

https://dspace.org/

The identification and characterisation of the causative gene mutation for keratolytic winter erythema (KWE) in South African families

Ngcungcu, Thandiswa

2017

http://hdl.handle.net/10539/23084

Ngcungcu, Thandiswa Goodness (2017) The identification and characterisation of the causative gene mutation for keratolytic winter erythema (KWE) in South African families, University of the Witwatersrand, Johannesburg, http://hdl.handle.net/10539/23084

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

The identification and characterisation of the causative gene mutation for keratolytic winter erythema (KWE) in South African families

Thandiswa Ngcungcu



A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment for the degree of Doctor of Philosophy

Johannesburg, 2017

Declaration

I Thandiswa Ngcungcu declare that this Thesis/Dissertation/Research Report is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Thandiswa Ngcungcu

21st day of June 2017 in Barcelona, Spain

Dedication

To my mother, Gladys Nolubabalo Ngcungcu, thank you for your never ending love and support through this journey. Thank you for all the sacrifices that you made so that I could live my dreams. Thank you for believing in me and always encouraging me to be the best that I can be. I love you!

Presentations arising from this study

Thandiswa Ngcungcu, Martin Oti, Peter Hull, Jan Sitek, Bjorn Haukanes, Fan Yang, Robert Bruccoleri, Tomasz Stokowy, Edward Oakeley, Bolan Linghu et al. "Duplicated enhancer upstream of the *CTSB* gene segregates with keratolytic winter erythema in South African and Norwegian families". Oral presentation at the European Human Genetics Conference (Barcelona, 21-24 May 2016).

Thandiswa Ngcungcu, Martin Oti, Peter Hull, Jan Sitek, Bjorn Haukanes, Fan Yang, Robert Bruccoleri, Tomasz Stokowy, Edward Oakeley, Bolan Linghu et al. "Non-coding variants segregate with disease in South African families with keratolytic winter erythema (KWE)". Oral presentation at the 7th Cross Faculty Graduate Symposium (University of the Witwatersrand, 01-02 March 2016).

Thandiswa Ngcungcu, Bolan Linghu, Fan Yang, Edward Oakeley, Frank Staedtler, Robert Bruccoleri, Thomas Morgan, Nanguneri Nirmala, Stine Buechmann-Moller, Marc Sultan, et al. "Non-coding variant segregates with disease in South African families with keratolytic winter erythema (KWE)". Poster presentation at the American Society of Human Genetics (Baltimore, 16-19 October 2015).

Thandiswa Ngcungcu, Bolan Linghu, Fan Yang, Edward Oakeley, Frank Staedtler, Robert Bruccoleri, Thomas Morgan, Nanguneri Nirmala, Stine Buechmann-Moller, et al. "Non-coding variant segregates with disease in South African families with keratolytic winter erythema (KWE)". Oral Presentation at the 16th Biennial Congress of the South African Society of Human Genetics (Pretoria, 16-19 August 2015).

Thandiswa Ngcungcu, Bolan Linghu, Fan Yang, Edward Oakeley, Frank Staedtler, Robert Bruccoleri, Thomas Morgan, Nanguneri Nirmala, Stine Buechmann-Moller, Marc Sultan, et al. "Non-coding variants segregate with disease in South African families with keratolytic winter erythema (KWE)". Oral Presentation at the Genomics of Rare Disease: Beyond the Exome conference (Hinxton, 29 April – 01 May 2015).

Thandiswa Ngcungcu, Bolan Linghu, Fan Yang, Edward Oakeley, Frank Staedtler, Robert Bruccoleri, Thomas Morgan, Nanguneri Nirmala, Stine Buechmann-Moller, Marc Sultan, et al. "Novel *COL14A1* variant identified in a family with keratolytic winter erythema". Poster presentation at the Novartis Institutes of Biomedical Research Visiting Scholar Research Day (Cambridge, MA, 05 August 2014).

Publications arising from this study

Thandiswa Ngcungcu, Martin Oti, Jan C. Sitek, Bjørn I. Haukanes, Bolan Linghu, Robert Bruccoleri,' Tomasz Stokowy, Edward J. Oakeley, Fan Yang, Jiang Zhu, Marc Sultan, Joost Schalkwijk, Ivonne M.J.J. van Vlijmen-Willems, Charlotte von der Lippe, Han G. Brunner, Kari M. Ersland,' Wayne Grayson, Stine Buechmann-Moller, Olav Sundnes, Nanguneri Nirmala, Thomas M. Morgan, Hans van Bokhoven, Vidar M. Steen, Peter R. Hull, Joseph Szustakowski, Frank Staedtler, Huiqing Zhou, Torunn Fiskerstrand, Michèle Ramsay. (2017) Duplicated enhancer region increases expression of *CTSB* and segregates with Keratolytic Winter Erythema in South African and Norwegian families. American Journal of Human Genetics, 100 (5), 737-750.

Abstract

Keratolytic winter erythema (KWE) is a rare autosomal dominant skin disorder characterized by recurrent episodes of palmoplantar erythema and epidermal peeling, and symptoms worsen in winter. KWE is relatively common in South African (SA) Afrikaners and was mapped to 8p23.1-p22 through a common haplotype in SA families. The aim of this study was to identify and characterize the causal mutation for KWE in SA families.

Targeted resequencing of 8p23.1-22 was performed in three families and seven unrelated controls. Reads were aligned to the reference genome using BWA. GATK and Pindel were used to call small and large structural variants, respectively. A 7.67 kb tandem duplication was identified upstream of the *CTSB* gene and encompassing an enhancer element that is active in a keratinocytes (based on H3K27ac data). The tandem duplication segregated completely with the KWE. The tandem duplication overlaps with a 15.93 kb tandem duplication identified in two Norwegian families at a 2.62 kb region encompassing the active enhancer suggesting that the duplication of the enhancer leads to the KWE phenotype.

Existing chromatin structure, CTCF binding and chromatin interaction data from several cell lines, including keratinocytes were analysed and three potential topological subdomains were identified, all containing the enhancer and *CTSB*, or *CTSB* and *FDFT1* or both genes and *NEIL2.* Additionally, we showed that the enhancer's activity correlated with *CTSB* expression, but not with *FDFT1* and *NEIL2* expression in differentiating keratinocytes and other cell lines. RNA polymerase II ChIA-PET interaction data in cancer cell lines showed that the enhancer interacts with *CTSB* but not *FDFT1* or *NEIL2*. These data suggest that the enhancer normally regulates *CTSB* expression. Relative gene expression and immunohistochemistry from palmar biopsies from South African and Norwegian participants (7 Affected and 7 Controls) showed a significantly higher expression of *CTSB,* but not *FDFT1* and *NEIL2,* in affected individuals compared to the controls and that CTSB was significantly more abundant in the granular layer of affected individuals compared to controls. We conclude that the enhancer duplication causes KWE by upregulating *CTSB* expression and causing an overabundance of CTSB in the granular layer of the epidermis.

Acknowledgements

Aspects of the Research performed by the candidate (T Ngcungcu) and by collaborators:

1. Sequencing and analysis
    a. Sequencing: Candidate
    b. Data Processing: Candidate and collaborators
    c. Filtering and analysis: Candidate
    d. Tandem duplication validation studies: Candidate
2. *In silico* analysis: Candidate and collaborators
3. Functional studies
    a. Patient recruitment and sample collection: Candidate and collaborators
    b. qPCR experiment design, lab work and analysis: Candidate and collaborators
    c. Immunohistochemistry, experiment design, lab work and analysis: Collaborators

I would like to thank the following people:

- My supervisor, Prof Michèle Ramsay: Thank you for all your support, guidance, encouragement, patience and wisdom over the past seven years. I couldn't have asked for a better supervisor and mentor.
- Dr Edward J. Oakeley and Dr Frank Staedtler for hosting me at Novartis in Basel and for their mentorship and guidance during my visit to Novartis Institutes for BioMedical Research (NIBR). Virginie Petitjean, for training me on DNA library preparation and sequencing and Marc Altorfer and Moritz Frei for their assistance in the laboratory.
- Dr's Bolan Linghu, Robert Bruccoleri, Fan Yang, Joseph Szustakowski for hosting me at Novartis Institutes for BioMedical Research, Cambridge, Massachusetts, and for training me on the bioinformatics analysis required for the study.
- Dr Torunn Fiskerstrand and the scientists in her group for their work on the Norwegian KWE study and their participation in relative gene expression studies and for training me to perform the assays and analyse the data.
- Prof Joost Schalkwijk and Ms Ivonne M.J.J. van Vlijmen-Willems for performing the histology and immunohistochemistry on the skin samples.
- Prof Huiqing Zhou and Dr Martin Oti for their role in the *in silico* functional studies of the KWE region.
- Dr Peter Hull for his willingness to answer my questions about KWE and his ongoing enthusiasm for the project over the years.
- Prof Wayne Graysen, Dr Karen Koch and Dr Laeeka Moosa for their help with the collection and processing of the skin biopsies.

- The participants who have been involved in the KWE study since the beginning and those who donated skin biopsies for this study, without whom this work would have not been possible.
- The postgraduate students and staff in the Division of Human Genetics and SBIMB for their help, support and on-going encouragement.
- Funders:
    - Novartis Next Generation Scientist Program and NIBR Visiting Scholar Program
    - NHLS Research Trust
    - Faculty Research Committee
    - National Research Foundation, South Africa
    - Wits Non-Communicable Disease Research Leadership Program (1D43TW008330-01A1)
- Lastly, I would like to thank my family and friends for your unwavering support.

Table of Contents

Declaration i

Dedication ii

Presentations arising from this study iii

Publications arising from this study v

Acknowledgements vii

List of Tables xiv

Abbreviations xv

Chapter 1 1

Introduction and Literature Review 1

1.1. The skin 1

1.1.1. The layers of the skin 2

1.1.2. The epidermis 4

1.1.2.1. Structure of the epidermis and epidermal differentiation 5

1.2. Keratolytic Winter Erythema 12

1.2.1. Clinical characterization of KWE 13

1.2.2. Nosology of KWE 14

1.2.3. Histological and microscopic findings 15

1.2.3.1. Light microscopy 16

1.2.3.2. Electron Microscopy 17

1.2.4. Age of onset 18

1.2.5. Aggravating factors and improvement 18

1.2.6. Genetics of KWE 18

1.2.6.1. Mode of inheritance and penetrance 18

1.2.6.2. Prevalence 19

1.2.7. Identification of the position of the KWE locus 19

1.2.8. Exclusion of genes in the KWE critical region 20

1.3. Alternative mutation finding strategies 22

1.3.1. Copy number variation 22

1.3.2. Targeted capture resequencing of the KWE critical region 24

1.3.3. Exome sequencing 24

1.4. Genome structure and gene regulation 25

1.4.1. Histone modifications 25

1.4.2. Gene regulation through chromatin structure 27

1.5. Aim and Objectives 28

Chapter 2 29
Materials and Methods 29
2.1. Participants 29
2.2. Laboratory methods 32
2.2.1. DNA extraction 32
2.3. Processing of NGS data 35
2.4. Small variant calls using the GATK pipeline 35
2.4.1. Pre-processing 36
2.4.2. Variant calling 37
2.4.3. Variant annotation 37
2.4.4. Variant Filtering 38
2.5. Structural variant calling using Pindel 38
2.6. Validation of potential causal variants 38
2.7. *In silico* functional analysis 39
2.7.1. Frequency of structural variants and overlapping regulatory elements 39
2.7.2. Identification of regulatory elements that overlap with the causal variant 39
2.7.3. Genomic architecture prediction 40
2.7.4. Known regulatory element interactions 40
2.7.5. Epigenomic profiling using public databases 40
2.8. In vitro functional analysis 41
2.8.1. Relative gene expression analysis 42
Chapter 3 45
Results 45
3.1. Coverage and small variant analysis 45
3.2. Structural variant analysis 45
3.2.1. A novel tandem duplication segregates with KWE 46
3.2.2. Validation of the tandem duplication 46
3.3. In silico functional analysis 49
3.3.1. Architecture of the tandem duplication 49
3.3.2. The tandem duplication overlaps with a tandem duplication in Norwegian KWE families ..... .50
3.3.3. Transcription factor binding sites in the duplicated enhancer region 50
3.3.4. Frequency of the two duplications and other duplications/deletions in the KWE critical region 52
3.3.5. Chromatin architecture in and around the duplicated regions 54
3.3.5.1. CTCF binding sites and interactions 54

3.3.5.2. Topologically associating domains using Hi-C data 56
3.3.6. Chromatin interactions within the duplicated region 57
3.3.7. Correlation of the duplicated enhancer activity with transcription of *CTSB, FDFT1* and *NEIL2* 59
3.4. In vitro functional analysis in skin biopsies 61
3.4.1. Relative gene expression for *CTSB*, *FDFT1* and *NEIL2* 61
3.4.2. Immunohistochemistry 63
3.4.2.1. Histology 63
3.4.2.2. Immunohistochemistry 64
3.5. Summary of the results 66
Chapter 4 67
Discussion 67
4.1. Exclusion of small variants as causal variants for KWE 68
4.2. Identification of a novel non-coding tandem duplication as the causal variant for KWE 68
4.2.1. Copy number variation of non-coding regions as an unusual mechanism for disease causation 69
4.3. Functional impact of the enhancer element 70
4.3.1. The enhancer occurs in the same topological domain as *CTSB*, *FDFT1* and *NEIL2* 70
4.3.2. The enhancer interacts with the promoter of *CTSB* 73
4.3.3. The enhancer's activity correlates with the expression of *CTSB* in keratinocytes and other cells 73
4.4. The duplication of the enhancer leads to the overexpression of its target gene, *CTSB* 74
4.5. Overexpression of *CTSB* and ectopic localisation of CTSB as a plausible cause for the KWE phenotype 76
4.5.1. CTSB affects keratinocyte cell-cell dissociation and keratinocyte migration 76
4.5.2. Absence of cathepsin inhibitors cause skin peeling disorders 77
4.5.3. Other cathepsin genes and their role in skin disease 79
4.6. Study Limitations 80
4.7. Future studies 81
Chapter 5 83
Conclusion 83
References 84
Appendices 90

List of Figures

Figure 1.1: A schematic representation of the three layers that make up human skin 2
Figure 1.2: A schematic representation of the layers that make up epidermis in palmoplantar skin.....8
Figure 1.3: Oudtshoorn shown on the map of Western Cape within the South Africa map ...12
Figure 1.4: KWE patients showing the erythema and peeling of palmar and plantar surfaces and the web space.....13
Figure 1.5: Haemotoxylin and Eosin (H&E) stained palm biopsy of a KWE patient. 15
Figure 2.1: Pedigrees of South African KWE families used for next generation sequencing...30
Figure 2.2: Pedigrees used for the validation of potential causal variants identified using NGS and for functional analysis 31
Figure 2.3: High level view of the workflow followed to identify, validate and characterize the KWE mutation.....33
Figure 2.4: GATK workflow for data NGS data processing and small variant calling. 36
Figure 2.5: Affected palm where the biopsy was taken.....42
Figure 3.1: Visualisation of the tandem duplication break points on the Integrative Genomics Viewer (IGV).....47
Figure 3.2: Mapping of the duplication breakpoints using gel electrophoresis and Sanger sequencing for the validation of the tandem duplication.....48
Figure 3.3: The tandem duplication lies upstream of the *CTSB* gene and overlaps with several regulatory elements 49
Figure 3.4: The South African and Norwegian tandem duplications overlap at an enhancer region active in keratinocytes 50
Figure 3.5: Transcription factors that bind to the enhancer element common in the South African and Norwegian tandem duplications.....51
Figure 3.6: CNVs in the KWE critical region reported in normal individuals overlapping with the duplicated region 53
Figure 3.7: Topological subdomains, CTCF binding sites and loops involving the enhancer and nearby genomic regions 55
Figure 3.8: NHEK Hi-C interactions around the duplicated region.....56
Figure 3.9: RNAPII ChIA-PET interactions between the enhancer and neighbouring *CTSB*.....58
Figure 3.10: Prediction of enhancer activity and transcription of nearby genes using ENCODE data.....60

Figure 3.11: Relative expression of three nearby genes that occur in the same domain with the duplicated enhancer 63
Figure 3.12: Haemotoxylin and Eosin (H&E) staining of hand palm biopsies of a healthy control (normal skin) and KWE patient (affected skin) 64
Figure 3.13: CTSB staining in a control individual and KWE case 65
Figure 4.1: ChIP-Seq analyses of H3K27ac during human primary keratinocytes (HKC) differentiation 69
Figure 4.2: CTCF binding sites, MCF-7 interaction loops and NHEK predicted CTCF interaction loops. 72
Figure 4.3: Activity of the enhancer and the three nearby genes during keratinocyte differentiation 74
Figure 4.4: StringDB output showing interactions of CSTA and CTSB with other proteins. 78

List of Tables

Table 1.1: Human Type I and Type II epithelial keratins 6

Table 1.2: Histone modifications and variations used in the ENCODE project to access the genome for regulatory elements* 26

Table 2.1: Participants used in the study shown according to the different experiments 29

Table 3.1: Interactions between the duplicated enhancer and nearby genes/regions based on CHIA-PET data 57

Table 3.2: Average CT values and fold changes for *CTSB*, *FDFT1* and *NEIL2* 62

Table 3.3: Number of granular layers and CTSB staining intensity in affected and control skin 65

Abbreviations

ΔCt	Change in thermal cycle
°C	Degrees celsius
µg	Microgram
µl	Micro litre
µM	Micro molar
3D	three-dimensional
A	Adenine
aCGH	Array comparative genomic hybridization
AFR	Africa
APSS	acral peeling skin syndrome
BDA2	brachydactyly type A2
BLAST	Basic Local Alignment Search Tool
BMP2	Bone morphogenetic protein 2
bp	Base pair
BWA	Burrows-Wheeler Aligner
C	Cytosine
C5	5-Methylcytosine
C8orf13	Family With Sequence Similarity 167 Member A (FAM167A)
CARD14	Caspase Recruitment Domain Family Member 14
CCL3L1	C-C Motif Chemokine Ligand 3 Like 1
cDNA	Complementary DNA
CE	Cornified envelope
CEBPB	CCAAT/Enhancer Binding Protein Beta
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag Sequencing
chr	Chromosome
cm	Centimetre
CNV	Copy number variant
COL14A1	collagen, type XIV
CpA	Cytosine-phosphate-adenine dinucleotide
CpG	Cytosine-phosphate-guanine dinucleotide
CpT	Cytosine-phosphate-thymine dinucleotide
CSTC /CST3	Cystatin C
CT	threshold cycle
CTCF	CCTC-binding factor
CTSA	cystatin A
CTSB	Cathepsin B
CTSD	Cathepsin D
CTSL	Catheosin L
ddH2O	Deionised distilled water
DEFB	Beta-defensin
DGV	Database of Genomic Variation
DHODH	Dihydroorotate Dehydrogenase (Quinone)
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSC	Desmocollins
DSG	Desmogleins

DUB3	Deubiquitinating enzyme 3
EDC	Epidermal differentiating complex
EDTA	Ethylenediaminetetraacetic acid
e.g.	Example
EKV	Erythrokeratoderma variabilis
EKVP	Erythrokeratodermia variabilis et progressiva
ENCODE	Encyclopedia of DNA Elements
ENGINES	ENtire Genome INterface
eQTL	Expression quantitative trait loci
ESP	NHLBI Exome Sequencing Project
EtBr	Ethedium bromide
FDFT1	Farnesyl-diphosphate farnesyltransferase
FSGS	Focal segmental glomerulosclerosis (FSGS
g	Gram
GATK	Genome Analysis Toolkit
GJB1	Connexin-43
GJB3	Connexin-31
GJB4	Connexin-30.3
H&E	Haemotoxylin and Eosin
H3K27ac	Acetylated histone H3 at lysine 27
H3K36me3	Trimethylated histone H3 at lysine 36
HCT-116	Human colon cancer cell line
HDAC	Histone de-acetyltransferases
HeLa-S3	Cervical carcinoma cell line
hg19	Human genome build 19
HAT	Histone acetyltransferases
HIV	Human immunodeficiency virus
HREC	Human Research Ethics Committee
HTS	Hass-type polysyndactyly
Indel	Insertion and deletion variant
JAK	Janus kinase
K	Keratin
K562	Chronic myeloid leukemia cell line
kb	Kilo base
kDA	Kilo Dalton
kg	Kilogram
KIF	Keratin intermediate filaments
KRT	Keratin
KWE	Keratolytic winter erythema
LCE	Late cornified envelope proteins
LD	Linkage disequilibrium
LMX1B	LIM Homeobox Transcription Factor 1 Beta
LSS	Laurin-Sandrow syndrome
m2	Metres squared
MAF	Minor allele frequency
MAPK	Mitogen-activated protein kinase
MCF-7	Michigan Cancer Foundation-7 breast cancer cell line
$MgCl_2$	Magnesium chloride

miRNA	MicroRNA
ml	Milli litre
mM	Milli molar
n	Number
NA^+	Sodium
NaOH	Sodium hydroxide
NB4	Acute promyelocytic leukemia cell line
ncRNA	Non-coding RNA
NEIL2	Nei like DNA glycosylase 2
ng	Nano gram
NGS	Next generation sequencing
NHEK	Primary Normal Human Epidermal Keratinocytes
ns	Non-synonymous
OMIM	Online Mendelian Inheritance in Man
ORF	Open reading frame
p	Short arm of a chromosome
PCR	Polymerase chain reaction
PPPK1B	Punctate palmoplantar keratoderma type IB
PSEK	Progressive symmetric erythrokeratoderma
PSORS2	Autosomal dominant psoriasis susceptibility locus 2
PSS4	Exfoliative ichthyosis
q	Long arm of a chromosome
QC	Quality control
r^2	Correlation coefficient
RefSeq	Reference sequence
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
RP1L1	Retinitis Pigmentosa 1-Like 1
RPLP0	Ribosomal Protein Lateral Stalk Subunit P0
rpm	Revolutions per minute
rs	Reference sequence
RSA	South Africa
s	Synonymous
S100	Calcium-binding proteins
SD	Standard deviation
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variants
SPPR	Small proline-rich proteins
SPSmart	SNPs-for-population-Studies-mart
SQL	Structured Query Language
STAT	Signal transducer and activator of transcription
STAT3	Signal transducer and activator of transcription 3
STR	Short tandem repeat
T	Thymine
TAD	Topologically associating domains
Taq	Thermus aquaticus polymerase
TDH	L-Threonine Dehydrogenase
TE	Tris/EDTA buffer

TF	Transcription factor
TFBS	Transcription factor binding site
TGF-α	Transforming growth factor-α
U	Uracil
UCSC	University of California, Santa Cruz
UK	United Kingdom
µl	Micro litre
USA	United States of America
UTR	Untranslated region
UV	Ultra violet
V	Volts
VEP	Variant Effect Predictor
vs.	Versus
WES	Whole exome sequencing
WHO	World Health Organisation

Chapter 1

Introduction and Literature Review

Keratolytic Winter Erythema (KWE; MIM 148370) is a rare autosomal dominant skin disorder of unknown aetiology. It is characterized by cyclical redness (erythema), epidermal layer thickening and the formation of dry blisters where centrifugal skin peeling originates at the palms and soles (Findlay and Morrison, 1978; Findlay et al., 1977; Hull, 1986). Interestingly, the symptoms appear to worsen in winter in many of the affected individuals. KWE globally rare, but fairly common in South African Afrikaners and Coloureds, as a result of genetic drift by founder effect in South African families (Hull, 1986; Starfield et al., 1997). In 1997, the KWE mutation was localised to chromosome 8p23.1-p22, between and including markers D8S550 and D8S1759, a region referred to as the KWE critical region (Appel et al., 2002; Starfield et al., 1997). Previous research approaches, such as candidate gene sequencing and gene expression analysis (Appel et al., 2002; Hobbs et al., 2012; Hull et al., 2013), proved insufficient in identifying the KWE causal mutation which remained elusive for twenty years following its localization to chromosome 8p. In this study, a new approach using targeted capture resequencing of the KWE critical region led to the identification of the causal mutation for KWE.

1.1. The skin

The skin, also known as the integument, is the largest organ making up approximately a sixth of the body weight (Hirobe, 2014; Wickett and Visscher, 2006). Hair, nails, skin oil, sweat gland are regarded as skin derivatives. The skin prevents water loss, synthesises vitamin D, and serves as the body's shield against environmental factors such as temperature fluctuations, radiation, infection, toxins and injury (Groenendaal et al., 2010; Nestle et al., 2009; Nithya et al., 2015; Wickett and Visscher, 2006). The skin is a flexible, continuously regenerating organ with varying thickness and varying metabolic state with age and environmental cues such as injury. The skin also harbours a complex microbiome (Grice et al., 2008) which helps to maintain skin health and varies between individuals depending on diet, environmental exposure and genetics. The microbiome may be protective against

pathogens or make individuals more susceptible to infection based on the composition and balance in bacteria and microflora on the skin. The skin is susceptible to a variety of diseases as a result of environmental exposures such as ultraviolet (UV) radiation which may lead to certain forms of cancer, exposure to pathogens which may lead to infection, allergens and harmful chemicals. Additionally skin may become dysregulated or abnormal due to genetic causes.

1.1.1. The layers of the skin

The skin is made up of three layers, namely the layer of subcutaneous fat, the dermis and the epidermis (Figure 1.1) (Hirobe, 2014), and disruption in the regulation of the formation and maintenance these layers may lead to disease.

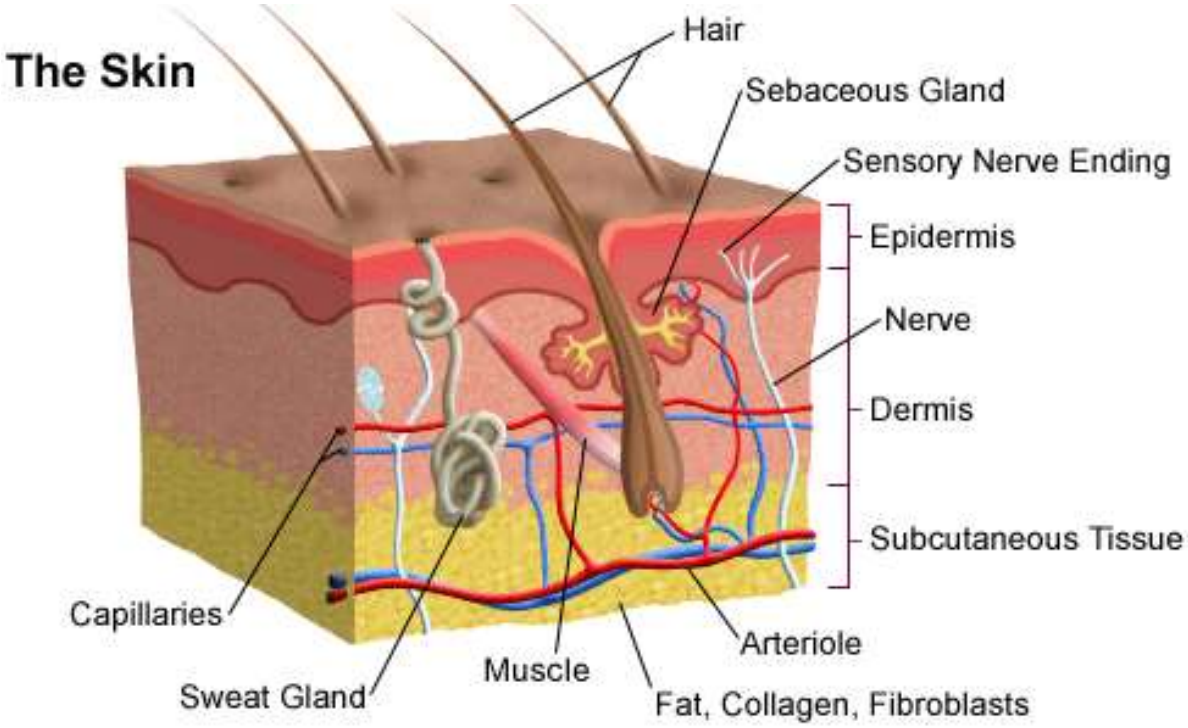


Figure 1.1: A schematic representation of the three layers that make up human skin
(From http://www.healthoncare.com/basic-human-skin-structure-functions-skin-care/)

The subcutaneous fat layer is the innermost layer of the skin that varies in thickness based on the location of the skin on the body. This layer attaches the rest of the skin to the muscle layers beneath through its connection to the dermis and covers the skeleton, muscles and other organs to maintain the structural integrity of the body. It also creates a physical barrier

between the environment and internal organs, bones and muscles by storing fat which acts as a cushion protecting bones and muscles from external physical pressure while also acting as an energy source for the organ (Osman et al., 2013). This layer also controls body temperature by acting as a heat and cold insulator. It made up of adipose tissue, and is the initial entry point of blood supply and nerves from the body to the skin.

The dermis is the multifunctional middle layer of the skin and is responsible for the skin's elasticity and strength which are achieved through the presence of different collagens and a wide range of proteins (Groenendaal et al., 2010; Osman et al., 2013). It also plays a major role in temperature regulation. The dermis has two collagen rich layers; the papillary and reticular layer (Osman et al., 2013). The papillary layer attaches the dermis to the epidermis whereas the reticular layer holds the major components of the dermis including the sweat glands and is responsible for the dermis's strength and elasticity (Osman et al., 2013). The dermis comprises mainly of fibroblasts, adipocytes and macrophages. It receives oxygen and nutrients through the vasculature which also serves to remove waste products. The blood vessels also serve in regulating body temperature in response to heat and cold by dilating or contracting to increase or decrease blood circulation to the surface of the skin thereby releasing or conserving heat.

The skin also regulates body temperature through skin appendages such as the sweat glands, hair follicles, sebaceous glands and sweat glands (Xie et al., 2016). Sweat glands produce sweat in response to stress and heat. Hair follicles stem from the epidermis and also regulate temperature, enhance skins sensation and act as physical barrier from injury. The number of hair follicles varies across the body and are absent on palmoplantar skin. The dermis also has nerve endings that sense environmental stimuli such as heat, cold, pain and pressure. Sebaceous glands produces oils that regulate the skins softness and moisture while acting as a barrier against alien substances and pathogens and rendering the skin waterproof.

1.1.2. The epidermis

The dermis and epidermis are separated by a basement membrane made of a complex of glycoproteins and proteoglycans (Breitkreutz et al., 2013). This membrane extends to the outer regions of the skin derivatives such as hair. The epidermis is the outmost, multi-layered layer of the skin made of stratified squamous epithelia and serves as the first line of defence against UV radiation, physical, thermal, pathogen and toxic insults from the environment (Hirobe, 2014). Unlike the dermis, the epidermis is avascular. The epidermis mainly comprises keratinocytes but also has Langerhans cells and lymphocytes (immune cells), Merkel cells (touch sensitivity) and melanocytes (produces melanin) (Nakatani et al., 2015; Wickett and Visscher, 2006). Keratinocytes and keratinocyte differentiation will be discussed in section 1.1.2.1.

Langerhans cells are antigen presenting dendritic immune cells located in the basal layer of the epidermis and in epithelia of digestive, respiratory and urogenital tracts. They protect the epidermis from infection but also allow tolerance of commensal bacteria by the skin. These cells perform their function by acquiring antigens and presenting the antigens to naïve T-Cells in the lymph nodes for the initiation of an adaptive immune response (Chomiczewska et al., 2009).

Epidermal Merkel cells are non-neural cells found in the dermal/epidermal border of highly tactile regions (eg. fingertips) of the skin (Morrison et al., 2009; Nakatani et al., 2015). The exact function of Merkel cells is unknown but they are thought to be important in the detection of shape, texture and curvature along with the ability to discriminate between two distinct points touching the skin at the same time (Morrison et al., 2009).

Melanocytes are found in the basal layer of the skin (section 1.1.2.1) and are responsible for skin pigmentation through their production of melanin. These cells contain melanosomes which contain melanin which are transferred from the melanocytes to the keratinocytes resulting in energy absorption from the sun to protect against UV radiation and provide colour to the skin. The process of melanin production and melanosome transfer happens

concurrently with epidermal renewal but may be accelerated by excess sun exposure (Wickett and Visscher, 2006).

1.1.2.1. Structure of the epidermis and epidermal differentiation

The epidermis accounts for 3.7–16.8% and 40.6–44.6% of the dorsal and palmoplantar skin respectively (Lee and Hwang, 2002) and the epidermis is primarily composed of keratinocytes (Wickett and Visscher, 2006). Depending on gender, age and the skin's location on the body, the thickness of the epidermis varies between 30-105 μm for skin across the body and 601–637 μm on palmoplantar skin (Lee and Hwang, 2002; Saager et al., 2015; Sandby-Moller et al., 2003; Whitton and Everall, 1973). The epidermis is not flat but rather has rete ridges that interlock the epidermis to the dermis.

Keratins

Keratinocytes primarily produce keratin which is a major structural protein of the epidermis and aids in the barrier function of the epidermis by forming the structural framework of keratinocytes and other epithelial cells (Hirobe, 2014). There are 54 known functional keratin genes that are divided into two groups (Moll et al., 2008, Schweizer et al., 2006). The first group, type I, has 28 keratins (17 epithelial and 11 hair keratins) which are small, acidic (pH 4.5-4.5) keratins, whose genes are located on chromosome 17q21.2 (except keratin 18 (KRT18); located at 12q13.13) (Schweizer et al., 2006, Moll et al., 2008). The second group, type II, has 26 keratins (20 epithelial and 6 hair keratins) which are large, basic to neutral (pH 5.5-7.5) keratins whose genes are located on chromosome 12q13.13 (Moll et al., 2008, Schweizer et al., 2006). Keratin sizes range from 44-66 kDa (Moll et al., 2008). Epithelial keratins are listed in Table 1.1. Additionally, keratins form heterodimers composed of a combination of keratins from type I and II (1 type I/1 type II), to make keratin intermediate filaments (KIF), some specific to each epidermal layer (Moll et al., 2008). KIFs form tonofilaments that attach desmosomes (composed of desmosomal cadherins) or hemidesmosomes (composed of α6β4-integrin-containing) which are intercellular connections that connect keratinocytes to each other and to the basal membrane (hemidesmosomes) (Blanpain and Fuchs, 2006; Smith et al., 2004). The disturbance of the keratinisation process by altered expression keratins and the other genes involved in

keratinisation results in a variety of skin disorders. Alteration of keratinisation at any of the layers of the epidermis can cause a range of disorders including keratolytic winter erythema (KWE) which will be discussed in section 1.2.

