

https://dspace.org/

The identification and characterization of the causative gene for Keratolytic winter erythema in South African families

Hobbs, Angela Wendy

2010-09-17
http://hdl.handle.net/10539/8734

Downloaded from a DSpace repository. https://dspace.org/

DECLARATION

I, Angela Wendy Hobbs, declare that the work contained in this thesis is my own original work. It is being submitted for the degree of Master of Science in Medicine at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

Signature:……. Date:…….

“To accomplish great things, we must not only act, but also dream; not only plan, but also believe.”

Anatole France

In loving memory of my father

Wayne Dennis Hobbs

1955 – 2001

“Always be and do the best that you can. Always remember I love you”

DAD

ABSTRACT

Keratolytic winter erythema (KWE) or Oudtshoorn skin disease is a rare monogenic autosomal dominant condition with an unknown cause. KWE is a disorder of epidermal keratinisation that involves the necrobiosis of the Malpigihian layer of the palmoplantar skin with the consequent dissection of the stratum corneum. This cutaneous disorder was first described by Findlay et al. (1977) and occurs with a high prevalence of 1/7200 in the South African Afrikaans-speaking white population and with a lower, but unspecified prevalence in the Coloured population. The primary objective of this study was to identify and characterise the causative gene for KWE, by examining plausible positional candidate genes. The KWE gene has been localized to chromosome 8p23.1-p22 in a region of 1.2 Mb.

In order to confirm and possibly reduce the size of the KWE critical region, haplotype analysis was done by genotyping 9 microsatellite markers positioned along chromosome 8p23.1–p22 between markers D8S1819 and D8S552 in four KWE families. The KWE critical region was validated and the positional candidate genes located within this region were prioritized according to their biological function. The most likely positional candidate genes were *CTSB*, *FDFT1*, *DUB-3* and the open reading frame, *C8orf49*.

In order to identify a potentially causative KWE mutation, the coding regions of each candidate gene was sequenced from genomic DNA. Each of the genetic variants identified was also observed in the control group or had previously been shown to be polymorphic, eliminating them all from causing KWE. The cDNA of the two most likely candidates, *FDFT1* and *CTSB*, was sequenced in order to identify deep intronic variants that might affect splicing and that would not be identified at a genomic DNA level. No such variation was observed. The relative expression profiles of *CTSB* and *FDFT1* in affected and non-affected palmoplantar skin was analysed using real-time RT-PCR. The relative expression of *CTSB* in the skin of patients did not differ significantly from controls ($p=0.68$). However, a trend was observed towards increased expression of *FDFT1* in the skin of KWE affected individuals ($p=0.063$). This observation prompted the analysis of the *FDFT1* promoter region through genomic sequencing. No genetic variants identified within the promoter region segregated with the KWE phenotype. The increased *FDFT1* expression is therefore unlikely to result from a mutation within the promoter region of this gene and may be in response to the disruption of the epidermal barrier in affected skin. There is a strong correlation between the severity of the KWE phenotype and the level of *FDFT1* expression.

Although none of the chosen positional candidate genes appear to harbour the KWE-causing mutation, they can be excluded from the list of possible positional candidates for KWE, taking us one step closer to discovering the molecular cause of KWE.

ACKNOWLEDGEMENTS

I would like to thank God for the guidance and strength that he has given me over the last two years. Without his constant presence, completing my thesis to the best of my ability would not have been possible.

I would like to thank:

my supervisor Professor Michèle Ramsay for all the time she took out of her busy schedule to assist me (always with a smile), for her priceless advice, support and encouragement;

all my colleagues in the laboratory for their support;

my mom, who even through her illness, has been a great support to me,

my two wonderful sisters, Samantha and Natalie, and my boyfriend Brendan, for their constant love and support;

the South African Medical Research Council and NHLS Research Trust for their financial assistance, the NHLS infrastructure for the use of their equipment and a big thank-you to Bavesh Kana and the TB laboratory for the use of the real-time Lightcycler.

And finally a very big thank-you to all the individuals who participated in this study.

TABLE OF CONTENTS

DECLARATION		**ii**
DEDICATION		**iv**
ABSTRACT		**v**
ACKNOWLEDGEMENTS		**vii**
TABLE OF CONTENTS		**viii**
LIST OF FIGURES		**xiv**
LIST OF TABLES		**xvii**
LIST OF ABBREVIATIONS		**xxi**
1. INTRODUCTION		
1.1	**THE SKIN**	**1**
1.1.1	**An overview of epidermal differentiation and the keratinisation process**	**3**
1.1.2	**The pathobiology of the Stratum Corneum**	**6**
1.1.2.1	*Lipid abnormalities in the stratum corneum*	*8*
1.1.2.2	*Protein abnormalities in the stratum corneum*	*9*
1.1.3	**Palmoplantar skin and trunk skin**	**11**
1.1.4	**Palmoplantar keratodermas: Inherited diseases of the epidermis**	**12**
1.2	**KERATOLYTIC WINTER ERYTHEMA (KWE)(MIM No.14837)**	**14**

1.2.1	The clinical description of KWE	14
1.2.2	The genetics of KWE	17
1.3	THE IDENTIFICATION OF THE KWE LOCUS	17
1.4	THE 8p23 POLYMORPHIC INVERSION	19
1.5	THE GENES WITHIN THE KWE CRITICAL REGION ANALYSED IN THIS STUDY	23
1.5.1	Farnesyl - diphosphate farnesyltransferase (*FDFT1)*	23
1.5.2	Cathepsin B (*CTSB)*	29
1.5.3	Deubiquitinating enzyme (*DUB-3)*	32
1.5.4	Open reading frame 49 (C8orf49)	38
1.6	MUTATION DETECTION STRATEGY	38
1.6.1	Analysing the coding regions of the candidate genes	38
1.6.2	Analysing the 5' untranslated regions (5' UTRs) of candidate genes	39
1.7	RELATIVE GENE EXPRESSION USING REAL-TIME PCR	42
1.8	PROJECT OBJECTIVE	44
2	MATERIALS AND METHODS	
2.1	SUBJECTS	45
2.2	NUCLEIC ACID EXTRACTION AND QUANTIFICATION	47
2.2.1	DNA	47
2.2.2	RNA	47

2.3	**THE POLYMERASE CHAIN REACTION (PCR)**	**48**
2.3.1	**Reverse transcriptase PCR**	**48**
2.3.2	**Conventional PCR**	**48**
2.3.3	**Real-time PCR**	**51**
2.3.4	**Primer design**	**51**
2.3.5	**PCR optimization**	**54**
2.4	**MICROSATELLITE GENOTYPING AND HAPLOTYPE ANALYSIS**	**59**
2.4.1	**Microsatellite marker selection**	**59**
2.4.2	**Setting up a microsatellite genotyping run**	**59**
2.4.3	**Haplotype analysis**	**59**
2.5	**CANDIDATE GENE DNA AND mRNA MUTATION SCREENING**	**60**
2.5.1	**PCR fragment purification**	**60**
2.5.2	**Cycle sequencing**	**61**
2.5.3	**Capillary electrophoresis**	**62**
2.5.4	**Sequence analysis**	**62**
2.5.4.1	*Ensembl*	*63*
2.5.4.2	*Polyphen (**Poly**morphisms **Phen**otyping)*	*63*
2.5.4.3	*SNPs3D*	*65*
2.5.4.4	*FastSNP (**F**unction **A**nalysis and **S**election **T**ool for **S**ingle nucleotide polymorphism*	*66*
2.6	**REAL-TIME PCR USING THE ROCHE LIGHTCYCLER AND THE SYBR GREEN METHOD OF DETECTION**	**67**

2.6.1 Determination of real-time PCR efficiencies of each system 67

2.6.2 Setting up the run 68

2.6.3 Statistical Analysis 69

3.1 THE CONSTRUCTION OF HAPLOTYPES 70

3.2. CANDIDATE GENE ANALYSIS 80

3.2.1 The Cathepsin B gene (*CTSB)* 80

3.2.1.1 Genomic DNA analysis 80

3.2.1.2 CTSB cDNA analysis 86

3.2.1.3 The relative quantification of CTSB expression in human epidermal palmar skin 87

3.2.2 Farnesyl-diphophate-farnesyl (*FDFT1)* 94

3.2.2.1 Genomic DNA analysis 94

3.2.2.2 Further variant analysis in a larger group of unaffected Individuals (n=15) 100

3.2.2.3 FDFT1 cDNA analysis 101

3.2.2.4 The relative quantification of FDFT1 expression in human epidermal palmar skin 102

3.2.3 *DUB-3* analysis 109

3.2.4 *C8orf49* analysis 115

4.1 THE FINE MAPPING OF THE KWE CRITICAL REGION BY ANALYSIS OF THE KWE FAMILY HAPLOTYPES 121

4.1.1 Family 23 and family 46 121

4.1.2	**Family 26**	**122**
4.1.3	**Family 36**	**122**
4.2	**MUTATIONAL ANALYSIS**	**125**
4.2.1	**CTSB as a KWE gene candidate**	**130**
4.2.2	**FDFT1 as a KWE candidate**	**133**
4.2.3	**DUB-3 and C8orf49 KWE gene candidates**	***141***
4.3	**CONCLUSIONS**	**143**
4.4	**LIMITATIONS AND FUTURE STUDIES**	**145**
	REFERENCES	**146**
	APPENDICES	**162**
	APPENDIX A	**162**
	APPENDIX B	**166**
	APPENDIX C.I	**169**
	APPENDIX C.II	**171**
	APPENDIX C.III	**173**
	APPENDIX C.IV	**175**
	APPENDIX D	**177**
	APPENDIX E.I	**179**
	APPENDIX E.II	**181**
	APPENDIX E.III	**182**
	APPENDIX F	**183**
	APPENDIX G	**185**

APPENDIX H **189**

APPENDIX I.I **190**

APPENDIX I.II **191**

LIST OF FIGURES

Figure 1.1: A schematic representation of the three layers that make up human skin 3

Figure 1.2: The morphological differences between trunk and palmoplantar human skin 12

Figure 1.3: The palms of an individual with KWE. The redness of the palm can be seen clearly as well as the edge where the redness arrests 16

Figure 1.4: The progress in defining the KWE critical region 19

Figure 1.5a: A schematic diagram of chromosome 8p23.1-22 showing the relative positions of the genes on a non-inverted chromosome (not drawn to scale**)** (NCBI 36 assembly of the human genome; November 2005) 21

Figure 1.5b: A schematic diagram of chromosome 8p23.1-22 showing the relative positions of the genes on a non-inverted chromosome (not drawn to scale**)** (NCBI 36 assembly of the human genome; November 2005) 24

Figure 1.6: A schematic overview of the *FDFT1* gene (A) and the FDFT1 protein (B) 24

Figure 1.7: The metabolic mevalonate pathway in eukaryotes that is responsible for the production of cholesterol 25

Figure 1.8: A schematic overview of the *CTSB* gene (A) and the cathepsin protein (B) 30

Figure 1.9: A schematic overview of the *DUB-3* gene (A) and the DUB-3 protein (B) 34

Figure 1.10: A schematic overview illustrating the differences between an upstream AUG (uAUG) and an upstream open reading frame (uORF) 40

Figure 2.1: The wedge biopsy taken from the palmar region of a KWE patient 46

Figure 2.2: An illustration of the method used to design primers for the conventional and real-time amplification of *CTSB* and *FDFT1* mRNA 53

Figure 2.3: Nucleotide sequence of the *FDFT1* promoter 54

Figure 3.1: Haplotypes constructed for KWE family 23 71

Figure 3.2: Haplotypes constructed for KWE family 26 73

Figure 3.3: Haplotypes constructed for KWE family 36 75

Figure 3.4: Haplotypes constructed for KWE family 46 77

Figure 3.5: An illustration of the founder KWE haplotype and the region of this haplotype that is conserved among KWE affected individuals 78

Figure 3.6: The confirmed KWE critical region 79

Figure 3.7: The genetic variants identified within *CTSB* 84

Figure 3.8: The standard curve produced to determine the efficiency of the *CTSB* and *GAPDH* real-time PCR assays 89

Figure 3.9: Calculation of the relative efficiency of the target gene *CTSB* system 90

Figure 3.10: Normalized *CTSB* expression in plantar skin from KWE affected and control individuals 93

Figure 3.11: Electropherograms of the genetic variants identified within the *FDFT1* gene 98

Figure 3.12: The relative frequency of the genotypes observed for the novel

variant identified within the 5' UTR of the *FDFT1* gene 100

Figure 3.13: L7L (T<C, rs1047643) genotypes group of unaffected individuals (n=15) individuals 101

Figure 3.14: The standard curve produced to determine the efficiency of the *FDFT1* and *GAPDH* real-time PCR assays 103

Figure 3.15: Calculation of the relative efficiency of the *FDFT1* real-time PCR system The slope value is 0.014 (m<0.1), reflecting optimal relative efficiency 105

Figure 3.16: Normalized *FDFT1* expression 108

Figure 3.17: The genetic variants identified within the coding region of *DUB-3* 114

Figure 3.20: The genetic variants identified within *C8orf49* 119

Figure 4.1: An illustration of the possible effect that the single nucleotide base change in the 5' UTR of *FDFT1* may have on the protein 134

LIST OF TABLES

Table 1.1: A list of diseases in which epidermal cytokine dysregulation has been reported to play a role 37

Table 2.1: A summary of the number of samples collected and their specific use in the present study 45

Table 2.2: The general PCR protocol used to set up the amplification systems for each microsatellite marker 49

Table 2.3: The general PCR protocol used to amplify the 5'UTRs and exons of the candidate genes as well as the mRNA of *FDFT1* and *CTSB* 50

Table 2.4: The optimized Long Template Expand protocol used for the amplification of the large 1.85 kb *DUB-3* fragment 50

Table 2.5: The general PCR protocol used for the real-time amplification of *FDFT1* and *CTSB* mRNA 51

Table 2.6: The sequence and annealing temperature (T_A°C) of the primers used to amplify the markers used in the linkage analysis study 55

Table 2.7: The sequence and annealing temperature (T_A°C) of the primers used to amplify *FDFT1*; and the size of the amplicons produced 56

Table 2.8: The sequence and the annealing temperature (T_A°C) of the primer pair used to amplify *DUB-3;* and the size of the amplicon produced. The sequence of the internal cycle sequencing primers is given 56

Table 2.9: The sequence and the annealing temperature (T_A°C) of the primer pairs used to amplify *C8orf49.* The size of the amplicon produced is given in

base pairs (bp) 57

Table 2.10: The sequence and annealing temperature (T_A°C) of the primers used to amplify *CTSB;* and the size of the amplicons produced (Mahurkar et al., 2006) 57

Table 2.11: The sequence and the annealing temperature (T_A°C) of the primers used to amplify *FDFT1* mRNA; and the size of the amplicons produced 58

Table 2.12: The sequence and the annealing temperature (T_A°C) of the primers used to amplify *CTSB* mRNA; and the size of the amplicons produced 58

Table 2.13: The sequence and the annealing temperature (T_A°C) of the primers used in the real-time amplification of *CTSB* and *FDFT1* mRNA 58

Table: 2.14: The thermal cycling conditions and the PCR components required for the cycle sequencing of each PCR fragment 61

Table 3.1: The genetic variants identified in *CTSB* 83

Table 3.2: A bioinformatic assessment of the potential functional impact of the genetic variants identified within *CTSB* 85

Table 3.3: Analysis of *CTSB* cDNA 87

Table 3.4: The data used to determine the amplification efficiency of the *CTSB* and *GAPDH* systems 88

Table 3.5: The C_q values obtained from the *CTSB* and *GAPDH* patient sample run. The standard deviation (SD) and coefficient of variance (CV) is shown 91

Table 3.6: The C_q values obtained from the *CTSB* and *GAPDH* control sample run. The standard deviation (SD) and coefficient of variance (CV) is shown 92

Table 3.7: The $\Delta\Delta C_q$ value and corresponding $2^{-\Delta\Delta Cq}$ of each patient and control sample 93

Table 3.8: Statistical analysis results 94

Table 3.9: The genetic variants identified in F*DFT1* 96

Table 3.10: The analysis of the genetic variants identified in the promoter region of *FDFT1* within KWE families 97

Table 3.11: A bioinformatic assessment of the potential functional impact of the genetic variants identified within *FDFT1* 99

Table 3.12: Genetic variants identified within of FDFT1 cDNA 102

Table 3.13: The data used to determine the amplification efficiency of the *FDFT1* system 103

Table 3.15: The C_q values obtained from the *FDFT1* and *GAPDH* patient run (Standard deviation (SD) and coefficient of variance (CV) shown 106

Table 3.16: The C_q values obtained from the *FDFT1* and *GAPDH* control sample run (Standard deviation (SD) and coefficient of variance (CV) shown 107

Table 3.17: The $\Delta\Delta C_q$ value and corresponding $2^{-\Delta\Delta Cq}$ of each patient and control sample 108

Table 3.18: Statistical analysis results 109

Table 3.19: The correlation between the increased level of *FDFT1* expression and the severity of the individuals 109

Table 3.20: The genetic variants identified in *DUB-3* 112

Table 3.21: The bioinformatic assessment of the functional impact the

characterised genetic variants identified within *DUB-3* may have on the protein 113

Table 3.22: The genetic variants identified within the open reading frame, *C8orf49* 117

Table 3.23: A bioinformatic analysis of the characterised genetic variants identified within the opening reading frame, C8orf49 120

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
bp	base pairs
°C	degrees Celsius
C8orf49	open reading frame 49 on chromosome 8
cM	centiMorgan
cDNA	complementary DNA
CE	cell envelope
CTSB	cathepsin B
ddH_2O	deionized distilled water
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
D8Sdi	dinucleotide microsatellite marker on chromosome 8
D8Stet	tetranucleotide microsatellite marker on chromosome 8
DUB-3	deubiquitinating gene 3
EDTA	ethylenediaminetetraacetic acid
e.g.	for example
6-FAM	6 – Carboxyfluorescein
FDFT1	farnesyl transferase
FISH	fluorescence *in situ* hybridization
5'UTR	five prime untranslated region

HEX	6 - carboxy - 2', 4, 4', 5', 7, 7' – hexachlorofluorescein
HI-DI	highly deionized formamide
kb	kilobase pairs
KWE	keratolytic winter erythema
LONRF1	LON peptidase N-terminal domain and ring finger 1
$MgCl_2$	magnesium chloride
ml	millilitre
mRNA	messenger RNA
mM	millimolar
NCBI	National Center for Biotechnology Information
NEIL2	nei like 2 gene
NF1	nuclear factor 1
No.	number
p	short arm of chromosome
PCR	polymerase chain reaction
PolyPhen	polymorphism phenotyping
RNA	ribonucleic acid
ROX	6 - Carboxy - X – rhodamine
Sp1	stimulatory protein 1
SRE-1	sterol regulatory element
3'UTR	three prime untranslated region
uAUG	upstream start codon

µg	microgram
µl	microliters
uORF	upstream open reading frame
UTR	untranslated region
UV	ultra violet

CHAPTER 1

INTRODUCTION

1.1 THE SKIN

The skin is the largest organ of the body, accounting for an eighth of the total body weight. This uniquely engineered organ is the bodies' first line of defense against heat, light, injury and infection (http://www.umm.edu/dermatology-info/anatomy.htm). The skin also regulates body temperature and stores essential substances such as water, fat, and vitamin D (Goldsmith, 1991). A primary function of the skin is to generate and maintain the cutaneous epidermal permeability barrier which prevents the loss of moisture from the skin and also prevents the entry of infectious or toxic substances into the body (Paulson, 1999; Altemus et al., 2001). The skin is continually in a process of regeneration and is metabolically active (McGrath et al., 2004). Human skin is made up of multiple layers of epithelial tissue that act as a protective shield to underlying organs and muscles.

Human skin is comprised of two main layers, the outermost epidermis and the innermost dermis. These two layers are separated by a layer of loose connective tissue known as the subcutaneous layer or hypodermis (Byrant and Nix, 2004). The epidermis is derived from the embryonic ectoderm and its primary function is in protection as the major barrier against the environment (Williams et al., 2005). This complex avascular membrane is composed of four main cell types namely keratinocytes, melanocytes, Langerhans cells and Merkel cells (McGrath et al., 2004; Williams et al., 2005). It is a homeostatic, self renewing tissue that expresses differentiation-specific genes sequentially as keratinocytes move outward from the basal to the granular cell layer

(Schmuth et al., 2001). Keratinocytes are the primary type of cell in the epidermis (McGrath et al., 2004) and are responsible for producing keratin (Freinkel and Woodley, 2001). Keratin is the protein present on the surface of the epidermis which makes this layer of skin dry and waterproof thus preventing excessive water loss (Mader, 2004). Melanocytes are cells that are located within the bottom layer of the epidermis which function to produce melanin, a pigment that absorbs energy from the sun protecting the skin from ultra violet radiation (Lin and Fisher, 2007). The Langerhans cells are recognized as the major fully-functional antigen-presenting cells in the skin (Merad et al., 2008) and function primarily to prevent unwanted substances from penetrating the skin (Scheibner et al., 1986). Merkel cells are specialized skin sensory receptors that initiate the perception of gentle touch (Haeberle et al., 2004).

The middle layer of the skin is the dermis, a fiber rich connective tissue portion of the skin (Layman, 2003) that is made up of two layers, namely the papillary layer and reticular layer (Williams et al., 2005). Unlike the epidermis, the dermis is richly supplied with blood and lymphatic vessels, contains the sweat and oil glands as well as hair follicles and nerve endings (Wilkes et al., 1973; Paulson, 1999). Collagen and elastin are the major components of the dermis (Byrant and Nix, 2004). Collagen supports the dermis, lending to its durability and elastin keeps the skin flexible (Layman, 2003). The subcutaneous layer or hypodermis is the fatty layer of the skin that bridges between the overlying dermis and the underlying body constituents (Marks and Miller, 2006). This layer of fatty tissue protects the body's organs against injury and conserves bodily heat (Marks and Miller, 2006). Figure 1.1 illustrates the position and composition of the three major layers of the skin.

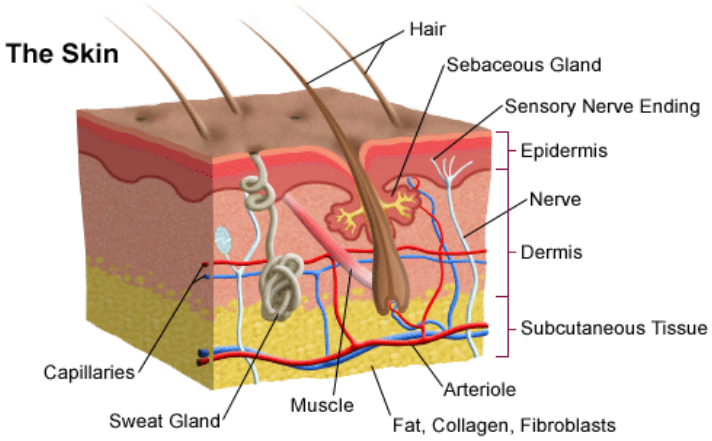


Figure 1.1: A schematic representation of the three layers that make up human skin (www.rush.edu/rumc/page-1098987353859.html).

1.1.1 An overview of epidermal differentiation and the keratinisation process

The epidermis is the outermost layer of the skin that functions to protect the body from mechanical injury, UV damage, dessication and immunological damage (Byrant and Nix, 2004; Marks and Miller, 2006). The epidermis is a stratified epithelium composed of five histologically distinct layers: the stratum basale, stratum spinosum, stratum granulosum, stratum lucidum and the stratum corneum (Elias, 2005). Together, these layers constantly rebuild and renew the skin's surface from within through the process of keratinisation (Ponec, 1983). Keratinisation is defined as the process by which epidermal cells, keratinocytes, undergo maturation and differentiation as they migrate from the innermost basal layer to the outermost stratum corneum (Kanitakis, 2002). The process of keratinisation allows for self renewal under both homeostatic and injury conditions. During keratinisation, epidermal cells differentiate into stratified squamous cells which ultimately form a highly resistant permeability barrier (Smith et al., 1982). This semi-permeable epidermal barrier is an extremely important feature of skin that functions to prevent

the loss of moisture from the skin and prevent the entry of infectious or toxic substances into the body (Segre, 2006).

Keratinisation begins in the deepest epidermal layer, the stratum basale (Marks and Miller, 2006). The basal cells are metabolically active and are attached to the basement membrane by rivet-like structures known as hemidesmosomes (Byrant and Nix, 2004). These cuboidal shaped cells are continually dividing to produce new keratinocytes which are progressively pushed to the skin's surface. As the new cells migrate upwards, they force older cells closer to the surface (Radjoda et al., 2006; Paulson, 1999).

The stratum spinosum is positioned directly above the basal layer and is the thickest layer of the epidermis. The function of the stratum spinosum is to initiate differentiation (Radoja et al., 2006). The stratum spinosum and the stratum basale are collectively termed the Malpighian layer (McGrath et al., 2004). The keratinocytes present in this layer are maturing cells with the ability to produce keratin, the most abundant protein produced during the complex process of epidermal differentiation (Schmuth et al., 2001). The keratins will aggregate to form tonofilaments that in a condensed state are referred to as desmosomes. Desmosomes are specialised cell structures that take part in cell-cell adhesion, providing a connection between intermediate filaments of adjacent cells' cytoskeletons (Alberts et al., 2003). Although it is generally accepted that cell-cell adhesions in the epidermis are primarily maintained by desmosomes, it is thought that additional connections may be mediated by intercellular lipids (Smith et al., 1982). From the stratum spinosum, the

migrating cells will pass through the stratum granulosum and the stratum lucidum. In these thin granular layers of the epidermis, lipids are produced inside lamellar bodies, keratins associate with filaggrin and are bundled into macrofibrils and a cornified envelope (CE) is assembled (Segre, 2006). The development of the cornified envelope begins with the disintegration of the nucleus, cytoplasm and other cellular organelles (Rossi et al., 2000). The absence of a cellular membrane allows for extracellular calcium to move freely into the cell. The transglutaminase enzyme, which functions to irreversibly cross link the CE proteins, is activated by this calcium influx. This results in the formation of a tough sac that surrounds the keratin proteins. Lipids are then extruded into the intercellular space present on the CE scaffold, completing the final step of epidermal barrier formation (Segre, 2006). When these keratinized cells reach the skins' surface, they become part of the outermost layer, the stratum corneum where they are shed and replaced by new underlying keratinocytes (McGrath et al., 2004).

The stratum corneum is the final product of epidermal differentiation. This horny layer serves to protect the body from harmful penetrating substances and continuously regulates water loss from the body (Kuechle et al., 2000; Stachowitz et al., 2002). The stratum corneum is a two-compartment system consisting of protein-enriched corneocytes embedded in a lipid-enriched intercellular matrix (Fulmer et al., 1986). The barrier nature of the stratum corneum depends wholly on its constituents; 70 – 80% is protein, predominantly alpha-keratin and 5 – 15% is lipids (Wilkes et al., 1973).

The process of keratinisation maintains the thickness of the stratum corneum by balancing the production of keratinocytes (Wertz et al., 1987; Cork et al., 2006). Abnormal or aberrant keratinisation may result in an unusually thick stratum corneum as well as a dry and scaly skin surface due to either hypoproliferation or a delay in cell shedding (Jackson et al., 1993). There are many skin diseases which are characterized by abnormal keratinisation including seborrheic dermatitis, icthyosis vulgaris, lamellar ichthyosis (Segre, 2006), erythrokeratolysis, follicular keratosis parakeratosis ichthyosis, psoriasis (Altemus et al., 2001), X-linked ichthyosis and Dowling-Degos disease (McLean and Irvine, 2007). Keratolytic winter erythema was first described by Findlay et al. in 1977 as being "a disorder of epidermal keratinisation that involves the necrobiosis of the Malpighian layer with the subsequent dissection of the stratum corneum." This statement highlights the importance in understanding the complex process of keratinisation and the pathobiology of the stratum corneum.

1.1.2 The pathobiology of the Stratum Corneum

The epidermis is a dynamic system whose metabolic activity is regulated largely by the integrity of the permeability barrier (Jackson et al., 1993). The most important function of the epidermis is to form and maintain the epidermal barrier which protects the organism against environmental damage (Reichelt et al., 1999). This barrier is found in the stratum corneum and consists of the protein-rich, highly cross-linked cornified envelope and the extracellular lipid lamellae (Elias and Menon, 1991). Disruption of the barrier induces secretion of preformed lamellar bodies of cells at the stratum granulosum-stratum corneum interface, increases synthesis of cholesterol, ceramide and fatty acids, and enhances the activity of lipid processing enzymes, resulting in the

formation of new lipid membrane sheets (Altemus et al., 2001). Disruption of this barrier also increases epidermal cell proliferation and DNA synthesis (Proksch et al., 1996). The stratum corneum is the end product of epidermal differentiation and involves many biochemical processes (Schmuth, 2002). The same processes that form the stratum corneum also expose it to pathological processes, including abnormal barrier function, abnormal desquamation and hyperproliferation (Jackson et al., 1993). Pathological desquamation results in the formation of an abnormally thick stratum corneum with a dry or scaly appearance (McGrath et al., 1993). Increased thickness of the stratum corneum is due to either hyperproliferation or abnormal retention of corneocytes which is due to intercellular detachment failure (Jackson et al., 1993).

This bi-layered lipid envelope of the stratum corneum contains a unique mixture of lipids namely, ceramides, cholesterol, cholesterol sulfate, sterol esters and fatty acids (Reichelt et al., 1999). These stratum corneum lipids are synthesized in the epidermis (Feingold, 1991) and are important factors in the formation of a functional epidermal permeability barrier (Elias and Menon, 1991; Pilgram et al., 2001). It has been demonstrated that during epidermal differentiation, characteristic changes in lipid composition occur. These changes are consistent with the formation of the epidermal permeability barrier (Bouwstra et al., 1999) and include a progressive depletion of glycol and sphingolipids and the increase of cholesterol, fatty acids and ceramides, the three key stratum corneum lipids. For this reason the stratum corneum lipids have become a popular research topic because of the increasing evidence suggesting a major role of these lipids in the maintenance of the epidermal barrier homeostasis and in epidermal barrier repair. Abnormal stratum corneum lipid composition has been

associated with many scaly skin disorders for example; high levels of cholesterol sulphate were found in recessive X-linked ichthyosis (Zettersten et al., 1998), increased levels of free cholesterol and decreased concentrations of esterfied cholesterol has been found in psoriasis patients (Schmidt et al., 1977) and abnormally high levels of n-alkanes have been found in congenital ichthyosis (Fulmer and Kramer, 1986). The scaly appearance of these diseases is the gross morphological presentation of hyperkeratosis which is the result of either an increase in cell proliferation or a delay in cell shedding (Segre, 2006).

1.1.2.1 Lipid abnormalities in the stratum corneum

The stratum corneum lipids, cholesterol, ceramides and free fatty acids play a major role in the formation of the permeability barrier and any abnormality of these lipids will primarily lead to transepidermal water loss and barrier damage (Jackson et al., 1993; Schmuth et al., 2001). These lipids are present in an equimolar ratio (Pilgram et al., 2001). The loss of polar lipids from the stratum corneum, such as cholesterol and ceramides, results in profound barrier damage while the effect of losing less essential non-polar lipids causes only minor barrier damage (Grubauer et al., 1989; Jackson et al., 1993). Any degree of barrier disruption results in the release of cytokines (Wood et al., 1992). These cytokines trigger epidermal hyperplasia and in reaching the dermis, initiate dermal inflammation (Jackson et al., 1993). Epidermal hyperplasia in turn results in abnormal desquamation (Grubauer et al., 1989). The overall result is a stratum corneum comprising of abnormal lipids and proteins with poor barrier function and impaired desquamation (Jackson et al., 1993). There is a number of genetically determined disorders that are

associated with lipid abnormalities of the stratum corneum including recessive X-linked ichthyosis (Koppe et al., 1978), Refsum disease, Harlequin ichthyosis (Dykes et al., 1978; Schmuth et al., 2001), atopic dermatitis and Lamellar Ichthyosis (Pilgram et al., 2001; Demerjian et al., 2006) and psoriasis (Jackson et al., 1993).

1.1.2.2 Protein abnormalities in the stratum corneum

Although abnormalities in the stratum corneum proteins affect the stratum corneum to a lesser extent than lipid abnormalities, there are a number of skin diseases that result from abnormalities in keratins, desmosomes, filaggrin, profilaggrin and cornified envelopes, the protein components of the stratum corneum (Jackson et al., 1993). Desmosomes are structures that play a major role in cell-cell adhesion (Whittock et al., 1999). They also provide anchoring points for intermediate filaments to help build a strong structural framework (Menon et al., 1992; McGrath et al., 2004). Desmosomes have been associated with the pathogenesis of a number of skin disorders including ichthyosis vulgaris, psoriasis, X-linked ichthyosis, Refsum disease and harlequin ichthyosis (Menon et al., 1992). In normal skin, the desmosomes are present throughout the epidermis, but are progressively degraded within the stratum corneum by a number of desmosome degrading proteases and glycosidases (Jackson et al., 1993). It is this progressive degradation of the desmosomes that enables cells to be shed from the skin's surface. Keratins are the major structural proteins of the vertebrate epidermis and its appendages, constituting up to 85% of a fully differentiated keratinocyte (Reichelt et al., 1999). The association of keratin mutations with genetic skin fragility disorders is now one of the

best-established examples of cytoskeleton disorders (Lane and McLean, 2004; Ramaekers and Bosman, 2004). This association has served as a paradigm for many other diseases and has been highly informative for the study of intermediate filaments and their associated components, in helping to understand the functions of this large family of structural proteins (Lane and McLean, 2004). The keratin diseases have shown that one of the major functions of intermediate filaments is providing physical resilience for epithelial cells (Fulmer and Kramer, 1985; Lane and McLean, 2004). Keratin proteins have been implicated in the pathogenesis of a number of skin diseases including epidermolytic hyperkeratosis (Kwak and Maverakis, 2006), epidermolysis bullosa simplex (Hovnanian et al., 1993), ichthyosis bullosa of Siemens (McLean et al., 1994) and white sponge naevus (Chao et al., 2003).

A disturbance in the degradation of profliggrin into filaggrin is responsible for a number of stratum corneum abnormalities (Jackson et al., 1993). The hydrolysis of filaggrin results in the release of amino acids and without these amino acids, the amount of water retained in the stratum corneum is decreased (Scott et al., 1988). The cracking, flaking and scaling phenotypic appearance of skin in certain skin diseases could be a direct result of the inflexibility and reduced elasticity of the stratum corneum that is attributed to its decrease in water content (Jackson et al., 1993; Simon et al., 1996).

The cornified envelope is a highly insoluble structure that is formed on the inner membrane of the keratinocytes in the final stage of epidermal differentiation and is made

up of proteins covalently linked together by transglutaminases (Simon et al., 1996). Any abnormality occurring in the cornified envelope that results in an alteration of intracellular lipid organisation, could lead to alterations in barrier function or desquamation although many studies involving the morphological analysis of cornified envelopes in skin diseases indicate that the cell envelopes appear normal (Simon et al., 1996).

1.1.3 Palmoplantar skin and trunk skin

There are two different kinds of human skin; hair-bearing trunk skin and non-hairy glabrous skin. Hair-bearing skin is the skin that covers the majority of the body and has both hair follicles and sebaceous glands but lacks encapsulated sense organs (Kelsell and Stephens, 1999). Glabrous or palmoplantar skin is found only on the palms and soles and is characterised by a thick compact epidermis (McGrath et al., 1993). A thick palmoplantar epidermis is essential due to the constant stresses that are applied to the hand and foot (Kelsell et al., 1999). Palmoplantar skin contains five times fewer melanocytes than trunk skin (Yamaguchi et al., 2007), is more durable under mechanical stress and contains sweat glands which are larger than those found on the rest of the skin's surface (Kelsell and Stevens, 1999). Palmoplantar skin contains a highly retractile and well-defined layer of the epidermis, known as the stratum lucidum, which is absent in the trunk skin (Yamaguchi et al., 1999). The process of keratinisation is well advanced in the stratum lucidum, an epidermal layer consisting of approximately five layers of clear, flat keratinocytes referred to as eosinophilic cells (Zaslau, 2005). Eosinophils are the predominant inflammatory cells in allergic reaction and primarily deal with parasitic infections (Rothenberg, 1998).

Epidermal ridges in the palm are either deep or shallow, whereas those of other body sites are relatively uniform in size (Yamaguchi et al., 1999). Palmoplantar stem cells are found above the deep epidermal ridges, whereas nonpalmoplantar stem cells are located above the papillary dermis (Yamaguchi et al., 1999). Figure 1.2 illustrates the morphological differences between glabrous palmoplantar skin and hairy trunk skin. Because KWE primarily affects palmoplantar skin it would be of extreme value and interest to perform studies that would identify the specific processes of epidermal differentiation and desquamation that differ between palmoplantar and trunk skin.

a) b)

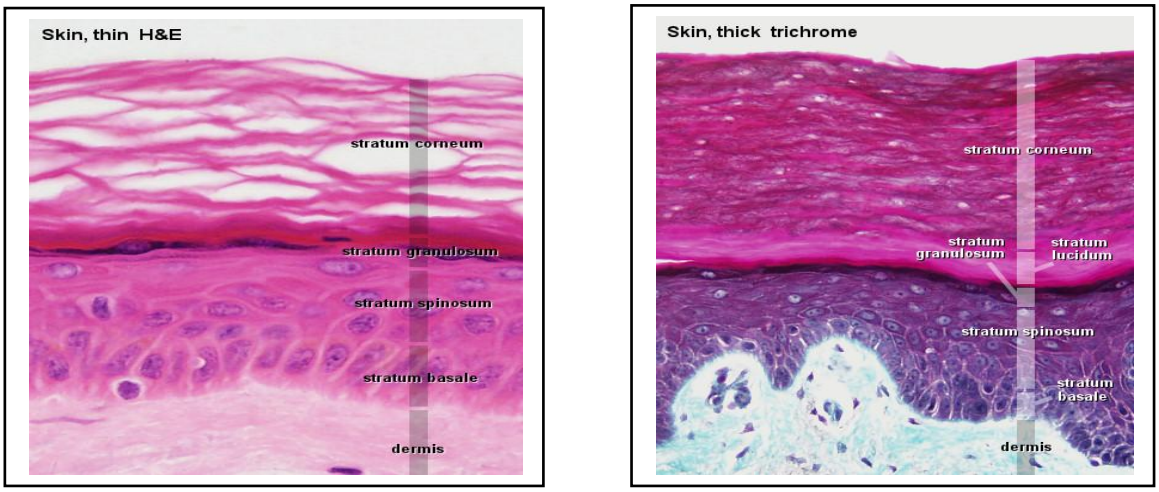


Figure 1.2: The morphological differences between human trunk (a) and palmoplantar (b) skin (http://www.lab.anhb.uwa.edu.au/mb140/CorePages/Integumentary/Integum.htm)

1.1.4 Palmoplantar keratodermas: Inherited diseases of the epidermis

The stratified epithelium of the epidermis represents the tissue location with the largest number of genetic skin diseases characterised (Khavari, 2000). To date, a large number of genes have been implicated in certain monogenic diseases of the epidermis; however there still remain a large number of diseases with unknown etiology. The palmoplantar

skin is a highly specialized tissue which is able to resist mechanical trauma and other physical stresses (Itin and Fistarol, 2005). The inherited palmoplantar keratodermas (PPKs) are an extremely heterogeneous and diverse group of epidermal diseases in which there is hyperkeratosis of the palms and soles (Freedberg et al., 2003). PPK can be divided into the following subgroups: disturbed gene functions in structural proteins (keratins); cornified envelope (loricrin, transglutaminase); cohesion (plakophilin, desmoplakin, desmoglein1), cell-to-cell communication (connexins) and transmembrane signal and transduction (cathepsin C) (Kimyai-Asadi et al., 2002). They can be inherited in both an autosomal dominant and recessive manner and are classified according to their phenotype (Hatsell and Kelsell, 1999). Simple keratodermas manifest lesions only on palmoplantar skin where complex keratodermas involve lesions present on trunk skin, nails, hair, teeth or sweat glands. The syndromic keratodermas are associated with abnormalities of other organs such as various cancers, cardiomyopathy, deafness and adrenal insufficiency (Kelsell et al., 1999). Most of the genes causing PPKs have been mapped on different chromosomes, and many mutations in keratin genes are responsible for these disorders (Kimonis et al., 1994; McLean et al., 1995; Starfield et al., 1997; Smith et al., 1998; Kelsell et al., 1999; Kimyai-Asadi et al., 2002; Whittock et al., 2002; Zhang et al., 2004; Itin and Fistard, 2005). The identification and mapping of the genes that underlie the PPKs will definitely reveal new insights into the biological interactions of the structural components of palmoplantar epidermis (Sybert et al., 1988).

1.2 KERATOLYTIC WINTER ERYTHEMA (KWE) (MIM No. 148370)

In the early twentieth century, South African dermatologists observed a handful of individuals whose soles and palms exhibited bouts of redness and peeling. This novel dermatosis had not been classified before. In 1977, a dermatologist by the name of George Findlay observed that this dermatosis is inherited in an autosomal dominant fashion and that the bouts occur more frequently in winter. After thorough investigation, Findlay et al. (1977) suggested that the novel dermatosis be named keratolytic winter erythema (KWE) to emphasize the three major features of this palmar plantar skin disease. KWE or Oudtshoorn skin disease is a rare monogenic autosomal dominant condition with an unknown cause. KWE is a disorder of epidermal keratinisation that involves the necrobiosis of the Malpigihian layer with the consequent dissection of the stratum corneum (Findlay et al., 1977). This cutaneous disorder occurs with a high prevalence of 1/7200 in the South African Afrikaans-speaking white population and with a lower, but unspecified prevalence in the Coloured population (the term Coloured is used in South Africa in reference to a person of mixed ancestry, including Khoisan, European, Malay and African). The high frequency of KWE in the Afrikaans-speaking white population is a result of genetic drift by founder effect (Findlay et al., 1977). Initially it was thought that KWE was unique to South Africa, however, members from a large German family have been described as displaying clinical symptoms of KWE (Starfield et al., 1997). In 1984, Peter Hull, a dermatologist, became fascinated by this rare skin disease and performed an extensive genealogical study and managed to trace all known South African

KWE families back to a single founder individual by the name of Francois Renier Duminy. Francois Renier Duminy was born in 1747 and was a French mariner and South African pioneer (Duminy, 2006).

1.2.1 The clinical description of KWE

KWE is characterized by intermittent and recurrent centrifugal peeling of the palmar and plantar surfaces (Findlay et al., 1977). Occasional cases of spread onto the trunk and forearms have been reported although this is uncommon. Initially this cyclic occurrence begins with erythema (redness) followed by the thickening of the keratin layer and the formation of dry blisters, from where the peeling develops. Peeling proceeds centrifugally and is arrested at major skin creases (Hull, 1986). A puzzling feature of this disorder is the onset and recurrence of the phenotype with cold weather (Findlay et al., 1977). Generally, the lesions are more severe in the winter months and improve significantly or even disappear during warmer summer conditions, indicating possible sensitivity to temperature or changes in climate humidity. This disorder is associated with palmoplantar sweat (hyperiridrosis) which occurs primarily in winter. Hyperiridrosis is a cold sweat which aggravates peeling and is known to have a distinct odour which is thought to be due to the strong smell of macerated keratin (Hull, 1986). The KWE phenotype demonstrates a large amount of variable penetrance across affected individuals (Starfield et al., 1997). In mild cases the only clinical evidence of KWE may be slight interdigital peeling while in severe cases the entire palm and sole is inflamed and covered with thick peeling skin. Another unusual manifestation of KWE is the complete or partial disappearance of the

phenotype in pregnant women. In some cases the features of KWE worsen prior to and during menstruation. Improvement of the phenotype has been noted in environmental conditions exhibiting high humidity (coastal areas) while a lack of humidity associated with the inland winter is an aggrevating factor (Hull, 1986; Starfield et al., 1997).



Figure 1.3: The palms of an individual with KWE. The redness of the palm can be seen clearly as well as the edge where the redness arrests (Photograph courtesy of Professor Michèle Ramsay)

1.2.2 The genetics of KWE

This disorder usually manifests within the first five years of life and displays large variability among KWE patients. It appears as if the severity of KWE is directly related to age of onset; the earlier the manifestation the more severe the phenotype tends to be (Hull, 1986). KWE follows a dominant mode of inheritance and affects both males and females equally. In 1977, Findlay et al. described the penetrance of KWE to be approximately 85%, however in a more recent study carried out by Hull in 1986 the penetrance was reported to be as high as 92.36%. The conflicting values indicate that the true penetrance of KWE in South African families will always be underestimated due to the minimal and transient disease expressivity in some cases and the fact that in some KWE patients, the manifestations improve with age (Starfield et al., 1997). The high prevalence of KWE in the Afrikaans-speaking white population, as previously mentioned, is attributed to founder effect.

