Determining the genetic basis of Keratoconus and implications for treatment

By

Layal Fadi Abi Farraj

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Medicine

December, 2015

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

On this page, I take a moment to look back to arduous years of work finally topped with this thesis, which could not have been possible without the help and support of so many great people. I would like to express my sincere appreciation to my supervisors Dr. Manir Ali, Dr. Carmel Toomes and Professor Chris Inglehearn for their constant guidance and support. Their encouragement, generous help and constructive criticism paved the way towards the fulfilment of this thesis.

I would like to extend my gratitude to the immeasurable support and friendship of the staff from the section of ophthalmology and neuroscience, in particular James, Dave, Jose, Ian Carr and Alastair Droop who had put time and effort in helping me accomplish the bioinformatics analysis. I would like also to express my special gratitude to Qais who has offered his help in writing scripts indispensable to this research. I am thankful for the encounter of Dr. Pirro Hysi who, at any moment, had shown willingness to answer my statistical questions.

Many thanks go to the ophthalmologists who helped in recruiting the patients of this study, in particular Salina who is always willing to provide help and clarification on clinical matters. Many thanks go equally to the patients and controls who participated in this study. I would like to acknowledge Fight For Sight for their financial support.

At the heart of this journey remains the presence of my parents. I carry for them immense thankfulness for they have illuminated my way with their love and belief in my potential. I have been surrounded by their continuous support despite the many miles away from home. Finally, I would like to massively thank my siblings and friends for showing support and love in everything I do, especially my partner Mazen whose love, patience, support and positive attitude helped me overcome the tough days.

Abstract

Keratoconus (KC) is a non-inflammatory, progressive thinning of the cornea resulting in a conically shaped protrusion. Its incidence in the population is 1/2000. There are known environmental causes, but familial aggregation, bilateralism and twin studies point to a genetic component, which is poorly understood. This project aims to identify these genetic causes by studying a cohort of familial cases and KC and wild type (WT) corneal tissue.

The family cohort consisted of 15 multiplex recessive KC families from consanguineous and endogamous ethnic backgrounds. DNA from at least one family member was subjected to autozygosity mapping and whole exome sequencing (WES). Tissue from 6 sporadic KC and 5 WT corneas was analysed by RNA-seq. Genomic DNA from the 6 KC patients who donated their corneas was also subjected to WES.

Autozygosity mapping and WES did not identify clear pathogenic alleles, though WES alone yielded long lists of variants segregating in families. An underpowered enrichment test on the WES results also yielded lists of variants potentially implicated in KC. RNA-seq generated a transcriptome profile of the normal anterior human cornea, and highlighted putative new cornea-specific transcripts. The differential expression (DE) data between KC/WT corneas generated lists of significantly DE genes. The combination of all these analyses identified 5 genes, *FLNB, ITGB4, KIAA0100, LAMA5* and *PCDH1,* harbouring coding variants that cosegregated with KC in the families, enriched in variation in the cohort and significantly down-regulated in KC corneas. These are now strong candidates to harbour variants increasing susceptibility to KC.

The work describe shows that recessive alleles of large effect are either rare or non-existent in the cohort studied, suggesting that KC is indeed a genetically complex disease. Combining the datasets highlights variants in 5 large structural proteins requiring further analysis to confirm their involvement in KC pathogenesis.

Table of Contents

List of Tables

List of Figures

Abbreviations

1000g	1000 genome project/database
AFR	African
AGAP3	Ankyrin Repeat And PH Domain 3
ALDH3	Aldehyde dehydrogenase
AMD	Age- Macular degeneration
AMR	Admixed American
ANGPTL7	Angiopoietin-Like 7
ANKRD36BP2	Ankyrin Repeat Domain 36B Pseudogene 2
AS-OCT	Anterior segment optical coherence tomography
AUST	American University of Science and Technology
BiB	Born in Bradford database
bp	Base pair
BSG	Basigin
CADD	Combined Annotation Dependent Depletion
CCL5	Chemokine (C-C Motif) Ligand 5
CCT	Central Corneal Thickness
cDNA	Complementary DNA
CDS	Coding sequence
CHRNA3	Cholinergic Receptor Nicotinic Alpha 3
CNV	Copy Number Variations
COL	Collagen
CXL	Collagen cross-linking
D	Diopter
DALK	Deep anterior lamellar keratoplasty
DAVID	The Database for Annotation Visualization and Integrated Discovery
DE	Differential expression
DHODH	Dihydroorotate Dehydrogenase
DL	Dua's Layer
DM	Descemet's membrane
DNA	deoxyribonucleic acid
ds cDNA	Double stranded complementary DNA
DSEK	descemet stripping automated endothelial keratoplasty
DST	Dystonin
EAS	East Asian
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic acid
ERCC1	DNA excision repair protein
ERN2	Endoplasmic reticulum to nucleus signalling 2
ExAc	The Exome Aggregation Consortium
FAT4	Cadherin-Related Family Member 11
FC	Fold change
FDFT1	Farnesyl-Diphosphate Farnesyltransferase 1
FDR	False Discovery Rate
FIN	Finnish
FLAK	Femtosecond laser-associated keratoplasty

FLG	Filaggrin
FLNB	Filamin B, beta
FPKM	Fragments Per Kilobase Of Exon Per Million Fragments Mapped
GATK	Genome Analysis Tool Kit
GBA3	Glucosidase, Beta, Acid 3
GSG1	Germ Cell Associated 1
GTF	Gene transfer format
GWAS	Genome-wide association study
GXYLT1P3	Glucoside xylosyltransferase 1 pseudogene 3
H2O2	Hydrogen peroxide
HNRNPL	Heterogeneous nuclear ribonucleoprotein L
IBD	Identity By Descent
ICAM-1	Intercellular adhesion molecule 1
ICRS	Intracorneal ring segment
IFNG	nterferon gamma
IGV	The Integrative Genomics Viewe
IL-n	Interleukine-n
ITGB4	Integrin beta 4
IVCM	In Vivo Confocal Microscopy
K	Keratometry
KC	Keratoconus
LAMA5	Laminin subunit alpha 5
LOD scores	Logarithm (base 10) of odds scores
LOX	Lysil oxidase
MAF	Minor Allele Frequency
MAP1LC3A	MAP1 light chain 3-like protein 1
MAP2K3	Mitogen-Activated Protein Kinase Kinase 3
MCAM	Melanoma cell adhesion molecule
MDA	Malondialdehyde
MECOM	The Integrative Genomics Viewer
MFN2	Mitofusin 2
MgCl2	Magnesium Chloride
miRNA	micro RNA
MMP-n	Matrix Mmetalloproteinases
mRNA	Messenger RNA
mt	Mitochondria
MUC(n)	Mucins
ncRNA	Non-coding RNA
NFE	Non-finnish European
NOP9	Nucleolar protein
NOX1	NADPH Oxidase 1
NT	Nitrotyrosine
OCT	optical coherence tomography
OR9G9	Olfactory Receptor, Family 9, Subfamily G, Member 9,
PCDH1	Protocadherin 1
PCR	Minor Allele Frequency
PDGF	Platelet-Derived Growth Factor

PK	Penetrating Keratoplasty
POAG	Primary Open-Angle Glaucoma
PVPI	Povidone-Iodine
QC	Quality Control
RAC1	Ras-related C3 botulinum toxin substrate 1
RE	Refractive error
RNA	Ribonucleic acid
RNF19B	Ring Finger Protein 19B
RNS	Reactive Nitrogen species
ROS	Reactive oxygen species
RPE	Retinal pigment epithelium
RPKM	Read Per Kilobase per Million mapped reads
rpm	Revolutions per minute
RPS6KA2	Ribosomal Protein S6 Kinase, 90kDa, Polypeptide 2
rRNA	Ribosomal RNA
RT-PCR	Reverse-transcriptase PCR
RUM	RNA-seq Unified Mapper
SAS	South Asian
SDS	Sodium dodecyl sulfate
SFRP1	Secreted frizzled-related protein 1
SNP	Single-nucleotide variants
sva	Mucin 4
TBE	Tris borate EDTA
TE	Sodium thiosulphate
TENM1	Teneurin transmembrane protein 1
TILK	Tuck in Lamellar keratoplasty
TIMP1	TIMP metallopeptidase inhibitor 1
TLR6	Toll-like receptor 6
TNF	Tumor Necrosis Factor
TSS	Transcript’s inferred start site
TTN	Titin
UTRs	Untranslated regions
UV	Ultaviolet radiations
VCF	Variant Call format
VSX1	Visual System Homeobox 1
WES	Whole Exome Sequencing\
WT	Wild-Type
ZFPM1	Zinc Finger Protein FOG Family Member 1
ZNF469	Zinc Finger Protein 469
ZNF738	Zinc Finger Protein738

1 General Introduction

1.1 The Human Eye

The eye is the sensory organ for vision. Its role in collecting and focusing the light onto specialized cells, which convert photons into neural signals, means that the eye is an exquisitely complicated organ. Vision is important for modern independent living and has also been subject to strong selection for the survival of many animal species, including humans, throughout their evolution.

The human eye is composed of three main layers (Figure 1-1) (reviewed by (Willoughby *et al.*, 2010). The outer layer comprises the cornea and the sclera; the middle layer is composed of the iris, the ciliary body and the choroid; and finally the inner layer of the eye consists of the retina. The optic nerve, part of the optical apparatus, consists of a bundle of nerve fibres extending to the visual cortex of the brain. The aqueous humour, the vitreous and the lens are the three transparent structures that make part of the visual system and are surrounded by the ocular layers (Justice, 1946). Figure 1-1 is adopted from National Cancer Institute - part of the National Institutes of Health – and shows the lateral view of the human eye, while depicting the three layers of the eye.

Good vision relies on the passage of refracted light rays through a clear and transparent cornea. The cornea also provides approximately two thirds of the focussing power, while the lens acts as a fine focus to form an image on the retina. Once the image hits the retina, light is converted by the retinal photoreceptors into electrical impulses that get transmitted to the brain through the optic nerve.

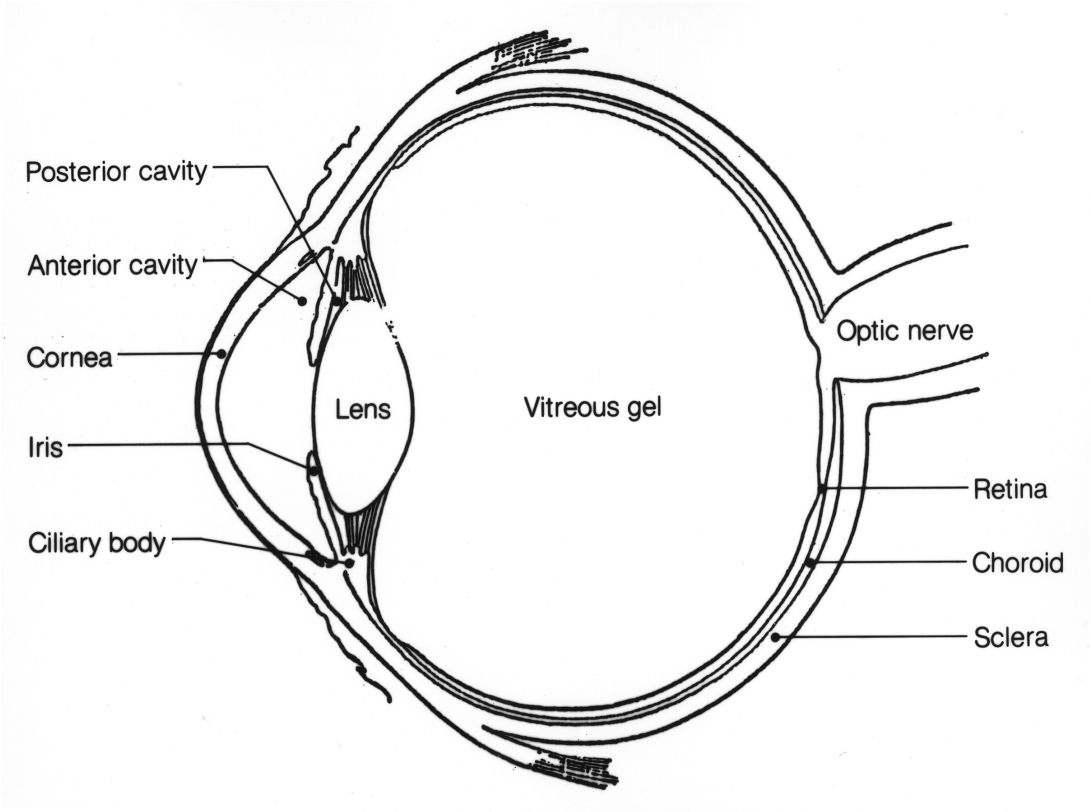


Figure 1-1: Schematic illustration showing lateral view of the human eye. Diagram showing all the major different parts of the eye. It depicts the cornea and the sclera in the anterior part of the eye, then the iris, the choroid and the ciliary body in the middle part of the eye, and the retina constituting the inner layer of the eye; as well as the three transparent layers: the aqueous, the lens and the vitreous gel. (Free Source: National Cancer Institute).

1.2 The Human Cornea

1.2.1 Structure and function of the adult cornea

The cornea is an avascular, transparent dome covering the anterior part of the human eyeball. A healthy human cornea will have a diameter, or white-to-white distance, in the average adult (18-45 years old) of 11.65±0.36 mm. This figure is slightly higher in males than in females and decreases with age (Gharaee *et al.*, 2014). The human cornea is 0.5 – 0.6 mm thick in the centre and slightly increases in thickness at the periphery (DelMonte and Kim, 2011). Along with the tear film, the cornea represents the main and the most refractive component of the ocular system (Maurice, 1957) accounting for approximately 2/3rd of the refractive power of the eye (DelMonte and Kim, 2011). Since it represents the first barrier of the eye, it is known to protect

the eye against infections and structural damage and to maintain mechanical strength. It is also characterized by its unique optical properties, with its ability to refract light while also absorbing the major part of the incoming ultraviolet (UV) radiation entering the optical system (Ringvold, 1998). The cornea represents one of the most sensitive tissues of the body due to heavy innervation via the rich sub-basal nerve plexus beneath the basal corneal epithelial layer (Oliveira-Soto and Efron, 2001).