Table 1.1: Human Type I and Type II epithelial keratins

Category	Number range	Location	pH
Human type I epithelial keratins	9-28	17q21.2	Acidic
Human type II epithelial keratins	1-8 and 71-80	12q13.13	Basic-neutral
Human type I epithelial keratins	KRT18	2q13.13	Acidic

Other regulators of epidermal differentiation

Environmental and intrinsic factors such as the innate immune system, a calcium gradient, steroid metabolism, growth factors and drug reactions affect differentiation (Banno and Blumenberg, 2014; Gilbert, 2000). Although keratins play a major role in differentiation and the formation of corneocytes, other groups of genes are also essential in differentiation and a few examples are given below.

The epidermal differentiation complex on chromosome 1q21 comprises: (1) cornified envelope precursors loricrin, involucrin, small proline-rich proteins (SPRR) and the late cornified envelope proteins (LCE); (2) calcium-binding proteins (S100) and (3) the S100 fused proteins (filaggrin, filaggrin-2, trichohyalin, trichohyalin-like protein, hornerin, repetin and cornulin) (Volz et al., 1993, Kypriotou et al, 2012). Desmosomes are essential intercellular junctions that allow keratinocytes to adhere to each other and the genes involved in maintaining desmosomes are also important in the regulation of keratinocyte differentiation. The cadherin, armadillo and plakin protein families are major constituents of desmosomes and mutations in the genes coding for these proteins have been shown to alter differentiation. The desmocollin and the desmoglein subfamilies form the cadherin family and are coded for by the desmocollins (*DSC*) and desmogleins (*DSG*) (Smith et al., 2004). Mutations in *DSG1* have been shown to cause autosomal dominant striate palmoplantar keratoderma (Rickman et al., 1999).

Transglutaminases are essential for cross linking proteins to make the cornified envelope (CE), and therefore its proper regulation is important in keratinisation (Nithya et al., 2015). Cathepsins which are found in lysomes and lamellar granules (cathepsin V) along with transglutaminases also play an important role in differentiation and the development of the cornefied envelope as they regulate transglutaminases (Eckhart et al., 2013). Some cathepsins activate key transglutaminases such as transglutaminases 3 (activated by cathepsin L) (Eckhart et al., 2013). Cathepsin D regulates the transglutaminase 1 which is important in forming the CE (Egberts et al., 2004). Calcium influx is also important in transglutamination as it activates the pathway (Eckhart et al., 2013).

Keratinisation of the epidermis

The epidermis has four distinct layers (five in palmoplantar skin) (Figure 1.2) with distinct keratinocyte morphology, gene expression, keratin and lipid composition (Nestle et al., 2009). The layers of the epidermis are the stratum basale, stratum spinosum, stratum granulosum and the stratum corneum (preceded by the stratum lucidum in palmoplantar skin) and these cells are connected to each other through desmosomes (calcium dependant cell adhesion glycoproteins) (King et al., 1991). The basal layer and the spinosum layers make up the Malpighian layer (Gilbert, 2000). The basal layer is connected to the basement membrane through the hemidesmosomes. The basement membrane allows nutrient diffusion from the dermis to the basal layer. The epidermis is constantly regenerating through a process of self-renewing and desquamation (shedding) and it does this through a balanced process of cell proliferation and cell death during a process known as keratinisation (Alonso and Fuchs, 2003). Keratinocytes are produced by the basal layer and differentiate through keratinisation to form the stratum corneum which comprises terminally differentiated dead keratinocytes also known as corneocytes that lack nuclei and cytoplasmic organelles (Bouwstra and Ponec, 2006). These corneocytes (dead keratinocytes) are the building blocks of the epidermis barrier (Eckhart et al., 2013).

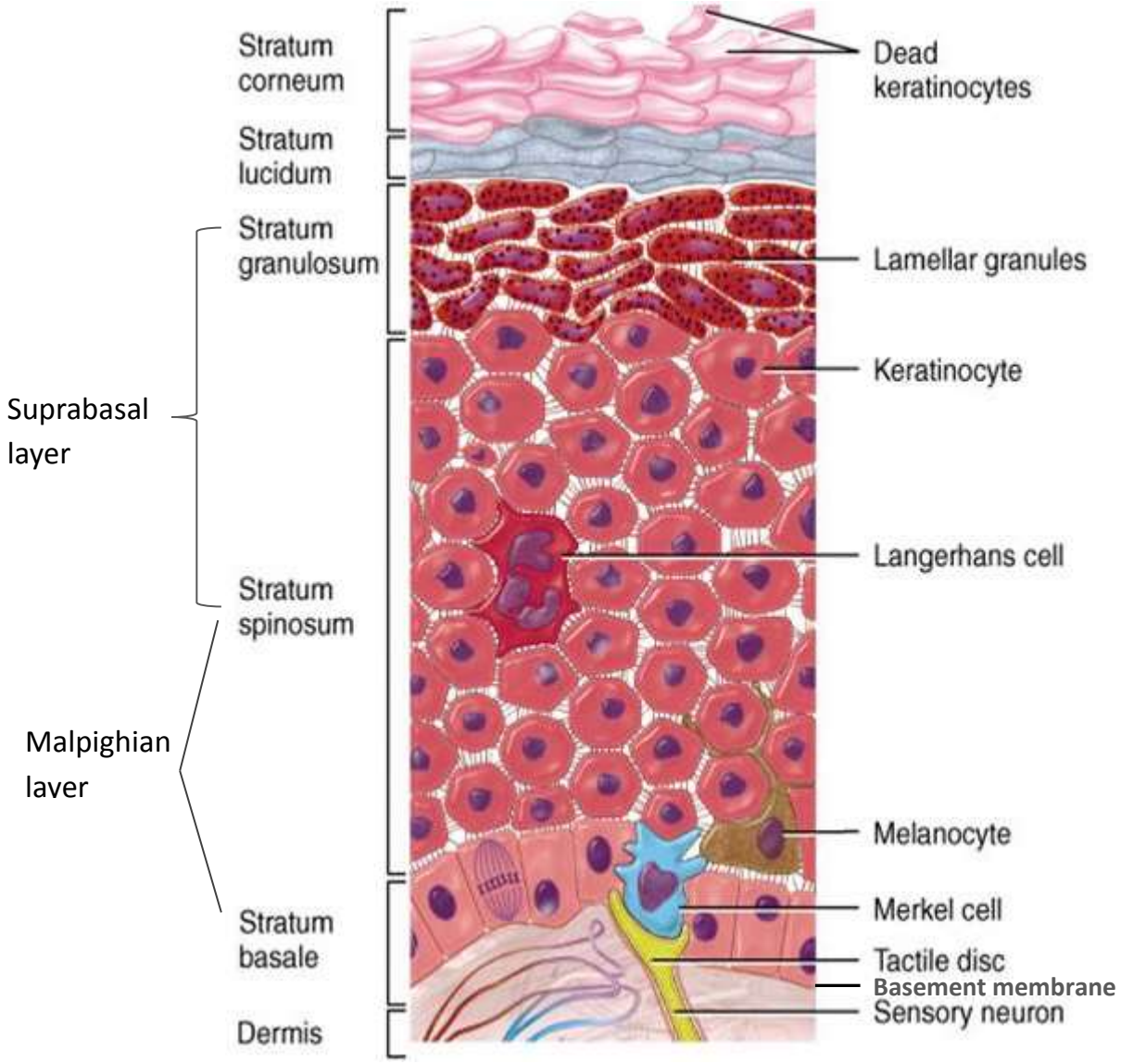


Figure 1.2: A schematic representation of the layers that make up epidermis in palmoplantar skin
The figure shows keratinocytes at different stages of differentiation and non-keratinocyte cells (Langerhans, Merkel and melanocytes) that overlap with the epidermis. Cells within the epidermis are joined together by desmosomes indicated as black lines connecting the cells. (From: http://www.pamperbay.co.uk/?p=54).

Stratum Basale

The stratum basale (basal or germinativum cell layer) is the innermost layer of the epidermis and contains multipotent keratinocyte stem cells and transient amplifying keratinocytes that give rise to the rest of the epidermis (Alonso and Fuchs, 2003; Banno and Blumenberg, 2014). The stem cells produce transforming growth factor-α (TGF-α) which

induces more cell division, producing basal daughter cells (Gilbert, 2000). The daughter cells either become stem cells or transiently amplifying keratinocytes which will commit to terminal differentiation (Alonso and Fuchs, 2003). Transient amplifying cells proliferate and after 3-6 rounds of mitosis, they withdraw from cell division to commit to cell differentiation by detaching from the basement membrane, initiating cell differentiation (Alonso and Fuchs, 2003). Additionally, integrin and laminin expression (forms part of the basement membrane) is inactivated and negative regulators of proliferation, lipid and steroid enzymes, are induced thereby initiating cell differentiation (Blanpain and Fuchs, 2006). The major keratins expressed in the basal epidermis layer are the K5 and K14 keratins, and K15 is the minor keratin (Blanpain and Fuchs, 2006). These keratins form a K5/K14 complex along with microtubules (tubulin) and microfilaments (actin), to form the keratinocytes cytoskeleton. The K5/K14 KIFs form tonofilaments that connect to α6β4-integrin-rich hemidesmosomes to anchor the basal layer to the laminin5-rich extracellular matrix of the basement membrane (Blanpain and Fuchs, 2006; Moll et al., 2008). K15 has also been shown to be expressed in the basal layer cells of young children (up to 1.5 years), and expression is decreased in adults, and it is localised to the rete ridges in adult skin (Moll et al., 2008; Pontiggia et al., 2008). As the basal keratinocytes move upward, the expression of K5, K14 and the integrin's reduce and these keratins are no longer expressed when the cells progress to the spinous layer cells (Fuchs and Green, 1980; Gilbert, 2000). These cells then migrate upward to become the new spinous layer of keratinocytes (Banno and Blumenberg, 2014).

Stratum spinosum

The cells in the 4-8 layer spinous layer appear prickly and form desmosomes that connect cells to each other (Fuch, 1990). These cells are no longer dividing but are metabolically active (Fuch, 1990). The cytoskeleton of the cells in this layer is formed by a complex containing K1/K10 KIFs which are in close proximity to the euchromatic nucleus and attached to the desmosomes. In palmoplantar skin, K2 forms a complex with K10 (K2/K10, expressed in upper stratum spinosum, stratum granulosum) (Moll et al., 2008; Wallace et al., 2012). Keratin 9 (K9) is also expressed in this layer with K9 forming dimers with K1 in palmoplantar skin (Moll et al., 2008; Wallace et al., 2012). The keratinocytes are anchored together by desmosomes and the cytoplasm also contains melanosomes. Additionally, these cells produce membrane coating proteins and involucrin (glutamine and lysine-rich envelope

protein incorporated into the inner surfaces plasma membranes) (Rice and Green, 1979 as cited by Fuchs, 1990). The involucrin is transported to the desmosmes where it binds to the plakins and initiates the formation of the CE (Kypriotou et al., 2012).

Stratum granulosum

As the cells in the basal layer proliferate, the cells in the spinous layer move upward to form the stratum granulosum (granular layer). These cells have a flattened appearance and undergo extensive cellular changes in preparation for forming the horny layer. The cells contain granules of keratin (Gilbert, 2000). The chromatin in the nuclei condenses then the nucleus disintegrates. The mitochondria, Golgi apparatus and ribosomes also disintegrate. Loricrin, a major structural protein of the CE is also synthesised in this superficial layer of the granular layer (Rice and Green, 1979 as cited by Fuchs, 1990). The cytoplasm is then occupied by lamellar granules (concentrated in the plasma membrane) and keratin filaments which become compact and associate with keratohyalin granules. The keratohyalin granules contain profilaggrin which is modified to filaggrin in the stratum corneum and loricrin cross-linked to involucrin (Nithya et al., 2015; Wickett and Visscher, 2006). The lamellar granules have an essential function in forming the epidermis permeability layer, allowing it to be waterproof. The lamellar granules primarily compact in the plasma with which they fuse to release hydrophobic glycophospholipid contents into the intercellular spaces within the granular layer and between the stratum granulosum and the stratum corneum. The formation of the cornified envelope starts in the granular layer due to the production of many of the CE's components in this layer (Nithya et al., 2015).

Stratum lucidum

The stratum lucidum, presents only in palmoplantar skin is an additional layer of clear, flat, dead keratinocytes filled with eleidin (a keratin intermediate). In palmoplantar skin, granular layer cells migrate to form the stratum lucidum which migrate to form the stratum corneum.

Stratum corneum

Granulated keratinocytes migrate upward to form the final product of terminal differentiation, the stratum corneum (horny layer). The stratum corneum is the outermost layer and functions as the barrier whereas the rest of the epidermis functions in generating and maintaining the stratum corneum (Bouwstra and Ponec, 2006). The keratinocytes in this layer, often referred to as corneocytes are large, flat cells filled with keratin (Nithya et al., 2015). They lack nuclei and organelles consist of keratin filaments embedded in a filaggrin (once cells reach the horny layer, profilaggrin is modified to make filaggrin) rich cytoplasmic matrix. Corneocyte plasma membranes appear much thicker due to the cross-linking by transglutaminination of several proteins from the epidermal differentiating complex (EDC) including involucrin (soluble), loricrin and small proline rich proteins (SPRPs) with the insoluble plasma membrane, the lipid and lamellae crosslinked to form the cornified envelope (Eckhart et al., 2013; Nithya et al., 2015; Wikramanayake et al., 2014). Transglutaminination is catalysed by transglutaminases 1,3 and 5 which are activated by a calcium influx (Eckhart et al., 2013; Nithya et al., 2015; Wikramanayake et al., 2014). Other major proteins in the CE include desmosomal proteins, KIFs, cystatin A and elafin (Nemes and Steinert, 1999; Nithya et al., 2015). Corneocytes outer membranes also have a layer of lipid which accounts for 15% of the cell content and comprised mainly of ceramides, cholesterol, and free fatty acids which all help build the physical barrier (Wilkes et al., 1973). The number of layers in the horny layer range from a few layers in trunk skin to 50 layers in palmar or plantar skin. The horny layer is constantly being pushed up by cells from the lower epidermal layer leading to the shedding of uppermost horny layer cells in a process known as desquamation.

The correct maintenance of the keratinocyte differentiation is crucial in the formation and maintenance of the stratum corneum. When this tight regulation is altered, keratinocyte related skin disorders may arise and these can affect the trunk skin and/or the palmoplantar skin. In this study, we discuss keratolytic winter erythema, a disorder where the formation pf the stratum corneum is altered leading to the disease phenotype.

1.2. Keratolytic Winter Erythema

Keratolytic Winter Erythema (KWE, OMIM: 148370) or Erythrokeratolysis hiemalis is a rare skin disorder of unknown aetiology. The disorder was first named and clinically defined by Findlay and colleagues (Findlay et al., 1977) and described in families residing in Oudtshoorn (Western Cape, South Africa, Figure 1.3), or whose roots could be traced to back to Oudsthoorn. KWE is therefore also known as "Oudtshoorn Skin" (Findlay and Morrison, 1978; Findlay et al., 1977; Hull, 1986). Although most of the patients were from Oudtshoorn, many now live across the country or have moved to other countries such as Zimbabwe, the United Kingdom and Canada (Hull, 2016, personal communication). KWE has also been described in Germany (Starfield et al., 1997) Denmark (Danielsen et al., 2001) the USA (in a family of Norwegian descent) (Huntington and Jassim, 2006) and in the UK (in a family linked to the South African pedigrees)(Amin et al., 2011).



Figure 1.3: Oudtshoorn shown on the map of Western Cape within the South Africa map
Oudtshoorn is shown as the yellow city on the Western Cape map (From http://www.drodd.com/html7/south-african-map.html)

1.2.1. Clinical characterization of KWE

KWE is characterized by intermittent and recurrent centrifugal peeling, with initial palmoplantar redness (erythema), keratin layer thickening and formation of superficial dry blisters where centrifugal skin peeling originates (Figure 1.4). The affected skin dissects off to form a peel which proceeds centrifugally to the major skin creases where a mild hyperkeratosis develops (Hull, 1986, Hull et al., 2013). Centrifugal peeling appears at multiple palmoplantar sites and is arrested at the major skin creases. The lesions start as red multi-forme like papules which enlarge and have red, sometimes thickened edges. Dry bulla which may be exudative and eventually dissect off from the centre outwards removing the stratum corneum as a though elastic scale that is adherent and not disturbed by friction but can be peeled off when moistened (Findlay and Morrison, 1978; Findlay et al., 1977). Once the stratum corneum peels off the palms and soles appear red and wrinkled.



a.



b.



c.

Figure 1.4: KWE patients showing the erythema and peeling of palmar and plantar surfaces and the web space

The peeling occurs on the palms (A), soles (B) and web spaces (C). (Photograph Credit: by Peter Hull).

The erythema and peeling may spread to the fingers/toes and web spaces, and occasionally dorsal regions may be affected. In all cases, the palms and soles are affected but in some cases, annular scaling erythema on the knees, forearms, hips, buttocks, legs and thighs, shoulders, spine and the dorsal have been reported (Findlay and Morrison, 1978; Findlay et al., 1977). A palmoplantar cold sweat with a distinct odour, accompanies aggravated peeling and occurs mainly in winter but may also be observed in the summer, or as a result of other triggers, for example excessive exposure to water or during infections (Findlay et al., 1977; Hull, 1986). Hyperkeratosis lasts for 6-8 weeks and itching at the affected regions has been reported in some cases (Findlay and Morrison, 1978; Findlay et al., 1977).

1.2.2. Nosology of KWE

There is no clear published systematic classification (nosology) of KWE, but expert dermatologists familiar with KWE have classed KWE as an erythrokeratoderma (Hull, 2016, personal communication). Erythrokeratodermas are a clinically and genetically heterogeneous group of disorders. This group of disorders has several overlapping features with KWE in that they rare inherited skin disorders, manifest at an early age and are characterised by well demarcated erythema and hyperkeratotic plaques and many show variable expressivity and incomplete penetrance (Hohl, 2000; Ishida-Yamamoto et al., 1997). As is the case for KWE, the cell nuclei are usually not fully disintegrated by the time the keratinocytes reach the final layer (see section 1.2.3) (Hohl, 2000; Ishida-Yamamoto et al., 1997). Symptoms may be influenced by environmental factors such as temperature, stress and mechanical pressure (Karadag et al., 2013; Mahajan et al., 2015). Initially, there were two main subtypes of erythrokeratodermas namely progressive symmetric erythrokeratoderma (PSEK) and erythrokeratoderma variabilis (EKV) (Hohl, 2000, Rogers, 2005) but these two disorders have now been consolidated as one disorder, erythrokeratodermia variabilis et progressiva (EKVP). This was mainly due to the fact that both disorders were identified in the same family and the people who had the different disorders shared common mutations (Richard et al., 2003; van Steensel et al., 2009). The plaques can either be migratory or stationary which led to initial characterisation of EKVP as two separate disorders, EKV and PSEK (Hohl, 2000; Ishida-Yamamoto et al., 1997). Most EKVPs are inherited in an autosomal dominant manner, but autosomal recessive cases (Gottfried et al., 2002) have also been reported. EKVP affects the trunk skin and

palmoplantar skin in some instances (Ishida-Yamamoto et al., 1997; Richard et al., 2000). EKVP has been shown to be caused by mutations in the gap junction protein alpha and beta genes (*GJA1*, *GJB2* and *GJB4*) (Gottfried et al., 2002; Richard et al., 2003; Richard et al., 2000; van Steensel et al., 2009; Wilgoss et al., 1999) and the loricrin gene (Ishida-Yamamoto et al., 1997).

1.2.3. Histological and microscopic findings

Haemotoxylin and Eosin (H&E) staining of affected skin clearly shows that the epidermis of affected skin is not normal as peeling and an abnormal parakeratotic layer were noted in the skin of affected individuals (Figure 1.5) (Hull et al., 2013). By evaluating 40 biopsies from 18 patients, Findlay and Morrsion (1978) conducted the first light microscopy evaluations to determine the pathogenesis of KWE at a cellular level. Two major processes were observed: (1) at the rear end of the lesion, necrobiosis of the Malphighian layer with the absence of a stratum granulosum occurs; and (2) the presence of spongiosis and vesicles outside the lesion.

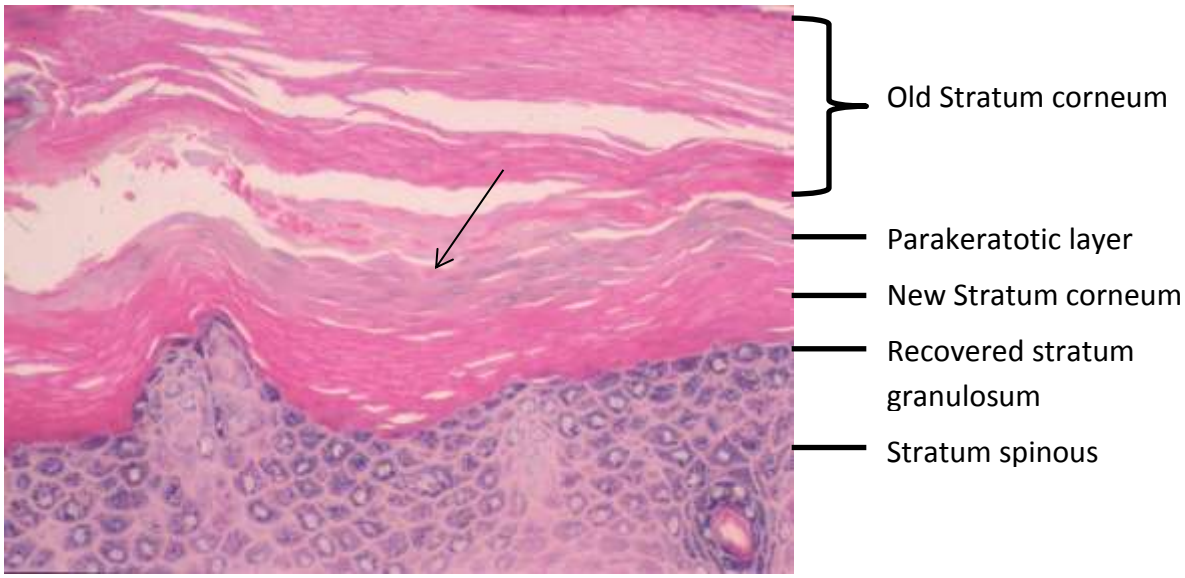


Figure 1.5: Haemotoxylin and Eosin (H&E) stained palm biopsy of a KWE patient. The layer of parakeratosis (indicated by the arrows) is sandwiched between the old orthokeratotic stratum corneum and the re-established orthokeratotic layer (Hull et al., 2013).

1.2.3.1. Light microscopy

In this section, the light microscopy findings by Findlay and Morrison (1978) will be summarized. The Malphighian layer does not appear normal as it appeared that the Malphighian thickness and rete ridge prolongation was significantly reduced when viewed under the microscope. The peeling of the skin is caused by cell death (necrobiosis) of some keratinocytes, the failure of keratinocytes in the Malphighian layer to undergo normal keratinization and the absence of an overlaying normal granular layer.

The granular layer thickens with the destruction of the Malphighian layer and the granular cells may undergo necrobiosis forming “granular cell ghosts”. These granular cell ghosts have a regular granular shape and size, but each lacks a nucleus and granules and forms a granular monolayer of granular ghost cells that occupy the stratum lucidum space. Necrobiosis is evidenced by visible nuclear changes including nuclear shrinkage and fragmentation which are associated with cell death.

The lesion spreads horizontally across the Malphighian layer and spongiosis and vesicles covered by a granular layer may also be observed just outside the lesion. A 2-8 cell layer parakeratotic lamina forms and this triggers epidermal recovery which results in the parakeratotic lamina being pushed out of the epidermis by the recovering layers below it. The recovery is achieved by basal cell proliferation, although maturation of these basal cells is halted due to the defects in the Malphighian layer. The basal cells often show spongiosis and necrobiosis and form block-like masses which engulf melanocytes. The parakeratotic layer then appears sandwiched between the original normal orthokeratotic stratum corneum and the new orthokeratotic stratum corneum below the parakeratotic layer. The split occurs above the newly formed stratum corneum and forms a substantial peel as shown in Figure 1.5 (Hull et al., 2013). As the lesion proceeds centrifugally, the centre of the lesion is restored.

1.2.3.2. Electron Microscopy

In this section, the electron microscopy findings by from Peter Hull's PhD thesis (1986) will be summarized. By evaluating biopsies from five patients, Hull noted a normal dermis and basement membrane. Hull concluded that there were no structural abnormalities, only the secondary changes discussed in this section including spongiosis, keratinocyte mitochondrial changes and the presence of perinuclear vacuoles.

Focal spongiosis (fluid accumulation) between the keratinocytes of affected individuals resulting in expanded intercellular spaces were observed. Cells outside these spongiotic areas appeared normal and seemed to be undergoing normal keratinisation. The spongiosis was most pronounced in the intact basal layer which remained intact with normal hemidesmosomes, and extended to the rest of the Malphighian layer and was less pronounced in the granular layer (Hull, 1986). The number of desmosomes was significantly decreased in the spongiotic spaces and desmosomes within the epidermis, connecting keratinocytes to each other, appeared stretched although they had normal morphology. Within the intercellular spongiotic space, desmosomes were often freely floating without any cellular attachment. The spongiotic areas also included membrane bound cytoplasmic material containing organelles but lacking the structural supportive tonofibrils. Langerhans cells within their dendrites along with Merkel cells were seen floating freely in the expanded intercellular spaces.

The keratinocytes also showed marked changes particularly in the mitochondria which showed intra-cristae swelling which is a result of a water influx into the outer chamber of the mitochondrial cristae. The movement of water from the mitochondrial matrix into the inter-membranous spaces resulted in a denser matrix and occasional inter-membranous blistering. Eventually progressive granular mitochondrial degeneration occurs, resulting in perinuclear lucid vacuoles. These vacuoles occupy a large section of the cytoplasm and they distort and displace the nuclei of the granular layer keratinocytes. These mitochondrial changes were only seen in keratinocytes as the mitochondria of the lymphocytes near the affected keratinocytes which are thought to be involved in the pathogenesis of the lesions, appeared normal. Similar mitochondrial changes in granular cells have been described in

progressive symmetric erthrokeratodermia which also has palmoplantar involvement (Nazzaro, 1986 as cited by Hull, 1986).

1.2.4. Age of onset

The age of onset varies from infancy to early adulthood and severity varies with the age of onset (Findlay et al., 1977; Hull, 1986). Most individuals show symptoms by the age of five and in severe cases, the disorder manifests in infancy (Hull, 1986). The condition improves with age and in adulthood, minimal or no scaling may be noted although there is considerable inter-individual variation (Findlay and Morrison, 1978; Findlay et al., 1977).

1.2.5. Aggravating factors and improvement

Symptoms often worsen in winter with the onset of the recurrent peeling generally starting with the onset of the cold season (March or April in South Africa) and subsiding as the weather becomes warmer (August – October, varies through the years based on the climate) (Findlay et al., 1977). In more severe cases, symptoms persist all year round (Findlay et al., 1977; Hull, 1986). Additionally, low humidity has also been associated with more severe lesions as individuals who have relocated to high humidity regions have reported improvement (Hull, 1986). Menstruation has also been reported as an aggravating factor in women whereas pregnancy drastically improves or eradicates symptoms for the duration of the pregnancy (Hull, 1986). Other aggravating factors include moisture, general anaesthesia, infections (particularly upper respiratory tract infections), antibiotics and topical steroids (Findlay and Morrison, 1978; Findlay et al., 1977; Hull, 1986; Ngcungcu et al., 2017).

1.2.6. Genetics of KWE

1.2.6.1. Mode of inheritance and penetrance

KWE is a highly penetrant, autosomal dominant disorder (Findlay and Morrison, 1978). The penetrance has been estimated from 85% (Findlay et al., 1977) to 92.6 % (Hull, 1986); although these figures may underestimate the true penetrance of KWE. Disease expressivity varies even within families and often subsides in adulthood. These factors may affect

penetrance if individuals are incorrectly diagnosed when the disease has improved or if individuals have very mild symptoms.

1.2.6.2. Prevalence

KWE is very rare across the world but relatively common in the white and Coloured Afrikaner populations of South Africa. Based on the known KWE families and the use of the 1980 South African census data, the prevalence of KWE was estimated at 1/13 307 among all South African whites, 1/7 208 among white Afrikaners (accounted for 54% of the white population at the time) and 1/90 483 among the Coloured population (Hull, 1986). KWE has been reported in a German family (Starfield et al., 1997), two Norwegian families (Ngcungcu et al., 2017), the USA (in a family of Norwegian descent) (Huntington and Jassim, 2006), in the UK (in a family linked to the South African pedigrees) and a spontaneous case in Denmark (Danielsen et al., 2001, Starfield et al., 1997, Hull, 1986). There is now an extensive African diaspora of cases across the world (Hull, personal communication).

Genetic drift by founder effect is the cause of the high prevalence of KWE in South African Afrikaners as all KWE families in South Africa can be traced back to a common ancestor, Captain Francois Renier Duminy, originally from France, who is thought to be the founder (Hull, 1986). Linkage studies in the German, Norwegian and Danish patients suggest that the causal mutations in these populations have a different origin from the South African founder mutation (Danielsen et al., 2001; Ngcungcu et al., 2017; Starfield et al., 1997).

1.2.7. Identification of the position of the KWE locus

Repetitive DNA, particularly short tandem repeat (STR) or microsatellites, are highly polymorphic markers dispersed across the genome that can be leveraged to identify regions linked to a disease by linkage analysis, thereby identifying the genetic region containing the causal mutation for a Mendelian disease. Linkage relies on the idea that markers and/or genes that lie in close proximity to each other on a chromosome will remain close to each other during meiosis without recombination happening between these markers (Pulst, 1999). Therefore, if a given microsatellite profile or haplotype is observed at a significantly

higher frequency in affected individuals, the disease mutation is close to those markers. A lod score (likelihood of linkage of given marker to disease) is calculated using different recombination fractions to get the maximum likelihood of linkage estimate for the recombination fraction at the highest lod score observed, with a lod score of 3.0 or higher considered as evidence for linkage and a lod score of -2 considered as evidence against linkage (Pulst, 1999). This analysis also takes into account the mode of inheritance and the penetrance of the trait. The same analysis can be performed using SNP markers.

Using microsatellite linkage analysis, the first major breakthrough in KWE genetics was the localisation of the KWE locus to the short arm of chromosome 8. Through linkage analysis, Starfield et al. (1997) the causative mutation was localised to chromosome 8p22-p23, between and including markers D8S550 and D8S552 (later refined to D8S550 and D8S1759 (Appel et al., 2002)), also known as the KWE critical region. The KWE critical region is a large (approximately 1.2 Mb), complex region with many genes. The region is flanked by two olfactory repeat regions which are prone to non-homologous recombination which results in chromosome rearrangements, including a common 4.7 Mb polymorphic 8p23 inversion (Giglio et al., 2002). The region has also been associated with certain forms of cancer such as prostate cancer (MacGrogan et al., 1994 as cited by Giglio et al., 2002). A common haplotype within the KWE critical region was identified in affected South African families, and therefore a single mutation is expected to be causal across all the South African families (Starfield et al., 1997). A different haplotype was identified within the same region in the German family suggesting tht the German mutations in the KWE critical region but is likely to be different from the South African mutation (Starfield et al., 1997).

1.2.8. Exclusion of genes in the KWE critical region

Since the localisation of the KWE mutation to the KWE critical region, a number of studies have interrogated the coding regions within the KWE critical region. Appel et al. (2002) investigated the KWE critical region by cloning and sequencing a large part of the region. Twelve transcripts were identified in the region using exome trapping but no pathogenic variants were identified in any of the coding regions (Appel et al., 2002).

Hobbs and colleagues (2012) several genes as candidate genes due to their roles in keratinocyte differentiation or maintenance of the skin's integrity. The coding regions, intron/exon junctions, and UTRs of cathepsin B (*CTSB*) and farnesyl-diphosphate farnesyltransferase (*FDFT1*), *C8orf49* and the deubiquitinating enzyme 3 (*DUB-3*) were sequenced (Hobbs, MSc dissertation 2008). Additionally, gene expression studies of two of the most biologically relevant genes within the region, *CTSB* and *FDFT1,* were done (Hobbs et al., 2012). No causal mutation was identified in any of the sequenced genes and these genes were excluded as candidate genes for KWE. After analysing gene expression levels, *FDFT1* showed higher expression levels in affected individuals, although this result was not significant (Hobbs et al., 2012).

Using a custom tiling array covering the KWE critical region, Hull et al. interrogated the CNVs in the critical region using 385,000 probes to cover the region (Hull et al., 2013). No CNVs were found for any of the genes within the region and several small non-coding CNVs were identified in the region but none of the identified CNVs appeared to have a role in KWE (Hull et al., 2013). Interestingly, a duplication upstream the *CTSB* gene was found in 3/5 patients but not in any of the controls. The authors noted that the variant may regulate *CTSB* expression or may also affect other genes in the region. Another CNV upstream *DUB3* was identified in 3/5 patients and was thought to regulate *DUB3*. The authors did not pursue these two variants any further after the study.