1.3 THE IDENTIFICATION OF THE KWE LOCUS

From 1985 until present, various students from the Ramsay laboratory have applied molecular techniques in a quest to identify the position of the KWE locus and identify the causative gene and mutation. In 1997, Michelle Starfield et al. performed a genome wide linkage analysis study in an attempt to localize the KWE locus to a specific chromosomal location. After the exclusion of more than 200 polymorphic markers, Starfield et al. found significant linkage to marker D8S550 which is located on the short arm of chromosome 8. She also observed that 11 out of the 14 families which she studied demonstrated a

complete 'founder' haplotype that segregated with KWE affected individuals and the remaining 3 families demonstrated parts of this haplotype. This confirmed the existence of a founder haplotype and a founder effect for this disease in South African families. In 2000, Starfield et al. concluded that the KWE locus lies between markers D8S265 and D8S552. In 2002, Appel et al. constructed a transcript map of the region between the markers D8S550 and D8S265. Altogether, 12 transcripts were identified in this region. Of the 12 transcripts, four were known genes namely *BLK*, *MTMR8*, *TDH* and *AMAC1.* After analysing the sequence of each of these genes, no KWE causing mutation was identified, making it unlikely that one of these was the KWE gene. This finding further strengthened the evidence that the region from marker D8S550 up until and including marker D8S265 does not form part of the KWE critical region. The region between and including markers D8S1759 and D8S552 is roughly 1.26Mb and contains seven characterised genes namely *FDFT1*, *CTSB*, *DUB-3*, *GATA4*, *NEIL2*, *BLK*, and *LONRF1*; five putative uncharacterised proteins; two zinc fingers; five beta defensin precursors and an open reading frame C8orf49.

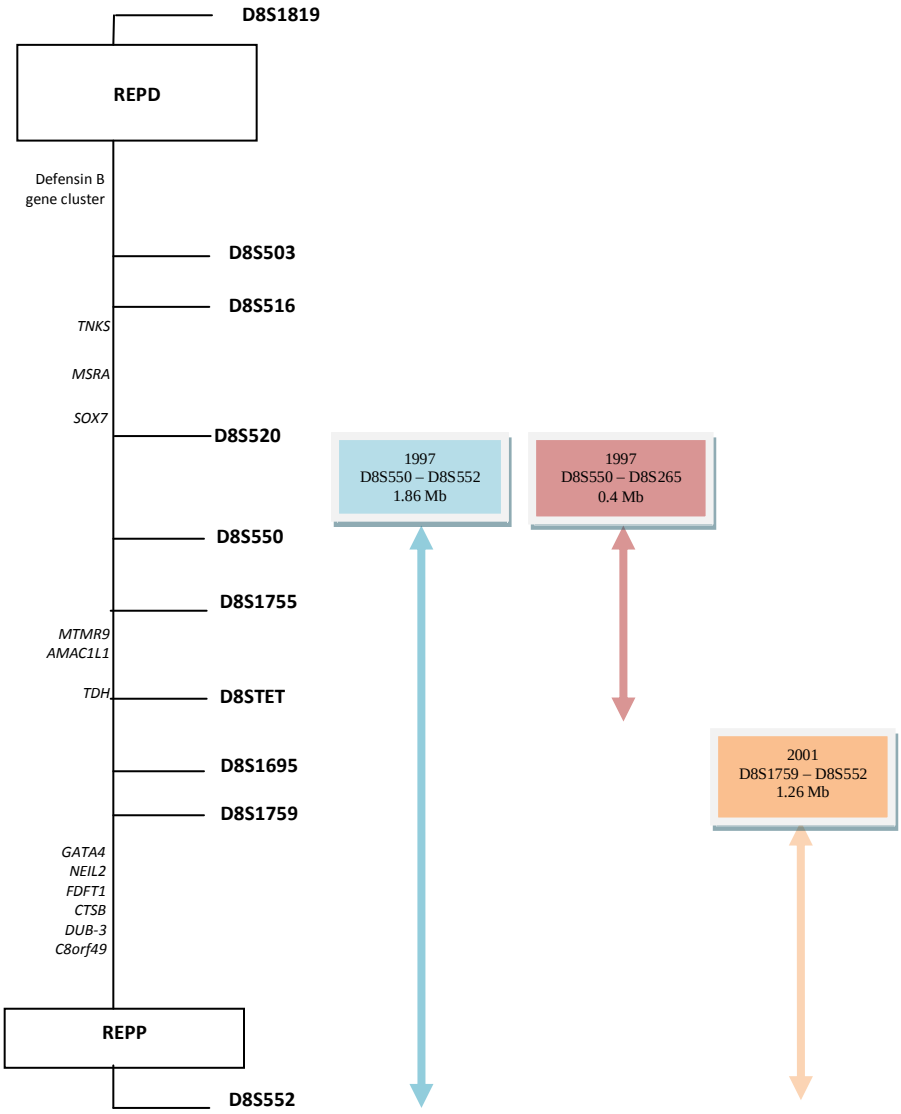


Figure 1.4: The progress in defining the KWE critical region (distance between markers not shown to scale). The red bars indicate the chromosomal region classified as the KWE critical region and the yellow bars are the chromosomal regions excluded from the KWE critical region. None inverted region.

1.4 THE 8p23 POLYMORPHIC INVERSION

The KWE critical region, located on the short arm of chromosome 8, is flanked by two olfactory receptor/defensin repeat regions, REPP (repeat proximal) and REPD (repeat

distal). REPD is approximately 0.4Mb and REPP is approximately 1.3Mb. The olfactory receptor (OR)–gene superfamily is the largest in the mammalian genome. Several of the human olfactory receptor genes appear in clusters with more than ten members located on almost all human chromosomes, and some chromosomes contain more than one cluster (Giglio et al., 2001). It is within these repeated sequences that break points occur resulting in a large number of chromosomal rearrangements. Different rearrangements, most of them recurring, are associated with the distal 8p region. Among them, inv dup(8p) (Floridia et al. 1996), del(8p22) (Devriendt et al. 1999; Giglio et al., 2002), small marker chromosomes der(8)(p23-pter) (Neumann et al.,1999) and a 8p23.1 duplication (Barber et al., 2008). The t(4;8)(p16;p23) translocation, in either the balanced or unbalanced form, has been reported several times (Giglio et al., 2002). A submicroscopic, common 8p23 polymorphic inversion has also been identified (Giglio et al., 2001). This 8p23 inversion involves the KWE critical region. Because this 8p inversion is polymorphic (Giglio et al., 2001) and has no known phenotypic effect in carriers, we know that it is unlikely to be the reason for the manifestation of KWE. The previous KWE linkage studies defined the KWE critical region as the region between and including markers D8S1759 and D8S552 (Starfield et al., 1997). When assuming the presence of the polymorphic 8p23.1-22 inversion, this region is much larger (Figure 1.5a and b) and there is evidence from three colour FISH experiments that KWE lies on an inverted chromosome (Dr P Willem., unpublished data). A haplotype study was performed with the aim to determine which markers segregate with affected individuals and hence, reduce this large region and secondly to validate the KWE critical region.

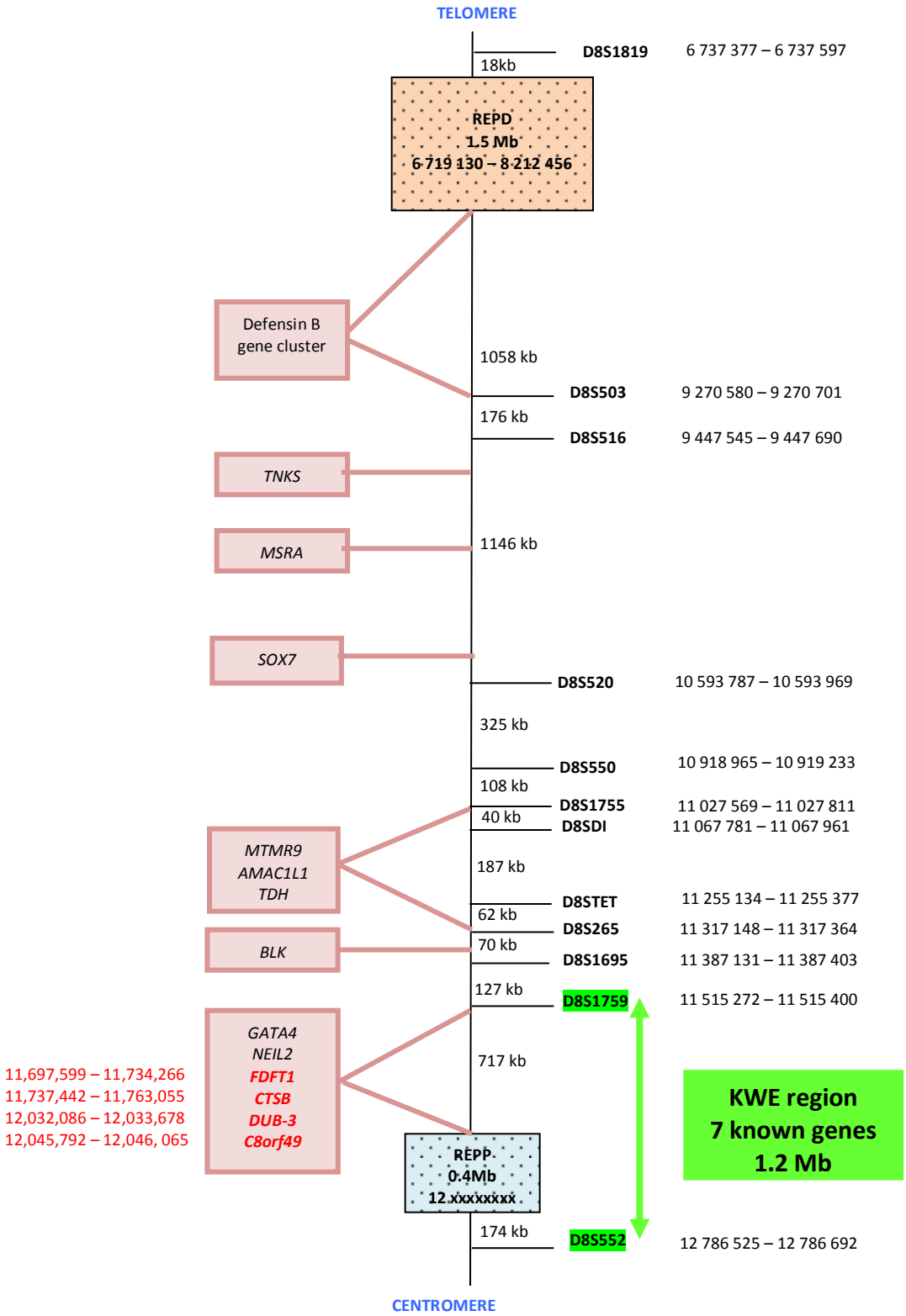


Figure 1.5a: A schematic diagram of chromosome 8p23.1-22 showing the relative positions of the genes on a non-inverted chromosome (not drawn to scale**)** (NCBI 36 assembly of the human genome; November 2005).

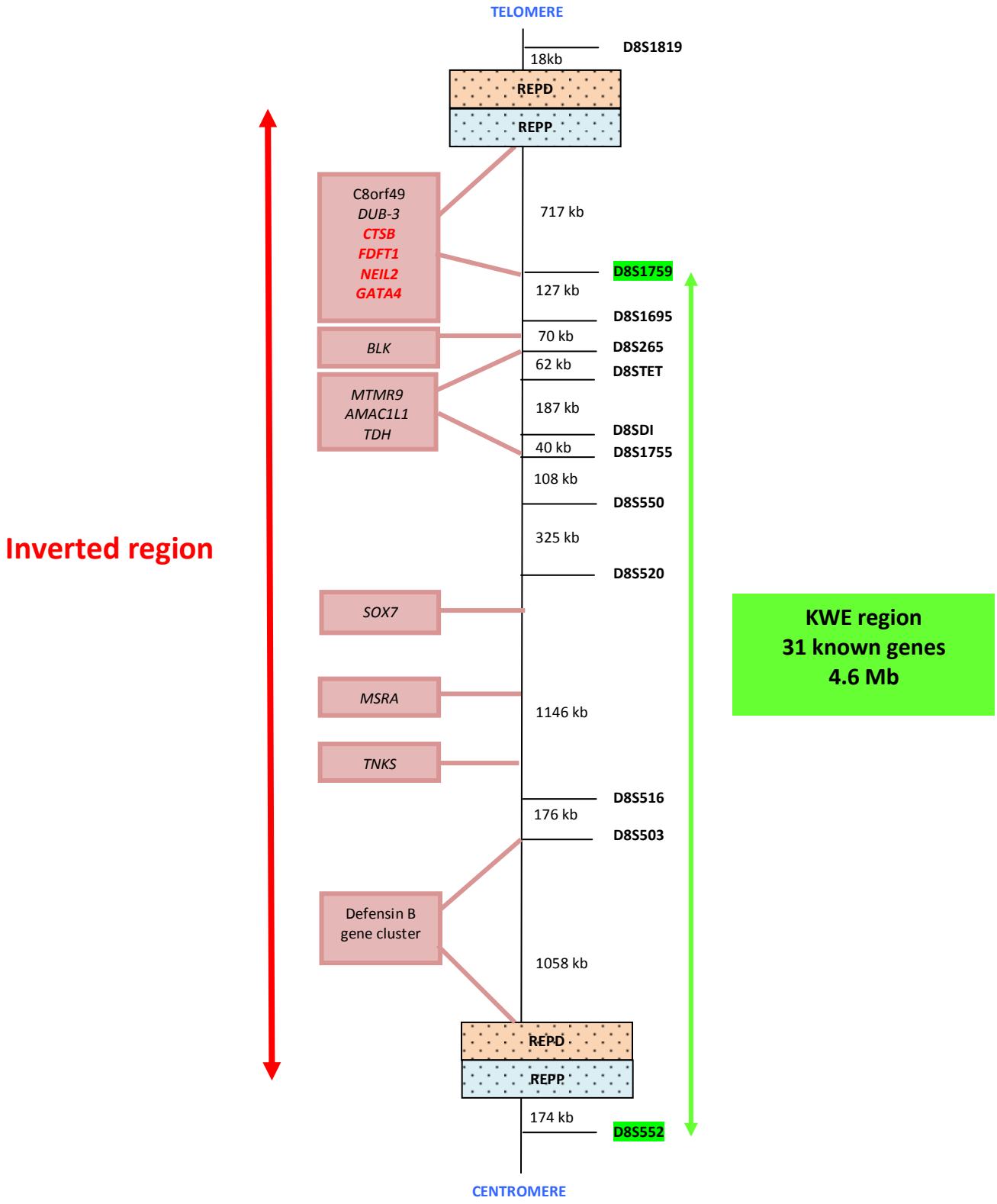


Figure 1.5b: A schematic diagram of chromosome 8p23.1-22 showing the relative positions of the genes on an inverted chromosome (not drawn to scale**)** (NCBI 36 assembly of the human genome; November 2005)

1.5 THE GENES WITHIN THE KWE CRITICAL REGION ANALYSED IN THIS STUDY

The genes present in the critical KWE region that were chosen to be the most likely KWE candidates were cathepsin B (*CTSB*) and farnesyl-diphosphate farnesyltransferase (*FDFT1*). *C8orf49* was analysed as well due to its location within the KWE critical region and the presence of upstream open reading frames in its 5'UTR. A fellow student demonstrated that the deubiquitinating enzyme gene (*DUB-3*), also located within the KWE critical region, is highly conserved among many species. For this reason, *DUB-3* was included in the analysis.

1.5.1 Farnesyl - diphosphate farnesyltransferase (*FDFT1*)

FDFT1 encodes a membrane-associated enzyme located at a branch point in the mevalonate pathway (Schechter et al., 1994), a pathway responsible for the production of critical end products such as cholesterol, dolichol, ubiquinone, steroid hormones and prenylated proteins in humans (Pandit et al., 2000). In 1993 Jiang et al. isolated *FDFT1* cDNA and the gene was localized to chromosome 8p23.1-p22 using FISH (Jiang et al., 1993; Hazra et al., 1995). *FDFT1* contains 8 exons and 7 intervening sequences (IVS). It has only one conserved domain named the Trans_IPPS_HH domain (Figure 1.6). The enzyme encoded by *FDFT1* is essential in the production of lipids that are present in many organs of the body including skin, both palmoplantar and non-palmoplantar.

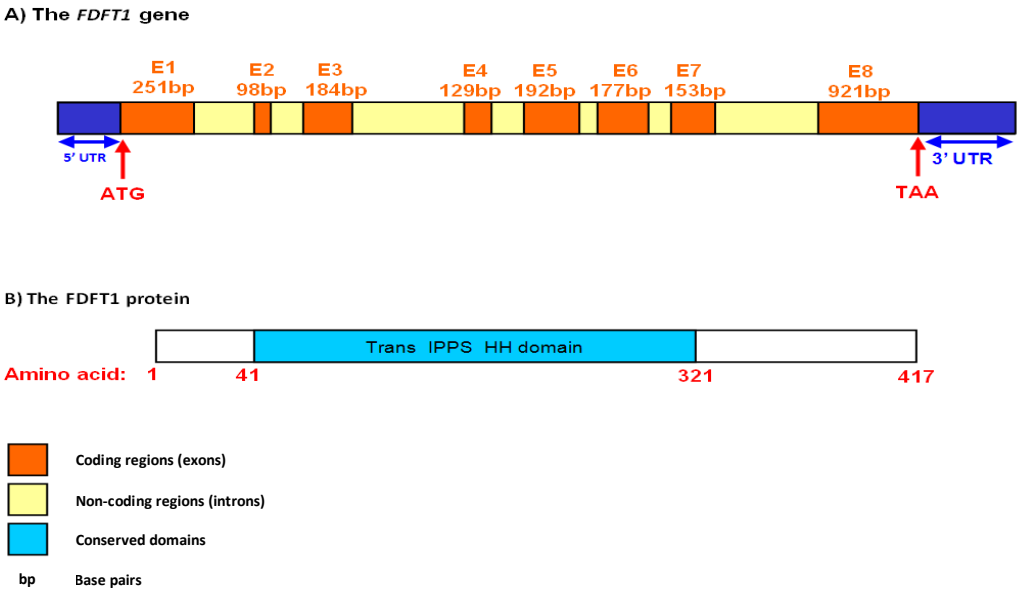


Figure 1.6: A schematic overview of the *FDFT1* gene (A) and FDFT1 protein (B)

FDFT1 encodes *farnesyl-diphosphate farnesyltransferase*, also referred to as *squalene synthase*. Squalene synthase (EC 2.5.1.21) is a 47-kDa membrane-associated enzyme that functions to catalyze the two-step dimerization of two farnesyl-diphosphate molecules into squalene (Figure 1.8), the first committed step in cholesterol formation in mammals (Pandit et al., 2000).

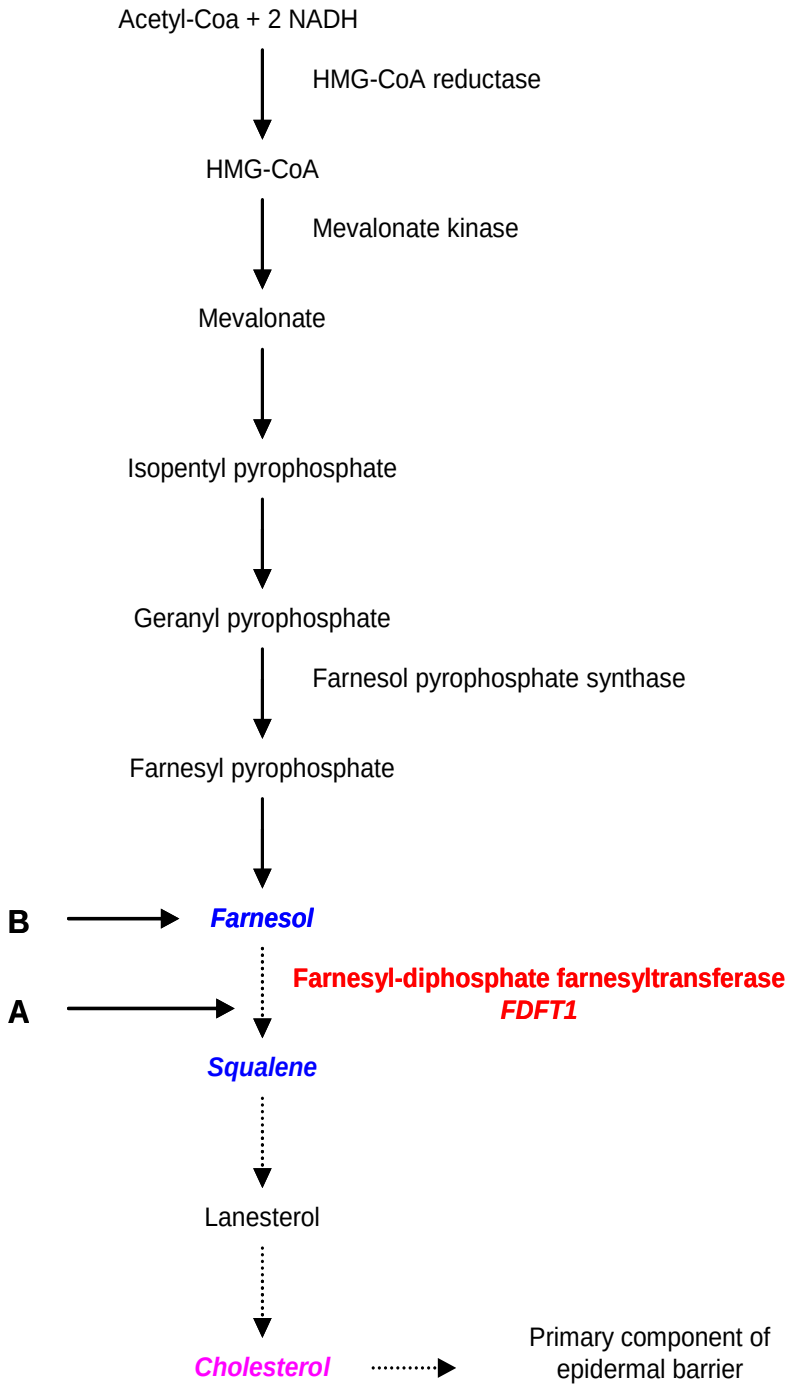


Figure 1.7: The metabolic mevalonate pathway in eukaryotes that is responsible for the production of cholesterol (Mason, 2003) (A and B are referred to in subsequent sections)

The final product produced via the mevalonate pathway is cholesterol, a product essential in many cell functions such as cell growth, cytokinesis and differentiation (Harris et al., 1997) Cholesterol, a component of intercellular lipids, is important for stratum corneum homeostasis (Harris et al., 2000; Tsuruoka et al., 2001) and is one of the three key lipid components that make up the stratum corneum. The evidence suggesting that cholesterol and cholesterol sulphate regulate desquamation (Haratake et al., 2006) and that cholesterol plays an important role in epidermal barrier functioning, is overwhelming and persuasive (Menon et al., 1985; Grubauer et al., 1987; Feingold et al., 1989; Zettersten et al., 1997). In order to achieve optimal barrier permeability and normal barrier recovery cholesterol, free fatty acids and ceramides are required in an equimolar distribution in the stratum corneum (Elias, 2005). In 2005, Peter Elias headed a study based on barrier disruption and its effect on cholesterol production. Their results indicate that barrier disruption initiates an increase in cholesterol synthesis. This increased cholesterol production is due to an associated increase in the activity, protein levels and mRNA levels of key enzymes in the cholesterol synthetic pathway; i.e., HMG CoA synthase, farnesyl diphosphate synthase and squalene synthase (farnesyl-diphosphate farnesyltransferase) (Figure 1.7). In the same study, it was demonstrated that a blockage of cholesterol synthesis, by inhibition of one of these enzymes, slowed barrier recovery and delayed the return of cholesterol to the stratum corneum. These results clearly indicate that epidermal cholesterol is required for lipid barrier formation, for the maintenance of this barrier and for barrier repair. Cholesterol has been shown to be involved in the late stages of epidermal differentiation by promoting cornified envelope formation (Schmidt et al.,

1991; Jans et al., 2004). In a separate study, Norlen (1999) demonstrated that a disturbance of the cholesterol/ceramide/fatty acid ratio which resulted in an increase of the cholesterol concentration caused a depressant effect on water permeability through the skin, however the mechanism remains unclear. A structural alteration of the protein or altered levels of the protein could have an effect on the cholesterol biosynthesis pathway due to a disruption occurring at point A in Figure 1.7. This disruption will ultimately affect the cholesterol biosynthetic pathway, disturbing the equimolar balance that is maintained between the stratum corneum lipids and could have an indirect effect on barrier permeability, function and repair. There are many skin diseases that are characterized by a barrier abnormality, namely psoriasis, contact dermatitis and atopic dermatitis (Elias et al., 1999) and the extent of the barrier abnormality parallels clinical severity (Sugarman et al., 2003; Prosch et al., 1996).

One of the many other possibilities that a mutation within *FDFT1* could assert is a change in the concentration of farnesol, an isoprenoid in the mevalonate pathway (Figure 1.7, B). The best known signal for induction of normal human keratinocyte (NHK) differentiation is the calcium gradient mechanism (Yuspa et al., 1989; Hanley et al., 2000). Other signals which control NHK differentiation are not well understood. Many studies have suggested that the RxR-heterodimerizing nuclear hormone receptors play an important role in the regulation of epidermal differentiation. Retinoic acid receptors are the most intensely studied members of this group. The over expression of a truncated inactive retinoic acid receptor (RAR) during fetal development results in aberrant epidermal development

(Saitou et al., 1995; Hanley et al., 2000) and both systematic and topical RAR ligands modulate epidermal growth and differentiation (Eichner et al., 1996). The peroxisome proliferator activator receptor (PPARx) accelerates epidermal maturation and farnesol stimulates epidermal keratinocyte differentiation via PPARx (Hanley et al., 2000). An increase in farnesol concentration increases the rate at which the cornified envelope, the final product of terminal differentiation, is formed (Hanley et al., 2000). This results in an increased expression of the two most abundant proteins expressed in the granular layer of the epidermis, loricrin and filaggrin (Yoneda et al., 1993).

An increased concentration of loricrin has no unusual morphology or evident pathology in any tissue including the epidermis (Yoneda et al., 1993). However, an overabundance of filaggrin monomers was found to cause the disruption of cell-cell adhesion in keratinocytes (Presland et al., 2001) and increase keratinocyte susceptibility to apoptotic cell death (Keuchle et al., 2000). The hydrolysis of filaggrin into amino acids is highly regulated because the amino acids contribute to the water-holding properties of the stratum corneum (Jackson et al., 1993). Filaggrin expression is disrupted in ichthyotic disorders, disorders characterized by a markedly thickened stratum corneum (Kuechle et al., 2000) which is one of the most noticeable characteristics of KWE.

Farnesol induces an increase in the protein and mRNA levels of involucrin and transglutaminase, early markers of keratinocyte differentiation (Hanley et al., 2000). Therefore an increase of farnesol would lead to an overabundance of these markers of

differentiation. It becomes apparent that a disruption in the mevalonate pathway at the point of squalene synthase, which results in an increase in the amount of farnesol, could disrupt the balance of involucrin, filaggrin and loricrin (Simon et al., 1996). These essential proteins are required for epidermal differentiation but an overabundance of any could result in abnormal keratinocyte differentiation and hypoproliferation. It has been reported that the analysis of the lipid composition of the outer stratum corneum is a promising approach to study the pathophysiology of inherited disorders of keratinisation (Kuster et al., 2002).

1.5.2 Cathepsin B (*CTSB*)

The molecular processors responsible for controlling epidermal keratinocyte proliferation and differentiation are not fully understood. Since epidermal homeostasis is highly dynamic with regard to proliferation, differentiation and migration, it is likely that proteases are essential in order to carry out these processes (Reinheckel et al., 2005). Cathepsins are a family of lysosomal proteases that are involved in intracellular protein degradation (Schwartz et al., 2002). Cathepsin B, C, F, H and L are ubiquitously expressed in mammalian tissue while cathepsin K, S, V and W are restricted to specific cell types (Reinheckel et al., 2005).

The protein encoded by *CTSB* is a lysosomal cysteine proteinase composed of a dimer of disulfide-linked heavy and light chains, both produced from a single protein precursor (Buth et al., 2004). It is also known as amyloid precursor protein secretase and plays a role

in the differential regulation of amyloid precursor protein (APP) proteolytic processing (Esch et al., 1990; Klein et al., 2008). *CTSB* is located on chromosome 8p23.1-p22, contains 11 exons, encodes 339 amino acids and belongs to the peptidase family C1, a family of proteins that function to hydrolyze other proteins (Berquin et al., 1995). *CTSB* has ubiquitous expression and is therefore referred as a "housekeeping" gene (Reinheckel et al., 2005). Cathepsin B consists of one major conserved domain referred to as the carboxydipeptidyl domain (Figure 1.8). It is through this conserved domain that the CTSB protein functions primarily as an exopeptidase. Together with the other cathepsins, cathepsin B is also involved in antigen processing (Deussing et al., 1998), proenzyme activation and apoptosis (Foghsgaard et al., 2001). Overexpression of the encoded protein has been associated with esophageal adenocarcinoma and tumors of the breast, colon, prostate and the lungs (Hughes et al., 1998).

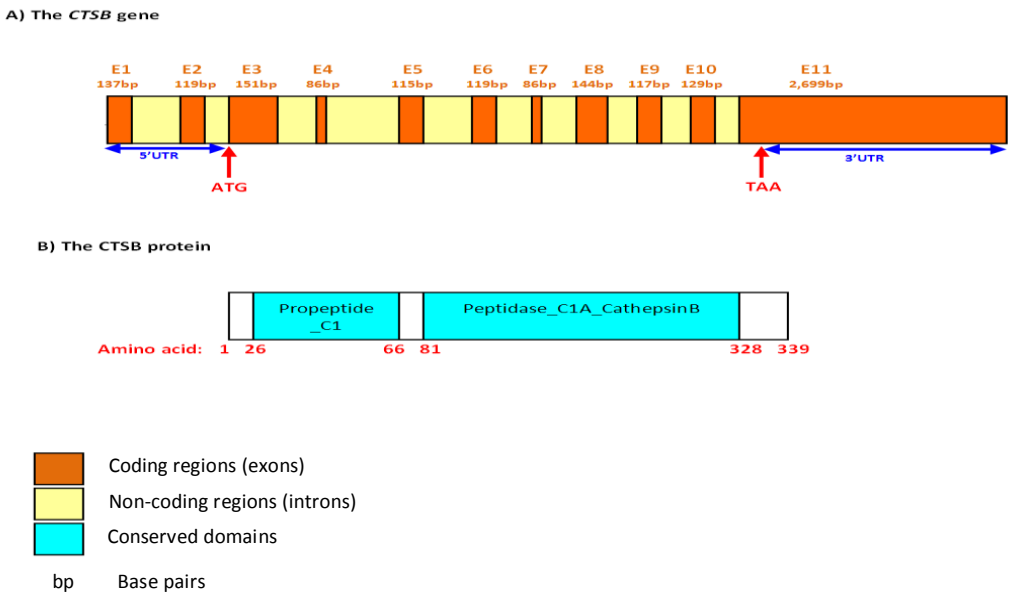


Figure 1.8: A schematic overview of the *CTSB* gene (A) and CTSB protein (B)

Cathepsin B is present within vesicles of cellular protrusions forming cell-cell contact sites keratinocytes of the stratum spinosum of human skin (Buth et al., 2004). This observation suggested that the secretion of *CTSB* into the pericellular spaces could be involved in the dissociation of cell-cell contacts enabling keratinocyte migration. It was demonstrated that the migration ability of keratinocytes was reduced due to the blockage of functional cathepsin B (Buth et al., 2004). The unobstructed migration of keratinocytes through the multilayered epidermis is essential in the process of epidermal differentiation and keratinisation. Also as mentioned before, migrating keratinocytes are essentially important in the repair mechanism of a disrupted epidermal barrier and failure of the keratinocytes to migrate to the wounded epidermis results in the skin becoming inflamed and scaly (Haratake et al., 2006).

The fact that many cathepsins have been implicated in disorders and processes involving the epidermis, desquamation and palmoplantar keratosis (Hughes et al., 1998), was a persuasive second reason to consider *CTSB* as the second most likely KWE candidate. Cathepsin C has been implicated in Papillon-Lefevre syndrome, a syndrome displaying palmoplantar keratosis (Cagli et al., 2005). This syndrome is caused by mutations occurring in the *CTSC* gene which introduce premature stop codons. *CTSC* encodes dipeptidyl-peptidase I, a lysosomal cysteine proteinase which functions primarily in immunity and inflammation (Selvaraju et al., 2003). Cathepsin G is thought to display chymotryptic activity and might play a role in desquamation (Schechter et al., 1983).

Cathepsin L is broadly expressed in epithelial cells of the skin and has a function that is specific to the regulation of keratinocyte proliferation (Reinheckel et al., 2005). Major insights into the function of *CTSL* in skin result from the analysis of $ctsl^{-/-}$ mice. Cathepsin L-deficient mice develop periodic hair loss, gingival acanthosis and epidermal hyperplasia with hyperkeratosis (Nishimura et al., 2002). Hyperproliferation of basal epidermal keratinocytes is the primary characteristic of the $ctsl^{-/-}$ phenotype (Reinheckel et al., 2005). Cathepsin V has a specific location in the stratum granulosum and in the root sheets of the hair follicle (Zeewen et al., 2007). This protease is secreted into the extracellular space where it associates with corneodesmosomes, suggesting a possible role in epidermal desquamation (Cheng et al., 2006). Cathepsin S is thought to have an immunogenic function in keratinocytes, which are an integral part of the skin's immune system (Schwartz et al., 2002) and Cathepsin E, an endolysosomal aspartic proteinase, is predominantly expressed in immune system cells. Deficiency of cathepsin E is associated with the development of atopic dermatitis, a pruritic inflammatory skin disease (Chou, 2002). The Cathepsin D-like proteinase is one of the enzymes leading to desquamation of the stratum corneum (Horikoshi et al., 1999).

1.5.3 Deubiquitinating enzyme (*DUB-3*)

Deubiquitinating enzymes, also termed DUBs, are a group of proteins capable of producing a free ubiquitin moiety from ubiquitin-fused precursors and ubiquitinylated proteins in cells (Wilkinson, 1997). DUBs possess conserved catalytic regions known as

the 'cys' and the 'his' domain. There are 2 distinct families of DUBs namely ubiquitin-specific processing proteases (UBPs or USPs) and ubiquitin C-terminal hydrolases (UCHs) (DeSalle and Pagano, 2001). The DUB enzymes act at several points in the ubiquitin pathway namely in polyubiquitin precursor processing, the removal of ubiquitin from substrates to rescue them from degradation and the removal of residual ubiquitin to aid proteasomal degradation (Burrows et al., 2004). The ubiquitin proteasome pathway plays an important role in many molecular cell mechanisms such as cell division and cell-cycle regulation, transcription, differentiation and development and apoptosis regulation (Schiffer, 2006), and also plays a central role in the degradation of short-lived and regulatory proteins that are important in basic cellular processes such as the regulation of the cell cycle, modulation of cell surface receptors and ion channels, and antigen processing and presentation (Ciechanover and Schwartz, 1996). The pathway employs an enzymatic cascade by which multiple ubiquitin molecules are attached to the protein substrate and polyubiquitin marks the protein for destruction.

DUB-3, a human member of the DUB family of deubiquitinating enzymes has been successfully cloned (Burrows et al., 2004) and is expressed at mRNA level in many different tissues including heart, spleen, kidney, colon, skeletal muscle and liver. DUB-3 is a cytokine-inducible protein that is expressed transiently and plays a role in blocking cell proliferation and inducing apoptosis. *DUB-3* consists of one exon, does not have a known 5’ or 3’UTR and encodes a protein that is 530 amino acids in length (Figure 1.9). *DUB-3* has two conserved domains, UBP5 and Peptidase_C19E. UBP5 represents the ubiquitin C-

terminal hydrolase which is thought to be responsible for the posttranslational modification, protein turnover and chaperone activity of *DUB-3.* Peptidase_C19E forms part of the Peptidase C19 family which contains ubiquitinyl hydrolases. These are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleaving isopeptide bonds (Pickard, 2001). The purpose of deubiquitination is thought to be a means by which ubiquitin conjugates are rescued from degradation and also a means of recycling ubiquitin (Wu et al., 2003). The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome (Scheffner et al., 1995).

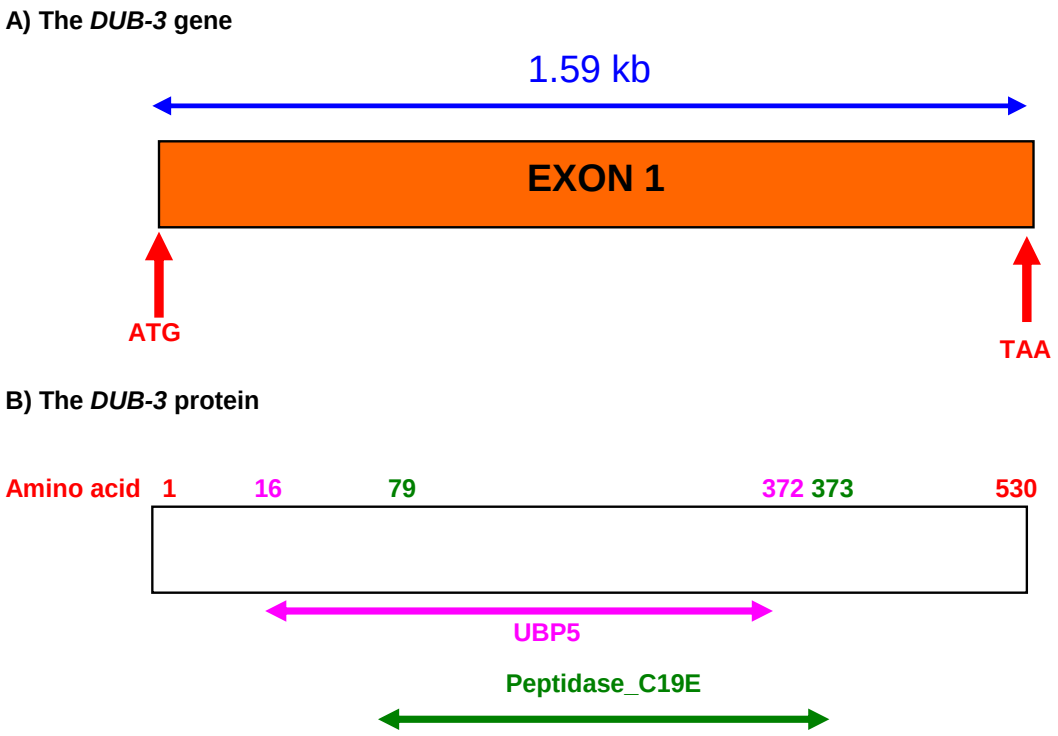


Figure 1.9: A schematic overview of the *DUB-3* gene (A) and the predicted functional domains of the protein (B)

The transient expression of *DUB-3* is regulated by the cytokines interleukin 4 (IL-4) and interleukin 6 (IL-6). Cytokines are polypeptide growth factors that function primarily to provide a cell-cell communication system between cells. Cytokines mediate and regulate immunity and inflammation and are produced *de novo* in response to an immune stimulus (Decker, 2000). Such stimuli include an increase or decrease in the expression of membrane proteins (including cytokine receptors), proliferation, and the secretion of effector molecules (Oyama et al., 1999). Cytokines are classified according to origin, function and chemical structure as interleukins (IL), colony-stimulating factors (CSF), interferons (IFN), growth factors (GF) and tumor necrosis factors (TNF) (Feliciani et al., 1996) and are very important for skin permeability repair (Wang et al., 2004). IL-4 has many roles in the immune system. Although IL-4 is not produced by keratinocytes or mucosal epithelium, they are important mediators in disorders where eosinophilia is prominent such as atopic dermatitis and autoimmune bullous disorders (Chan et al., 2001). IL-6 is a pleiotropic cytokine that is the major mediator of the response to tissue injury and infection and is involved in the growth and differentiation of numerous cell types, including those of dermal and epidermal origin (Sehgal, 1990; Gallucci et al., 2004). IL-6 is produced by keratinocytes, monocytes, Langerhans cells, fibroblasts and melanocytes (Grossman et al., 1989; Paquet and Pierard, 1996) and stimulates the proliferation and differentiation of keratinocytes, hepatocytes and nerve cells. IL-6 plays a role in the immune system where it can promote Th2 development and B-cell differentiation. Increased levels of IL-6 have been associated with a number of skin diseases namely psoriasis (Grossman et al., 1989), scleroderma and systemic lupus

erythematosus (Koch et al., 1993). Psoriasis is a common skin disease that is characterized by increased epidermal thickness, rapid proliferation of keratinocytes, altered keratinocyte differentiation and inflammation. The histopathology of psoriasis holds some similarity to that of KWE. It has been hypothesized that the phenotype of psoriasis results from an increase in local cytokine production by keratinocytes (Grossman et al., 1989). In 2004, Gallucci et al. demonstrated that the over expression of IL-6 in the skin of normal rats induces epidermal proliferation and inflammation.

Cytokines in the skin

Not only does skin function as a physical barrier that protects the body from the outside environment but it also plays an important role in modulating interactions between the environment and the body. As previously mentioned (Section 1.1) keratinocytes are the major cell type found in skin. The keratinocytes are organized as stratified epithelia, the morphology dependant on function. Palmoplantar skin is more prone to mechanical stress and for this reason is more resistant to trauma than trunk skin (Feliciani et al., 1996). This difference between palmoplantar and trunk skin is associated with differences in cytokines. The cytokines that are produced by these epithelial cells function in the maintenance of normal homeostatic mechanisms in the skin. Skin cytokines have a pro-inflammatory and anti-inflammatory function and can therefore induce proliferative changes in the skin that result from injury (Feliciani et al., 1996). Any dysfunction or abnormality in the cytokine balance in the skin can contribute to inflammatory diseases in

the skin. Epidermal cytokine dysregulation has been reported to play a role in many diseases, many of which are skin diseases or are diseases that affect the skin (Table 1.1).

Table 1.1: A list of diseases in which epidermal cytokine dysregulation has been reported to play a role (Feliciani et al., 1996)

Psoriasis	IL-1, IL-6, IL-8, TGF-a, PDGF, EGF
Cutaneous T-cell lymphoma	IL-1, IL-8
Kaposi's sarcoma	IL-1, bFGF
Atopic dermatitis	IL-4, IL-5
Pemphigoid	IL-4, IL-5
Pemphigus	IL-1, TNF-a
Lupus erythematosus	IL-1, IL-6, TNF-a
Melanoma	IL-6, IL-8, MGSA
Squamous cell carcinoma	IL-1, EGF, TGF-a

DUB-3 is an attractive candidate gene, not only because of its location within the human genome, but also due to the its role in blocking cell proliferation and inducing apoptosis (Burrows et al., 2004). Apoptosis is a highly conserved basic physiological process in which dead unwanted cells are removed without inducing an inflammatory response. In self-renewing tissue like the epidermal layers of the skin, cell numbers are regulated by a delicate balance between proliferation, differentiation and cell death (Reefman et al., 2005). In some human skin diseases such as psoriasis, apoptosis is thought to play a role in the development of the disease (Chen et al., 2006). The induction of apoptosis results in an increase in the amount of dead skin cells which need to be removed from the skin's stratum corneum. If for some reason, the balance between proliferation, differentiation and apoptosis is disrupted, it could lead to the development of abnormalities. It has been demonstrated that the transient expression of *DUB-3* is regulated by the cytokines

interleukin 4 (IL-4) and interleukin 6 (IL-6). Any form of stress to the palmoplantar surfaces of the skin will result in the stimulation and release of cytokines, including IL-4 and IL-6, which will stimulate the transient expression of *DUB-3.* The constitutive expression of *DUB-3* will block growth factor-dependent cell proliferation and induce apoptosis. *DUB-3* may function to regulate immune responses by influencing cell proliferation and survival (Burrows et al., 2004).

1.5.4 Open reading frame 49 (*C8orf49*)

C8orf49 is a putative uncharacterized protein and for this reason there is minimal information available relating to this open reading frame. It is hypothesized that *C8orf49* encodes a single-pass membrane protein made up of 230 amino acids. This single exon open reading frame was predicted using a GeneWise/Exonerate model (Curwen et al., 2004) and was analysed in this study because of its location within the human genome (KWE critical region) and due to the presence of two upstream open reading frames (uORFs) in the *C8orf49* mRNA 5'UTR.

1.6 MUTATION DETECTION STRATEGY

1.6.1 Analysing the coding regions of candidate genes

Analysing the coding regions of candidate genes is the most common way of detecting genetic variants that could be responsible for the production of an abnormal protein and the subsequent manifestation of a disease phenotype (Burchard et al., 1999). Variation in

the coding region of a gene, the region that is translated into the protein, is expected to have a bigger effect on the protein than a variant occurring in a region that is non-coding i.e. introns. However, there are exceptions. It is important to include the exon–intron boundaries in analyses in order to detect splice site mutations. Because the promoter region of a gene is the region of DNA that facilitates the transcription of that particular gene, it is important to detect whether there are any genetic variants that occur within this region that may affect the transcription of the gene. There are a number of diseases that manifest due to malfunction of the promoter of a gene, either through direct mutation of a promoter sequence or mutation in a transcription factor (Kulozik et al., 1991; Burchard et al., 1999).

