. Epub ahead of print. PMID: 34309119.

27. Fridman H, Yntema HG, Mägi R, Andreson R, Metspalu A, Mezzavila M, Tyler-Smith C, Xue Y, Carmi S, Levy-Lahad E, Gilissen C, Brunner HG. The landscape of autosomal-recessive pathogenic variants in European populations reveals phenotype-specific effects. Am J Hum Genet. 2021 Apr 1;108(4):608-619. doi: 10.1016/j.ajhg.2021.03.004. Epub 2021 Mar 18. PMID: 33740458; PMCID: PMC8059335.

28. Alkuraya FS. Homozygosity mapping: one more tool in the clinical geneticist's toolbox. Genet Med. 2010 Apr;12(4):236-9. doi: 10.1097/GIM.0b013e3181ceb95d. PMID: 20134328.

29. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet. 2012 Aug 10;91(2):275-92. doi: 10.1016/j.ajhg.2012.06.014. PMID: 22883143; PMCID: PMC3415543.

30. Belkadi A, Pedergnana V, Cobat A, Itan Y, Vincent QB, Abhyankar A, Shang L, El Baghdadi J, Bousfiha A; Exome/Array Consortium, Alcais A, Boisson B, Casanova JL, Abel L. Whole-exome sequencing to analyze population structure, parental inbreeding, and familial linkage. Proc Natl Acad Sci U S A. 2016 Jun 14;113(24):6713-8. doi: 10.1073/pnas.1606460113. Epub 2016 May 31. PMID: 27247391; PMCID: PMC4914194.

31. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7. doi: 10.1073/pnas.74.12.5463. PMID: 271968; PMCID: PMC431765.

32. McGinn S, Gut IG. DNA sequencing - spanning the generations. N Biotechnol. 2013 May 25;30(4):366-72. doi: 10.1016/j.nbt.2012.11.012. Epub 2012 Nov 16. PMID: 23165096.

33. Beck TF, Mullikin JC; NISC Comparative Sequencing Program, Biesecker LG. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. Clin Chem. 2016 Apr;62(4):647-54. doi: 10.1373/clinchem.2015.249623. Epub 2016 Feb 4. PMID: 26847218; PMCID: PMC4878677.

34. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet. 2011 Sep;48(9):580-9. doi: 10.1136/jmedgenet-2011-100223. Epub 2011 Jul 5. PMID: 21730106.

35. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013 Dec;98(6):236-8. doi: 10.1136/archdischild-2013-304340. Epub 2013 Aug 28. PMID: 23986538; PMCID: PMC3841808.

36. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. Trends Genet. 2018 Sep;34(9):666-681. doi: 10.1016/j.tig.2018.05.008. Epub 2018 Jun 22. PMID: 29941292.

37. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013 Mar;14(2):178-92. doi: 10.1093/bib/bbs017. Epub 2012 Apr 19. PMID: 22517427; PMCID: PMC3603213.

38. Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. HomozygosityMapper--an interactive approach to homozygosity mapping. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W593-9. doi: 10.1093/nar/gkp369. Epub 2009 May 21. PMID: 19465395; PMCID: PMC2703915.

39. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006. doi: 10.1101/gr.229102. PMID: 12045153; PMCID: PMC186604.

40. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, Powell CC, Nassar LR, Maulding ND, Lee CM, Lee BT, Hinrichs AS, Fyfe AC, Fernandes JD, Diekhans M, Clawson H, Casper J, Benet-Pagès A, Barber GP, Haussler D, Kuhn RM, Haeussler M, Kent WJ. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res. 2021 Jan 8;49(D1):D1046-D1057. doi: 10.1093/nar/gkaa1070. PMID: 33221922; PMCID: PMC7779060.

41. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012 Aug;40(15):e115. doi: 10.1093/nar/gks596. Epub 2012 Jun 22. PMID: 22730293; PMCID: PMC3424584.

42. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002 Apr;12(4):656-64. doi: 10.1101/gr.229202. PMID: 11932250; PMCID: PMC187518.

43. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D514-7. doi: 10.1093/nar/gki033. PMID: 15608251; PMCID: PMC539987.

44. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R,

Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME; Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May;581(7809):434-443. doi: 10.1038/s41586-020-2308-7. Epub 2020 May 27. Erratum in: Nature. 2021 Feb;590(7846):E53. PMID: 32461654; PMCID: PMC7334197.

45. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014 Apr;11(4):361-2. doi: 10.1038/nmeth.2890. PMID: 24681721.

46. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL, Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL, Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C, Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC, Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Seidel MG, Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yüksel Z, Helbig I, Mungall CJ, Haendel MA, Robinson PN. The Human Phenotype Ontology in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D1207-D1217. doi: 10.1093/nar/gkaa1043. PMID: 33264411; PMCID: PMC7778952.

47. Scacheri CA, Scacheri PC. Mutations in the noncoding genome. Curr Opin Pediatr. 2015 Dec;27(6):659-64. doi: 10.1097/MOP.0000000000000283. PMID: 26382709; PMCID: PMC5084913.

48. Jan A, Basit S, Wakil SM, Ramzan K, Ahmad W. A novel homozygous variant in the dsp gene underlies the first case of non-syndromic form of alopecia. Arch Dermatol Res. 2015 Nov;307(9):793-801. doi: 10.1007/s00403-015-1590-y. Epub 2015 Jul 7. PMID: 26148547.

49. Choi HJ, Weis WI. Crystal structure of a rigid four-spectrin-repeat fragment of the human desmoplakin plakin domain. J Mol Biol. 2011 Jun 24;409(5):800-12. doi: 10.1016/j.jmb.2011.04.046. Epub 2011 Apr 22. PMID: 21536047; PMCID: PMC3107870.

50. Roncagalli R, Cucchetti M, Jarmuzynski N, Grégoire C, Bergot E, Audebert S, Baudelet E, Menoita MG, Joachim A, Durand S, Suchanek M, Fiore F, Zhang L, Liang Y, Camoin L, Malissen M, Malissen B. The scaffolding function of the RLTPR protein explains its essential role for CD28 co-stimulation in mouse and human T cells. J Exp Med. 2016 Oct 17;213(11):2437-2457. doi: 10.1084/jem.20160579. Epub 2016 Sep 19. PMID: 27647348; PMCID: PMC5068240.

51. Singh B, Schwartz JA, Sandrock C, Bellemore SM, Nikoopour E. Modulation of autoimmune diseases by interleukin (IL)-17 producing regulatory T helper (Th17) cells. Indian J Med Res. 2013 Nov;138(5):591-4. PMID: 24434314; PMCID: PMC3928692.

52. Alazami AM, Al-Helale M, Alhissi S, Al-Saud B, Alajlan H, Monies D, Shah Z, Abouelhoda M, Arnaout R, Al-Dhekri H, Al-Numair NS, Ghebeh H, Sheikh F, Al-Mousa H. Novel CARMIL2 Mutations in Patients with Variable Clinical Dermatitis, Infections, and Combined Immunodeficiency. Front Immunol. 2018 Feb 9;9:203. doi: 10.3389/fimmu.2018.00203. PMID: 29479355; PMCID: PMC5811477.

53. Marangi G, Garcovich S, Sante GD, Orteschi D, Frangella S, Scaldaferri F, Genuardi M, Peris K, Gurrieri F, Zollino M. Complex Muco-cutaneous Manifestations of CARMIL2-associated Combined Immunodeficiency: A Novel Presentation of Dysfunctional Epithelial Barriers. Acta Derm Venereol. 2020 Jan 23;100(1):adv00038. doi: 10.2340/00015555-3370. PMID: 31709449.

54. Schober T, Magg T, Laschinger M, Rohlfs M, Linhares ND, Puchalka J, Weisser T, Fehlner K, Mautner J, Walz C, Hussein K, Jaeger G, Kammer B, Schmid I, Bahia M, Pena SD, Behrends U, Belohradsky BH, Klein C, Hauck F. A human immunodeficiency syndrome caused by mutations in CARMIL2. Nat Commun. 2017 Jan 23;8:14209. doi: 10.1038/ncomms14209. PMID: 28112205; PMCID: PMC5473639.

55. Shayegan LH, Garzon MC, Morel KD, Borlack R, Vuguin PM, Margolis KG, Demirdag YY, Pereira EM, Lauren CT. CARMIL2-related immunodeficiency manifesting with photosensitivity. Pediatr Dermatol. 2020 Jul;37(4):695-697. doi: 10.1111/pde.14173. Epub 2020 Apr 27. PMID: 32342551; PMCID: PMC7599087.

56. Sorte HS, Osnes LT, Fevang B, Aukrust P, Erichsen HC, Backe PH, Abrahamsen TG, Kittang OB, Øverland T, Jhangiani SN, Muzny DM, Vigeland MD, Samarakoon P, Gambin T, Akdemir ZH, Gibbs RA, Rødningen OK, Lyle R, Lupski JR, Stray-Pedersen A. A potential founder variant in CARMIL2/RLTPR in three Norwegian families with warts, molluscum contagiosum, and T-cell dysfunction. Mol Genet Genomic Med. 2016 Sep 17;4(6):604-616. doi: 10.1002/mgg3.237. PMID: 27896283; PMCID: PMC5118205.