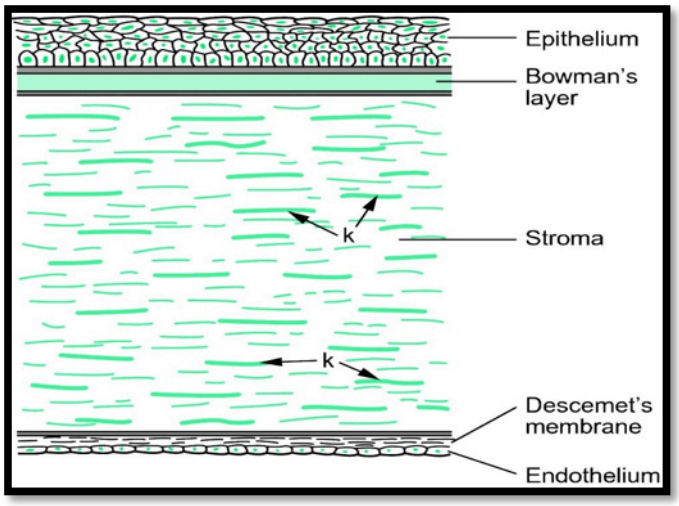


Figure 1-2: Schematic diagram depicting the human corneal layers. k stands for keratocytes. (Millodot: Dictionary of Optometry and Visual Science, 7th edition. © 2009 Butterworth-Heinemann http://medical-dictionary.thefreedictionary.com/cornea**).**

The human cornea is organized into five layers; three cellular (epithelium, stroma and endothelium) and two interface layers (Bowman and Descemet membranes) (Figure 1-2).

The *outer epithelium* is derived embryologically at around 5 to 6 weeks of gestation from surface ectoderm (Vasileva, 1957, DelMonte and Kim, 2011). It is a non-keratinized, stratified, squamous epithelium representing 10% (40-50 µm) of the corneal thickness. It is composed of 4-6 cell layers. These consist of superficial flat polygonal cells for 2-3 layers that lie in direct contact with the tear film on the corneal surface, and then less flat, wing cells or suprabasal cells for another 2-3 layers followed by a single cell layer of basal columnar cells (20 µm tall). The basal columnar cells are attached to the basement membrane by a hemidesmosomal system for enhanced

adherence to the underlying corneal layers. The outer epithelium is covered by a tear film and is responsible for protecting the rest of the eye from debris and bacteria, smoothing out corneal surface micro-irregularities and contributing to the refractive power of the cornea (DelMonte and Kim, 2011).

The *anterior limiting lamella,* also known as *Bowman's layer* is localized beneath the basement membrane of the epithelium and just anterior to the stroma. It is attached to the stroma and often considered as a simple extension of this layer (Komai and Ushiki, 1991). The Bowman's layer is a 15μm thick acellular membrane that helps the cornea maintain its shape. It consists of collagen fibres of types I, V, VI, XII and XXIV that run in various directions (Ihanamaki *et al.*, 2004).

The *stroma* originates from the migrating neural crest cells during the 7th week of gestation. It comprises a series of collagen-rich lamellae and a keratocyte matrix, and constitutes about 80-85% of the corneal thickness. This layer consists mainly of water, collagen (types I, III, V, VI, XII, XIII and XXIV), salts, proteoglycans and proteins (Ihanamaki *et al.*, 2004). The stroma provides the mechanical strength of the cornea, mainly through its collagen fibrils. The main type of collagens found in the stroma are types I and V which are arranged in long, thin, flat bundles called fibrils that run parallel to one another to form sheets called lamellae (Fini and Stramer, 2005). The lamellae run along the cornea from one side to the other (Figure 1-3) (Mathew *et al.*, 2008). The precise organization of collagen and extracellular matrix (ECM) in the stroma provides the cornea with its transparency (Hassell and Birk, 2010). The keratocytes provide stromal homeostasis to allow optimal light refraction and transparency by maintaining the synthesis of ECM molecules including crystallins, collagens, proteoglycans and matrix metalloproteinases (Jester *et al.*, 1999).

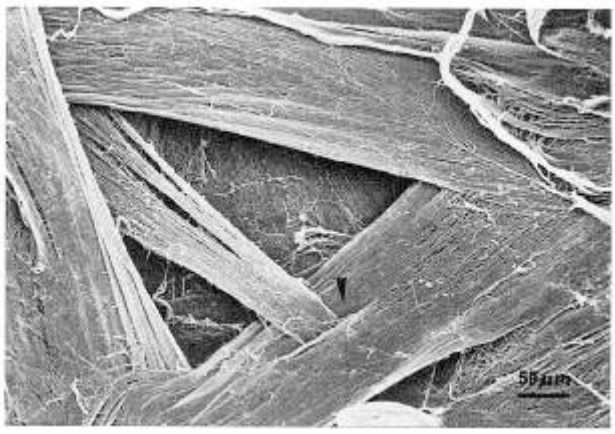


Figure 1-3: Collagen fibrils in lamellae in paracentral region of the stroma. Scanning electron micrograph shows interweaving lamellae running in different directions. The arrowhead shows a small lamellae sheet crossing another lamellae (Reused from (Radner *et al.*, 1998) by the permission of Elsevier Copy Right Clearance Centre, License number: 3771890898791).

Pre-Descemet's layer (also called *Dua's layer*) is an acellular layer about 10.1±3.60 µm in thickness (Figure 1-4) (Dua *et al.*, 2013). It is composed of collagen types I, IV and VI. The Dua's layer is made of 5 to 8 thin lamellae for which the structure seems more compact than the adjacent anterior deep stromal layer (Dua *et al.*, 2013).

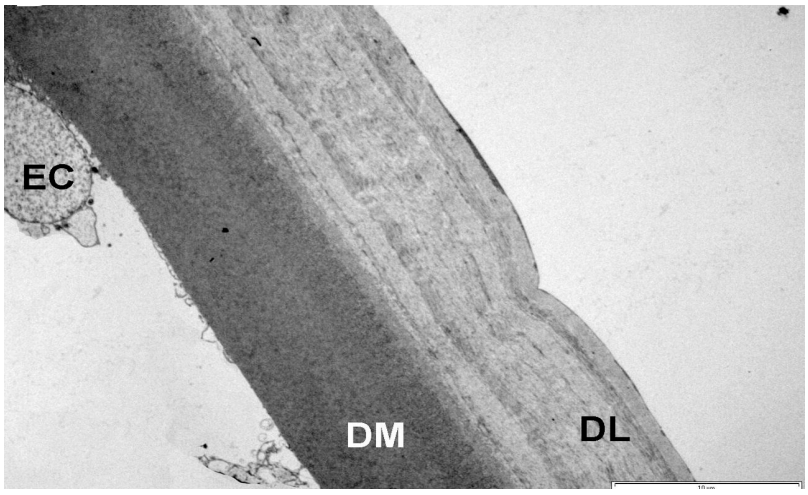


Figure 1-4: Transmission electron micrograph showing Dua's layer (DL) and Descemet's membrane (DM). Part of an endothelial cell (EC) is shown posterior to DM. DL is located in the anterior part of DM and is seen to be made of multiple thin layer. (Reused from (Dua *et al.*, 2013) by the permission of Elsevier Copy Right Clearance Centre, License number: 3764670382072).

The *posterior limiting lamella*, also known as *Descemet's membrane,* is a strong non-cellular layer located under the stroma and loosely attached to it.

Descemet's membrane measures a mean of 10.97±2.36 µm of thickness in the adult cornea (Dua *et al.*, 2013). It is composed of collagen types IV, VI and VIII (Ihanamaki *et al.*, 2004). Descemet's membrane is divided into two distinct parts: the anterior part adjacent to the stroma, is secreted by the underlying endothelial cells at the 8th week *in utero* (in the womb) and represents around 3µm of thickness of a highly organised collagen structure. The posterior part, adjacent to the endothelium, has a vague and undefined ultrastructural texture compared to the anterior part, and is secreted after birth (DelMonte and Kim, 2011).

The *endothelium* is the innermost corneal cell layer. It is a monocellular layer that is responsible for maintaining tissue integrity and corneal deturgescence by retaining the balance of fluids. As the endothelium ages, it undergoes a decrease in cell density and change in topography. The cells, which are 10µm thick at birth, start to flatten becoming 4µm thick at adulthood (Watsky *et al.*, 1989).

The strength and power of corneal refractive performance is determined by the curvature of the cornea, which can be measured by Keratometry (K). K-readings can be retrieved using a corneal topographer and the anterior segment imaging can be achieved using an optical coherence tomography (OCT) device or an anterior segment optical coherence tomography (AS-OCT) imager.

The normal corneal mean horizontal K-readings measured by AS-OCT for adult corneas of 32 years old mean age are 43 ±1.73 D (where D, the Diopter, is the unit of K-readings describing the refractive power of the cornea) with a radius of curvature of 7.85 ± 0.31 mm (Sorbara *et al.*, 2010).

1.3 Diseases of the Cornea

Being the first light barrier, the cornea serves not only to refract the light to the lens then the retina, but also as a filter, screening out the most damaging UV light and thus preventing lens and retina UV radiation injury. The cornea also serves as a shield to protect the rest of the eye from dust and bacteria. It follows a systematic mechanism of wound healing, allowing it to cope with minor injuries and abrasions (Maycock and Marshall, 2014). Thus corneal clarity and transparency is essential for its main function and any damage in its structure and/or composition can lead to a vision disturbance.

The cornea can be affected by a variety of corneal diseases arising from infections, inherited degenerations or ectasias. This last group, which includes Keratoconus, have varying degrees of genetic and environmental causes of disease and include some conditions for which the cause is unknown.

1.3.1 Infectious diseases

Infectious conditions include keratitis (Ong and Corbett, 2015), ocular herpes (Zhu and Zhu, 2014) and herpes zoster (Shingles) (Shaikh and Ta, 2002). Microbial keratitis is the most common form of keratitis accounting for 55-65% of cases (Dart *et al.*, 1991, Toshida *et al.*, 2007). The common cause of microbial keratitis is bacterial infection. This can be the result of infection by *Pseudomonas*, *Serratia marcescens*, *Enterobacteriaceae* and *Mycobacterium,* all belonging to the Gram-negative bacteria and *Staphylococcus* and *Streptococcus* species and *Nocardia,* belonging to the Gram-positive bacteria (Keay *et al.*, 2006, Dart *et al.*, 2008, Shah *et al.*, 2011). Fungal keratitis is another cause for microbial keratitis and is caused by filamentary fungi such as *Fusarium*, *Aspergillus* and microsporidia, or by yeast *Candida* (Srinivasan *et al.*, 1997, Basak *et al.*, 2005, Sharma *et al.*, 2007, Loh *et al.*, 2009). Additionally, Acanthamoeba keratitis is caused by *Acanthamoeba,* the most commonly found protozoa in soil and fresh water (Stehr-Green *et al.*, 1989, Schaumberg *et al.*, 1998).

1.3.2 Inherited degenerations

Corneal dystrophies include a range of rare, inherited degenerative conditions that are progressive and affect the cornea alone (Calhoun, 1951, Sacchetti *et al.*, 2016). Corneal dystrophies were previously classified into 3 groups according to the predominant anatomical location of the alteration at first occurrence of the disease: first the superficial corneal dystrophies, second the stromal corneal dystrophies and third the posterior corneal dystrophies (Franceschetti, 1954). Recently, corneal dystrophies were subject to a re-classification by the International Committee for Classification of Corneal Dystrophies (IC3D) according to clinical signs, genetic basis and pathological examinations of the diseased cornea. According to these criteria, the corneal dystrophies were categorised into 4 groups (Weiss *et al.*, 2008, Vincent, 2014, Weiss *et al.*, 2015):

(1) Epithelial and sub-epithelial dystrophies (e.g. Epithelial Basement Membrane Dystrophy (EBMD) (MIM:121820), Epithelial Recurrent Erosion Dystrophies (EREDs) (MIM:122400), Subepithelial Mucinous Corneal Dystrophy (SMCD) (MIM:612867), Meesmann Corneal Dystrophy (MECD) (MIM:122100), Lisch Epithelial Corneal Dystrophy (LECD) (MIM:300778), Gelatinous Drop-like Corneal Dystrophy (GDLD) (MIM:204870)).
(2) Epithelial-stromal TGFBI dystrophies (e.g. Reis–Bücklers Corneal Dystrophy (CDRB) (MIM:608470), Thiel–Behnke Corneal Dystrophy (TBCD) (MIM:604649), Lattice Corneal Dystrophy, type 1 (Classic) (LCD1) (MIM:122200), Granular Corneal Dystrophy, type 1 (Classic) (GCD1) (MIM:121900), Granular Corneal Dystrophy, type 2 (GCD2) (MIM:607541))
(3) Stromal dystrophies (e.g. Macular Corneal Dystrophy (MCD) (MIM:217800), Schnyder Corneal Dystrophy (SCD) (MIM:608370), Congenital Stromal Corneal Dystrophy (CSCD) (MIM:610048), Fleck Corneal Dystrophy (FCD) (MIM:136800), Posterior Amorphous Corneal Dystrophy (PACD) (MIM:612868), Central Cloudy Dystrophy

of François (CCDF) (MIM:217600), Pre-Descemet Corneal Dystrophy (PDCD))

(4) Endothelial dystrophies (e.g. Fuchs Endothelial Corneal Dystrophy (FECD) (MIM:136800), Posterior Polymorphous Corneal Dystrophy (PPCD) (PPCD1 MIM:122000; PPCD2 MIM:609140; PPCD3 MIM:609141), Congenital Hereditary Endothelial Dystrophy (CHED) (MIM:217700), X-linked Endothelial Corneal Dystrophy (XECD)).

1.3.3 Ectasias

In general, corneal ectasias are bilateral, non-inflammatory conditions characterized by progressive corneal steepening and thinning that causes visual impairment. The genetic basis or environmental trigger that causes these conditions is often unknown. Corneal ectasias include Keratoconus, Pellucid Marginal Degeneration, Keratoglobus, Postkeratorefractive ectasia and wound ectasia after penetrating keratoplasty-PK (Ophthalmology, 2013). Since keratoconus is the main focus of this thesis, the next section is designated to describe in details all the aspects of this ectatic disorder.

1.4 Keratoconus

1.4.1 Definition and introduction

Keratoconus (KC) [MIM:148300] is derived from the Greek words kerato, which means cornea, and konos, which means cone. The condition is characterised by a conical-shaped protruded cornea that results from progressive corneal thinning. KC was originally reported by Duddell in 1736, whithout being called KC, in a treatise on his work in ophthalmology (Etzine, 1954, Grzybowski and McGhee, 2013), and was first described and referred to as staphyloma diaphanum by Burchard Mauchart in 1748 (Nielsen *et al.*, 2013). It was distinguished from other corneal ectatic diseases by John Nottingham in 1854 (Grzybowski and McGhee, 2013).