1.6.2 Analysing the 5' untranslated regions (5'UTRs) of candidate genes

With the recent and continuing discoveries of new mechanisms of gene regulation, it becomes clear that our understanding of this complex and ubiquitous biological process remains incomplete (Morris and Geballe , 2000). There have been many recent discoveries of upstream open reading frames (uORFs) in the 5'UTRs of genes and their possible role in regulating mRNA translation. This prompted the inclusion of the analysis of the candidate genes' 5'UTRs. In the process of translation, the small subunit of the ribosome binds to the 7-methyl guanosine cap at the 5' of the mRNA and scans the 5'UTR until it encounters the first AUG codon (Kozak, 1999). This interaction between the ribosome and the mRNA initiates translation. The 5'UTR of vertebrate mRNAs vary considerably in length and contain a variety of features which affect the efficiency of translation of the main coding

sequence (Meijer et al., 2002; Crowe et al., 2006). These features include the presence of upstream AUG codons. These AUG codons are termed upstream AUGs if they do not have an associated in-frame stop codon and in the presence of an in-frame stop codon, they are classified as upstream open reading frames or uORFs (Morris and Geballe, 2000). Figure 1.10 highlights the differences between an uAUG and an uORF. Although there is a mass of data that suggests that uORFs and uAUGs regulate gene expression by modifying mRNA translation (Morris and Geballe, 2000; Meijer and Thomas, 2002), an accurate characterization of the features involved in the activity of the uORF has only been made for a few genes (Iacono et al., 2005; Morris and Geballe, 2000) and the mechanisms by which they can modulate mRNA translation have only been partially described (Iacono et al., 2005).

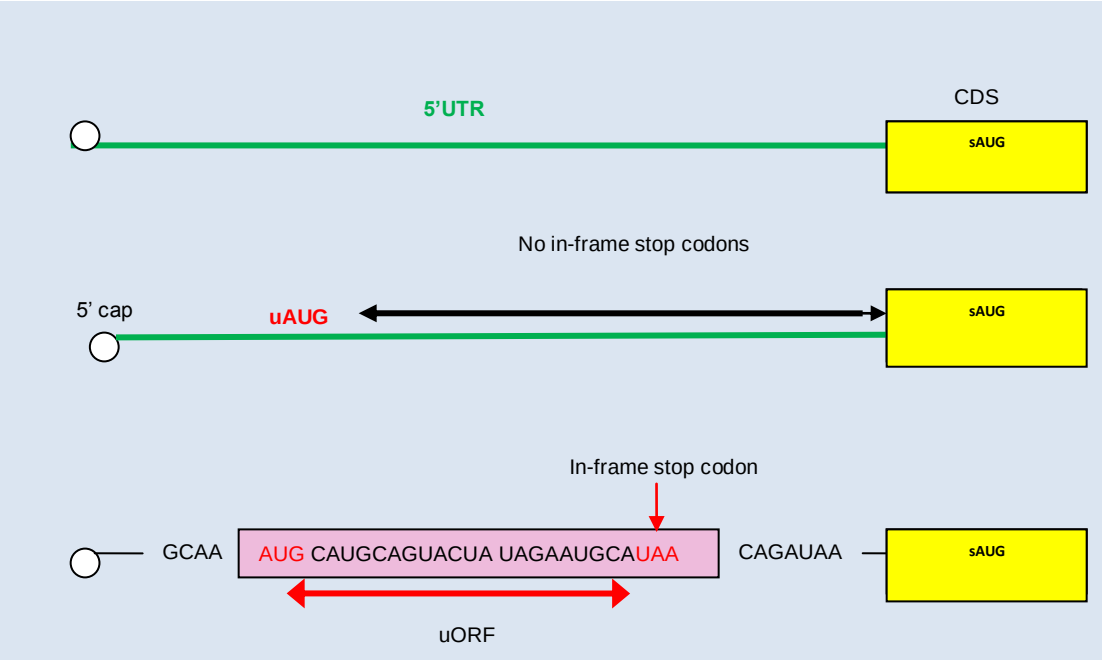


Figure 1.10: A schematic overview of mRNA illustrating the differences between an upstream AUG (uAUG) and an upstream open reading frame (uORF) as defined by Iacono et al. (2005) CDS – coding sequence; sAUG – start codon in coding region

The mechanisms postulated as to how uORFs and uAUGs can modulate mRNA translation always have an end result of a decrease in the efficiency of translation of the main coding sequence. In some cases it has been reported that these above mentioned mechanisms completely eliminate the translation of the main coding sequence although it is more common that there is between a 2 and 50-fold reduction in protein levels (Iacono et al., 2005). For this reason, it has been postulated that uAUGs and uORFs may represent a method of post-transcriptional regulation, through the repression of translation of the main coding sequence. There are a number of diseases that are associated with uORFs and uAUGs. These diseases result due to mutations that either eliminate or introduce uAUGs, resulting in a consequent increase or decrease in translational efficiency. Also mutations may change the splicing patterns of the 5'UTR to alter the number of uAUGs. For example, the main coding sequence of the oncogene *mdm2* in normal human cells contains two upstream open reading frames in its 5'UTR. The elimination of these uORFs raises the translational efficiency of the transcript by over 10-fold in He-La cells, therefore these uORFS have a fundamental role in regulating the expression of the *mdm2* gene (Jin et al., 2003). Also, it was determined that the translation of *TPO* mRNA is strongly inhibited by the presence of several uAUG codons in the 5'UTR. Directed mutagenesis of all the uAUGs in the *TPO* mRNA restored full translational efficiency (Ghilardi et al., 1998), again highlighting the important role of uORFs in translational control.

1.7 RELATIVE GENE EXPRESSION USING REAL-TIME RT-PCR

Real-time RT-PCR was developed because of the need to quantify differences in mRNA expression and because of the availability of only small amounts of mRNA in some procedures (http://www.ambion.com/techlib/basics/rtpcr/index.html). Real-time RT-PCR is a technique for collecting data throughout the PCR process as it occurs, thus combining amplification and detection into a single step (Wong and Medrano, 2005). The real-time PCR technique, where the fluorescent signal reflecting the PCR accumulation is detected in every amplification cycle, has simplified the quantification of nucleic acid using PCR (Ciskos et al., 2007). PCR reactions are characterised by the cycle where the target amplification is first detected. This cycle is referred to as the cycle threshold (C_q) (Wong and Medrano, 2005). Several different types of real-time PCR are being marketed to the scientific community at this time, each with their advantages and disadvantages. There are currently four different chemistries available for real-time PCR namely; TaqMan® (Applied Biosystems, Foster City, CA, USA), Molecular Beacons, Scorpions® and SYBR® Green (Roche). Each of these methods allows detection of PCR products via the generation of a fluorescent signal (http://www.ambion.com/techlib/basics/rtpcr/-index.html). TaqMan probes, Molecular Beacons and Scorpions depend on Förster Resonance Energy Transfer (FRET) to generate the fluorescence signal via the coupling of a fluorogenic dye molecule and a quencher moiety to the same or different oligonucleotide substrates. (http://www.appliedbiosystems.com/support/tutorials/pdf/rtpcr_vs_tradpcr.-pdf). In this project, the SYBR green method was used to monitor cDNA synthesis. This is the simplest method for detecting and quantify PCR products in real-time reactions.

SYBR green is a fluorogenic dye that exhibits little fluorescence when in solution, but emits a strong fluorescent signal upon binding to double-stranded DNA (Dorak, 2006). Therefore, as the PCR product accumulates, the fluorescence increases. This method is sensitive and easy to use, however, SYBR green binds to any double-stranded DNA in the reaction, including primer-dimers, which results in an overestimation of the target concentration. Therefore this method requires extensive PCR optimization (http://www.ambion.com/techlib/basics/rtpcr/index.html).

There are two ways that are commonly employed to quantify the results obtained from real-time PCR; the standard curve method (absolute quantification) and the comparative threshold method (relative quantification) (Ferreira et al., 2006). Absolute quantification uses serially diluted standards of known concentrations to create a standard curve. This curve is then used as a reference standard for extrapolating quantitative information for mRNA targets of unknown concentrations (Wong and Medrano, 2005). Relative quantification (RQ) determines the changes in mRNA levels of a target gene across multiple samples (Cikos et al., 2007) and expresses it relative to the levels of a reference gene across the same samples (Wong and Medrano, 2005). A calibrator corrects for differences in detection sensitivity between the target and reference gene caused by differences in primer annealing and dye extinction coefficients. It also corrects for sample-to-sample variation (personal communication with a Roche technician). In this way, the calibrator acts to normalise the data between the target samples and reference samples as well as between run-to-run repeats. There are many different methods that are used to relatively quantify real-time PCR results, namely the Comparative $\Delta\Delta C_q$ method, the Pfaffl

method, Q-gene, the Gentle et al. method and the Liu and Saint method (Wong and Medrano, 2005). In this study, real-time data were analysed using the Comparative $\Delta\Delta C_q$ method.

1.8 PROJECT OBJECTIVES

The aim of this project was to identify and characterise the gene that causes keratolytic winter erythema. The primary objectives of the project included:

1. To validate the KWE critical region and identify possible informative recombination events within KWE families in an attempt to further reduce the size of the KWE critical region.

2. To identify the most likely candidate genes in the KWE critical region (8p22.1-p23) and to perform thorough analysis through the sequencing of the 5'UTRs, exons and flanking sequences of the candidate genes and the mRNA of the two most likely candidates, *FDFT1* and *CTSB.*

3. To compare the gene expression profiles of *CTSB* and *FDFT1* in non-affected and KWE affected palmar skin.

CHAPTER 2

MATERIALS AND METHODS

2.1 SUBJECTS

A total of 45 samples were obtained from KWE affected (N=19) and unaffected (N=26) individuals, sampled from South Africa (Cape Town, Johannesburg) The unaffected or control group included individuals who are directly related to the patients and also individuals who are unrelated to the KWE families (KWE family pedigrees listed in Appendix A). The samples were collected from three different tissues, namely peripheral blood, saliva and palmar skin. The nucleic acids were extracted from these samples and used for the various aims of the study. Table 2.1 gives a description of how the samples were used in the study.

Table 2.1: A summary of the number of samples collected and their specific use in the present study

Tissue	No. of samples obtained	Nucleic acid extracted	Sample use in study
Blood	25	DNA and RNA	Haplotype analysis, DNA and mRNA sequencing
Saliva	4	DNA and RNA	DNA and mRNA sequencing
Epidermal Palmar skin	16	RNA	Quantitative analysis of gene expression
	45		

Once informed written consent was obtained (consent forms can be found in Appendix I), 10ml of venous blood was taken from adult participants and approximately 2ml of saliva were collected from participants younger than the age of ten. The blood samples were collected in EDTA tubes and the saliva samples were collected in the

Oragene·DNA/Oragene·RNA saliva kit collection vials (DNA Genotek). The skin biopsies were performed for the extraction of RNA for expression studies. A local anaesthetic was subcutaneously administered to the palm. A wedge biopsy measuring approximately 4mm in length was taken (Figure 2.1). The purpose of the wedge biopsy was to minimize the sampling of dermal tissue and to get as much epidermal tissue as possible from the biopsy. The biopsies were cut into smaller chunks and placed in cryo tubes which were immediately stored in liquid nitrogen. Once in the laboratory they were kept frozen at -80°C until needed. Mild cauterisation was performed on the wound to minimize bleeding and accelerate healing. Ethics clearance for the use and collection of these samples was obtained from the Human Research Ethics Committee, ethics approval number M050706 (See Appendix H).



Figure 2.1: The wedge biopsy taken from the palmar region of a KWE patient

2.2 NUCLEIC ACID EXTRACTION AND QUANTIFICATION

2.2.1 DNA

DNA was extracted from the whole blood samples using the salting out method (Miller et al., 1988). This extraction was carried out by the diagnostic division in the laboratory. The DNA from the saliva samples was extracted following the protocol provided by the OraGene·DNA kit manufacturers (Complete protocol in Appendix C.I).

2.2.2 RNA

Total RNA was extracted from the whole blood samples using the commercially available QIAamp RNA blood kit (Qiagen)(Complete protocol in Appendix C.II). The average total RNA yield per 1.5ml of blood collected in EDTA tubes is 1-5µg of total RNA. Total RNA from the saliva samples was extracted following the protocol provided by the OraGene·RNA kit manufacturers (Complete protocol in Appendix C.III). Total RNA was extracted from the palmar skin biopsies using the RNeasy Micro Plus kit supplied by Qiagen (Complete protocol in Appendix C.IV).

The quality of the DNA and RNA was evaluated by running it on a 0.8% agarose gel. The concentration (ng/µl) and the purity ($OD_{260/280}$) of the DNA and RNA was measured using the Nanodrop® ND-1000 Spectrophotometer (Nanodrop Technologies). DNA samples were diluted to a final concentration of 100ng/µl using 1X TE buffer and stored at -20°C. RNA samples were stored at -80°C.

2.3 THE POLYMERASE CHAIN REACTION (PCR)

This technique, first described by Mullis et al. (1987), involves the amplification of a specific region of genomic DNA. Amplification is achieved by successive cycles of heating and cooling to temperatures specific for denaturation of double stranded DNA, annealing of the primers and the extension of the new product. Conventional PCR was used to amplify the microsatellite markers, the coding sequences and 5' untranslated regions (5'UTRs) of the candidate genes as well as the mRNA of *CTSB* and *FDFT1*. The mRNA obtained from the palmar skin was amplified using real-time PCR in order to perform a quantitative analysis of two of the candidate genes expression in affected and unaffected skin.

2.3.1 Reverse transcriptase PCR

Prior to PCR, the total RNA extracted from the blood, saliva and skin samples was reverse transcribed into mRNA using the ImProm-II™ Reverse Transcriptase kit (Promega). The oligo$(dT)_{15}$ primers supplied with the kit were used to amplify total mRNA (Appendix D). The cDNA of *CTSB* and *FDFT1* was amplified using gene-specific primers in a conventional PCR system.

2.3.2 Conventional PCR

Conventional PCR was used to amplify each microsatellite marker (Protocol in Table 2.2), the DNA of *CTSB*, *FDFT1*, *DUB-3* and *C8orf49* and the cDNA of the two most likely

candidate genes, *CTSB* and *FDFT1* (Protocol in Table 2.3). The optimised long template expand protocol used to amplify the 1.85kb *DUB-3* fragment is listed in Table 2.4. PCR was carried out in a GeneAmp® PCR System 2700 (Applied Biosystems).

Table 2.2: The general PCR protocol used to set up the amplification systems for each microsatellite marker

PCR Components	**Initial concentration**	**Final concentration**	**Volume (µl)**
DNA	N/A	30ng/µl	1
dNTPs (Bioline)	1.25mM	0.15mM	2.5
10X Buffer (Applied Bioystems)	10X	1.25X	2.5
$MgCl_2$ (Applied Biosystems)	25mM	2.5mM	2
Primer forward – *labeled*	10mM	0.5mM	1
Primer reverse	10mM	0.5mM	1
Amplitaq Polymerase (Applied Biosystems)	1U/µl	0.01U/ µl	0.2
ddH_20	N/A	N/A	9.8
TOTAL			**20**
Thermal cycling conditions			
Cycle step	**Temperature (°C)**	**Time**	**No. of cycles**
Initial Denaturation	95	5 minutes	1
Denaturation Annealing Extension	95 X[1] 72	15 seconds 30 seconds 30 seconds	25
Final extension	72	7 minutes	1
Hold	4	∞	

[1] X is the annealing temperature appropriate for every primer pair

Table 2.3: The general PCR protocol used to amplify the 5'UTRs and exons of the candidate genes as well as the cDNA of *FDFT1* and *CTSB*

PCR Reagent	Initial concentration	Final concentration	Volume (µl)
DNA/cDNA	N/A	100ng/µl	1
dNTPs (Bioline)	1.25mM	0.125mM	2.5
10X buffer (Applied Biosystems)	10X	1X	2.5
$MgCl_2$ (Applied Biosystems)	25mM	2mM	2
Forward primer	10mM	0.4mM	1
Reverse Primer	10mM	0.4mM	1
Amplitaq Gold polymerase (Applied Biosystems)	5U/µl	0.04U/µl	0.2
ddH_2O	N/A	N/A	14.8
TOTAL			25
Thermal cycling conditions			
Cycle step	**Temperature (°C)**	**Time**	**No. of cycles**
Initial Denaturation	95	5 minutes	1
Denaturation Annealing Extension	95 X[1] 72	15 seconds 30 seconds 30 seconds	35
Final extension	72	7 minutes	1
Hold	4	∞	

[1] X is the annealing temperature appropriate for every primer pair

Table 2.4: The optimized Long Template Expand protocol used for the amplification of the large 1.85kb *DUB-3* fragment

PCR components	Stock concentration	Final concentration	Volume (µl)
DNA	N/A	100ng/µl	3
Long expand template buffer 2 (Roche)	10X	1X	2.5
Betaine (Sigma)	5M	1M	5
dNTPs (Bioline)	1.25mM	0.125mM	2.5
Primer Forward	10mM	0.8mM	2
Primer Reverse	10mM	0.8mM	2
Long expand template enzyme (Roche)	5U/µl	0.04U/µl	0.2
ddH_20	N/A	N/A	7.8
TOTAL			**25**
Thermal cycling conditions			
Cycle step	**Temperature (°C)**	**Time**	**No. of cycles**
Initial denaturation	95	7 minutes	1
Denaturation Annealing Extension	95 58 72	90 seconds 90 seconds 150 seconds	35
Final Extension	72	10 minutes	1
Hold	4	∞	

2.3.3 Real-time RT-PCR

Real-time RT-PCR was used to amplify and quantitfy the mRNA expression of *CTSB* and *FDFT1* in the epidermal palmar skin of KWE affected and unaffected individuals. The real-time PCR protocol in listed below in Table 2.5. Real-time PCR carried out on LightCycler ® 2.0 Real-Time PCR machine (Roche).

Table 2.5: The general PCR protocol used for the real-time amplification of *FDFT1* and *CTSB* cDNA

PCR components	Stock concentration	Final concentration	Volume (µl)
cDNA	N/A	500ng/µl	2
SYBR green mix (Applied Biosystems)	N/A	N/A	4
Primer Forward	10mM	0.8mM	1
Primer Reverse	10mM	0.8mM	1
ddH_20	N/A	N/A	12
TOTAL			**20**

Thermal cycling conditions

Analysis mode	Cycles	Segment	Temperature (°C)	Time	Acquisition mode
Hot Start	1		95	10 min	None
PCR	45	Denaturation Annealing Extension	95 X[1] 72	10 sec 10 sec 25 sec	None None Single
Melting curve	1	Denaturation Annealing Melting	95 65 95 (slope = 0.1)	0 15 sec 0	None None Continuous
Cooling	1		40	30 sec	None

[1] X is the annealing temperature appropriate for every primer pair

2.3.4 Primer design

The primer sequences used for the amplification of the microsatellite markers were taken directly from the Ensembl website (www.ensembl.org) and are listed in Table 2.6. The primers used to amplify the coding sequences and 5'UTRs of *DUB-3*, *FDFT1* and *C8orf49* were designed using Primer3 (http://frodo.wi.mit.edu/primer3/), a computational

program which enables the user to design a compatible primer pair for a specified region, e.g. flanking each exon. The entire genomic sequence of each candidate gene, taken from the NCBI 36 assembly of the human genome (November 2005), was copied into Primer3. This program has different input parameters that can be controlled and adjusted according to what is required by the researcher in respect to designing primers. Once the program had listed the possible primer pairs for each of the gene's exons and 5'UTRs, the primer sequences were copied into e-PCR Reverse, a program offered by NCBI (http://www.ncbi.nlm.nih.gov/sutils/e-pcr/). e-PCR Reverse BLASTS the primer sequence against the entire human genome to determine whether the chosen primer will anneal elsewhere in the genome. The primer sequences and annealing temperatures for *FDFT1*, *DUB-3*, and *C8orf49* are listed in Tables 2.7–2.9 respectively. The coding sequence and the 5'UTR of *CTSB* was amplified using the primer sequences and annealing temperatures published in Mahurkar et al., 2006 (Table 2.10). The overlapping primer pairs used to amplify *CTSB* and *FDFT1* mRNA were designed using Primer3. The primer pairs generated by Primer3 were BLASTed against a human mRNA reference sequence using the nucleotide BLAST tool offered by NCBI (http://www.ncbi.nlm.nih.gov/). These primers were designed in such a way that they contained sequence from two adjacent exons (Figure 2.2). This way of designing primers reduced the possible problem of amplifying contaminating DNA. The primer sequences and annealing temperatures of these overlapping primers are listed in Tables 2.11 and 2.12 for *FDFT1* and *CTSB* mRNA respectively. The real-time PCR oligonucleotide primers were manually designed for each

of the genes to ensure maximal efficiency and sensitivity. The forward and reverse primers designed for real-time PCR were designed in the same way as depicted in Figure 2.2. One of the requirements of real-time PCR primers is that they amplify a region smaller than 200bp. Table 2.13 lists the primers used in real-time amplification of *CTSB* and *FDFT1* (F = Forward primer sequence, R = Reverse primer sequence).

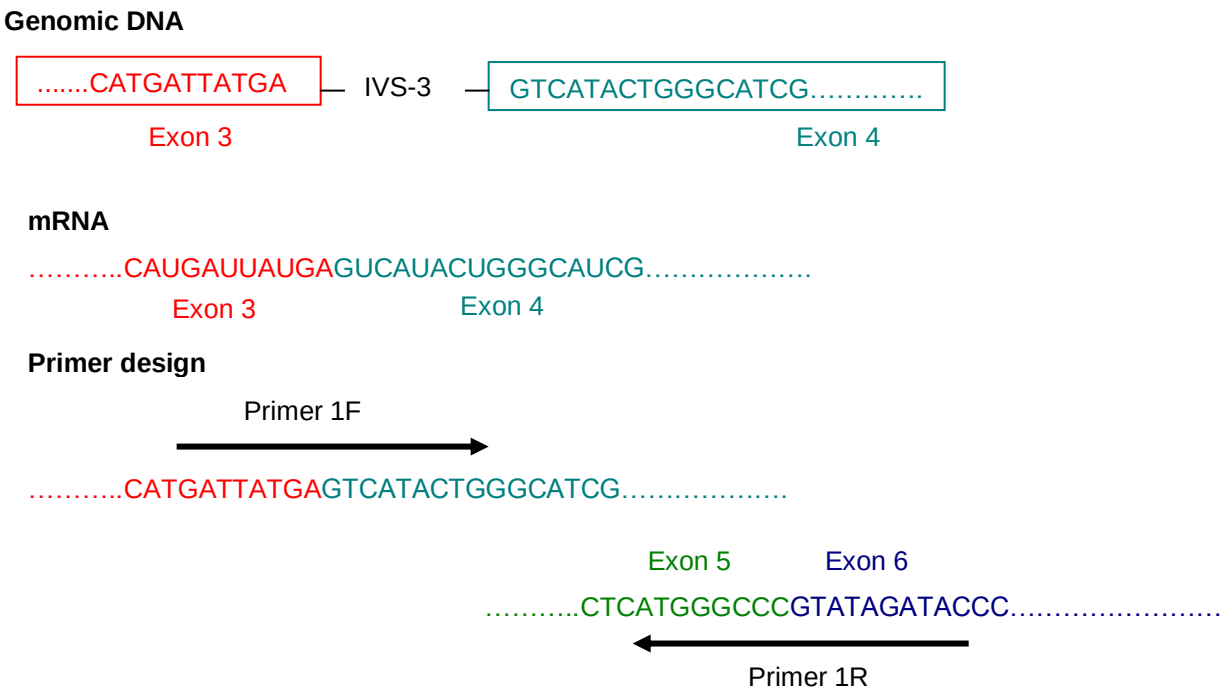


Figure 2.2: An illustration of the method used to design primers for amplification of *CTSB* and *FDFT1* mRNA

For the amplification of the *FDFT1* promoter (Figure 2.3), primers were designed to cover 600bp upstream to the start codon of the *FDFT1* reference gene. This region includes a 69bp region, located 131 nucleotides 5' to the transcription start site, which is thought to be the promoter region of this gene (Guan et al., 1995). This 69bp region contains a sterol

regulatory element (SRE-1) and two potential NF-1 binding sites. Two CCAAT boxes, another SRE-1 element and two Sp1 sites flank this region.

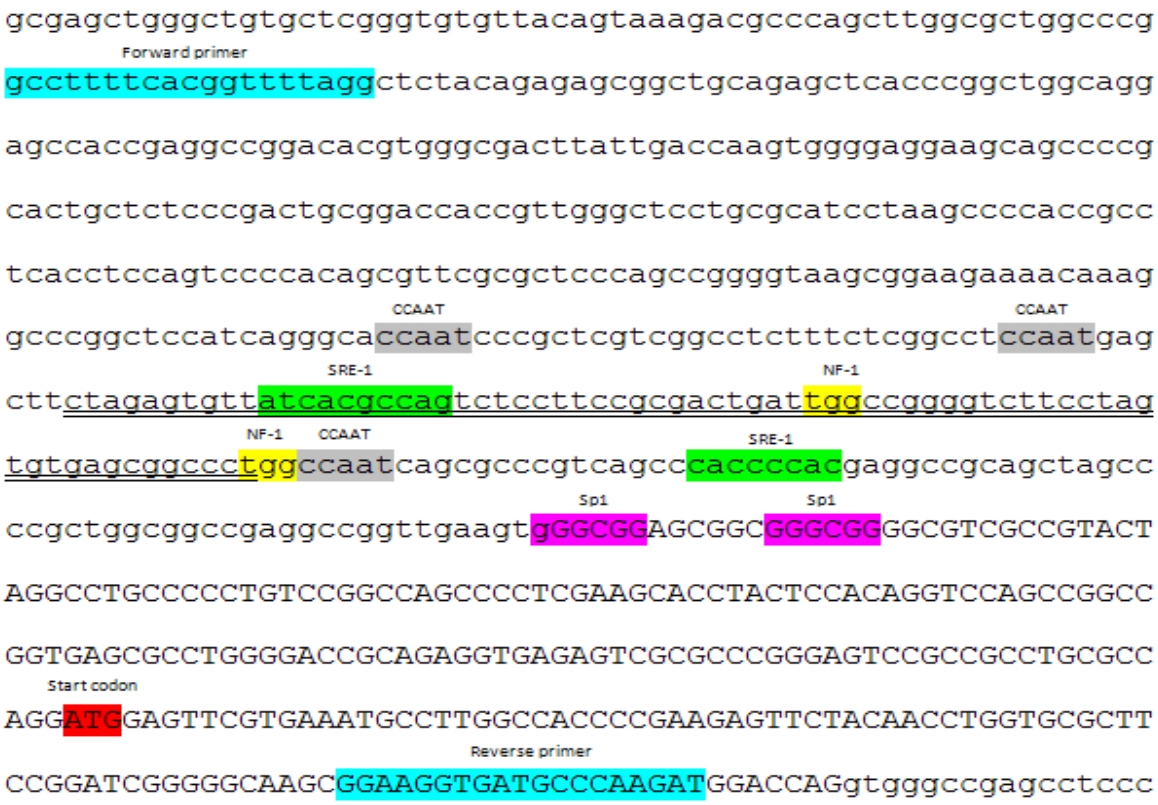


Figure 2.3: Nucleotide sequence of the *FDFT1* promoter. The putative transcriptional and regulatory elements are highlighted and identified above the sequence. The 69bp region is double underlined.

2.3.5 PCR optimisation

Throughout the study, PCR optimisation was required. Optimisation included setting up temperature gradients which tested temperatures 5-6°C above and below the initial calculated annealing temperature, varying the volume and/or concentration of the

general PCR components, increasing or decreasing the number of cycles of the reaction and/or the length of time that the particular cycle occurred for. Optimisation also included the addition of betaine, a PCR enhancer which improves the amplification of DNA and mRNA and also enhances the specificity of the polymerase chain reaction. For each PCR experiment set up, a blank sample (Xµl DNA replaced with Xµl ddH_20) was added in order to accurately detect contamination. The expiry dates of all reagents used in the PCR reaction were checked on a weekly basis to eliminate expired reagents as a cause of failed PCR. Temperature gradients were performed using the GeneAmp® PCR system 9270 (Applied Biosystems). Any changes made to the general conventional and real-time PCR protocols due to optimisation, are indicated in the tables which list the sequences and annealing temperatures of the primers used in the amplification of each gene (Tables 2.6 –2.13).

Table 2.6: The sequence and annealing temperature (T_A°C) of the primers used to amplify the markers used in the linkage analysis study

DNA Marker	Primer sequence 5' → 3'	T_A °C	Amplicon size (bp)	Changes made to general protocol
D8S1819 F D8S1819 R	TCACTGAGGGACTTGGC- FAM-6 CGTGCTGAGAATGAGACC	53	221	Add 2 µl DNA, Add 3.5 µl 5M betaine
D8S1755 F D8S1755 R	ACCGCATCTGGTCAACT – FAM-6 GCATCTCCATTGGAAGATTT	55	242	N/A
D8SDI F D8SDI R	TAGGGCTCTTCCGAGAAACA – HEX GCCTTCCCTCAGTTCTCAAG	59	180	N/A
D8STET F D8STET R	TGGATGTCTCATGGGCATTT – HEX CCCAACCTAAGAGCAGAACA	59	240	N/A
D8S265 F D8S265 R	ACCTCTTTCCAGATAAGCCC – HEX CCAATGGTTTCGGTTACTGT	57	217	N/A
D8S1695 F D8S1695 R	AACCCAGCATCCTACAAAG- FAM-6 CATCTGGAACCCATGAG	56	221	N/A
D8S1759 F D8S1759 R	GAGACTGACAATCTCCTCGTCTTAT- FAM-6 CTATTGCCTAGCTTAGCACATTTGA	60	123	Add 3.5 µl 5M betaine
D8S2657 F D8S2657 R	CAGAATACACAGATGGAGGG – HEX GAACCCAGGTGACTTCTTCC	57	236-240	N/A
D8S552 F D8S552 R	AGGATTGTAATTTCCTTGC – FAM-6 GGGACTTTTTGAAGGTTTG	53	168-180	N/A

Table 2.7: The sequence and annealing temperature (T_A°C) of the primers used to amplify *FDFT1*; and the size of the amplicons produced

Exon	Primer sequence 5' → 3'	T_A °C	Amplicon size (bp)	Changes made to general PCR protocol
Promoter F Promoter R	GCCTTTTCACGGTTTTAGG ATCTTGGGCATCACCTTCC	60	680	Add 2 µl DNA; Add 5 µl 5M betaine
1F 1R	CTCCCACTCCCACTCCCACTCCTG ACTGCCAGCTCTGTCTAATAC	59	254	Add 5 µl 5M betaine; extend thermal cycling times (30sec, 60sec, 60sec)
2F 2R	CTCCCACTCCCACTCCCACTCCTG ATGACCCCCTCCGCCCTCTTCTCT	66	397	Add 5µl 5M betaine
3F 3R	AATCTTTGGCTGTTTGTTC GAGTCTTGGCTACCGGAGTTTTTA	54	385	N/A
4F 4R	CCGCCATTGTTTGCCTTGTG ACTGCCAGCTCTGTCTAATAC	57	264	N/A
5F 5R	TGCATTCTTACCCATTCTCTTTT TTTTTCTGTCACCCGCAACC	56	394	N/A
6F 6R	GCC CAG CCT GTA TCA TAG TTC TTA TAG CAC AGC AAA ATC CAC CCT TAC	58	361	Add 5µl 5M betaine
7F 7R	GTT TTC TTA CCT TTT TGT GAT A TGC TCG GCT CCT GTG AAT AGA CTT	55	405	N/A
8F 8R	CAGTGGAGGGTTGTTGTAAGTAGC AAAATAGGAGCCGAACAGAAAGAA	58	719	N/A

Table 2.8: The sequence and the annealing temperature (T_A°C) of the primer pair used to amplify *DUB-3;* and the size of the amplicon produced. The sequence of the internal cycle sequencing primers is given

Region	Sequence 5' → 3'	T_A°C	Amplicon size (bp)
Exon 1F Exon 1R	TTGGTTTATCGCACACCTGA GGGTGTATTTGTGCGTGTGT	58	1853
Int.seq[(a)] 1F Int.seq[(a)] 2F Int.seq[(a)] 3F	GGGCTCCAGAATATGGGAAA CCACGGGATTTCAGACACTT GGTCACTGCCTGTAGCATCA	N/A	N/A
Int.seq[(a)] 1R Int.seq[(a)] 2R Int.seq[(a)] 3R	GACATGACCAGGACTGTGGA TGATGCTACAGGCAGTGACC CTCGCCTGTCTGTGTCTTCA	N/A	N/A

(a): Internal cycle sequencing primers are designed specifically for cycle sequencing and are not included in a PCR system. For this reason, an annealing temperature and amplicon size is not applicable.

Table 2.9: The sequence and the annealing temperature (T_A°C) of the primer pairs used to amplify *C8orf49.* The size of the amplicon produced is given in base pairs (bp)

Primer pair[(a)]	Sequence 5' → 3'	T_A°C	Amplicon size (bp)	Changes to general PCR protocol
1F[(b)]	GCCACTTACACTTCAGCTTGG	60	672	Addition of 3.5 μl 5M betaine
1R	TTCTTTTGGAGTGGGTCTG			
2F	TAGGACAAGCCCCGCACT	60	599	N/A
2R	GGTGTGTGTGGGGGTTAGTTT			
3F	GCTCCTCCAGATAGCAGTGGCAGT	56	638	Addition of 3.5 μl 5M betaine
3R	AAGTGAATGGCTGACATAAGGA			
4F	CTCCATTTCCTCTGGGATGA	60	630	N/A
4R	GAGGTGGAGGTTGCAGTGA			
5F	CCTCTTCAGGACGAATTTGC	57	494	N/A
5R	TGGAGGATAGATGTGGTCGG			

(a) – C8orf49 is one large open reading frame therefore there is only one exon present. Exon 1 is 1.97 kb in size and overlapping primer pairs were designed in order to sequence the entire exon.

(b) – Primer pair 1 covers the 5'UTR.

Table 2.10: The sequence and annealing temperature (T_A°C) of the primers used to amplify *CTSB;* and the size of the amplicons produced (Mahurkar et al., 2006)

Exon	Primer sequence 5' → 3'	T_A°C	Amplicon size (bp)	Changes made to general PCR protocol
Exon 1F*	GCGGCGGAAGGTGGCGGGAG	47	419	Addition of 5 μl 5M betaine; increase cycles to 35
Exon 1R*	CGCGGACTCCCACCCCAGCC			
Exon 2F*	GCCTTATCGTCCGCCTCTGT	66	397	N/A
Exon 2R*	GTGGGAGGAGGCAAGGGGAT			
Exon 3F	AGACGGTGCCCCTGTGTGTG	61	1027	N/A
Exon 4R	GGCCTTCACTCTCCCACTTC			
Exon 5F	TGTGCTCATTTCCTTGTTAG	54	633	Addition of 5 μl 5M betaine
Exon 5R	AAAAGAAGATTGCTAAGATTAT			
Exon 6F	TGCCTACAAGGTCTGAGGTG	59	300	N/A
Exon 6R	TCAAAGTCCCCTCCACTGAG			
Exon 7F	TGGTCTGGAGAGCTGGTGGT	64	620	N/A
Exon 8F	TGCCCCAGGTCTTCTCCGTG			
Exon 8R	GCCCTGACCTCTTCGCTGCA		292	N/A
Exon 9R	CCCTCTTCCCCAGCCCCTCA		955	N/A
Exon 10F	TGCCCCAGGTCTTCTCCGTG	61	380	N/A
Exon 10R	GTGCTCATTTCCTTGTTAG			
Exon 11F	GCCTTATCGTCCGCCTCTGT	62	640	N/A
Exon 11R	TGCCCCAGGTCTTCTCCGTG			

* Exon 1 and a part of Exon 2 make up the 5'UTR of *CTSB*

Table 2.11: The sequence and the annealing temperature (T_A°C) of the primers used to amplify *FDFT1* cDNA; and the size of the amplicons produced

Primer pair	Primer sequence 5' → 3'	T_A°C	Amplicon size (bp)	Changes made to general PCR protocol
1F 1R	CCTCGAAGCACCTACTCCAC AGCGAGTCCTGGTCCATCTT	61	253	Addition of 5 µl 5M betaine; change cycle times (30sec, 30sec, 60 sec)
2F 2R	TCTACAACCTGGTGCGCTTC AAGGGAGATCGTTGGGAAGT	61	343	N/A
3F 3R	GATGGGGAAATGCGCAAC ACCTGCTCCAAACCTCTTGA	61	568	Addition of 5 µl 5M betaine
4F 4R	CAGGAGTGGGACAAGTACTGC GGTGGAGATGATCTGCCTTG	63	640	N/A
5F 5R	TGTGCTATTCCACAGGTGATG ATGAGGCACCCTTTCCTTTC	61	580	Addition of 5 µl 5M betaine

Table 2.12: The sequence and the annealing temperature (T_A°C) of the primers used to amplify *CTSB* cDNA; and the size of the amplicons produced

Primer pair	Primer sequence 5' → 3'	T_A°C	Amplicon size (bp)	Changes made to general PCR protocol
1F 1R	ATTGGCAGGTGGATCTAGGA TAACTCTCTGGGGTGGCTTG	61	250	Addition of 3.5 µl 5M betaine; change cycle time (30sec, 60sec,60sec)
2F 2R	CAAGCCACCCCAGAGAGTTA TCCTGACTTGTAGAGCAGGAAG	62	598	N/A
3F 3R	CAGGACAAGCACTACGGATACA TGCATTTCTACCCCGATCTC	62	414	N/A

Table 2.13: The sequence and the annealing temperature (T_A°C) of the primers used in the real-time amplification of *CTSB* and *FDFT1* cDNA

Gene	Primer sequence 5' → 3'	T_A°C	Amplicon size (bp)
CTSB F *CTSB* R	GTGTGGGGACGGCTGTAAT TGTGAGCACCACGTCAACG	63	105
FDFT1 F *FDFT1 R*	AGGACTTCCCAACGATCTCC GGCAGTACTTGTCCCACTCC	61	150

2.4 MICROSATELLITE GENOTYPING AND HAPLOTYPE ANALYSIS

2.4.1 Microsatellite marker selection

In total, nine microsatellite markers located on chromosome 8p23.1-p22 between markers D8S1819 and D8S552 (Figure 1.6) were selected. Seventeen individuals were genotyped for the full set of microsatellite markers. The polymerase chain reaction was used to amplify each marker. The forward primer of each primer pair was fluorescently labeled with either a 6-FAM (blue) or HEX (green) label.

2.4.2 Setting up a microsatellite genotyping run

One microliter of the PCR product was added to 0.25μl of the ROX 500 size standard (Applied Biosystems) and 9.25μl HI-DI to give a total volume of 10.50μl. The sample was spun down and heated at 95°C for 3 minutes and then put on ice for 3 minutes. The sample plate was run in the ABI 3130xl genetic analyzer using the settings of Genemapper. Once the run was complete, the genotype for each individual was inferred from the graph produced for each marker.

2.4.3 Haplotype analysis

Haplotypes were constructed manually from the genotypes, by assuming the minimum number of recombinations and by tracing the segregation of alleles in the families. Disease associated haplotypes were identified from alleles that were transmitted from an affected

parent to an affected child in each family. Non-disease associated alleles were identified as the alleles that were not transmitted to affected children. The disease associated haplotype from each family was compared in order to identify similarities between the haplotypes of the different families since they are expected to be similar by descent, because of a known founder effect. Visual inspection allowed for easy recognition of the regions conserved between the disease haplotypes of the different families.

2.5 CANDIDATE GENE DNA AND mRNA MUTATION SCREENING

2.5.1 PCR fragment purification

Two methods were employed to obtain a pure PCR product in preparation for DNA/cDNA sequencing. The enzyme-based method was used when one distinct fragment was seen on the gel and involved Exonuclease I (New England Biolabs) and Shrimp alkaline phosphatase (New England Biolabs). These enzymes digested all the remaining genomic DNA, unused primers and dNTPs that may interfere with cycle sequencing (Enzyme based purification protocol in Appendix E.I). The second method applied was the physical excision of the relevant sized fragment from the gel. This method was used when, even after PCR optimization, non specific fragments were present on the gel. The PCR product was run on a 2% agarose gel and the fragment was carefully excised under UV light guidance and subsequently purified using a Nucleospin Extract II column (Machery-Nagel). This method was taken directly from the PCR Clean-up Gel Extraction User Manual (Appendix E.II).

2.5.2 Cycle Sequencing

In cycle sequencing, a forward and a reverse reaction was set up for each PCR fragment. In this way, the DNA sequences of the forward and reverse strands could be compared, thereby confirming the integrity of the sequences. Each reaction was prepared according to Table 2.14.

Table: 2.14: The thermal cycling conditions and the PCR components required for the cycle sequencing of each PCR fragment

PCR components	Stock concentration	Final concentration	Volume (µl)
Purified PCR product	N/A	N/A	6
Forward or reverse primer	10µM	0.5µM	1
BigDye v3.1 (Applied Biosystems)	N/A	N/A	4
5X sequencing BigDye buffer	5X	0.5X	2
ddH_20	N/A	N/A	8
TOTAL			**20**
Thermal cycling conditions			
Temperature	**Time**	**No. of cycles**	
96 °C 50 °C	30 seconds 15 seconds	30	
60 °C	4 minutes	1	
4 °C	10 minutes	1	
24 °C	Hold		

Cycle sequencing was performed in the GeneAmp® PCR system 9700 (Applied Biosystems). Once cycle sequencing was complete, the excess dye terminators (ddNTPs) were removed from the sample using a Qiagen DyeEx 2,0 Purification column (Qiagen) which uses gel-filtration chromatography to trap the unincorporated ddNTPs in the gel matrix and elute the cycle sequencing product. If the ddNTPs are not removed, they will interfere with the reading of the nucleotide sequence (Complete Qiagen DyeEx2.0 protocol in Appendix E.III). Once the cycle sequencing products had been purified, they

were vacuum dried for one hour using the Eppendorf Concentrator 5301 (Eppendorf) to concentrate the pellet. The vacuum dried samples were stored at -20°C until capillary electrophoresis could be performed.

2.5.3 Capillary electrophoresis

Capillary electrophoresis in the ABI 3130*xl* Genetic analyzer (Applied Biosystems) was used to separate and analyze the cycle sequenced products. Since the ABI 3130*xl* Genetic analyzer can be used for sequence and microsatellite analysis, a sequence analysis profile needed to be established (Appendix F.IV). Once the sequence analysis profile had been defined, the vacuum dried samples were prepared by resuspending them in 10µl of an ultra-pure form of Formamide called HiDi™ Formamide (Applied Biosystems). Each resuspended sample was added to a well of a 96 well plate, spun down in a centrifuge and denatured for two minutes at 95°C. Once the sample plate had been correctly inserted into the Genetic analyzer, the capillary electrophoresis run was started (Complete protocol in Appendix F). Once the samples had been separated by electrophoresis, the sequences were ready to be analyzed using the Sequencing Analysis v5.2 (Applied Biosystems) software.

2.5.4. Sequence analysis

The reference sequence of each candidate gene was converted into a sequence text file using the EditSeq™ module of Lasergene. The sequence text file was imported into a new

file in SeqMan™, another module of Lasergene. This reference sequence and each sample's electropherogram generated from the ABI 3130xl Genetic Analyzer were assembled and aligned in SeqMan™. Once the genetic variants had been identified, they were studied using the computational programs Ensembl(http://www.ensembl.org/index.html), SNPs3D http://www.snps3d.org/), FastSNP (http://fastsnp.ibms.sinica.edu.-tw/pages/input_CandidateGeneSearch.jsp) and PolyPhen (http://genetics.bwh.harvard.edu/pph/data/index.html).

2.5.4.1 Ensembl

The Ensembl genome annotation website allows one to link to information to determine whether a particular genetic variant was previously described and commonly found in certain population groups, and therefore classified as a SNP, or if it is a novel variant. Information such as population and individual genotypes and the allele frequencies for each SNP is available as well as whether the SNP is synonymous or non-synonymous.

*2.5.4.2 PolyPhen (**Poly**morphism **Phen**otyping)*

PolyPhen is an web based tool which predicts the possible molecular effect that non-synonymous SNPs could have on a human protein by making use of straightforward physical and comparative considerations. The prediction is based on empirical rules which are applied to the sequence, phylogenetic and structural information characterising the substitution. Several studies demonstrate that the impact of amino acid allelic variants on protein structure/function can be reliably predicted through the analysis of multiple

sequence alignments and 3D protein structures. For a given amino acid substitution in a human protein, PolyPhen performs several steps, namely the sequence-based characterisation of the substitution site, the calculation of the PSIC profile scores for the two amino acid variants and the calculation of structural parameters and contacts. PolyPhen will classify a particular variant as possibly damaging, probably damaging, benign or unknown. The major predictive tool used by Polyphen is residue conservation. Mutations of "conserved" residues are more likely to be pathogenic than mutations of non-conserved residues (Trovo et al., 2004). This is because more highly conserved residues tend to be more important for protein function and structure. It is unlikely that a amino acid substitution will occur in the active site of an enzyme because it could have a highly deleterious effect on the stability and structure of the protein. To identify whether the particular substitution may be incompatible with the spectrum of substitutions observed at this position in the family of homologous proteins, PolyPhen identifies homologues of the input sequence (target protein) via a BLAST search tool. After the BLAST search, the set of hits is filtered to retain hits that have a sequence identity to the input sequence in the range between 30 - 94% and the length of the alignment with the query sequence must not be smaller than 50 amino acids. The resulting multiple alignment is used by the PSIC software (**P**osition-**S**pecific **I**ndependent **C**ounts) to calculate the resulting profile scores which are logarithmic ratios of the likelihood of a given amino acid occurring at a particular position to the likelihood of this amino acid occurring at any position. The absolute value of the difference between profile scores of both allelic variants in the polymorphic position is calculated and are termed PSIC scores. Large PSIC

scores indicate that the particular substitution is rarely or never observed in the particular protein family while small PSIC scores indicate that the particular variant occurs frequently in the protein family (http://genetics.bwh.harvard.edu/pph/pph_help.html). PolyPhen also predicts whether the substitution occurs within the active domains of the protein, affects the protein's interaction with ligands in the crystallographic structure, leads to a hydrophobic change and/or electrostatic charge change in the protein, results in the breakage of a disulphide bond or whether the substitution results in alternate splicing.