The KWE mutation therefore remained elusive, even after the interrogation of the most likely genes in the region and after assessing the CNVs in the region. The failure to identify the KWE causal mutation was due to a number of reasons including the fact that it could have been in a non-coding region, since this was not investigated, and it could have been missed due to technical factors which meant that it was not detected in all affected individuals. Alternatively, although very unlikely (high linkage score in the linkage study (Starfield et al., 1997), the mutation may not within KWE critical region. KWE shows incomplete penetrance and therefore some patients may have been designated as not having the disease which would lead to the exclusion of mutations found in these individuals. Although CNV analysis and expression studies have been carried out, the

methods that were used to interrogate the KWE critical region and gene expression may have not been sensitive enough to detect the mutation.

1.3. Alternative mutation finding strategies

Since previous attempts to identify the mutation responsible for KWE were unsuccessful and included candidate gene sequencing of a single or a few genes, a new approach was required. With advances in technology, we are now able to sequence chromosome regions, whole exomes and even whole genomes using next generation sequencing (NGS). NGS has become an essential tool in disease gene identification as it allows for massively parallel sequencing at high speed and a lower per base cost (Ari and Arikan, 2016). Whole genome sequencing, targeted capture resequencing and exome sequencing strategies have been used in several studies to determine the causative mutations for a range of Mendelian disorders, mainly for autosomal recessive traits, but also for inherited autosomal dominant disorders. Most of the causal variants discovered using NGS have been small variants within the coding region of a gene. With the appropriate bioinformatics analysis, NGS sequencing can successfully detect small variants as well as large structural variants, including copy number variants.

1.3.1. Copy number variation

Copy number variants (CNVs) are defined as DNA segments that are 1 kb or more and are present in variable copy numbers when compared to the reference genome (Redon et al., 2006). They are spread across the genome, accounts for about 12% of the human genome, and have been shown to be associated with at least 3000 genes (Kehrer-Sawatzki, 2007). CNVs are found in both gene coding and non-coding regions of the genome, including regions that overlap with regulatory elements. They contribute to genomic diversity, which can result in phenotypic diversity (Kehrer-Sawatzki, 2007). CNVs may have dramatic phenotypic consequences if they change gene dosage, disrupt genes to make them non-functional or interfere with long range gene regulation, thus influencing gene expression. For instance, if a gene suppressor is deleted or an activator binding site is duplicated, then the transcription of the gene regulated by these sites may increase, whereas if an activator

binding site is deleted then the gene may not be expressed or its transcription will be reduced, depending on the importance of the activator in initiation of transcription.

Like all other genetic variants, CNVs may have medical significance with 14.5% of all characterised CNVs overlapping with OMIM morbid genes (Redon et al., 2006). CNVs have been found to be polymorphic in breakpoints of DiGeorge and Williams Beuren syndromes (Redon et al., 2006). CNVs have also been implicated in microdeletion and duplication disorders and have been shown to confer risk to complex disease traits such as HIV infection, where low copy number of the *CCL3L1* gene makes the person more susceptible to HIV, and the *DEFB4* gene that predisposes individuals to Crohn's disease (Kehrer-Sawatzki, 2007; Redon et al., 2006).

CNVs may also overlap with regulatory elements such as enhancers and repressors which are often found in the non-coding regions of the genome. Copy number variation of these elements has been shown to alter phenotypes. Duplications of non-coding regulatory regions are an uncommon disease mechanism but enhancer duplications have been previously described, primarily in limb malformation disorders and some sex reversal phenotypes (Dathe et al., 2009; Klopocki et al., 2011; Lohan et al., 2014). Polydactylies and syndactylies, including Hass-type polysyndactyly (HTS) (larger duplication), Laurin-Sandrow syndrome (LSS) and brachydactyly type A2 (BDA2), are caused by duplications of enhancers in noncoding regions (Dathe et al., 2009; Lohan et al., 2014).

SNP genotyping arrays and comparative genomic hybridization arrays can be used for CNV detection. NGS can also be used to detect CNVs but because NGS methods have short reads (75-150 bp depending on the assay) the usual analysis strategies may not be able to easily detect large CNV. NGS can however be used to detect evidence of CNV through split reads, for example where a read has two parts of the genome close to each other that are usually far apart in the genome suggesting the presence of a deletion or duplication. Conventional NGS analysis tools such as GATK may miss CNVs so alternative alignment and variant calling tools should be considered when looking for CNVs.

1.3.2. Targeted capture resequencing of the KWE critical region

To date, all studies that have searched for the KWE mutation have focused on genes or coding regions (exons) of genes within the KWE critical region and all have failed to identify the mutation that causes KWE. The KWE mutation may therefore, not be in a gene or coding region, but in a non-coding region such as an intron or an intergenic region. By sequencing the entire KWE critical region on chromosome 8 (8p23.1-p22) using target-capture of the KWE critical region, all DNA sequence differences between affected and unaffected individuals could be identified and all variants within the KWE critical region including single nucleotide variants (SNVs), insertions and deletions (indels) and CNVs could be detected. Mutations within coding, regulatory, intronic and intergenic regions will be detected and can be compared between affected and unaffected individuals to identify mutations unique to KWE affected individuals. By sequencing the entire KWE critical region, the KWE causal mutation should be identified if it is indeed in this region and is detectable given the limitations of the approach. Whole genome sequence approaches should also detect the same mutations in this region but performing targeted capture resequencing of the KWE critical region is more cost effect and would yield more accurate due to the high coverage that is possible with targeted capture resequencing (Dapprich et al., 2016).

1.3.3. Exome sequencing

A vast majority of all pathogenic mutations in monogenic disorders are located in the exons (Takeichi et al., 2013). The first exome sequencing study used to identify the causal mutation for a Mendelian disorder was published in 2010. In this study, mutations in the *DHODH* gene were shown to be responsible for Miller syndrome (Ng et al., 2010). Exome sequencing has identified causal mutations for several skin disorders that are caused by de novo and recessive mutations (Takeichi et al., 2013). Through several studies, exome sequencing has also been shown to be very successful in identifying causal mutations for dominant traits. Although exome sequencing has been successful in identifying disease genes in inherited disorders, the identification procedures and analyses are also complicated and as a result, pathogenic variants may be missed in some studies. Examples of the successful identification of causal genes for dominant traits include the identification of mutations in the *CARD14* gene responsible for autosomal dominant psoriasis susceptibility locus 2 (PSORS2) (Jordan et al., 2012) and the identification of *LMX1B* mutation involved in focal segmental

glomerulosclerosis (FSGS, also autosomal dominant) (Boyer et al., 2013). Exome sequencing uses a capture design which allows for sequencing of exons and 3' and 5' untranslated regions (UTRs) of all known genes. Some genes are regulatory genes (e.g. transcription factors) and mutations in these regulatory genes may affect regulation of target genes.

1.4. Genome structure and gene regulation

The genome is a complex structure regulated by chromatin folding and several epigenetic mechanisms including histone modifications, DNA methylation and micro-RNAs (Costa, 2008; Steiner et al., 2016). It is comprised of DNA wrapped around histones forming the nucleosomes that are densely compacted in a tightly regulated spatial manner to form chromatin. The nucleosome is made of 147bp DNA wrapped over 1.65 turns around nucleosomes (four pairs of core histones) (Luger et al., 1997; Radman-Livaja and Rando, 2010). The core histones, namely H2A, H2B, H3 and H4 (Lu et al., 2006), are charged. H2A and H2B have positive charges (due to lysines at the C-terminus) and the other two have negative charges (Latchman, 2005). These charges determine how tightly bound the DNA is to the histones and since DNA is negatively charged, it associates with the positively charged histones. Each of the core histones has 2 subunits (Lu et al., 2006). The H1 histones maintain the structure of the nucleosome by clamping DNA to the core proteins to prevent DNA movement away from the core histones (Luger et al., 1997). Several histone modifications influence chromatin conformation and the regulation of gene expression. These alterations determine the level of DNA binding to the histones (Lu et al., 2006). When the DNA is free from the nucleosomes and in a more open configuration, it becomes available for binding of transcription factors and hence gene transcription.

1.4.1. Histone modifications

Several mechanisms that alter histones and influence gene regulation are acetylation, methylation and phosphorylation (Lu et al., 2006), and have been described and evaluated in the ENCODE project (Table 1.2). These histone modifications and the DNA it is associated with can be detected using Chromatin Immunoprecipitation Sequencing (ChIP-Seq) which uses an antibody specific to that histone modification to detect the DNA fragments that associate with that particular histone modification (Landt et al., 2012). The modifications can

be used to determine the chromatin state and the presence of regulatory elements including enhancers, insulators, promoters, gene bodies and inactive sites, just to name a few. Only histone acetylation will be discussed because it will be referred to again in the subsequent chapters.

Table 1.2: Histone modifications and variations used in the ENCODE project to access the genome for regulatory elements*

Histone modification	Putative functions
H2A.Z	Histone protein variant (H2A.Z) associated with regulatory elements with dynamic chromatin
H3K4me1	Mark of regulatory elements associated with enhancers and other distal elements, but also enriched downstream of transcription starts
H3K4me2	Mark of regulatory elements associated with promoters and enhancers
H3K4me3	Mark of regulatory elements primarily associated with promoters/transcription starts
H3K9ac	Mark of active regulatory elements with preference for promoters
H3K9me1	Preference for the 5' end of genes
H3K9me3	Repressive mark associated with constitutive heterochromatin and repetitive elements
H3K27ac	Mark of active regulatory elements; may distinguish active enhancers and promoters from their inactive counterparts
H3K27me3	Repressive mark established by polycomb complex activity associated with repressive domains and silent developmental genes
H3K36me3	Elongation mark associated with transcribed portions of genes, with preference for 3' regions after intron 1
H3K79me2	Transcription-associated mark, with preference for 5' end of genes
H4K20me1	Preference for 5' end of genes

*(ENCODE Project Consortium, 2012)

Histone acetyltransferases (HATs) and histone de-acetyltransferases (HDACs) are histone acetylation and deacetylation enzymes, respectively, that are involved in nucleosome folding and determine gene expression by exposing or compacting DNA (Mercurio et al., 2010). HATs acetylate the lysine groups in positively charged histones to reduce the positive charge in the protein thus reducing the association between the positively charged protein and negatively charged DNA (Latchman, 2005). This mechanism also decreases the binding affinity of H1 allowing the DNA to dissociate from the histones to allow transcription factors to bind to allow gene expression (Latchman, 2005). This modification is associated with open

chromatin near regulatory regions, including active enhancers. Histone mark, H3K27ac associates with active enhancer elements and this modification is often cell specific (Creyghton et al., 2010). Histone acetylation can be reversed by HDACs which remove the acetyl group, increasing the positive charge of the protein thus attracting the DNA back to the histones and inhibiting gene expression (Latchman, 2005). Epigenetic errors can occur during histone modification and as a result can lead to the activation or suppression of gene expression depending on the nature and histone modifying enzyme causing the errors. Errors could happen during histone acetylation and histone deacetylation with the HAT and HDAC enzymes regulating the mechanisms, respectively (Mercurio et al., 2010).

1.4.2. Gene regulation through chromatin structure

Gene expression is highly regulated by chromatin conformation and looping which determines how parts of the genome are accessible to each other and are therefore able to interact. This is achieved through highly conserved topologically associating domains (TADs) which can be identified using different chromatin conformation capture technologies such as 4C, 5C, Hi-C, CaptureC and ChIA-PET (Lupiáñez et al., 2016). TADs are defined as linear, discrete genomic units that loop into three-dimensional (3D) structures that bring together regions of the genome which are usually far from each other and favours chromatin interactions internally within the loop and discourage interactions with regions outside the loop (Ciabrelli and Cavalli, 2015; Lupiáñez et al., 2016). The looping allows for protein mediated interactions between enhancers and promoters to occur, thus allowing for the activation of transcription (Lupiáñez et al., 2016). The boundaries of TADs have insulators (Ciabrelli and Cavalli, 2015) and are demarcated are by CCTC-binding factor (CTCF) binding. CTCF is a multifunctional protein essential in gene regulation that acts as a transcription regulator by acting on promoters and facilitating long range chromatin interactions (Steiner et al., 2016). When TADs are altered, gene regulation may also be altered as enhancers can either be blocked off from their target genes or be placed in the close proximity to non-target genes resulting in mis-expression or enhancer adoption (Lupiáñez et al., 2015; Spielmann and Mundlos, 2013).

1.5. Aim and Objectives

The aim of this study was to identify and characterize the causal mutation for keratolytic winter erythema (KWE) in South African families.

Objectives

1. To use a NGS approach to search for the KWE mutation: target-capture followed by NGS of the 8p KWE critical region applying a variant filtering process for possible candidate causative mutations including the following criteria: mode of inheritance, segregation with affection status, mutation frequency, functional impact or biological relevance, and modelling for incomplete penetrance.
2. To validate potential causal variants by PCR and Sanger sequencing in additional KWE families.
3. To perform functional studies to determine the biological impact of validated potential causal variants in skin tissue from cases and unaffected controls.

Chapter 2

Materials and Methods

2.1. Participants

A total of 167 samples from affected and control Afrikaner individuals were included in this study and the disease status of these individuals is summarized in Table 2.1. Forty-two individuals, including 23 affected individuals and 19 unaffected individuals, from three KWE families (A, B and C, Figure 2.1) and a singleton from a different KWE family (Family E: III-4, Figure 2.2) were sequenced. An additional seven random white Afrikaners, not related to individuals in the study were included in the targeted resequencing as unrelated, ethnically matched controls (samples provided by the NHLS, Division of Human Genetics). For the validation studies, the 49 sequenced individuals (including the 7 random controls) and 23 additional KWE family members from 4 families (11 affected and 12 unaffected individuals from families D, E, F and G, Figure 2.2) and 89 unaffected and unrelated ethnically matched controls were used. A nuclear family from a larger KWE family comprising of an affected mother and her two sons was used for the functional experiments (Family H: I-2, II-1 and II-2, Figure 2.2).

Table 2.1: Participants used in the study shown according to the different experiments

KWE Family		Affected	Not Affected	8p23.1-22 sequencing	Validation	Functional studies
A, B and C		23	19	All	All	None
D, E, F and G		11	12	None	All	None
H		3	0	None	2	All
Controls	DNA only	0	89	None	All	None
	Biopsy only	3	0	None	None	Yes

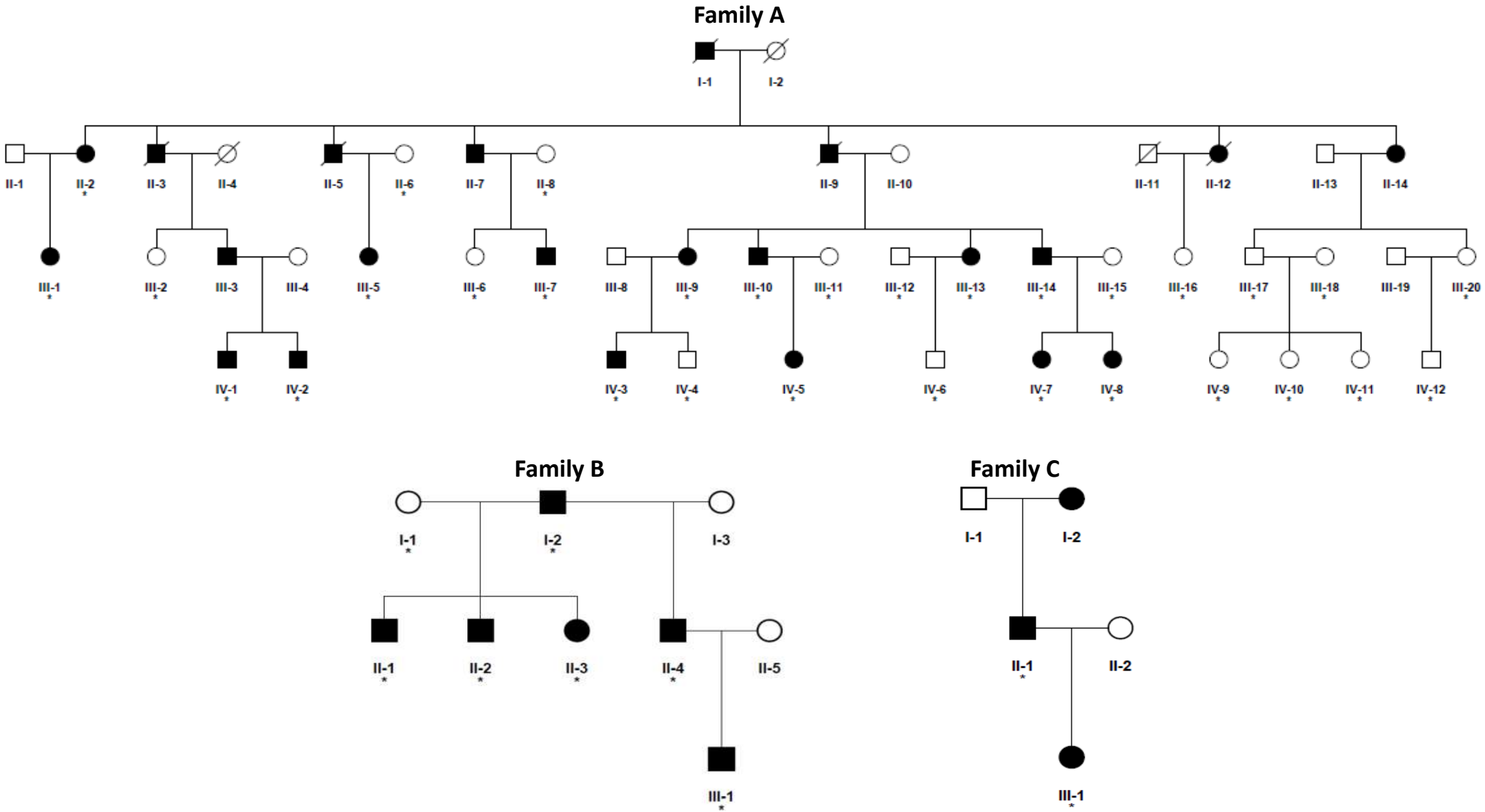


Figure 2.1: Pedigrees of South African KWE families used for next generation sequencing
The individuals whose samples were subjected to NGS sequencing are marked with an asterisk (*).

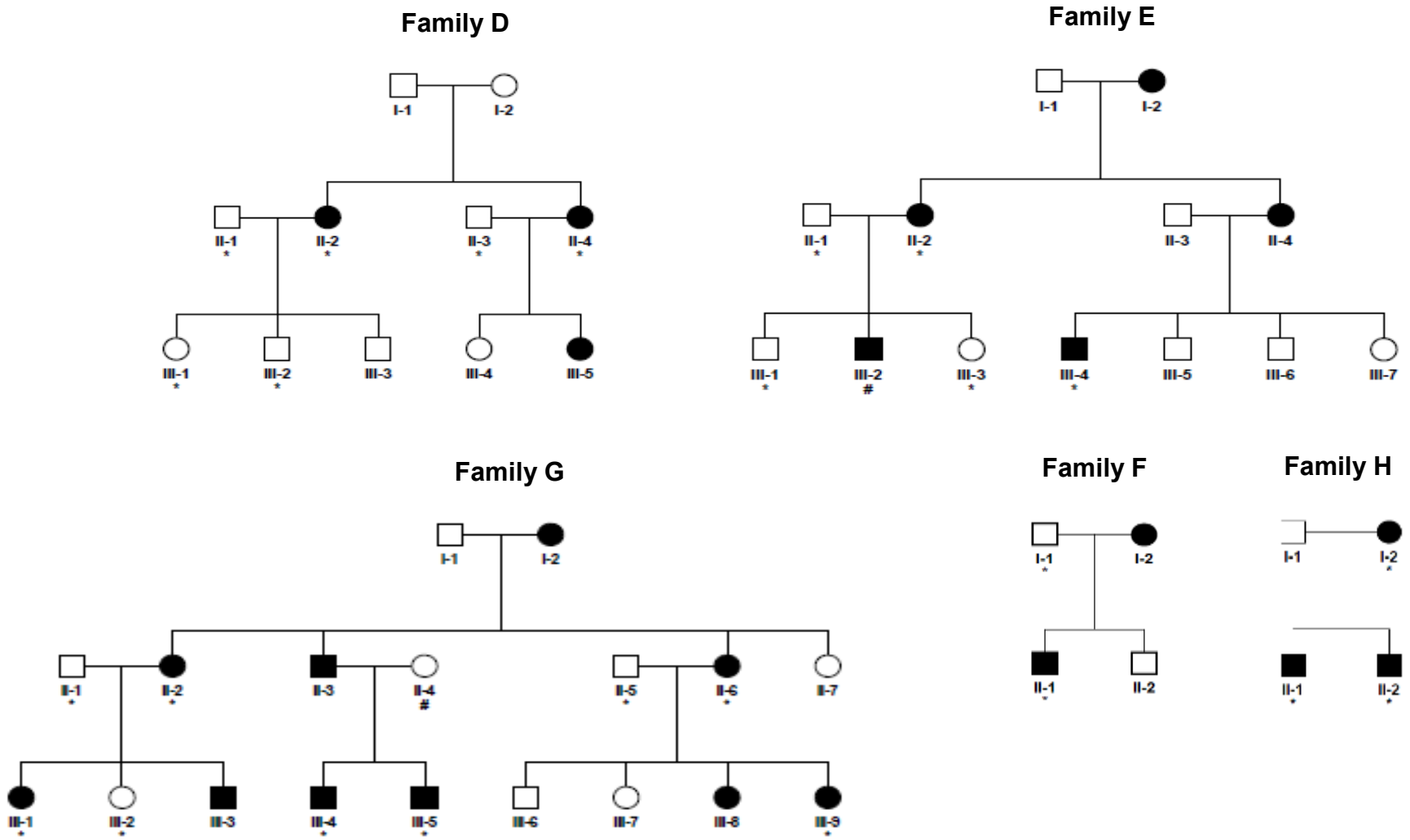


Figure 2.2: Pedigrees used for the validation of potential causal variants identified using NGS and for functional analysis
The individuals whose samples were used for validation are marked by an asterisk (*). Individuals marked with a hash tag (#) indicate individuals that were included in NGS sequencing. Family H was also used for the functional studies.

All DNA samples (except Family H) and phenotype data were collected with informed consent prior to the present study, as part of the ongoing studies (1985-2007) in the Division of Human Genetics (HREC Protocol numbers: M050706 and M070423). Next generation sequencing was carried out under an existing protocol approved by the HREC (Protocol number: M070423) but amended and extended to allow for next generation sequencing in Switzerland. The use of existing data/samples and the collection of new data/ samples (Family H) for this PhD project were approved by the HREC (Protocol number: M140530, Appendix A). Export permits were granted by the South African Department of Health to transport samples to Novartis (Basel, Switzerland) for sequencing, Haukeland University Hospital (Bergen, Norway) and Radboud University Medical Center (Nijmegen, The Netherlands) for qPCR and immunohistochemistry, respectively. Materials transfer agreements were signed by both the senders and recipients.

2.2. Laboratory methods

The project had three major methodological components, namely: next generation sequencing; analysis and validation of NGS results; *in silico* and *in vitro* functional analysis. Figure 2.3 summarizes the different experiments conducted and the analyses performed.

2.2.1. DNA extraction

The DNA used for the NGS portion of the study was already available for this study, it had been extracted from blood using the salting out method (Miller et al., 1988). The DNA from the two affected sons from family H was extracted from brush buccal swab samples using the Gentra Puregene DNA purification kit (Appendix B). DNA quality was determined by running DNA samples on a 0.8% agarose gel and all degraded samples were excluded from the study. DNA was quantified using the Nanodrop® ND-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and normalized to 100 ng/µl with 1X Tris Ethylenediaminetetraacetic acid **(**TE) buffer (salting out DNA) or Hydration buffer (buccal swab samples) and stored at 4°C. Additionally, the DNA used for the NGS experiments was also quantified using the Qubit to determine the concentration of double stranded DNA and normalised to 1µg (targeted capture) using TE buffer.

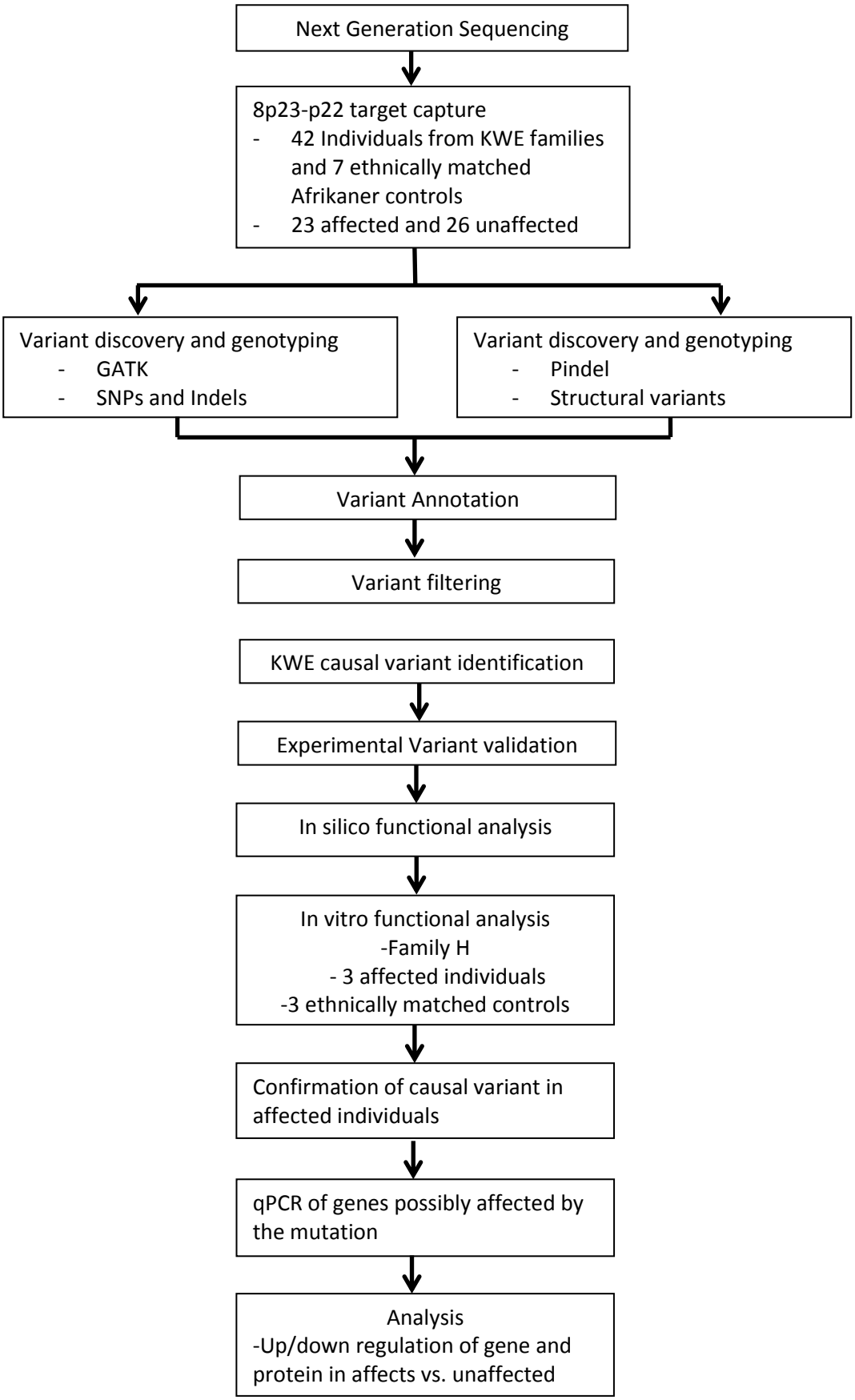


Figure 2.3: High level view of the workflow followed to identify, validate and characterize the KWE mutation.

2.2.2. Next generation sequencing using targeted resequencing

A microgram (1µg) of DNA was fragmented by shearing using the single tube method on the Covaris E210 Water Bath Sonicator (Covaris Inc., Woburn, Massachusetts, USA). Fragmented DNA was anayled using the Agilent Bioanlyser to determine fragment size (target ~700bp, sample was repeated if fragments were significantly smaller). The resulting fragments were repaired to remove overhangs to produce blunt 5' ends and the 3' ends of the fragments were adenylated (single A). A ligation step ligated adapter sequences to the ends of each fragment which enabled the DNA to bind to the sequencing flow cell. The adapter sequences also contained bases complementary to the sequencing primers used in the sequencing step. Index sequences were also ligated and used as identifiers for each sample. A number of clean-up steps removed non-ligated fragments which were then enriched for by PCR and followed by appropriate clean-up steps as per protocol. In this reaction, the PCR primers annealed to the ends of the adapters on either side of the ligated fragments and only fragments with adapters on both sides will be amplified. The products were used to assess the quality of the fragments by means of the appropriate Agilent DNA chips. Where the ligated libraries met quality control measures, libraries were stored for the KWE critical region DNA capture.

To enrich specifically for the KWE critical region, DNA libraries were prepared with the Illumina TruSeq DNA Sample preparation kit v.2 (#FC-121-2001, Illumina Inc.,San Diego, CA, USA) and the target region (KWE critical region: hg19 build, chr8:11 477 641-12 742 458) was captured following the NimbleGen SeqCap EZ SR protocol (#6266304001, Roche NimbleGen, Madison, WI, USA). The NimbleGen SeqCap EZ library was designed to have primers that span the KWE critical region. The kit contained biotinylated long oligonucleotide probes in redundant tiling manner with up to 38X base coverage specific to the target region. The libraries were sequenced on the Illumina HiSeq 2000 following a 101 bp paired-end sequencing protocol. Libraries were pooled so that each lane had eight libraries from eight samples.

2.3. Processing of NGS data

Illumina sequencers generate the sequencing output as per-cycle BCL base call files that are not compatible with many downstream processing protocols. These BCL files were combined and converted into FASTQ files which are compatible with downstream processing using bcl2fastq.

The sequencing data generation and quality control (QC) check was conducted on FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to generate a QC report. FastQC assesses the sequence quality, read length, GC content, overexpressed sequences and sequence duplication. Runs were considered to be of good quality when the 25th percentile Phred score (which determines the probability of each base being called correctly) of all bases was above Q26. If the 25th percentile Phred score for more than one third of the bases were below Q20 then the data were considered to be of low quality, were analysed with caution, and flagged for a repeat experiment to generate new data. The flow cell quality traces for “per base sequence quality” are shown in Appendix C.

2.4. Small variant calls using the GATK pipeline

If the run quality was acceptable, the raw sequencing data was processed to prepare a data set for analysis (Figure 2.4). This was done in a preprocessing step, followed by a variant discovery step. Picard tools and the Genome Analysis Toolkit (GATK) (DePristo et al., 2011; McKenna et al., 2010) were used to process the data except where noted (http://www.broadinstitute.org/gatk/guide/topic?name=best-practices).

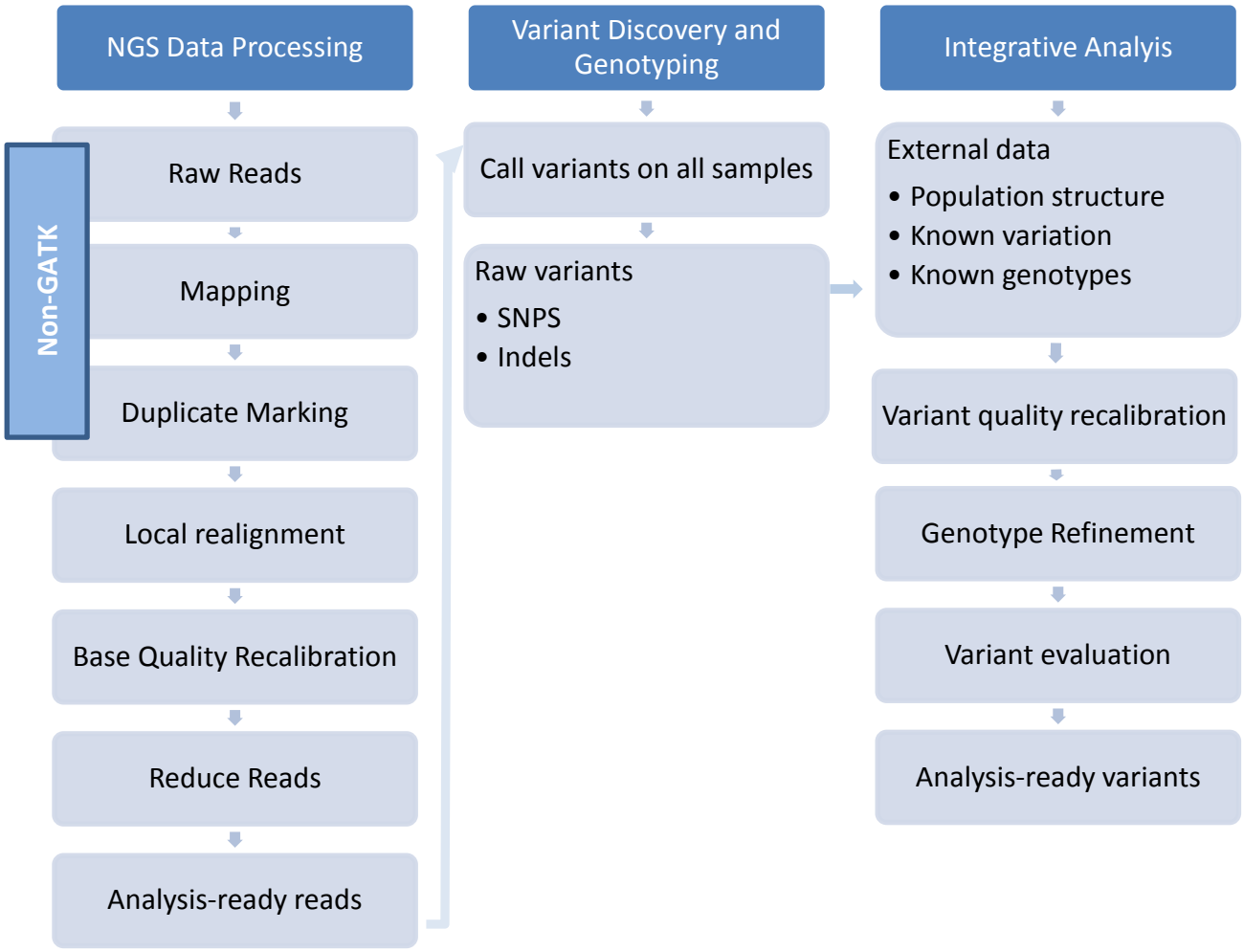


Figure 2.4: GATK workflow for data NGS data processing and small variant calling. (Adapted from http://gatkforums.broadinstitute.org/gatk/discussion/1186/old-version-of-best-practices-from-gatk-2-0-retired)

2.4.1. Pre-processing

The sequencing output results in millions of reads that are in FASTQ format that have Illumina adapter sequences at the 3' and 5' end and indexes to identify the reads. Some of these reads are of low quality. The adaptors and the low quality reads were removed using FASTQC and the resulting files were in FastQC format. The reads were mapped to the human genome (hg19) using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). The BAM files created by BWA were then sorted and indexed using SAMtools (Li et al., 2009). Duplicated reads were identified, marked as duplicates and removed using Picard tools (https://broadinstitute.github.io/picard/). The resulting de-duplicated reads were also sorted and indexed. All subsequent processes were done on GATK. Using the IndelTargetCreator, possible true indels assigned as mismatches by BWA were identified and

re-aligned using the IndelRealigner. This was followed by the recalibration and scoring of variants by the BaseRecalibrator. The final output of the pre-processing steps was analysis-ready reads in the form of recalibrated BAM files that were interrogated in the variant discovery process.