KC is a non-inflammatory, generally asymmetric, progressive thinning of the cornea resulting in a conically shaped protrusion (Figure 1-5). The cone shape is a consequence of structurally weakened tissue and can result in irregular corneal scarring, astigmatism and myopia, eventually leading to loss of vision if left untreated (Rabinowitz, 1998). According to the study by Krachmer *et al.*, the impaired vision is primarily due to abnormalities in light refraction so that the image cannot be focussed onto the retina (Krachmer *et al.*, 1984).

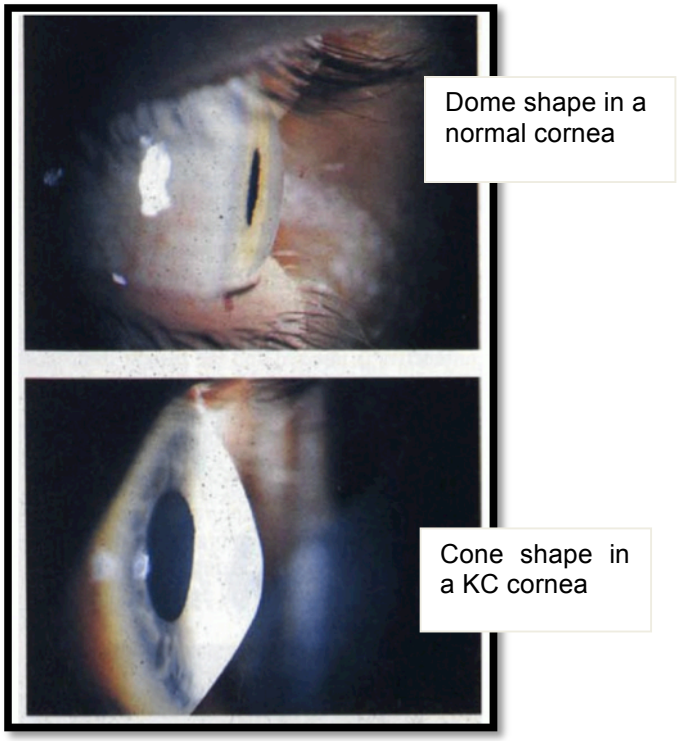


Figure 1-5: Photograph of a normal and KC eye in profile (Reused from (Lawless *et al.*, 1989) by the permission of Copy Right Clearance Centre, License number: 3795370830303).

KC is generally detected at puberty and evolves throughout life to stabilize in the third and fourth decade (Rabinowitz, 1998). In its early stages, KC is not easy to diagnose since it is difficult to distinguish its mild or early stages from irregular myopic astigmatism; however, recent advances in computerized corneal topography have made early diagnosis easier (Maeda *et al.*, 1994). The incidence of KC varies from one population to another, but several reports have stated that KC affects approximately one patient among 2000 individuals in the general population (Krachmer *et al.*, 1984, Kennedy *et al.*, 1986, Zadnik *et al.*, 1996, Rabinowitz, 1998, Rabinowitz, 2003). KC is

reported to be the principal cause of corneal transplantation in the western world (Zadnik *et al.*, 1996, Legeais *et al.*, 2001, Cassidy *et al.*, 2013, Gomes *et al.*, 2015, Parker *et al.*, 2015, van Dijk *et al.*, 2015).

1.4.2 Epidemiology and association with other diseases

The prevalence of KC varies from one population to another, ranging from 0.3 per 100 000 in Russia (Gorskova and Sevost'ianov, 1998) to 3,333 per 100 000 in Lebanon (Waked *et al.*, 2012). This variability may be specific to each population since there is a four-fold increased incidence of KC in Asians living in the UK compared to Caucasians (Pearson *et al.*, 2000). Similarly, according to Georgiou et al., this observation may have a genetic basis as Asians are 7.5 times more likely to develop KC than white affected individuals (Georgiou *et al.*, 2004).

It does not appear that KC is more prevalent in one gender than another. Some studies have found a greater prevalence in males (Weed and McGhee, 1998, Pearson *et al.*, 2000, Owens and Gamble, 2003, Wagner *et al.*, 2007), other studies have found higher incidences in females (Krachmer *et al.*, 1984, Harold A. Stein, 2013), and still other studies have found no differences (Kennedy *et al.*, 1986, Li *et al.*, 2004).

KC has been shown to be associated with a number of ophthalmic and genetic systemic conditions (Edwards *et al.*, 2001, Sugar and Macsai, 2012). The types of diseases associated with KC fall into four categories. The first category belongs to the diseases of abnormal retinal function (e.g. Lebers Congenital Amaurosis (Hameed *et al.*, 1999, Damji *et al.*, 2001, McMahon *et al.*, 2009), Retintis Pigmentosa (Bruna, 1954, Knoll, 1955), Bardet-Biedl syndrome (Francois *et al.*, 1982), albinism (Rao *et al.*, 2008) and cone dystrophy (Fogla and Iyer, 2002, Yeh and Smith, 2008)). The second category falls into the diseases associated with atopy, eczema and eye rubbing (e.g. Down syndrome (Koppen *et al.*, 2010, Sabti *et al.*, 2015), Turner syndrome (Macsai *et al.*, 1997, Pinna *et al.*, 2005), hyper-IgE

syndrome (Kim and Netto, 2004) and Mulvihill–Smith syndrome (Rau and Duncker, 1994)). The third category of diseases associated with KC belongs to the diseases harbouring abnormal collagen elasticity and connective tissue disorders (e.g. Brittle cornea syndrome (Lechner *et al.*, 2014, Davidson *et al.*, 2015), Congenital Hip Dysplasia (Nucci and Brancato, 1991), Osteogenesis Imperfecta (Beckh *et al.*, 1995), Marfan Syndrome (Dudakova and Jirsova, 2013) and Joint Hypermobility (Woodward and Morris, 1990). The fourth category of diseases associated to KC belong to the category of intellectual impairment (e.g. Crouzon syndrome (Perlman and Zaidman, 1994), Hyperornithinemia (Chen and Furr, 1983) and Noonan syndrome (Ascaso *et al.*, 1993, Lee and Sakhalkar, 2014).

KC can also be associated to other corneal dystrophies (e.g. Posterior Polymorphous Corneal Dystrophy (Heon *et al.*, 2002b, Vincent *et al.*, 2013a), Granular dystrophy (Rho *et al.*, 2014, Wilson *et al.*, 2014), Fuch's Endothelial Dystrophy (Salouti *et al.*, 2010, Mazzotta *et al.*, 2014) and Lattice Dystrophy (Hoang-Xuan *et al.*, 1989), implying some common defective pathways and tightly linked protein networks.

Interestingly, KC is also associated with diseases caused by chromosomal abnormalities, such as translocations and chromosomal deletions. Amongst these diseases are Downs and Turner Syndromes, chromosome 7;11 translocation (Morrison *et al.*, 2001), chromosome 22q11.2 deletion (Saffra and Reinherz, 2015) and the chromosome 13 ring anomaly (Heaven *et al.*, 2000).

1.4.3 Risk factors of KC

KC is suspected to have a multifactorial basis to disease onset, having both a genetic and an environmental component (Edwards *et al.*, 2001, Sugar and Macsai, 2012, Patel and McGhee, 2013). In an attempt to find possible causative factors for KC using patients' questionnaires, it was found that contact lens wear, allergy, eye rubbing, atopy and family history all

contributed to KC (Rabinowitz, 2003). It was suggested that the environmental contribution may trigger KC in genetically-susceptible individuals (Davidson *et al.*, 2014). However the relative contribution of each of these factors remains to be established.

Atopy in the form of allergy, eczema and asthma has been reported to be a risk factor for KC by many research groups (Crews *et al.*, 1994, Kaya *et al.*, 2007, Nemet *et al.*, 2010, Shneor *et al.*, 2013). A negative association has been reported by several other groups, but the controls group were not age and sex matched, and atopy did not include all three sub-categories (Lowell and Carroll, 1970, Gasset *et al.*, 1978, Wachtmeister *et al.*, 1982, Bawazeer *et al.*, 2000). It has long been believed that eye rubbing caused by the atopy itch triggers the onset of KC (Coyle, 1984). However, atopy is not the only provocative factor for eye rubbing in KC as KC patients with a history of eye rubbing present with a much higher incidence than KC patients with atopy (Rahi *et al.*, 1977).

Researchers over the years have associated eye rubbing with KC development and considered it to be a KC risk factor. Individual case reports (Coyle, 1984, Gritz and McDonnell, 1988, Zadnik *et al.*, 2002, Jafri *et al.*, 2004, Koenig, 2008), as well as case-control studies (Bawazeer *et al.*, 2000, Kim and Joo, 2008, Gordon-Shaag *et al.*, 2013), associate habitual and chronic eye rubbing to the appearance and/or progression of KC. It is not the cause of all KC development as some KC patients develop the disease without any history of eye rubbing. However, eye rubbing may be a risk factor for genetically predisposed patients (McMonnies, 2007), since a number of patients develop KC after a history of eye rubbing (Kennedy *et al.*, 1986).

Excessive exposure to UV may also be a risk factor of KC as rabbits and mice exposed to UV show signs of KC in their corneas (Podskochy *et al.*, 2000, Newkirk *et al.*, 2007). The rabbit and mouse corneas showed

evidence of cell apoptosis in all layers, and additional stromal thinning and collagen degeneration was shown in mice.

Sporadic KC is the most commonly reported form of disease, but familial cases have been reported with a rate ranging of 5 to 27.9% (Ihalainen, 1986, Kennedy *et al.*, 1986, Rabinowitz, 2003, Weed *et al.*, 2008, Millodot *et al.*, 2011, Shneor *et al.*, 2013). Gordon-Shaag et al. recently established strong evidence for a genetic component in the development of KC by associating consanguinity and endogamy with KC (Gordon-Shaag *et al.*, 2013). In addition, consanguinity and ethnic influence has been reported to be a risk factor for KC, where incidence of KC was much higher in first cousin than in second cousin married relatives, as well as a four fold increase in frequency Asians over Caucasians in the UK (Pearson *et al.*, 2000).

1.4.4 KC signs and symptoms - Clinical features

The hallmark of KC diagnosis is irregular astigmatism, which can be detected by AS-OCT in order to detect the corneal curvature that steepens with the development of KC (Figure 1-6). A K-reading for corneal curvature in the normal cornea of between 43 and 45 D is indicative of mild KC. A range between 45 to 52 D is indicative of moderate KC and a K-reading higher than 52 D is considered to be advanced KC (Sinjab, 2012). In addition to conicity in advanced KC, patients exhibit varying degrees of blurred vision, photophobia and eye-strain that is accompanied by monocular diplopia (seeing a double image in one eye), multiple ghost images, flaring and glares around light (Lee *et al.*, 1995).

In mild KC, loss of visual acuity often manifests as an inability to reach a 6/6 score on a snellen test. Irregular astigmatism increases with progression of the disease and “scissor” shadows start to appear while conducting a retinoscopy. Apical thinning of the apex of the cone is another characteristic of an early stage KC (Rabinowitz, 1998).

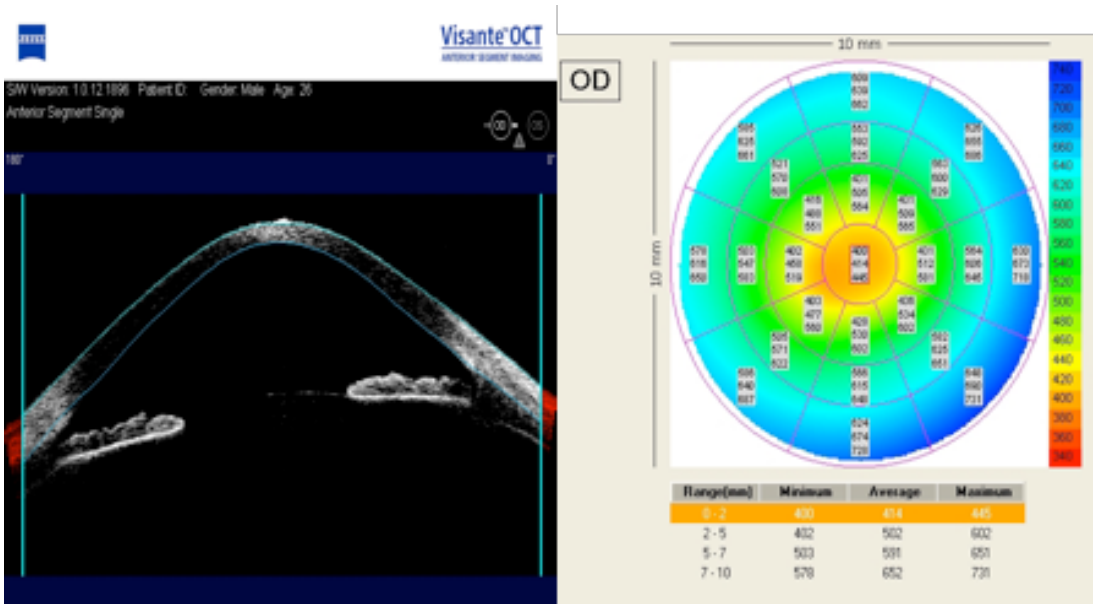


Figure 1-6: Visante OCT corneal mapping of a KC cornea. The left image shows the apical corneal thinning in KC cornea. The right image shows the central thinning detected by the pachymetry map, which reports the corneal thickness in multiple points of the cornea (Provided by Salina Siddiqui, Saint James's University Hospital, Leeds, UK)

In moderate and advanced KC, the thinning continues to progress and the basal layer of the epithelium presents with an accumulation of ferritin particles that can be seen by slit lamp examination. These particles form an iron, yellow-brown to olive-green coloured ring encircling the cornea, which is called Fleischer's ring (Iwamoto and DeVoe, 1976, Maguire and Bourne, 1989) (Figure 1-7, 1-8). It is localised inside and between epithelial cells as viewed using a cobalt blue or Perl's staining and photographed under UV light.