2.5.4.3 SNPs3D

SNPs3D is a web based tool which predicts molecular functional effects of non-synonymous SNPs based on structure and sequence analysis. SNPs3D performs a multiple alignment of the input sequence with a family of homologous proteins. Using information gained from the alignment, the entropy and Position Specific Scoring Matrix (PSSM) score is calculated. Sequence entropy is a measure of the variability at a given site (Liao et al., 2005). An amino acid substitution that occurs within the active protein domain (highly conserved regions of the protein) tend to have low entropy because of the intolerance to substitutions in these highly conserved regions. Amino acid substitutions that occur in regions of the protein that are more tolerant to substitutions tend to have a high entropy. Entropy calculations take into account the nature of the residue that is substituted as well as the nature of the substitution. The PSSM score is the exchangeability score generated from the multiple alignment. A higher PSSM score indicates that the substitution is

tolerable, indicating that the substitution occurred at a site near the surface of the protein and not within an active protein domain. Substitutions that occur at the surface of the protein are unlikely to affect the stability or function of the protein and are therefore more likely to occur (Yue et al., 2006). If the protein of interest is well characterised, SNPs3D is able to predict whether the amino acid substitution will result in the introduction of backbone strain, a buried charged or polar residue or electrostatic repulsion within the protein. It will also predict whether the amino acid substitution disrupts a disulfide bond and whether there is a loss of hydrophobic effect (kcal/mol), hydrogen bonds, a polar-polar interaction (kcal/mol), a polar-charge interaction (kcal/mol) and the loss of charge-charge interaction (kcal/mol).

*2.5.4.4. FastSNP (**F**unction **A**nalysis and **S**election **T**ool for **S**ingle **N**ucleotide **P**olymorphisms)*

FastSNP predicts the possible molecular effect that a non-synonymous genetic variant may have on the protein structure and function, predicts whether the variant changes exonic splicing enhancers or exonic splicing silencing motifs and also predicts whether the variant causes a change in the active protein domain. FastSNP prioritizes SNPs according to twelve phenotypic risks and putative functional effects, such as changes to the transcriptional level, pre-mRNA splicing, protein structure, etc. A unique feature of FastSNP is that the prediction of functional effects is always based on the most up-to-date information, which FastSNP extracts from eleven external Web servers using a team of re-configurable Web wrapper agents.

2.6 REAL-TIME PCR USING THE ROCHE LIGHTCYCLER AND THE SYBR GREEN METHOD OF DETECTION

The $\Delta\Delta C_q$ method of relative quantification using real-time quantitative PCR with SYBR Green I detection was adapted and optimised to estimate the quantitative expression of *CTSB* and *FDFT1* mRNA, the two most compelling KWE positional candidates. This method assumes that the target and reference genes are amplified with efficiencies close to 100% and that the PCR efficiency of the target and reference gene systems do not differ by more than 5% of each other, in other words, the relative efficiency must be optimal (Ferreria et al., 2006). Therefore, in order for the $\Delta\Delta C_q$ calculation to be valid, these assumptions need to be validated.

2.6.1 Determination of real-time PCR efficiencies of each system

In order to determine the amplification efficiencies of each gene (reference and targets), the calibrator cDNA sample was diluted into five different concentrations (500 ng/µl, 250 ng/µl, 100 ng/µl, 50 ng/µl, 25 ng/µl) and the C_q value was measured (in triplicate) at each dilution. The average C_q value at each dilution was plotted against the log of the calibrator concentration. The amplification efficiency can be calculated from the gradient of the slope produced and Equation 1.

$$E = 10^{1/slope} - 1 \times 100 \quad \text{.....Equation 1}$$

Where E= PCR efficiency expressed in percentage

To measure the relative efficiency, the same diluted samples were amplified in triplicate using primers for both the target genes and the reference gene. The average C_q value was calculated for each system and the ΔC_q ($C_{qtarget\ gene} - C_{qreference\ gene}$) was determined. This value was plotted against the log of the calibrator concentration. A slope of <0.1 indicates optimal relative efficiency of the systems. Once the amplification and relative efficiencies are validated, the data will be used to determine the differential expression of the target genes in the samples in relation to the calibrator sample using Equations 2 and 3.

$\Delta\Delta C_q = [(C_q target - C_q reference)]$ Sample X - $[(C_q target - C_q reference)]$ calibrator…..Equation 2

$2^{-\Delta\Delta Cq}$ = N-fold change in genes expression……Equation 3

2.6.2 Setting up the run

Once the real-time reaction mixture for each gene was set up according to Table 2.6, the capillaries were placed into a capillary holder and 18µl of the reaction mix was added to each capillary followed by the cDNA. cDNA was not added to the negative control. Each sample, both patient and control, as well as the calibrator was run in triplicate for both the reference gene (*GAPDH*) and the target gene (*FDFT1* and *CTSB*) assays.

2.6.3 Statistical Analysis

To determine whether any change in the relative expression of the target genes observed in the patients are significant ($p<0.05$), the data will be assessed using two different statistical tests, namely the Wilcoxon two-sample test (a non parametric one way ANOVA test) as well as the non-parametric Mann-Whitney U test.

CHAPTER 3

RESULTS

3.1 THE CONSTRUCTION OF HAPLOTYPES

The position of the KWE critical region on the short arm of chromosome 8 has been redefined over the years, but due to instability in the genome remains to be confirmed (Starfield et al., 1997; Starfield et al., 2000; Appel et al., 2001). For this reason, the confirmation of the KWE critical region became the first priority of this study. Extensive KWE associated haplotypes were constructed for the South African KWE families 23, 26, 36 and 46 (Figures 3.1–3.4). A total of nine markers, spanning a region of roughly 1.9Mb, was genotyped. The allele sizes are used for the notation of the alleles in the construction of the haplotypes which are documented in Table 2.1. The haplotypes constructed in this study are based on the evidence that the KWE critical region is inverted due to the polymorphic 8p23 inversion. The order of the markers is D8S1819 – inversion breakpoint - D8S1759 – D8S1695 – D8S265 – D8Stet – D8Sdi – D8S1755 – inversion breakpoint - D8S2657 – D8S552 assuming a non-inverted chromosome (Figure 1.6a). The haplotypes were constructed in order to visually determine the most likely location of the KWE locus with respect to the markers analyzed. The founder KWE haplotype 214-258-118-221-231-182-243-226-173 is boxed and its segregation with affection status in family members is indicated. In the case of recombination in ancestors, the section differing from the founder haplotype is excluded from the box.

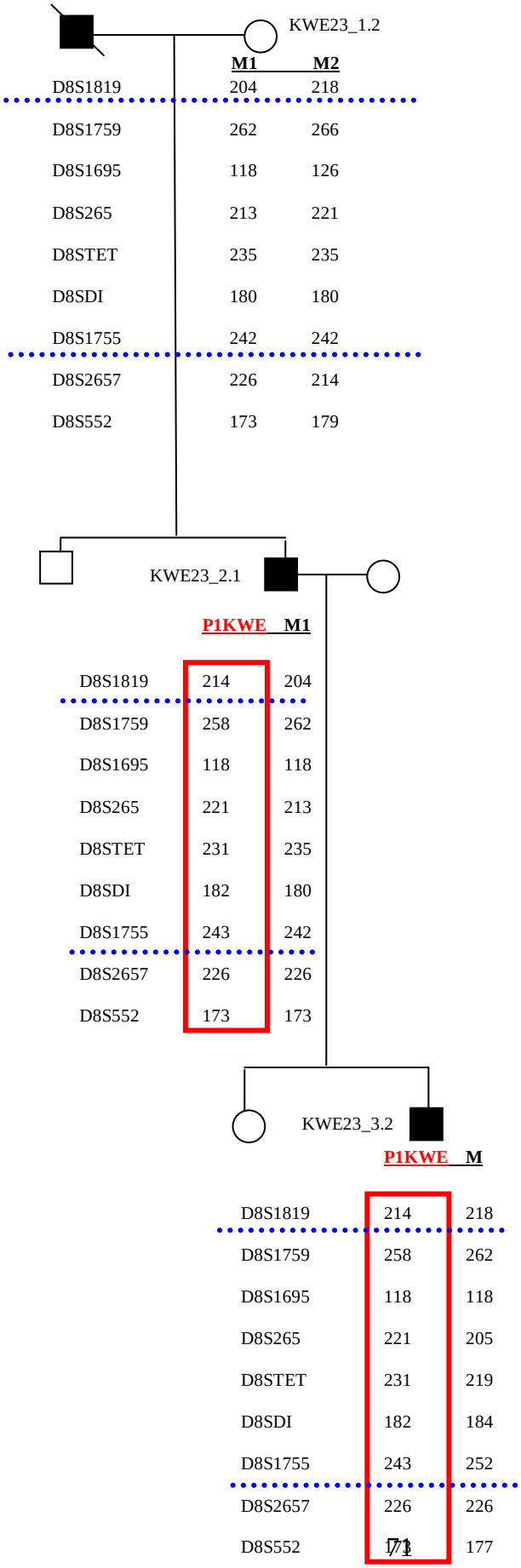


Figure 3.1: Haplotypes constructed for KWE family 23. The blue dotted lines indicate where the distal and proximal breakpoints of the polymorphic 8p23.1 inversion occur.

Although no information was available for many members of this family, the majority of the markers were informative and facilitated the construction of the haplotypes. The entire proposed KWE founder haplotype was observed in both the KWE affected individuals 2.1 and 3.2 and was not observed in the unaffected individual 1.2. No informative recombination events were detected in this family.

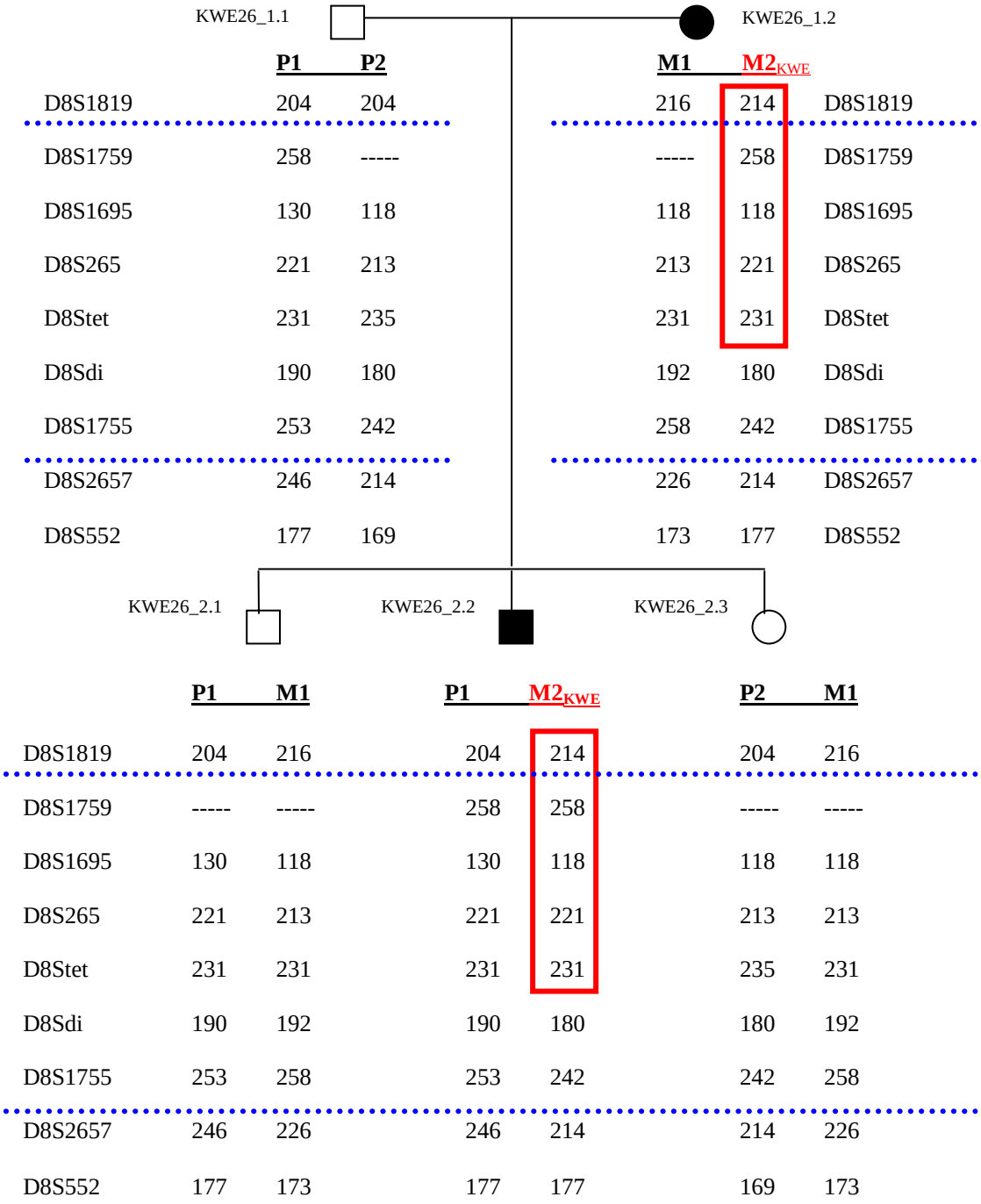


Figure 3.2: Haplotypes constructed for KWE family 26. Only part of the founder KWE haplotype is present in the affected members of this family (-----no results obtained for this marker)

The affected child, individual 2.2, inherited the same alleles from his affected mother and the unaffected children inherited the alleles from the other chromosome 8 from their mother. This confirms that the M2 haplotype segregates with KWE as both the affected mother and affected son carry the M2 haplotype. There were no informative recombination events observed in this family. It is clear that only part of the founder KWE haplotype (D8S1819-D8S1759-D8S1695-D8S265-D8Stet) is conserved in both the affected mother and the affected child. There is part of the KWE founder haplotype that is not conserved in the two affected individuals (region including marker D8Sdi up until and including marker D8S552) and is proposed to be the result of an ancestral cross-over event, and is thus considered as evidence to exclude it from the KWE critical region, placing the KWE gene between markers D8S1819 and D8Sdi.

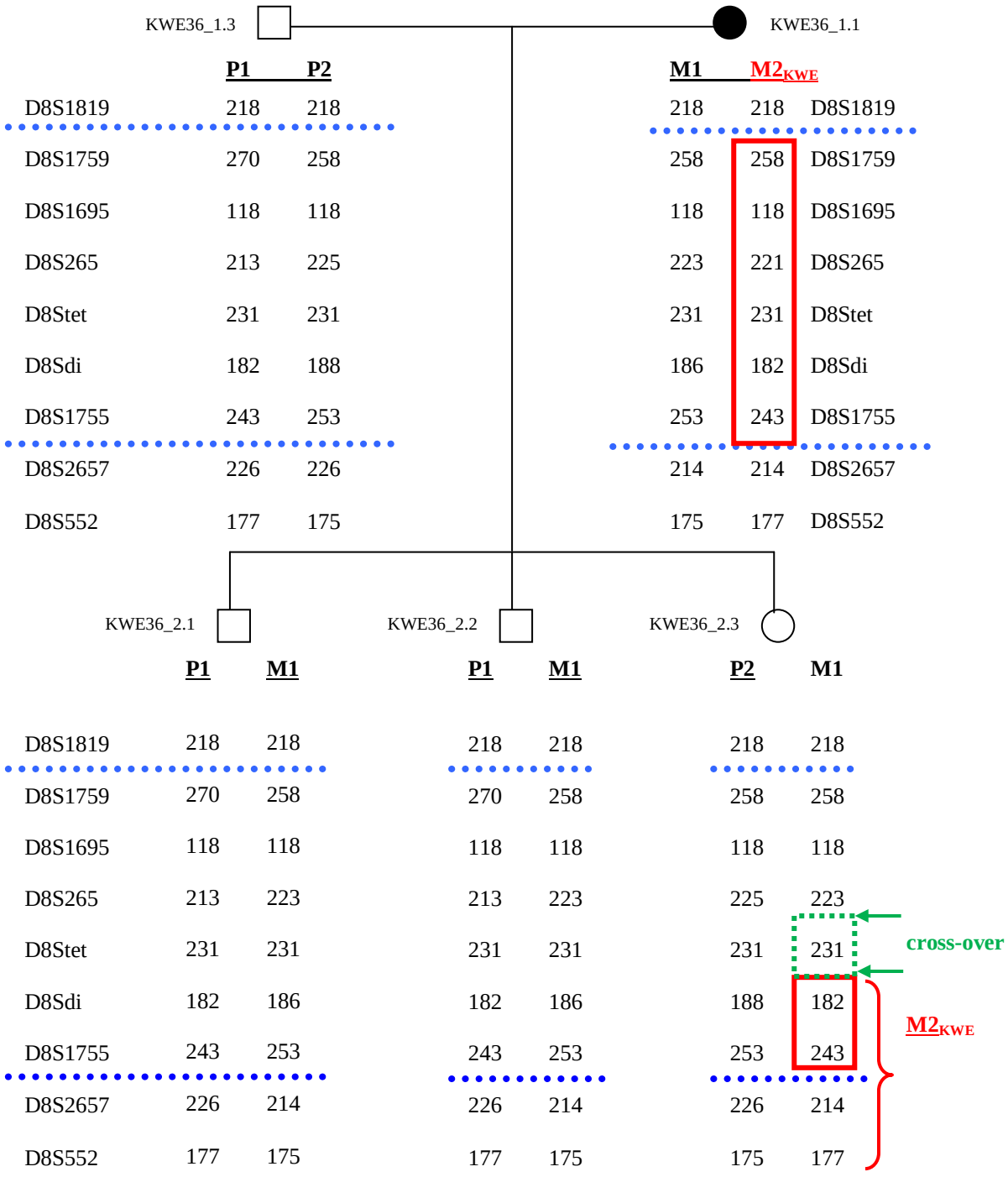


Figure 3.3: Haplotypes constructed for KWE family 36.

Only a section of the founder haplotype is present in the affected individual in this family, individual 1.1. It is clear that the alleles of the markers positioned on the outer side of the repeat boundaries are not part of the founder KWE haplotype. The founder KWE haplotype appears to be conserved in the inverted region in the affected family members. The unaffected individual 2.3 has inherited part of the KWE founder haplotype from her affected mother. Unfortunately, the affected mother and unaffected daughter have a number of uninformative marker genotypes (D8S1819, D8S1759, D8S1695 and D8Stet) which make it difficult to identify the recombination breakpoint. By looking at the two maternal alleles and the informative markers (D8Sdi, D8S265 and D8S2657) it becomes clear that the region between markers D8Sdi and D8S2657, which individual 2.3 inherited from her affected mother and which would be expected to segregate with the KWE allele, are not segregating with the KWE phenotype. However, the uninformativeness of the D8Stet marker in each member of this family makes it unclear as to whether the 231 allele present in the unaffected child (bracketed by ⬚) was inherited from the M1 or $M2_{KWE}$ haplotype. This prevents one from being able to exclude or include the region around this marker from the KWE critical region. The exact position of the cross-over remains unclear; it either took place between markers D8S265 and D8Stet or between markers D8Stet and D8Sdi. Based on the knowledge of the founder haplotype, the recombination that is observed in individual 2.3 places the KWE locus above D8Sdi (Figure 3.3) since this individual is unaffected.

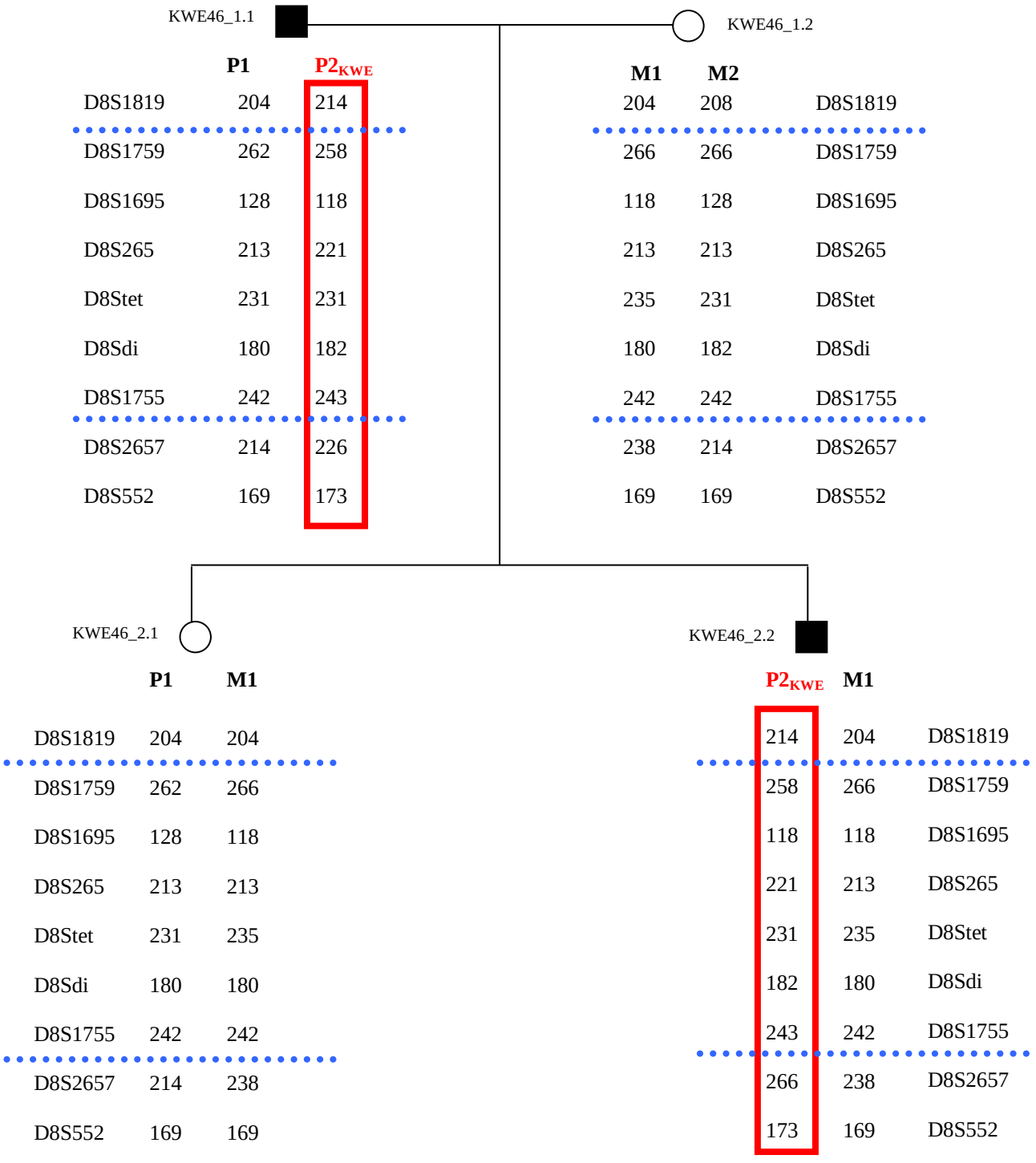


Figure 3.4: Haplotypes constructed for KWE family 46.

The majority of the markers for family 46 were informative. The affected father in this pedigree has the complete founder KWE haplotype, as seen in family 23, which was inherited by his affected son. No recombination events were observed in this family.

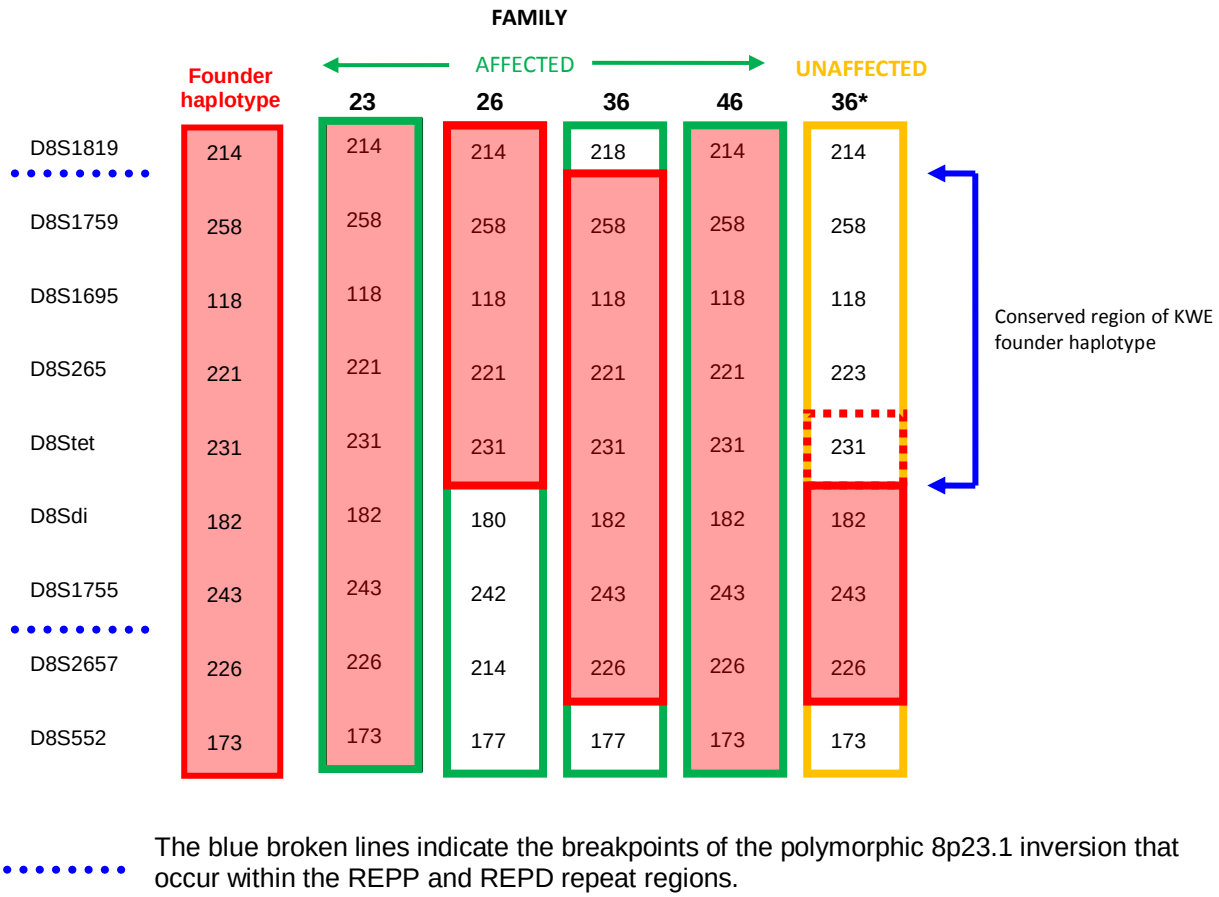


Figure 3.5: An illustration of the founder KWE haplotype and the region of this haplotype that is conserved among KWE affected individuals. An informative recombination event is observed in the unaffected individual denoted 36*, who has inherited part of the KWE founder haplotype from her affected mother

Figure 3.6 highlights the critical KWE region according to the results obtained from the haplotype analysis. The KWE critical region, a region roughly 0.1kb, is the region between markers D8S1819 and D8Sdi and contains eight known genes, one pseudogene, four open reading frames, a gene for a zinc finger protein and four beta defensin precursors.

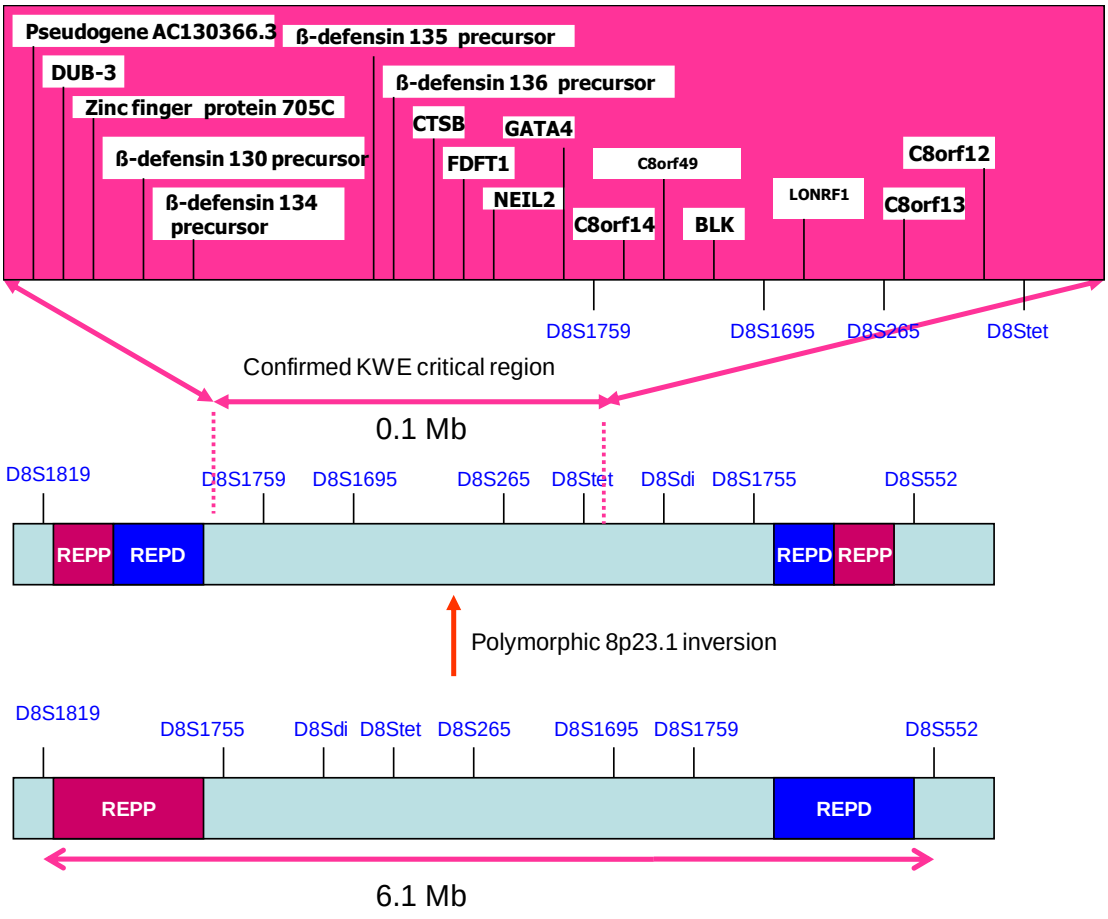


Figure 3.6: The confirmed KWE critical region

3.2. CANDIDATE GENE ANALYSIS

The validated KWE region contains many genes that are positional KWE candidates due to their chromosomal location. The genes chosen as the most likely candidates from this region were *CTSB*, *FDFT1*, *DUB-3* and *C8orf49*. The intronic genetic variants identified within each candidate gene (Tables 3.1, 3.9, 3.20 and 3.22) were observed in both the patient and control groups and were therefore excluded from any further analysis. The genetic variants within the candidate genes were identified through sequencing. The non-synonymous variants identified within *CTSB* were analysed using the bioinformatic tools SNPs3D, FastSNP and Polyphen. All the coding variants identified through DNA sequencing within *FDFT1* were synonymous therefore limiting variant analysis to FastSNP. The genetic variants identified within *DUB-3* and *C8orf49* were analysed using only FastSNP as SNPs3D and PolyPhen do not provide predictive data for these genes.

3.2.1 The Cathepsin B gene (*CTSB*)

3.2.1.1 Genomic DNA analysis

Fifteen genetic variants were identified within this gene (Table 3.1), three of which occurred in the coding region. One variant, which has been previously described as a SNP, was observed within the 5'UTR of *CTSB*. This variant was only observed in the control group and does not appear to create or abolish an upstream start codon (uAUG) or upstream open reading frame (uORF).

***L26V* (rs12338, G>C)**

The exchange of lysine with valine at amino acid position 26 was predicted to cause a possibly damaging structural effect on the *CTSB* protein (Table 3.2). This variant does not occur in the active/binding domain of the protein and is not conserved across orthologues but is predicted to occur in a site that is intolerant to variation. The normal splicing of the gene is not affected by this transversion and the variant C allele (Figure 3.7a) is found in both the patient and control groups. This variant has a PSIC score of 0.114 and the entropy and PSSM scores are 2.27 and 2 respectively (Table 3.2).

***S53G* (rs175855418, A>G)**

This variant was predicted to be a low risk non-synonymous polymorphism that leads to a conservative change in the protein. The regular splicing of the gene is not affected by this substitution (Table 3.2). The amino acid exchange involves serine and glycine at amino acid position 53. This transversion has a PSIC score of 0.494 and an entropy and PSSM score of 3.5 and -1 respectively. The variant G allele (Figure 3.7b) is present in both the patient and control group.

***T140T* (rs13332, A>C)**

This synonymous variant has been previously identified in many population groups and is predicted to be a low risk variant that leads to a conservative change in the protein exerting no damaging effect on the structure or function of the protein (Table 3.2). The variant C allele (Figure 3.7c) is present in both the patient and control group.

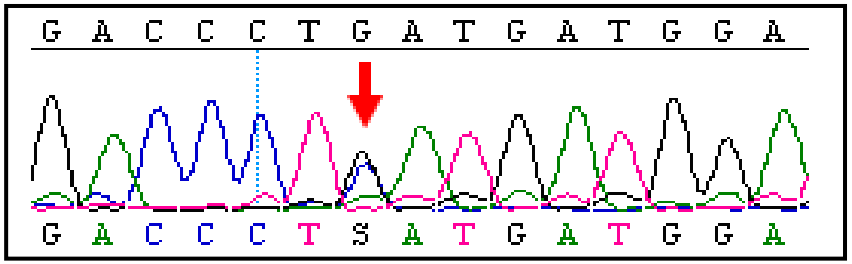
The intronic 19bp duplication (novel)

Although this duplication occurs in the intronic region of the gene, it is of interest because it may affect the splicing of the gene that can only be observed through the analysis of the gene's mRNA. Both KWE affected and unaffected individuals carry this duplication.

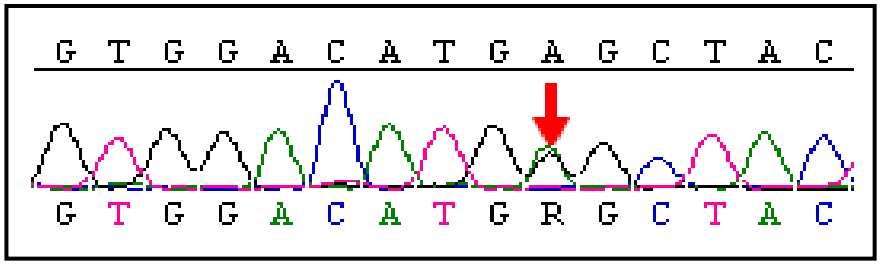
Table 3.1: The genetic variants identified in *CTSB* by sequencing.

Genetic Variant	SNP C>T rs17154017	cSNP G>C rs12338	SNP c>t	SNP t>c	cSNP A>G rs17855418	SNP c>t	cSNP A>C rs13332	SNP c>t
Position	**5'UTR**	**Exon 3 L26V**	**IVS3-196**	**IVS 4 -96**	**Exon 4 S53G**	**IVS5+20**	**Exon 6 T140T**	**IVS7+128**
Patients								
KWE46_1.1	C/C	G/C	c/c	t/c	A/G	c/t	A/C	c/c
KWE23_2.1	C/C	G/C	c/c	t/c	A/G	c/c	A/C	c/c
KWE36_1.1	C/C	G/C	c/t	c/c	A/G	c/t	C/C	c/t
KWE50_1.1	C/C	G/C	c/c	t/c	A/G	c/t	C/C	c/c
KWE50_2.1	C/C	G/C	c/c	t/c	A/G	c/c	C/C	c/c
Controls								
KWE46_2.1	C/T	G/C	c/c	t/c	A/G	c/t	A/C	c/c
KWE23_1.2	C/C	G/C	c/c	t/t	A/G	c/c	A/A	c/c
KWE36_2.1	C/T	G/C	c/t	t/c	A/G	c/c	A/C	c/t
Control#1	C/T	G/C	c/c	t/c	A/G	c/t	A/C	c/c
Control#2	C/C	G/C	c/t	t/t	A/G	c/c	A/C	c/c
Genetic Variant	**SNP c>a**	**19bp Duplication**	**SNP C>G rs709821**	**SNP A>G rs8898**	**SNP C>G rs709822**	**SNP C>T rs1803352**	**SNP C>G rs4987129**	
Position	**IVS7+134**	**IVS7 21 267 – 21 280**	**3'UTR**	**3'UTR**	**3'UTR**	**3'UTR**	**3'UTR**	
Patients								
KWE46_1.1	c/a	WT/dup	C/C	A/G	C/C	C/T	C/C	
KWE23_2.1	c/a	WT/dup	C/C	A/G	C/G	C/C	C/G	
KWE36_1.1	c/a	dup/dup	C/G	A/A	C/G	C/T	C/G	
KWE50_1.1	c/a	WT/dup	C/C	A/G	C/G	C/C	C/G	
KWE50_2.1	c/a	WT/WT	C/C	A/G	C/G	C/C	C/G	
Controls								
KWE46_2.1	c/a	WT/dup	C/C	A/G	C/C	C/T	C/C	
KWE23_1.2	c/c	WT/WT	C/C	G/G	C/G	C/T	C/G	
KWE36_2.1	c/a	WT/dup	C/G	A/G	C/G	C/C	C/C	
Control#1	c/c	WT/WT	C/C	G/G	C/G	C/T	C/G	
Control#2	c/c	WT/WT	C/C	A/G	C/G	C/C	C/C	

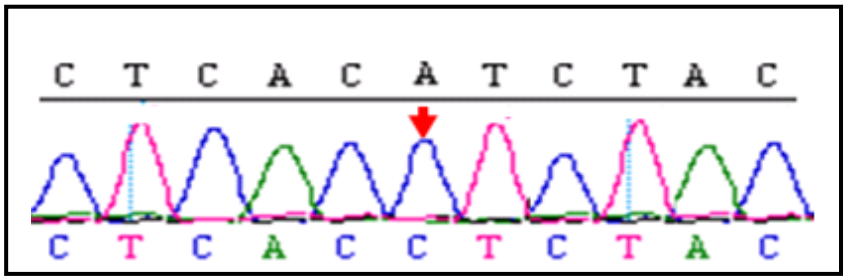
a) rs12338 G>C exchange



b) rs17855418 A>G exchange



c) rs13332 A>C exchange



d) Intronic 19bp duplication

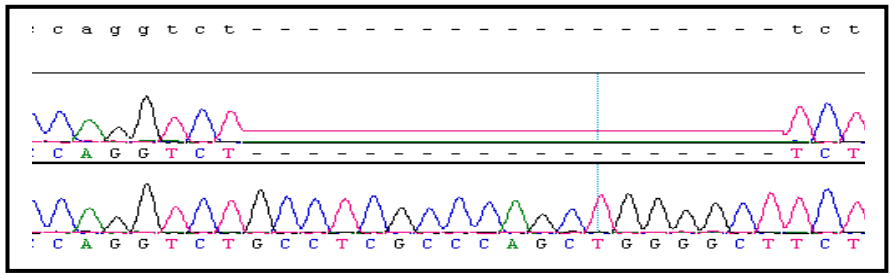


Figure 3.7: The genetic variants identified within *CTSB.*

Table 3.2: A bioinformatic prediction of the potential functional impact of the genetic variants identified within *CTSB.*

			Variants	
Bioinformatic tool		***L26V (G>C)*** ***rs12338***	***S53G (A>G)*** ***rs17855418***	***T140T ()A>C)*** ***rs13332***
Polyphen				Polyphen does not predict data for synonymous variants
	Is the amino acid conserved across species?	No	No	
	Does the substitution occur in an active protein domain?	No	No	
	PSIC score	0.114	0.494	
FastSNP	Risk	Medium	Low	low
	Possible functional effect	missense (conservative)	missense (conservative)	missense (conservative)
	Does the substitution occur in an active protein domain?	No	No	No
	Does the substitution affect regular splicing of the gene?	No	No	No
	Does the substitution change a transcription binding site?	No	No	No
	Does the substitution affect an exon enhancer site?	No	No	No
				SNPs3D does not predict data for synonymous variants
SNPs3D	Does the substitution introduce backbone strain?	No	No	
	Does the substitution cause a change in the hydrophobicity of the protein?	No	No	
	Does the substitution result in a change in the electrostatic charge?	No	No	
	Does the substitution result in the breakage of a disulphide bond?	No	No	
	Is there a loss of hydrogen bonds due to the substitution?	No	No	
	Entropy score	2.27	3.5	
	PSSM score	2	-1	
	OVERALL prediction	Benign	Benign	Benign

3.2.1.2 CTSB cDNA analysis

In order to identify genetic variants that are often not observed at a DNA level, such as deep intronic variants that may cause aberrant splicing of a gene and large insertions or deletions, the cDNA of *CTSB* was screened. Total RNA was extracted from two peripheral blood samples and two saliva samples obtained from KWE affected individuals as well as unaffected controls. The total RNA was reverse transcribed into cDNA which was used in the downstream applications of RT-PCR and sequencing.

There are five transcript variants encoding the same *CTSB* protein (NM_001908.3 (1), NM_147783.2 (2), NM_147782.2 (3), NM_147781.2 (4) and NM_147780.2 (5)). Transcript 5 is the transcript most commonly found in humans and is generally referred to as the *CTSB* reference sequence. Transcripts 1-4 each has one or two additional exons which are spliced out. After splicing, the five transcripts differ only in the sequence of their 5' untranslated regions. This difference in the cDNA's 5'UTR sequence made it possible to distinguish which of the five transcripts each of the patients and controls had. Although many transcripts encode *CTSB*, transcript 5 was the only transcript observed in the patient and control groups. Table 3.3 lists the genetic variants that were observed in the cDNA. These variants occur in the coding region of the gene and were previously identified through the mutational screening of *CTSB* genomic DNA. Due to the location of these variants in the coding region of the gene, one would expect to see them in the cDNA. No additional variants were identified at the cDNA level.

Table 3.3: Mutation analysis of *CTSB* cDNA

Genetic variant	G>C GTG>CTG	A>G AGC>GGC	A>C ACA>ACC
Position	L26V	S53G	T140T
Patients			
KWE50_1.1(blood)	G/C	A/G	C/C
KWE50_2.1(saliva)	G/C	A/G	C/C
Controls			
Control#1 (saliva)	G/C	A/G	A/C
Control#2 (blood)	G/C	A/G	A/C

3.2.1.3 The relative quantification of CTSB expression in human epidermal palmar skin

During amplification, an amplification profile for each sample is created by the real-time thermal cycler. This amplification profile indicates which samples were successfully amplified (peak formed) and which samples failed to amplify. The amplification profile obtained for the patient and control samples for both the target gene system (*CTSB*) and reference gene system (*GAPDH*) are shown in Appendix G. The C_q values are automatically calculated by the relative quantification software and are used to determine the N-fold differential expression of the target gene in the samples. This differential expression is relative to the calibrator sample and normalized by the reference gene. Details of the $2^{-\Delta\Delta Cq}$ method have been previously described (Section 2.8). In this study, the data are presented as the quantity of *CTSB* mRNA (relative quantity) in KWE affected and unaffected individuals normalized to the internal control gene (*GAPDH*) and relative to the calibrator sample. The real-time PCR data is represented as C_q values which are determined by the RQ software as the cycle

at which the first significant increase in PCR product (fluorescence) is detected. The amplification efficiencies of each target and reference gene system as well as the relative efficiency was determined before the quantity of *CTSB* mRNA in the target sample could be determined using the $2^{-\Delta\Delta Cq}$ method.

To determine the efficiency of the *CTSB* and *GAPDH* PCR systems, the five calibrator dilutions (500ng/µl, 250ng/µl, 100ng/µl, 50ng/µl and 25ng/µl) were run in triplicate in each assay. The resulting C_q values at each dilution were averaged and this value was plotted against the log of the calibrator dilution. Table 3.4 lists the data used to construct the standard curve.

Table 3.4: The data used to determine the amplification efficiency of the *CTSB* and *GAPDH* systems

Calibrator dilutions (ng/µl)	Log calibrator concentration		$C_q$values		Average C_q value
***CTSB* assay**		**1**	**2**	**3**	
500	2.7	27.74	27.64	27.77	27.72
250	2.4	28.68	28.79	28.74	28.73
100	2	30.09	29.97	30.16	30.07
50	1.7	31.12	31.26	21.19	31.19
25	1.4	32.10	32.13	32.16	32.13
***GAPDH* assay**		**1**	**2**	**3**	
500	2.7	22.05	22.98	22.02	22.35
250	2.4	23.92	23.03	23.98	23.64
100	2	24.64	24.57	24.55	24.58
50	1.7	25.66	25.63	25.77	25.69
25	1.4	26.92	27.12	27.02	27.02

The data generated for both the *CTSB* and *GAPDH* amplification systems produced straight lines with slopes of -3.48 and -3.43 respectively. The standard curves are depicted in Figure 3.8.