2.4.2. Variant calling

The GATK UnifiedGenotyper module was used for variant calling to identify sites that were different from the reference sequence thereby identifying single nucleotide variants (SNVs) and indels. A minimum base quality score of 20 was used as a cut-off. The UnifiedGenotyper module uses a Bayesian genotype likelihood model to estimate the most likely genotypes and allele frequencies for each position in the sequenced samples. The module then outputs a genotype for each base for every sample or only outputs the variant sites, based on the user preferences. The module uses analysis-ready reads from the pre-processing steps and outputs raw, unfiltered, highly sensitive variants. Variant quality recalibration uses machine learning to determine good and bad variants then annotates genotypes taking into account external data such as known genotypes (e.g. 1000 Genomes Project data (The 1000 Genomes Project Consortium, 2015)) and rescores genotypes based on read coverage and orientation producing a recalibrated VCF file. The VCF file was then processed for genotype refinement to improve genotype call accuracy and filter out low confidence genotype calls resulting in a refined VCF file containing analysis ready variants. VCFs were loaded onto the Vertica Analytic Database which is a commercial structured query language (SQL) relational database system (Lamb et al., 2012). All the variant filtering was performed using SQL queries on this database and additional queries were run on Postgres (open source) when Vertica was not available.

2.4.3. Variant annotation

Known and novel variant status was assigned based on comparison to publicly available data (such as dbSNP, the 1000 Genomes Project, and NHLBI Exome Sequencing Project (ESP)) and in-house Novartis sequence data. The Ensembl Variant Effect Predictor (VEP) tool was used to determine the position of each variant in exons, introns, intergenic regions or within regulatory regions (McLaren et al., 2016). This tool assesses potential functional impact (e.g.

deleterious variants) or the biological relevance of the mutations. Non-synonymous variants, synonymous, splice variants and regulatory variants were prioritized as potential causal variants.

2.4.4. Variant Filtering

The following criteria were applied to identify the causal variant: (*1*) within the KWE critical region; (*2*) segregate with disease across all three families following an autosomal dominant mode of inheritance; (*3*) novel; and (*4*) alter a protein's amino acid sequence and be likely to disrupt the function of that protein. Incomplete penetrance was accounted for in step 2, and possible causal variants were considered even if they were present in a maximum of two unaffected family members with an affected parent, child and/or sibling, assuming 90% penetrance. Variants present in a minimum of 80% of the affected group but not in the unaffected group were also considered, accounting for possible incorrect affected status. Non-coding variants were also considered, following criteria 1-3.

2.5. Structural variant calling using Pindel

BWA was used to align the reads using the "mem" method. The BAM files resulting from the alignment were passed through the version of Pindel (Ye et al., 2009) downloaded from the Pindel github site (https://github.com/genome/pindel) (Appendix D). The results were parsed into individual components for each structural variant. These components were loaded into a relational database which included de-identified information about each participant. Simple SQL queries were used against the database that counted the number of times each variant occurred, grouped by disease status. For variant filtering, the sample filtrating protocol as outlined in section 2.4.4 was used.

2.6. Validation of potential causal variants

Sequence data interpretation may lead to incorrect genotyping calls and therefore as an additional quality control step, the potential causal variant was validated in all sequenced individuals (n=49, shown in Figure 2.2 and including seven unrelated ethnically matched controls). An additional four KWE families (including 11 affected and 12 unaffected

individuals) and 89 ethnically matched controls (random Afrikaners with no history of KWE) were used for validation. Samples used in the functional studies were also validated for the causal variant (Figure 2.2). Because the disorder is an autosomal dominant disorder, one variant that is unique to all affected individuals was expected to remain after validation. A tandem duplication that segregates with KWE was validated using PCR and Sanger sequencing. Primer design, PCR essay and condition and sequencing protocol are outlined in appendix E.

2.7. *In silico* functional analysis

Published data from various sources such as ENCODE (ENCODE Project Consortium, 2012) and the Epigenome project (Kundaje et al., 2015; Zhu et al., 2013) were used to perform the *in silico* functional analysis. Where possible, data from keratinocytes or from a keratinocyte cell line were used. In the absence of keratinocyte data, data from other cells types were used. All the data in this section, unless otherwise stated was retrieved using ENCODE data (ENCODE Project Consortium, 2012) and visualised using the UCSC genome browser (Kent et al., 2002).

2.7.1. Frequency of structural variants and overlapping regulatory elements

The frequency and novelty of large structural variants was evaluated through DGV (MacDonald et al., 2014) which annotates known structural variants from literature and databases, such as the 1000 Genomes Project. Variants that were present in DGV were excluded from further analysis. Large structural variants were also assessed through UCSC genome browser using ENCODE data to determine if variants overlapped with any regulatory regions.

2.7.2. Identification of regulatory elements that overlap with the causal variant

At the first pass, all "Regulation" tracks were activated, so that all regulatory features that overlap with the structural variants could be detected. Additionally, transcription factor binding sites overlapping with the regulatory element within the KWE causal mutation were identified from the ENCODE factorbook ChIP-seq data (Wang et al., 2013). ENCODE

factorbook ChIP-seq experimentally interrogated the binding of 167 transcription factors across several cell lines in 837 experiments.

2.7.3. Genomic architecture prediction

The genomic architecture of the region containing the duplication was evaluated using available Hi-C and ChIP-seq data (Rao et al., 2014). The potential pathogenic effect of the identified mutation may depend on the topological domain structure of the genomic region. Domain demarcations are defined by CTCF binding and these are conserved between cell types. CTCF binding sites around the identified mutation were determined in normal human epidermal keratinocytes (NHEK) and in a breast cancer cell line (MCF-7) using CTCF ChIP-seq data. Like the NHEKs, the MCF-7 cell line is of ectodermal origin and was therefore used in the absence of keratinocyte data. The MCF-7 cell line was used to determine the genomic architecture around the identified mutation using CTCF chromatin interaction data from Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET)(Li et al., 2010) as no CTCF interaction data exists for epidermal keratinocytes (ENCODE Project Consortium, 2012).

2.7.4. Known regulatory element interactions

RNA polymerase II (RNAPII)-targeted ChIA-PET interaction data was used to identify contacts between regulatory elements such as enhancers and promoters and to determine possible interactions between the regulatory elements and their predicted target genes. No RNAPII-targeted ChIA-PET interaction data exists for epidermal keratinocytes; therefore all published RNAPII-targeted ChIA-PET interaction data from five cancer cell lines (breast cancer (MCF-7), chronic myeloid leukemia (K562), cervical cancer (HeLa-S3), colorectal cancer (HCT116) and promyelocytic (NB4) cell lines) were used to determine possible interactions.

2.7.5. Epigenomic profiling using public databases

Epigenomic data from ENCODE (ENCODE Project Consortium, 2012) and Epigenome Roadmap (Roadmap Epigenomics Consortium et al., 2015) project data were used to search

for regulatory elements and annotated non-coding variants. NHEK cell lines were prioritized as they are most similar to the cells affected in KWE. DNaseI-seq data from NHEK were used to identify DNaseI hypersensitive sites and define them as open chromatin could contain potential regulatory elements. The presence of regulatory elements in the regions of interest was further validated by integrating histone modification (H3K36me3/H3K27ac) data in 47 different human cell types. The average H3K27ac signal across the duplicated regions was calculated to measure the overall regulatory activities of the structural variants, and the H3K36me3 signal across the gene body was calculated to measure the transcriptional activity of nearby genes. Assuming an enhancer's activity will correlate with the expression of its regulated genes, Pearson correlations between H3K27ac signal of the tandem duplication and expression (H3K36me3) of nearby genes were calculated and used to predict the target gene/s of the tandem duplication as the nearby gene with the highest correlation.

2.8. In vitro functional analysis

This portion of the study was done in collaboration with our Norwegian colleagues. Samples were collected from South African (n=3 affected and n=3 controls) and Norwegian (n=3 affected and n=3 controls) participants. Single 4 mm palmar skin punch biopsies were taken from affected individuals and ethnically matched healthy controls by a dermatologist. Biopsies were taken from the thenar region of the palm in regions that did not appear to be going through a peeling cycle (Figure 2.5). The biopsies were cut in two by a histopathologist, perpendicular to the skin surface. One half was used for relative qPCR expression studies and the other was processed for histology. Half of each biopsy was placed in 1ml dispase II protease (1.2 U/µl) for 4 hours to separate the epidermis from dermis. The epidermis was then placed in RNALater and stored at 4°C for 24 hours and then at -20°C until RNA extraction. The other half of the biopsy was placed in 10% formalin for immunohistochemistry processing.



Figure 2.5: Affected palm where the biopsy was taken
The biopsy was taken from the thenar region of the hand where the skin did not appear to be going through a peeling cycle. (Image Credit: Prof Michele Ramsay)

2.8.1. Relative gene expression analysis

The RNALater tubes containing the epidermis samples were thawed on ice. Each epidermis was homogenized using the TissueLyser II (Qiagen, Hilden, Germany) and total RNA was extracted using the RNeasy® kit (Qiagen, Hilden, Germany) as per manufactures instructions. RNA quality and quantity were determined using the Experion™ automated electrophoresis station (Bio-Rad, Hercules, CA, USA) and the NanoDrop® ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) , respectively. One of the South African KWE samples was excluded from further analysis due to low RNA quality and concentration. Generation of first-strand cDNA was performed using the SuperScript® VILO™ cDNA Synthesis Kit (ThermoFisher Scientific, Waltham, MA, USA), with an input of 250ng of RNA per sample (along with a blank/negative control and a no reverse transcriptase control). The cDNA samples were then stored at -20°C.

The gene expression assays were run using 25ng cDNA and calibrator sample was used for each assay. The Applied Biosystems™ TaqMan® Gene Expression Assays (ThermoFisher Scientific, Waltham, MA, USA) for *CTSB* (Hs00947433_m1), *FDFT1* (Hs00926054_m1), nei like DNA glycosylase 2 (*NEIL2*, Hs00979610h) and ribosomal protein lateral stalk subunit P0 (*RPLP0*,Hs99999902m1, endogenous/ housekeeping gene control) were used to determine the gene expression levels of the respective genes. Although three housekeeping genes are usually recommended for normalisation (Bustin et al., 2010), *RPLP0* was shown to be an appropriate and sufficient housekeeping gene in the epidermis as it's expression was consistent during keratinocyte differentiation whereas the expression of commonly used housekeeping genes such as *GAPDH* altered with differentiation (Minner and Poumay, 2009). All the samples were run in triplicate on the on a 96 well plate using the ABI Prism 7900HT sequence detector system (Applied Biosystems, Foster City, CA, USA). The average threshold cycle (C_T) values for each sample were used to determine gene expression using the ΔCt-method (Winer et al., 1999). Differences in gene expression levels between affected individuals and unaffected controls were determined, and P values were calculated by Mann-Whitney U test on R (R Development Core Team, 2008). All the protocols for RNA extraction, cDNA conversion and qPCR are shown in Appendix F. Reagent preparations protocols are listed in Appendix G.

2.8.2. Histology and immunohistochemistry

On the samples were collected, biopsies were fixed in 10% buffered formalin for 4-8 hours immediately after collection then embedded in paraffin blocks before shipping from South Africa and Norway to our colleagues in The Netherlands. All histology and immunohistochemistry were done by Prof Joost Schalkwijk and his PhD student, Ivonne van Vlijmen-Willems. Paraffin sections were cut for histopathology (Haemotoxylin and Eosin (H&E) staining) and immunohistochemistry. Antisera against CTSB (R&D Systems), FDFT1 (Sigma) and NEIL2 (Sigma) was used for immunohistochemistry. Antisera for CTSB, FDFT1 and NEIL2 were raised in goat, rabbit and mouse respectively. Deparaffinized and dehydrated sections (6μm) were pretreated for 10 minutes in 10 mM sodium citrate buffer, pH 6.0 for antigen retrieval. Endogenous peroxidase was blocked with 3% hydrogen peroxide. Background staining was reduced by using 5% normal serum of the animal in which the biotinylated conjugated serum was raised. Further staining and embedding was

according to standard protocols, using the avidin-biotin-peroxidase complex and amino-ethyl-carbazole as a peroxidase substrate. CTSB staining was evaluated by two blinded observers. CTSB staining intensity was scored on a scale from 0 to 3 points; 0 indicating no visible staining and 3 strong staining. Staining was scored separately for the stratum spinosum and the stratum granulosum. Scores for 7 patients and 7 controls were analyzed by a 2-tailed Mann Whitney U-test.

Chapter 3

Results

3.1. Coverage and small variant analysis

Targeted capture sequencing of the KWE critical region on 8p23.1-22 was done in all members of the KWE related samples (n=42, 23 affected) and in seven unrelated, ethnically matched controls. The average read depth across the critical region was 795x and 92.3% of the target region was covered at >10x read depth.

Small variant analysis was carried out in all sequenced individuals using Vertica. A total of 3 690 rare or novel variants were found within the KWE critical region (variant present in at least one affected individual) and 410 of these variants were novel. No coding or regulatory small variants that segregated well with KWE were identified. However, a novel intronic variant in the *FDFT1* gene located at chr8: 11 672 190 (C>T) was identified in all affected individuals but not in any of the unaffected individuals. The *FDFT1* variant reference allele (C) is conserved in all primates but does not appear to overlap with any regulatory features in keratinocytes. There is no evidence to suggest that this variant causes KWE but it may be in linkage disequilibrium with the causal variant.

3.2. Structural variant analysis

When no small variants were found to be the likely cause of KWE, structural variants were evaluated in the target capture sequencing data. No structural variants overlapping with the coding regions within the KWE critical region segregated with the disease. From the target capture sequencing of the data, one variant, a tandem duplication within the KWE critical region was identified and showed strong segregation with KWE. This variant will be described in detail in the next section.

3.2.1. A novel tandem duplication segregates with KWE

A novel, non-coding 7.67 kb tandem duplication was identified in the KWE critical region on chr8:11 729 286-11 736 955 in 21/23 affected South African patients using the targeted sequencing data. Individuals with the tandem duplication had split reads that aligned to two parts of the genome, with the junction sequence of the duplication clearly defined in the reads (Figure 3.1). Pindel correctly identified the tandem duplication in 21/23 affected individuals but did not assign the tandem duplication to the remaining two affected individuals. Although the program did not detect the duplication in these two individuals, on manual inspection of the data they did have a small number of reads that suggested that they did in fact have the duplication, but had too few reads to be called with confidence by Pindel as having the mutation. Although the segregation of this variant was not perfect, it did show the best segregation pattern among the structural variants.

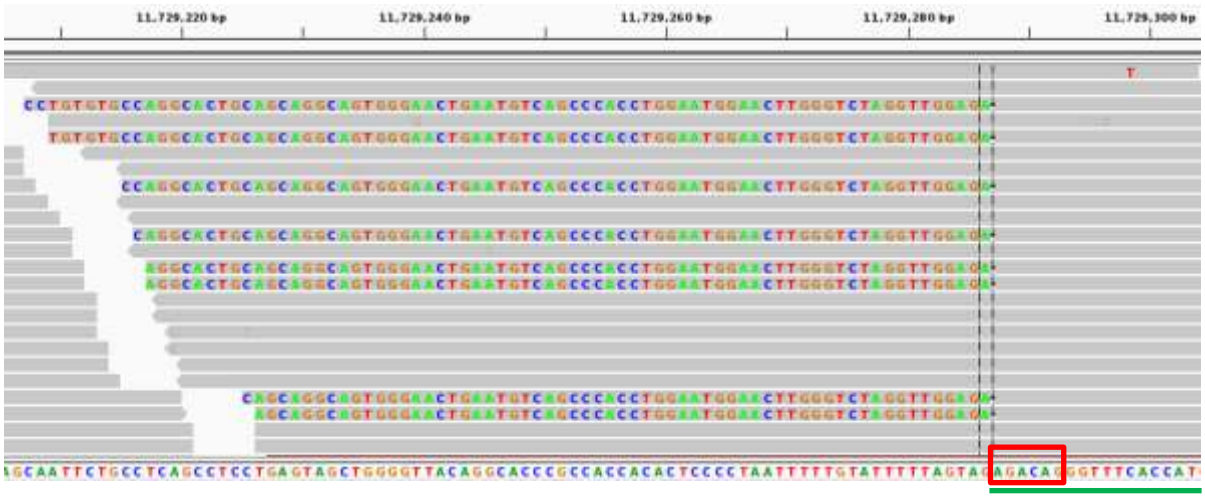
3.2.2. Validation of the tandem duplication

Through validation, using PCR and Sanger sequencing (Figure 3.2), the tandem dupliacationwas confirmed in all affected individuals (23 sequenced individuals with KWE and 11 additional KWE affected individuals). This included the two individuals that were previously called by Pindel as not having the tandem duplication. The duplication was not detected in any of the unaffected individuals (n=127, 96 unrelated ethnically matched control samples and 31 unaffected KWE family members). We used the junction sequence which is only present in people with the tandem duplication to design an assay to detect the duplication.

A.

TGTTGGTCTATTATTTTGCCTCCTGTGTGCCAGGCACTGCAGCAGGCAGTGGGAACTGAATGTCAGCCCACCTGGAATGGA
ACTTGGGTCTAGGTTGGAGAagacagggtttcaccatgttggccaggctggtctcgaactcctgacctcgtgatccacctgcctcggccccccaaagtgc
tgggattacaggcgtgagcc

B.



C.

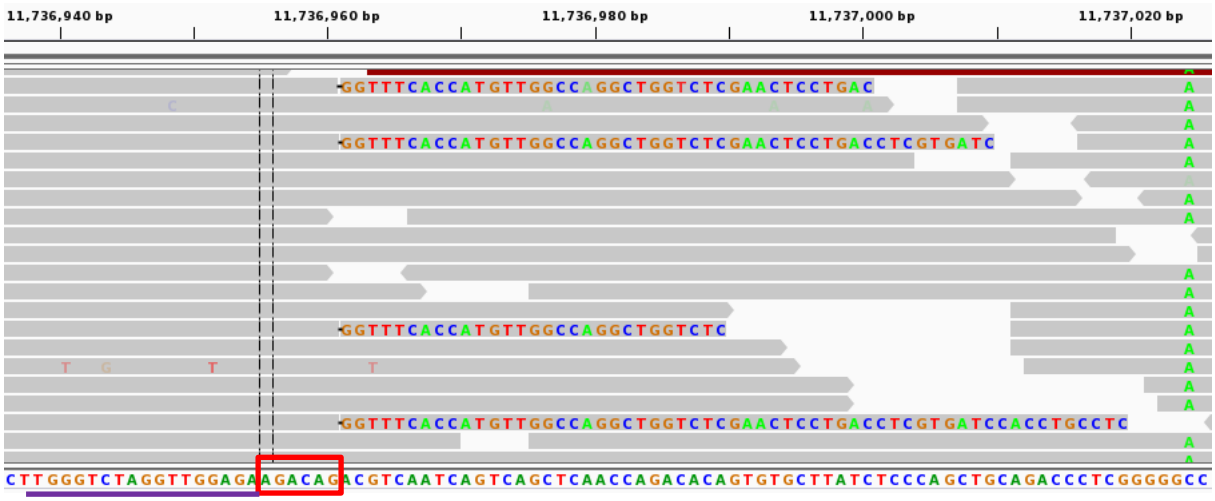
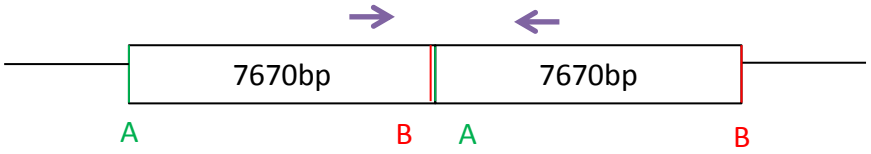


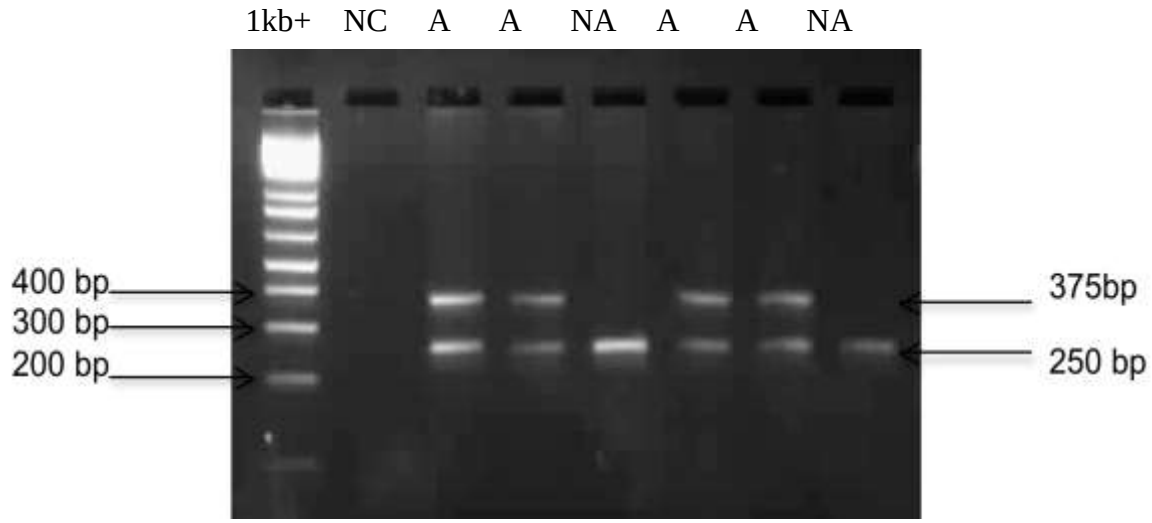
Figure 3.1: Visualisation of the tandem duplication break points on the Integrative Genomics Viewer (IGV)

(A) The sequence shows the junction sequence of the duplicated region, indicated by the change in case. Capital bases indicate the sequence at the end of the duplicated region whereas the small bases indicate the sequence at the start of the duplicated region. The reads as viewed on IGV are shown in (B) and (C) where the grey bars show reads that align with the reference sequence (shown at the bottom of each image) and the coloured bases indicate the regions of mismatch from the reference sequence. The vertical broken lines in (B) and (C) indicate the start and end of the duplication. (C) The vertical broken line is at position chr8: 11 736 955 instead of where the base mismatch starts (chr8: 11 736 961) because according to Pindel this is where the breakpoint occurs. The start and end sites of the duplication have a "AGACA" sequence in common (red box at the bottom of B and C), and therefore these bases align perfectly with the reference sequence in C and therefore IGV does not show these bases as a mismatch, erroneously placing the duplication breakpoint at position 11 736 961 instead of 11 736 955.

A.



B.



C.



Figure 3.2: Mapping of the duplication breakpoints using gel electrophoresis and Sanger sequencing for the validation of the tandem duplication

The junction sequence of the duplicated region is unique and is only present in individuals with the duplication. (A) An illustration of the duplication as it would appear on the chromosome. In the normal chromosome, region A is usually far from region B, which are the start and end of the duplication but these two regions are brought close to each other forming the junction sequence. Primers were placed on either side of the junction which is unique to people with the duplication. (B) Gel electrophoresis of the PCR products using the primer pairs (in A) for the duplication junction region and a control amplicon from the *LEP* gene. The control amplicon (250 bp) is present in all samples and the duplication junction amplicon (375 bp) is present only in affected individuals (lanes 3, 4, 6 and 7). A=Affected, NA=not affected, NC=Negative blank control. (C) shows a fragment of the sequencing output of the 375 bp amplicon from B and clearly shows the junction sequence as indicated by the "AGACAG" sequence at the start of the junction sequence, indicated by the purple block.

3.3. In silico functional analysis

3.3.1. Architecture of the tandem duplication

The tandem duplication is located in an intergenic region 3627 bp upstream of the *CTSB* gene transcription start site as shown in Figure 3.3. Data from the ENCODE project annotated in the UCSC genome browser was investigated to determine whether the duplicated region included functional genomic elements. Based on DNAse I hypersensitivity site data, parts of the duplicated region lies in an open chromatin conformation in keratinocytes. Based on H3K4me1 histone modification data, a marker for regulatory elements, including inactive elements, the region overlaps with two regulatory elements. Based on H3K27ac data from NHEK, a marker for active enhancers, one of these regulatory elements is an enhancer element that is active in normal human epidermal keratinocytes (NHEK), the cell type affected in KWE.

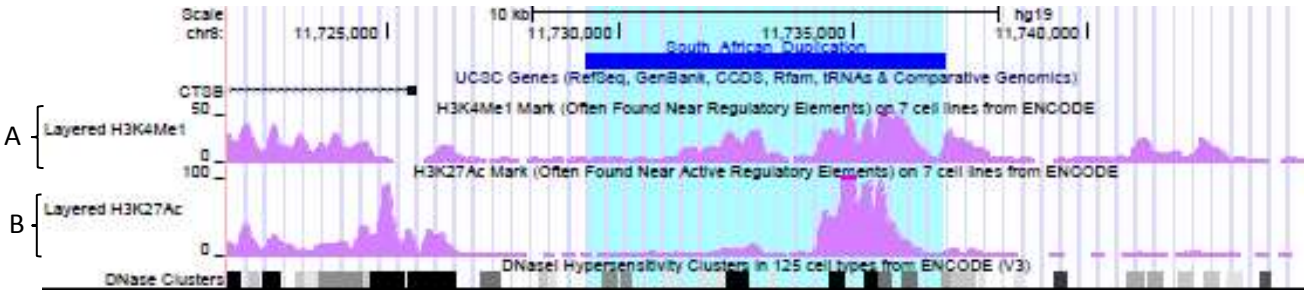


Figure 3.3: The tandem duplication lies upstream of the *CTSB* gene and overlaps with several regulatory elements

The tandem duplication (highlighted in turquoise) lies upstream of the *CTSB* gene within the KWE critical region. The duplicated region overlaps with histone markers associated with regulatory elements. H3K4Me1 is associated with regulatory elements and the pink peaks in (A) show that the duplicated region overlaps with two regulatory elements.
The duplication overlaps with the H3K27ac mark shown as the pink peak associated with active enhancers, suggesting that the region overlaps with an enhancer that is active in normal human epidermal keratinocytes (NHEK), the closest target cell type for KWE. Image generated using the UCSC genome browser (Kent et al., 2002).

3.3.2. The tandem duplication overlaps with a tandem duplication in Norwegian KWE families

During the course of the PhD study, we were contacted by Dr Torunn Fiskerstrand from Norway who was working with a dermatologist in Oslo who had identified patients with a phenotype compatible with a diagnosis of KWE. Independently, they had identified a 15.93 kb tandem duplication (chr8: 11 734 333-11 750 263) with a 95bp triplication in the middle of the duplication (chr8:11 744 352-11 744 446) in two Norwegian KWE families (Figure 3.4). This was done by whole genome SNP profiling, subsequent whole genome sequencing in two individuals and copy number variation analysis (Ngcungcu et al., 2017, Appendix H). The South African and Norwegian tandem duplications overlap at a 2.622 kb region. This 2.62 kb overlapping duplicated region co-localizes with an active enhancer (detected by H3K27ac profiling) identified in NHEK cells (keratinocytes) (Figure 3.4). These data strongly suggest that the duplication of the enhancer within the 2.62 kb overlapping region is the causal genetic mutation for the KWE phenotype in both populations.

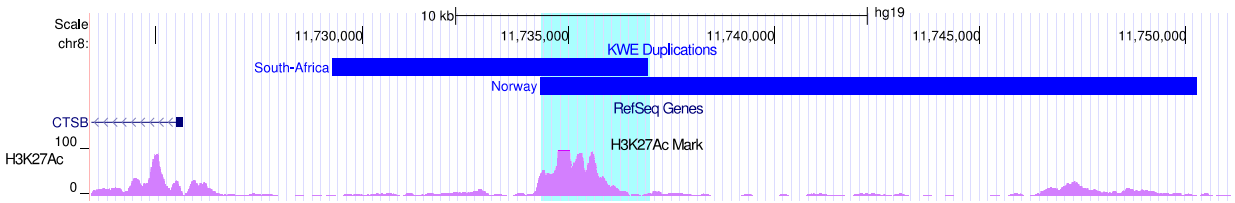


Figure 3.4: The South African and Norwegian tandem duplications overlap at an enhancer region active in keratinocytes
The South African (7.67 kb) and Norwegian (15.93 kb) tandem duplications are displayed as blue horizontal bars, located upstream of the *CTSB* gene. The 2.62 kb overlap (chr8: 11 734 333-11 736 955) between the two duplications (turquoise shading) is clearly positioned at the site of an H3K27ac histone mark (pink peaks, enhancer activity) in the NHEK cell line. Image created using the UCSC genome browser (Kent et al., 2002).

3.3.3. Transcription factor binding sites in the duplicated enhancer region

Several transcription factors are predicted to bind to the duplicated enhancer region across various cell lines, and many are active in keratinocytes (Figure 3.5) (Kouwenhoven et al., 2015; Wang et al., 2013). These include signal transducer and activator of transcription 3 (STAT3) and CCAAT/Enhancer Binding Protein Beta (CEBPB) transcription factors, which are important for keratinocyte differentiation and function (Hauser et al., 1998; Zhu et al., 1999).



Figure 3.5: Transcription factors that bind to the enhancer element common in the South African and Norwegian tandem duplications
Schematic overview showing the 2.62 kb overlap with the active enhancer in the South African (7.67 kb) and Norwegian (15.93kb) tandem duplications (blue horizontal bars). The enhancer region includes binding sites for several transcription factors. The different transcription factors are displayed as bars with the strength of binding represented by the shade of colour from light grey to black with black indicating the strongest evidence for transcription factor binding. Transcription factor names are shown to the left of each bar. Image created using the UCSC genome browser (Kent et al., 2002).

3.3.4. Frequency of the two duplications and other duplications/deletions in the KWE critical region

No duplications of similar size or a duplication of the same size as the 2.62 kb overlap region have been reported in the Database of Genomic Variation (DGV). However, several non-pathogenic duplications, deletions and inversions that encompass the duplicated regions (South African and Norwegian), and the *CTSB* and *FTFD1* genes which are two of the closest genes to the duplicated regions (Figure 3.6), have been described in apparently healthy individuals. These larger duplication/deletion variants therefore do not appear to cause KWE.