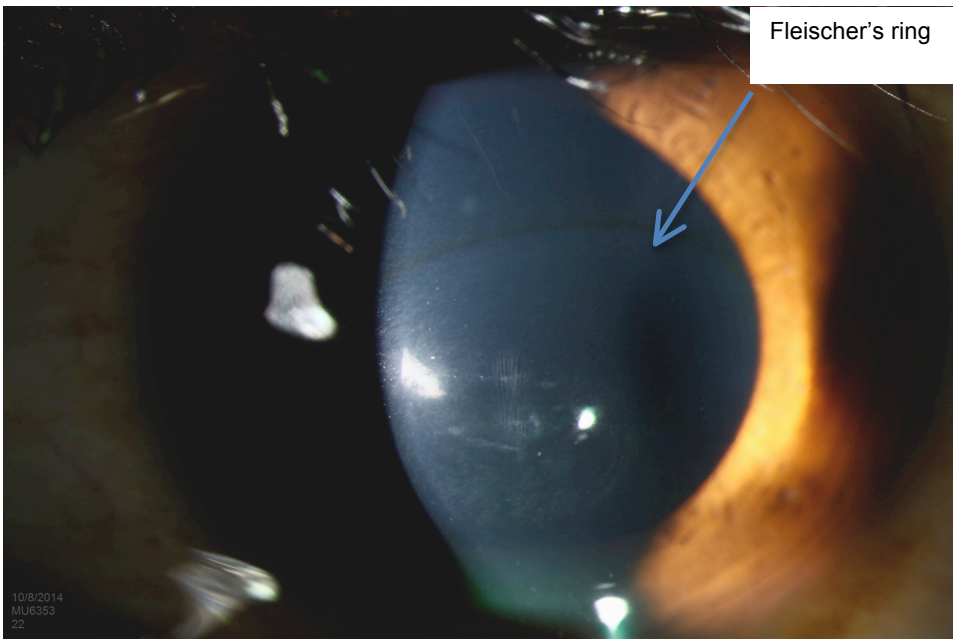


Figure 1-7: Fleischer's ring in moderate KC. The arrow points to corneal iron deposit ring (Photographed by Stefani Karakas, CRA, retrieved under a free license from Ophthalmic Atlas Images by EyeRounds.org, The University of Iowa)

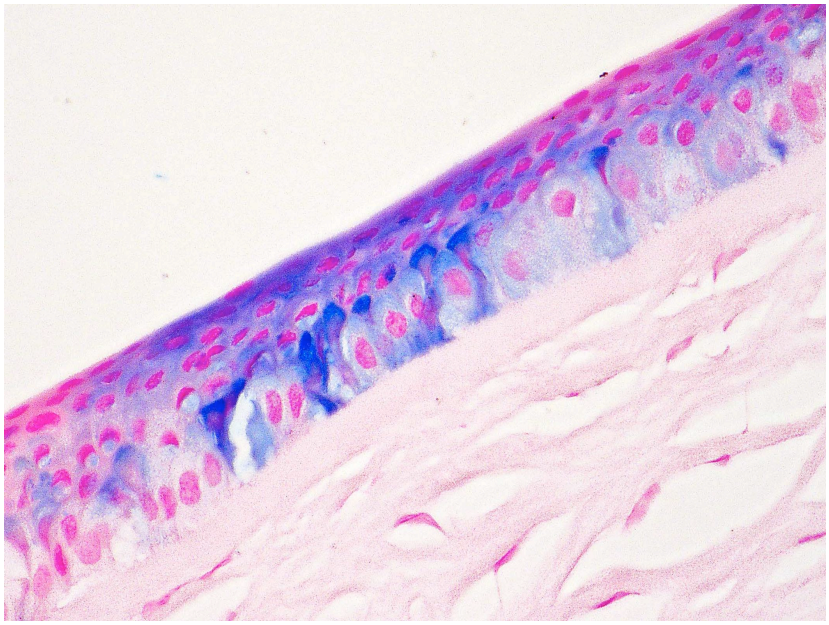


Figure 1-8: Perl's stain shows intraepithelial iron deposits - Fleischer's ring. The blue stain represents the iron molecules (Provided by Dr Hardeep S Mudhar, Consultant Ophthalmic Histopathologist, Royal Hallamshire Hospital, Sheffield, UK).

Another clinical sign that is required for a KC diagnosis are the Vogt's lines (Vogt's striae) (Figure 1-9) (Krachmer *et al.*, 1984). These fine vertical and

deep stromal striae are detected by slit lamp examination in the deep stroma, and sometimes in Descemet's membrane, and are due to stretching within the deeper corneal lamellae (Von Der Heydt, 1930). Another characteristic sign is the visibility of corneal nerves in addition to deep corneal opacities. The apical thinning in the very early stages of KC can evolve, during later stages, into *Munson's sign* which represents the bulging of the lower eyelid on down gaze by the protruded cornea (Figure 1-10) (Krachmer *et al.*, 1984, Li *et al.*, 2004). Rizzuti's sign is another typical sign of advanced KC. It is detected in advanced stages and described as the penlight test for KC. Rizzuti's sign manifests while illuminating the cornea from the temporal side, the beam of light sharply focuses near the nasal limbus (Figure 1-11) (Rizzuti, 1970). In severe cases, acute hydrops appear over the longer term, followed by deep stromal scarring, rupturing of the Descemet's membrane and corneal oedema as a result of aqueous fluid passing into the stroma, leading in turn to opacification (Thota *et al.*, 2006, Poyales-Galan *et al.*, 2009).

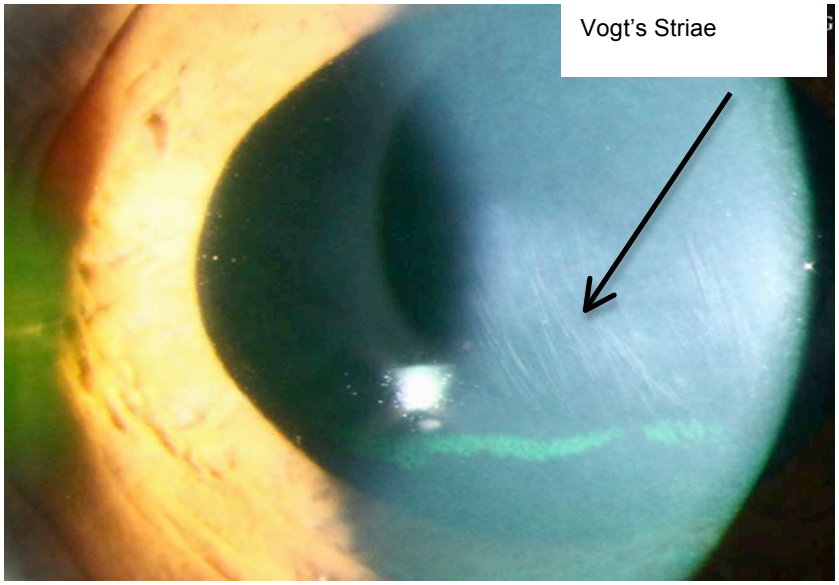


Figure 1-9: Vogt's Striae in Keratoconus. The arrow points to the Vogt's Striae in a KC cornea (Photographed by Toni Venckus, CRA, retrieved under a free license from Ophthalmic Atlas Images by EyeRounds.org, The University of Iowa)

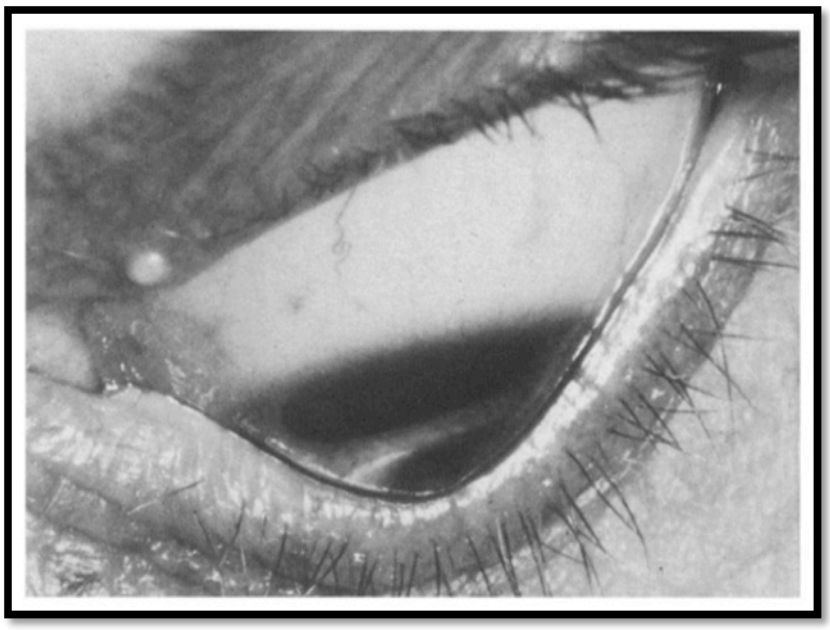


Figure 1-10: Munson's sign. The lower lid bulges forward on down gazing (Reused from (Krachmer *et al.*, 1984) by the permission of Copy Right Clearance Centre, License number: 3795371398260).

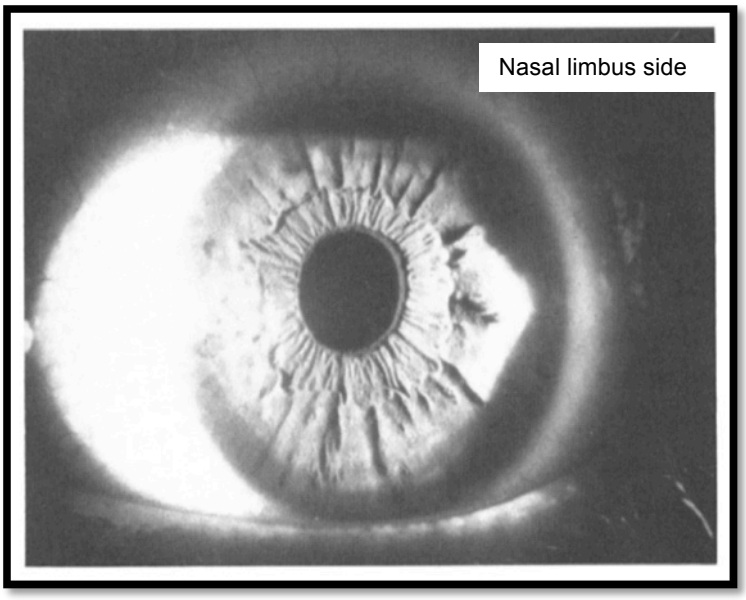


Figure 1-11: Rizzuti's sign. The beam of light illuminating the cornea from the temporal side sharply focuses near the nasal limbus (Reused from (Krachmer *et al.*, 1984) by the permission of Copy Right Clearance Centre, License number: 3795371398260).

1.4.5 Corneal structure and histopathological findings in KC corneas

In KC, the structure of all corneal layers (the epithelium, Bowman's layer, stroma, Descemet's membrane and the endothelium) is altered (details to follow), preventing the cornea from executing its normal function. The etiology of KC is unknown so identifying, the mechanisms of corneal alterations in KC will enable better comprehension of the pathogenesis and characteristics of the disease.

The main pathological features known to be hallmarks of the disease are the breaks in, or complete absence of, Bowman's membrane. The breaks are due to fragmentation and stromal thinning (Figure 1-12) (Kenney *et al.*, 2000) accompanied by lower cell and innervation density (Niederer *et al.*, 2008).

In vivo confocal microscopy (IVCM) of the non-keratinized stratified squamous epithelium reveals degeneration of the basal epithelial cells along with an approximately 25% reduction in cell density compared to control corneas (Niederer *et al.*, 2008).

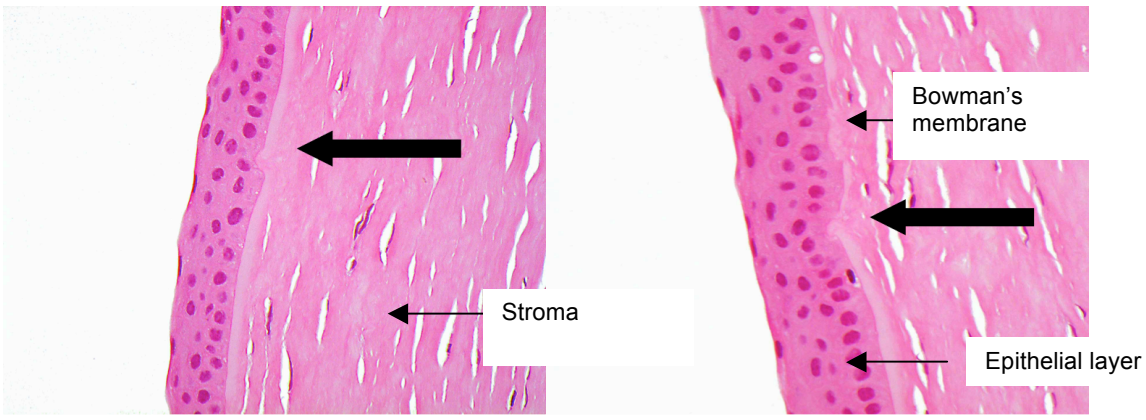


Figure 1-12: Angulated breaks in Bowman's membrane. Histopathological sections of KC corneas show thinning, scarring and breaks at the level of Bowman's membrane (Reproduced with the permission of Dr Hardeep S Mudhar, Consultant Ophthalmic Histopathologist, Royal Hallamshire Hospital, Sheffield, UK).

Bowman's layer shows discontinuities in the form of breaks at the early stages of KC. These breaks may increase in more advanced stages of disease and may lead to a total loss of this layer. Consistent with this loss, epithelial cells and stroma become directly in contact (Figure 1-13) due to the absence of the anchoring lamellae that extend from the epithelium into Bowman's layer (Kenney *et al.*, 2000, Morishige *et al.*, 2007). There also appears to be an increased visibility of corneal nerves, consistent with nerve thickenings due to breaks in Bowman's membrane (Bron, 2001, Brookes *et al.*, 2003).

Figure 1-13: Breaks in Bowman's membrane compared to normal. Histological differences between intact and altered Bowman's layer (taken from Global Keratoconus Foundation website).

The stroma in KC corneas shows alterations in the collagen-rich lamellar structure, causing mechanical weakness which distorts corneal function and shape (Bron, 2001, Muller *et al.*, 2001, DelMonte and Kim, 2011). These alterations can be considered to be the leading cause of corneal thinning and cone formation (Morishige *et al.*, 2007). Furthermore, stromal thinning has been shown to be directly related to stromal lamellar slippage, which could cause biomechanical instability leading to molecular stress of the cell. This slippage is due to a lack of cohesion between collagen fibres and the non-collagenous ECM, which were reported to be broken down and degraded in KC (Meek *et al.*, 2005). In addition to the decrease in number of collagen lamellae and the loss of arrangement of fibrils (Section 1.2.2) in the

anterior stroma, KC corneas demonstrate a decrease in keratocyte density compared to normal ones (39.2% and 27.5% decrease in anterior and posterior keratocyte densities respectively) (Niederer *et al.*, 2008), consistent with the keratocyte apoptosis reported in KC (Kim *et al.*, 1999, Kaldawy *et al.*, 2002, Matthews *et al.*, 2007).