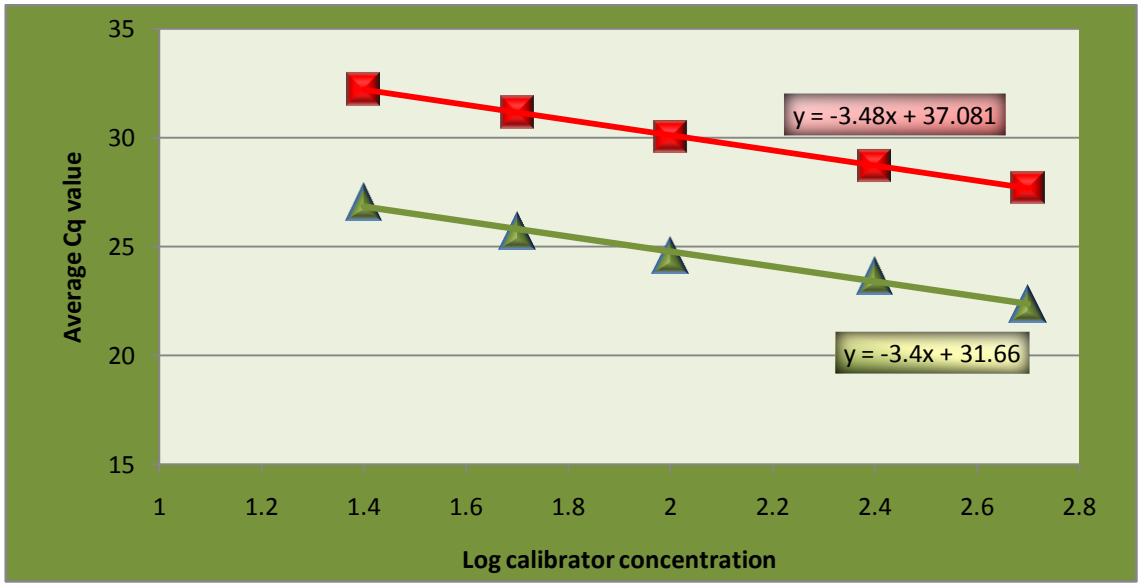


Figure 3.8: The standard curve produced to determine the efficiency of the *CTSB* (■) and *GAPDH* (▲) real-time PCR assays.

The amplification efficiencies (E) of the target and reference gene systems were calculated using the slope value (Figure 3.8) and Equation 1.

***CTSB* amplification efficiency:** $= 10^{1/slope} - 1 \text{ X } 100\%$

$= 10^{1/-3.48} - 1$

$= 10^{0.287} - 1$

$= 1.936 - 1 \text{ X } 100\%$

= 93.6 %

***GAPDH* amplification efficiency**: $= 10^{1/slope} - 1 \text{ X } 100\%$

$= 10^{1/-3.43} - 1$

$= 10^{0.294} - 1$

$= 1.968 - 1 \text{ X } 100\%$

= 96.7 %

96.7% - 93.6% = 3.1% Therefore the amplification efficiencies of these systems are relatively close to 100% and are less than 5% apart.

To calculate the efficiency of the target gene system relative to the reference gene system, the data tabulated in Table 3.4 were used. The average C_q value at 500 ng/µl in the *GAPDH* assay was subtracted from the average C_q value at 500 ng/µl in the *CTSB* assay. The resulting value was plotted against the log of the calibrator concentration (log of 500). This was done for each dilution. The gradient of the graph produced should be less than 0.1 in order for the data to be regarded as efficient and accurate (Ferreria et al., 2006). The graph (Figure 3.9) and gradient were determined using Microsoft Excel 2007.

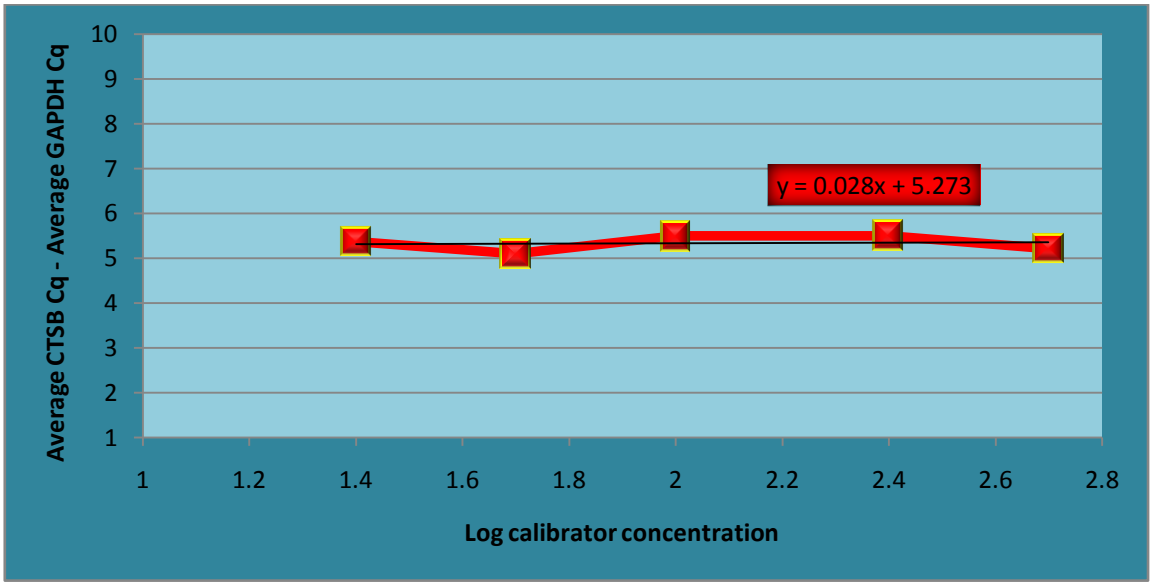


Figure 3.9: Calculation of the relative efficiency of the target gene *CTSB* system

The slope value is **0.029** (<0.1), which reflects optimal relative efficiency.

Once it was determined that the *CTSB* and *GAPDH* amplification systems had optimal and similar efficiencies, the quantity of *CTSB* mRNA in the target samples, relative to the calibrator sample and normalized by the reference system, was calculated using the $\Delta\Delta C_q$ calculation (Equation 1).

Table 3.5: The C_q values obtained from the *CTSB* and *GAPDH* patient run. The standard deviation (SD) and coefficient of variance (CV) are shown

Patients - *CTSB*		C_q values		Average C_q value	SD	CV (%)
Experimental run (A)	**1**	**2**	**3**			
KWE50_1.2	30.94	32.23	30.82	31.4	0.31	1.19
KWE50_2.2	31.4	30.89	31.68	31.32	0.4	1.27
KWE50_3.2	31.69	31.08	31.53	31.43	0.32	1
KWE50_3.3	25.1	25.11	25.15	25.12	0.03	0.13
Calibrator	30.62	30.63	30.61	30.64	0.01	0.03
Experimental run (B)	**1**	**2**	**3**			
KWE20_3.11	21.76	21.89	21.85	21.8	0.07	0.32
KWE20_2.5	19.59	19.31	19.52	19.5	0.15	0.77
KWE23_2.1	22.21	22.42	22.22	22.3	0.12	0.54
KWE45_3.6	23.14	23.01	23.19	23.1	0.34	1.5
Calibrator	27.72	27.72	27.67	27.7	0.03	0.11
Patient - *GAPDH*		**C_q values**		**Average C_q value**	**SD**	**CV (%)**
Experimental run (A)	**1**	**2**	**3**			
KWE50_1.2	26.26	26.45	26.29	26.33	0.11	0.39
KWE50_2.2	26.24	26.29	26.48	26.34	0.17	0.81
KWE50_3.2	26.61	26.53	26.70	26.61	0.13	0.48
KWE50_3.3	21.48	21.34	21.14	21.32	0.085	0.32
Calibrator	25.65	25.73	25.68	25.69	0.04	0.16
Experimental run (B)	**1**	**2**	**3**			
KWE20_3.11	16.86	16.94	16.91	16.9	0.04	0.26
KWE20_2.5	15.19	15.36	15.24	15.3	0.09	0.60
KWE23_2.1	17.68	17.72	17.56	17.6	0.08	0.45
KWE45_3.6	18.79	18.86	18.87	18.84	0.04	0.21
Calibrator	22.69	22.76	22.59	22.68	0.09	0.40

Table 3.6: The C_q values obtained from the *CTSB* and *GAPDH* control sample run. The standard deviation (SD) and coefficient of variance (CV) is shown

Control - *CTSB*		C_q values		Average C_q value	SD	CV (%)
Experimental run (A)	**1**	**2**	**3**			
Control #1	26.71	26.75	26.69	26.72	0.03	0.11
Control #2	29.01	28.98	28.95	28.89	0.03	0.10
Control #3	27.13	27.33	27.09	27.18	0.13	0.47
Control #4	30.88	30.02	30.92	30.6	0.31	1.00
Calibrator	30.32	30.39	30.34	30.36	0.036	0.12
Experimental run (B)	**1**	**2**	**3**			
Control #5	23.69	24.58	24.42	24.2	0.47	1.94
Control #6	25.02	25.09	24.81	25.0	0.15	0.60
Control #7	25.70	25.76	26.10	25.9	0.22	0.85
Control #8	24.60	23.83	23.78	24.07	0.46	1.91
Calibrator	27.86	27.72	28.59	28.10	0.47	1.67
Controls - *GAPDH*		**C_q values**		**Average C_q value**	**SD**	**CV (%)**
Experimental run (A)	**1**	**2**	**3**			
Control #1	22.25	22.32	22.50	22.45	0.11	0.5
Control #2	24.60	24.52	24.58	24.56	0.42	0.17
Control #3	23.03	23.04	23.04	23.03	0.005	0.02
Control #4	25.83	25.92	25.83	25.86	0.05	0.21
Calibrator	25.18	25.23	25.20	25.21	0.025	0.01
Experimental run –(B)	**1**	**2**	**3**			
Control #5	20.12	20.20	20.30	20.2	0.09	0.45
Control #6	20.79	20.82	20.84	20.8	0.03	0.14
Control #7	21.02	20.99	20.90	20.97	0.06	0.29
Control #8	19.73	19.73	19.80	19.75	0.04	0.20
Calibrator	23.03	22.95	23.06	23.01	0.06	0.26

The data tabulated in the Tables 3.5 and 3.6 were used to calculate the $\Delta\Delta C_q$ value using equation 2. The N-fold differential expression in the *CTSB* gene of the affected and unaffected samples are expressed as $2^{-\Delta\Delta Cq}$. From the $\Delta\Delta C_q$ value, the $2^{-\Delta\Delta Cq}$ value is determined. The $\Delta\Delta C_q$ and corresponding $2^{-\Delta\Delta Cq}$ value for each patient and control is listed in Table 3.7. In a study completed by Ferreira et al., in 2006, increased mRNA expression was defined as N-fold ≥2.0, "normal" expression was an N-fold ranging from 0.5001 to 1.9999, and decreased mRNA expression was N-fold ≤0.5.

Table 3.7: The $\Delta\Delta C_q$ value and corresponding $2^{-\Delta\Delta Cq}$ of each patient and control sample

Sample	$\Delta\Delta C_q$ value	$2^{-\Delta\Delta Cq}$
KWE50_1.2	0.11	**1.89**
KWE50_2.2	0.02	**1.98**
KWE50_3.2	-0.14	**1.13**
KWE50_3.3	-0.76	**1.69**
KWE20_3.11	-0.12	**1.08**
KWE20_2.5	-0.82	**1.76**
KWE23_2.1	-0.32	**1.25**
KWE45_3.6	-0.76	**1.69**
Control #1	-0.88	**1.84**
Control #2	-0.82	**1.77**
Control #3	-1	**2**
Control #4	-0.41	**1.33**
Control #5	-0.96	**1.95**
Control #6	-0.82	**1.77**
Control #7	-1	**2**
Control #8	-0.41	**1.33**

To visually assess the differential expression of *CTSB* in each patient and control sample, a simple bar graph was constructed using the data in Table 3.7.

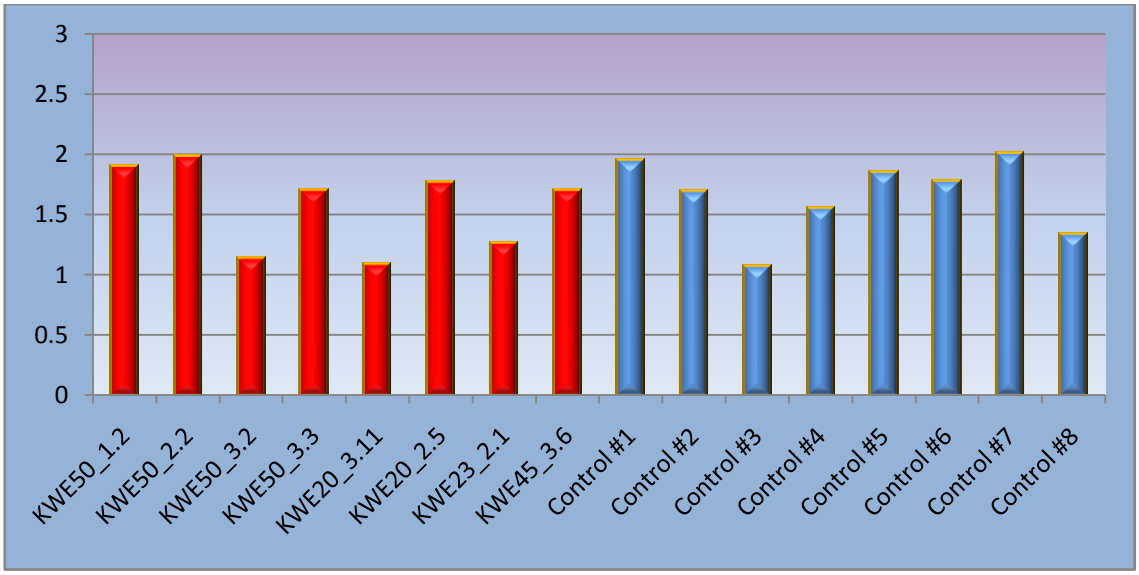


Figure 3.10: Normalized *CTSB* expression in plantar skin from KWE affected and control individuals.

The Wilcoxon two-sample test (a non-parametric one way ANOVA test) and a non-parametric Mann-Whitney U test showed that there is no significant difference in the expression of *CTSB* in patient samples when compared to the expression observed in the control samples.

Table 3.8: Statistical analysis results

.

Statistical test	p-value
Two-tailed test	
Wilcoxon	0.680
Mann-Whitney U	0.707

3.2.2 *Farnesyl-diphosphate-farnesyl (FDFT1)*

3.2.2.1 Genomic DNA analysis

Ten genetic variants were identified within this gene (Table 3.9), three of which occurred within the coding region. Two genetic variants were identified in the 5'UTR of this gene. The one 5'UTR variant (rs11549153) occurs in both the patient and control groups and the second variant (Figure 3.11a), found only in the patient group, has not been previously described. This novel variant creates an uAUG in the gene's 5' untranslated region. All the coding genetic variants observed in *FDFT1* have been previously described as synonymous polymorphisms. The one variant, L7L, was found to segregate with the KWE affected individuals, however, when analysed in a larger control group, the L7L variant was observed. The novel 5'UTR variant was also analysed in a larger group of unaffected individuals. Two genetic variants were observed within the promoter/upstream region of the *FDFT1* gene, one of which is

novel. These variants are not located within a putative transcriptional element (SP-1, SBE-1 and NF1) and were identified in the control group. Table 3.10 lists data collected from sequencing the *FDFT1* promoter in KWE families.

***L7L* (rs1047643, T>C)**

The variant C allele (Figure 3.11b) was observed only in the patient group. However, as this variant results in a silent transversion, it is unlikely to be relevant in the manifestation of KWE. It has been predicted that this low to medium risk SNP may result in a change in an exon splicing enhancer (ESE) motif, however it does not affect the binding of transcription factors or splicing and is predicted to have no effect on the stability or structure of the protein (Table 3.11). This variant is also found in unaffected individuals (Section 3.2.2.2).

***L211L* (rs904011, T>C)**

This low risk synonymous genetic variant has been previously described as a SNP and does not occur in the active site of the protein. The predictions show that a change from a tyrosine to a cytosine nucleotide at this position is unlikely to affect the binding of a transcription factor or affect the splicing of the gene (Table 3.11). The variant C allele is present in the patient and control groups (Figure 3.11c).

***L324L* (rs9205, G>C)**

This low risk synonymous genetic variant has been previously described as a SNP and The predictions show that the transition from a guanine to a cytosine nucleotide at this position is unlikely to affect the binding of a transcription factor or affect the splicing of the gene. For this reason it is predicted that this variant does not affect the structure or function of the protein (Table 3.11). The variant C allele is present in both the patient and control groups (Figure 3.11d).

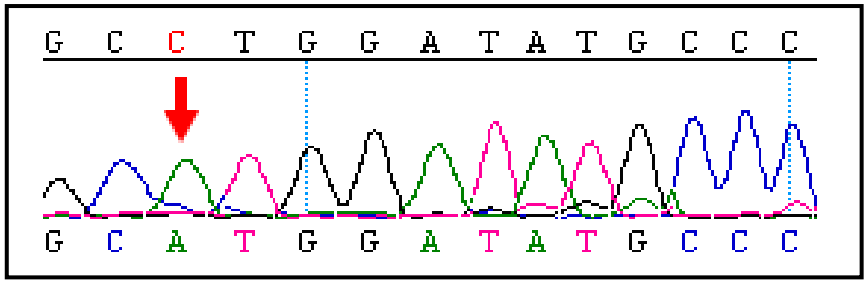
Table 3.9: The genetic variants tested in F*DFT1* by DNA sequencing.

Genetic Variant	Promoter A>G	Promoter T>C	5'UTR G>C rs11549153	5'UTR C>A novel	cSNP T>C rs1047643	SNP c>t	cSNP T>C rs904011	SNP g>c	SNP c>t	cSNP G>C rs9205	SNP t>g	3'UTR T>C rs3258	3'UTR A>C rs3188986
Position					Exon 1 L7L	IVS2+50	Exon 5 L211L	IVS6+9	IVS6+30	Exon 7 L324L	IVS7+49		
Patients													
KWE46_1.1	G/G	C/C	G/C	C/A	T/C	c/t	C/C	c/c	c/c	G/G	t/g	C/C	A/A
KWE23_2.1	G/G	C/C	G/C	C/A	T/C	c/t	C/C	c/c	c/c	G/G	t/g	T/C	A/C
KWE36_1.1	A/G	C/C	G/C	C/C	T/C	c/c	C/C	c/c	c/c	G/G	t/t	T/C	A/A
KWE50_1.2	G/G	C/C	G/C	C/C	T/C	c/t	C/C	c/c	c/c	G/G	t/g	T/C	A/C
KWE50_2.1	G/G	C/C	C/C	C/A	T/C	c/c	C/C	c/c	c/t	G/C	t/t	C/C	A/C
Controls													
KWE46_2.1	G/G	C/C	G/G	C/C	T/T	c/c	C/C	c/c	c/t	G/C	g/g	T/C	C/C
KWE23_1.2	G/G	C/C	G/G	C/C	T/T	c/t	C/C	c/c	c/c	G/G	g/g	T/C	C/C
KWE36_3.1	G/G	C/C	G/C	C/C	T/T	c/t	C/C	c/c	c/c	G/G	t/g	T/T	A/A
Control#1	G/G	C/C	C/C	C/C	T/T	c/t	C/C	c/c	c/c	G/G	g/g	T/C	C/C
Control#2	G/G	C/C	G/G	C/C	T/T	c/t	C/C	c/c	c/c	G/G	t/g	C/C	A/A
KWE associated haplotype	G	C	C	C	C	c	C	c	c	G	t	C	A

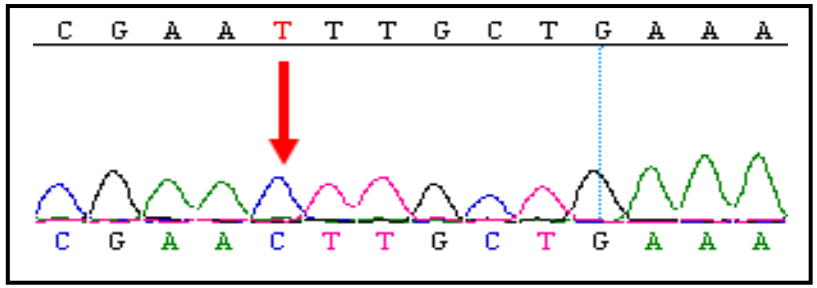
Table 3.10*:* The analysis of the genetic variants identified in the promoter region of *FDFT 1* within KWE families (additional information)

Genetic Variant			A>G Novel upstream	T>C rs17524895 upstream	G>C rs11549153 5'UTR	C>A Novel 5'UTR	cSNP T>C rs1047643 Exon 1
Family 23	**Affection status**	**Individual**					
KWE23_1.2	Unaffected	Grandmother	G/G	C/C	G/G	C/C	T/T
KWE23_2.1	Affected	Son	G/G	C/C	G/C	C/A	T/C
KWE23_3.2	Affected	Grandson	G/G	C/C	C/C	C/A	C/C
Family 26							
KWE26_1.1	Unaffected	Father	A/G	C/C	G/C	C/C	T/T
KWE26_1.2	Affected	Mother	G/G	C/C	C/C	C/C	C/C
KWE26_2.1	Unaffected	Son	A/G	C/C	G/C	C/C	T/C
KWE26_2.2	Affected	Son	A/G	C/C	G/C	C/C	T/C
KWE26_2.3	Unaffected	Daughter	G/G	C/C	G/C	C/C	T/C
Family 45							
KWE45_2.1	Affected	Father	A/G	C/C	G/C	C/A	T/C
KWE45_2.3	Unaffected	Mother	A/G	C/C	G/G	C/C	T/T
KWE45_3.1	Unaffected	Son	A/A	C/C	G/G	C/C	T/T
KWE45_3.2	Affected	Son	G/G	C/C	G/C	C/A	T/C
Family 46							
KWE46_1.1	Affected	Father	G/G	C/C	G/C	C/A	T/C
KWE46_1.2	Unaffected	Mother	G/G	C/C	G/G	C/C	T/T
KWE46_2.1	Unaffected	Daughter	G/G	C/C	G/G	C/C	T/T
KWE46_2.2	Affected	Son	G/G	C/C	G/C	C/A	T/C
Family 48							
KWE48_1.2	Affected	Mother	G/G	C/C	G/C	C/C	T/C
KWE48_2.2	Unaffected	Son	G/G	C/C	G/C	C/C	T/T
KWE associated haplotype			**G**	**C**	**C**	**C**	**C**

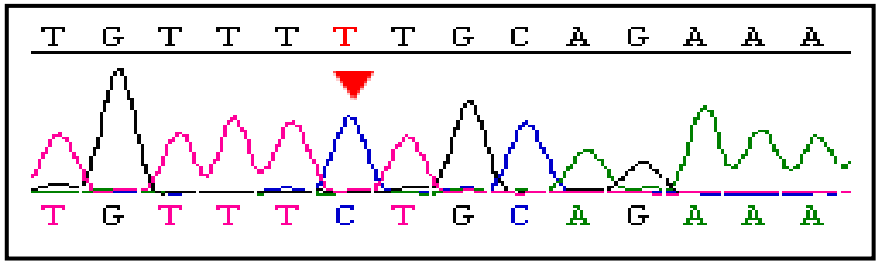
a) Novel 5'UTR variant C>A exchange



b) rs1047643 T>C exchange



c) rs904011 T>C exchange



d) rs9205 G>C exchange

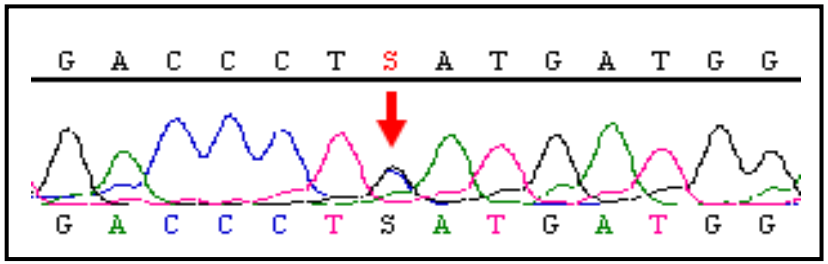


Figure 3.11: Electropherograms of the genetic variants identified within the *FDFT1* gene

Table 3.11: A bioinformatic assessment of the potential functional impact of the genetic variants identified within *FDFT1.*

Bioinformatic			*Variants*	
tool		*L7L (T>C)*	*L211L (T>C)*	*L324L (G>C)*
		rs1047643	*rs904011*	*rs9205*
FastSNP	Risk	Low	Low	Low
	Possible functional effect	missense (conservative)	missense (conservative)	missense (conservative)
	Does the substitution occur in an active protein domain?	No	No	No
	Does the substitution affect regular splicing of the gene?	No	No	No
	Does the substitution change a transcription binding site?	No	No	No
	Does the substitution affect an exon enhancer site?	**Yes**	No	No
	OVERALL prediction	Possibly damaging	Benign	Benign

3.2.2.2 Further variant analysis in a larger group of KWE affected and unaffected individuals

The novel 5'UTR and L7L variants were present in the KWE affected individuals and absent from the control group. Before any comprehensive conclusion can be made as to whether these variants play a role in the manifestation of KWE, a larger group of unaffected individuals (n=15) were sequenced for these variants in order to eliminated small sample size as the reason for the absence of these variants in unaffected individuals. Fifteen unaffected individuals were analysed. The DNA samples used for this analysis were samples that had been obtained from KWE families in previous studies as well as some of those collected for the present study.

Although the novel 5'UTR variant (C>A) was present in only three of the five affected individuals analysed, it is of interest due to its location within the 5'UTR of the gene and its absence in the initial small study in unaffected individuals. Figure 3.12 illustrates the frequency of the genotypes in a group of fifteen unaffected individuals.

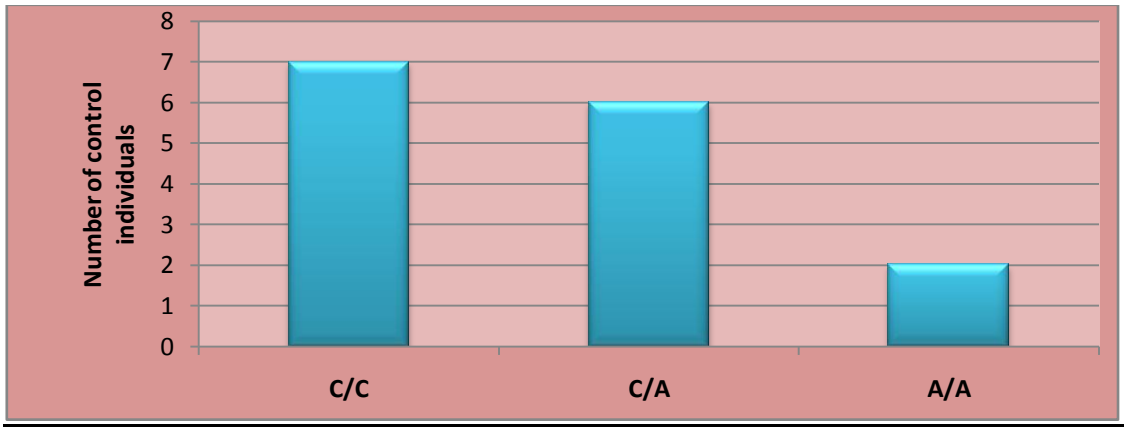


Figure 3.12: The relative frequency of the genotypes observed in unaffected individuals (n=15) for the novel variant identified within the 5'UTR of the *FDFT1* gene.

The L7L *(T<C)* variant was the only variant that segregated exclusively with the KWE affected individuals in the initial small study. Although L7L is a silent synonymous variant, it was decided to further investigate whether this variant was present in a larger group of unaffected individuals (n=15). Figure 3.13 illustrates the frequency of the genotypes in the larger group of unaffected individuals.

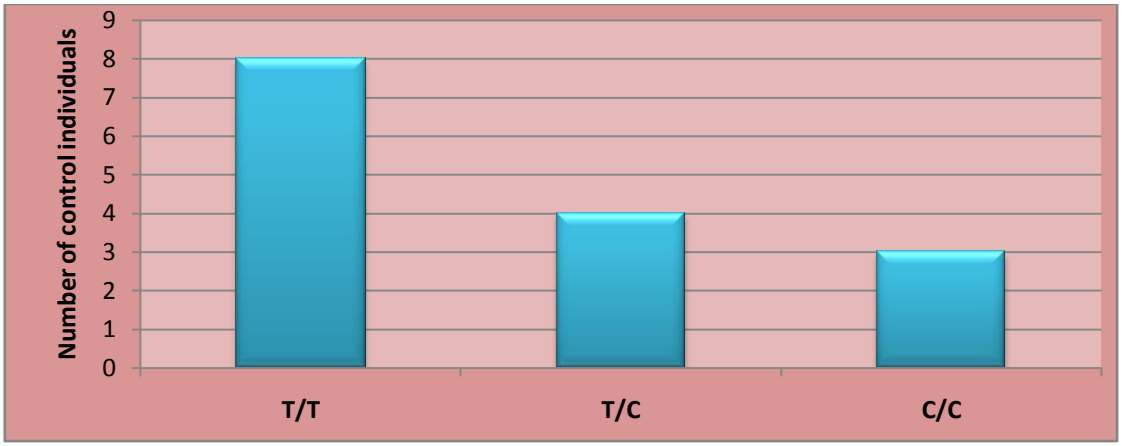


Figure 3.13: L7L (T>C, rs1047643) genotypes in a group of unaffected (n=15) individuals.

Both variants are common polymorphisms and are therefore not the cause of KWE.

3.2.2.3 FDFT1 cDNA analysis

There are four transcript variants for *FDFT1* that encode different *FDFT1* proteins (uc003wuk (1), uc003wuj (2), uc003wui (3) and uc003wug (4) (http://www.ensembl.org/index.html). Transcript 3 is the transcript most commonly found in humans and is referred to as the *FDFT1* reference sequence. Due to differences in splicing, transcripts 1, 2 and 4 encode proteins that differ slightly from each other and from the reference protein. After splicing, the four transcripts produce different cDNAs which can be distinguished through sequencing. Through the analysis of the cDNA sequences obtained for each patient and control, only transcript 3 was

observed. No additional variation was identified at a cDNA level. Table 3.12 lists the genetic variants that are observed in *FDFT1* cDNA. All the coding variants identified through the screening of *FDFT1* at a genomic level are also observed in the cDNA

Table 3.12: Genetic variants identified within *FDFT1* cDNA

Genetic variant	T>C CTT>CTC	T>C TTG>CTG	G>C CTG>CTC
Position	L7L	L211L	L324L
Patients			
KWE50_1.1 (blood)	T/C	C/C	G/G
KWE50_2.1 (saliva)	T/C	C/C	G/C
Controls			
Control#1 (blood)	T/T	C/C	G/G
Control#2 (saliva)	T/T	C/C	G/G
KWE associated haplotype	C	C	G

3.2.2.4 The relative quantification of FDFT1 expression in human epidermal palmar skin

In this study, the data are presented as the N-fold differential expression of *FDFT1* cDNA in KWE affected and unaffected individuals normalized to the internal control gene (*GAPDH*) and relative to the calibrator sample. The real-time PCR data are represented as C_q values which are determined by the RQ software as the cycle at which the first significant increase in PCR product (fluorescence) is detected. The amplification efficiencies of each target and reference gene system as well as the relative efficiency was determined before the quantity of *FDFT1* cDNA (relative units) in the target sample could be determined using the C_q equation (Equation 2).

To determine the efficiency of the *FDFT1* PCR system, the five calibrator dilutions (500ng/µl, 250ng/µl, 100ng/µl, 50ng/µl, and 25ng/µl) were run in triplicate in this assay. The resulting C_q values at each dilution were averaged and this value was plotted against the log of the calibrator dilution. Table 3.13 lists the data used to construct the graph.

Table 3.13: The data used to determine the amplification efficiency of the *FDFT1* system

Calibrator dilutions (ng/µl)	Log calibrator dilution		C_q values		Average C_q value
***FDFT1* assay**		**1**	**2**	**3**	
500	2.7	19.24	19.20	19.15	19.20
250	2.4	20.58	20.57	20.26	20.14
100	2	21.86	21.64	21.05	21.52
50	1.7	22.43	22.72	22.59	22.58
25	1.4	23.60	23.79	23.84	23.74

The data generated for the *FDFT1* amplification system produced a straight line with a slope of -3.40. The standard curve is depicted in Figure 3.14. The *GAPDH* standard curve is also indicated for comparative reasons.

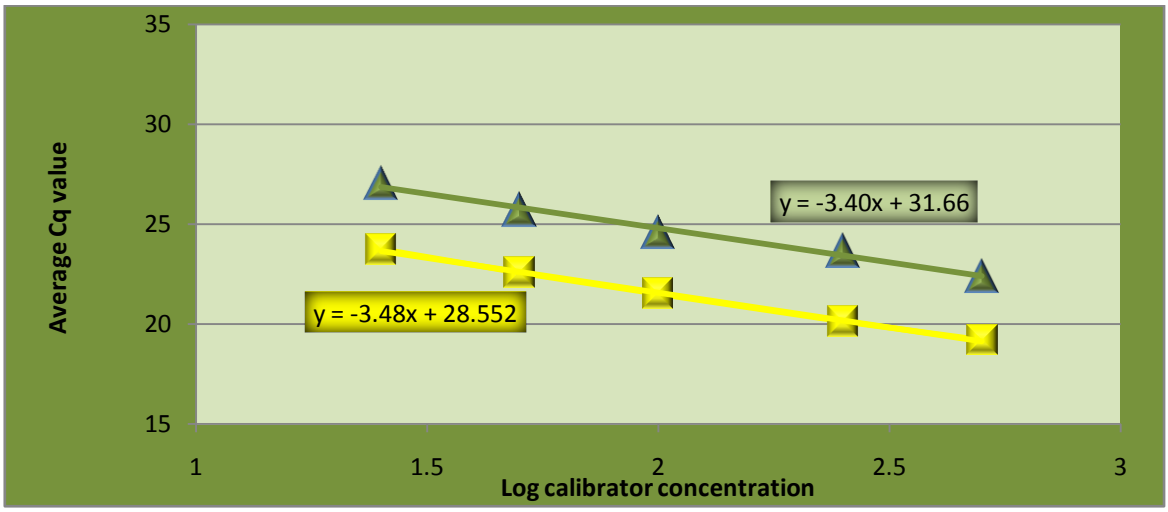


Figure 3.14: The standard curve produced to determine the efficiency of the *FDFT1 (* (▭) and *GAPDH* (△) real-time PCR assays.

The amplification efficiencies (E) of the target and reference gene systems were calculated using the slope value (Figure 3.14) and the equation $E = 10^{1/slope} - 1 \times 100\%$.

***FDFT1* amplification efficiency**: $= 10^{1/slope} - 1 \times 100\%$

$= 10^{1/-3.48} - 1$

$= 10^{0.290} - 1$

$= 1.948 - 1 \times 100\%$

= 94.8 %

***GAPDH* amplification efficiency**: = 96.7% (Section 3.2.1.3.1)

96.7% - 94.8% = 1.9%. Therefore the amplification efficiencies of these systems are relatively close to 100% and are less than 5% apart.

To calculate the efficiency of the target gene system relative to the reference gene system, the data tabulated in Tables 3.4 (*GAPDH*) and 3.13 were used. The average C_q value at 500 ng/µl in the *GAPDH* assay was subtracted from the average C_q value at 500 ng/µl in the *FDFT1* assay. The resulting value was plotted against the log of the calibrator concentration (log of 500). This was done for each dilution. The slope of the graph produced should be less than 0.1 in order for the data to be regarded as efficient and accurate (Ferreira et al., 2006).

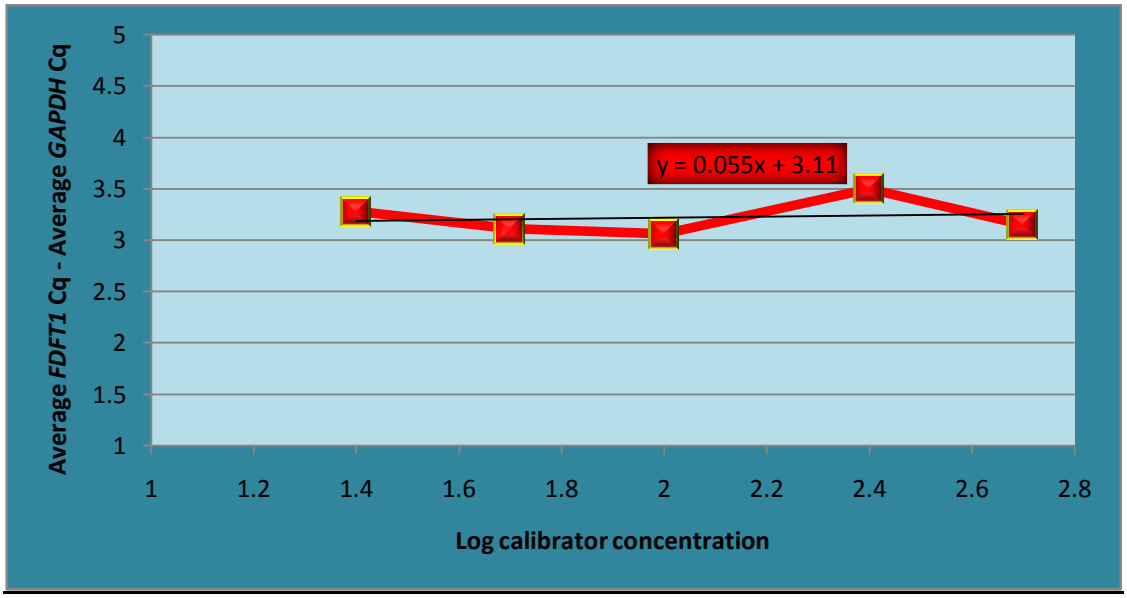


Figure 3.15: Calculation of the relative efficiency of the *FDFT1* real-time PCR system. The slope value is 0.055 (m<0.1), reflecting optimal relative efficiency.

Once it was determined that the *FDFT1* and *GAPDH* amplification systems had optimal and similar efficiencies, the quantity of *FDFT1* mRNA in the target samples, relative to the calibrator sample and normalised by the reference system, was calculated using the $\Delta\Delta C_q$ calculation (Equation 2).

Table 3.15: The C_q values obtained from the *FDFT1* and *GAPDH* patient run (Standard deviation (SD) and Coefficient of variance (CV) shown)

Patient *FDFT1*		**C_q values**		**Average C_q value**	**SD**	**CV (%)**
Experimental run - (A)	**1**	**2**	**3**			
KWE50_1.2	31.02	31.14	31.54	31.23	0.27	0.87
KWE50_2.2	32.03	31.68	31.86	31.85	0.18	0.55
KWE50_3.2	31.78	31.69	31.70	31.72	0.38	1.18
KWE50_3.3	25.94	25.66	25.09	25.9	0.22	0.84
Calibrator	26.6	27.15	26.89	26.88	**0.28**	1.04
Experimental run - (B)	**1**	**2**	**3**			
KWE20_3.11	21.32	21.46	21.54	21.44	0.11	0.51
KWE20_2.5	19.79	19.88	19.95	19.87	0.08	0.4
KWE23_2.1	22.19	22.33	22.46	22.33	0.14	0.63
KWE45_3.6	22.46	22.55	22.67	22.56	0.11	0.49
Calibrator	26.86	27.01	27.13	27.0	**0.14**	0.52
Patient GAPDH		**C_q values**		**Average C_q value**	**SD**	**CV (%)**
Experimental run - (A)	**1**	**2**	**3**			
KWE50_1.2	26.79	26.98	26.98	26.92	0.11	0.47
KWE50_2.2	27.31	27.39	27.35	27.35	0.04	0.15
KWE50_3.2	27.12	27.44	27.29	27.28	0.16	0.52
KWE50_3.3	21.81	21.83	21.77	21.80	0.03	0.16
Calibrator	21.65	21.71	21.69	21.68	0.02	0.09
Experimental run - (B)	**1**	**2**	**3**			
KWE20_3.11	17.52	17.32	17.48	17.44	0.11	0.63
KWE20_2.5	15.92	15.87	15.91	15.9	0.03	0.19
KWE23_2.1	18.18	18.23	18.26	18.22	0.04	0.22
KWE45_3.6	18.66	18.47	18.46	18.53	0.11	0.59
Calibrator	22.28	22.16	22.01	22.15	0.14	0.63

Table 3.16: The C_q values obtained from the *FDFT1* and *GAPDH* control sample run (Standard deviation (SD) and Coefficient of variance (CV) shown)

Controls FDFT1		C_q values		Average C_q value	SD	CV (%)
Experimental run – (A)	**1**	**2**	**3**			
Control #1	27.24	27.52	27.49	27.42	0.15	0.55
Control #2	29.54	29.92	30.28	29.91	0.37	1.24
Control #3	27.58	28.12	27.6	27.76	0.31	1.1
Control #4	31.53	32.30	31.72	31.85	0.3	0.95
Calibrator	27.40	27.35	27.39	27.38	**0.03**	0.2
Experimental run – (B)	**1**	**2**	**3**			
Control #5	24.63	24.60	24.79	24.67	0.10	0.41
Control #6	25.22	25.24	25.33	25.26	0.06	0.24
Control #7	25.42	25.31	25.58	25.44	0.14	0.55
Control #8	23.85	23.86	24.04	23.92	0.11	0.46
Calibrator	26.96	27.01	27.13	27.1	0.09	0.33
Control GAPDH		C_q values		Average C_q value	SD	CV (%)
Experimental run - (A)	**1**	**2**	**3**			
Control #1	22.47	22.54	22.70	22.57	0.12	0.52
Control #2	27.01	27.08	27.24	27.11	0.12	0.44
Control #3	23.17	23.21	23.22	23.20	0.03	0.14
Control #4	26.07	26.09	26.22	26.13	0.08	0.32
Calibrator	22.24	22.33	22.42	22.33	0.09	0.40
Experimental run - (B)	**1**	**2**	**3**			
Control #5	20.22	20.05	20.08	20.12	0.09	0.45
Control #6	20.90	20.77	20.74	20.80	0.09	0.43
Control #7	20.80	20.76	20.84	20.8	0.04	0.19
Control #8	19.06	19.03	18.98	19.02	0.04	0.21
Calibrator	22.24	22.20	22.15	22.20	0.05	0.23

The data in Tables 3.15 and 3.16 were used to calculate the $\Delta\Delta C_q$ value using equation 2. The N-fold differential expression of the *FDFT1* gene in the affected and unaffected samples are expressed as $2^{-\Delta\Delta Cq}$. From the $\Delta\Delta C_q$ value, the $2^{-\Delta\Delta Cq}$ value is determined using equation 3. The $\Delta\Delta C_q$ and corresponding $2^{-\Delta\Delta Cq}$ value for each patient and control is listed in Table 3.17. Increased mRNA expression is defined as N-fold ≥2.0, "normal" expression is an N-fold ranging from 0.5001 to 1.9999, and decreased mRNA expression is N-fold ≤0.5 (Ferreira et al., 2006).

Table 3.17: The $\Delta\Delta C_q$ value and corresponding $2^{-\Delta\Delta Cq}$ of each patient and control sample

Sample	ΔΔ C_q value	$2^{-\Delta\Delta Cq}$
KWE50_1.2	-0.89	**2.36**
KWE50_2.2	-0.70	**1.62**
KWE50_3.2	-1.24	**1.85**
KWE50_3.3	-1.1	**1.67**
KWE20_3.11	-0.85	**2.14**
KWE20_2.5	-0.88	**1.84**
KWE23_2.1	-0.74	**1.77**
KWE45_3.6	-0.82	**1.84**
Control #1	-0.20	**1.15**
Control #2	-0.30	**1.23**
Control #3	-0.49	**1.52**
Control #4	0.67	**1.33**
Control #5	-0.46	**1.38**
Control #6	0.65	**1.35**
Control #7	-0.26	**1.20**
Control #8	0	**1**

To visually assess the differential expression of *FDFT1* in each patient and control sample, a simple bar graph was constructed using the data in Table 3.17.

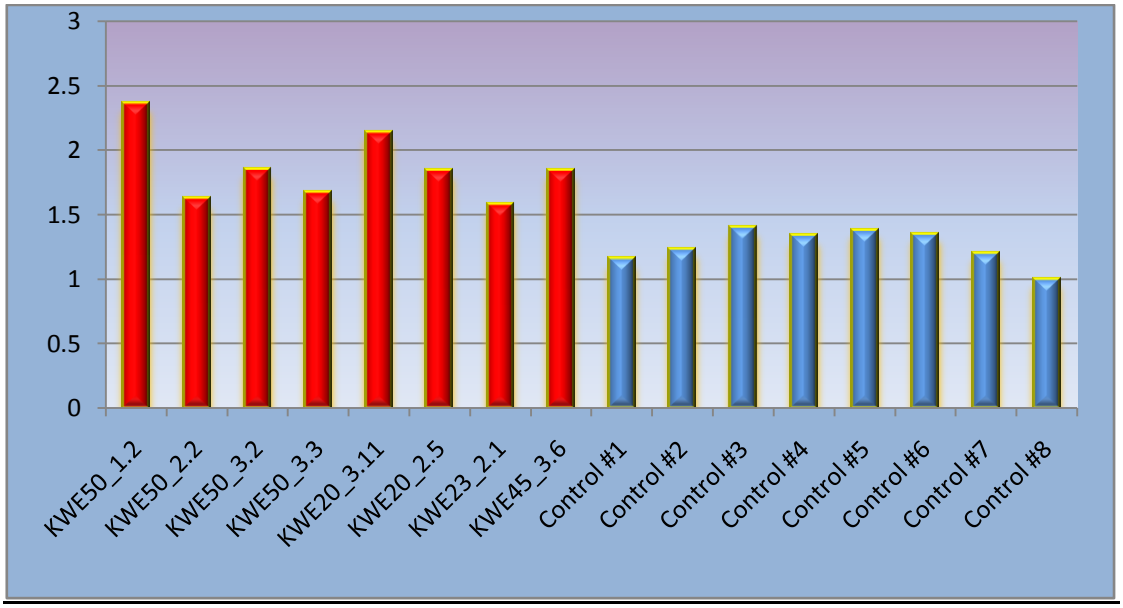


Figure 3.16: Normalized *FDFT1* expression in plantar skin from KWE affected and control individuals.