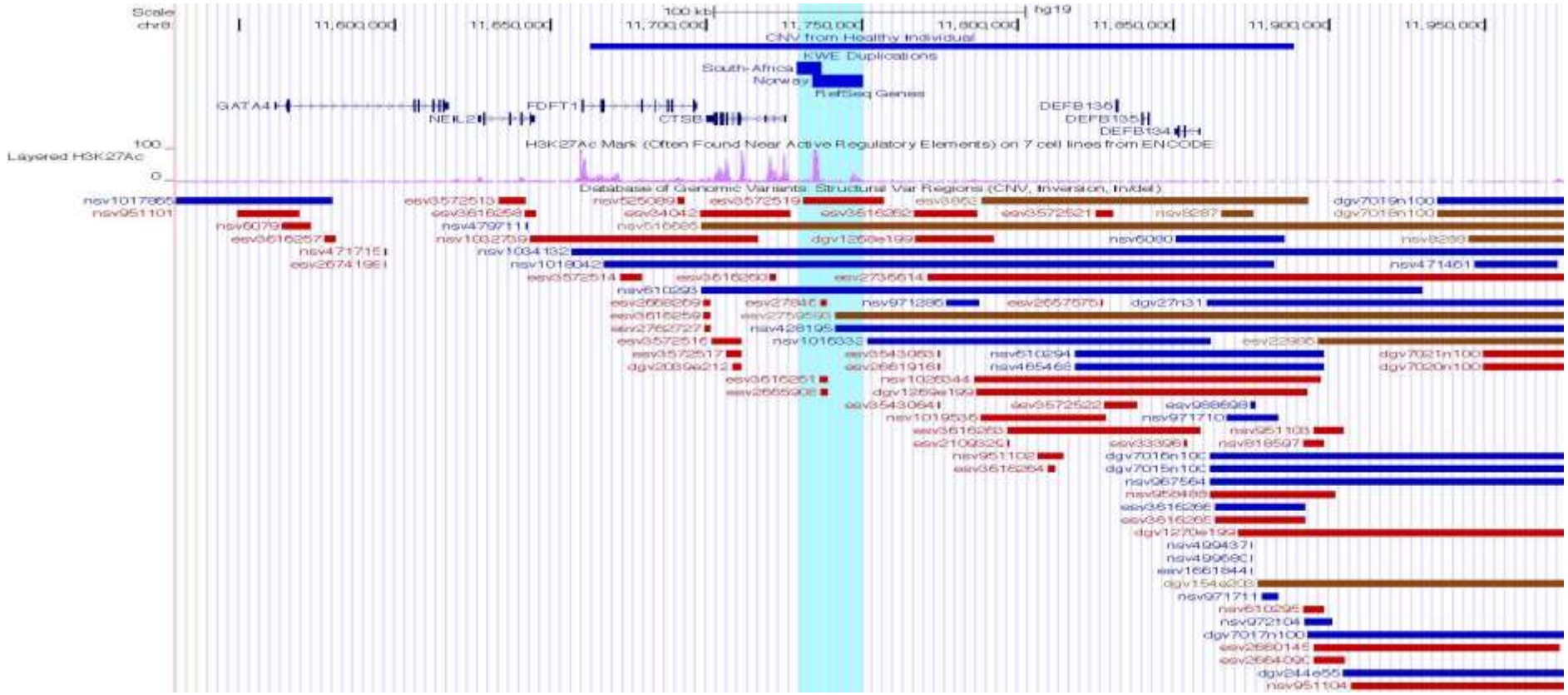


Figure 3.6: CNVs in the KWE critical region reported in normal individuals overlapping with the duplicated region

Schematic overview of the KWE critical region on chr8:11 525 326-11 987 757, with scale shown in the upper panel. Directly underneath, a large duplication identified in a Dutch cohort of 1416 healthy students (Bralten et al., 2011) and the South African (7.67 kb) and Norwegian (15.93kb) tandem duplications are displayed as blue horizontal blue bars. In the Database of Genomic Variants, no duplications (blue bars) of similar size to those described in South African and Norwegian patients (turquoise shade) have been reported. Several larger duplications (including the one in the Dutch cohort) are shown encompassing the region of the enhancer at chr8: 11 734 333-11 736 956 (large pink peak within the turquoise shade), but all of them include the *CTSB* gene and even the *FDFT1* gene. Bar colour coding: Red=deletion, blue=duplication and brown=inversion. Image created using the UCSC genome browser (Kent et al., 2002).

3.3.5. Chromatin architecture in and around the duplicated regions

3.3.5.1. CTCF binding sites and interactions

CCTC-binding factor (CTCF) binding often acts as insulators and restrict interactions between different regions of the genome (Bell et al., 1999). Therefore, conserved CTCF binding sites are important in gene regulation as they may act as a barrier between an enhancer and a gene. We assessed CTCF binding in the region around the enhancer in NHEK (keratinocyte) and in MCF-7 (breast cancer) cell lines. Figure 3.7 shows the CTCF binding in the two cell types. Notably, the CTCF binding in NHEK and MCF-7 is very similar, suggesting that the CTCF looping interactions would also be similar between the two cell types.

CTCF interaction loops reveal parts of the genome that are in the same domain and are possibly interacting with each other and this binding also reveals regions that are cut off from other regions due to the binding of CTCF. ChIA-PET CTCF looping data are not available in NHEK, but since most of the CTCF binding positions that were examined between NHEK and MCF-7 are similar, we used interaction loops in MCF-7 to predict potential interaction loops in keratinocytes, thereby defining topological domains in the region. Data from the MCF-7 cell line pointed to larger domains that include the *CTSB, FDFT1* and *NEIL2* genes as well as the enhancer. Based on this domain structure, the enhancer would be able to associate with any of the regions overlapping with this domain including the promoters of *CTSB*, *FDFT1* and *NEIL2*. A smaller subdomain within this region that contains only the *CTSB* gene and the enhancer was also observed. The smaller domain is particularly interesting since there are several lines of evidence for the presence of this smaller domain in replication studies using the MCF-7 cell line. This domain is much smaller and only includes the duplicated regions and the *CTSB* gene suggesting that under certain conditions, the enhancer only interacts with the *CTSB* gene and not with the other two genes in the larger domain.

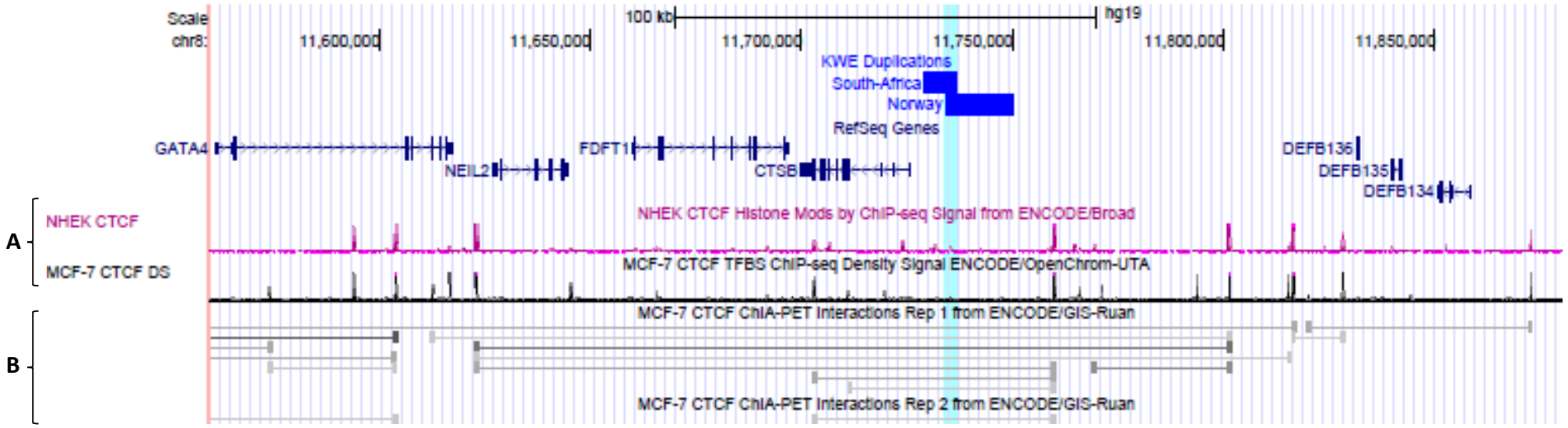


Figure 3.7: Topological subdomains, CTCF binding sites and loops involving the enhancer and nearby genomic regions
Schematic overview of the KWE critical region on chr8:11 560 000-11 880 000, with scale shown in the upper panel. The South African (7.67 kb) and Norwegian (15.93 kb) tandem duplications are displayed as blue horizontal bars, and the 2.67 kb overlap (chr8: 11 734 333-11 736 956) is marked across all horizontal panels (turquoise shading). (A) Shows that the CTCF binding sites are very similar in the NHEK (pink) and the breast cancer MCF-7 cell lines (black). (B) CTCF interaction data from ChIA-PET for the MCF-7 cell line showed that *CTSB, FDFT1* and *NEIL2* occur in the same topological domain with the enhancer, and a smaller subdomain exists, including only the *CTSB* gene and the enhancer.

3.3.5.2. Topologically associating domains using Hi-C data

Hi-C data reveal long and short-range chromosomal interactions which indicate DNA interactions. Low resolution NHEK Hi-C heat maps (Rao et al., 2014) show that the duplicated enhancer lies within a chromatin subdomain that includes the *CTSB* and *FDFT1* genes (Figure 3.8) suggesting that the region encompassing the enhancer may associate with the *CTSB* and *FDFT1* genes in keratinocytes. This domain overlaps with the larger domain discussed in 3.3.7.1 suggesting that under certain conditions, the enhancers interaction is limited to *CTSB* and *FDFT1*.

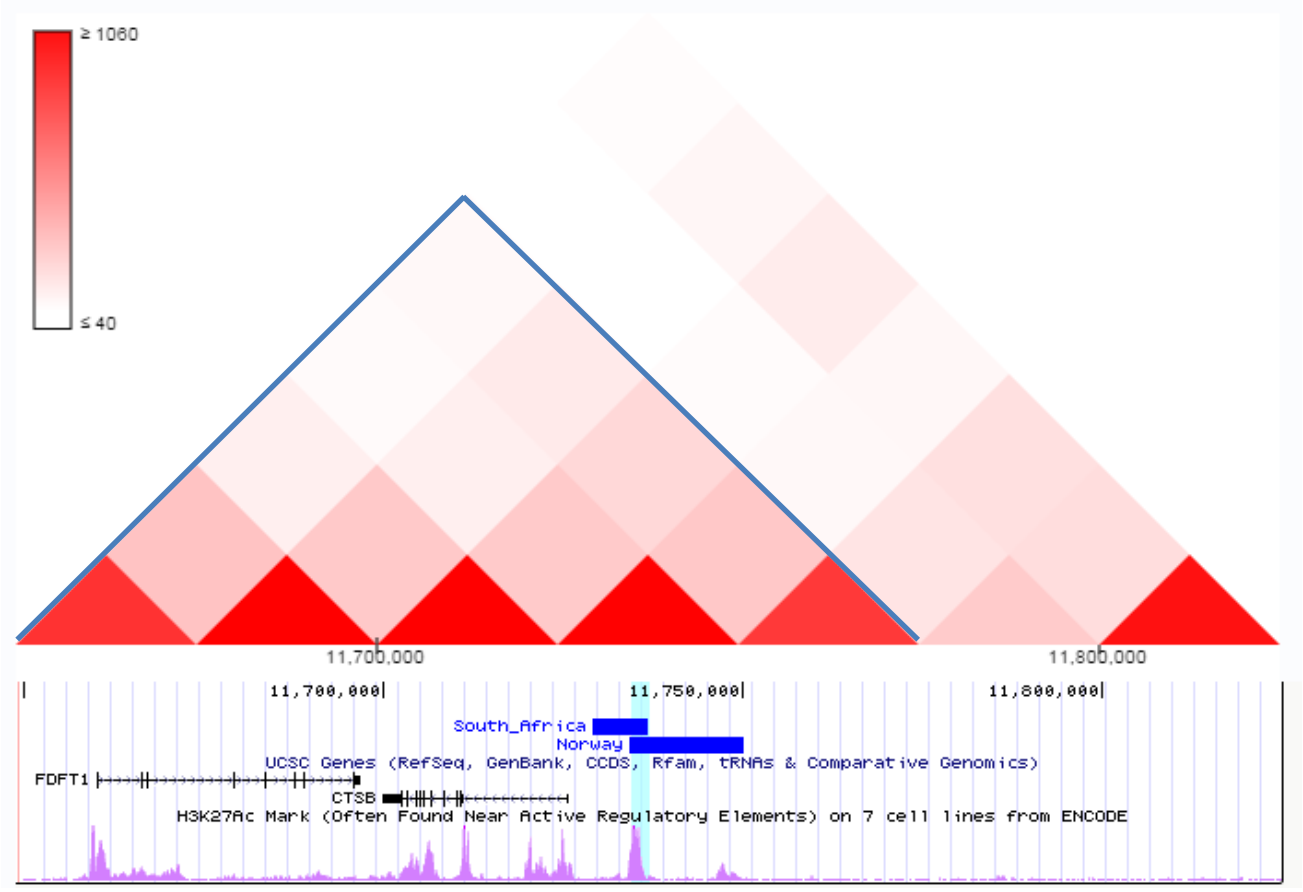


Figure 3.8: NHEK Hi-C interactions around the duplicated region
The heat map in the upper panel shows the topologically associating domains retrieved from HI-C data. The lower panel shows the region surrounding the duplicated region and the genes overlapping with the topological domain. The heat map shows that the duplications lie in the same topological domain (indicated by the blue lines) with *CTSB* and *FDFT1*. The shading from white to deep red indicates the intensity which is based on the number of reads that suggest that the given regions interact. Only data with at least 40 reads is shown. Darker shades indicate that there is more evidence for the interactions (Figure drawn using 3D Genome Browser at http://www.3dgenome.org).

3.3.6. Chromatin interactions within the duplicated region

To further examine possible intra-domain interactions within the region around the duplications, RNAPII ChIA-PET data from the MCF-7, K562, HeLa-S3, HCT116 and NB4 cell lines were analysed. Only MCF-7 and K562 interaction data are shown because enhancer interactions were only seen in these two cell lines. All enhancer interactions are shown in Table 3.1 and Figure 3.9. Interactions between the enhancer and the *CTSB* promoter and gene body, but not with the promoters of *FDFT1 or NEIL2* (Figure 3.9) were identified. In the K562 cells, the enhancer also appears to interact with the L-Threonine Dehydrogenase (*TDH*) pseudogene. Virtually all the observed RNAPII interactions in this region fall within the larger topological subdomain demarcated by strong CTCF binding in the MCF-7 cell line. The majority of the RNAPII-mediated chromatin interactions involving the *CTSB* promoter lie within the smaller subdomain that includes only the enhancer and the *CTSB* gene. An interaction between the region of the duplicated enhancer and *CTSB* gene expression was noted in cancer cell lines using CHIA-PET data, thereby adding to the evidence that *CTSB* is a possible target gene for the duplicated enhancer. Additionally, there was no evidence that the enhancer interacts with *FDFT1, NEIL2* and other genes in the region.

Table 3.1: Interactions between the duplicated enhancer and nearby genes/regions based on CHIA-PET data

Enhancer/gene interaction	Target gene
chr8:11724798..11726931- chr8:11734409..11737375	*CTSB*
chr8:11727352..11730707- chr8:11734493..11737265	*CTSB*
chr8:11200722..11203684- chr8:11734586..11737466	*TDH, BC038546/LOC101929290*
chr8:11710084..11711593- chr8:11733792..11735545	*CTSB*
chr8:11732515..11734330- chr8:11735926..11737604	Intergenic, between CTSB and the enhancer
chr8:11717778..11719934- chr8:11735920..11738113	*CTSB*
chr8:11724372..11727512- chr8:11735176..11737878	*CTSB*
chr8:11724372..11727512- chr8:11735176..11737878	*CTSB*
chr8:11724109..11726850- chr8:11734821..11736918	*CTSB*
chr8:11734290..11736871- chr8:11743194..11744743	Intergenic, between *CTSB* and the enhancer

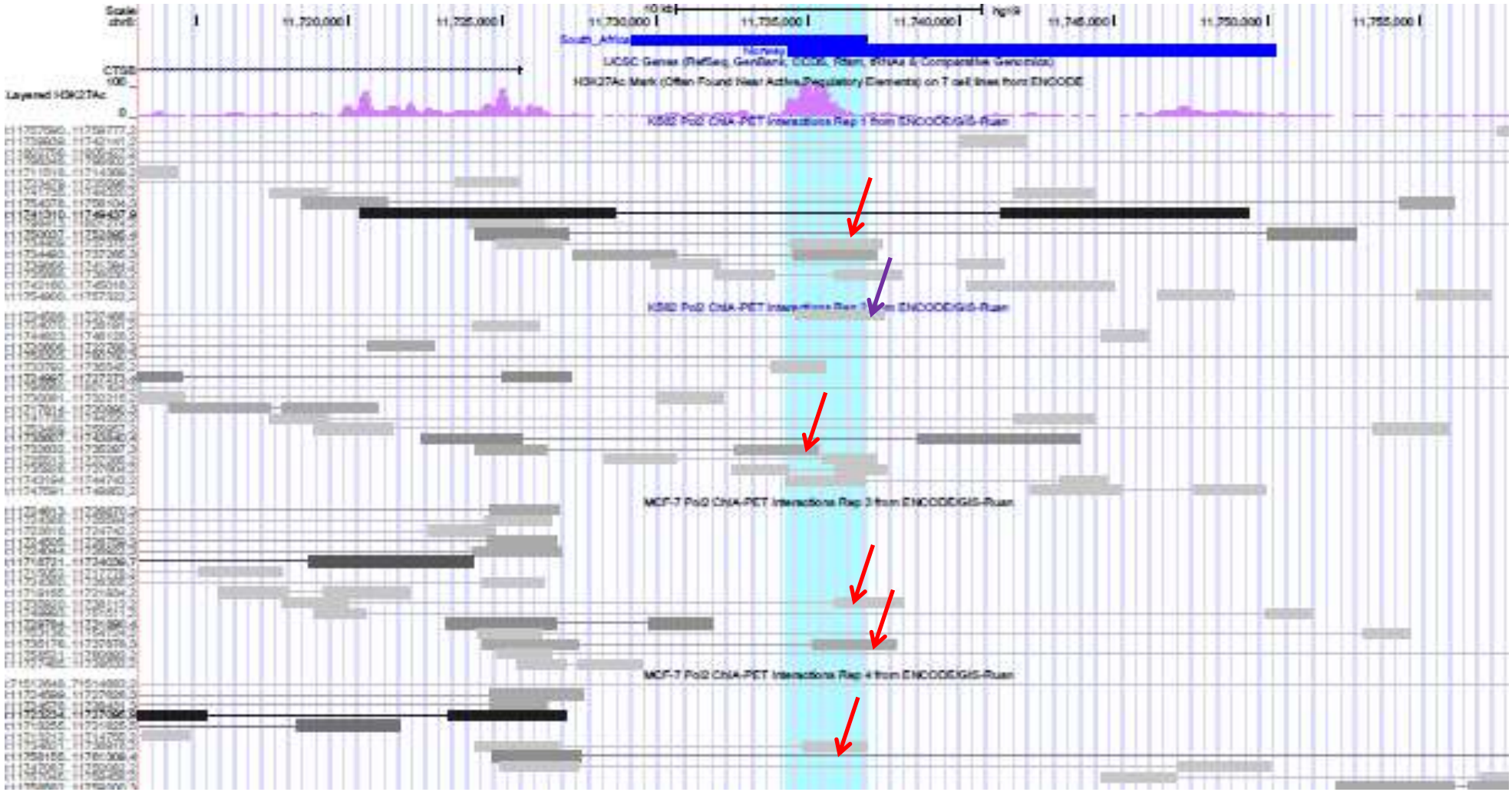


Figure 3.9: RNAPII ChIA-PET interactions between the enhancer and neighbouring *CTSB*

K562 and MCF-7 RNAPII ChIA-PET interaction data show interactions between the enhancer and *CTSB*, but not with *FDFT1* or *NEIL2*. The arrows indicate the interactions the between the enhancer and *CTSB* (not all shown). The purple arrow points to the enhancer's interaction with the *TDH* gene.

3.3.7. Correlation of the duplicated enhancer activity with transcription of *CTSB, FDFT1* and *NEIL2*

Based on the available data, the target and function of the duplicated enhancer, active in human keratinocytes, is unclear. Using histone markers to predict gene regulation through enhancer activity (H3K27ac) and transcription (H3K36me3), we examined the correlation of enhancer activity in the region with gene expression of three nearby genes, *CTSB*, *FDFT1* and *NEIL2*. There was a strong correlation between the enhancer's activity and *CTSB* across different cell lines including keratinocytes ($P<<10^{-9}$ and R=0.74) (Figure 3.10). The correlation between *FDFT1* transcription and duplicated enhancer activity is not correlated across different cell types (P<0.67 and R=-0.06). There was no evidence of *NEIL2* transcription in keratinocytes and no strong correlation between the enhancer and *NEIL2* transcription across different cell lines (P<-0.45 and R=-0.11).

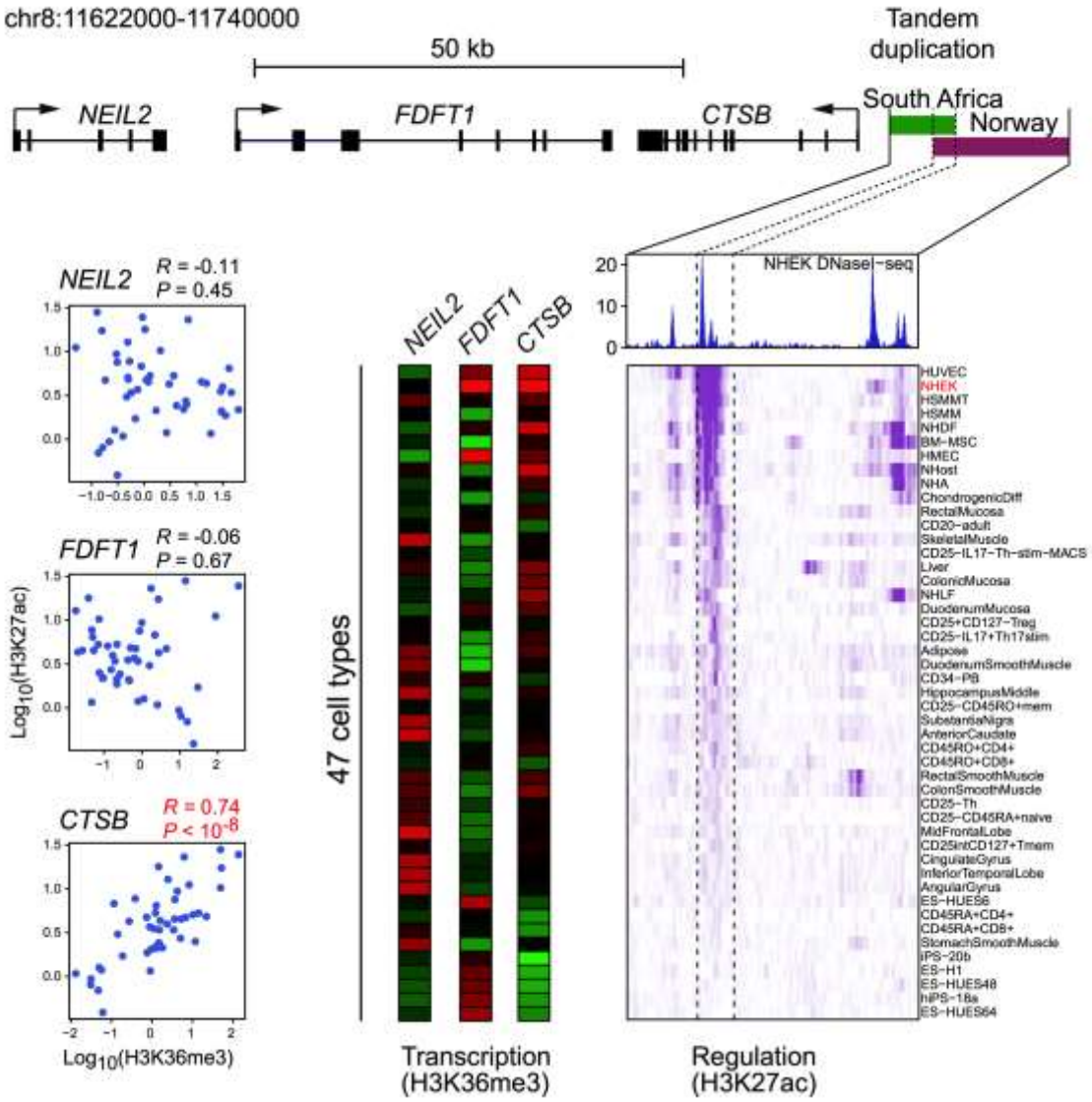


Figure 3.10: Prediction of enhancer activity and transcription of nearby genes using ENCODE data

The region of the South African (7.67 kb, green bar) and Norwegian (15.93kb, purple bar) tandem duplications is displayed, with the scale in the upper panel. The enhancer within the overlapping region was identified based on a DNase I hypersensitivity assay, shown here for keratinocytes (NHEK, blue peaks). Beneath this panel, the histone modification H3K27ac levels (ChIP-Seq data) across the enhancer region in 50 human tissue and cell types is shown as a regulation heat map (purple). Darker shades of purple indicate higher signal of the histone marker (H3K27ac), correlated to enhancer activity in the different tissues. Transcription of three nearby genes (*CTSB, FDFT1* and *NEIL2)* was predicted by H3K36me3 levels (ChIP-Seq data) across the genes in the same 50 different cell types. The transcription data are displayed as three parallel vertical bars in which red colour denotes high and green denotes low transcription activity. The correlations between H3K27ac and H3K36me3 levels are shown in the plots to the left. There was a highly significant positive correlation between the enhancer's activity and *CTSB* expression ($p<10^{-8}$).

3.4. In vitro functional analysis in skin biopsies

The functional analysis was done in both the South African and Norwegian KWE cases (n=7 in each group) and controls (n=7 in each group) in collaboration with Drs Friskerstand and Sitek and their groups from Norway and Prof Joost Schalkwijk from The Netherlands. Samples were collected from the palmar skin of affected and control individuals. In the affected individuals, the biopsy was taken from non-lesioned skin that did not appear to be peeling at the time of sampling (Figure 2.5). Each biopsy was cut in half and one half was used for the relative gene expression studies and the remaining half was used for histology and immunohistochemistry.

3.4.1. Relative gene expression for *CTSB*, *FDFT1* and *NEIL2*

We compared the relative gene expression levels of *CTSB, FDFT1* and *NEIL2* in palmar epidermis of affected individuals (n=6 [2 South Africans and 4 Norwegians]) to unaffected individuals (n=7 [3 South Africans and 4 Norwegians]), using qRT-PCR (Table 3.2). One affected South African sample was excluded from further analysis due to low RNA quality and quantity. Each cDNA sample was run in triplicate and the average CT values, expression ratio's (ΔCt) and fold change were calculated for each sample (Table 3.2). The test genes (*CTSB*, *FDFT1* and *NEIL2*) were normalised to one housekeeping gene, *RPLP0*. *RPLP0* is an appropriate exogenous control for keratinocytes (Minner and Poumay, 2009). Because of the small sample size in the individual Norwegian and South African groups, the data were analysed together. The relative expression level of *CTSB* was significantly higher in the skin from the affected individuals compared to controls (p=0.001) (Figure 3.11). There was no statistically significant difference in the expression of *FDFT1* (p=0.29) or *NEIL2* (p=0.44) between the two groups. Although there was no significant difference in the expression of *FDFT1* and *NEIL2* in the combined group, visually, the expression of the two genes appears to be slightly higher in the South African affected samples compared to the controls. Due to the small sample size, we were unable to determine if these differences were statistically different in the South African group.

Table 3.2: Average CT values and fold changes for *CTSB*, *FDFT1* and *NEIL2*

Sample	Status*	Sex	Test_gene_CT**	*RPLP0*_CT**	Fold change	Average fold change	Standard errors of the mean
CTSB							
NOR _Ctr1	N	F	26.21	21.86	1.39	1.16	0.19
NOR _Ctr2	N	F	26.02	21.72	1.45		
NOR _Ctr3	N	F	25.73	21.72	1.78		
NOR _Ctr4	N	F	25.75	21.78	1.81		
SA_Ctr1	N	F	28.7	22.48	0.38		
SA _Ctr2	N	F	27.45	22.24	0.77		
SA _Ctr3	N	M	27.96	22.19	0.52		
OSL_DII-1	A	F	24.55	22.62	7.46	5.95	0.93
OSL_DII-6	A	F	24.65	22.77	7.77		
OSL_DIII-6	A	F	24.22	22.05	6.29		
OSL_EI-2	A	F	23.83	22.12	8.67		
SA_KWE2	A	M	24.57	21.55	3.49		
SA_KWE3	A	M	26.19	22.37	2.02		
FDFT1							
NOR _Ctr1	N	F	29.4	21.86	1.33	1.21	0.22
NOR _Ctr2	N	F	28.74	21.72	1.9		
NOR _Ctr3	N	F	28.81	21.72	1.82		
NOR _Ctr4	N	F	28.74	21.78	1.98		
SA_Ctr1	N	F	31.45	22.48	0.49		
SA _Ctr2	N	F	31.24	22.24	0.48		
SA _Ctr3	N	M	31.25	22.19	0.46		
OSL_DII-1	A	F	29.55	22.62	2.02	1.62	0.13
OSL_DII-6	A	F	30.09	22.77	1.54		
OSL_DIII-6	A	F	29.19	22.05	1.74		
OSL_EI-2	A	F	29.05	22.12	2.01		
SA_KWE2	A	M	28.98	21.55	1.43		
SA_KWE3	A	M	30.33	22.37	0.99		
NEIL2							
NOR _Ctr1	N	F	29.91	21.86	2.17	1.41	0.3
NOR _Ctr2	N	F	29.75	21.72	2.21		
NOR _Ctr3	N	F	29.24	21.72	3.15		
NOR _Ctr4	N	F	30.7	21.78	1.19		
SA_Ctr1	N	F	32.85	22.48	0.44		
SA _Ctr2	N	F	32.54	22.24	0.46		
SA _Ctr3	N	M	33.2	22.19	0.28		
OSL_DII-1	A	F	30.86	22.62	1.9	1.96	0.22
OSL_DII-6	A	F	30.97	22.77	1.95		
OSL_DIII-6	A	F	30.33	22.05	1.85		
OSL_EI-2	A	F	29.64	22.12	3.11		
SA_KWE2	A	M	29.55	21.55	2.24		
SA_KWE3	A	M	32.01	22.37	0.73		

* N=Not affected, A=Affected; ** CT= Cycle threshold for each of the test genes or the endogenous control

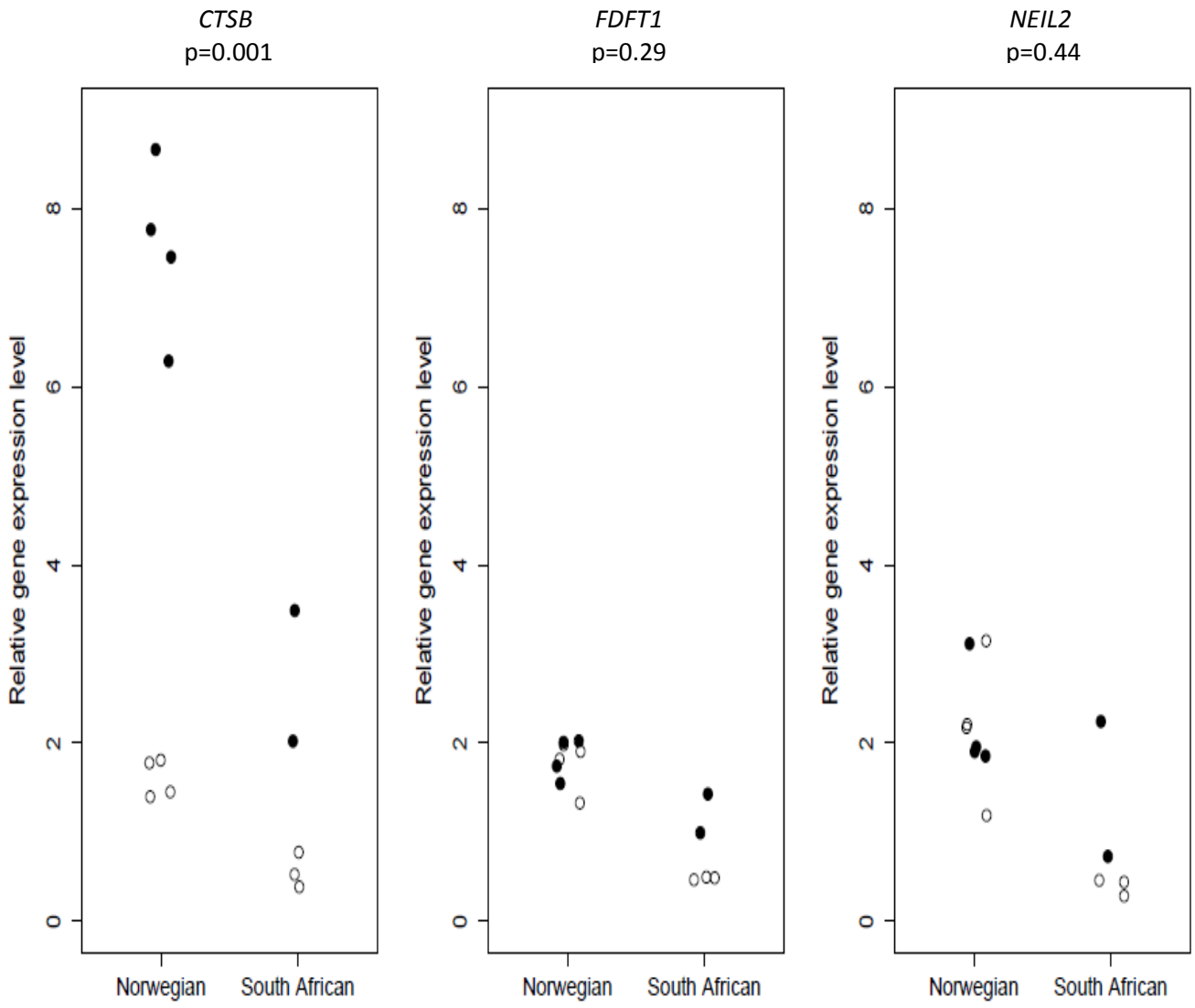


Figure 3.11: Relative expression of three nearby genes that occur in the same domain with the duplicated enhancer
Relative expression was calculated using the ΔCt method. Solid circles represent KWE patients and hollow circles the controls. In the combined South African and Norwegian sample, the expression of *CTSB* is significantly higher in affected individuals compared to controls (p=0.001), whereas there was no significant difference in the expression of *FDFT1* (p=0.29) and *NEIL2* (p=0.44) in affected compared to unaffected individuals.