Descemet's membrane, the posterior stromal non-cellular layer, is rarely affected in KC (Rabinowitz, 1998). However in advanced stages of disease, breaks in this layer can be detected (Poyales-Galan *et al.*, 2009).

The endothelium is usually not affected in KC (Rabinowitz, 1998) though in advanced stages of disease, lower endothelial cell density was observed under IVCM (Niederer *et al.*, 2008) consistent with increased apoptosis in these cells (Kaldawy *et al.*, 2002).

1.4.6 Phenotypic spectrum of KC and morphology

The introduction of computerized corneal topography and imaging (such as *in vivo* confocal microscopy) has led to the detection of minimal variations in corneal diseases and has permitted the detection of early and mild KC. This presents advantages over mainstream diagnostic methods in that such methods usually detect patients at more advanced stages of disease development. In particular, abnormalities in corneas not falling within the KC diagnosis range, yet which are not entirely normal, also constitute a form of KC referred to as "forme-fruste". This is usually detected in unaffected members of families with individuals that have KC. KC corneas presenting as form-fruste, present at the central corneal zone an intermediate level of epithelial thickness between that of normal and KC corneas (Lafond *et al.*, 2001, Rabinowitz *et al.*, 2014, Temstet *et al.*, 2015).

In previous years, advanced KC used to be sub-classified into three categories according to the morphology of the cone, these being Nipple,

Oval and Globus KC. All three categories were divided according to the shape of the cornea after KC onset. For nipple KC, the cone is near-central. For oval KC, the cone shape is oval and the corneal apex is usually situated under the corneal midline. For the Globus KC, the cone presents irregularly shaped and could affect over 75% of the cornea (Sinjab, 2012, Romero-Jimenez *et al.*, 2015).

1.4.7 Possible pathways involved in KC pathogenesis

To date, several mechanisms have been suggested to contribute to the onset and progression of KC, including inflammation, apoptosis of stromal cells, enzyme/inhibitor imbalances and oxidative damage.

1.4.7.1 Inflammation in KC

KC has always been thought to be a non-inflammatory disorder (Rahi *et al.*, 1977, Krachmer *et al.*, 1984, Edwards *et al.*, 2001). However this notion has been challenged recently (Lema and Duran, 2005, Lema *et al.*, 2009, Lema *et al.*, 2010, Cheung *et al.*, 2013, Cheung *et al.*, 2014, Galvis *et al.*, 2015b, McMonnies, 2015). Indeed proteomic analysis of KC patients' tear film revealed high levels of pro-inflammatory cytokines IL-4, -5, -6, and -8 (Interleukin-4, -5, -6 and -8), TNF-α and β (Tumour Necrosis Factor-alpha and Beta), ICAM-1 (intercellular adhesion molecule-1), MMP-1, -3, -7, -9 and 13 (Matrix Metalloproteinase-1, -3, -7, -9 and -13) and decreased levels of IL-12 and -17 (Interleukin-12 and -17) (Duran and Lema, 2003, Lema and Duran, 2005, Seppala *et al.*, 2006, Lema *et al.*, 2009, Pannebaker *et al.*, 2010, Jun *et al.*, 2011, Balasubramanian *et al.*, 2012).

In addition, there is an increase of tear protein level of MMP-13, IL-6 and TNF-alpha in KC corneas as well as in normal eyes in response to eye rubbing (Balasubramanian *et al.*, 2013), suggesting that the causal link between KC and eye rubbing might be directly related to the increased levels of these cytokines. More specifically, a wound healing *ex vivo* analysis of KC and normal corneas conducted by Cheung et al., revealed higher

levels of IL-1α, IGF-1 (insulin-like growth factor 1), TNF-α, and TGF-B1 (TGFβ-1) in KC corneas compared to the normal corneas (Cheung *et al.*, 2013, Cheung *et al.*, 2014).

1.4.7.2 Apoptosis

Apoptosis has been reported in kerarocytes of keratoconic corneas by several studies (Kim *et al.*, 1999, Kaldawy *et al.*, 2002, Sevost'ianov *et al.*, 2002, Chwa *et al.*, 2006, Matthews *et al.*, 2007). Wilson et al. and Helena et al., suggested that this pathway is triggered by mechanical injury of the corneal epithelium (Wilson *et al.*, 1996b, Helena *et al.*, 1998). As previously described in Section 1.4.6, KC has been associated with eye rubbing, and also with atopic diseases from which the corneal injury might arise, as well as the use of poorly fitted contact lenses (Macsai *et al.*, 1997). Under these conditions it is thought that the wounded epithelial KC cells release cytokines TNF-α (Tumor Necrosis Factor alpha) and IL-1 (Interleukin 1) which induce apoptosis in the underlying stromal cellular layer (Cheung *et al.*, 2013). KC corneas have indeed been shown to overexpress IL-1 receptors, and IL-1 induces keratocytes to undergo apoptosis through the activation of Fas ligand (Bureau *et al.*, 1993).

Another hypothesis for a mechanistic basis for KC is the downregulation of TIMP-1 in the early stages of disease, which might also be a stimulus for keratocytes apoptosis in KC corneas (Guedez *et al.*, 1998) since a progressive decrease of TIMP-1 expression during the progression of KC has been recorded (Cristina Kenney and Brown, 2003). Other factors that have been proposed as possible stimuli for keratocyte apoptosis in KC corneas are the imbalance between TIMP-1 and TIMP-3 levels (Matthews *et al.*, 2007) and the up-regulation of Cathepsins (Chwieralski *et al.*, 2006).

1.4.7.3 Extracellular Matrix degradation and remodelling

Two mechanisms have been shown to cause KC corneal thinning and tissue degradation (Section 1.4.4): lamellae slippage and ECM remodelling (Meek

et al., 2005, Chaerkady *et al.*, 2013). A noticeable reduction in collagen lamellae has been detected in KC patients compared to controls under light and electron microscopy (Takahashi *et al.*, 1990). Additionally, stromal degeneration and lack of epithelial integrity have been detected in KC corneas, along with a decrease in collagens types I, III, V and XII, lumican and keratocan proteins (Chaerkady *et al.*, 2013). Given that the corneal stroma is composed of 75% collagen type I, with types III, IV and V intertwined into the lamellae (Newsome *et al.*, 1982), and that the epithelium and sub-epithelial stroma are composed mainly of collagen type XII (Chaerkady *et al.*, 2013, Cheung *et al.*, 2013), loss of these proteins is consistent with the corneal thinning and remodelling observed.

Lamellae slippage is thought to be caused by alterations in the stromal organization due to an increase in the interfibrillar distance. The increase in distance promotes excessive contact between the collagen lamellae and the proteoglycans, and the proteoglycans are documented as exhibiting abnormalities in their configuration with the progression of the disease (Akhtar *et al.*, 2008). Moreover, a loss in lamellae connecting the stroma to Bowman's membrane has been detected by second-harmonic imaging, along with a reduction and sometimes a loss in interweaving lamellae (Morishige *et al.*, 2007). Anterior KC corneas express fibronectin and tenascin glycoproteins that are not usually found in wounded or normal corneas, confirming the involvement of anterior cornea in KC pathogenesis (Tuori *et al.*, 1997, Filenius *et al.*, 2003).

1.4.7.4 Imbalance in proteinase activity and proteinase inhibitor level

Increased proteinase activity and decreased proteinase inhibitor function have also been proposed to play a role in the pathogenesis of KC (Sawaguchi *et al.*, 1989, Sawaguchi *et al.*, 1990, Brown *et al.*, 1993, Opbroek *et al.*, 1993, Smith *et al.*, 1995, Zhou *et al.*, 1998, Collier, 2001). KC corneas present increased levels of gelatinase activity (Cristina Kenney and Brown, 2003), which may correlate with increased levels of lysosomal enzymes such as acid lipases, acid phosphatases and esterases

(Sawaguchi *et al.*, 1989), increased levels of cathepsin B and G activity (Zhou *et al.*, 1998), and increased activity of gelatinolytic enzymes such as the matrix metalloproteinases (MMPs) (Zhou *et al.*, 1998). Among the MMPs, MMP-2 is suspected to be involved in KC (Smith and Easty, 2000); this is based on the imbalance between this enzyme and its inhibitor, TIMP-1 (Brown *et al.*, 2004). In addition to the reduced levels of the gelatinase inhibitor, TIMP-1, two other inhibitors of degradative enzymes are decreased in KC, α1-proteinase and α-2 macroglobulin (Whitelock *et al.*, 1997a, Whitelock *et al.*, 1997b, Zhou *et al.*, 1998, Brown *et al.*, 2004). This decrease correlates with increased levels of corresponding degradative enzymes, (Brown *et al.*, 1993, Maruyama *et al.*, 2001) favouring tissue breakdown and leading to stromal thinning as well as fibroblast apoptosis (Chwa *et al.*, 2006, Balasubramanian *et al.*, 2012).

At the molecular level, the overexpression of cathepsin K, G and B and the upregulation of MMP-2 and MMP-9 could be caused by an increase in IL-1 (Kamolmatyakul *et al.*, 2004). In turn, cathepsins activate MMP-1, MMP-9 and MT-MMP1 known as MMP14, which contributes to MMP-2 activation. An increase in MMP-2 can also be caused by low levels of TIMP-1 (at KC early stages) (Meijers *et al.*, 1994). The basement membrane is then more prone to degradation by the stimulated MMP-1, MMP-2 and MMP-9 as well as by the increase in cathepsin levels (Balasubramanian *et al.*, 2010).

Another hypothesis proposed is that the ECM changes in KC corneas might be due to activation of the protease cascade involving the plasmin system and the cyclooxygenase and MMPs, by overexpression of cytokines IL-1, IL-6, IL-8, TNF-α, TGFβ and PDGF (Platelet-Derived Growth Factor) in KC corneas (Cheung *et al.*, 2013, Galvis *et al.*, 2015a).

1.4.7.5 Oxidative Stress

Several studies have investigated the role of oxidative stress in KC corneas and proposed that oxidative damage plays a major role in KC pathogenesis

(Behndig *et al.*, 2001; Buddi *et al.*, 2002; Gondhowiardjo and Haeringen, 1993; Gondhowiardjo *et al.*, 1993; Kenney *et al.*, 2005). KC corneas are deficient in antioxidant capability as well as glutathione (defence against chemicals and oxidative stress), leading to a deficit in processing reactive oxygen species (ROS). This in turn accelerates oxidative damage (Arnal *et al.*, 2011, Wojcik *et al.*, 2013b). This is known as the "Cascade Hypothesis" (Cristina Kenney and Brown, 2003).

Consistent with this, large amounts of cytotoxic by-products were shown to be present in KC corneas compared to normal ones, with malondialdehyde (MDA) and nitrotyrosine (NT) as the primary end products emerging respectively from lipid peroxidation and nitric oxide pathways (Figure 1-14) (Buddi *et al.*, 2002). These finding support the hypothesis that KC corneas undergo oxidative stress (Buddi *et al.*, 2002, Chwa *et al.*, 2006).

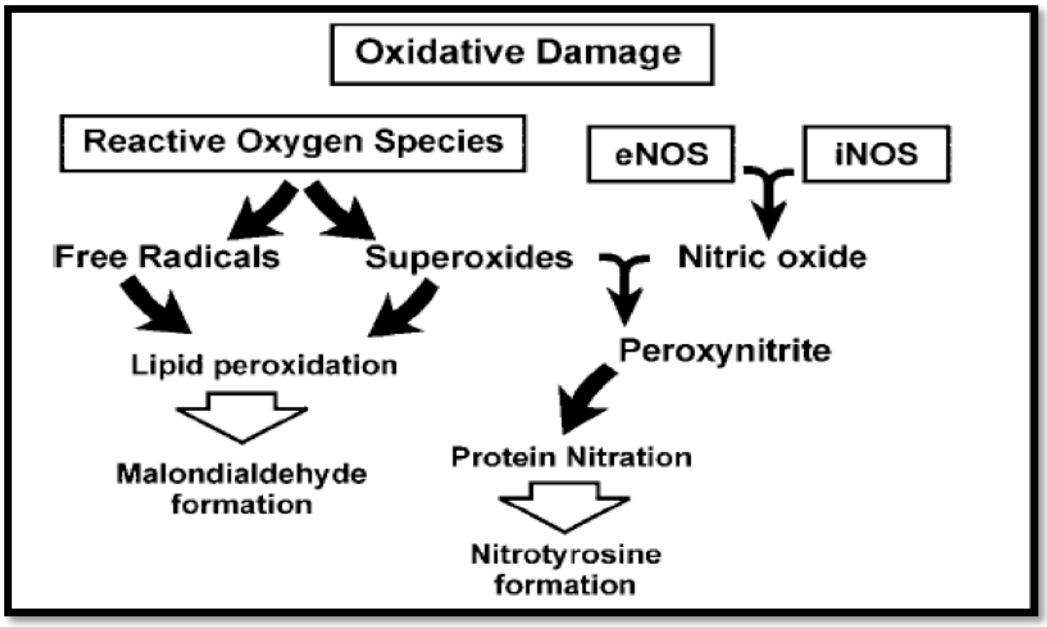


Figure 1-14: Schematic representation of oxidative stress pathways. eNOS is the endothelium isoform of the constitutive nitric oxide synthase (cNOS), and iNOS is the inducible nitric oxide synthase (taken from (Buddi *et al.*, 2002). No permission needed for re-use in Doctoral dissertation).

Indeed, high levels of ROS, reactive nitrogen species (RNS) and cytotoxic aldehydes were reported in KC patients corneas, also consistent with a deficiency in antioxidant capacity (Buddi *et al.*, 2002, Chwa *et al.*, 2006, Arnal *et al.*, 2011, Wojcik *et al.*, 2013b).

Since corneas are constantly generating ROS through UV light absorption, they therefore possess defence mechanisms to protect them from oxidative stress. These mechanisms are represented by protective enzymes against aldehyde formation such as aldehyde dehydrogenase (ALDH3) (Kays and Piatigorsky, 1997), and against ROS formation such as antioxidant enzymes including superoxide dismutase (SOD), catalase, glutathione peroxidase and glutathione reductase (Figure 1-13) (Rao *et al.*, 1987). Antioxidant enzymes are shown to be present in almost every layer of normal corneas (Cejkova *et al.*, 2004a, Cejkova *et al.*, 2004b). In KC corneas, the free radical scavengers ALDH3 and SOD are underexpressed compared to normal corneas (Abedinia *et al.*, 1990, Gondhowiardjo *et al.*, 1991, Behndig *et al.*, 2001). Furthermore, increased levels of catalase, an enzyme involved in hydrogen peroxide (H_2O_2) elimination, were reported in KC corneas (Kenney *et al.*, 2005), suggesting therefore a defence mechanism against elevated quantities of H_2O_2 (Figure 1-15). These alterations in protective enzymes are consistent with the accumulation of toxic by-products such as MDA and NT reported in KC corneas (Buddi *et al.*, 2002).