The Wilcoxon two-sample test (a non-parametric one way ANOVA test) and a non-parametric Mann-Whitney U test showed that there seems to be a trend towards increased expression of *FDFT1* in patients, however, it is not significant. An interesting

observation is that the more severely affected individuals have a higher level of expression than the mildly affected individuals (Table 3.19).

Table 3.18: Statistical analysis results

Statistical test	p-value
Two-tailed test	
Wilcoxon	0.063
Mann-Whitney U	0.078

Table 3.19: The correlation between the increased level of *FDFT1* expression and the severity of the individual.

Individual	Severity*	Phenotype of patient	$2^{-\Delta\Delta Cq}$
KWE50_1.2	+++	Severe peeling on entire palmar surface, very red palms	2.36
KWE50_2.2	+	Palms reddish, one or two sites where peeling is present	1.62
KWE50_3.2	++	Peeling moderate covers most of palms, red palms	1.85
KWE50_3.3	+	Very mild phenotype, scarce peeling on palms	1.67
KWE20_3.11	+++	Severe peeling that covers entire palm,	2.14
KWE20_2.5	++	Moderate peeling, very red palms	1.84
KWE23_2.1	+	One or two sites on the palm where peeling occurs	1.77
KWE45_3.6	++	Moderate peeling, red palms	1.84

*+++: severe, ++: mild-severe, + mild

3.2.3 *DUB-3* analysis

Six coding genetic variants were identified within *DUB-3* (Table 3.20). Three of the variants observed are novel and there is no predictive data available from FastSNP, Polyphen and SNPs3D regarding the possible effect that the novel variants will have on the protein's structure and function. The analysis of *DUB-3* was difficult because of the limited amount of information and data available regarding this deubiquitinating gene.

K40N(G>C)

This is a novel variant that occurs in the UPB5 conserved domain of the protein and results in a lysine to asparagine exchange. The variant C allele (Figure 3.17a) is observed in both the patient and control groups and is predicted not to have an effect on the normal splicing of the gene or change a transcription factor binding site (Table 3.21). It is unlikely that this benign variant would result in the manifestation of the KWE phenotype.

***P69L* (rs3988859, C>T)**

This low to medium risk missense mutation is predicted to cause a conservative change in the protein with the substitition of proline with lysine. It is predicted to be a benign polymorphism that has no known effect on the protein's structure or function. Predicted data suggest that the variant will not affect the normal splicing of the gene and will not alter a transcription binding site (Table 3.21). The variant T allele (Figure 3.17b) is present in both the patient and control groups.

E244K (G>A)

This novel variant involves the substitution of glutamic acid with lysine. The variant A allele (Figure 3.17c) is observed in both the patient and control groups. There is no predictive data available from SNPs3D and FastSNP regarding this novel variant. The exchange of an acidic amino acid, glutamic acid, with a basic one, lysine, is a major change between amino groups, however because this variant does not segregate with affected individuals, it can not be the cause of KWE manifestation.

***D332D* (rs13274255, C>T)**

The synonymous variant is predicted to be a very low risk variant that will have no effect on the structure or function of the protein. According to the predictive bioinformatic tools, the transversion from a cytosine to a thymine nucleotide will not alter a transcription factor binding site or alter the splicing of the gene. The variant T allele (Figure 3.17d) is present in both the patient and control groups.

A403V(C>T)

This novel genetic variant results in the exchange of alanine with valine and does not occur within the conserved domains of the protein. Both valine and alanine are hydrophobic, non-polar amino acids and so an exchange from one to the other is not expected to cause a deleterious effect on the structure and/or stability of the protein. The variant T allele (Figure 3.17e) is present in the patient and control groups.

***K458R* (rs12543578, G>A)**

This low to medium risk polymorphism involves the substitution of lysine with arginine and the variant A allele (Figure 3.17f) is present in both the patient and control groups. Predictive data suggest that this variant will not alter a transcription binding site but does result in the diminishment of an ESE motif which may affect the splicing regulation of the gene and is predicted to be possibly damaging (Table 3.21).

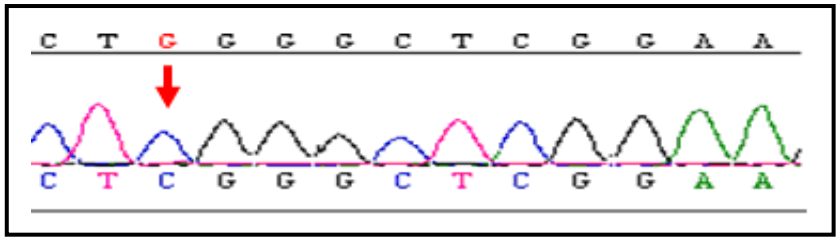
Table 3.20: The genetic variants identified in *DUB-3* by DNA sequencing

Genetic Variant	cSNP G>C novel	cSNP C>T rs3988859	cSNP G>A novel	cSNP C>T rs13274255	cSNP C>T novel	cSNP G>A rs12543578
Position	K40N	P69L	E244K	D332D	A403V	K458R
Patients						
KWE46_1.1	G/C	C/T	G/A	C/C	C/T	G/A
KWE23_2.1	G/C	C/T	G/A	T/T	C/C	G/A
KWE36_1.1	G/C	C/T	G/A	T/T	C/T	A/A
KWE50_1.2	G/C	C/T	G/A	C/C	C/T	G/A
KWE50_2.1	G/C	C/T	G/A	C/C	C/T	G/A
Controls						
KWE46_2.1	G/G	C/T	G/A	T/T	C/C	G/A
KWE23_1.2	C/C	T/T	A/A	T/T	C/C	A/A
KWE36_2.1	G/G	C/T	G/G	C/T	C/T	A/A
Control#1	C/C	T/T	A/A	T/T	C/C	A/A
Control#2	G/C	T/T	G/G	C/C	C/C	A/A

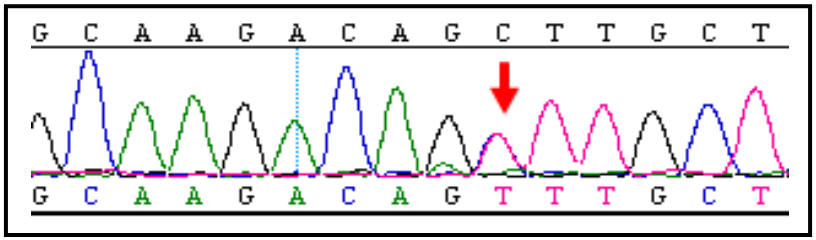
Table 3.21: The bioinformatic assessment of the functional impact the characterised[•] genetic variants identified within *DUB-3* may have on the protein

Bioinformatic			***Variants***	
tool		***P69L (C>T)***	***D332D (C>T)***	***K458R (G>A)***
		rs3988859	***rs13274255***	***rs12543578***
FastSNP	Risk	Low-medium	Very very low	Low-medium
	Possible functional effect	missense (conservative)	missense (conservative)	missense (conservative)
	Does the substitution occur in an active protein domain?	No	No	No
	Does the substitution affect regular splicing of the gene?	No	No	No
	Does the substitution change a transcription binding site?	No	No	No
	Does the substitution affect an exon enhancer site?	No	No	**Yes**
	OVERALL prediction	Benign	Benign	Possibly damaging

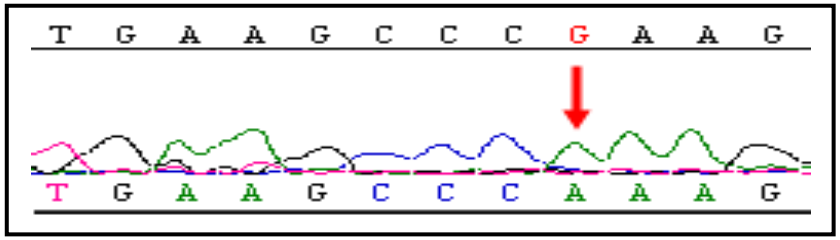
a) Novel variant G>C exchange



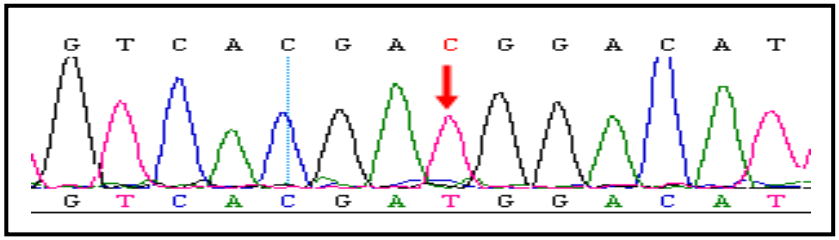
b) rs3988859 C>T exchange



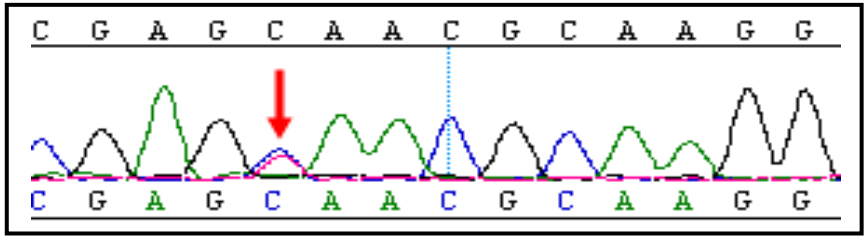
c) Novel variant G>A exchange



d) rs13274255 C>T exchange



e) Novel variant C>T exchange



f) rs12543578 G>A exchange

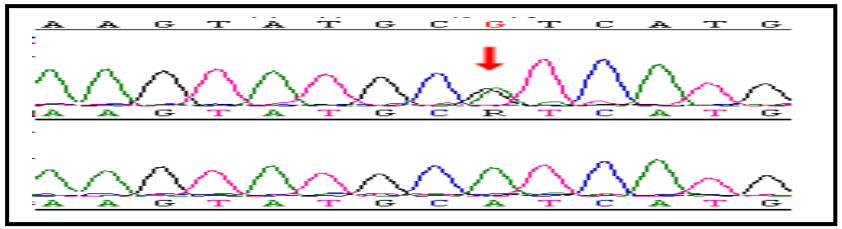


Figure 3.17: The genetic variants identified within the coding region of *DUB-3*

3.2.4 *C8orf49* analysis

No genetic variants were observed in the 5'UTR of the open reading frame however, several predicted nonsynonymous variants and a predicted nonsense variant were observed (Table 3.22). Predictive data was not available from SNPs3D or PolyPhen for this open reading frame.

***G26R* (rs804285, G>C)**

This medium to high risk missense mutation involves the substitution of glycine with arginine and is predicted to cause a non-conservative change in the protein that may affect the protein's structure. It is predicted that the introduction of a cytosine nucleotide will affect an exon enhancer site (Table 3.23). The variant C allele (Figure 3.18a) is found in both the patient and control groups.

***I81V* (rs809204, G>A)**

This low to medium risk missense mutation involves the substitution of isoleucine with valine and results in a conservative change in the protein. It is predicted that this variant will not affect splicing of the open reading frame and that the substitution of a guanine nucleotide with alanine will not alter a transcription factor binding site (Table 3.23). The variant A allele (Figure 3.18b) is present in both the patient and control groups.

***R95G* (rs13281294, T>C)**

This medium to high risk missense mutation that involves the substitution of arginine with glycine is predicted to cause a non-conservative change in the protein. The variant C allele (Figure 3.18c) is present in both the control and patient groups.

***G113G* (rs2740431, T>A)**

The synonymous variant is predicted to be a very low risk variant that that will have no effect on the function or structure of the protein. The variant A allele (Figure 3.18d) is present in both the patient and control groups.

***L138L* (rs811966, C>T)**

The synonymous variant is predicted to be a low risk variant that will not affect the normal functioning or structure of the putative protein in any way (Table 3.23). The variant T allele (Figure 3.18e) is present in both the patient and control groups.

***V159I* (rs36018280, G>A)**

This low to medium risk missense mutation that involves the substitution of valine with isoleucine is predicted to cause a conservative change in the protein that should not have a damaging effect on the structure or function of the protein (Table 3.23). The variant A allele (Figure 3.18f) is present in both the patient and control groups.

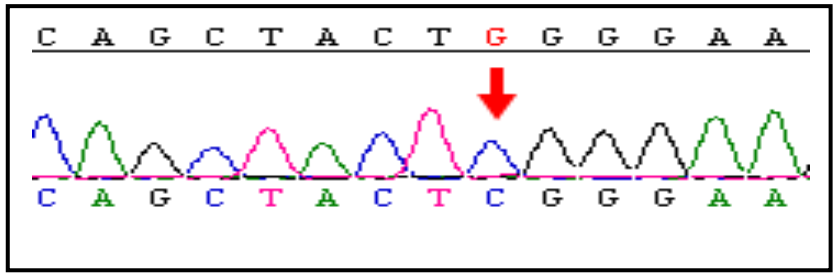
***C194X* (rs809203)**

This very high risk nonsense mutation introduces a premature stop codon that could result in the production of a truncated protein. Predictive data suggests that this variant is damaging to the protein (Table 3.23). The variant A allele (Figure 3.18g) has been previously described and is observed in both the patient and control groups.

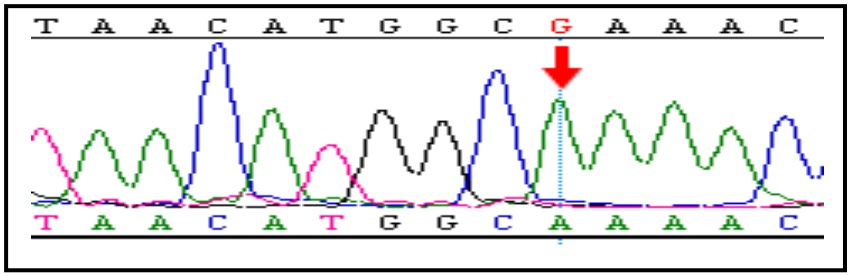
Table 3.22: The genetic variants identified within the open reading frame, *C8orf49*

Genetic Variant	cSNP G>C rs804285	cSNP G>A rs809204	cSNP T>C rs13281294	cSNP T>A rs2740431	cSNP C>T rs811966	cSNP G>A rs36018280	cSNP T>A rs809203	3ÚTR C>T rs7015453	3'UTR T>C rs2686206	3'UTR G>A novel
Position	**G26R**	**I81V**	**R95G**	**G113G**	**L138L**	**V159I**	**C194X**	**C1222T**	**T1551C**	**G1697A**
Patients										
KWE46_1.1	C/C	A/A	T/T	A/A	C/C	A/A	T/T	C/T	T/T	A/A
KWE23_2.1	C/C	G/A	T/T	A/A	C/T	G/A	T/A	C/T	T/T	G/A
KWE36_1.1	C/C	A/A	C/C	T/T	C/T	G/A	T/T	C/T	T/T	G/A
KWE50_1.2	C/C	A/A	T/T	A/A	C/C	A/A	T/T	C/T	T/T	A/A
KWE50_2.1	C/C	G/A	T/T	A/A	C/T	G/A	T/A	C/T	T/T	G/A
Controls										
KWE46_2.1	C/C	G/A	T/T	A/A	C/T	G/A	T/A	C/C	T/C	G/A
KWE23_1.2	C/C	A/A	T/T	T/T	C/T	A/A	T/T	C/T	C/C	G/G
KWE36_2.1	C/C	A/A	C/C	T/T	C/C	G/G	T/T	C/C	T/T	G/G
Control#1	C/C	A/A	T/T	T/T	C/T	G/G	T/T	C/T	C/C	G/G
Control#2	C/C	G/A	C/C	A/A	C/C	G/G	T/A	C/C	C/C	G/A

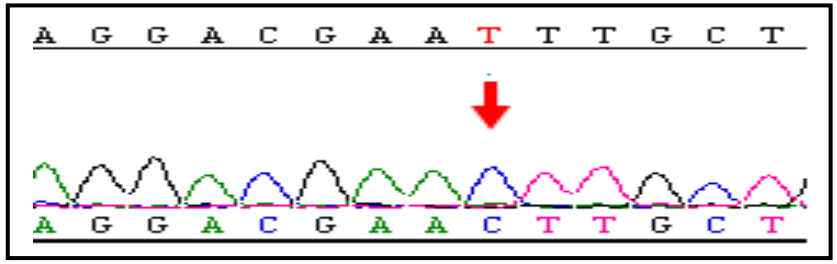
a) rs804285 G>C exchange



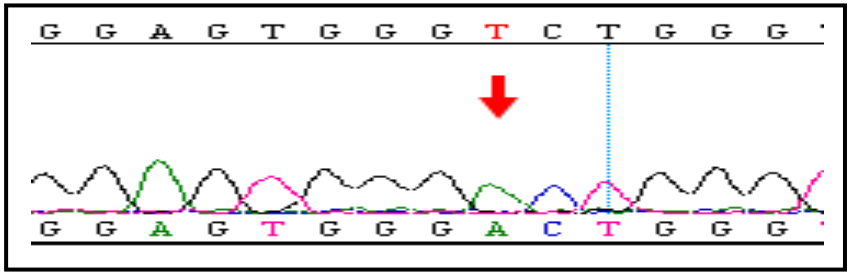
b) rs809204 G>A exchange



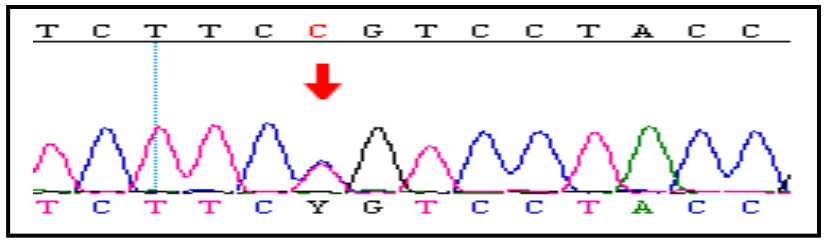
c) rs13281294 T>C exchange



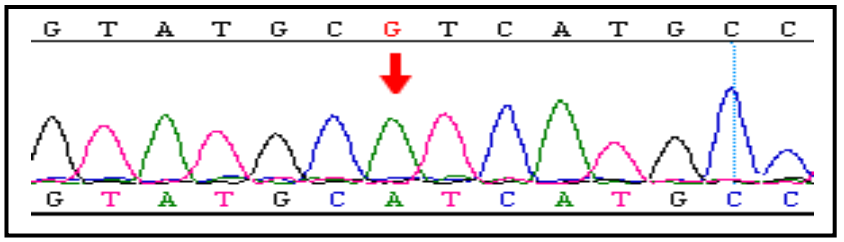
d) rs2740431 T>A exchange



e) rs811966 C>T exchange



f) rs36018280 G>A exchange



g) rs809203 T>A exchange

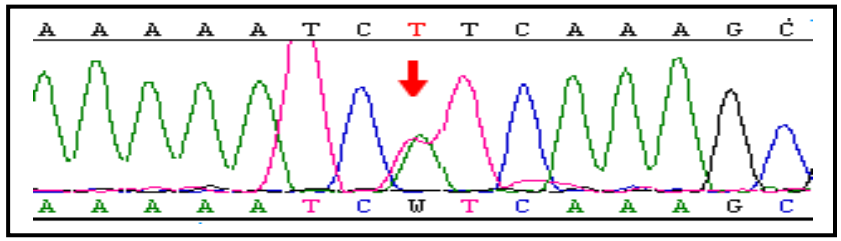


Figure 3.18: The genetic variants identified within *C8orf49*

Table 3.23: A bioinformatic assessment of the potential functional impact the genetic variants identified within the open reading frame, C8orf49 may have on the protein

Bioinformatic				***Variants***				
tool		***G26R (G>C)***	***I81V (G>A)***	***R95G (T>C)***	***G113G (T>A)***	***L138L (C>T)***	***V159I (G>A)***	***C194X (T>A)***
		rs804285	***rs809204***	***rs13281294***	***rs2740431***	***rs811966***	***rs36018280***	***rs809203***
FastSNP	Risk	Low-medium	Low-medium	Low-medium	Very Low	Low-medium	Low-medium	Very high
	Possible functional effect	Missense (conservative)	missense (conservative)	missense (conservative)	Synonymous	Missense (conservative)	Missense (conservative)	**Nonsense**
	Does the substitution occur in an active protein domain?	No	No	No	No	No	No	No
	Does the substitution affect regular splicing of the gene?	No	No	No	No	No	No	No
	Does the substitution change a transcription binding site?	No	No	No	No	No	No	No
	Does the substitution affect an exon enhancer site?	No	No	No	No	No	No	No
	OVERALL prediction	Benign	Benign	Benign	Benign	Benign	Benign	Damaging

CHAPTER 4

DISCUSSION

4.1 THE FINE MAPPING OF THE KWE CRITICAL REGION BY ANALYSIS OF THE KWE FAMILY HAPLOTYPES

The main objective of the study was to identify the gene that is responsible for the manifestation of KWE. To achieve this objective, it was necessary to validate the previously determined KWE critical region.

In the present study, the haplotypes were constructed twice for each family. The first set of haplotypes was constructed under the assumption that there was no inversion involving the KWE critical region. The second set of haplotypes (illustrated in Figures 3.1–3.4), was constructed under the assumption that the KWE mutation lies on an inverted chromosome. This was done because of three colour FISH experimental data that suggest that the KWE mutation lies on an inverted chromosome (P. Willem, personal communication). In both analyses, the complete founder KWE haplotype was present in 2 of the 4 KWE South African families and parts of this haplotype were present in the remaining two families.

4.1.1 Family 23 and family 46

The affected members of these two families demonstrated the complete founder haplotype. It is clear that this haplotype is conserved and stably inherited by affected children from their affected parent. Unaffected family members did not inherit this

ancestral haplotype. Due to a lack of informative recombinations in these families, no constructive information with regard to the position of the KWE locus was obtained.

4.1.2 Family 26

Although no recombination events were observed in the two generations of family 26 represented in the pedigree, this family did prove to be informative because the affected members of this family inherited only part of the founder haplotype. The part of the founder haplotype that is not conserved among the KWE affected individuals can be excluded from the KWE critical region. Regions of the founder haplotype that are not conserved are unlikely to harbour the gene, which in a mutant state, will result in the manifestation of KWE. The haplotype for this family shows that the region between and including markers D8Sdi and D8S552 does not segregate with the KWE phenotype and can be excluded from the KWE critical region. This family places the KWE gene in the region between and including markers D8S1819 and D8Stet. It is important to keep in mind that this family, as illustrated by Figure 3.2, does not demonstrate a cross-over, and that the box illustrates which part of the founder haplotype is conserved. There may have been a cross-over that occurred in an earlier generation of this family which resulted in the inheritance of only part of the founder haplotype by KWE affected individuals.

4.1.3 Family 36

Family 36 proved to be the most informative family analysed in this study. The different allele observed for marker D8S1819 in this family may be a result of a number of reasons. Firstly, it is possible that a recombination event occurred in an earlier

generation of this family and this recombination was stably inherited through the consecutive generations. The region between markers D8S1819 and D8S1759 is termed the REPP repeat region and is a region that is known to demonstrate a high frequency of recombination. This may be a second reason for the variant allele observed at marker D8S1819. In order to determine with confidence whether the variant 218 allele present in this family is due to an earlier recombination associated with KWE or if it is as a result of other factors, it would be necessary to identify and genotype microsatellite repeat markers located between markers D8S1819 and D8S1759 in the area before and after the REPP region.

An informative recombination event was observed in unaffected individual 2.3 who has part of the founder haplotype. The part of the founder haplotype that is present in an unaffected individual can be excluded from the KWE critical region since KWE is highly penetrant, but there may be a small number of individuals who harbour the mutation but remain clinically unaffected. It is accepted that individuals who are clinically affected definitely have the KWE causal mutation. In families, like family 36, where the one breakpoint is premised on a clinically unaffected individual, a measure of uncertainly remains with regard to the validity of the breakpoint suggested in that case. The exact distal boundary of the KWE critical region could not be defined with confidence due to the uninformativity of marker D8Stet (represented by allele 231 in individual KWE36_2.3 of family 36). This uninformativity makes it difficult to exclude the region around this marker from the KWE critical region. The affected mother is homozygous for the 231 allele, therefore it cannot be determined whether the 231 allele present on the maternal allele of unaffected individual 2.3 forms part of the M1

or the $M2_{KWE}$ haplotype. For this reason, marker D8Stet was not excluded from the KWE region. The uninformativeness of this marker inhibits any further defining and reducing of the KWE critical region.

The uninformativeness of marker D8Stet does prevent the confident determination of the exact distal boundary of the KWE region. However in order to prevent the exclusion of possible candidate genes, D8Stet was included in the KWE critical region for this study. In future KWE studies, it would be necessary to identify additional repeat markers located between markers D8Sdi and D8Stet as well as between markers D8Stet and D8S265. If the markers render informative results it may be possible to identify a distal boundary for the KWE critical region. Additional markers identified between markers D8S1819 and D8S1759 will possibly identify a proximal boundary. The lack of more informative recombinations observed within these families forces one to rely on the part of the founder haplotype that is conserved among KWE affected individuals. Because KWE is a highly penetrant disease, any part of the ancestral haplotype that appears in unaffected individuals is excluded from the KWE critical region. The validated KWE critical region is the region between and including markers D8S1759 (expanding proximally toward the REPP repeat region) and D8Stet (Figure 3.6) and is therefore defined as the region between markers D8S1819 and D8Sdi.

4.2 MUTATIONAL ANALYSIS

There are many different types of variation found within the human genome namely, insertions, deletions and single base changes. The most commonly found variation within the human genome is single base changes (http://science.education.nih.gov/supplements/nih1/Genetic/guide/genetic_variation1.htm).

Genetic variations present in the 5'UTR of positional candidate genes should not be overlooked. Many studies have successfully demonstrated that variants that occur within the 5'UTR of a gene could affect the downstream translation of the gene, increasing a person's susceptibility to developing a particular disease. A novel genetic variant present in the 5'UTR of the CD28 gene is associated with an increase in an individual's susceptibility to HIV infection (Soraino et al., 2004). In 2008, Ivanov et al. demonstrated that the direct mutation of the uAUG of uORFs present in the 5'UTR of certain genes, resulted in a substantial increase in the gene's expression. In many cases, it was shown that an upstream open reading frame located in a gene's 5'UTR regulates the downstream expression of the gene by either increasing or decreasing its level of expression. Genetic variants that directly affect uORFs, by abolishing their start or stop codons, could affect the expression of the protein. For these above mentioned reasons, the analysis of the variants observed in the 5'UTR of the candidate genes has become an essential part of mutational analysis. Genetic variants present in the coding, non-coding and regulatory regions of a gene may be implicated in disease and for this reason, require thorough analysis.

Single base changes that occur in the coding region of a gene are classified as either non-synonymous or synonymous. Synonymous variations (also termed silent substitutions) are the substitution of one nucleotide for another such that the amino acid sequence produced is not altered (Angelico-Sato and Toha, 2005). This type of substitution rarely has a deleterious effect on the protein's structure or stability and is unlikely to result in the manifestation of a disease phenotype. Non-synonymous variations involve the exchange of one amino acid with another and are of major interest because this type of variation currently accounts for approximately half the known gene lesions responsible for inherited diseases (Ng and Henikoff, 2003). Non-synonymous variations are commonly termed non-synonymous single nucleotide polymorphisms (nsSNPs) (Dobson et al., 2006). Some nsSNPs are associated with the manifestation of a disease while others are not associated with any change in phenotype and are regarded as benign or neutral nsSNPs. But how does one decide if a particular genetic variant is likely to be pathogenic? Several studies have attempted to predict the functional consequences of a nsSNP based on particular attributes of the polymorphism (Dobson et al., 2006). Some attributes depend on the sequence information such as the location of the SNP and also if the protein sequence containing the nsSNP has a known 3D structure. Many rules by which a nsSNP is predicted to be deleterious or benign have emerged from a number of different studies. Some of the rules suggest that the majority of disease associated nsSNPs are located in the surface pockets of protein structures (Stitziel et al., 2004), affect protein stability (Wang and Moult, 2001) and are conserved across species (Saunders and Baker, 2002). There are

many computational tools available that predict whether a particular nsSNP is possibly deleterious or benign.

Polyphen, FastSNP and SNPs3D are the bioinformatics tools that were used to analyse the possible molecular effect that the non-synonymous coding variants could have on the protein. Before PolyPhen predicts the possible molecular effect a variant could have on the protein, it takes into account the position of the variant within the protein and whether this position is conserved across orthologues. Theoretically, if a variant occurs at a position that is highly conserved across orthologues, it is more likely to be involved in many important aspects of protein function and structure (Russel et al., 2003). The majority of amino acids that are located within the active domains of proteins are highly conserved. Polyphen also calculates a PSIC score which indicates how different the two amino acids, involved in the amino acid exchange, are in side chain structure and charge. A low PSIC score indicates that the amino acids involved in the substitution are similar in charge and structure and that this type of substitution occurs commonly in the particular protein family. A high PSIC score indicates that the two amino acids involved in the substitution are different in respect to charge and polarity and that this type of substitution occurs rarely in the particular protein family. The substitution of amino acids with large differences in their side chain electrostatic charge could have a possibly damaging effect on the protein structure.

The charge of an amino acid side chain is extremely important in terms of protein structure since ion binding can stabilize a protein structure and an unpaired charge can disrupt the protein structure (Voet et al., 2001). Also, charged amino acids are highly

hydrophilic and are found on the outside of a protein (Affabuddin and Kundu, 2007). Before predicting what effect a certain amino acid substitution will have on a protein, SNPs3D first determines whether the substitution introduces backbone strain, changes the hydrophobicity of the protein, results in a change in electrostatic charge, breaks a disulphide bond or results in the loss of hydrogen bonds within the protein, all aspects which are extremely important in terms of protein stability and structure. The introduction of backbone strain, loss of hydrogen bonds, breakage of disulphide bonds (http://www.massey.ac.nz/~wwbioch/Prot/thirds/tertext.htm) and a change in electrostatic charge will all have a negative impact on the stability of the protein. A single amino acid is often termed a residue and the linked series of carbon, nitrogen, and oxygen atoms are known as the protein backbone. Any strain on the protein's backbone could have a negative effect on the protein's structure in terms of accelerated unfolding or misfolding of the protein (Ventura et al., 2002; Macallister et al., 2005). Disulphide bonds play an important role in increasing the conformational stability of a protein by constraining the unfolded conformation of the protein thereby decreasing conformational entropy (Zavodszky et al., 2000). Hydrogen bonds play an important role in determining the three-dimensional structure of proteins. The overall effect of hydrogen bonds in proteins is that of providing structural rigidity and not stability (Rose, 1993; Darby and Creighton, 1995) although it has been demonstrated that hydrogen bonds have a conserved role in protein folding (www.physorg.com/news10776.html). Although hydrogen bonds contribute little to a proteins stability, they do function to align molecular groups into specific orientations that give the protein its defined structure. Hydrophobic interactions are the most

important non-covalent forces responsible for phenomena such as structure stabilization (Privalov and Gill, 1988) and folding of proteins (Dill, 1990). A change in the hydrophobicity of the protein due to amino acid substitutions could have a negative effect on the protein structure. SNPs3D also calculates the entropy of a particular amino acid substitution. A substitution occurring in the site of the protein that is not tolerable to substitutions (an active domain or highly conserved region) will have a low entropy while a substitution occurring in the sites of the protein that are more tolerable of amino acid substitutions has a high entropy.

FastSNP predicts whether a particular amino acid substitution will affect the regular splicing of the gene in any way, changes a transcription start site or affects an exon enhancer site. A genetic variant that affects the regular splicing of a gene can result in the production of an alternate or truncated protein. This usually has a large effect on the function of the protein. The altering of a transcription start site can result in the over or under expression of a gene or inhibit gene expression completely. An exonic splicing enhancer (ESE) is a DNA sequence motif consisting of 6 bases within an exon that directs, or enhances, accurate splicing of pre-mRNA into messenger RNA (mRNA) (Cartegni et al., 2003). The presence of exonic splicing enhancers is essential for proper identification of splice sites by the cellular machinery. Computational methods are available which identify ESEs. ESEs are clinically significant because synonymous point mutations previously thought to be silent mutations located in an ESE can lead to exon skipping and the production of a non-functioning protein (Blencowe, 2000). Because KWE is a rare autosomal dominant condition, a non-synonymous genetic variant which occurs in the coding region of a candidate gene and which is present in a heterozygous

form in only the patient group, would be a potential disease causing mutation. It is thought that KWE is a result of a founder effect and for this reason, all affected KWE patients should carry the same disease causing mutation. Since KWE is highly penetrant, genetic variants that are found in unaffected individuals are unlikely to be the cause of KWE and are therefore excluded from further analysis.

4.2.1 *CTSB* as a KWE gene candidate

The *CTSB* cDNA transcripts identified in both the patient and control groups had an identical sequence to the transcript translated from transcript NM_147780. This transcript has two uORFs, one which is 25 amino acids in length and the other which is 46 amino acids in length. There was one variant identified within the *CTSB* genes 5' UTR (rs17154017). This variant does not affect either of the uORFs in any way and is observed in only the control group. It is not clear whether the uORFs affect the downstream expression of this gene, but because both the patient and control groups have the same *CTSB* transcript in both blood and saliva tissue it is unlikely that any possible change in the expression of *CTSB* due to the presence of the uORFs will result in the manifestation of KWE or increase one's susceptibility to developing the disease.

After sequencing the cDNA of the patients and controls, it was clear that the 19 bp duplication present in IVS-7 has no functional effect on the *CTSB* protein. Intronic duplications and deletions that result in the manifestation of disease are those which affect the regular splicing and ultimately the coding region of the gene. However, the

majority of intronic duplications and deletions observed in the human genome are polymorphic and have no phenotypic effect in carriers. This study cannot predict whether this novel 19bp duplication may increase an individual's susceptibility to developing other diseases however, the duplication does not appear to increase an individual's susceptibility to developing KWE as not all the patients carried this duplication. Therefore it is unlikely that this duplication plays any role in the development of KWE.

The genetic variants identified within the coding region of *CTSB* were computationally classified as low to medium risk SNPs that would not have a deleterious effect on the function and/or structure of the *CTSB* protein. Each of these variants occurs at a site in the protein that tends to be more tolerant to substitutions (demonstrated by high entropy scores; Table 3.2) and are not highly conserved across orthologues. According to predictive data, none of the observed variants occurs in the active protein domain, affects the regular splicing of the gene, changes a transcription start site or affects an exon enhancer site. None of the substitutions introduces protein backbone strain, changes the hydrophobicity of the protein, results in the breakage of a disulphide bridge or causes a loss of hydrogen bonds. Taking these predictive data into consideration along with the observation that each variant allele is also present in some of the unaffected individuals, these variants are regarded as benign variants that play no role in the manifestation of KWE. As previously mentioned, KWE is an autosomal dominant condition and the pathogenic mutation is expected to be observed in a heterozygous state in KWE patients only. The intronic variants identified within this gene are observed in both the patient and control group and do not affect

the regular splicing of the gene. Through the analysis of CTSB cDNA it was possible to determine that there were no deep intronic variants present in the intronic regions of this gene that could affect the regular splicing of the gene or the translation of the protein. Partial or whole exon deletions would also be identified through cDNA analysis. The cDNA produced from each patient and control individual were identical to the reference CTSB cDNA. The variants identified within the non-coding region of the gene are predicted to be benign in respect to KWE. In past KWE studies, the promoter sequence of this gene was analysed but no genetic variants were identified that segregated with the patient cohort (Silke Arndt, unpublished data).

The quantitative analysis of *CTSB* expression in affected and non-affected palmar skin was performed with the hope to determine whether there was differential expression of this gene in KWE patients. The amplification and relative efficiency of the *CTSB* and reference gene system fell within the range theoretically accepted as optimal (Ogino, 2001). The variances among the set of data collected for each patient and control individual (reported as a %) were all relatively low statistically confirming that there is little variance among each set of data (C_q values) which is ideal. Variance between a set of data should be less than 5% (Ellis et al., 2009) to obtain results of the highest accuracy. Low CV values demonstrate good reproducibility and precision of the data and indicate that the average value is a good representative of that set of data (Jebbink et al., 2003).

The relative quantity of *CTSB* cDNA in the patients did not differ significantly from the relative amount of *CTSB* cDNA found in the unaffected individuals. There does not

appear to be a significant difference between the relative expression of *CTSB* in the patients and controls (Figure 3.11). There is no significant difference between the expression of *CTSB* in the patients in comparison to the controls.

At this point, the entire 5'UTR, coding region and intron/exon boundaries of *CTSB* has been thoroughly analysed through the mutational analysis of genomic DNA. Intronic variants that could cause aberrant splicing were indirectly analysed through the analysis of the CTSB cDNA. The promoter region was indirectly screened for genetic variants through the quantitative analysis of the expression of *CTSB* as a variant in the promoter region of the gene can affect the level of the gene's expression (Section 3.4). Although in the literature, *CTSB* appears to be an excellent candidate gene for KWE, this study has not identified any genetic variations within this gene that may result in the manifestation of the disease. It is unlikely that *CTSB* is the KWE gene.

4.2.2 *FDFT1 as a KWE gene candidate*

There were two genetic variants that were identified within the 5'UTR of the *FDFT1* gene, one of which has been previously described as a SNP (rs11549153) and was present in both the patient and control groups and one which was a novel variant found only in the patient group. The SNP identified in both the patient and control group required no further analysis however the novel variant required a more thorough analysis. Because this variant is only observed in three of the five KWE affected individuals it cannot be the pathogenic KWE causing mutation. However, there may be a possibility that the presence of the variant may increase an individual's

susceptibility to develop KWE. With genetic variants that occur in the 5'UTR of genes, the first question to answer is, "Does this variant affect an uORF in any way?". When analysing the 5'UTR of the *FDFT1* transcript that was observed in both the patient and control group, it was clear that this transcript does not contain an uORF. This variant results in the replacement of cytosine with adenine, creating an upstream AUG codon. Although this uAUG is positioned only 3 nucleotides away from the cAUG, there may be a possibility that during translation the ribosome will attach to the upstream AUG instead of the coding AUG (Kozak, 1999). This will result in the production of a protein that is two amino acids longer. Figure 4.1 illustrates this possibility.

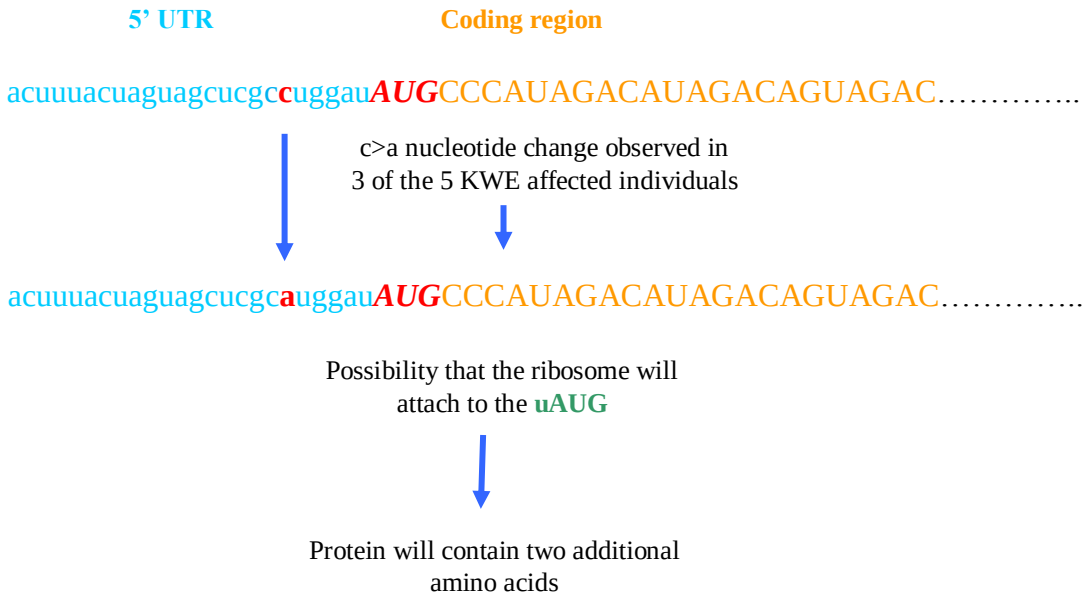


Figure 4.1: An illustration of the possible effect that the single nucleotide base change in the 5' UTR of *FDFT1* may have on the protein

Although Figure 4.1 demonstrates only a possible scenario, it is interesting that this particular variant occurs in the more severely affected KWE patients. Is there a possibility that the variant A allele will result in a slight increase or decrease of the expression of *FDFT1,* which will in turn increase an individual's susceptibility to developing KWE? When analysing this variant in a larger group of unaffected (n=15) individuals, the variant A allele was observed frequently in a heterozygous state (40%). This variant can therefore be classified as a polymorphism that is highly unlikely to play any role in the manifestation of KWE.

The coding variants (rs1047643, rs904011 and rs9205) observed in *FDFT1* are all predicted to be synonymous low risk variants that are unlikely to affect the structure and function of the protein. None of these variants affect the regular splicing of the gene or affect a transcription binding site. Because synonymous variants do not involve an exchange of amino acids they are often regarded as neutral variants that will not have a deleterious effect on the protein. FastSNP predicts that rs1047643 may affect the splicing regulation of the gene by diminishing an ESE motif (Table 3.10). This variant, which occurs in exon 1 of the *FDFT1* gene, segregates with KWE patients and is not found in the control group. In 2006, McVeaty et al. demonstrated that the disruption of an exon splicing enhancer in exon 3 of the *MLH1* gene is the cause of HNPCC (human non polyposis colorectal cancer) in a Quebec family. Could the variant disrupt the ESE motif to an extent that it results in the production of an abnormal or non-functioning *FDFT1* protein that is responsible for the manifestation of the KWE phenotype?

In 2006, Kimchi-Sarfaty et al. demonstrated that silent mutations are not always silent and that some minor genetic variants can profoundly affect normal cell activity if the variant causes the cell to read a different DNA codon. While the same protein is produced, the presence of the rare codon may slow protein folding, resulting in an altered protein conformation. This may affect the function as well as the rate at which the protein is synthesized. Although it is interesting that the L7L variant was not observed in the control group, there was a possibility that this was completely coincidental due to the small sample size used in the present study. In order to provide further information on the L7L variant, a larger group of unaffected (n=15) individuals was sequenced for this variant and it was found that this variant was present in a heterozygous and homozygous form in unaffected individuals. It is therefore highly unlikely that this variant will deleteriously affect the splicing of this gene and it was excluded from being the mutation responsible for the manifestation of KWE. The cDNA of *FDFT1* from two different tissues (blood and saliva) was analysed in order to identify the presence of deep intronic variants such as large insertions and deletions or splicing variants that are not identifiable on a genomic DNA level. No additional variants were identified. The *FDFT1* cDNA transcripts obtained from the different tissues were identical to each other and to the reference sequence. There were no differences in the variants that were identified in the RNA extracted from whole blood and the RNA extracted from saliva. Each variant identified in the coding genomic DNA was observed in the cDNA, in both the saliva and whole blood RNA. It does not appear that a particular variant occurs only in one type of tissue.

When considering the *FDFT1* DNA variation in affected individuals, the following patient KWE-associated haplotype was determined: G-C-C-C-G-C-c-C-c-c-G-t-C-A (Table 3.9). However this haplotype was also present in unaffected individuals. No unique KWE associated haplotype was observed, suggesting that *FDFT1* is unlikely to be the gene responsible for the manifestation of KWE.