57. Wang Y, Ma CS, Ling Y, Bousfiha A, Camcioglu Y, Jacquot S, Payne K, Crestani E, Roncagalli R, Belkadi A, Kerner G, Lorenzo L, Deswarte C, Chrabieh M, Patin E, Vincent QB, Müller-Fleckenstein I, Fleckenstein B, Ailal F, Quintana-Murci L, Fraitag S, Alyanakian MA, Leruez-Ville M, Picard C, Puel A, Bustamante J, Boisson-Dupuis S, Malissen M, Malissen B, Abel L, Hovnanian A, Notarangelo LD, Jouanguy E, Tangye SG, Béziat V, Casanova JL. Dual T cell- and B cell-intrinsic deficiency in humans with biallelic RLTPR mutations. J Exp Med. 2016 Oct 17;213(11):2413-2435. doi: 10.1084/jem.20160576. Epub 2016 Sep 19. PMID: 27647349; PMCID: PMC5068239.

58. Yonkof JR, Gupta A, Rueda CM, Mangray S, Prince BT, Rangarajan HG, Alshahrani M, Varga E, Cripe TP, Abraham RS. A Novel Pathogenic Variant in CARMIL2 (RLTPR) Causing CARMIL2 Deficiency and EBV-Associated Smooth Muscle Tumors. Front Immunol. 2020 Jun 18;11:884. doi: 10.3389/fimmu.2020.00884. PMID: 32625199; PMCID: PMC7314954.

59. De Vita G, Alcalay M, Sampietro M, Cappelini MD, Fiorelli G, Toniolo D. Two point mutations are responsible for G6PD polymorphism in Sardinia. Am J Hum Genet. 1989 Feb;44(2):233-40. PMID: 2912069; PMCID: PMC1715414.

60. Luzzatto L, Ally M, Notaro R. Glucose-6-phosphate dehydrogenase deficiency. Blood. 2020 Sep 10;136(11):1225-1240. doi: 10.1182/blood.2019000944. PMID: 32702756.

61. Moiz B, Arshad HM, Raheem A, Hayat H, Karim Ghanchi N, Beg MA. Frequency of G6PD Mediterranean in individuals with and without malaria in Southern Pakistan. Malar J. 2017 Oct 24;16(1):426. doi: 10.1186/s12936-017-2069-4. PMID: 29065882; PMCID: PMC5655902.

62. Luzzatto L, Seneca E. G6PD deficiency: a classic example of pharmacogenetics with on-going clinical implications. Br J Haematol. 2014 Feb;164(4):469-80. doi: 10.1111/bjh.12665. Epub 2013 Dec 28. PMID: 24372186; PMCID: PMC4153881.

63. Midha MK, Wu M, Chiu KP. Long-read sequencing in deciphering human genetics to a greater depth. Hum Genet. 2019 Dec;138(11-12):1201-1215. doi: 10.1007/s00439-019-02064-y. Epub 2019 Sep 19. PMID: 31538236.

64. High KA, Roncarolo MG. Gene Therapy. N Engl J Med. 2019 Aug 1;381(5):455-464. doi: 10.1056/NEJMra1706910. PMID: 31365802.

65. Tang R, Xu Z. Gene therapy: a double-edged sword with great powers. Mol Cell Biochem. 2020 Nov;474(1-2):73-81. doi: 10.1007/s11010-020-03834-3. Epub 2020 Jul 21. PMID: 32696132.

66. Koller U. Ex-vivo-Stammzellgentherapie an der Haut : Reif für klinische Anwendungen? [Ex vivo stem cell gene therapy of the skin : Ready for clinical use?]. Hautarzt. 2020 Feb;71(2):85-90. German. doi: 10.1007/s00105-019-04529-7. PMID: 31965203.

67. Barman A, Deb B, Chakraborty S. A glance at genome editing with CRISPR-Cas9 technology. Curr Genet. 2020 Jun;66(3):447-462. doi: 10.1007/s00294-019-01040-3. Epub 2019 Nov 5. PMID: 31691023.

68. Zhang JH, Adikaram P, Pandey M, Genis A, Simonds WF. Optimization of genome editing through CRISPR-Cas9 engineering. Bioengineered. 2016 Apr;7(3):166-74. doi: 10.1080/21655979.2016.1189039. PMID: 27340770; PMCID: PMC4927198.