3.4.2. Immunohistochemistry

3.4.2.1. Histology

Haemotoxylin and Eosin (H&E) staining of the patient samples showed a thicker epidermis and stratum corneum compared to control palmar skin and the morphology was largely normal, with all epidermal layers present and an orthokeratotic stratum corneum (Figure 3.12). As the biopsies were taken from non-lesioned skin, there was no overt blistering or stratum corneum splitting visible, nor was parakeratosis evident (Findlay and Morrison,

1978). Morphology was otherwise unremarkable with no evidence of the changes seem in the active phase which is characterized by basaloid hyperplasia, spongiosis, and horny layer showing parakeratosis and a split. However, the epidermis from the affected skin was much thicker than the epidermis from control skin. Additionally, the number of granular layers was increased in affected skin compared to the skin of controls (Table 3.3).

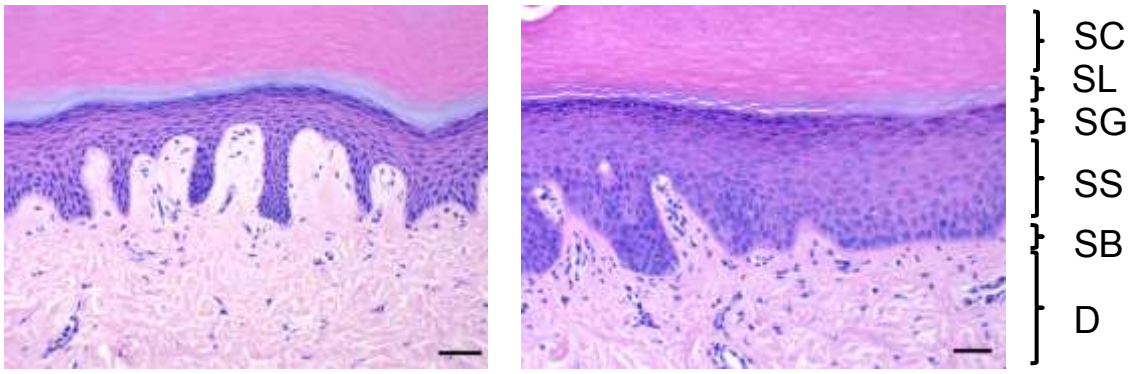


Figure 3.12: Haemotoxylin and Eosin (H&E) staining of hand palm biopsies of a healthy control (normal skin) and KWE patient (affected skin)
The epidermis in the affected individuals was notably thicker in affected skin compared to normal skin. SC: stratum corneum, SL: stratum lucidum, SG: stratum granulosum, SS: stratum spinosum, SB: stratum basale, D: dermis. The scale bar is 50µm.

3.4.2.2. Immunohistochemistry

As a part of a collaboration with our Dutch colleagues, the protein abundance and localisation of CTSB, FDFT1 and NEIL2 were determined in the biopsies from affected (3 South African and 4 Norwegian) and control (3 South African and 4 Norwegian) individuals. CTSB was found to be present in the stratum spinosum of control and affected skin in a granular pattern throughout the epidermis, consistent with a lysosomal localization. However, the amount of CTSB in the stratum granulosum was either absent or weak in controls but always present in the stratum granulosum of affected skin, varying from weak (3 out of 7)to strong (4 out of 7) (Figure 3.13, Table 3.3). Semi-quantitative scores (Table 3.3) for CTSB stratum granulosum staining were significantly higher in affected skin (1.7 ±1.0 (mean±SD)) compared to controls (0.3±0.5 (mean±SD)) (p=0.006). FDFT1 stained very feint staining and no staining was observed for NEIL2 in both affected individuals and controls.

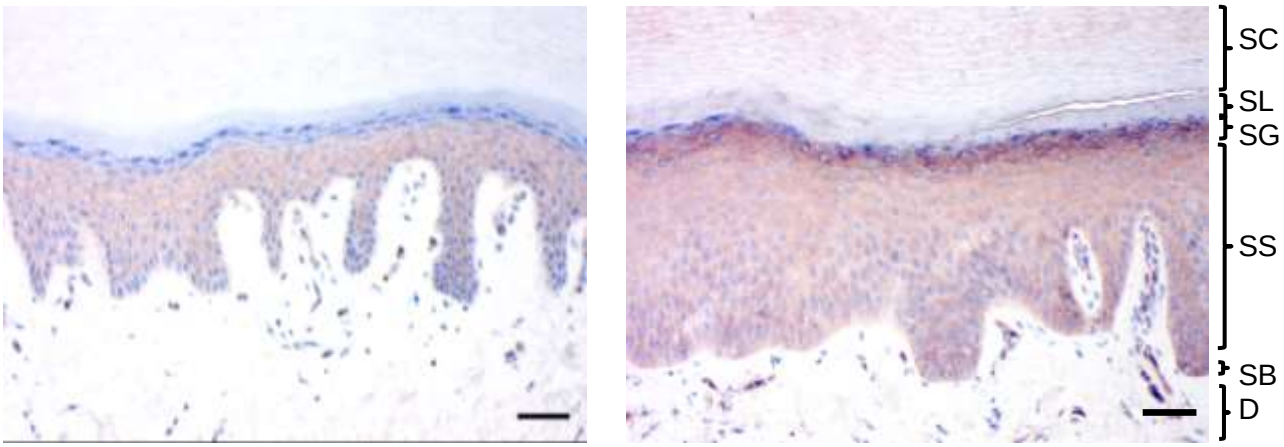


Figure 3.13: CTSB staining in a control individual and KWE case
CTSB is present in a granular pattern in the stratum spinosum of control and KWE skin. Note that CTSB staining is absent in the stratum granulosum of control skin, whereas it is strongly expressed in the stratum granulosum of KWE skin. SC: stratum corneum, SG: stratum granulosum, SS: stratum spinosum, SB: stratum basale. Note the increased epidermal thickness of the KWE skin. SC: stratum corneum, SG: stratum granulosum, SS: stratum spinosum, SB: stratum basale. Note the increased epidermal thickness of the KWE skin. Magnification= 20X. The scale bar is 50µm.

Table 3.3: Number of granular layers and CTSB staining intensity in affected and control skin

Sample	# of granular layers	CTSB staining intensity
NOR _Ctr1	2	0
NOR _Ctr2	3	0
NOR _Ctr3	3-4	1
NOR _Ctr4	4	1
SA_Ctr1	4	0
SA _Ctr2	4	0
SA _Ctr3	3	0
NOR_DII-1	6	2.5
NOR_DII-6	4-5	2.5
NOR_DIII-6	4	1
NOR_EI-2	6	0.5
SA_KWE2	5	2.5
SA_KWE3	5	2.5
SA_KWE1	4	0.5

*Intensity of CTSB staining in the granular layer of the epidermis

3.5. Summary of the results

NGS data revealed that there were no small variants or structural variants overlapping with the coding regions that segregated exclusively with KWE across all families. However, a large intergenic 7.69 kb tandem duplication overlapping with an enhancer known to be active in keratinocytes segregated with all affected individuals in all the South African KWE families studied. This tandem duplication overlaps with a 15.93 kb Norwegian tandem duplication at the enhancer element across a 2.62 kb region. This enhancer appears to regulate *CTSB* gene expression in normal keratinocytes and other cells. In the presence of the duplicated enhancer, *CTSB* expression was significantly upregulated in the palmar skin of individuals with KWE and there was an overabundance of CTSB in the granular layer of affected individuals. We therefore showed that the duplication of the enhancer leads to *CTSB* overexpression and CTSB overabundance in affected individuals.

Chapter 4

Discussion

The genetic cause of KWE, an autosomal dominant disorder, eluded researchers for decades. KWE was first described four decades ago (Findlay et al., 1977). In 1985, Peter Hull, a PhD candidate at the time, did a genealogical study on South African KWE families using their family histories including, birth, death and baptism records and found that the KWE families, including the Coloured families, could be traced back a single founder, Captain Francois Renier Duminy, a French mariner who settled in South Africa in the late 1700s (Hull, 1986). Hull, and Findlay et al. (1977) before him, suggested that KWE was common in South Africa due to genetic drift by founder effect, thereby suggesting the genetic mutation in South African families would be the same mutation across all these families. Subsequently, the keratin genes were excluded as candidates of KWE in South African patients (Spurdle and Starfield, unpublished data cited by Starfield et al., 1997). Two decades ago, the causal mutation was localized to 8p23.1-22 (KWE critical region) using linkage analysis (Starfield et al., 1997). The data showed that affected individuals share a common haplotype, confirming that the cause was due to a single founder mutation in South African families (Starfield et al., 1997). Since then, a number of studies have interrogated the coding regions in the KWE critical region and excluded coding mutations in several genes including the *FDFT1* and *CTSB* (Appel et al., 2002; Hobbs et al., 2012).

The aim of this study was to identify and characterize the causal mutation for KWE in South African families. Targeted resequencing of 8p23.1-22 was performed in three KWE families and seven unrelated, ethnically matched controls. No plausible novel, coding or regulatory small variants (SNVs or small indels) that segregated exclusively with the disease were identified. However, a tandem duplication that segregates completely with KWE and overlaps with an enhancer element active in keratinocytes was identified in all affected South African individuals. This tandem duplication overlaps with a tandem duplication identified in affected Norwegians and there is an enhancer element in the region of overlap.

4.1. Exclusion of small variants as causal variants for KWE

In this study, we performed target capture resequencing of 8p23.1-22 and found no plausible KWE causative small variants (SNVs or small indels) that were novel and coding or in regulatory regions, and that segregated exclusively with the disease. However, a novel intronic variant within the *FDFT1* gene was identified in all sequenced affected individuals but was excluded as a possible plausible causal variant because it did not overlap with any regulatory elements and there was insufficient evidence to suggest that the variant was causal.

4.2. Identification of a novel non-coding tandem duplication as the causal variant for KWE

We identified a novel 7.67 kb tandem duplication in a non-coding region within the KWE critical region, upstream of the *CTSB* gene that segregated completely with the disease in seven KWE families. The duplication overlaps with a novel, independently identified tandem duplication in two Norwegian families with the KWE phenotype. The South African and Norwegian duplications have a 2.62 kb region in common that encompasses an active enhancer element (based on H3H27Ac histone ChIP-seq data) that has been shown to be active in normal human epidermal keratinocyte cell lines (NHEK) (Figure 3.3 and 3.4), differentiating keratinocytes (Figure 4.1) (Ngcungcu et al., 2017) and several other cell lines used in the ENCODE and Epigenome Roadmap projects (Figure 3.10). Both the South African and Norwegian duplications are novel as no duplications with the same break points have been reported (Figure 3.6). Interestingly, larger duplications that include both duplications and *CTSB* only or *CTSB* and *FDFT1,* have been reported in apparently healthy individuals and have no obvious health implications, suggesting that it is the duplication of the smaller target region that is implicated in the disorder (Smith et al., 2004).

Several transcription factors were shown to bind to the enhancer across various cell lines (Figure 3.5) (Kouwenhoven et al., 2015; Wang et al., 2013) and are active in keratinocytes (Ngcungcu et al., 2017). The signal transducer and activator of transcription 3 (STAT3) and CCAAT/Enhancer Binding Protein Beta (C/EBPB) proteins along with p63 are important in

keratinocyte differentiation and function and mutations or altered regulation of these genes are known to alter keratinocyte differentiation (Hauser et al., 1998; Smith et al., 2004; Zhu et al., 1999).

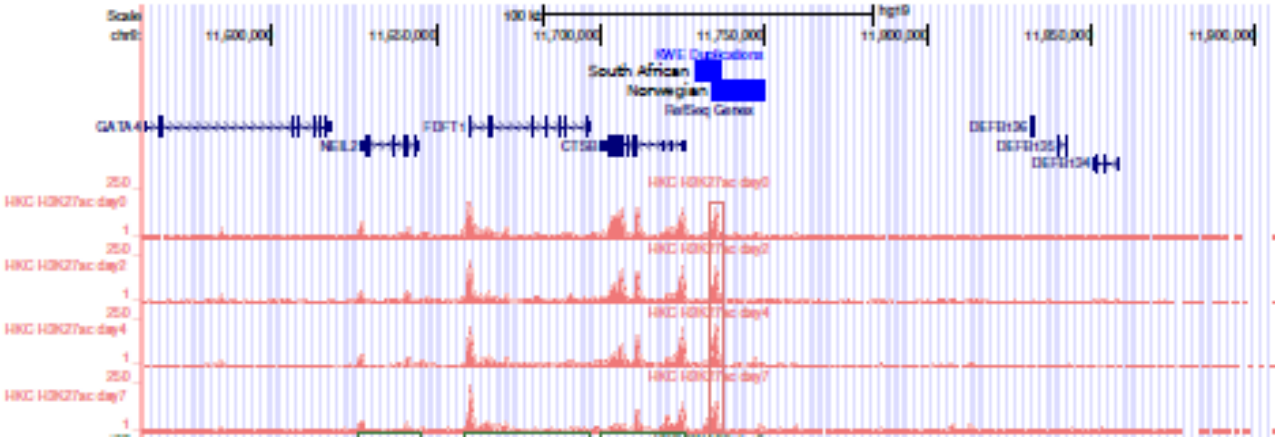


Figure 4.1: ChIP-Seq analyses of H3K27ac during human primary keratinocytes (HKC) differentiation.
Schematic overview of the KWE critical region on chr8:11 560 500-11 905 700, with the scale shown in the upper panel. The South African (7.67 kb) and Norwegian (15.93 kb) tandem duplications are displayed as blue horizontal bars. The H3K27ac histone modification signal (pink peaks) represents enhancer activity. The region corresponding with the 2.62 kb overlap between the duplications is indicated by the block over the H3K27ac peaks. Each line in the pink panel indicates different days of keratinocyte differentiation. From the first pink line: Day 0, day 2, day 4 and the last line is day 7 of keratinocyte differentiation. The enhancer is active throughout the first seven days of keratinocyte differentiation.

4.2.1. Copy number variation of non-coding regions as an unusual mechanism for disease causation

Because the enhancer is duplicated in the South Africa and in the Norwegian patients and is active in the cell type closest (non-palmoplantar epidermal keratinocytes) to the cell types affected in KWE patients (palmoplantar epidermal keratinocytes), we proposed that the duplication of the enhancer is the genetic cause for KWE. Copy number variation (CNV) including large insertions, deletions, duplications and inversions occurs across the genome and can either be benign or result in phenotype variation or cause disease. In cases where CNVs result in a diseased state, the coding regions are usually involved. Duplications of genes or portions of genes have been implicated in a several of Mendelian disorders but the duplication of non-coding regulatory region is a far less common disease mechanism although they have been previously reported, primarily in limb malformation disorders

(Dathe et al., 2009; Klopocki et al., 2011; Lohan et al., 2014). Amongst these are duplications that overlap with an enhancer known as the Zone Polarizing activity (ZPA) regulatory sequence (ZRS) which regulates expression of the sonic hedgehog (SHH) gene (Lohan et al., 2014) . Through ZRS regulation, *SSH* is expressed in the posterior region of the limb bud (ZPA) and determines the anterior-posterior axis in the limb bud during development. Several studies have shown that microduplications that overlap with the ZRS result in several polydactylies and syndactylies (Dathe et al., 2009; Lohan et al., 2014). Lohan and colleagues (2014) showed that overlapping microduplications caused Hass-type polysyndactyly (HTS) (larger duplication) and Laurin-Sandrow syndrome (LSS). HTS and LSS have overlapping phenotypes but LSS is more severe, with the smaller duplications causing the more severe phenotype (Lohan et al., 2014). Dathe and colleagues (2009) identified a 5.5 kb tandem duplication downstream of *BMP2* that was associated with an autosomal-dominant trait, brachydactyly type A2 (BDA2). BDA2 is a limb malformation characterized by hypoplastic middle phalanges of the second and fifth fingers. The duplication contains highly conserved sequences that are proposed to function as a cis-regulatory element regulating *BMP2* expression in the limb.

4.3. Functional impact of the enhancer element

4.3.1. The enhancer occurs in the same topological domain as *CTSB*, *FDFT1* and *NEIL2*

Using publically available data, we interrogated the genomic region around the enhancer in an effort to determine the target gene for the enhancer in the absence of the duplicated enhancer. Hi-C data showed that the enhancer occurs in the same topologically associating domain as *CTSB* and *FDFT1* (Figure 3.8). These data are low resolution and fine mapping of interesting regions is often required. Using CTCF binding sites in the NKEK and MCF-7 cell lines and CTCF interaction loops in MCF-7 cell lines, we showed that the enhancer lies in the same large topological domain with *CTSB, FDFT1* and *NEIL2* (Figure 3.7). These data also showed that a smaller subdomain within the larger domain exists and only includes the enhancer and the *CTSB* gene suggesting that in some instances, the enhancer may only be able to interact with *CTSB* but not the other two genes in the larger domain. NHEK CTCF binding data was available but CTCF interaction data were never generated. Oti and colleagues (2016), designed a tool that is able to predict CTCF interactions using CTCF

binding site data (Oti et al., 2016) and this tool predicted similar loops to those experimentally determined in MCF-7 cells (Figure 4.2) (Ngcungcu et al., 2017). These data showed that 1) the enhancer and *CTSB* and 2) the enhancer and *CTSB*, *FDFT1* and *NEIL2* were in the same domain and therefore, under certain conditions, the enhancer could in theory be able to interact with either of the three genes or exclusively with *CTSB*.

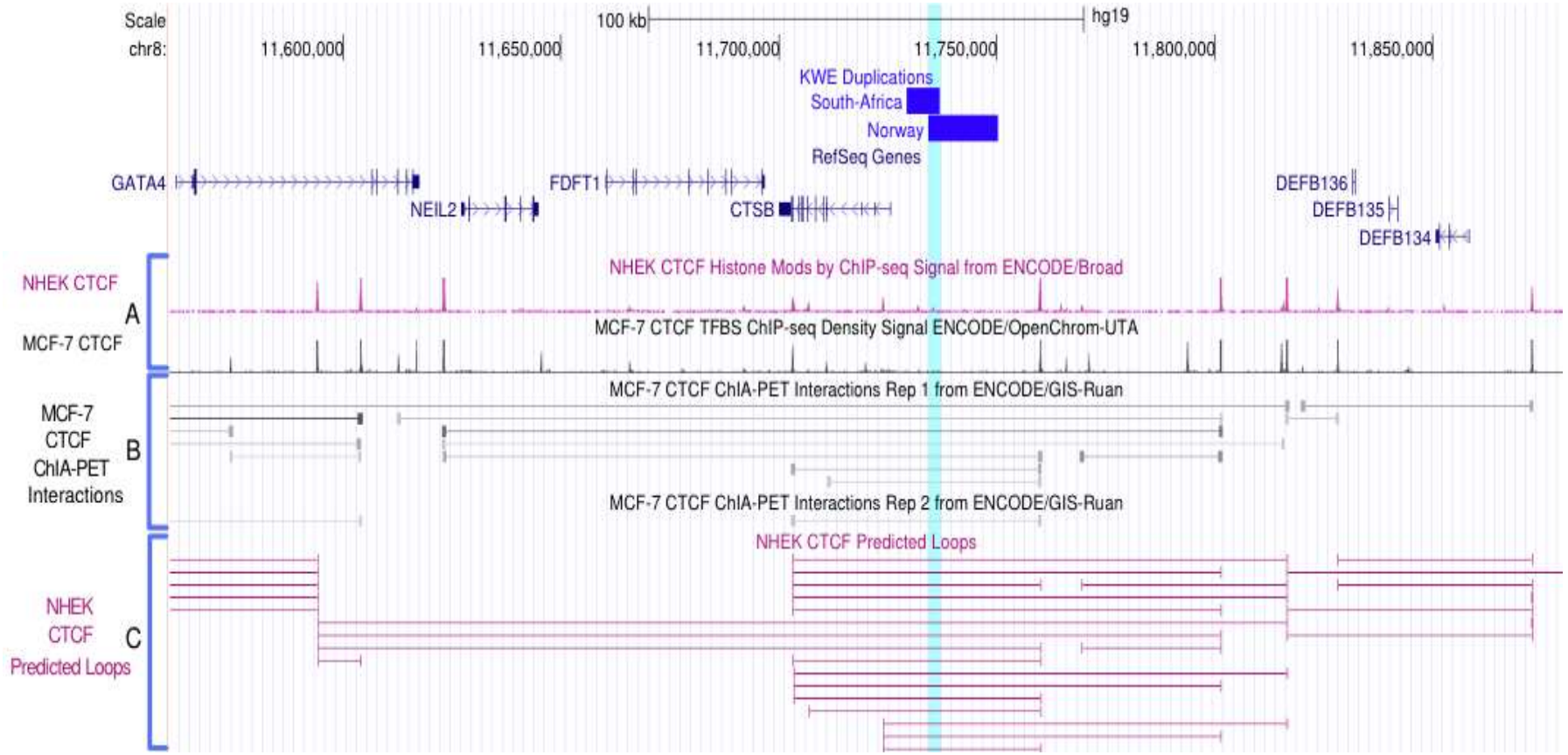


Figure 4.2: CTCF binding sites, MCF-7 interaction loops and NHEK predicted CTCF interaction loops.
Schematic overview of the KWE critical region on chr8:11 560 000-11 880 000, with the scale shown in the upper panel. The South African (7.67 kb) and Norwegian (15.93 kb) tandem duplications are displayed as blue horizontal bars, and the 2.62 kb overlap (chr8:11 734 333-11 736 955) is marked across all horizontal panels (turquoise shading). (A) CTCF binding sites in the keratinocyte (NHEK, pink peaks) and the breast cancer (MCF-7, black peaks) cell lines are highly similar. (B) ChIA-PET CTCF data from MCF-7 cells showing that *CTSB*, *FDFT1* and *NEIL2* may occur in the same topological domain with the enhancer (Rep 1 (Repetition 1)). A smaller subdomain exists that only includes the CTSB gene and the enhancer (Rep 1 and 2). (C) Using the CTCF binding sites identified in NHEK cells, CTCF interaction loops (and hence subdomains) in these cells were predicted (Oti et al., 2016). These data indicate a larger subdomain extending from the strong CTCF binding site close to the *DEFB136* gene to the *GATA4* gene, and a smaller subdomain, including only the enhancer and *CTSB*.

4.3.2. The enhancer interacts with the promoter of *CTSB*

We examined RNA polymerase II (RNAPII) binding which is required for transcription and interacts with the promoter, and indirectly with the enhancer. RNAPII interactions were determined using ChIA-PET data which identify regions of the genome that interact with each other through RNAPII binding. It is expected that using this data, based on the interaction of the enhancer to a particular gene, we could determine the target gene for the enhancer. Because these experiments had only been performed in the cancer cell lines, data from five cancer cell lines were analysed. The enhancer only showed interactions in the leukemia (K562) and the breast cancer (MCF-7) cell lines. Using these data, it was clear that the enhancer interacts with the *CTSB* gene but not with the other two genes in the same domain as the enhancer, this was observed in 2 replicates in both cell lines. In one replicate of the K562 cell line, one other interaction was noted with the L-Threonine Dehydrogenase (*TDH*) gene, which is a pseudogene (Table 3.1 and figure 3.9) (Edgar, 2002). Only one line of evidence suggested that the enhancer interacts with the *TDH* gene and because this is a pseudogene, it is unlikely that its altered enhancer interaction could lead to a disease.

4.3.3. The enhancer's activity correlates with the expression of *CTSB* in keratinocytes and other cells

Using H3K27ac and RNAPII ChIP-seq data, our collaborators analysed enhancer activity (H3K27ac) and gene expression (RNAPII) for *CTSB*, *FDFT1* and *NEIL2* in differentiating epidermal keratinocytes over seven days (Figure 4.3) (Kouwenhoven et al., 2015; Ngcungcu et al., 2017). The enhancer's activity was highly correlated, but not with statistical significance, with *CTSB* expression during differentiation of normal keratinocytes but not with *FDFT1* and *NEIL2* expression. Additionally, we analysed ENCODE and Epigenome Roadmap data from 47 cell lines and observed a strong correlation between the enhancers activity and *CTSB* transcription (H3K36me3 histone mark) across the cell lines, including NHEK (Figure 3.10).

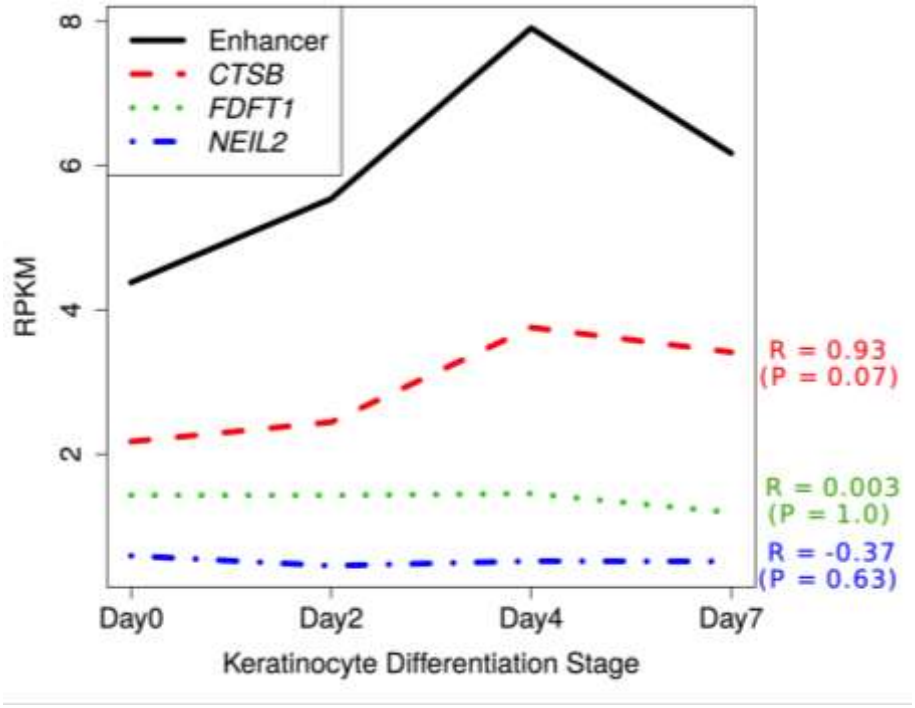


Figure 4.3: Activity of the enhancer and the three nearby genes during keratinocyte differentiation.
The enhancer's activity and gene expression were determined from H3K27Ac and RNAPII ChIP-seq data, respectively. There was a strong correlation between the enhancer's activity (black line) and the expression of *CTSB* (red broken line) although it was not statistically significant (R=0.93, p=0.07). There was no significant correlation between the enhancer activity and *FDFT1* (green broken line; R=0.003, p=1) and *NEIL2* (blue broken line: R=-0.37, p=0.63) (Ngcungcu et al., 2017).

Taken together, the presence of *CTSB* in the same domain as the enhancer along with the RNAPII interaction data and the correlation of the enhancer's activity with *CTSB* expression in keratinocytes and other cell types, suggest that the enhancer regulates the expression of *CTSB*. It was therefore important to perform *in vitro* studies in KWE affected skin to determine how the enhancer affects the expression of its target gene*, CTSB,* and/or the other genes in the same domain as the enhancer.

4.4. The duplication of the enhancer leads to the overexpression of its target gene, *CTSB*

Different lines of evidence suggest that the normal enhancer regulate*s CTSB* but it is still unclear which gene or genes are dysregulated in the presence of the duplication. The enhancer may dysregulate the expression of its target gene or the expression of a different

gene through enhancer adoption, probably of one of the genes in the same domain with the enhancer. We collected palmar skin biopsies from affected and control individuals and performed qPCR, histology and immunohistochemistry to determine the expression levels of the three genes in the same topological domain as the enhancer and the site and level of expression of their products. In KWE cases, the biopsy was taken from non-lesioned skin.

We analyzed the South African and Norwegian data together because the South African sample size was too small. In the combined South African and Norwegian group, the expression of *CTSB* was significantly higher in affected individuals compared to controls (Figure 3.11). The expression levels for *CTSB* are much higher in the Norwegian group compared to the South African group. It is unclear why this difference exists. It may be a confounding factor due to the fact that the RNA for the South African and Norwegian samples were extracted at different times and the RNA concentrations (for the test genes and the reference gene) in the South African samples were lower compared to the Norwegian samples. The differences in expression levels in the two groups may also be biological, as the South African and Norwegian duplications are different in size and position of the breakpoints. Lohan and colleagues (2014) showed that overlapping tandem duplications of different sizes caused similar phenotypes but the severity of the phenotype varied based on the size of the duplication. The smaller duplication caused LSS which is a more severe phenotype compared to HSS which is caused by a larger tandem duplication (Lohan et al., 2014). The different sizes of the duplications may cause different chromatin conformational changes in response to environmental stimuli and may therefore affect gene expression in a different way. The KWE phenotype appears to be more severe in South Africans compared to Norwegians (Hull, 2015, personal communication). CTSB abundance was significantly higher in the granular layer of affected individuals compared to controls (Table 3.3, Figure 3.13) whereas FDFT1 stained very faintly and no staining was observed for NEIL2. CTSB was present in the spinous layer of the epidermis in both affected and control skin.

Interestingly, in a previous South African study that examined gene expression of *CTSB* and *FDFT1* in wedge biopsies (without separation of the epidermis from the dermis), Hobbs and

colleagues (2012) did not observe an increase in the expression of *CTSB* in affected skin (Hobbs et al., 2012). Although the finding of the previous study did not observe an increase in *CTSB* in affected individuals, the evidence presented in the current study clearly show that *CTSB* expression is significantly increased in affected individuals and this increased expression is the likely altered pathway that leads to KWE. The authors concluded that the elevated expression of *FDFT1* that they observed in their study was likely due to a response to an intracellular cholesterol deficit caused by the inflammation associated with KWE during a peeling cycle (Hobbs et al., 2012). This elevation was not observed in the current study possibly due to the differences in the cycle of the skin peeling where the biopsies were collected (lesioned skin, in the Hobbs study vs non-lesioned skin in the current study).

4.5. Overexpression of *CTSB* and ectopic localisation of CTSB as a plausible cause for the KWE phenotype

The *CTSB* gene encodes cysteine cathepsin B (CTSB) which is a lysosomal protease of the papain enzyme family (Li et al., 2016; Turk et al., 2012). CTSB is one of eleven known human cathepsin subtypes (Li et al., 2016) and has homologues in other mammals. CTSB shows endopeptidase and carboxydipeptidase activity and is localised in lysosmal compartments of different tissue including the epidermis (Nagler et al., 1997; Rowan et al., 1993; Tholen et al., 2013). CTSB and many other cathepsins are ubiquitously expressed and are involved in protein degradation and turnover (Tholen et al., 2013; Turk et al., 2012). *CTSB* is associated with several cancers and other disorders including pancreatitis (Aggarwal and Sloane, 2014; Mahurkar et al., 2006; Yang et al., 2016). *CTSB* and its product are biologically plausible candidates for KWE because of its role, and that of other cathepsins and cathepsin inhibitors, in keratinocyte differentiation and disorders of keratinization.

4.5.1. CTSB affects keratinocyte cell-cell dissociation and keratinocyte migration

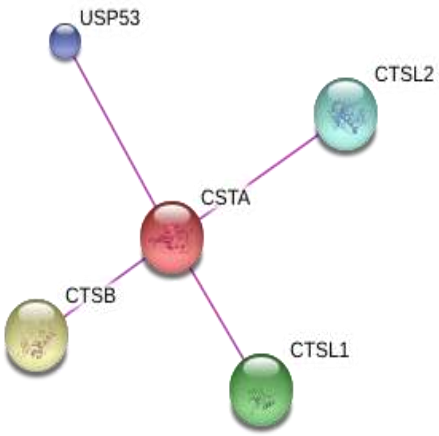
CTSB is abundant throughout the epidermal layers and secreted into pericellular spaces by keratinocytes (Büth et al., 2004). CTSB is contained within lysosomes in the cytosol of basal and spinous cells. Within the spinous layer, CTSB containing lysosomes can also be found close to desmosomes of the spinous layer keratinocytes and CTSB is therefore thought to be

secreted into the extracellular matrix between the keratinocytes in this layer (Büth et al., 2004). This is achieved through the migration of CTSB containing lysosomes from the region near the nucleus to the cell periphery then into the extracellular region during keratinocyte migration (Büth et al., 2004). It was also shown that CTSB was proteolytically active on the cell surfaces of migrating keratinocytes and that it has a role in cell-cell dissociation and degradation of the extracellular matrix allowing for cell migration during wound healing (Büth et al., 2007; Büth et al., 2004). When CTSB was inhibited in wounded cells, keratinocyte migration was impaired thereby suggesting that CTSB has a role in keratinocyte migration (Büth et al., 2007; Büth et al., 2004). In the case of KWE affected skin, CTSB was highly abundant in the granular layer rather than the spinous layer. Overexpression of CTSB in the granular layer may be causing accelerated keratinocyte dissociation and migration, thereby affecting the formation of the stratum corneum.

4.5.2. Absence of cathepsin inhibitors cause skin peeling disorders

Keratinocyte cornification and desquamation is controlled by a fine tuned balance between different proteases (including several cathepsins), and their inhibitors (Brocklehurst and Philpott, 2013; Zeeuwen, 2004). Interestingly, loss of function mutations in the gene encoding cystatin A (*CSTA*), an important inhibitor of CTSB and other cathepsins (Figure 4.4) cause phenotypes that have overlapping characteristics with KWE. CSTA is a thiol proteinase inhibitor and is a member of the stefins family of cystatins (cystatin family I) (Pavlova et al., 2000). CTSA has been shown to interact with several cathepsins including CTSB, and different isoforms of CTSL (Figure 4.4). CSTA inhibits CTSB through a two-step mechanism that involves an initial weak CTSB-CSTA link followed by a second tighter link that results in the conformational change of CTSB through the removal of the essential occluding loop in CTSB (Pavlova et al., 2000). Although CSTC (CST3 in Figure 4.4) inhibits CTSB in the same manner as CSTA but does so more efficiently (Pavlova et al., 2000), we only focused on the CSTA because the absence of this inhibitor leads to phenotypes that have overlapping characteristics with KWE.