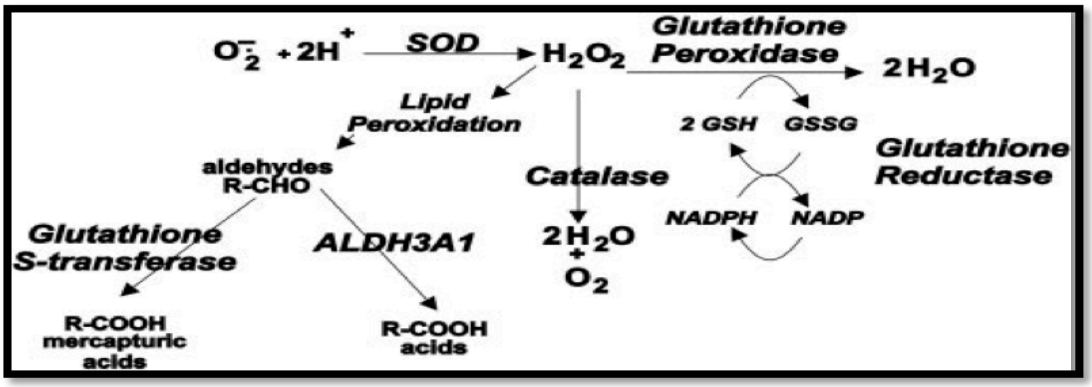


Figure 1-15: Schematic representation of the reaction of reactive oxygen species with antioxidant enzymes (Reused from (Kenney *et al.*, 2005) by the permission of Copy Right Clearance Centre, License number: 3780400109886).

Oxidative stress in KC patients has long been thought to be restricted to the defective corneas (Jun *et al.*, 2011). However, a recent study reported high levels of oxidative stress index in KC patients' sera compared to controls (Toprak *et al.*, 2014).

Additional evidence for the presence of oxidative stress in KC corneas comes from the finding of increased mitochondrial DNA (mtDNA) damage in KC corneas compared to normal ones (Atilano *et al.*, 2005). The mtDNA damage also leads to increased levels of ROS production and may be directly correlated with tissue degeneration in corneal fibroblasts (Yakes and Van Houten, 1997, Chwa *et al.*, 2006, Chwa *et al.*, 2008).

1.4.8 Treatments for KC

1.4.8.1 Contact lenses

The type of contact lenses used by KC patients depends on the severity of the disease. For KC patients with mild disease and a visual acuity of 6/12 or better, spectacles are an option. However, with increased irregular astigmatism, spectacle lenses would not suffice to correct the resultant astigmatism (Jhanji *et al.*, 2011).

Moderate KC patients can also achieve good visual acuity with soft lenses. However, patients with more severe KC wear rigid gas permeable contact lenses. The rigid contact lenses used have single base curve lens or multiple base curve lens for moderately advanced KC patients (Garcia-Lledo *et al.*, 2006, Rathi *et al.*, 2013). For severe KC cases a scleral contact lens is required (Segal *et al.*, 2003, Visser *et al.*, 2007a, Visser *et al.*, 2007b).

1.4.8.2 Collagen cross-linking (CXL)

Collagen cross-linking (CXL) is a parasurgical approach in the treatment of KC and is less invasive than corneal transplantation. It provides a way to reduce and slow the progression of KC (Wollensak *et al.*, 2003, Cinar *et al.*, 2014, Li *et al.*, 2014, Vinciguerra *et al.*, 2014, Bonnel *et al.*, 2015, Khan *et al.*, 2015, Sedaghat *et al.*, 2015, Seyedian *et al.*, 2015, Sykakis *et al.*, 2015). This approach involves applying topical riboflavin (vitamin B_2) drops to the KC cornea, followed by a 30 minute UV-A (370nm) exposure. This technique helps to crosslink the collagen, hence stiffening the cornea and preventing further thinning and deforming (Jhanji *et al.*, 2011).

1.4.8.3 Intracorneal ring segment (ICRS)

Intracorneal rings do not stop the progression of KC yet they do have a role in stabilizing the KC ectasia (Burris *et al.*, 1991). They are implanted in the corneal stroma, without invasion of the central optical zone, in order to modify the corneal curvature. Three models of intracorneal rings are available for myopia correction: the Ferrara intracorneal ring, Bisantis segments, and Intacs. More recently the ICRS include semicircular segments, intrastromal corneal ring segments and Intacs microthin prescription inserts (Jhanji *et al.*, 2011). The first implantation of Intacs was conducted by Colin et al. in 1997 (Colin *et al.*, 2000).

1.4.8.4 Corneal Transplantation

1.4.8.4.1 Penetrating Keratoplasty

Penetrating keratoplasty (PKP) is mainly used for advanced cases of KC and has been the main treatment for KC in developed countries (Patel *et al.*, 2005, Ghosheh *et al.*, 2008, Guerin *et al.*, 2008). It consists of removing of all the corneal thickness and replacing it with a donor whole cornea. The graft survival after PKP is estimated to be 89-95% at 5-10 years (Registry, 1993).

1.4.8.4.2 Lamellar Keratoplasty

Lamellar keratoplasty is used instead of PKP in cases where Descemet's membrane is still intact and there is no hydrops or significant corneal scarring (Jhanji *et al.*, 2011). The deep anterior lamellar keratosplasty (DALK) preserves the Descemet's membrane and the endothelium, thus preventing a later transplant rejection, but removes the anterior part of the cornea and replaces it with a donor anterior cornea (Anwar and Teichmann, 2002). Three types of DALK exist: air assisted lamellar keratoplasty using intrastromal air injection (Shimazaki *et al.*, 2002, Noble *et al.*, 2007), the DALK using Melles technique (air guided deep stromal dissection) (Melles *et al.*, 1999, Melles *et al.*, 2000) and DALK using the Anwar's big-bubble technique (Foroutan and Dastjerdi, 2007, Parthasarathy *et al.*, 2008).

1.4.8.4.3 Femtosecond laser-associated keratoplasty

Femtosecond laser-associated keratoplasty (FLAK) is a femtosecond-laser assisted graft. The laser helps in creating a precise incision shape in the cornea of the KC patient undergoing the corneal transplantation thus matching the recipient's corneal dimensions (Slade, 2007, Jhanji *et al.*, 2011).

1.4.8.4.4 "Tuck in" Lamellar Keratoplasty (TILK)

"Tuck in" Lamellar Kerastoplasty or TILK is a one-stage technique used in extreme corneal ectasia and pellucid marginal degeneration. A procedure in which peripheral intrastromal pockets are created in the host cornea, where the peripheral overhang of the donor cornea is tucked (Vajpayee *et al.*, 2002, Kaushal *et al.*, 2008).

1.4.8.4.5 TILK Microkeratome-associated Lamellar Keratoplasty

TILK Microkeratome-associated Lamellar Kerastoplasty involves shaving off the surface of both donor and host KC corneas before transplantation with a microkeratome, to reduce the irregular interface induced by manual dissection (Barraquer, 1972, Haimovici and Culbertson, 1991, Springs *et al.*, 2002).

1.4.9 Genetics of KC

The qualification for a genetic predisposition to KC is based on reports of familial aggregation with prevalence of 23% approximately and increases of 3.34% of prevalence of KC in first-degree relatives (Section 1.4.2) (Wang *et al.*, 2000, Rabinowitz, 2003, Burdon and Vincent, 2013). In addition, the reports of bilateralism of the disorder (Bilgin *et al.*, 2013, Rogers and Attenborough, 2014) strongly suggest a genetic contribution in the development of KC. Moreover, in twin studies, Valluri et al studied refractive error (RE), which is a major component for KC (astigmatism) (Section 1.4.1),

in a set of 20 monozygotic and 19 dizygotic twin pairs. Valluri et al. suggested that RE presents significant genetic basis by showing that the mean difference in refractive error is significantly lower (p=0.001) in monozygotic twins (RE=0.41) compared to dizygotic twins (RE=1.53) (Valluri *et al.*, 1999). More directly, Tuft et al. investigated KC in 13 pairs of monozygotic twins and 5 pairs of dizygotic twins by studying the severity scores deduced from keratometry values. Tuft et al. showed an increased concordance (p=0.035) for the severity of KC in monozygotic twins (mean difference in severity scores = 1.4) than in dizygotic twins (mean difference in severity scores = 3.0), which further suggests genetics involvement in KC pathogenesis (Valluri *et al.*, 1999, Tuft *et al.*, 2012).

Furthermore, central corneal thickness (CCT), which has been associated with KC and thought to be a KC endophenotype (Patel and McLaughlin, 1999), has been shown by Toh et al. in a set of 256 of twin pairs, to have a major genetic factor with a heritability score of 0.95 (Toh *et al.*, 2005). Thus, the high heritability of CCT and its association with KC provide another indication of the importance of genetic component contribution in KC development.

To date, and in order to determine the genes mutated in KC, established techniques such as linkage analysis, genome wide association studies (GWAS) and candidate gene approaches based on functional evidence have been tried on genomic DNA from KC patients and controls deriving from different populations, with some successes (Burdon and Vincent, 2013, Abu-Amero *et al.*, 2014a). However, recent advances in next generation sequencing have also revealed some interesting findings.

1.4.9.1 Linkage analysis studies

In most reports, KC is transmitted as an autosomal dominant trait (Rabinowitz *et al.*, 1992, Tyynismaa *et al.*, 2002, Brancati *et al.*, 2004, Burdon *et al.*, 2008) but autosomal recessive (Wang *et al.*, 2000, Fullerton *et*

al., 2002) and X-linked inheritance have been reported (Hammerstein, 1971).

Linkage analyses on large KC families with a dominant mode of inheritance or dominant with incomplete penetrance have identified several chromosomal loci. To date, 12 studies have mapped 17 loci in large families with multiple affected KC members (Table 1.1) (Wheeler *et al.*, 2012). The high number of loci associated with KC suggests genetic heterogeneity in KC. However it should be noted that, among all the 17 loci reported and displayed in Table 1-1 along with their respective LOD scores, only three chromosomal regions, 5q21, 5q32 and 14q11, have been replicated in more than one study (Tang *et al.*, 2005b, Li *et al.*, 2006, Bisceglia *et al.*, 2009).

Table 1-1: Linkage analysis studies mapping KC to different chromosomal loci.

Locus	LOD scores	Population	Reference
1q36.32-q36.21 **8q13.1-q21.11**	Two-locus LOD score 3.4	Australian (British Descent)	(Burdon *et al.*, 2008)
2p24	5.13	Caucasian, Arab, Caribbean African	(Hutchings *et al.*, 2005)
2q13-q14.3 **20p13-p12.2**	-	Ecuadorian	(Nowak *et al.*, 2013)
3p14-q13	3.09	Italian	(Brancati *et al.*, 2004)
4q **5q32.2** **12p** **14q11.2**	-	African-American; Asian; Hispanic	(Li *et al.*, 2006)
5q14.1-q21.3	3.53	American (Caucasians)	(Tang *et al.*, 2005b)
5q21.2 **5q32-q33** **14q11.2** **15q23.2**	-	Italian	(Bisceglia *et al.*, 2009)
9p34	4.5	African-American; Asian; Hispanic	(Li *et al.*, 2006)
13q32	4.1	Ecuadorian	(Gajecka *et al.*, 2009) (Czugala *et al.*, 2012b)
14q24.3	3.58	British (Mixed)	(Liskova *et al.*, 2010b)
16q22.3-q23.1	4.1	Finnish	(Tyynismaa *et al.*, 2002)
20q12	-	Tasmanian, Australian	(Fullerton *et al.*, 2002)

The majority of these studies have not yet led to the identification of KC-associated genes at the relevant loci. However, studies detailed below have identified some potential genes mutated in KC (*miR184 and DOCK9, CAST,*

IL1RN and *SLC4A11*) highlighted through linkage analysis in four large extended families.

More recently, using a combination of linkage analysis and NGS, Hughes et al. identified heterozygous variants in 3 genes (*miR184, IREB2, DNAJA4*). The study was conducted on an autosomal dominant Northern Irish extended family with a phenotype of KC associated with anterior capsular cataract (Hughes *et al.*, 2003, Hughes *et al.*, 2011). The presence of variant r.57c>u in the seed region of *miR184*, together with a prior knowledge of the abundance of this particular *miRNA* in the cornea and lens epithelia, made it a good candidate for involvement in KC in this family (Hughes *et al.*, 2011). The *IREB2* variant was detected in the 3'UTR of the gene and *DNAJA4* was not found to play any role in the eye, so these changes were considered unlikely to be causative. Furthermore, a second family presenting dominant congenital cataract with corneal problems, which the authors termed EDICT syndrome (endothelial dystrophy, iris hypoplasia, congenital cataract, and stromal thinning) presented the same exact mutation reported for KC (Iliff *et al.*, 2012a) (Iliff *et al.*, 2012b). *miR184* was then screened in a cohort of 780 KC patients in order to determine whether this microRNA was involved in the pathogenesis of isolated KC cases. Two rare variants were found in the screened cohort, which led the authors to the conclusion that *miR184* may indeed be involved in KC pathogenesis but accounts for only 0.25% of isolated KC cases (Lechner *et al.*, 2013a).

A variant in *DOCK9* (Dedicator of Cytokinesis) was identified following linkage analysis in a large dominant Ecuadorian family by Gajecka et al., and candidate gene sequencing (Gajecka *et al.*, 2009). The heterozygous variant c.2262a>c, p.Gln754His has been shown to segregate in the family and is absent in ethnically matched controls; however this result has not been replicated by other studies (Karolak *et al.*, 2015a). This putative mutation was shown to cause aberrant splicing leading to exon skipping in one of the two DOCK9 protein isoforms (Karolak *et al.*, 2015b). This gene is

expressed in both KC and normal corneas. The precise role of *DOCK9* in normal cornea development has yet to be clarified (Czugala *et al.*, 2012a).

Li et al. screened *CAST* (Calpastatin, OMIM 114090) by looking at the candidate genes identified in a 5Mb linked region on chromosome 5q15. They reported a SNP (rs4434401) that was associated with both familial (262 cases) and sporadic (304 cases) KC in a Caucasian population (Li *et al.*, 2013b).