The quantitative analysis of *FDFT1* expression in affected and non-affected palmar skin was performed with the hope to determine whether there was differential expression of this gene in KWE patients. The amplification and relative efficiency of the *FDFT1* and reference gene system fell within the range theoretically accepted as optimal (Meijerink et al., 2001). The variance among the set of data collected for each patient and control individual (reported as a %) were all relatively low statistically confirming that there is little variance among each set of data (C_q values) which is ideal. There appears to be a trend in a higher expression of *FDFT1* in the patients in comparison to the controls (Figure 3.18). There appears to be a trend towards an increased expression of FDFT1 cDNA in the patient group. Because of this potentially exciting finding, the promoter region of *FDFT1* was sequenced in order to search for mutations that may impact on the regulation of *FDFT1* expression. The 200bp region upstream to the start of transcription has been well characterised and shown to include a sterol regulatory element (SRE-1) and two potential NF-1 binding sites. A CCAAT box, a SRE-1 element and two Sp1 sites were identified 3' to this region (Guan et al., 1995). However, after sequencing the promoter region and surrounding regions, the variants that were identified in the patients were also observed in the control individuals and did not occur in any essential regulatory elements. Therefore it is unlikely that a

genetic variant present in the currently defined promoter region of this gene is the cause of the increased expression observed in the KWE patients. The question remains, “What is causing the increased expression of *FDFT1* in the KWE patients?” *FDFT1* produces an enzyme which plays an essential role in the pathway responsible for the synthesis of epidermal cholesterol (Proksch et al., 1991). Cholesterol biosynthesis occurs actively in the skin and it is thought that the epidermis is one of the most active tissue sites for cholesterol synthesis. Cholesterol synthesis occurs in both the proliferating basal and differentiating suprabasal cellular layers of the epidermis (Feingold et al., 1983; Proksch et al., 1996). One of the most important functions of epidermal cholesterol is its role in the maintenance of the skin barrier. Elias et al. (2005) reported that a disruption of the epidermal barrier results in an increase in the production of cholesterol via the mevalonate pathway (Figure 1.8). This increase in epidermal cholesterol is due to an increase in the activity, protein levels and mRNA of key enzymes that play a role in this pathway, including *FDFT1*. Elias et al. (2005) describes that a disruption in the balance between the stratum corneum lipids (cholesterol/ceramide/fatty acid ratio) in the epidermis causes barrier disruption and ultimately results in an increase in the production of epidermal cholesterol. In many other studies it was also shown that a disruption of the skin barrier will result in an increase in the epidermal synthesis of cholesterol (Grubauer et al., 1987; Proksch et al., 1996). Skin barrier disruption is also caused by physical, chemical, pathological and environmental factors (Fluhr et al., 2004). Denda et al. (1996) showed that repeated disruptions of the skin barrier induced cutaneous pathology including inflammation and epidermal hyperplasia. The degree of barrier disruption is directly related to the degree of increase observed in epidermal cholesterol synthesis (Menon et al., 1985;

Prosch et al., 1996). In some cutaneous disorders such as psoriasis, contact dermatitis and atopic dermatitis, it has been demonstrated that these disorders are associated with barrier disruption (Fluhr et al., 2004) and the more severe the barrier disruption, the higher the increase in epidermal cholesterol synthesis. Interestingly, it was shown that the disruption of the epidermal skin barrier results in an increase in protein prenylation, a function in which *FDFT1* is a key enzyme. Prenylation, also known as isoprenylation, is the covalent attachment of isoprenyl groups to the carboxyl cysteine residues of proteins (Denda et al., 1996). Isoprenyl groups are intermediates of the cholesterol biosynthesis pathway. Denda et al. (1996) concluded that the prenylation of proteins in the epidermis is stimulated by injury, suggesting that one or more of the prenylated proteins may be important in epidermal proliferation. Therefore, barrier disruption will result in an increase in protein prenylation and cholesterol biosynthesis in the epidermis. The present study shows that there is a trend in a higher level of *FDFT1* expression in KWE patients in comparison to the controls, an observation which could be due to the fact that KWE patients do experience epidermal barrier disruption, which is known to increase mRNA levels of the enzymes within the mevalonate pathway.

One of the main causes of epidermal barrier disruption is environmental humidity (Fluhr et al., 2004). Psoriasis, contact dermatitis and atopic dermatitis phenotypes tend to worsen in the colder months when the humidity is low (Sauer and Hall et al., 1996). Even if the damage to the epidermal skin barrier is small, constant repetition under low environmental humidity could induce epidermal hyperplasia and inflammation (Fluhr et al., 2004). Many studies have demonstrated that severity of skin conditions

change in areas of different environmental humidity. It is commonly reported that the severity of skin disorders worsen in winter (Rycroft and Smith, 1980) when there is a decrease in humidity (Sauer and Hall, 1996). Denda et al. (1998) showed that low humidity stimulates epidermal DNA synthesis and amplifies the hyperproliferation response to barrier disruption. Also, that there is the seasonal aggravation of cutaneous disorders that are attributable to decreased humidity (Denda et al., 1998). In a separate study, it was demonstrated that low humidity increases epidermal DNA synthesis and accelerates stratum corneum desquamation (Sato et al., 1998). With KWE it is known that the phenotype clears up and almost disappears during warmer summer months and at coastal regions where the humidity is higher. Although many studies have proven that low humidity does induce epidermal barrier disruption, the mechanism by which humidity influences cutaneous disorders is poorly understood. Therefore it is possible that KWE patients experience epidermal barrier disruption that will result in a response to increase epidermal cholesterol via the mevalonate pathway, which is why the increased levels of *FDFT1* expression is seen. The finding that the more severely affected individuals tend to have higher levels of *FDFT1* expression compared to the less severely affected individuals supports the statement made by Proksch et al. that the level of increase in total epidermal cholesterol synthesis correlates directly with the extent of disturbance in barrier function (Proksch et al., 1990).

At this point it appears that *FDFT1* is not the KWE causing gene and there is a good chance that the higher level of *FDFT1* expression observed in KWE patients is due to barrier disruption which leads to upregulation of *FDFT1* expression. The promoter

region, entire 5'UTR, coding region and intron/exon boundaries of *FDFT1* have been thoroughly analysed through the mutational analysis of genomic DNA. Intronic variants that could cause aberrant splicing were indirectly analysed through the analysis of the FDFT1 cDNA. Although in the literature, *FDFT1* appears to be an excellent candidate gene for KWE, this study has not identified any genetic variations within this gene that may result in the manifestation of the disease. *FDFT1* is not the KWE causing gene.

4.2.3 *DUB-3 and C9orf49 KWE gene candidates*

Three of the six genetic variants identified within *DUB-3* have been previously described as SNPs in many different population groups. The remaining three variants are novel. The majority of the genetic variants identified within *DUB-3* were computationally predicted to be benign (with the exception of K458R) exerting no deleterious or damaging effect on the structure or function of the protein. Each variant was observed frequently (>70%) in the control group, an observation that would not be expected with the KWE causing mutation. For KWE, an autosomal dominant and highly penetrant disease, the pathogenic mutation would be observed in the patient group only. For this reason, the variants identified within *DUB*-3 are excluded from being possibly pathogenic.

The K458R variant is predicted to be a low to medium risk variant that may affect the splicing regulation of the gene by affecting an ESE motif. The transition from a guanine to an adenine nucleotide will result in the loss of an exon enhancer site. As already discussed, the disruption of an ESE motif could result in the production of a non-

functional and abnormal protein. It is unlikely that a substitution at this particular site in the protein will have a deleterious effect on the structure or function of the protein. The fact that this variant A allele is present in each of the control individuals strengthens the idea that although this particular substitution does affect an exon enhancer site, the result will not be KWE. Further analysis into the disruption of the ESE motif is not necessary for this project. A variant, which is present in only KWE affected individuals, and that affects an ESE motif would be of interest and would require further investigation.

In the present study, *DUB-3* mRNA was not analysed. This gene is an excellent candidate unfortunately certain challenges make it difficult to perform a thorough analyse. Firstly, *DUB-3* lacks an established 5'UTR. This makes it difficult to design primers that will cover the entire mRNA sequence. The presence of highly homologous (>98% homology) sequences that surround *DUB-3* create a second challenge. These highly homologous sequences make it difficult to design primers that are specific for the *DUB-3* sequence, which make downstream applications such as reverse transcriptase-PCR and real-time PCR challenging. Once these problems are overcome, it would be essential to continue a thorough analysis of this potential candidate gene.

The C8orf49 protein product and its function are not known, but it was considered as an important positional candidate gene because of the presence of two upstream open reading frames in its 5'UTR and because it was one of the transcripts that had not been analysed in previous studies. The genetic variants identified within *C8orf49*, with the exception of C194X, were predicted to be low to medium risk polymorphisms that

do not affect the transcription start site, do not occur in the active domain and do not affect ESEs. From predictive data one can conclude that the variants identified will not have a deleterious effect on the structure or function of the putative protein. Each variant was observed in the unaffected individuals which strengthens the proposal that these variants are benign in respect to the manifestation of KWE. The C194X (rs809203) variant is predicted to be a very high risk nonsense mutation. The replacement of a thymine nucleotide with an adenine, introduces a premature stop codon into the cDNA sequence, resulting in the production of a truncated protein. The majority of nonsense mutations are damaging to a protein and are known to cause serious phenotypic effects in patients. However, this variant has been found to be polymorphic in many different population groups and does not result in a phenotypic manifestation in the individuals who carry this variant. How important is this putative protein in the human body if a truncated form of the protein has no known effect on the individuals who have this variant? Because this variant is observed in the control individuals (40%) and only two of the five KWE affected individuals, no further research and analysis is necessary for the present study.

4.3 CONCLUSIONS

It is highly unlikely that *CTSB*, *FDFT1*, *DUB-3* and *C8orf49* play a role in the manifestation of KWE. The higher level of *FDFT1* expression observed in the patients could be due to the epidermal barrier disruption associated with the KWE phenotype. Genes located within the KWE critical region that have to date not been analysed include *NEIL2* and *LONRF1* as well as the many defensin genes that make up the cluster

located in the KWE critical region. *NEIL2* belongs to a group of DNA glycosylases which initiate the first step in base excision repair. This is done by cleaving bases that have been damaged by reactive oxygen species and introducing a DNA strand break via the associated lyase reaction (Bandaru et al., 2002). *LONRF1* is known as LON peptidase N-terminal domain and ring finger 1 and was only recently mapped to the KWE critical region. There is very little information available regarding this protein-coding gene. There is a group of defensins located within the KWE critical region. Defensins are a family of small cationic, antibiotic peptides that contain six cysteines in disulfide linkage (Ganz and Lehrer, 1995) and are found in both vertebrates and invertebrates. These peptides contribute to host defense against fungi and bacteria and may participate in tissue inflammation and endocrine regulation during infection. (Ganz and Lehrer, 1995). Defensin genes have been implicated in a number of diseases such as Crohn's disease (Wehkamp et al., 2005), are thought to play a role in the development of acne (Philpott, 2003) and have been found to be expressed in human skin during inflammatory conditions such as psoriasis (Harder et al., 1999) and wound healing (Sorensen et al., 2006). These genes that are positioned within the KWE critical region are positional candidate genes and need to be thoroughly analysed to determine whether they could be the cause of the manifestation of KWE. This study has contributed significantly to the search for the KWE causing gene since this excluded four highly compelling positional candidates, now focusing attention on the remaining positional candidate genes for future studies. The quest for the KWE causing gene continues.

4.4 LIMITATIONS AND FUTURE STUDIES

Limitations of the study

Only two of the families that were genotyped in the haplotype analysis showed informative cross-overs which aided in the validation of the KWE critical region. The other two families were uninformative. Additional polymorphic markers could have been typed to refine the candidate region further. Expression studies were only performed for two of the four candidate genes. It is an expensive and time consuming procedure that requires extensive optimization. *DUB-3* is a good candidate for expression studies but because of the high homology of this gene with flanking sequences it would be difficult to design primers specific for this gene.

Future studies

In future studies, the KWE critical region needs to be examined thoroughly. There are two positional candidate genes located within the KWE critical region that need to be examined in future studies, namely *LONRF1* and *NEIL2* as well as the defensin genes. The highly conserved regions of the KWE critical region should also be examined. This could identify potential non-coding transcripts that may cause the KWE phenotype. Using the new generation high throughput sequencing technology, the entire KWE critical region could be sequenced in several KWE patients and controls. Bioinformatic analyses could be done in order to identify miRNA genes and other potential RNAi features in the critical region. Lastly, global expression analysis could be done using microarray technology to examine expression patterns in affected and unaffected skin. This could result in the identification of affected pathways and many possible candidate genes.

REFERENCES

Aftabuddin MD, Kundu S (2007) Hydrophobic, Hydrophilic, and Charged Amino Acid Networks within Protein. Biophys J 93:225-231

Alberts B, Bray D (2003) Essential Cell Biology: An introduction to the Molecular Biology of the Cell. Garland Science, USA pp 481-513

Altemus M, Rao B, Dhabhar FS, Ding W, Granstein RD (2001) Stress-induced changes in skin barrier function in healthy women. J Invest Dermatol 117:309-17

Appel S, Filter M, Reis A, Hennies HC, Bergheim A, Ogilvie E, Arndt S, Simmons A, Lovett M, Hide W, Ramsay M, Reichwald K, Zimmermann W and Rosenthal A (2002) Physical and transcriptional map of the critical region for keratolytic winter erythema (KWE) on chromosome 8p22-p23 between D8S550 and D8S1759. Eur J Hum Genet 10:17-25

Bandaru V, Sunkara S, Wallace SS, Bond JP (2003) A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. DNA Repair 1:517–29

Barber JC, Maloney VK, Huang S, Bunyan DJ, Cresswell L, Kinning E, Benson A, Cheetham T, Wyllie J, Lynch SA (2008) 8p23.1 duplication syndrome; a novel genomic condition with unexpected complexity revealed by array CGH. Eur J Hum Genet 16:18-27

Berquin IM, Cao L, Fang D, Sloane BF (1995) Identification of the two new exons and multiple transcription start points in the 5'-untranslated region of the human cathepsin-B encoding gene. Gene 159:143-149

Blencowe BJ (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic disease. TIBS 25: 106–110

Bouwstra JA, Gooris GS, Dubbelaar FE, Ponec M (1999) Cholesterol sulfate and calcium affect stratum corneum lipid organization over a wide temperature range. J Lipid Res 40:2303-2312

Burchard EG, Silverman EK, Rosenwasser LJ, Borish L, Yandava C, Pillari A, Weiss ST, Hasday J, Lilly CM, Ford JG, Drazen JM ( ) Association between a sequence variant in the IL-4 gene promoter and FEV(1) in asthma. AJRCCM 160:919–22

Burrows JF, McGrattan MJ, Rascle A, Humbert M, Baek KH, Johnston JA (2004) DUB-3, a cytokine-inducible deubiquitinating enzyme that blocks proliferation. J Biol Chem 279:13993-14000

Buth H, Wolters B, Hartwig B, Meier-Bornheim R, Veith H, Hansen M (2004) HaCaT keratinocytes secrete lysosomal cysteine proteinases during migration. Eur J Cell Biol 83:781-795.

Byrant RA, Nix DP (2004) Acute and Chronic wounds. Elsevier Health Sciences, United Kingdom. Pages 32-45

Cagli NA, Hakki SS, Dursun R (2005) Clinical, genetic and biochemical findings in two siblings with Papillon-Lefevre Syndrome J Periodontol 76:2322-2329

Chan LS, Robinson N, Xu L (2001) Expression of interleukin-4 in the epidermis of transgenic mice results in a pruritic inflammatory skin disease: an experimental animal model to study atopic dermatitis. J Invest Dermatol 117:977-983

Chao SC, Yang T, Lee J (2003) A novel mutation in the keratin 4 gene causing white sponge naevus. Br J of Dermatol 148:1125-1128

Chen J, Lu J, Liao Z, Zuo C, Huang J (2006) Apoptosis of psoriatic keratinocytes and the severity of patients condition. Med Sci 31:936-940

Cheng T, Hitomi K, van Vlijmen-Willems IM, de Jongh GJ, Yamamoto K, Nishi K, et al. (2006) Cystatin M/E is a high affinity inhibitor of cathepsin V and cathepsin L by a reactive site that is distinct from the legumain-binding site. A novel clue for the role of cystatin M/E in epidermal cornification. J Biol Chem 281:15893-15899

Chou H, Lai HY, Tam MF, Chou MY, Wang SR, Han SH, et al. (2002) cDNA cloning, biological and immunological characterization of the alkaline serine protease major allergen from Penicillium chrysogenum. Int Arch Allergy Immunol 127:15-26.

Ciechanover A, Schwartz AL (1996) The ubiquitin-proteasome pathway: The complexity and myriad functions of proteins death. Annu Rev Biochem 65:801–847

Ciskos S, Bukovska A, Koppel J (2007) Relative quantification of mRNA:comparison of methods currently used for real-time PCR data analysis BMC Mol Biol 8:113-127

Cork MJ, Robinson DA, Vasilopoulos Y, Ferguson A, Moustafa MM, MacGowan A, Duff GW, Ward SJ, Tazi-Ahnini R (2006) New perspectives on epidermal barrier dysfunction in atopic dermatitis: Gene-environmental interactions. J Allergy Clin Immunol 118:3-21

Crowe ML, Wang XQ, Rothnagel JA (2006) Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. BMC Genomics 7:16

Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SM, Clamp M (2004) The Ensembl automatic gene annotation system. Genome Res 14:942-950

Darby N, Creighton TE (1995) Disulfide bonds in protein folding and stability *Methods Mol Biol* **40:** 219–252

Decker JM (2000) Introduction to immunology. Blackwell Science LTD, United States pp 89

DeSalle LM and Pagano M (2001) Regulation of the G1 to S transition by the ubiquitin pathway. Elsevier 490:179-189

Demerjian M, Crumrine DA, Milstone LM, Williams ML, Elias PM (2006) Barrier dysfunction and pathogenesis of neutral lipid storage disease with ichthyosis (Chanarin-Dorfman syndrome). J Invest Dermatol 126:2032-2038

Denda M, Brown BE, Elias PM, Feingold KR (1997) Epidermal injury stimulates prenylation in the epidermis of hairless mice. Arch Dermatol Res 289:114-110

Denda M, Sato J, Masuda Y, Tsuchiya T, Kuramoto M, Elias PM (1998) Exposure to a dry environment enhances epidermal permeability barrier function. *J Invest Dermatol* 111:858–863

Deussing J, Roth W, Sattig P, Peters C. Ploegh HL, Villadangos JA (1998) Cathepsins B and D are dispensible for major histocompatibility complex class II mediated antigen presentation. Proc Natl Acad Sci 95:4516-4521

Devriendt K, Matthijs G, Van Dael R, Gewillig M, Eyskens B, Hjalgrim H (1999) Delineation of the critical deletion region for congenital heart defects, on chromosome 8p23.1. Am J Hum Genet 64:1119-1126

Dill KA (1990) Dominant forces in protein folding. Biochem 29:7133–7155

Dobson R (2006) Predicting deleterious nsSNPs: an analysis of sequence and structural attributes. BMC Bioinf 7:217

Dykes PJ, Marks R, Davies MG, Reynolds DJ (1987) Epidermal metabolism in heredopathia atactica polyneuritiformis (Refsum's disease). J Invest Dermatol 70:126-129

Eckert RL, Rorke EA (1989) Molecular biology of keratinocyte differentiation. Environ Health Perspect 80:109-16

Eichner R, Gendimenico GJ, Kahn M, Mauon JP, Capetcla RJ, Mezick JA (1996) Br J Dermatol 135:687-695

Elias PM (2005) Stratum corneum defensive functions: an integrated view. J Invest Dermatol 125:183-200

Elias PM, Menon GK (1991) Structural and lipid biochemical correlates of the epidermal permeability barrier. Adv Lipid Res 24:1-26

Esch FS, Keim PS, Beattie EC, Blacher RW, Culwell AR, Oktersdorf T, McClure D, Word PJ (1990) Cleavage of amyloid beta peptide during constitutive processing of the precursor. Science 248: 1122-1124

Feingold KR (1991) The regulation and role of epidermal lipid synthesis. Adv Lipid Res 24:57-82

Feliciani C, Gupta AK, Sauder DN (1996) Keratinocytes and cytokine/growth factors. Crit Rev Oral Biol Med 7:300-318

Ferreira ID, Rosario VE, Cravo PVL (2006) Real-time quantitative PCR with SYBR Green I detection for estimating copy numbers of nine drug resistance candidate genes in *Plasmodium falciparum.* Malaria Journal 5:1-16

Findlay GH, Nurse GT, Heyl T, Hull PR, Jenkins T, Klevansky H et al. (1977) Keratolytic winter erythema or 'oudtshoorn skin': a newly recognized inherited dermatosis prevalent in South Africa. S Afr Med J 52:871-874

Floridia G, Piantanida M, Minelli A, Dellavecchia C, Bonaglia C, Rossi E (1996) The same molecular mechanism at the maternal meiosis I produces mono- and dicentric 8p duplications. Am J Hum Genet 58:785-796

Fluhr JW, Mao-Qiang M, Brown BE, Wertz PW, Crumrine D, Sundberg JP, Feingold KR and PM Elias (2003) Glycerol Regulates Stratum Corneum Hydration in Sebaceous Gland Deficient (Asebia) Mice. J Invest Dermatol 120:728-737

Foghsgaard L, Wissing D, Mauch D, Lademann U, Bastholm L, Boes M, Elling F, Leist M, Jaattela M (2001) Cathepsin B acts as a Dominant Execution Protease in Tumor Cell Apoptosis Induced by Tumor Necrosis factor. J Cell Biol 153:999-1010

Fradette J, Larouche D, Fugere C, Guignard R, Beauparlant A, Couture V, Caouette-Laberge, Roy A, Germain L (2003) Normal Human Merkel cells are present in epidermal cell populations isolated and cultured from glabrous and hairy skin sites. J Invest Dermatol 120:313-317

Freedberg (2003) Fitzpatricks dermatology in general medicine 6th ed. McGraw-hill, New York pp 312-316

Freinkel RK, Woodley D (2001) The biology of the skin. Patthenon publishing group, UK pp19-26

Fulmer AW, Kramer GJ (1986) Stratum corneum lipid abnormalities in surfactant-induced dry scaly skin. J Invest Dermatol 86:598-602.

Gallucci RM, Sloan DK, Heck JM, Murray AR, O'Dell SJ (2004) Interleukin 6 indirectly induces keratinocyte migration. J Invest Dermatol 122:764-772

Ganz T, Lehrer RI (1995) Defensins. *Pharmacol Ther 66:*191–205

Ghilardi N, Wiestner A, Skoda RC (1998) Thrombopoietin production is inhibited by a translational mechanism. Blood 92:4023-4030.

Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 68:874-883

Giglio S, Calvari V, Gregato G, Gimelli G, Camanini S, Giorda R, et al. (2002) Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. Am J Hum Genet 71:276-285

Goldsmith LA (1991) Biochemistry and Physiology of the skin, 2nd ed. Oxford University Press, New York pp 64-101

Grossman RM, Krueger J, Yourish D, Granelli-Piperno A, Murphy DP, May LT, et al (1989) Interleukin 6 is expressed in high levels in psoriatic skin and stimulates proliferation of cultured human keratinocytes. Proc Natl Acad Sci 86:6367-6371

Grubauer G, Feingold KR, Elias PM (1987) Relationship of epidermal lipogenesis to cutaneous barrier function. J Lipid Res 28:746-752

Hanley K, Komuves LG, Ng DC, Schoonjans K, He SS, Lau P (2000) Farnesol stimulates differentiation in epidermal keratinocytes via PPARalpha. J Biol Chem 275:11484-11491

Haratake A, Komiya A, Horikoshi T, Uchiwa H, Watanabe S (2006) Acceleration of de novo cholesterol synthesis in the epidermis influences desquamation of the

stratum corneum in aged mice. Skin Pharmacol Physiol 19:275-282

Harder J, Bartels J, Christophers E, Schroder JM (1999) A peptide antibiotic from human skin. Nature 387 6636: 861

Harris IR, Farrell AM, Grunfeld C, Holleran WM, Elias PM, Feingold KR (1997) Permeability barrier disruption coordinately regulates mRNA levels for key enzymes of cholesterol, fatty acid, and ceramide synthesis in the epidermis. J Invest Dermatol 109:783-787

Harris IR, Hoppner H, Siefken W, Farrell AM, Wittern KP (2000) Regulation of HMG-CoA synthase and HMG-CoA reductase by insulin and epidermal growth factor in HaCaT keratinocytes. J Invest Dermatol 114:83-87

Hatsell SJ and Kelsell DP (2000) The diffuse palmoplantar keratodermas. ACTA 9:12-15

Hazra TK, Kow YW, Hatahet Z, Imhoff B, Boldogh I, Mokkapati SK, Mitra S, Izumi T (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. J Mol Biol 34:30417-30420

Hoath SB, Leahy DG (2003) The organization of human epidermis: functional epidermal units and phi proportionality. J Invest Dermatol 121:1440-6.

Horikoshi T, Arany I, Rajaraman S (1999) Isoforms of cathepsin D and human epidermal differentiation. Br J Dermatol 141:453-459

Hovnanian A, Pollack E, Hilal L, Rochat A, Prost C, Barrandon Y, et al. (1993) A missense mutation in the rod domain of keratin 14 associated with recessive epidermolysis bullosa simplex. Nat Genet 3:327-332

Hull PR (1986) Keratolytic winter erythema (Oudtshoorn disease): clinical, gentic and ultrastructural aspects. PhD thesis, University of the Witwatersrand, Johannesburg

Iacono M, Mignone F, Pesole G (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene 349:97-105

Itin PH and Fistarol SK (2005) Palmoplantar keratodermas. Clin Dermatol 23:15-22

Ivanov IP, Loughran G, Atkins JF (2008) uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. Proc Natl Acad Sci 105:10079-10084

Jackson SM, Williams ML, Feingold KR, Elias PM (1993) Pathobiology of the stratum corneum. West J Med 158:279-85

Jans R, Atanasova G, Jadot M, Poumay Y (2004) Cholesterol depletion upregulates involucrin expression in epidermal keratinocytes through activation of p38. J Invest Dermatol 123:564-573

Jebbink J, Bai X, Rogers BB, Dawson BD, Scheuermann RH and Domiati-Saad R (2003) Development of Real-Time PCR Assays for the Quantitative Detection of Epstein-Barr Virus and Cytomegalovirus, Comparison of TaqMan Probes, and Molecular Beacons. J Med Diagnos 5:1

Jiang G, McKenzie TL, Conrad DG, Shechter I (1993) Transcriptional regulation by lovastatin and 25-hydroxycholesterol in HepG2 cells and molecular cloning and expression of the cDNA for the human hepatic squalene synthase. J Biol Chem 268:818-824

Jin X, Turcott E, Englehardt S, Mize GJ, Morris DR (2003) The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties. J Biol Chem 278:25716-25721

Kelsell DP, Stevens HP (1999) The palmoplantar keratodermas:much more than palms and soles. Arch Dermatol 14:107-113

Kelsell DP, Stevens HP, Purkis PE, Talas U, Rustin MH, Leigh IM (1999) Fine genetic mapping of diffuse non-epidermolytic palmoplantar keratoderma to chromosome 12q11-q13: exclusion of the mapped type II keratins. Exp Dermatol 8:388-391

Khavari PA (2006) Genetic correlation of inherited epidermal disorders. Human Journal of gene therapy 16:2277-2282

Kimchi-Sarfaty C, Oh JM, Kim I-W, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM (2007) A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. Science 315: 525–528

Kimonis V, DiGiovanna JJ, Yang JM, Doyle SZ, Bale SJ, Compton JG (1994) A mutation in the V1 end domain of keratin 1 in non-epidermolytic palmar-plantar keratoderma. J Invest Dermatol 103:764-769

Kimyai-Asadi A, Katcher LB, Jih MH (2002) The molecular basis of hereditary palmoplantar keratodermas. J Am Acad Dermatol 47:327-343

Klein DM, Felsenstein KM, Brenneman DE (2009) Cathepsin B and L differentially regulate Amyloid Precursor Protein processing. J Pharmacol Exp Ther 328:813-281

Koch AE, Kronfeld-Harrington LB, Szekanecz Z (1993) In situ expression of cytokines and cellular adhesion molecules in the skin of patients with systemic sclerosis. Their role in early and late disease. Pathobiology 61:239-246

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. Gene 234:187-208

Koppe JG, Marinkovic-ilsen A, Rijken Y, deGroot WP, Jobsis AC (1978) X-linked ichthyosis: a sulfatase deficiency. Arch Dis Child 53:803-806

Kuechle MK, Presland RB, Lewis SP, Fleckman P, Dale BA (2000) Inducible expression of filaggrin increases keratinocyte susceptibility to apoptotic cell death. Cell Death Differ 7:566-73

Kulozik AE, Bellan-Koch A, Bail S, Kohne E, Kleihauer E ( ) Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. Blood 77:2054–2058

Kuster W, Reis A, Hennies HC (2002) Epidermolytic palmoplantar keratoderma of Vorner: re-evaluation of Vorner's original family and identification of a novel keratin 9 mutation. Arch Dermatol 294:268-272.

Kwak J, Maverakis E (2006) Dermatology Online Journal 12:6

Lane EB, McLean WHI (2004) Keratins and skin disorders. J pathology 355-366

Layman DP (2003) Biology Demystified. McGraw-Hill publishing, United Kingdom pp 217-220

Liao H, Yeh W, Chiang D, Jernigan RL, Lustig B (2005) Protein sequence entropy is closely related to packing density and hydrophobicity. Protein Eng Des Sel 18:59-64

Lind L, Lundstrom A, Hofer PA, Holmgren G (1994) The gene for diffuse palmoplantar keratoderma of the type found in northern Sweden is localized to chromosome 12q11-q13. Hum Mol Genet 3:1789-1793.

Mader S (2003) Biology 8th ed. McGraw-Hill publishing, New York pp 321-323; 456-457

Makhurkar S, Idris MM, Reddy DN, Bhaskar S, Rao GV, Thomas V, Singh L, Chandak GR (2006) Association of cathepsin B gene polymorphisms with tropical calcific pancreatitis. Gut 55:1270-1275

Mason JC (2003) Statins and their role in vascular protection. Clin Sci 105:251-266

Marks JG, Miller J (2006) Lookingbill and Marks Principles of dermatology 4th ed. Saunders Elsevier, New York, pp 12-13

McAllister C, Karymov MA, Kawano Y, Lushnikov AY, Mikheikin A, Uversky VN, Lyubchenko YL (2005) Protein Interactions and Misfolding Analyzed by AFM Force Spectroscopy. *J Mol Biol* 354:1028–1042

McGrath JA, Eady RAJ, Pope FM (2004) Rooks Textbook of Dermatology 7th ed. Blackwell publishing, United Kingdom pp 102-113

McGrath ME, Palmer JT, Bromme D, Somoza JR (1998) Crystal structure of human cathepsin S. Protein Sci 6:1294-1302

McLean WHI, Irvine AD (2007) Disorders of keratinisation: from rare to common diseases of skin and other epithelial tissues. Ulster Med J 76:72-82

McLean WHI, Morley SM, Lane EB, Eady RAJ, Griffiths WAD, Paige DG, Harper JI, Higgins C, Leigh IM (1994) Ichthyosis Bullosa of Siemens – A disease involving Keratin 2e. J of Invest Dermatol 103:277-281

McVety S, Li L, Gordon PH, G Chong G and Foulkes WD (2005) Disruption of an exon splicing enhancer in exon 3 of *MLH1* is the cause of HNPCC in a Quebec family. L Med Genet 43:153-156

Meijer HA, Thomas AA (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. Biochem J 367:1-11

Meijerink J, Mandigers C, van de Locht L, Tonnissen E, Coodsaid F and Raemaekers J (2001) A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. J Mol Diagn 3:55-61

Menon GK, Feingold KR, Moser AH, Brown BE, Elias PM (1985) De novo sterologenesis in the skin. II. Regulation by cutaneous barrier requirements. J Lipid Res 26:418-427

Menon GK, Lee S, Elais PM, Feingold KR (1992) Localization of calcium in murine epidermis following disruption and repair of the permeability barrier. Cell Tissue Res 270:503-512

Merad M, Manz MG, Karsunky H, Wagers A, Peters W, Charo I, Weissman IL, Cyste JG, Engleman EG (2003) Langerhans cells renew in the skin throughout life under steady-state conditions. Nat immunol 3:1135-1141

Merad M, Ginhoux F, Collin M (2008) Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. Nat Rev Immunol 8:935-947

Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 20:8635-8642

Mullis KB, Ferre F, Gibbs RA (1987) The polymerase chain reaction. Birkhauser, pp 375-376

Ng PC, Henikoff (2003) Predicting Deleteriuos Amino acid sunstitutions. Genome Res 11:863-874

Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812-3814

Nishimura F, Naruishi H, Naruishi K, Yamada T, Saski J, Peters C, Uchiyama Y, Murayama Y (2002) Cathepsin-L, a key molecule in the pathogenesis of drug-induced and I-cell disease-mediated gingival overgrowth: a study with cathepsin-L deficient mice. Am J Pathol 161:2047-2052

Norlen L (2001) Skin barrier structure and function: the single gel phase model. J Invest Dermatol 117:830-836

Ogino S (2001) A method to compensate for different amplification efficiencies. J Mol Diagn 3:189-190

Oyama N, Iwatsuki K, Homma Y, Kaneko F (1999) Induction of transcription factor AP-2 by inflammatory cytokines in human keratinocytes. J Invest Dermatol 113:600-666

Pandit J, Danley DE, Schulte GK, Mazzalupo S, Pauly TA, Hayward CM, et al. (2000) Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. J Biol Chem 275:30610-30617

Paquet P, Pierard GE (1996) Interleukin-6 and the skin. Int Arch Allergy Immunol 109:308-317

Paulson DS (1999) Topical antimicrobial testing and evaluation. Taylor and Francis LTD, United States pp 16-22

Philpott MP (2003) Defensins and acne. Mol Immunol 40:457-462

Pickart CM (2001) Mechanisms underlying ubiquitination. Annu Rev Biochem 70:503-533

Pilgram GS, Vissers DC, van der Meulen H, Pavel S, Lavrijsen SP, Bouwstra JA, et al. (2001) Aberrant lipid organization in stratum corneum of patients with atopic dermatitis and lamellar ichthyosis. J Invest Dermatol 117:710-7

Presland RB, Kuechle MK, Lewis SP, Fleckman P, Dale BA (2001) Regulated expression of human filaggrin in keratinocytes results in cytoskeletal disruption, loss of cell-cell adhesion, and cell cycle arrest. Exp Cell Res 270:199-213

Privalov PL, Gill SJ and Murphy KP (1990) Science 260: 297-298

Proksch E, Brasch J, Sterry W (1996) Integrity of the permeability barrier regulates epidermal Langerhans cell density. Br J Dermatol 134:630-638.

Radoja N, Gazel A, Banno T, Yano S, Blumenberg M (2006) Transcriptional profiling of epidermal differentiation. Physiol Genomics 27:65-78

Ragone R, Colonna G (1994) Enthalpy-Entropy balance and convergence temperatures in protein unfolding. J Biol Chem 269:4047-4049

Ramaekers FC, Bosman FT (2004) The cytoskeleton and disease. J Pathol 204:351-354

Reefman E, Limburg PC, Kallenberg CGM, Bijl M (2005) Apoptosis in human skin: idea of pathogenesis of various diseases tolerance for therapy. NY Acad Sci 1051:52-63

Reichelt J, Doering T, Schnetz E, Fartasch M, Sandhoff K, Magin AM (1999) Normal ultrastructure, but altered stratum corneum lipid and protein composition in a mouse model for epidermolytic hyperkeratosis. J Invest Dermatol 113:329-334

Reinheckel T, Hagemann S, Dollwet-Mack S, Martinez E, Lohmuller T, Zlatkovic G, et al. (2005) The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. J Cell Sci 118: 3387-95

Rose GD (2003) Hydrogen bonding, hydrophobicity, packing and protein folding. Annu Rev Biophys 22:381-415

Rossi A, Catani MV, Candi E, Bernassela F, Puddu P, Melino G (2000) Nitric oxide Inhibits Cornified Envelope Formation in Human Keratinocytes by inactivating transglutaminases and activating protein I. J Invest Dermatol 115:731-739

Rothenberg ME (1998) Eosinophilia. NEJM 338:1592-1600

Russell SJ, Blandl T, Skelton NJ (2003) Stability of cyclic β-hairpins: Asymmetric contributions from side chains of a hydrogen-bonded cross-strand residue pair. J Am Chem Soc 125:388–395

Rycroft R and Smith WDL (1980) Low humidity occupational dermatoses. *Contact Dermatitis* 6:488–492

Saitou M, Sugai S, Tanaka T, Shimouchi K, Fuchs E, Narumiya S, et al. (1995) Inhibition of skin development by targeted expression of a dominant-negative retinoic acid receptor. Nature 1995 374:159-162

Sauer GC and Hall JC (1996) Seasonal skin diseases In: Sauer GC, Hall JC, eds. *Manual of Skin Diseases*. Lippincott-Raven,Philadelphia pp. 23–28

Saunders CT, Baker D (2002) Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Biol 322:891–901

Sato J, Denda M, Ashida Y and Koyama J (1998) Loss of water from the stratum corneum induces epidermal DNA synthesis in hairless mice. Arch Dermatol Res 290:634-637

Schechter I, Conrad DG, Hart I, Berger RC, McKenzie TL, Bleskan J, Patterson D (1994) Localization of the squalene synthase gene (*FDFT1*) to human chromosome 8p22-p23.1. Genomics 20:116-118

Scheffner M, Nuber U, Huibregtse JM (1995) Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature 373:81-83

Scheibner A, Hollis DE, McCarthy WH, Milton GW (1986) Effects of sunlight exposure on Langerhans cells and melanocytes in human epidermis. Photodermatol 3:15-25

Schiffer D (2006) The Ubiquitin-Proteasome System. Springer, Netherlands pp 183-187

Schmidt C, Penneys NS, Ziboh VA, Kiem I, Schlossberg J (1977) Cholestrol and cholesteryl ester content in normal and pathologic scale. J Invest Dermatol 68:206-209

Schmuth M, Yosipovitch G, Williams ML, Weber F, Hintner H, Ortiz-Urda S (2001)

Pathogenesis of the permeability barrier abnormality in epidermolytic hyperkeratosis. J Invest Dermatol 4:837-847

Scott IR, Richards S, Harding C, Liddell JE, Curtis CG (1988) Does catabolism of stratum corneum proteins yield functionally active molecules? Ann N Y Acad Sci 548:125-136.

Segre JA (2006) Epidermal barrier formation and recovery in skin disorders. J Clin Invest 5:1150-1158

Sehgal PB (1990) Interleukin-6: Molecular pathophysiology. J Invest Dermatol 94:2S-6S

Selvaraju V, Markandaya M, Venkata P, Prasad S, Sathyan P, Sethuraman G, Srivastava SC, Thakker N, Kumar A (2003) Mutation analysis of the cathepsin C gene in Indian families with Papillon-Lefevre syndrome. BMC Med Gen 4:1-8

Simon M, Haftek M, Sebbag M, Montezin M, Girbal-Neuhauser E, Schmitt D, et al. (1996) Evidence that filaggrin is a component of cornified cell envelopes in human plantar epidermis. Biochem J 317:173-177

Smith WP, Christensen MS, Nacht S, Gans EH (1982) Effect of lipids on the aggregation and permeability of human stratum corneum. J Invest Dermatol 78:7-11

Soriano A, Martinez C, Garcia F, Plana M, Palou E, Aroslegui J, Leujeune M, Joseph J, Miro J, de Romero J, Ridriguez C, Barrasa A, Lorenzo JI, Lozano F, Gatell JM, Gallart T (2002) A novel genetic variant of CD28 is associated with susceptibility to HIV infection. Retrovir Oppor Infect Abstract 5

Stachowitz S, Alessandrini F, Abeck D, Ring J, Behrendt H (2002) Permeability Barrier Disruption Increases the level of serine palmitoyltransferase in human epidermis. J Invest Dermatol 119:1048-1052

Starfield M, Hennies HC, Jung M, Jenkins T, Wienker T, Hull P (1997) Localization of the gene causing keratolytic winter erythema to chromosome 8p22-p23, and evidence for a founder effect in South African Afrikaans-speakers. Am J Hum Genet 61:370-378

Stitziel NO, Binkowski TA, Tseng YY, Kasif S, Liang J (2004) topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association. Nucleic Acids Res 1:32

Sugarman JL, Fluhr JW, Fowler AJ, Bruckner T, Diepgen TL, Williams ML (2003) The objective severity assessment of atopic dermatitis score: an objective measure using permeability barrier function and stratum corneum hydration with computer-assisted estimates for extent of disease. Arch Dermatol 139:1417-1422

Sunyaev S, Ramensky V, Koch I, Lathe W, Kondrashov AS, Bork P (2001) Prediction of deleterious human alleles. Hum Mol Genet 10:591–597

Sybert VP, Dale BA, Holbrook KA (1988) Palmar-plantar keratoderma. A clinical, ultrastructural, and biochemical study. J Am Acad Dermatol 18:75-86

Trovo AB, Goloni-Bertollo EM, Moncini UM, Rahal P, Azevedo WF, Tajara EH (2004) Mutational analysis of the GAP-related domain of the neurofibromatosis type I gene in Brazilian NF2 patients. Gen Mol Biol 27:326-330

Tsuruoka H, Khovidhunkit W, Brown BE, Fluhr JW, Elias PM, Feingold KR (2002) Scavenger receptor class B type I is expressed in cultured keratinocytes and epidermis. J Biol Chem 25:2916-2922

Voet D, Voet JG and Pratt CW (2001) Fundementals of Biochemistry. John wiley and Sons, Inc, United States pp 79-81

Wang Z, Moult J (2001) SNPs, protein structure, and disease. Hum Mutat 17:263–270

Wang XP, Schunck M, Kallen KJ, Neumann C, Trautwein C, Rose-John S, et al (2004) The interleukin-6 cytokine system regulates epidermal permeability barrier homeostasis. J Invest Dermatol 123:124-131

Wehkamp J, Schmid M, Fellermann K and Stange EF (2004) Defensin deficiency, intestinal microbes, and the clinical phenotypes of Crohn's disease. J Leuk Biol 77:460-465

Wertz PW, Swartzendruber DC, Madison KC, Downing DT (1987) Composition and Morphology of epidermal cyst lipids. Soc Inves Dermatol 89:419-425

Whittock NV, Ashton GHS, Dopping-Hepenstal PJC, Gratian MJ, Keane FM, Eady RAJ, McGrath JA (1999) Striate Palmoplantar Keratoderma resulting from desmoplakin haploinsufficiency. J Invest Dermatol 113:940-946

Wilkes GL, Brown IA, Wildnauer RH (1973) The biomechanical properties of skin. CRC Crit Rev Bioeng 1:453-495

Wilkinson C (1997) Regulation of ubiquitin-dependant processes by deubiquitinating enzymes. Faseb J 11:1245-1256

William J, Berger D, Elston D (2005) Andrews diseases of the skin: Clinical Dermatology 10th ed. Saunders, United States pp 2-3

Wong ML, Medrano JF (2005) Real-time PCR for mRNA quantification. Biotechniques 39:1-11

Wood LC, Jackson SM, Elias DM, Feingold KR (1992) Cutaneous barrier perturbation stimulates cytokine production in the epidermis of mice. J Clin Invest 90:482-487

Wu PY, Hanlon M, Eddins M, Tsui C, Rogers RS, Jensen JP, et al (2003) A conserved catalytic residue in the ubiquitin-conjugating enzyme family. EMBO J 5241-5250

Yamaguchi Y, Itami S, Tarutani M, Hosokawa K, Miura H, Yoshikawa K (1999) Regulation of keratin 9 in nonpalmoplantar keratinocytes by palmoplantar fibroblasts through epithelial-mesenchymal interactions. J Invest Dermatol 112:483-488

Yoneda K, Steinert PM (1993) Overexpression of human loricrin in transgenic mice produces a normal phenotype. Proc Natl Acad Sci 90:10754-10758

Yue P, Moult J (2006) Identification and Analysis of Deleterious Human SNPs. J Mol Biol 356:1263–1274

Yuspa SH, Kilkenny AE, Steinert PM, Roop DR (1989) Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro. J Cell Biol 109:1207-1217.