A.



B.

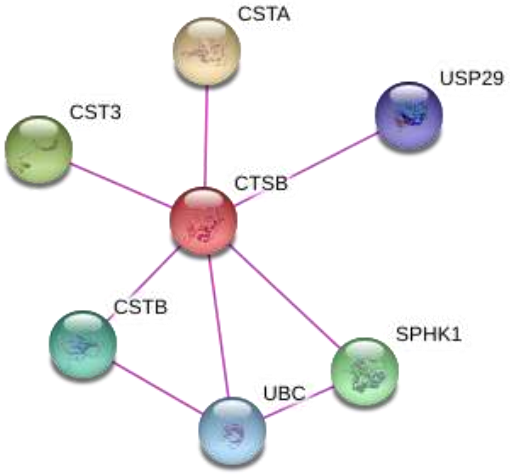


Figure 4.4: StringDB output showing interactions of CSTA and CTSB with other proteins. The pink line indicates that the interactions between the query protein (CSTA in A and CTSB in B) and other proteins were determined experimentally. Only high confidence interactions scores (0.7) are shown (Szklarczyk et al., 2015).

CTSA has an important role in desmosome-mediated cell-cell adhesion in the lower levels of the epidermis (Blaydon et al., 2011). In the epidermis, CSTA is expressed in stratum spinosum, stratum granulosum and the stratum corneum (Blaydon et al., 2011; Palungwachira et al., 2002) and secreted in normal sweat (Kato et al., 2005). Homozygous mutations in the CSTA gene lead to the absence of the protein causing exfoliative ichthyosis (PSS4) (Blaydon et al., 2011). Like KWE, affected individuals present as early as infancy, the palms and soles are affected and symptoms worsen when affected arrears are exposed to moisture. The disorder however, is non-erythematous and like the Norwegian KWE cases, the condition does not appear to be affected by seasonal change. Interestingly, the strongest synthesis of CSTA in normal skin occurs in the granular layer (Blaydon et al., 2011), this is the same layer where CTSB shows the strongest signal in KWE skin (Figure 3.13). This layer is subjected to apoptosis and temporary degradation during a KWE exacerbation (Hull et al., 2013), which could be explained by secretion of cathepsin B rising above a threshold triggering the lysosomal apoptosis pathway (de Castro et al., 2016). With the abundance of CSTA in the granular layer, one would expect that CTSB is significantly inhibited in this layer as seen in the normal skin of our control samples where CTSB was present at much lower levels in this layer compared to the controls. This suggests that CTSA is essential to maintain

low CTSB abundance in this layer for normal differentiation to occur. In the presence of the enhancer duplication, the abundance of CTSB is drastically increased and there likely is not enough CSTA in this layer to sufficiently inhibit CTSB leading to the inappropriate proteolysis of proteins required to successfully complete terminal keratinocyte differentiation. Homozygous mutations in CSTA also cause acral peeling skin syndrome (APSS), which like exfoliative ichthyosis is exacerbated by moisture and has skin desquamation phenotypes very similar to KWE (Krunic et al., 2013). Additionally, the deficiency of cystatin M/E, another cysteine protease inhibitor that also targets CTSB, causes an ichthyosis phenotype in mice (Zeeuwen et al., 2002).

These data suggest that the tight regulation of the expression pattern and level of CTSB and its inhibitors is crucial for regulating terminal keratinocyte differentiation. Moisture was a common trigger in the two human examples described above. Moisture is also a major factor in the presentation of KWE along with triggers that are more common in factors that occur in the winter (possibly cold conditions) which may play an important role in the pathways involved in protein regulation in the epidermis. In the case of KWE, chromatin conformation is important and this is likely altered in the presence of the duplication under environmental stressors. It may need a particular conformation that would lead to both enhancers having access to *CTSB* and may impact the *CTSB* promoter leading to the overexpression of *CTSB*.

4.5.3. Other cathepsin genes and their role in skin disease

Loss of function mutations in other cathepsin genes have also been shown to cause skin phenotypes both in humans and animals (Büth et al., 2004; Egberts et al., 2004; Horikoshi et al., 1998; Nishimura et al., 2002; Reinheckel et al., 2005; Roth et al., 2000; Schwarz et al., 2002; Toomes et al., 1999) with a pertinent example being the Papillon Lefèvre syndrome characterized by palmoplantar hyperkeratosis, periodontal disease and mutations in cathepsin C (Toomes et al., 1999). *CTSL* has been shown to regulate epidermal homeostasis as recessive mutations in *CTSL* have been shown to cause murine keratinocyte and hair follicle epithelial cell hyper proliferation (Roth et al., 2000). Notably, altered expression of

cathepsins can alter the localization of epidermal differentiation-related proteins. Murine CTSD, which plays a role in intracellular and extracellular catabolism, was shown to regulate transglutaminase 1, an important cornified envelope crosslinking protein (Egberts et al., 2004). In the absence of CTSD, the activity of transglutaminase 1 was significantly reduced and the concentration of other crucial differentiation and cornified envelope proteins such as keratins K1 and K5, loricrin, involucrin and filaggrin were altered (Egberts et al., 2004). The absence of *CTSD* also disrupted the structure of the epidermis, particularly the stratum corneum (Egberts et al., 2004). Similar to other cathepsins, altered expression and localization of CTSB may in turn alter expression of other genes and localization of proteins important in keratinocyte differentiation and maintenance of epidermal homeostasis leading to the KWE phenotype.

4.6. Study Limitations

There were a few minor limitations in the study and these are based on the limitations that we had in terms of the functional studies that we were able to do. The first limitation was that we had a small sample size for the functional studies. This was mainly due to the invasiveness of collecting the skin biopsy samples needed for the study. It was therefore difficult to recruit a large number of people for this experiment. As a result we had a small sample size for the gene expression analysis and immunohistochemistry. Although it is clear that *CTSB* is involved in KWE pathogenesis, it is still unclear whether *FDFT1* has a biological role in affected South African individuals. A larger sample size would help resolve this matter.

The second limitation was that we only examined the gene expression and protein abundance patterns of the three genes in the same topological domain as the enhancer. If there were additional interactions with the enhancer originating from beyond the topological domain, our assays would not detect these changes.

The final limitation was the cyclical nature of the skin peeling. In the current study, we collected skin biopsies from palmar skin that did not appear to be undergoing a peeling cycle in an effort to standardise the disease state across the different samples. In part, because of this, we were unable to confidently compare the gene expression findings from this study with the finding previously done in South Africa samples. Because KWE has several stages throughout the peeing cycles, we would need to collect samples from these different stages and evaluate the gene expression and protein abundance during the different stages of the peeling cycle to fully understand the disease at the genetic and protein level.

4.7. Future studies

Further functional studies will be necessary to determine how the overabundance of CTSB in the granular layer of the skin leads to the KWE phenotype and why there is variable expressivity of the disease phenotype in different individuals, even within the same family. Using cultured palmoplantar skin keratinocytes from affected individuals and controls would make it possible to examine a range of potential triggers for the erythema and peeling of the skin. Such studies could include exposure to stressors and gene expression studies in a cellular model of affected and unaffected cell cultures subjected to these stressors.

Further, chromatin conformation experiments may be used to determine the validity of our hypothesis that both enhancers interact with *CTSB* under certain conditions or in the presence of environmental stimuli. The conformation experiments would also answer the question of whether the chromatin conformation of the region around the enhancer is altered and creates different topological domains when the enhancer is duplicated.

It would be interesting to do RNA-seq to see which genes and pathways are affected by the presence of the enhancer duplication and the overabundance of CTSB in the granular layer. Using a genome wide analysis would allow us to determine pathways that are involved in keratinisation and the genes that are affected by CTSB. Now that we know which gene and protein cause KWE, we may be closer to getting an effective treatment for KWE, however

much work would be needed prior to developing an intervention where KWE can be effectively treated. It is unlikely that general CTSB inhibition on its own would be effective since CTSB appears to be an important protein in keratinisation. In this study we showed that CTSB is abundant in the stratum spinosum of normal skin and therefore inhibiting CTSB would also lead to inhibition of CTSB in this layer, likely altering the process of keratinisation. It is therefore important to determine the pathways that are affected by the overexpression of CTSB and the overabundance and ectopic localisation in the granular layer. RNA-seq would likely give us a clue of the other genes involved in keratinisation.

Chapter 5

Conclusion

Twenty years after the localization of the elusive KWE mutation to the KWE critical region (8p23.1-22), we have finally identified the causal mutation for KWE. Using next generation sequencing (targeted resequencing), we amplified and sequenced the entire KWE critical region (8p23.1-22) and identified a 7.67 kb tandem duplication that segregated completely with the disease in South African families. Furthermore, this tandem duplication overlaps with and independently identified tandem duplication in two Norwegian families. The two duplications overlap at an enhancer element that that is active in keratinocytes suggesting that the duplication of the enhancer element leads to the KWE phenotype.

Using available public data from keratinocyte and breast cancer cell lines, we showed that the enhancer is active in a keratinocyte cell line (NHEK) and in differentiating primary keratinocytes. The duplicated enhancer occurs in the same topological domain with *CTSB* and two other genes and there was some evidence of the presence of a smaller domain that only includes the enhancer and *CTSB* suggesting that under certain conditions, the enhancer may interact exclusively with *CTSB* and not the other two genes in the domain. Significant shifts in the topology of the region, possibly triggered by environmental factors, may drive the cells with the duplicated enhancer to favour increased expression of *CTSB* or any of the other genes in the domain. We also showed that the enhancer interacts with *CTSB* in two cancer cell lines but not with the other two genes in the same topological domain and that the expression of *CTSB* correlates with the activity of the enhancer in keratinocytes and other cell lines. All these experiments provide evidence suggesting that *CTSB* is the target gene for the enhancer in the absence of the duplication. We showed that *CTSB* is overexpressed in affected skin but not in the skin of healthy individuals. This overexpression of *CTSB* leads to the ectopic localisation and overabundance of the CTSB protein in the granular layer of the epidermis leading to KWE. Although the role of CTSB in the epidermis is not well understood, we are now one step closer to fully understanding KWE pathogenesis and finding an effective treatment.

References

AGGARWAL, N. & SLOANE, B. F. 2014. Cathepsin B: Multiple roles in cancer. *Proteomics Clin Appl,* 8**,** 427-437.

ALONSO, L. & FUCHS, E. 2003. Stem cells in the skin: waste not, Wnt not. *Genes Dev,* 17**,** 1189-200.

AMIN, A. N., DEGIOVANNI, C. V., FARRANT, P. B., et al. 2011. Photodynamic therapy for the treatment of keratolytic winter erythema. *Clin Exp Dermatol,* 36**,** 668-669.

APPEL, S., FILTER, M., REIS, A., et al. 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.

ARI, Ş. & ARIKAN, M. 2016. Next-Generation Sequencing: Advantages, Disadvantages, and Future. *Plant Omics: Trends and Applications.* Springer.

BANNO, T. & BLUMENBERG, M. 2014. Keratinocyte Detachment-Differentiation Connection Revisited, or Anoikis-Pityriasi Nexus Redux. *PLoS One,* 9**,** e100279.

BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell,* 98**,** 387-96.

BLANPAIN, C. & FUCHS, E. 2006. Epidermal Stem Cells of the Skin. *Annu Rev Cell Dev Biol,* 22**,** 339-373.

BLAYDON, D. C., NITOIU, D., ECKL, K. M., et al. 2011. Mutations in CSTA, encoding Cystatin A, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion. *Am J Hum Genet,* 89**,** 564-571.

BOUWSTRA, J. A. & PONEC, M. 2006. The skin barrier in healthy and diseased state. *Biochim Biophys Acta,* 12**,** 11.

BROCKLEHURST, K. & PHILPOTT, M. P. 2013. Cysteine proteases: mode of action and role in epidermal differentiation. *Cell Tissue Res,* 351**,** 237-244.

BUSTIN, S. A., BEAULIEU, J.-F., HUGGETT, J., et al. 2010. MIQE précis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Mol Biol,* 11**,** 74-74.

BÜTH, H., LUIGI BUTTIGIEG, P., OSTAFE, R., et al. 2007. Cathepsin B is essential for regeneration of scratch-wounded normal human epidermal keratinocytes. *Eur J Cell Biol,* 86**,** 747-61.

BÜTH, H., WOLTERS, B., HARTWIG, B., et al. 2004. HaCaT keratinocytes secrete lysosomal cysteine proteinases during migration. *Eur J Cell Biol,* 83**,** 781-795.

CHOMICZEWSKA, D., TRZNADEL-BUDZKO, E., KACZOROWSKA, A., et al. 2009. The role of Langerhans cells in the skin immune system. *Pol Merkur Lekarski,* 26**,** 173-177.

CIABRELLI, F. & CAVALLI, G. 2015. Chromatin-Driven Behavior of Topologically Associating Domains. *J Mol Biol,* 427**,** 608-625.

COSTA, F. F. 2008. Non-coding RNAs, epigenetics and complexity. *Gene,* 410**,** 9-17.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

DANIELSEN, A. G., WEISMANN, K. & THOMSEN, H. K. 2001. Erythrokeratolysis hiemalis (keratolytic winter erythema): a case report from Denmark. *JEADV,* 15**,** 255-256.

DAPPRICH, J., FERRIOLA, D., MACKIEWICZ, K., et al. 2016. The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics,* 17**,** 016-2836.

DATHE, K., KJAER, K. W., BREHM, A., et al. 2009. Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am J Hum Genet,* 84**,** 483-492.

DE CASTRO, M. A., BUNT, G. & WOUTERS, F. S. 2016. Cathepsin B launches an apoptotic exit effort upon cell death-associated disruption of lysosomes. *Cell Death Discov,* 2**,** 16012.

DEPRISTO, M. A., BANKS, E., POPLIN, R., et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet,* 43**,** 491-498.

ECKHART, L., LIPPENS, S., TSCHACHLER, E., et al. 2013. Cell death by cornification. *Biochim Biophys Acta,* 12**,** 20.
EDGAR, A. J. 2002. The human L-threonine 3-dehydrogenase gene is an expressed pseudogene. *BMC Genet,* 3**,** 18-18.
EGBERTS, F., HEINRICH, M., JENSEN, J. M., et al. 2004. Cathepsin D is involved in the regulation of transglutaminase 1 and epidermal differentiation. *J Cell Sci,* 117**,** 2295-2307.
ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.
FINDLAY, G. H. & MORRISON, J. G. 1978. Erythrokeratolysis hiemalis--keratolytic winter erythema or 'Oudtshoorn Skin'. A new epidermal genodermatosis with its histological features. *Br J Dermatol* 98**,** 491-495.
FINDLAY, G. H., NURSE, G. T., HEYL, T., 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.
FUCHS, E. & GREEN, H. 1980. Changes in keratin gene expression during terminal differentiation of the keratinocyte. *Cell,* 19**,** 1033-42.
GILBERT, S. F. 2000. The Epidermis and the Origin of Cutaneous Structures. *Developmental Biology.* 6th Edition ed. Sunderland (MA): Sinauer Associates.
GOTTFRIED, I., LANDAU, M., GLASER, F., et al. 2002. A mutation in GJB3 is associated with recessive erythrokeratodermia variabilis (EKV) and leads to defective trafficking of the connexin 31 protein. *Hum Mol Genet,* 11**,** 1311-6.
GRICE, E. A., KONG, H. H., RENAUD, G., et al. 2008. A diversity profile of the human skin microbiota. *Genome Res,* 18**,** 1043-50.
GROENENDAAL, W., VON BASUM, G., SCHMIDT, K. A., et al. 2010. Quantifying the composition of human skin for glucose sensor development. *J Diabetes Sci Technol,* 4**,** 1032-40.
HAUSER, P. J., AGRAWAL, D., HACKNEY, J., et al. 1998. STAT3 activation accompanies keratinocyte differentiation. *Cell Growth Differ,* 9**,** 847-55.
HIROBE, T. 2014. Keratinocytes regulate the function of melanocytes. *Dermatologica Sinica,* 32**,** 200-204.
HOBBS, A., ARON, S., HARTSHORNE, S., et al. 2012. Exclusion of CTSB and FDFT1 as positional and functional candidate genes for keratolytic winter erythema (KWE). *J Dermatol Sci,* 65**,** 58-62.
HOHL, D. 2000. Towards a better classification of erythrokeratodermias. *Br J Dermatol,* 143**,** 1133-1137.
HORIKOSHI, T., ARANY, I., RAJARAMAN, S., et al. 1998. Isoforms of cathepsin D and human epidermal differentiation. *Biochimie,* 80**,** 605-612.
HULL, P. 1986. *Keratolytic Winter Erythema (Oudtshoorn disease): clinical, genetic and ultraspectural aspects.* PhD, University of the Witwatersrand.
HULL, P. R., HOBBS, A., ARON, S., et al. 2013. The elusive gene for keratolytic winter erythema. *S Afr Med J,* 103**,** 961-965.
HUNTINGTON, M. K. & JASSIM, A. D. 2006. Genetic heterogeneity in keratolytic winter erythema (Oudtshoorn skin disease). *Arch Dermato,* 142**,** 1073-1074.
ISHIDA-YAMAMOTO, A., MCGRATH, J. A., LAM, H., et al. 1997. The Molecular Pathology of Progressive Symmetric Erythrokeratoderma: A Frameshift Mutation in the Loricrin Gene and Perturbations in the Cornified Cell Envelope. *Am J Hum Genet,* 61**,** 581-589.
KARADAG, A., BILGILI, S., CALKA, O., et al. 2013. Erythrokeratodermia variabilis: Two case reports. *Indian Dermatol Online J,* 4**,** 340-343.
KATO, T., TAKAI, T., MITSUISHI, K., et al. 2005. Cystatin A inhibits IL-8 production by keratinocytes stimulated with Der p 1 and Der f 1: biochemical skin barrier against mite cysteine proteases. *J Allergy Clin Immunol,* 116**,** 169-76.
KEHRER-SAWATZKI, H. 2007. What a difference copy number variation makes. *Bioessays,* 29**,** 311-3.
KENT, W. J., SUGNET, C. W., FUREY, T. S., et al. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.
KING, I. A., MAGEE, A. I., REES, D. A., et al. 1991. Keratinization is associated with the expression of a new protein related to the desmosomal cadherins DGII/III. *FEBS Lett,* 286**,** 9-12.

KLOPOCKI, E., LOHAN, S., BRANCATI, F., et al. 2011. Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am J Hum Genet,* 88**,** 70-75.
KOUWENHOVEN, E. N., OTI, M., NIEHUES, H., et al. 2015. Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. *EMBO Rep,* 16**,** 863–878.
KRUNIC, A. L., STONE, K. L., SIMPSON, M. A., et al. 2013. Acral peeling skin syndrome resulting from a homozygous nonsense mutation in the CSTA gene encoding cystatin A. *Pediatr Dermatol,* 30**,** e87-e88.
KUNDAJE, A., MEULEMAN, W., ERNST, J., et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-330.
KYPRIOTOU, M., HUBER, M. & HOHL, D. 2012. The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the 'fused genes' family. *Exp Dermatol,* 21**,** 643-9.
LAMB, A., FULLER, M., VARADARAJAN, R., et al. 2012. The Vertica Analytic Database: C-Store 7 Years Later. *Proceedings of the VLDB Endowment,* 5**,** 1790-18021.
LANDT, S. G., MARINOV, G. K., KUNDAJE, A., et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res,* 22**,** 1813-1831.
LATCHMAN, D. S. 2005. *Gene regulation a eukaryotic perspective,* New York; Abingdon, UK, Taylor & Francis.
LEE, Y. & HWANG, K. 2002. Skin thickness of Korean adults. *Surg Radiol Anat,* 24**,** 183-9.
LI, G., FULLWOOD, M. J., XU, H., et al. 2010. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol,* 11**,** 2010-11.
LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics,* 25.
LI, H., HANDSAKER, B., WYSOKER, A., et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.
LI, Y.-Y., FANG, J. & AO, G.-Z. 2016. Cathepsin B and L inhibitors: a patent review (2010 - present). *Expert Opin Ther Pat***,** 1-14.
LOHAN, S., SPIELMANN, M., DOELKEN, S. C., et al. 2014. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin Genet* 86**,** 318-325.
LU, Q., QIU, X., HU, N., et al. 2006. Epigenetics, disease, and therapeutic interventions. *Ageing Res Rev,* 5**,** 449-67.
LUGER, K., MADER, A. W., RICHMOND, R. K., et al. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.
LUPIÁÑEZ, D. G., KRAFT, K., HEINRICH, V., et al. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell,* 161**,** 1012-25.
LUPIÁÑEZ, D. G., SPIELMANN, M. & MUNDLOS, S. 2016. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet. ,* 32**,** 225-237.
MACDONALD, J. R., ZIMAN, R., YUEN, R. K., et al. 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res,* 42**,** 29.
MAHAJAN, V. K., KHATRI, G., CHAUHAN, P. S., et al. 2015. Progressive Symmetric Erythrokeratoderma Having Overlapping Features With Erythrokeratoderma Variabilis and Lesional Hypertrichosis: Is Nomenclature "Erythrokeratoderma Variabilis Progressiva" More Appropriate? *Indian J Dermatol,* 60**,** 410-1.
MAHURKAR, S., IDRIS, M. M., REDDY, D. N., et al. 2006. Association of cathepsin B gene polymorphisms with tropical calcific pancreatitis. *Gut,* 55**,** 1270.
MCKENNA, A., HANNA, M., BANKS, E., et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20**,** 1297-1303.
MCLAREN, W., GIL, L., HUNT, S. E., et al. 2016. The Ensembl Variant Effect Predictor. *Genome Biol,* 17**,** 016-0974.
MERCURIO, C., MINUCCI, S. & PELICCI, P. G. 2010. Histone deacetylases and epigenetic therapies of hematological malignancies. *Pharmacol Res,* 62**,** 18-34.

MILLER, S., DYKES, D. & POLESKY, H. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res,* 16**,** 1215.
MINNER, F. & POUMAY, Y. 2009. Candidate Housekeeping Genes Require Evaluation before their Selection for Studies of Human Epidermal Keratinocytes. *J Invest Dermatol,* 129**,** 770-773.
MOLL, R., DIVO, M. & LANGBEIN, L. 2008. The human keratins: biology and pathology. *Histochem Cell Biol,* 129**,** 705-733.
MORRISON, K. M., MIESEGAES, G. R., LUMPKIN, E. A., et al. 2009. Mammalian Merkel cells are descended from the epidermal lineage. *Dev Biol.,* 336**,** 76-83.
NAGLER, D. K., STORER, A. C., PORTARO, F. C., et al. 1997. Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry,* 36**,** 12608-15.
NAKATANI, M., MAKSIMOVIC, S., BABA, Y., et al. 2015. Mechanotransduction in epidermal Merkel cells. *Pflugers Arch,* 467**,** 101-108.
NEMES, Z. & STEINERT, P. M. 1999. Bricks and mortar of the epidermal barrier. *Exp Mol Med,* 31**,** 5-19.
NESTLE, F. O., DI MEGLIO, P., QIN, J.-Z., et al. 2009. Skin immune sentinels in health and disease. *Nat Rev Immunol,* 9**,** 679-691.
NGCUNGCU, T., OTI, M., SITEK, J. C., et al. 2017. Duplicated enhancer region increases expression of *CTSB* and segregates with Keratolytic Winter Erythema in South African and Norwegian families *Am J Hum Genet,* 100**,** 737-750.
NISHIMURA, F., NARUISHI, H., NARUISHI, K., et al. 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.
NITHYA, S., RADHIKA, T. & JEDDY, N. 2015. Loricrin – an overview. *JOMFP,* 19**,** 64-68.
OSMAN, O. S., SELWAY, J. L., HARIKUMAR, P. E., et al. 2013. A novel method to assess collagen architecture in skin. *BMC Bioinf,* 14**,** 1471-2105.
OTI, M., FALCK, J., HUYNEN, M. A., et al. 2016. CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics,* 17**,** 252.
PALUNGWACHIRA, P., KAKUTA, M., YAMAZAKI, M., et al. 2002. Immunohistochemical localization of cathepsin L and cystatin A in normal skin and skin tumors. *J Dermatol,* 29**,** 573-9.
PAVLOVA, A., KRUPA, J. C., MORT, J. S., et al. 2000. Cystatin inhibition of cathepsin B requires dislocation of the proteinase occluding loop. Demonstration by release of loop anchoring through mutation of His110. *FEBS Letters,* 487**,** 156-160.
PONTIGGIA, L., BIEDERMANN, T., MEULI, M., et al. 2008. Markers to Evaluate the Quality and Self-Renewing Potential of Engineered Human Skin Substitutes In Vitro and after Transplantation. *J Invest Dermatol,* 129**,** 480-490.
R DEVELOPMENT CORE TEAM 2008. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
RADMAN-LIVAJA, M. & RANDO, O. J. 2010. Nucleosome positioning: How is it established, and why does it matter? *Dev Biol.,* 339**,** 258-266.
RAO, S. S., HUNTLEY, M. H., DURAND, N. C., et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell,* 18**,** 1665-1680.
REDON, R., ISHIKAWA, S., FITCH, K. R., et al. 2006. Global variation in copy number in the human genome. *Nature,* 444**,** 444-454.
REINHECKEL, T., HAGEMANN, S., DOLLWET-MACK, S., et al. 2005. The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. *J Cell Sci,* 118**,** 3387-3395.
RICHARD, G., BROWN, N., ROUAN, F., et al. 2003. Genetic heterogeneity in erythrokeratodermia variabilis: novel mutations in the connexin gene GJB4 (Cx30.3) and genotype-phenotype correlations. *J Invest Dermatol,* 120**,** 601-609.
RICHARD, G., BROWN, N., SMITH, L. E., et al. 2000. The spectrum of mutations in erythrokeratodermias--novel and de novo mutations in GJB3. *Hum Genet,* 106**,** 321-9.

RICKMAN, L., SIMRAK, D., STEVENS, H. P., et al. 1999. N-terminal deletion in a desmosomal cadherin causes the autosomal dominant skin disease striate palmoplantar keratoderma. *Hum Mol Genet,* 8**,** 971-6.
ROADMAP EPIGENOMICS CONSORTIUM, KUNDAJE, A., MEULEMAN, W., et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-330.
ROTH, W., DEUSSING, J., BOTCHKAREV, V. A., et al. 2000. Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and pertubation of hair follicle cycling. *FASEB J,* 14**,** 2075-2086.
ROWAN, A. D., FENG, R., KONISHI, Y., et al. 1993. Demonstration by electrospray mass spectrometry that the peptidyldipeptidase activity of cathepsin B is capable of rat cathepsin B C-terminal processing. *Biochem J,* 294**,** 923-7.
SAAGER, R. B., BALU, M., CROSIGNANI, V., et al. 2015. In vivo measurements of cutaneous melanin across spatial scales: using multiphoton microscopy and spatial frequency domain spectroscopy. *J Biomed Opt,* 20**,** 066005.
SANDBY-MOLLER, J., POULSEN, T. & WULF, H. C. 2003. Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits. *Acta Derm Venereol,* 83**,** 410-3.
SCHWARZ, G., BOEHNCKE, W. H., BRAUN, M., et al. 2002. Cathepsin S activity is detectable in human keratinocytes and is selectively upregulated upon stimulation with interferon-gamma. *J Invest Dermatol,* 119**,** 44-49.
SMITH, C., ZHU, K., MERRITT, A., et al. 2004. Regulation of desmocollin gene expression in the epidermis: CCAAT/enhancer-binding proteins modulate early and late events in keratinocyte differentiation. *Biochem J,* 380**,** 757-765.
SPIELMANN, M. & MUNDLOS, S. 2013. Structural variations, the regulatory landscape of the genome and their alteration in human disease. *Bioessays,* 35**,** 533-43.
STARFIELD, M., HENNIES, H. C., JUNG, T., et al. 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.
STEINER, L. A., SCHULZ, V., MAKISMOVA, Y., et al. 2016. CTCF and CohesinSA-1 Mark Active Promoters and Boundaries of Repressive Chromatin Domains in Primary Human Erythroid Cells. *PLoS One,* 11.
SZKLARCZYK, D., FRANCESCHINI, A., WYDER, S., et al. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res,* 43**,** 28.
THE 1000 GENOMES PROJECT CONSORTIUM 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.
THOLEN, S., BINIOSSEK, M. L., GANSZ, M., et al. 2013. Deletion of cysteine cathepsins B or L yields differential impacts on murine skin proteome and degradome. *Mol Cell Proteomics,* 12**,** 611-25.
TOOMES, C., JAMES, J., WOOD, A. J., et al. 1999. Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. *Nat Genet,* 23**,** 421-424.
TURK, V., STOKA, V., VASILJEVA, O., et al. 2012. Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim Biophys Acta, Proteins Proteomics,* 1824**,** 68-88.
UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., et al. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40**,** 22.
VAN STEENSEL, M. A., ORANJE, A. P., VAN DER SCHROEFF, J. G., et al. 2009. The missense mutation G12D in connexin 30.3 can cause both erythrokeratodermia variabilis of Mendes da Costa and progressive symmetric erythrokeratodermia of Gottron. *Am J Med Genet,* 15**,** 657-61.
WALLACE, L., ROBERTS-THOMPSON, L. & REICHELT, J. 2012. Deletion of K1/K10 does not impair epidermal stratification but affects desmosomal structure and nuclear integrity. *J Cell Sci,* 125**,** 1750-8.
WANG, J., ZHUANG, J., IYER, S., et al. 2013. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res,* 41**,** D171-6.
WHITTON, J. T. & EVERALL, J. D. 1973. The thickness of the epidermis. *Br J Dermatol,* 89**,** 467-76.

WICKETT, R. R. & VISSCHER, M. O. 2006. Structure and function of the epidermal barrier. *Am J Infect Control,* 34**,** S98-S110.
WIKRAMANAYAKE, T. C., STOJADINOVIC, O. & TOMIC-CANIC, M. 2014. Epidermal Differentiation in Barrier Maintenance and Wound Healing. *Adv Wound Care (New Rochelle),* 3**,** 272-280.
WILGOSS, A., LEIGH, I. M., KELSELL, D. P., et al. 1999. Identification of a Novel Mutation R42P in the Gap Junction Protein β-3 Associated with Autosomal Dominant Erythrokeratoderma Variabilis. *J Invest Dermatol,* 113**,** 1119-1122.
WINER, J., JUNG, C. K., SHACKEL, I., et al. 1999. Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. *Anal Biochem,* 270**,** 41-9.
XIE, J., YAO, B., HAN, Y., et al. 2016. Skin appendage-derived stem cells: cell biology and potential for wound repair. *Burns Trauma,* 4.
YANG, W.-E., HO, C.-C., YANG, S.-F., et al. 2016. Cathepsin B Expression and the Correlation with Clinical Aspects of Oral Squamous Cell Carcinoma. *PLoS One,* 11**,** e0152165.
YE, K., SCHULZ, M. H., LONG, Q., et al. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics,* 25**,** 2865-2871.
ZEEUWEN, P. L. 2004. Epidermal differentiation: the role of proteases and their inhibitors. *Eur J Cell Biol,* 83**,** 761-73.
ZEEUWEN, P. L., VAN VLIJMEN-WILLEMS, I. M., HENDRIKS, W., et al. 2002. A null mutation in the cystatin M/E gene of ichq mice causes juvenile lethality and defects in epidermal cornification. *Hum Mol Genet,* 11**,** 2867-75.
ZHU, J., ADLI, M., ZOU, JAMES Y., et al. 2013. Genome-wide Chromatin State Transitions Associated with Developmental and Environmental Cues. *Cell,* 152**,** 642-654.
ZHU, S., OH, H.-S., SHIM, M., et al. 1999. C/EBPβ Modulates the Early Events of Keratinocyte Differentiation Involving Growth Arrest and Keratin 1 and Keratin 10 Expression. *Molecular and Cellular Biology,* 19**,** 7181-7190.

Web Resources

1000 Genomes, http://browser.1000genomes.org

3D Genome Browser at http://www.3dgenome.org.

OMIM, http://www.omim.org/

dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP

DGV, http://dgv.tcag.ca/dgv/app/home

Ensembl VEP, http://grch37.ensembl.org/Homo_sapiens/Tools/VEP

Primer3 (v 0.4.0), http://frodo.wi.mit.edu/

SPSmart ENGINES v5.1.1 – http://spsmart.cesga.es/engines.php

UCSC Genome Browser, http://genome.ucsc.edu

Appendices

Appendix A: Ethics Certificate



R14/49 Ms Thandiswa Ngcungcu et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M140530

NAME: (Principal Investigator)	Ms Thandiswa Ngcungcu et al
DEPARTMENT:	School of Pathology Dvisision of Human Genetics National Health Laboratory Services
PROJECT TITLE:	The Identification and Characterisation of the Causative Gene Mutation for Keratolytic Winter Erythema (KWE) in South African Families (BEC20140502)
DATE CONSIDERED:	30/05/2014
DECISION:	Approved unconditionally
CONDITIONS:	Application for Collection and Storage of Biological Samples at Sydney Brenner Institute for Molecular Bioscience Biobank
SUPERVISOR:	Prof Michele Ramsay

APPROVED BY:

Professor A Woodiwiss, Chairperson, HREC (Medical)

Prof A Dhai, Chairperson, BEC

DATE OF APPROVAL: 21/08/2014

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature _____ Date _____

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B: Gentra Puregene DNA purification protocol for buccal swabs

Sample Collection and Handling

1. Pipette 300µl Cell Lysis Solution (Puregene Kit) into a 1.5 ml centrifuge tube.
2. Scrape inside of mouth with provided sterile nylon bristle cytology brush (~ 10 strokes).
3. Place brush into 1.5 ml centrifuge tube containing 300 µl Cell Lysis Solution and swirl gently to facilitate removal of lysate from the collection brush head into the solution. Remove brush from Cell Lysis Solution, scraping the brush on the sides of the tube.
4. Proceed with DNA purification or store sample (Samples are stable in Cell Lysis Solution for at least 2 years at room temperature).