Nowak et al., identified through linkage analysis two novel loci 2q13-q14.3 and 20p13-p12.2 (whole-genome multipoint nonparametric linkage-NPL score= 2.395 and 2.409 respectively) segregating in a single family (9 KC patients, 9 healthy family members), and screened four candidate genes *IL1A, IL1B, IL1RN* (Interleukins 1-alpha, 1 Beta and Interleukin-1 Receptor Antagonist Protein) and *SLC4A11* in these regions (Sodium bicarbonate transporter-like protein 11). The screening resulted in the identification of a substitution in *IL1RN* (c.214+242C>T) and a novel deletion in *SLC4A11* (c.2558+149_2558+203del54). The *IL1RN* substitution was detected in nine KC affected family members and 3 unaffected members (P=0.004525), in other KC families (P=0.54) and in 10 healthy ethnically matched controls of 22 screened controls. The *SLC4A11* deletion was detected in 8 out of 9 affected KC members (P=0.00761), in 2 healthy family members and in one control out of the 22 screened ethnically matched controls. Variants in *IL1RN* and *SLC4A11*, were observed to be significantly enriched in KC more than unaffected family members, suggesting a possible involvement of these two genes in KC pathogenesis (Nowak *et al.*, 2013).

Although a few of the loci reported by researchers were reproduced independently by other groups, the majority of the regions reported to be linked by KC in the literature were not reproduced by other researchers. That might be largely due to the use of conventional linkage approaches to map complex disease traits in small family cases numbers (Altmuller *et al.*, 2001).

It might also be due to the inclusion under the KC patients' umbrella of all the different subclinical forms of KC (i.e. "forme-fruste KC", patients with thin corneas with no ectasia and simple astigmatism). The likely increased presence of phenocopies and reduced penetrance that will result may complicate linkage analysis, preventing an accurate outcome (Altshuler *et al.*, 2008). However, taking this data at face value, based on the evidence that multiple linkages have been identified, with each linkage (multiple loci in some cases) being assigned to a unique family, KC can be considered to be a genetically heterogeneous condition.

1.4.9.2 Reported KC candidate genes

Among the approaches used to understand the complex unknown aetiology of KC pathogenesis (Kriszt *et al.*, 2014), researchers have screened a number of candidate genes selected on the basis of their known role, biological functions and expression patterns in corneal and disease pathways. The potential candidate genes in KC are ones that are associated with corneal dystrophies, ocular development, ECM, collagens, apoptosis, oxidative stress related pathways and connective tissues disorders.

1.4.9.2.1 Genes reported in other corneal dystrophies and ocular development

KC has been reported to be associated with other ocular dystrophies. KC and PPCD (posterior polymorphous corneal dystrophy) association suggests a potential common genetic pathway. Mendelian mutations in Visual System Homeobox 1 (*VSX1*; OMIM 605020) and Zinc Finger E-box Binding Homeobox 1 (*ZEB1*; OMIM 189909) genes have been shown to cause PPCD (Liskova *et al.*, 2007b, Vincent *et al.*, 2009, Liskova *et al.*, 2010a, Vincent *et al.*, 2013b). Since PPCD and KC are often found to occur in the same patients, *VSX1* and *ZEB1* were screened for mutations in KC patients (Heon *et al.*, 2002a, Lechner *et al.*, 2013c). Putative mutations were found to be associated with the KC phenotype. Several studies assessed the association of *VSX1* variants with KC (Table 1-2) in several populations

(Table 1-2); however the variants found in *VSX1* account for only 2-3% of KC cases and most of the variants identified are common polymorphisms (Abu-Amero *et al.*, 2014a). The role of *VSX1* variants in KC pathogenesis therefore requires further clarification and to date *VSX1* has not been shown to be involved in KC pathogenesis (Aldave, 2005, Aldave *et al.*, 2006, Romero-Jimenez *et al.*, 2010, Kok *et al.*, 2012, Karolak *et al.*, 2015a). As for the *ZEB1* gene, Muszynska et al. reported putative mutations in *ZEB1* in unrelated KC European patients (Muszynska *et al.*, 2011).

Table 1-2: *VSX1* variants reported in KC patients in various populations by different research groups.

Gene	Zygosity	Mutation	Population	Reference
	Dominant	R166W, L159M, G160D	-	(Heon *et al.*, 2002a)
	Dominant	D144E, G160D, P247R, L17P	Italian	(Bisceglia *et al.*, 2005)
	-	None	-	(Aldave, 2005)
	Dominant	None	English	(Liskova *et al.*, 2007a)
	-	None	Mixed	(Tang *et al.*, 2008)
	Dominant	N151S, G160V	Korean	(Mok *et al.*, 2008)
	Dominant	D144E	Jewish	(Eran *et al.*, 2008)
	Dominant	Q175H	Indian	(Paliwal *et al.*, 2009)
	-	p.G160D	European	(Dash *et al.*, 2010)
	-	None	Slovenian	(Stabuc-Silih *et al.*, 2010)
	-	None	North Indian	(Tanwar *et al.*, 2010)
	Recessive	Q175H	Indian	(Paliwal *et al.*, 2011)
	Dominant Recessive	None	Saudi Arabian	(Abu-Amero *et al.*, 2011b)
	-	p.G239R	Italian	(De Bonis *et al.*, 2011)
	Dominant	p.R166W p.H244R	Iranian	(Saee-Rad *et al.*, 2011)
VSX1	-	Leu17Val Val199Leu Gly160Val	Korean	(Jeoung *et al.*, 2012)
	-	rs56157240 rs12480307 rs6050307	Han Chinese	(Wang *et al.*, 2013a)
	-	c.546A>G (rs12480307), c.627+23G>A (rs6138482), c.627+84T>A (rs56157240) c.504-24C>T (IVS3-24C)	South Indian	(Verma *et al.*, 2013)
	-	c.731A>G (p.His244Arg)	Australian	(Vincent *et al.*, 2013b)
	-	None	Greek	(Moschos *et al.*, 2015)
	-	H244R, rs6138482 c.546A>G (rs12480307)	Iranian	(Dehkordi *et al.*, 2013)
	-	c.-264_-255delGGGGTGGGGT, c.627 + 23G > A c.809-6_809-5insT, c.*200G > T	Polish	(Karolak *et al.*, 2015a)
	Dominant	p.Leu268His p.Ser251Thr	Indian	(Shetty *et al.*, 2015a)

1.4.9.2.2 Extracellular Matrix genes

The transforming growth factor beta induced gene (*TGFBI*; OMIM 190180) encodes an ECM protein, mutations in which are the cause of several corneal dystrophies. The investigation of the *TGFBI* gene by direct sequencing in KC patients revealed contradictory results in different populations. The most convincing data is the identification of a heterozygous nonsense mutation (G535X) in one patient out of 30 from a Chinese cohort (Guan *et al.*, 2012).

1.4.9.2.3 Collagens

One theory of KC pathogenesis is that the alteration in collagen structure/function is due to a genetic factor involved in collagen modification in KC corneas. Thus, collagen genes have been investigated in KC patients without much success in identifying a variant responsible for KC development. *COL4A3* (OMIM 120070) and *COL4A4* (OMIM 120131) showed no pathological mutations in 104 European unrelated KC patients, though a significant increase in allele frequency was detected for the D326Y variant ($p<0.0001$) of *COL4A3* and for M1237V ($p<0.0001$) and F1644F ($p=0.002$) variants in *COL4A4* in KC patients (Stabuc-Silih *et al.*, 2009). The same result was not confirmed by Wang et al. and Kokolakis et al., where they rejected the association of *COL4A3* and *COL4A4* with KC (Wang *et al.*, 2013a) (Kokolakis *et al.*, 2014). Likewise, the role of *COL4A1* (OMIM 120130) and *COL4A2* (OMIM 120090) (Karolak *et al.*, 2011) and *COL8A1* (OMIM 120251) and *COL8A2* (OMIM 120252) (Aldave *et al.*, 2007) in KC has been excluded by Sanger DNA sequencing. Recently, Saravani et al., associated SNP rs2229813 and the GA+AA genotypes in *COL4A4* as a risk factor for KC in the Iranian population (Saravani *et al.*, 2015).

1.4.9.2.4 Apoptosis related pathways

As previously described, atopy is a risk factor for KC and association between KC and atopy has been reported to be seen in as many as 50% of KC cases (Droitcourt *et al.*, 2011) (Section 1.4.5). In addition, eye rubbing

has been associated with KC (Section 1.4.5) and it has been shown that mechanical injury can induce keratocyte apoptosis in KC corneas by the *IL1* (Interleukin 1) pathway (Wilson *et al.*, 1996a). In parallel, *FLG* (Filaggrin; OMIM 135940) mutations are thought to cause atopic dermatitis, with the *FLG* protein expressed in the corneal epithelium (Liang *et al.*, 2015). Thus *FLG* and *IL1* have been considered as KC candidate genes and investigated in KC patients. Additionally, Droitcourt et al. screened *FLG* in 89 KC patients and found two heterozygous loss-of-function *FLG* mutations (R501X and 2282del4) (Droitcourt *et al.*, 2011).

For the *IL1* genes, which lie within a 70 kb region on chromosome 2q13, *IL1B* (OMIM 147720) promoter polymorphisms rs1143627 (c.−31t>c) (p=0.025) and rs16944 (c.−511c>t) (p=0.022) have been reported to increase the risk of KC in 100 unrelated Korean patients, while the *IL1A* (OMIM 147760) intronic polymorphism c.+376c>a (rs2071376) was significantly different between KC patients and controls (p=0.034) (Kim *et al.*, 2008). The same *IL1B* (rs16944) polymorphism to was screened in a 121 Turkish KC patients and 121 controls, and association between *IL1B* and KC was rejected (p=0.569) (Palamar *et al.*, 2014). However, Mikami et al., studied the polymorphisms reported by Kim et al. of both *IL1B* and *IL1A* polymorphisms in 169 Japanese patients and 390 controls and replicated the association of the *IL1B* promoter polymorphism rs1143627 (c.−31t>c) (p=0.014, corrected p=0.043) (Mikami *et al.*, 2013). Similarly, and very recently, the same variants in *IL1B* and *IL1A* were investigated in a 101 unrelated Chinese Han patients and 101 controls by Wang et al., who established a significant association (rs1143627 (C>T p=0.017), rs16944 (A>G p=0.002) and rs2071376 (A>C p=0.017)) between *IL1* and risk of KC (Wang *et al.*, 2015b). Functional analysis must be conducted in order to confirm a strong link between the pathogenesis of KC and *IL1*.

1.4.9.2.5 Oxidative Stress genes

In an effort to identify genes responsible for KC, Udar *et al.* screened the *SOD1* (Superoxide dismutase1, OMIM 147450) gene as a candidate gene for KC since oxidative stress has been hypothesised to play a role in KC development (Kenney *et al.*, 2000, Atilano *et al.*, 2005, Wojcik *et al.*, 2013b). *SOD1* encodes the cytosolic Cu/Zn-SOD, a protein that binds copper and zinc ions and responsible for destroying free superoxide radicals in the body (You *et al.*, 2010). As previously mentioned, KC corneas have underlying defects in their ability to process ROS and undergo oxidative damage that is translated into an increased oxidative stress (Cristina Kenney and Brown, 2003). Udar et al. reported a 7-base deletion in intron 2 of the *SOD1* gene in two pedigrees within which three KC patients were available for testing (Udar *et al.*, 2006). However this finding was not consistent with other cohorts (Stabuc-Silih *et al.*, 2010, De Bonis *et al.*, 2011, Saee-Rad *et al.*, 2011, Al-Muammar *et al.*, 2015, Moschos *et al.*, 2015). Even if the link with *SOD1* was proven, it would only account for 1% of KC patients.

Recently, Synowiec et al. suggested the genes *MUTYH* (mutY DNA glycolase, OMIM 608456) and *hOGG1* (8-oxoguanine DNA glycosylase, OMIM 601982) that play a role in oxidative DNA damage repair as candidates for involvement in KC pathogenesis. They therefore screened both genes in 205 Polish Subpopulation patients and 220 controls, but didn't find any association ($p<0.05$) between the mentioned genes and KC pathogenesis (Synowiec *et al.*, 2015).

Wojcik et al. hypothesised that base excision repair genes (BER) may play a role in KC pathogenesis, since KC corneas have lost their ability to effectively cope with oxidative stress and BER is usually the mechanism by which the damage caused by oxidative stress is removed. Thus, 4 BER genes were checked for mutations by genotyping 5 polymorphisms in 284 KC patients and 353 controls: g.46438521G>C in 3' near gene *NEIL1* (Nei endonuclease VIII-like 1, OMIM 608844), exonic c.2285T>C in *PARP-1* (poly(ADP-ribose) polymerase-1, OMIM 613867), c.–1370T>A in the 5' near

gene *POLG* (DNA polymerase γ, OMIM 174763) and exonic c.580C>T and c.1196A>G in *XRCC1* (X-ray repair cross-complementing group 1, OMIM 194360). This study concluded that changes in the *XRCC1* and *POLG* genes may constitute a risk factor in KC since the A/A genotype of the c.–1370T>A polymorphism of *POLG* and the A/G genotype and the A allele of the c.1196A>G polymorphism of the *XRCC1* were associated with increased risk of KC (Wojcik *et al.*, 2014b).

1.4.9.2.6 Other candidate genes

Additionally, some studies looked up the regions that were previously reported in linkage analysis and screened the genes lying in the mapped region that seemed to have a particular involvement in corneal pathways. De Bonis et al. screened *SPARC* (Secreted Protein Acidic and Rich in Cysteine, OMIM 182120) and *LOX* (Lysyl Oxidase, OMIM 153455) localized on chromosome 5q23.2 and 5q31.3-q32, respectively, in regions that were previously implicated by linkage analysis in familial KC (Li et al., 2006, Bisceglia et al., 2009). De Bonis et al., excluded the involvement of *LOX* in KC pathogenesis but reported a novel potential deleterious heterozygous mutation p.M92I in *SPARC* in one patient out of 302 Italian KC patients studied and absent in 200 screened controls, suggesting a possible role for this gene in KC pathogenesis that needs further investigation (De Bonis et al., 2011).

The *STK24* (Serine/Threonine Kinase 24, OMIM 604984) and *IPO5* (Importin 5, OMIM 602008) genes lie in the 13q32 locus identified by Czugala et al. in an Ecuadorian family, *IOP5* (c.2377-132A>C) and *STK24* (c.1053+29G>C) and both have been reported in introns in 10 KC affected patients out of 10, in 1 unaffected family member and in 1 family member of unknown phenotype. In addition, the *IOP5* intronic change has been identified in 2 KC patients belonging to different KC families than the original one investigated and in 5 controls out of 105. The *STK24* intronic change has been identified in 1 control out of 105 (Czugala et al., 2012b).

Therefore, Karolak et al. investigated the involvement of both genes but concluded that neither gene was mutated in KC patients from the Polish population (Karolak *et al.*, 2015a).