Zaslau S (2006) Board Buster Step 1. Lippincott Williams and Wilkins, United States pp 345

Zavodszky M, Chen C-W, Huang J-K, Zolkiewski M, Wen L, Krishnamoorthi R (2000) Disulfide bond effects on protein stability: Designed variants of *Cucurbita maxima* trypsin inhibitor-V. Protein Sci 10:149–160

Zeeuwen PL, Ishida-Yamamoto A, van Vlijmen-Willems IM, Cheng T, Bergers M, Iizuka H, et al (2007) Colocalization of cystatin M/E and cathepsin V in lamellar granules and corneodesmosomes suggests a functional role in epidermal differentiation. J Invest Dermatol 127:120-128

Zettersten E, Man MQ, Sato J, Denda M, Farrell A, Ghadially R, et al. (1998) Recessive x-linked ichthyosis: role of cholesterol-sulfate accumulation in the barrier abnormality. J Invest Dermatol 111:784-90

Zhang BR, Yin XZ, Xia K, Ding MP, Hu ZM, Zheng M, et al. (2004) Mutation analysis of keratin 9 gene in a pedigree with epidermolytic palmoplantar keratoderma. 21:570-573

Electronic references

Applied Biosystems:
http://www.ambion.com/techlib/basics/rtpcr/index.html;
http://www.appliedbiosystems.com/support/tutorials/pdf/rtpcr_vs_tradpcr.pdf Accessed April 2007

Ensembl Genome Browser: http://www.ensembl.org/index.html Accessed June 2007

FastSNP: A functional analysis and Selection Tool for SNP in largescale association study:
http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp Accessed March 2008

Massey University, Institute for Molecular Biosciences
http://www.massey.ac.nz/~wwbioch/Prot/thirds/tertext.htm

National Institutes of Health, National Human Genome Research Institution
http://science.education.nih.gov/supplements/nih1/Genetic/guide/genetic_variation1.htm Accessed July 2009

National Center for Biotechnology Information
http://www.ncbi.nlm.nih.gov/ Accessed July 2008/2009

PolyPhen: http://genetics.bwh.harvard.edu/pph/data/index.html Accessed March 2007

PhysOrg: www.physorg.com/news10776.html

Rush University Medical centre: www.rush.edu/rumc/page-1098987353859.html. Accessed March 2007

School of anatomy and human biology – The University of Western Australia:
http://www.lab.anhb.uwa.edu.au/mb140/CorePages/Integumentary/Integum.htm Accessed March 2007

SNPs3D: http://www.snps3d.org/ Accessed March 2008

University of Maryland Medical centre:
http://www.umm.edu/dermatology-info/anatomy.htm Accessed March 2007

APPENDIX A

South African KWE family pedigrees

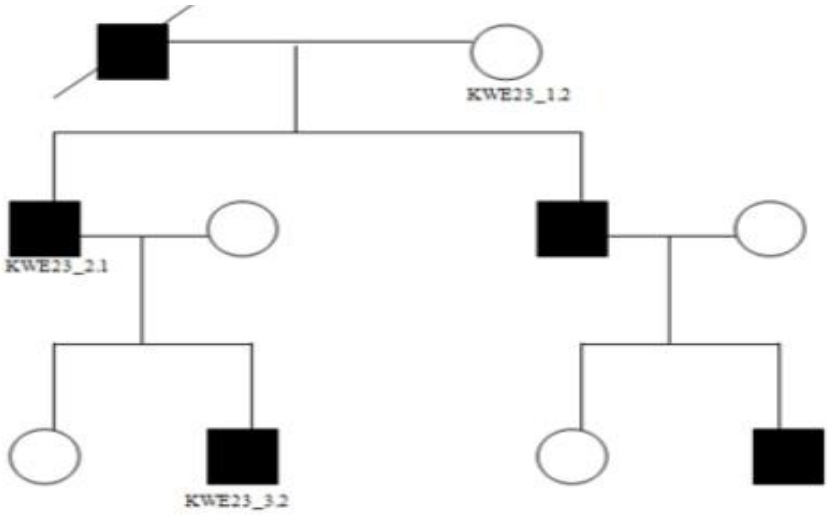


Figure 2.1 A: Family 23 pedigree

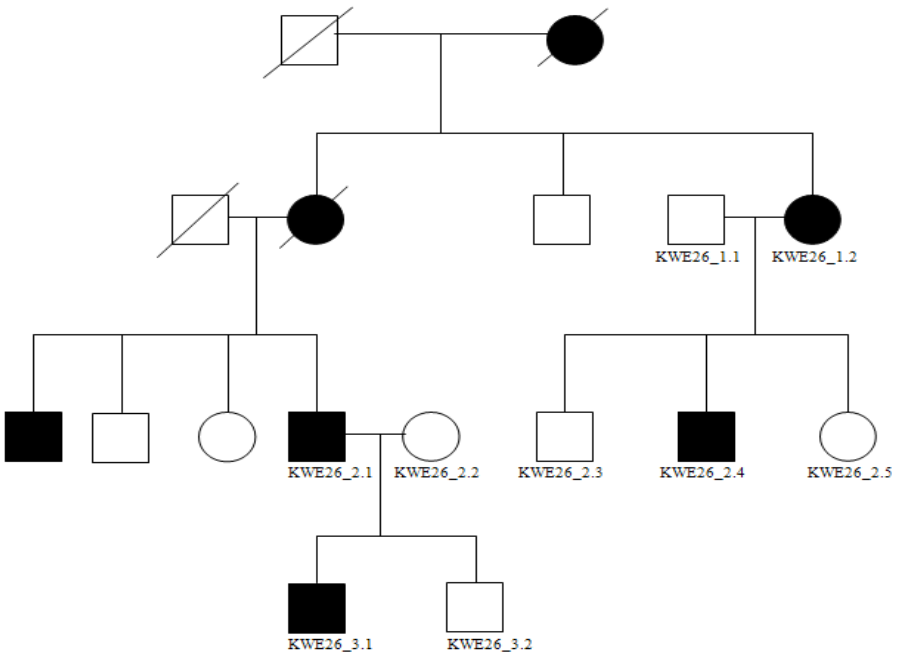


Figure 2.1 B: Family 26 pedigree

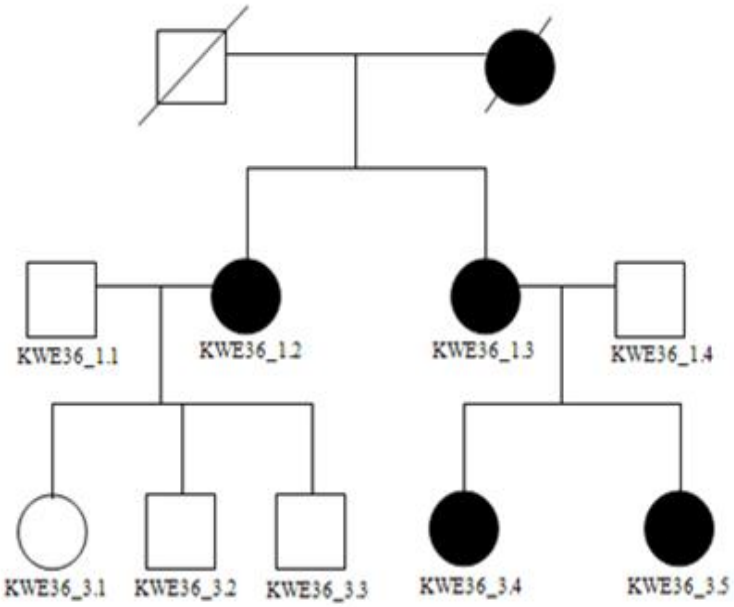


Figure 2.1 C: Family 36 pedigree

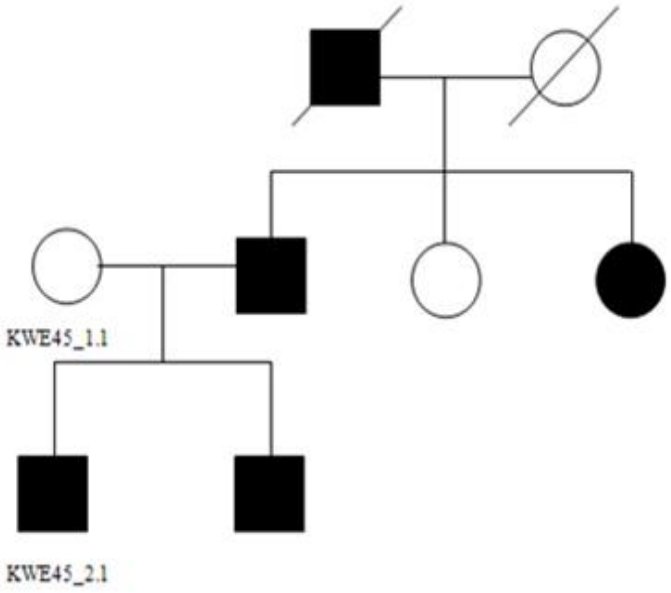


Figure 2.1 D: Family 45 pedigree

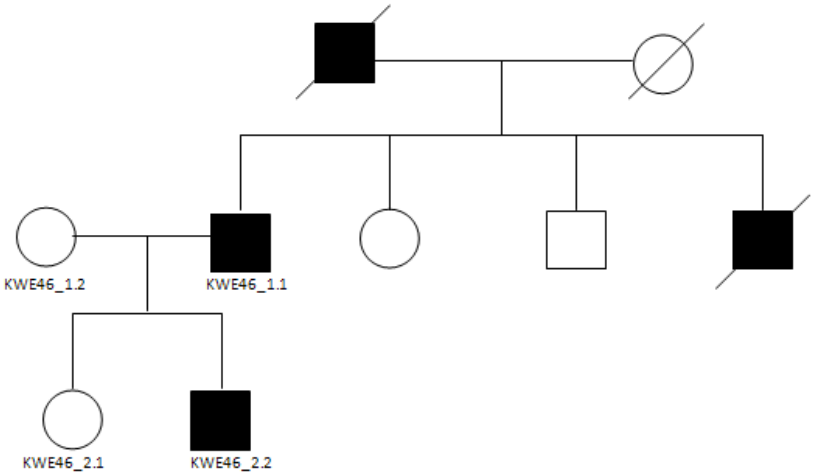


Figure 2.1 E: Family 46 pedigree

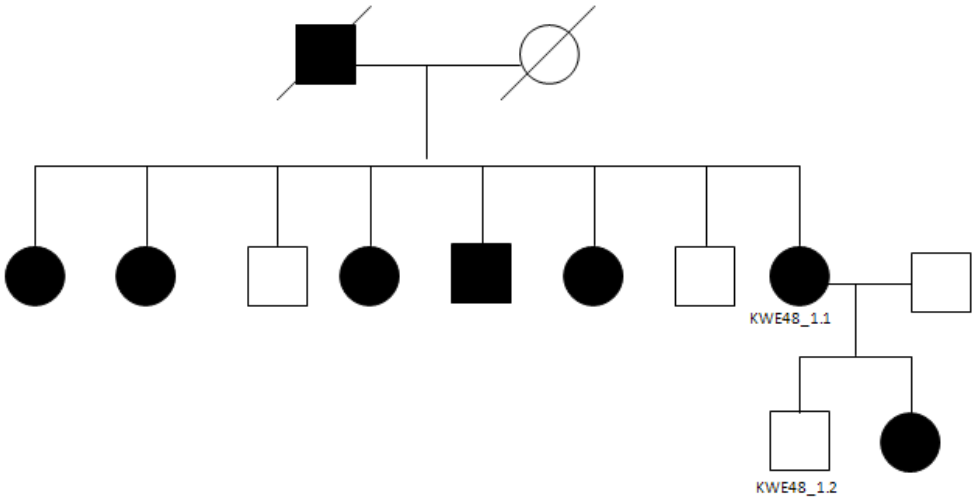


Figure 2.1 F: Family 48 pedigree

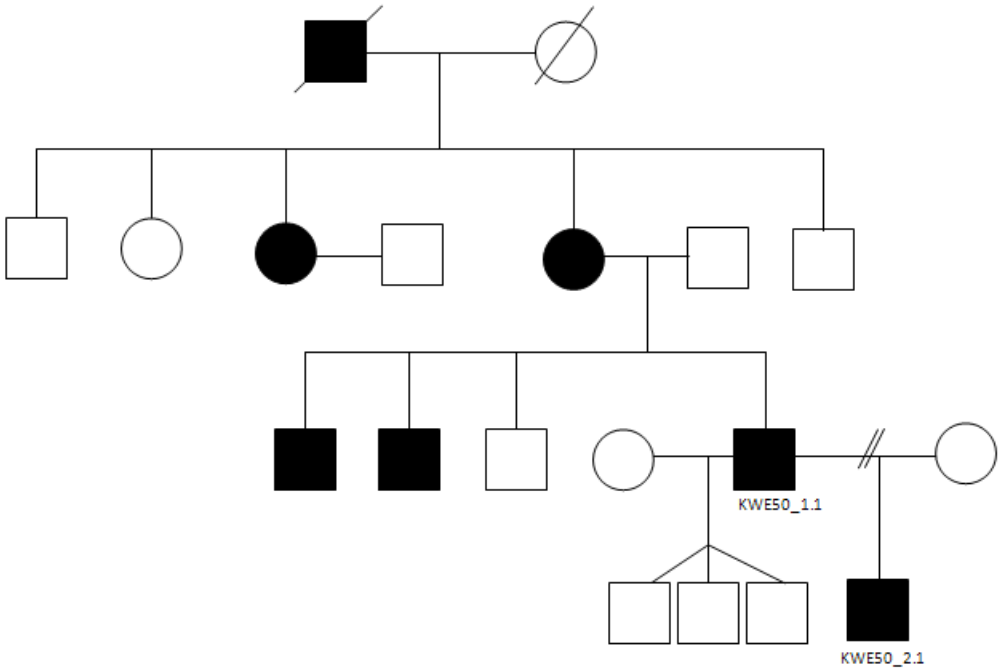


Figure 2.1 G: Family 50 pedigree

APPENDIX B

Reagents used in the study

- **2% Agarose Gel**

  2g agarose

  100ml 1xTBE

  Mix together, and boil until agarose is completely dissolved. Allow to cool for 30 minutes. Add 3µl Ethidium Bromide (10mg/µl) to the solution. Pour into gel casting tray to set.

- **dNTP mix (for PCR)**

  10µl 100mM dATP

  10µl 100mM dCTP

  10µl 100mM dGTP

  10µl 100mM dTTP

  760µl ddH2O

- **70% Ethanol**

  Add 70ml absolute ethanol to 30ml dH_2O to make a total volume 1L

- **Ficoll-Bromophenol Blue Loading Dye**

  50g 50% sucrose crystals

  0.1ml 0.5M EDTA (pH7.0)

  0.1g 0.1% bromophenol blue dye

  10g 10% ficoll powder

  Make up to 100ml with ddH_2O. Aliquot into 1.5ml eppendorf tubes. Store at 4°C.

- **1kb or 1kb+ Molecular Weight DNA Marker (1μg/μl)**

  250μl ladder (0.25μg/μl stock, GIBCO BRL)

  125μl Ficoll-Bromophenol Blue Loading Dye

  2.1ml 1xTE

  Store in smaller aliquots at 4°C

- **10x Tris Borate EDTA buffer (TBE)**

  109.02g Tris base

  55.64g Boric acid

  7.44g NaEDTA

  Dissolve all reagents in 1L dH_2O. Adjust the pH to 8.3 with HCl. Autoclave and store at room temperature. Dilute 10-fold before use

- **1x Tris-EDTA (TE) Buffer**

  10ml 1M Tris-HCl (pH8.0)

  2ml 0.5M EDTA

  Make up to 1L with dH2O. Autoclave and store at room temperature.

APPENDIX C: Extractions

C1

DNA extraction from saliva using the Oragene·DNA kit (DNA Genotek)

2.2.2.1 Saliva sample collection

Step 1

The participant delivers saliva into the Oragene·DNA collection vial.

Step 2

The collection vial is closed. By capping the vial, the container releases the DNA-conserving fluid which mixes with the saliva. DNA is now stable for long-term storage at room temperature.

- 500µl of the mixed Oragene·DNA/saliva sample was transferred to a 1.5mL microcentrifuge tube. 20µl of the Oragene·DNA purifier (OG-L2P, supplied) was added to the tube and vortexed for a few seconds to ensure adequate mixing.
- The sample was incubated on ice for 10 minutes and then centrifuged for 5 minutes at 13,000 rpm. The clear supernatant was carefully transferred into a clean microcentrifuge tube and the pellet containing impurities was discarded.
- 500µl of 100% ethanol was added to the supernatant. The sample was gently inverted 10 times to ensure sufficient mixing of the ethanol. The sample was left to stand at room temperature for 10 minutes to allow the DNA to fully precipitate.
- The sample was centrifuged for 2 minutes at 13,000rpm. The supernatant was carefully removed as to avoid disturbing the DNA-containing pellet. The

pellet was washed with 70% ethanol at left to stand at room temperature for 1 minute.

- Once the ethanol had been completely removed, 100µl of TE buffer was added to dissolve the pellet. The expected concentration of the fully hydrated DNA is 20 to 200ng/µl.

APPENDIX C.II

RNA extraction from whole blood using the QIAamp RNA blood Mini kit (QIAGEN)

Equipment and Reagents to be supplied by User

Protocol

- Mix 1 volume of whole human blood with 5 volumes of Buffer EL in a nunc tube.
- Incubate for 15 minutes on ice. Mix by vortexing briefly twice during incubation.
- Centrifuge at 400 x g for 10 minutes at 4°C, and remove supernatant.
- Add 2 volumes of Buffer EL per volume of whole blood used in step 1 to the cell pellet and resuspend the cells by vortexing.
- Centrifuge at 400 x g for 10 minutes and remove supernatant.
- Add 350µl Buffer RLT (with ®-ME) to pelleted leukocytes. Vortex.
- Pipet the lysate directly into a QIAshredder spin column in a 2 ml collection tube (provided) and centrifuge for 2 min at full speed.
- Discard QIAshredder spin column and save homogenized lysate.
- Add 350µl of 70% ethanol to the homogenized lysate and mix by pipetting.
- Pipet sample into a QIAamp spin column in a 2 ml collection tube. Centrifuge for 15s at 8000 x g.
- Transfer the QIAamp spin column into a new 2ml collection tube. Add 70 µl Buffer RW1 to the QIAamp spin column and centrifuge for 15s at 8000 x g.
- Place the QIAamp spin column into a new 2 ml collection tube. Add 500 µl Buffer RPE to the QIAamp spin column and centrifuge for 15s at 8000 x g.

- Carefully open the QIAamp spin column and add 500 µl of Buffer RPE. Close the cap and spin at 20 000 x g for 3 minutes.
- Place the QIAamp spin column into a new 2ml collection tube and centrifuge at full speed for 1 minute.
- Transfer the QIAamp spin column into a 1.5ml microcentrifuge tube and pipet 30 – 50µl of RNase-free water directly onto the QIAamp membrane.
- Centrifuge for 1 minute at 8000 x g to elute. Repeat if >0.5ml whole blood was processed.

APPENDIX C.III

RNA extraction from saliva using the ORAgene·RNA kit (DNA Genotek)

Steps prior to purification of RNA from saliva/ Oragene·RNA samples

- Shake samples vigorously when received in lab.
- Samples may be stored at room temperature for up to 8 weeks or stored at -20°C indefinitely.
- Prior to purification, incubate the entire sample at 50°C overnight in an air incubator.

Initial purification of an aliquot of Oragene·RNA samples

- Remove 500µl of the saliva sample into a 1.5mL microcentrifuge tube.
- Incubate the aliquot at 90 °C and cool to room temperature.
- Add 20µl to the sample and incubate on ice for 10 minutes.
- Centrifuge at maximum speed (>13 000 x g) for 3 minutes.
- Carefully remove the supernatant to a fresh tube and discard the pellet.
- Add 1000µl of cold 95% ethanol. Mix thoroughly by inversion.
- Incubate at -20°C for 30 minutes.
- Centrifuge at maximum speed for 3 minutes.
- Carefully remove and discard the supernatant.
- Dissolve the pellet in 350µl of Buffer RLT by vigorous vortexing. Ensure that the pellet is completely dissolved.
- Add 350µl 70% ethanol. Mix well by vortexing.

Proceed to step 5 of the Qiagen RNeasy Micro kit

- Transfer the sample onto a RNeasy MinElute spin column in a 2ml collection tube. Centrifuge for 15 seconds at 8000 x g.
- Discard the flow-through. Reuse the collection tube.
- Add 350µl of Buffer RW1. Centrifuge for 15 seconds at 8000 x g.
- Discard the flow-through. Reuse the collection tube.
- Add 10µl DNase I stock solution to 70µl Buffer RDD. Mix gently.
- Add the DNase I incubation mix (80µl) directly to the RNeasy MinElute spin column membrane and incubate for 15 minutes.
- Add 350µl Buffer RW1 to the RNeasy MinElute spin column and centrifuge for 15 seconds at 8000 x g. Discard the flow-through and the collection tube.
- Place the RNeasy MinElute spin column into a fresh 2ml collection tube. Add 500µl Buffer RPE to the spin column.
- Centrifuge for 15 seconds at 8000 x g and discard the flow-through.
- Add 500µl 80% ethanol to the RNeasy MinElute spin column and centrifuge for 2 minutes at 8000 x g.
- Place the RNeasy MinElute spin column into a 2ml collection tube. Centrifuge for 5 minutes at full speed with the lid of the spin column open.
- Place the RNeasy MinElute spin column into a fresh 1.5mL collection tube. Add 25µl RNase-free water directly to the centre of the spin column membrane.
- Incubate for 5 minutes at room temperature. Close the lid and centrifuge for 1 minute at full speed to elute the RNA.

APPENDIX C.IV

Extraction of total RNA from epidermal skin biopsies using the RNeasy Micro kit from Qiagen

Things to do before starting: Add 10µl β-Mercaptoethanol to 1ml RLT buffer

Method used in the study: Disruption using a mortar and pestle followed by homogenisation using a needle and syringe.

- Place tissue in liquid nitrogen and grind thoroughly using the mortar and pestle. Decant tissue powder into a RNAse-free 2ml microcentrifuge tube and add 350µl Buffer RLT Plus.
- Pass the lysate 5 times through a blunt 20-gauge needle. Centrifuge the lysate at full speed for 3 minutes.
- Remove the lysate carefully by pippetting it into a gDNA eliminator column place in a collection tube. Centrifuge for 30 seconds at 10 000 rpm. Discard the column and save flow through.
- Add 350µl 70% ethanol to the flow through. Carefully pipette the sample into a RNeasy MinElute column and centrifuge for 15 seconds at 10 000 rpm. Discard flow-through.
- Add 700µl Buffer RW1 to the column. Centrifuge for 15 seconds at 10 000 rpm. Discard flow-through. Add 500µl Buffer RPE to the column, centrifuge for 15 seconds at 10 000 rpm. Discard flow-through.
- Add 500µl 80% ethanol to the column. Centrifuge for 2 minutes at 10 000 rpm. Discard flow-through and collection tube.

- Place RNeasy MinElute column into a clean 2ml collection tube and centrifuge at full speed for 5 minutes. Discard collection tube.
- Place the column into a clean 1.5ml collection tube and add 14µl RNase-free water to the middle of the column. Centrifuge for 1 minute at full speed to elute RNA.

APPENDIX D

Improm-II Reverse transcription system (Promega)



This protocol outlines the synthesis of cDNA for subsequent amplification using PCR.

1. **Target RNA and Primer Combination and Denaturation**
    - Thaw total RNA and the Improm-II Reverse transcriptase system components on ice.
    - On ice, add 1µl Oligo dT primer to 4µl of RNA for a final volume of 5µl per RT reaction.
    - Close each tube tightly and spin for 10 seconds.
    - Place the tubes into a preheated 70°C heat block for 5 minutes.
    - Immediately chill in ice-water for at least 5 minutes.
    - Spin each tube for 10 seconds in a microcentrifuge to collect the condensate and maintain the original volume.
    - Keep the tubes closed and on ice until the reverse transcription reaction mix is added.
2. **Reverse transcription**
    - Prepare the reverse transcription reaction mix by combining the following components of the ImProm-II Reverse Transcription System in a sterile 1.5ml microcentrifuge tube on ice.

Component	Final concentration	Volume (µl)
Nuclease-free water		5.5
Improm-II™ 5X Reaction buffer		4
$MgCl_2$		2
dNTP mix		1
Recombinant RNasin® Ribonuclease Inhibitor		1.5
Improm-II™ reverse Transcriptase		1
Total		15

- For the negative control, make a mix that does not contain Improm-II™ reverse Transcriptase.
- Add 15µl of the reverse transcription reaction mix to each reaction tube containing the 5µl RNA/primer for a final reaction volume of 20µl per tube.
- Place the tubes in a PCR machine and run it following the thermal cycling conditions listed in the table below.

Thermal cycling conditions		
	Temperature (°C)	Time
Annealing	25	5 minutes
Extension	42	1 hour
Inactivate Reverse transcriptase	70	15 minutes
Hold	4	∞

3. PCR amplification

- ImProm-II Reverse Transcription reaction conditions support PCR amplification. No dilution of the cDNA is necessary. Add heat-inactivated reverse transcription reaction products directly to the PCR mix.

APPENDIX E.I

Exonuclease – Shrimp Alkaline Phosphatase cleanup of PCR products

This protocol is used for 25µL PCR products that require removal of excess dNTPs and primers prior to sequencing. The final volume of the reaction is 30µl after ExoSAP treatment (accounting for 5µL being run on an agarose gel previously).

Materials

- Shrimp Alkaline Phosphatase (1U/µL) (Roche Cat. No. 1 758 250)
- Exonuclease I (20U/µL) (New England Biolabs Cat. No. M0293S)
- PCR products in 0.2mL PCR tubes

Protocol

- Prepare a master mix of Exonuclease I/Shrimp Alkaline Phosphatse (ExoSAP) for use on PCR products; make mix for the required number of samples. Prepare master mix immediately before use and keep on ice. Only remove enzymes from the freezer immediately prior to use and use in a cold box, return enzymes to the freezer immediately after taking aliquots.

ExoSAP Mix	**1X**
Exonuclease I	0.025µL
SAP	0.250µL
Milli Q water	9.725µL
Final Volume	**10µL**

- Add 10µL of the ExoSAP mix to each PCR sample. Vortex the sample gently
- Spin the sample down for 5 – 10 seconds in an Eppendorf microcentrifuge

- Incubate samples at 37˚C for 30 minutes and then 80˚C for 5 minutes in a PCR machine. Store samples at 4˚C or -20˚C until needed.

APPENDIX E.II

Nucleospin® Extract II Protocol (Macherey-Nagel)

DNA extraction from agarose gel

- Excise the DNA fragment from the agarose gel, using UV light for guidance.
- Weigh the excised DNA fragment and place it into a clean Eppendorf tube.
- Add 200µl buffer NT for each 100 mg agarose gel. This buffer lyses the gel.
- Incubate samples at 50°C for 10 minutes, vortexing every 2 minutes until the gel is completely dissolved.
- Place a Nucleospin Extract II column into a collection tube and load the samples.
- Centrifuge the column for 1 minute at 14 000. Discard the flow-through. Place the column back into the column.
- Add 600µl buffer NT3 (that previously had 28 ml Ethanol added) and centrifuge for 1 minute at 14 000 rpm.
- Discard the flow-through in the collection tube and place column back into collection tube. Centrifuge for 2 minutes at 14 000 rpm.
- Incubate the column with the cap open for 2-5 minutes at 70 °C and discard the collection tube. Place the column into a clean Eppendorf tube.
- Add 15–50µl ddH_2O (this volume must equal the volume of starting material – PCR product, prior to adjusting it with 1 X TE). Incubate for 1 minute at room temperature. Centrifuge for 1 minute at 14 000 rpm. Discard column.

APPENDIX E.III

Qiagen DyeEx 2.0 Spin Kit (Qiagen)

- Gently vortex the spin column to resuspend the resin.
- Loosen the cap by ¼ turn and snap off the bottom of the spin column.
- Place the spin column in a collection tube.
- Centrifuge for 3 minutes at 3000 rpm. Transfer the spin column in to a clean Eppendorf tube.
- Slowly add 20µl cycle sequencing product directly onto the center of the gel bed.
- Centrifuge for 3 minutes at 3000 rpm. Discard the spin column.

APPENDIX F

Automated sequencing with the 3130xl Genetic Analyzer (Applied Biosystems 3130xl Genetic Analyzers: Getting started guide. 2004)

- Add 10µl HiDi™ Formamide (Applied Biosystems) to the vacuum dried samples and mix.
- Transfer the samples into a MicroAmp® optical 96-well reaction plate (Applied Biosystems) and cover it with a MicroAmp® optical 96-well plate septa.
- Make sure that the samples are added in multiple of 16. If the total number of samples was not a multiple of 16, then 10 µl HiDi™ Formamide must be added to the wells to make the sample number a multiple of 16.
- Centrifuge the plate and denature for 2 minutes at 95°C using the GeneAmp® PCR system 9700 (Applied Biosystems).
- Load the MicroAmp® optical 96-well reaction plate into the plate base. Snap plate retainer into position and ensure that the holes of the plate's retainer align with those of the septa.
- Press the tray button and allow the autosampler to stop in a forward position. Load the MicroAmp® optical 96-well reaction plate into either the A or B position of the Genetic Analyzer.
- Check that the levels of the water and anode and cathode buffer reservoirs are above the minimum requirement.

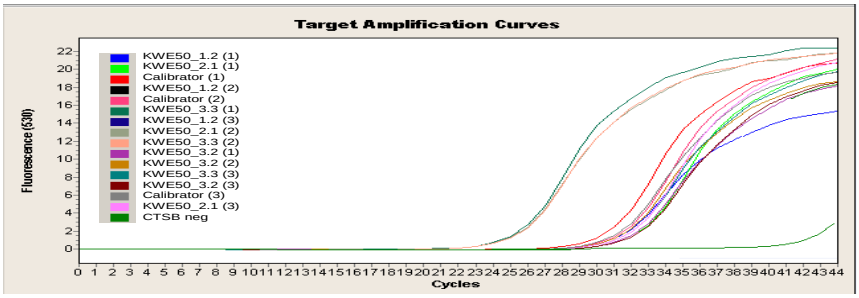
- Prepare a sample sheet. Select priority as 100. The instrument protocol is Z_seq_POP7_36 and for analysis 3130POP7_BDTv3-36 is used.Link the sample sheet to the correct plate (Either plate A or B).
- Press the “Analyze” button.

Sequence Analysis v5.2 software (Applied Biosystems)

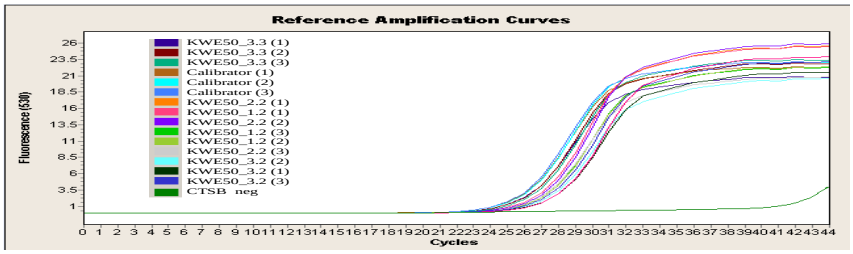
- Select the samples to be analyzed.
- Click “Add selected samples”.
- Check the Base-calling (BC) box and press the play button.
- Once analyzed, a blue or yellow block will form around the BC box. A blue block indicates that a sample produced a good quality sequence and a yellow block indicates poor sequence quality of a sample.

APPENDIX G
Experimental Run A

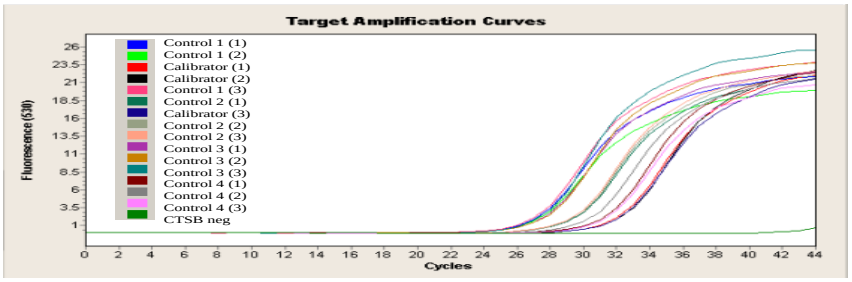
CTSB Patients



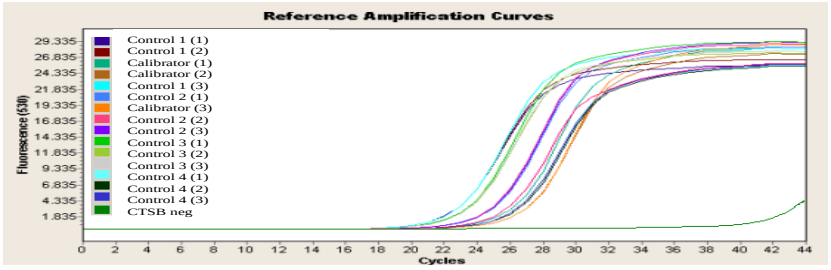
CTSB Patients (*GAPDH)*



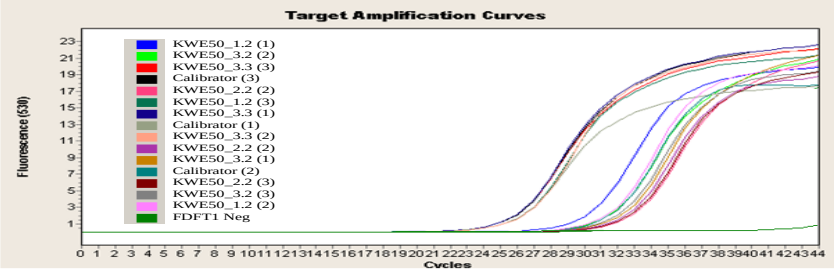
CTSB Controls



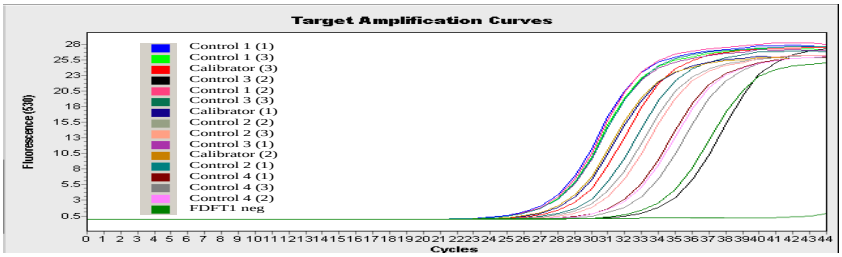
CTSB Controls (*GAPDH)*



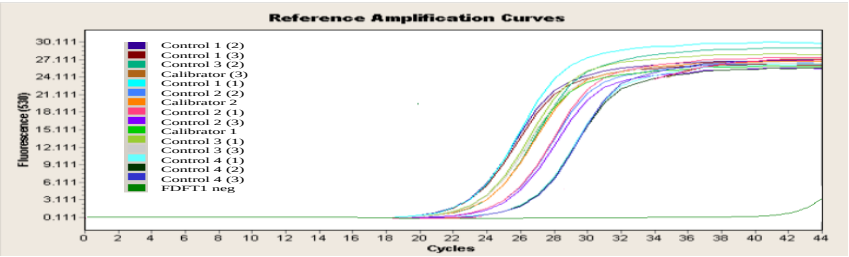
***FDFT1* Patients**



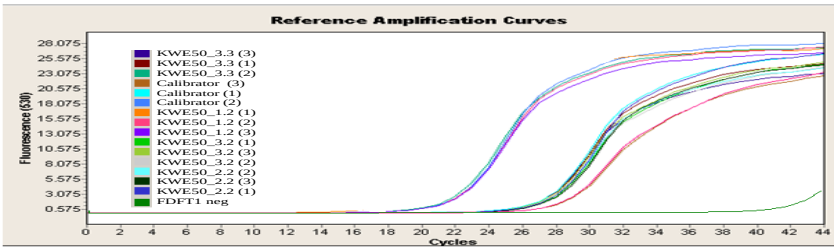
***FDFT1* Controls**



***FDFT1* Patients (*GAPDH*)**

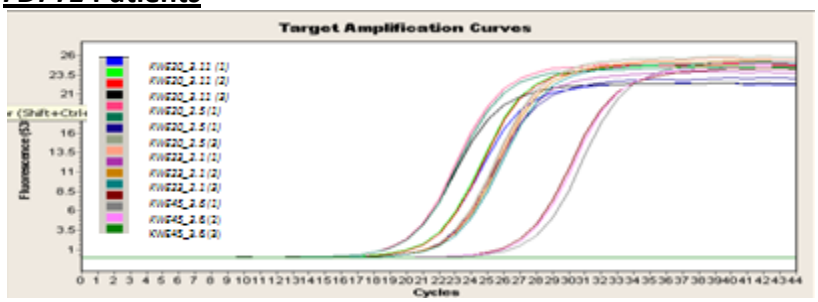


***FDFT1* Controls (*GAPDH*)**

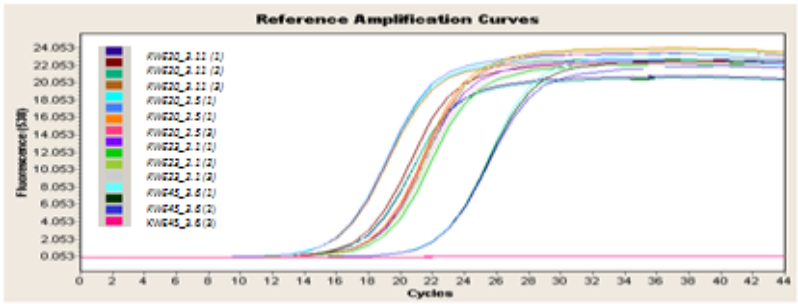


Experimental run B

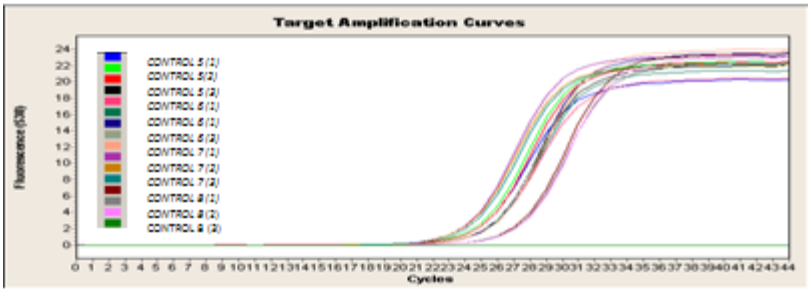
FDFT1 Patients



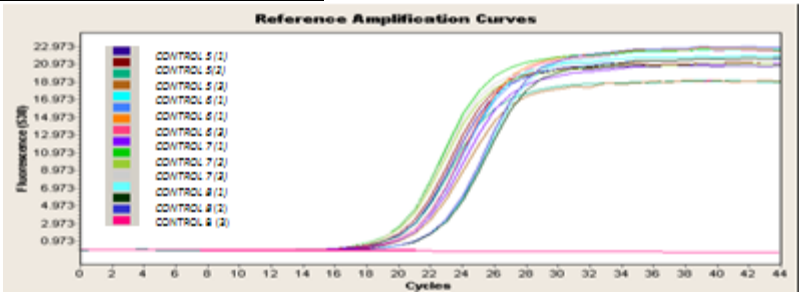
FDFT1 Patients (GAPDH)



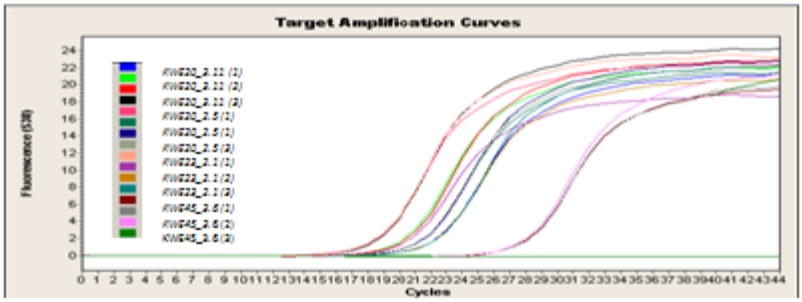
FDFT1 Controls



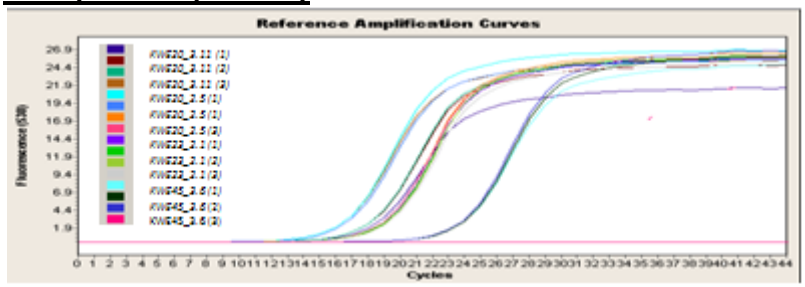
FDFT1 Controls (*GAPDH*)



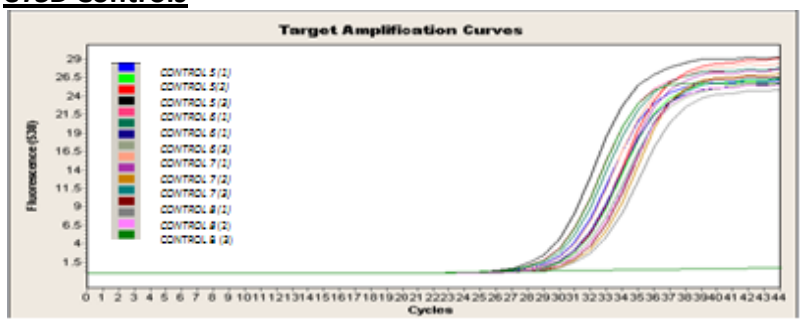
***CTSB* Patients**



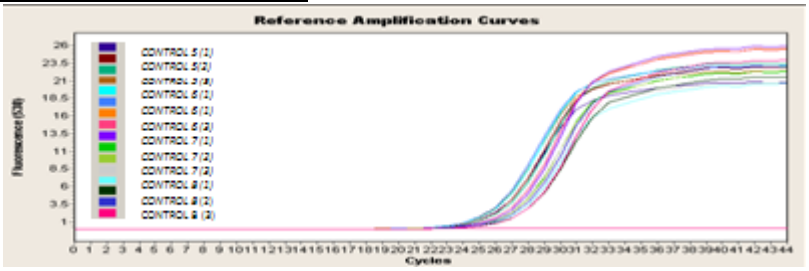
***CTSB* patients (*GAPDH*)**



***CTSB* Controls**



***CTSB* Controls (*GAPDH*)**



APPENDIX H

Ethics Clearance Certificate

APPENDIX I

Information and consent form for blood sampling

The National Health Laboratory Service
University of the Witwatersrand, School of Pathology Division of
Human Genetics

Hospital Street, Johannesburg 2001 P.O.Box 1038, Johannesburg 2001
Telephone: + 27-11-489-9224/9223/9211 Telefax: +27-11-489-9226 or +27-11-489-9209

Prof D Viljoen 489-9239 Prof A Krause 489-9219 Prof A Christianson 489-9212 Prof M Ramsay 489-9214
Prof H Soodyall 489-9208 Dr T Lane 489-9221 Ms T Wessels 489-9243

KERATOLYTIC WINTER ERYTHEMA STUDY

Information and consent form for blood sampling

We would like to invite you to participate in a genetics study aimed at finding out what causes keratolytic winter erythema. Participation is entirely voluntary and you will not be disadvantaged in any way if you choose not to participate.

Keratolytic winter erythema is associated with hyperkeratosis (thickening of the outer layer of skin), erythema (reddening of the skin) and recurrent peeling of the palms and soles. It is inherited as an autosomal dominant trait (only one parent needs to be affected for the offspring to be at risk to inherit the trait); occurs in successive generations and is observed in several individuals from the same family (parents and children). The aim of the study is to identify and understand the genetic basis of keratolytic winter erythema. This may in the future lead to improved treatment for the disorder, although this may take several decades.

To do these studies we need to take a blood sample from individuals with keratolytic winter erythema and their family members, both unaffected and affected. We will also need blood samples from control subjects who are unrelated to these families, but who live in the same area and have a similar ethnic background. An appropriately qualified person will take the blood samples. There will be a little discomfort. We will take the equivalent of about two or three tablespoons (10-30ml) of blood from each individual. All samples will be coded and any information obtained from this study will be completely confidential. You will not be disadvantaged in any way if you decide not to participate. No information will be reported back as this study is likely to take many years. The test will not involve any costs for the participants and participation is completely voluntary.

Should you require further information, please do not hesitate to contact:
Professor Michèle Ramsay, Division of Human Genetics
National Health Laboratory Service
School of Pathology, PO Box 1038, Johannesburg, 2000
(011) 489 9214

I, ……, have read and understood this information sheet and agree to participate in the study.
Signature:…… Date:……

I have fully explained the procedure and the purpose of the study and answered all questions to the best of my ability.
Signature:…… Date:……

APPENDIX I.II

Information and consent form for skin sampling

The National Health Laboratory Service
University of the Witwatersrand, School of Pathology
Division of Human Genetics

Hospital Street, Johannesburg 2001 P.O.Box 1038, Johannesburg 2001
Telephone: + 27-11-489-9224/9223/9211 Telefax: +27-11-489-9226 or +27-11-489-9209

Prof D Viljoen 489-9239 Prof A Krause 489-9219 Prof A Christianson 489-9212 Prof M Ramsay 489-9214
Prof H Soodyall 489-9208 Dr T Lane 489-9221 Ms T Wessels 489-9243

KERATOLYTIC WINTER ERYTHEMA STUDY

Information and consent form for skin biopsy

We would like to invite you to participate in a genetics study aimed at finding out what causes keratolytic winter erythema. Participation is entirely voluntary and you will not be disadvantaged in any way if you choose not to participate.

Keratolytic winter erythema is associated with hyperkeratosis (thickening of the outer layer of skin), erythema (reddening of the skin) and recurrent peeling of the palms and soles. It is inherited as an autosomal dominant trait (only one parent needs to be affected for the offspring to be at risk to inherit the trait); occurs in successive generations and is observed in several individuals from the same family (parents and children). The aim of the study is to identify and understand the genetic basis of keratolytic winter erythema. This may in the future lead to improved treatment for the disorder, although this may take several decades.

To do these studies we need to take a skin biopsy from selected individuals with keratolytic winter erythema and control subjects. This procedure does cause some discomfort and does hurt a little. It involves the removal of a small piece of skin (3mm in diameter) from the palmar region. After the procedure a steri-strip will be placed over the biopsy region to prevent any infections and to help the wound heal better. No stitches will be needed. The procedure will be carried out by a qualified dermatologist. All samples will be coded and any information obtained from this study will be completely confidential. You will not be disadvantaged in any way if you decide not to participate. No information will be reported back as this study is likely to take many years. The test will not involve any costs for the participants and participation is completely voluntary.

Should you require further information, please do not hesitate to contact:
Professor Michèle Ramsay, Division of Human Genetics
National Health Laboratory Service
School of Pathology, PO Box 1038, Johannesburg, 2000
(011) 489 9214

I, ……, have read and understood this information sheet and agree to participate in the study.
Signature:…….. Date:……
I have fully explained the procedure and the purpose of the study and answered all questions to the best of my ability.
Signature:…….. Date:……