DNA Purification from buccal brush using Puregene™ DNA purification kit
(Expected yield: 0.2µg DNA)

Step 1: Cell Lysis

1. Incubate sample @ 65°C for 15-60 minutes. For maximum yield add 3µl Proteinase K Solution (10mg/ml) to the cell lysate, mix by inverting 25 times, incubate @ 55°C for 1 hour or overnight.

Step 2: RNase treatment

1. Add 1.5µl RNase A Solution (Puregene kit) to the cell lysate in the microfuge tube.
2. Invert the tube ~25 times and incubate @ 37°C for 15-60 minutes.

Step 3: Protein Precipitation

1. Cool sample to room temperature by placing on ice for 1 minute.
2. Add 100µl Protein Precipitation Solution (Puregene kit) to the sample.
3. Vortex vigorously at high speed (20 sec) to ensure uniform mixing of the Protein Precipitation Solution and the sample.
4. Place tube on ice for 5 minutes.
5. Centrifuge @ 13 000 – 16 000 x g for 3 minutes. The protein should precipitate and form a tight white pellet. If pellet is not tight, repeat steps 3 – 5.

Step 4: DNA Precipitation

1. Pour the supernatant (containing the DNA) into a clean 1.5 ml microfuge tube containing 300µl 100% Isopropanol (2-propanol) and 0.5µl Glycogen Solution (20mg/ml, Genta® catalog number R-5010). (The pellet that is left behind in the first tube contains the proteins and can be discarded).
2. Mix by inverting gently 50 times and incubate at room temperature for at least 5 minutes.
3. Centrifuge at 12 000 x g for 5 minutes. (In the case of high yield, the DNA may be visible as a small white pellet).

4. Pour off supernatant and drain tube on clean paper towel. Add 300µl 70% Ethanol and invert tube several times to wash DNA.
5. Centrifuge at 13000 – 16 000 x g for 1 minute. Carefully pour off supernatant. (The pellet might be loose, therefore care must be taken not to pour the pellet out of the tube)
6. Invert tube and drain on paper towel. Allow to air dry 10 – 15 minutes

Step 5: DNA hydration

1. Add 20 µl DNA Hydration Solution (Puregene kit) (20µl will give a concentration of 50ng/µl if the yield is 1µg DNA)
2. Rehydrate DNA by incubating sample for 1 hour at 65°C or overnight at room temperature. Tube must be tapped periodically to aid dispersing of DNA
3. Store DNA at 4°C (For long term storage sample must be placed at -20°C / -70°C).

DNA Concentration and Quality Assessment

1. The DNA concentration is determined by spectrophotometry using the Nanodrop ND-1000.

Sample Storage

1. The sample is stored at 4°C until required.

References:

Puregene™ DNA purification kit manual

Appendix C: Per base sequence quality

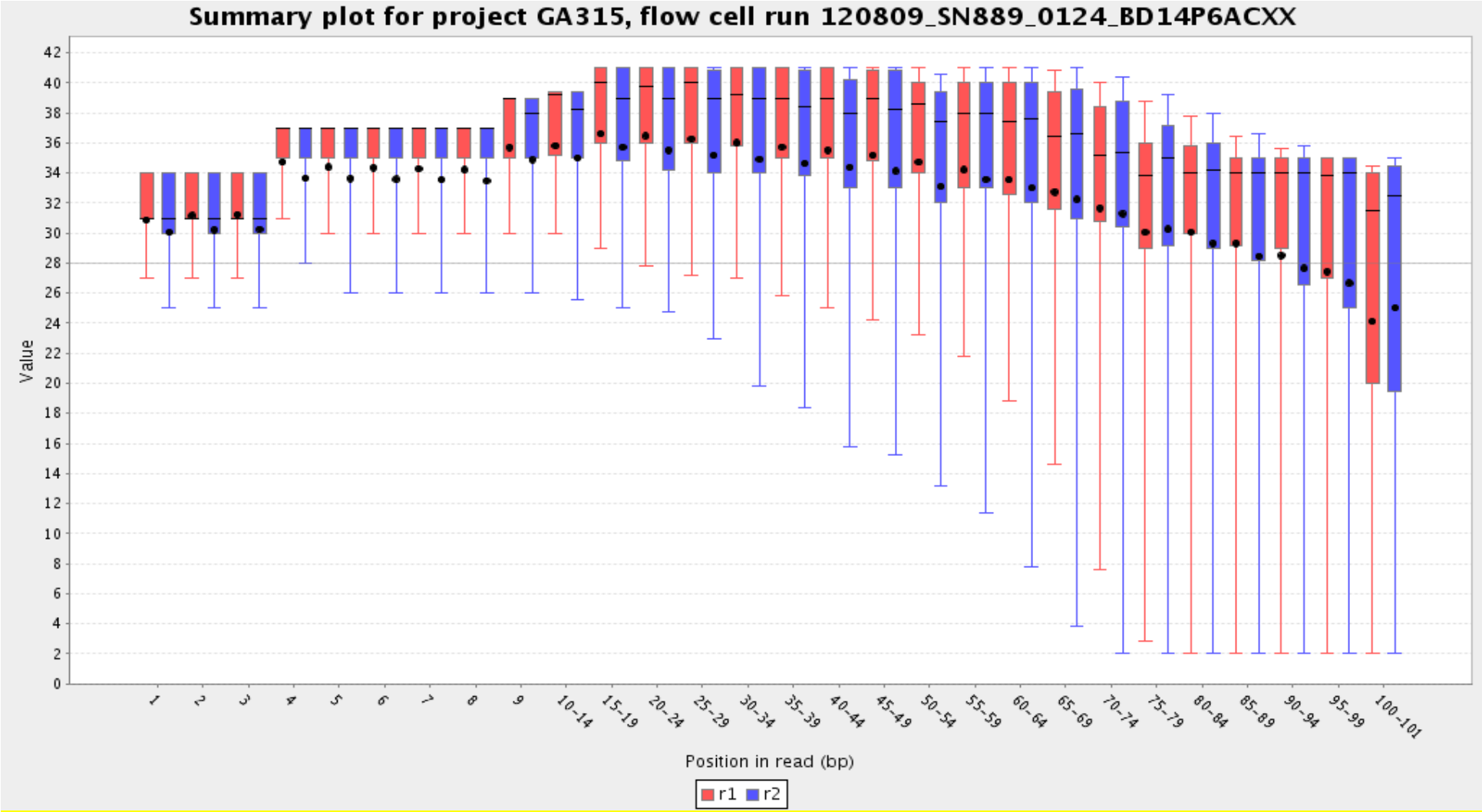


Figure C2: Per base sequence quality: Targeted Capture Sequencing

Appendix D: Pindel programming

Using sequence reads from the targeted linkage region, BWA was used to align the reads using "mem" method. The BAM files resulting from the alignment were passed through the version of Pindel downloaded from the Pindel github site (https://github.com/genome/pindel) on August 26, 2014 (commit da6a1ec336fa7159bce0813eb48c19ba76d40116). The options (-T 2 -s -x 6 --report_long_insertions --report_breakpoints --IndelCorrection) were used. The output results were parsed into individual components for each structural variant. These components were loaded into a relational database that included de-identified information about each patient. Simple SQL queries were issued against the database that counted the number of times each variant occurred grouped by disease status.

Appendix E: Protocol for validating the tandem duplication

E1. Primer design

A. Tandem duplication junction sequence

PCR primers were for the tandem duplication junction sequence (TDJS) and the control primer sets were designed using Primer 3 (version 0.4.0) software (Untergasser et al., 2012). The primer sets and PCR product size for the junction sequence amplicon are shown in Table E1. The Primer 3 input sequence was created by merging ~1200 bp of sequence from the start of the duplicated regions and ~1200 bp before the end of the duplicated region to create a 2328 bp sequence with the junction of the two ends of the duplicated region, indicated by square bracket. The sequence starts with bases from the end of the duplication and this is merged with the start of the duplication thereby creating the junction sequence seen in individuals with the duplication (the sequence is shown below). DNA sequences were retrieved in FASTA format from NCBI using the coordinates (GRCh37/hg19) of the duplicated region. UCSC *In-Silico* PCR was used to determine if the primer pairs would yield a PCR product using the default settings. Because these primers are far away from each other on the normal chromosome, no PCR product was predicted by the program as expected (Kent et al., 2002). The junction sequence of the duplicated region is unique and would only be present in individuals with the duplication.

Table E1: PCR primer sets and PCR product sizes for the junction sequence amplicon

Amplicon	Direction	Primer sequence	PCR product size (kb)
Junction sequence amplicon	Forward	5’ CTAGGCTTGCAGTGTTGGTC 3’	375
	Reverse	5’ GTTAAATCAGGCTGGGCGAG 3’	
Control	Forward	5’ AGCCAAGGCAAAATTGAGG 3’	250
	Reverse	5’ TCCAGCCGATCTCTCTGTTC 3’	

Input sequence for Primer3

GAATCATAGTGAGTTTAGACTTTTCTGGAGGCTTAGGGGTAGGCAATGAGGATGAGGAGG
GGGTGACCAAGAGATAGAGGAGAGATTTACTTGGTAGAAATGCATCCAGAACACAGGCAT
TGCCTTTCAGTTAACGTTCTTTGGTGACCCTTCCAGGCTAGATTTCATTTCACTCTGAAC
TTCCCATCTTTTTGGTTCCCTTTCGCCCTCATTCTCAGATTTCTCTCCATTCCTTTCAGT
GGCAAAACAACTCATTTCTTCCCTGTTCCCCCAAGAACAGCCTCCCTCATGATCTGCAGG
CTCAGAGCTGACTTTCGTCCCACAAGATGTCATCTTCCCCAACCAAGTCTCTCTACGCTA
AAGATAGTTCCTTCACTGGTTAGAGTTCTGGGGCCAGCTGTTCTCTCTCTTTTTAATTTC
ACTTCTCTAGAAAAGCTTCTTAAGAAATTGCAACCTTGCTTAGCTCAACGGAGGGGGTCA
GGGCAGTACTGCTGGTGGGGTCATGTGGGCCTGAACACCCCATCTTCTTGTGGCTTCCAT
GGTTCCCGATATGCCTATAGGAAGCAGACGGGCTGGCCAAAGGGGCGCCCAGCAGGGAGA
TAATGGGCGTGGGGTGGGCCCCTTTTTCTCTGTTGGCCTGGTCTGGTGCTGGTGTCCCAT
AAGCTCTCCCAGACGAATAACCTTGTTCTACAAACGTACCCACTTAGGCAATTTGCCCTA
AAGTTTCATCTTAAAAATGCATGTGAAATTGGACTTGTACTCCAGAGATATCCATGTTTG
TATTCATGTAAAAATAATGTCCTTCTTAATTATCTGGGGGTGGTGGTGTGTGCCTTTAGT
GCCAGCTACTTGGAAGGCTGAGGCAGGAGAATCACTTGGACCAAGGAGGCAGAGGTTGCA
GTGAGCTGAGATCGCGCCATTGCACTCCAGCCTGGGTGACAGAGAGAGACTCTGTCCCAA
AAAATAAAATAAAATAAAAATAAATACATAAAATAAAATAAAATAAAAGTCCTTCTCTAG
GCTTGCAGTGTTGGTCTATTATTTTGCCTCCTGTGTGCCAGGCACTGCAGCAGGCAGTGG
GAACTGAATGTCAGCCCACCTGGAATGGAACTTGGGTCTAGGTTGGAG[AA]GACAGGGT
TTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCGTGATCCACC
TGCCTCGGCCCCCCAAAGTGCTGGGATTACAGGCGTGAGCCACAGGGCCCAGCCTAATTT
TGTATTTTTTGTAGAGACGGTGTTTTCCCATGTTGCCCAGGCTGATCTCAAGCTCTTGAG
CTCAAATGATCCTGCTGTCTCCGCCTCTCAAAGTGCTAGGATTACAGGTGTGAGCCACCT
CGCCCAGCCTGATTTAACTTTTTAATTTTTAAAAATTATCCTTGAATACGCTACAGTGCC
CAACTATTCAGTGTTGAAATGAACTGGATTTCTTATTTGGCTGTGATTGTTTCCAGTATT
TTCAGAAATATGTCTCCATTCCTATCTAAACTCTAAGTTGGTTGGTGTAGGGTGCTGTCT
TATTCATTTATATTTCTCCTTCCTCTCTTCTCAAAGGACTTTCGTGCCCTTTACCCCTCC
TTCCAGCAGGGAGGTATGAGCCGTGCCTCTGGGCTCCTTCTTCTTCCTCCCGACAGCTCC
CTTTCTGTGGTGTGACTGTCATCAGAGGGAGGGAAATTACAGTGCCTTGGCTTCAGAGAA
CTAAGTTTCTGGTTCTAGCCTTATGAGGGCTGACCACATTGTTTTTTTTTTTTTTCTATT
GGCTTCCAGAAGGATTCCTCTTTATTCTCCTACAGTGTCTCCTTTGTGGCTTATCCTACT
GCGAGAGCACTTGTTTAACTGCGAGCAAAAGAGATGGCTGCTCCTCTAATACCGGTGGGG
GCCTCGACTTTCACATGAACTCCCTCGGAAGGGAAGCCATGTCTGCAGCCAGGGTCTCAG
CAGATGTCCACAGCCCCCAAAGCTTTGCCTCCAGAGGCCACCACTTCCCACCATTTCTGT
TTCTGGTGGGGCCCCTTCAATTCTAACTATGGCCTCCCCACCACAGAAGAGCAGGGTTTA
CTCACTTGTCCTCTTCCATGAAGAGCAAATGTATTATCTCACTTGGGGTTTCAGTTGCCC
TTGTAGTTTTATTTATTTTTATTTTTCATTTATTTAATTAAAAAAAAAGTTTGAGACAGG
GTCTCGCTCTGTCACCCAGGCTGGAGTGCAGTGGCAGGATCTTGGCTCACGGCAACCTTC
ACCTCCCGGGCTCAAGCAATCCTCCCACCTCAGCCTCTGGACTACCTGGGATCACAGACG

B. Primer design: Control primer set

A control PCR primer set (primers from the *LEP* gene) was included in the PCR reaction to determine PCR efficiency. The control primer set serves to show that the lack of the amplification of the TDJS fragment is because the junction is absent and not because the PCR failed, since control amplicon is expected in all individuals. PCR product sequence and size were determined using UCSC In-Silico PCR (Kent et al., 2002) (Figure E1).

A.

GAATTAGCTGCAAGGGGAACTAGGAAAAGCTTCTTTAAGGATGGAGAGGCCCTAGTGGAATGGGGAGA
TTCTTCCGGGAGAAGYGATGGATGCRCAGTTGGGCATCCCCACAGACRGACTGGAAAGAAAAAAGGCCT
GGARGAATCAATGTGCAAYGYATGTGTGTTCCCTGGTTCAAGGGCTGGGAACTTTCTCTAMAGGGCCAG
GTAGAAAACATTTTAGGCTTTCTAAGCCAAGGCAAAATTGAGGATATTACRTGGGTASTTATACAACAAR
AATAAACAATTTACACAATTTTTTGTTGACAGAATTCAAAACTTTATAGACACAGAAATGCAAATTTCCTG
TAATTTTYCCRTGAGAACTATTCTTCTTTTGTTTTGTTTTGCGACAGGGTTGY**[R]**CTGATCCTCYYGCCTCA
GTCTCCCTAAGTGCTGAGATGTTGCAGGAAGTCAGGGACCCCGAACAGAGAGATMRGCTGGAGCCRTG
GCAGAGGAACATAAATTTTGAAGATKTCATTTTAATATGGACACTTWTCAGTTCCCAAATAATAYTTTTAT
AATTTTTTATGCCTGTCTTTGCTTTAATCTCTTAATCCTGTTATCTYCATAAGCTAAGGATGTACRTCACCTC
AGGACCACTGTGATAATTGTGTTAACTGTACAGATTGAYTGCAAAACATGTGTGTTTGAACAATATGAAA
TCAGTGCACCTTGAAAAAGAGCAGAATAACAGCAATTTTTAGGGAACAAGGGAAGACAACTATAAGGTC
TGACTGCCTGCRGGGTCGGGCAAAGGGAGCCA

B.

Forward and reverse primers highlighted in yellow and blue respectively

>chr7:127878613+127878862 250bp AGCCAAGGCAAAATTGAGG TCCAGCCGATCTCTCTGTTC
AGCCAAGGCAAAATTGAGGatattacatgggtacttatacaacaagaata
aacaatttacacaatttttgttgacagaattcaaaactttatagacaca
gaaatgcaaatttcctgtaattttcccatgagaactattcttctttgtt
ttgttttgcgacagggttgcgctgatcctcccgcctcagtctccctaagt
gctgagatgttgcaggaagtcagggacccGAACAGAGAGATCGGCTGGA

Primer melting temperatures

Forward: 60.2 C agccaaggcaaaattgagg
Reverse: 60.5 C tccagccgatctctctgttc

Figure E1: rs7799039 control flanking sequence (A) and UCSC in-silico PCR output (B).

E2. Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify DNA fragments that encompass the tandem duplication junction sequence and the control fragment. PCR reaction mixtures are shown in Table E2 and PCR conditions are shown in Table E3.

Table E2: Concentrations and volumes of reagents for PCR reactions

Reagent	Concentration	Volume/ sample
KAPA Taq ReadyMix	2X	25
TD_R Primer	10 µM	2
TD_L Primer	10 µM	2
Con_R Primer	10 µM	2
Con_L Primer	10 µM	2
ddH_2O	-	15
DNA	100ng/µl	2
Total	-	50

Table E3: PCR run conditions

Step	Temperature (°C)	Time	
Initialisation	95	3 min	
Denaturation	95	30 sec	34 cycles
Annealing	54	30 sec	
Extension	72	30 sec	
Final extension	72	5 min	
Hold	4	∞	

Gel electrophoresis on a 3% agarose gel was used to determine if the PCR reaction was successful, check if the correct DNA fragment (according to size) was amplified and to check for contamination. PCR products were loaded to the solidified 3% agarose gels along with Fermentas GeneRuler™ 50 bp DNA ladder by mixing 3 µl Ficoll loading dye with 2 µl of the PCR product. The gels were run at 5V/cm for 45 min on Sigma-Aldrich gel tanks and analysed using Syngene's G-Box and GeneSnap (version 6.08) gel documentation system.

E3: Post PCR Sequencing protocol

E3.1. Post PCR clean-up

- Add water to each tube so that the final volume is 100 µl.
- Transfer all the tubes contents to a clean PCR Clean-up Filter Plates than vacuum until dry (approximately 10 minutes).
- Blot the bottom of the plate with a paper towel.
- Add 100 µl of ddH2O to each well then vacuum until dry.
- Blot the bottom of the plate with a paper towel.
- Add 25 µl of ddH2O to each well then place the tubes on a plate shaker for 10 minutes at low speed
- Move the contents of the well, now containing clean PCR product into clean tubes then proceed to cycle sequencing.

E3.2. Cycle sequencing

- Set up the cycle sequence reaction.
  - One tube for forward and one tube for reverse for each PCR product.
  - Remember product is light sensitive
- Cycle sequencing PCR mix (1/8th reaction)

2 µl	Cleaned PCR product
1 µl	Bigdye v3.1 (in -20°C freezer with Taq)
1.5 µl	5x Bigdye buffer
1 µl	Primer (forward OR reverse)
4.5 µl	ddH2O
10 µl	

NB: If sequencing reaction fails either dilute PCR product 1/10 or add 1u PCR product and 2 µl Big Dye.

- Cycle sequencing PCR program:

30 sec	@	96ºC	25 cycles
15 sec	@	50ºC	
4 min	@	60ºC	
Hold	@	24ºC	

E3.1. Post PCR clean-up

- Add 40 µl injection solution to each tube so that the final volume is 50 µl.
- Transfer all the tubes contents to a clean PCR Clean-up Filter Plates than vacuum until dry (approximately 10 minutes).
- Blot the bottom of the plate with a paper towel.
- Add 25 µl of injection solution to each well then vacuum until dry.
- Blot the bottom of the plate with a paper towel.

- Add 10 µl of ddH2O to each well then place the tubes on a plate shaker for five-seven minutes at low speed. The cycle sequencing is ligh sensive
- Move the contents of the well, now containing clean PCR product into clean tubes then proceed to cycle sequencing.

E4. Sequencing run

- After denaturing and chilling plate, samples were sequenced on the Applied Biosystems 3130xl Genetic Analyzer sequencer using default setting and selecting Z_Seq POP7-36 Ultra (for shorter sequences 400-500 bp) for instrument protocols.

Appendix F: *In vivo* studies

F1: RNA Extraction Protocol

1. Tissue homogenisation

1.1. The TissueLyser II was used to homogenise the epidermal tissue.

1.2. Remove RNA*later* stabilized tissues from the reagent using forceps. Determine the amount of tissue. Do not use more than 30 mg.

1.3. Place the tissues in 2 ml microcentrifuge tube containing one stainless steel bead (3–7 mm mean diameter).

1.4. Place the tubes at room temperature (15–25°C) then add 600 µl Buffer RLT to each tube (add ME to the Buffer RLT before use).

1.5. Place the tubes in the TissueLyser Adapter Set 2 x 24.

1.6. Operate the TissueLyser for 2 min at 30 Hz, this was done three time. Disassemble the adapter set, rotate the rack of tubes so that the tubes nearest to the TissueLyser are now outermost, and reassemble the adapter set. Operate the TissueLyser for another 2 min at 20–30 Hz.

2. RNA extraction

2.1. Total RNA was extracted using the RNeasy® Mini kit

2.2. Centrifuge the lysate for 3 min at full speed. Carefully remove the supernatant by pipetting, and transfer it to a new microcentrifuge tube (not supplied). Use only this supernatant (lysate) in subsequent steps. In some preparations, very small amounts of insoluble material will be present after the 3 min centrifugation, making the pellet invisible.

2.3. Add 1 volume of 70% ethanol to the cleared lysate, and mix immediately by pipetting. Do not centrifuge. The volume of lysate may be less than 350 µl or 600 µl due to loss during homogenization and centrifugation in the previous steps. Precipitates may be visible after addition of ethanol. This does not affect the procedure.

2.4. Transfer up to 700 µl of the sample, including any precipitate that may have formed, to an RNeasy spin column placed in a 2 ml collection tube (supplied). Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10,000 rpm).

2.5. 2.4. Discard the flow-through. Reuse the collection tube in step 6.If the sample volume exceeds 700 µl, centrifuge successive aliquots in the same RNeasy spin column. Discard the flow-through after each centrifugation.

2.6. Add 700 µl Buffer RW1 to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10,000 rpm) to wash the spin column membrane.

2.7. Discard the flow-through. Reuse the collection tube in step 2.8. After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Be sure to empty the collection tube completely.

2.8. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10,000 rpm) to wash the spin column membrane. Discard the flow-through and reuse the collection tube in step 2.8. (Buffer RPE is supplied as a concentrate. Ensure that ethanol is added to Buffer RPE before use).

2.9. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently, and centrifuge for 2min at ≥8000 x *g* (≥10,000 rpm) to wash the spin column membrane. Note: After

centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Otherwise, carryover of ethanol will occur.

2.10. Place the RNeasy spin column in a new 2 ml collection tube (supplied), and discard the old collection tube with the flow-through. Close the lid gently, and centrifuge at full speed for 1 min. Perform this step to eliminate any possible carryover of Buffer RPE, or if residual flow-through remains on the outside of the RNeasy spin column after step 2.11.

2.11. Place the RNeasy spin column in a new 1.5 ml collection tube (supplied). Add 30–50 µl RNase-free water directly to the spin column membrane. Close the lid gently, and centrifuge for 1 min at ≥8000 x *g* (≥10,000 rpm) to elute the RNA.

2.12. Perform quality (Table F1) and quantity (Figure F1) check then store the RNA at -80°C.

3. Check RNA Quality and quantity

3.1. The RNA quantity was determined using the NanoDrop ND-1000

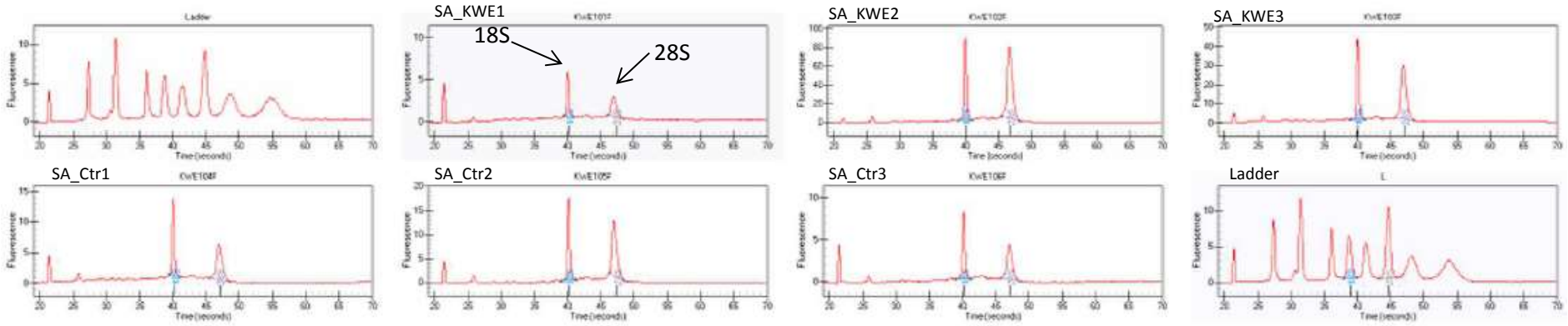
Table F1: RNA quantity and quality for all extracted samples

Sample ID	ng/ul	260/280	260/230	Total RNA (ug)
NOR _Ctr1	51	2,09	1,54	1,3
NOR _Ctr2	97	2,11	1,04	2,9
NOR _Ctr3	128	2,09	1,58	3,8
NOR _Ctr4	121	2,10	0,8	3,6
OSL_DII-1	208	2,10	1,89	6,3
OSL_DII-6	129	2,09	1,89	3,9
OSL_DIII-6	136	2,12	0,85	4,1
OSL_EI-2	92	2,08	1,32	2,8
SA_KWE1	13,5	2,07	0,12	0,40
SA_KWE2	143,1	2,11	1,32	4,29
SA_KWE3	80,4	2,08	0,62	2,41
SA_Ctr1	29,8	2,04	0,61	0,89
SA _Ctr2	39,9	2,12	0,73	1,20
SA _Ctr3	18,9	1,95	0,39	0,57

3.2. The RNA quanlity was determined using the Experion™

Total RNA was determined using the automated electrophoresis Experion™ station.

A. South African samples



B. Norwegian samples

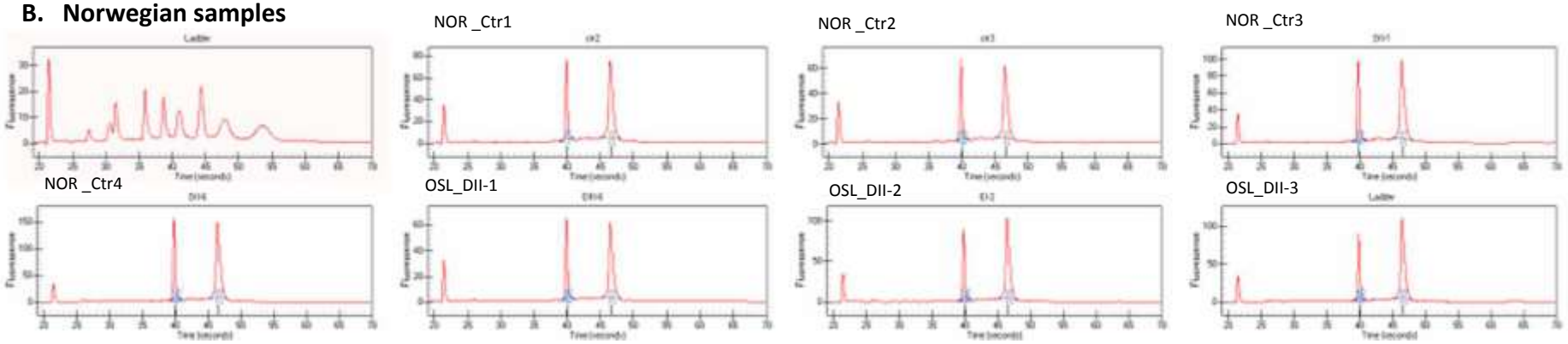


Figure F1: Quality check of the extracted RNA showing the ribosomal 18S and 28S subunits along with the ladder. Sample SA_KWE1 showed the lowest quality. There was evidence of a degree of degradation of the South African samples (lower peaks, particularly for the 28S peak), but the samples were still usable. The x axis is time (seconds) and the y axis is the fluorescence.

F2. First strand cDNA synthesis

1.1. For a single reaction, combine the following components in a tube on ice. Include a negative control.

Table F2: Component and volumes for first strand cDNA synthesis

Component	Quantity
5X VILO™ Reaction Mix	5 µL
10X SuperScript™ Enzyme Mix	2.5 µL
RNA	250 ng
DEPC-treated water*	x µL
Total	25 µL

***Calculated per sample based on the volume of input RNA (see Table F3)**

Table F3.Template dilution for first strand cDNA synthesis

Sample ID	ng/ul	Template 250 ng	ddH2O
NOR _Ctr1	51	4.9	12.6
NOR _Ctr2	97	2.6	14.9
NOR _Ctr3	128	2.0	15.5
NOR _Ctr4	121	2.1	15.4
OSL_DII-1	208	1.2	16.3
OSL_DII-6	129	1.9	15.6
OSL_DIII-6	136	1.8	15.7
OSL_EI-2	92	2.7	14.8
SA_KWE1	13	17.5	0.0
SA_KWE2	143	1.7	15.8
SA_KWE3	80	3.1	14.4
SA_Ctr1	30	8.3	9.2
SA _Ctr2	40	6.3	11.3
SA _Ctr3	19	13.2	4.3

1.2. Gently mix tube contents and incubate at 25°C for 10 minutes.

1.3. Incubate tube at 42°C for 60 minutes.

1.4. Terminate the reaction at 85°C at 5 minutes. 5. Store at –20°C until use.

*Sample NOR _Ctr3 was used for the standard curve and three separate tubes were ran for the first strand cDNA generation. All three samples were later mixed together in one tube. 25 µl of the cDNA was added to 25 µl ddH2O and was used for serial dilutions for the three test genes and the endogenous control. The serial dilution was as follows: Serial dilution: 500, 250, 0125, 62.5, 13.5.

F3: Relative gene expression assays

Work on ice.

Table F4: Reaction mix for gene expression essay for triplicate reactions per sample

PCR reaction mix component	Quantity per triplicate
20xTaqMan® Gene Expression Assay (Probe)	1.5 µL
2x TaqMan® Gene Expression Master Mix	15 µL
cDNA template	3.3 µL
RNase-free water	13.5 µL
Total	3.3 µL

Make enough reaction to include triplicate cDNA negative control and qPCR negative control

- Add 10 µl x3 per sample in a 384 well plate.
- Sample NOR _Ctr3 was used for the standard curve and three separate tubes were ran for the first strand cDNA generation. All three samples were later mixed together in one tube. 25 µl of the cDNA was added to 25 µl ddH2O and was used for serial dilutions for the three test genes and the endogenous control. The seral dilution was run in triplicate for each gene.
- qPCR was run on the Applied Biosystems 7900HT Fast Real-Time PCR System.
- Data was initially processed and viewed using SDS software.

Appendix G: Reagents and solutions

Primer Dilutions

ddH_2O	90 µl
100 µM/primer	10 µl
	100 µl of 10 µM of primer

10X TBE buffer

Tris	108 g
Boric Acid	55 g
EDTA	7.44 g

Filled solution up to 1L, brought buffer to pH 8.0 and autoclaved it.

3% Agarose Gel

10X TBE buffer	40 ml
ddH_2O	360 ml
Agarose	12 g

Added 12µl (3 µl/ 100 ml of gel) EtBr when agarose had melted.

Appendix H

Link to the paper below:

http://www.cell.com/ajhg/abstract/S0002-9297(17)30144-1

Thandiswa Ngcungcu, Martin Oti, Jan C. Sitek, Bjørn I. Haukanes, Bolan Linghu, Robert Bruccoleri,, Tomasz Stokowy, Edward J. Oakeley, Fan Yang, Jiang Zhu, Marc Sultan, Joost Schalkwijk, Ivonne M.J.J. van Vlijmen-Willems, Charlotte von der Lippe, Han G. Brunner, Kari M. Ersland,, Wayne Grayson, Stine Buechmann-Moller, Olav Sundnes, Nanguneri Nirmala, Thomas M. Morgan, Hans van Bokhoven, Vidar M. Steen, Peter R. Hull, Joseph Szustakowski, Frank Staedtler, Huiqing Zhou, Torunn Fiskerstrand, Michèle Ramsay. (2017) Duplicated enhancer region increases expression of CTSB and segregates with Keratolytic Winter Erythema in South African and Norwegian families. American Journal of Human Genetics, 100 (5), 737-750.