In addition, an increase in copy number variation of mitochondrial DNA (mtDNA), but not haplogroup, has been shown to be associated with KC (Hao *et al.*, 2015b). In addition, Abu-Amero et al. sequenced the full mitochondrial genome in 26 Saudi patients genome and found 10 synonymous variants in 10 different KC patients and absent in controls. Four changes were present in mitochondrial complex I, two variants in $tRNA^{Leucine2}$, one variant in $tRNA^{Histidine}$, one variant in $tRNA^{Glutamine}$, one variant in $tRNA^{Tryptophane}$ and one variant in $tRNA^{Aspargine}$ (Abu-Amero *et al.*, 2014b). In case these results are confirmed in a large cohort and screened in a larger number of cohorts, they highlight a possible involvement of mitochondrial sequence variations in KC pathogenesis.

1.4.9.3 Association studies

Several GWASs were recently conducted, through which a number of genes and loci were shown to be significantly associated with KC (Burdon *et al.*, 2011, Bykhovskaya *et al.*, 2012, Li *et al.*, 2012b, Bae *et al.*, 2013, Lu *et al.*, 2013, Lechner *et al.*, 2014, Sahebjada *et al.*, 2014). The first KC GWAS was conducted in 2011. By combining two genome-wide scans undertaken in parallel on Australian, American KC patients' cohorts, Burdon et al. highlighted a SNP (rs3735520) at the *HGF* (Hepatocyte Growth Factor, OMIM 142409) locus, which was associated ($P= 9.9\times10^{-7}$) with KC (Burdon *et al.*, 2011). The significant association ($p=8.1\times10^{-3}$) of *HGF* (rs5745752) with increased risk KC was later confirmed by Sahebjada et al. in an Australian cohort of 157 KC patients and 673 healthy controls (Sahebjada *et al.*, 2014).

In 2012, Li et al. identified a SNP (rs4954218) located near *RAB3GAP1* (RAB3 GTPase activating protein subunit 1, OMIM 602536) $P= 1.6 \times 10^{-7}$ in

a study including 222 Caucasian KC patients and 3324 controls (Li *et al.*, 2012b). This same finding was replicated in 2013 in an Australian Caucasian cohort (524 KC patients and 2761 controls) and in a combined meta-analysis by showing a significant association of SNP rs4954218 near *RAB3GAP1* ($P = 9.26 \times 10^{-9}$) (Bae *et al.*, 2013).

A meta-analysis conducted in 2013 by Lu et al. combined data from multiple GWASs of CCT to identify 16 new CCT loci. All of the CCT loci highlighted in this analysis were then tested for association in 874 KC cases and 6,085 controls. Six of the CCT associated SNPs were also significantly associated with KC, including: Forkhead Box O1-*FOXO1* (OMIM 136533, SNP rs2721051), Fibronectin Type III Domain Containing 3B-*FNDC3B* (OMIM 611909, SNP rs4894535), Retinoid X receptor Alpha-Collagen type V alpha *RXRA-COL5A1* (OMIM 180245; 120215, SNP rs1536482), Multiple PDZ domain protein-Nuclear Factor I/B *MPDZ-NFIB* (OMIM 603785; 600728, SNP rs1324183), Collagen type V alpha 1 *COL5A1* (OMIM 120215, SNP rs7044529) and Zinc Finger Protein 469 *ZNF469* (OMIM 612078, SNP rs9938149) (Lu *et al.*, 2013). The *MPDZ-NF1B* locus (SNP rs1324183) was replicated (p=0.005) in a Han Chinese cohort (210 KC patients and 191 controls) (Hao *et al.*, 2015a). Moreover, Lechner et al. reported in a cohort of significant enrichment of potentially pathogenic heterozygous alleles in *ZNF469* associated (P = 0.00102) with KC using Sanger sequencing on European ethnicity patients (112 patients and 784 controls) (Lechner *et al.*, 2014).

Furthermore, a locus on chromosome 5q has been associated with KC (Rabinowitz *et al.*, 1992, Li *et al.*, 2006) and confirmed recently (Bykhovskaya *et al.*, 2012). For this locus, Bykhovskaya et al. extrapolated from published GWAS data in a Caucasian cohort to support their hypothesis that the *LOX* gene, encoding the protein Lysyl Oxidase, is involved in KC pathogenesis. They noted that *LOX* is located under their linkage peak at 5q23.2, and identified two KC associated SNPs (rs10519694 and rs2956540) in *LOX*, following a combined family based and case-control

association study (Meta p= 4.0×10^{-5} and 4.0×10^{-7}) (Bykhovskaya *et al.*, 2012).

The previously mentioned recent GWAS studies (Burdon *et al.*, 2011, Bykhovskaya *et al.*, 2012, Li *et al.*, 2012b, Bae *et al.*, 2013, Li *et al.*, 2013b, Lu *et al.*, 2013, Sahebjada *et al.*, 2014) were successful in confirming existing KC associated loci and highlighting a number of new genetic associations in candidate genes implicated in KC pathogenesis. However, while this approach is able to detect the influence of common alleles in the form of polymorphic SNPs on KC, these will generally have only modest effect sizes. In contrast, this approach will miss rare variants with more significant effect sizes. Nonetheless rare variants with major effect may contribute significantly to the genetics of KC, a phenomenon has been invoked to explain the missing heritability in a range of common multifactorial diseases like KC (Eichler *et al.*, 2010, Zuk *et al.*, 2012).

1.5 Project Aim

The aim of this project is to identify the genetic basis of KC in human patients. The approach used initially was to collect KC families belonging to highly consanguineous and endogamous populations with multiple KC affected members, then explore the genomic DNA of KC patients and controls via whole exome next generation sequencing (WES) in order to highlight genetic variants that might account for KC pathogenesis. The second aim of this project was to establish a transcriptome profile via RNA-seq for KC and normal human corneas, in order to build the normal human corneal transcriptome profile and determine the extent of differential gene expression in KC corneas. In a third approach, WES and RNA-seq data were combined in samples obtained from the same patient to look for genetic variants that exist in genes, which are differentially expressed.

2 Materials and Methods

2.1 DNA sampling in Patients

All patients included in the study were diagnosed by an ophthalmologist following a topographical examination. The local Yorkshire KC patients and healthy controls were identified and blood or saliva sampled by Dr. Aine Rice and Miss Salina Siddiqui (the Eye Department, St. James's University Hospital, Leeds). DNA was extracted by the Yorkshire Regional Genetics Service on a research basis. Blood from the Lebanese KC families (patients and healthy family members) and healthy controls belonging to the same ethnicities as the patients were sampled by qualified phlebotomists during a field trip to Lebanon, some of the Lebanese controls were collected in the UK by contacting the British Druze society and personal contacts. DNA was extracted at the American University of Science and Technology (AUST) laboratories and subsequently brought to Leeds, through a collaboration with Dr. Gretta Abou-Sleymane from AUST as well as Lebanese ophthalmologists: Dr. Naji Waked, Dr. Bassam Barakat and the late Dr. Ghassan Shayya. The Lebanese ethical approval to sample patients and extract DNA is under the governmental licence number 3087-Minister of Health-decree number 1/211-19/4/2005. Informed consent was obtained from all subjects involved in the study using a process that was approved by the Leeds East Research Ethics Committee (Reference number 10/H1306/63 under the title: "Investigation of human inherited corneal dystrophies: genetic, tissue and transplant analyses"). The research followed the tenets of the Declaration of Helsinki.

2.2 DNA extraction

2.2.1 From Blood

Blood samples for DNA extraction (2-6ml depending on age) were collected by venepunture from the antecubital fossa and stored in BD vacutainer® EDTA blood collection tubes (BD Biosciences, Oxford UK). Genomic DNA was extracted in Lebanon at the AUST laboratory using the phenol-

chloroform extraction. This protocol involved adding 1000µl of Red Blood Cell Lysis Buffer (0.155M Ammonium Chloride, 0.01M Potassium Bicarbonate and 0.1mM EDTA) to every 350µl of Blood and mixing by inversion. The samples were incubated for 2 minutes at room temperature, mixed by inversion and centrifuged for 2 minutes at 12,000 rpm. The supernatants were discarded into disinfectant then each pellet was resuspended in 190µl of 0.2M of ammonium acetate and vortexed briefly. Next, 13µl of 10% SDS detergent and 2.5µl of proteinase K (20mg/ml H2O) (Sigma P-0390) was added and the mix was briefly vortexed and incubated for 1hour at 55ºC. Once the samples had cooled, 60µl phenol/chloroform/isoamyl alcohol was added then the mix was vortexed for 30 seconds and centrifuged for 2 minutes at 12,000 rpm. The aqueous layer was carefully removed to a new 1.5ml microcentrifuge tube, and 500µl of cold 100% ethanol was added in order to precipitate the DNA. The samples were then mixed, and incubated for 15 minutes at -20ºC. Subsequently, the samples were centrifuged for 2 minutes at 12,000 rpm in a microcentrifuge, and the supernatants were discarded. The pellets were redissolved in 90 µl of 10:1 TE buffer, vortexed, and incubated at 55ºC for 10 minutes. 10 µl of protein precipitation solution (2M sodium acetate) was added and the samples were mixed. The DNA was precipitated by 250 µl of cold 100% ethanol, mixed, and centrifuged for 1 minute at 12,000 rpm. The supernatants were discarded and the pellets were rinsed with 500µl of 80% ethanol, centrifuged for 1 minute at 12,000 rpm and rewashed with 500µl of 70% ethanol. Pellets were again centrifuged for 1 minute at 12,000 rpm in a microcentrifuge. The ethanol was removed and the DNA pellets were left to air-dry. Once dry the DNA pellets were dissolved in 100µl of 10:1 TE buffer and left overnight at room temperature. DNA samples were later stored at -20˚C.

For Leeds samples, DNA was extracted from peripheral blood samples by Yorkshire Regional Genetics, Saint James's University Hospital using a salt precipitation technique. The salt precipitation technique briefly corresponds to extracting DNA from fresh non frozen blood samples, by aliquotting 3ml of

blood into polypropylene tubes and add 9ml of Red Cell Lysis solution. The samples were then mixed for 10 minutes at RT and then centrifuged at 3,200 rpm for 10 minutes. The supernatant was removed. The remaining white cells pellet was re-suspended in 3ml of White Cell Lysis solution, and cells were lysed by pipetting. Add 1ml of Protein Precipitation Solution (10M Ammonium Acetate) to the mix in order to remove any contaminating proteins, and samples were vortex mixed for 20 seconds then centrifuged at 3,200rpm for 10 minutes. The supernatant containing the DNA was transferred into a fresh tube and 3ml of isopropanol was added for DNA precipitation. The samples were then centrifuged at 3,200rpm for 10 minutes followed by two washes in 70% ethanol. The precipitated DNA was air dried and redissolved in 200μl of 1xTE buffer, incubated at 55°C for 4 hours in a shaking water bath to complete re-suspension then left for 1hour at RT before storage.

2.2.2 From Saliva

From saliva, genomic DNA was collected using Oragene kits (Oragene.ONE ON-500, DNA Genotek) and extracted following the manufacturer's protocol. Briefly this involved transferring 750μl of saliva to a 1.5 ml tube, then adding 30μl of PT-L2P solution supplied with the kits. The samples were incubated on ice for 10 minutes then centrifuged at room temperature for 10 minutes at 3500 x g. 600μl of the supernatants were then transferred to a fresh tube, and the pellets were discarded. 720μl of 100% ethanol was added and the mix was inverted 10 times. Subsequently, the samples were incubated for 10 minutes at room temperature then centrifuged for 10 minutes at 3500 x g. The supernatant was discarded and 500μl of 70% ethanol was carefully added to the pellet, which was then left to stand for 1 minute at room temperature. The ethanol was completely removed without disturbing the pellet and smear. Once dry, the pellet was re-dissolved in 50μl of TE. The samples were incubated at room temperature overnight to ensure complete rehydration and transferred to a 1.5mL tube. DNA samples were later stored at -20˚C.

2.3 Human corneal tissue collection and RNA extraction

Under the same Leeds East Research Ethics Committee ethical approval referenced in section 1 (Reference number 10/H1306/63), the stroma, Bowman's layer and epithelium tissues of five normal cadaveric human corneas were donated by five different previously healthy individuals. In addition, six KC corneas obtained with the consent from six different patients undergoing corneal transplantation at the eye department at St James's Hospital were obtained. The healthy donated corneas from cadavers were used in descemet stripping automated endothelial keratoplasty (DSEK), a treatment for corneal endothelial dystrophy in which the inner endothelial layer of the cornea is transplanted, leaving the anterior part of the cornea. The enucleation procedure from cadavers donor's eye (WT corneas) takes place as soon as possible to a maximum of 24 hours after death. Before enucleation, the eye of the donor is cleaned with 3% PVPI (Povidone-Iodine) and 0.3% $Na_2S_2O_3$ (Sodium thiosulphate) then excised and preserved in Minimum Essential Medium Eagle 2% FBS (Fetal Bovine Serum) at 34 degrees for a period of time that can go no longer than 28 days. The keratoconus corneas were surgically removed in theatre from patients treated by deep anterior lamellar keratoplasty (DALK) where all corneal tissue is removed down to Descemet's membrane, leaving only the healthy endothelial layer. Healthy and KC corneal tissue samples were therefore equivalent, representing the entire cornea except the endothelial cell layer. The enucleation of both KC and WT corneas was done following the "Standards for the retrieval of human ocular tissue used in transplantation, research and training, The Royal College of Ophthalmologists" (Figueiredo, 2013).

Blood samples were also collected from the six KC patients sampled in this way, from which genomic DNA samples were extracted as described above. Corneal tissues were preserved in RNAlater (Ambion) from which RNA was extracted using the Trizol reagent (Invitrogen), following the manufacturers' protocols. Briefly this involved adding 1mL of TRIzol® Reagent per 50–100 mg of tissue sample then homogenizing the sample by vortexing. The

homogenized samples were incubated for 5 minutes at room temperature to permit complete dissociation of the nucleoprotein complex and 0.2 mL of chloroform were added to each sample. Samples were then mixed vigorously by hand for 15 seconds and incubated for 2–3 minutes at room temperature, then centrifuged at 12,000x g for 15 minutes at 4°C. The aqueous phase where the RNA resides was removed from the sample by angling the tube at 45° and pipetting the solution out into a new tube where 0.5 mL of 100% isopropanol was added. Samples were then incubated at room temperature for 10 minutes, and centrifuged at 12,000x g for 10 minutes at 4°C. The supernatants were discarded and the pellets were washed with 1mL of 75% ethanol, briefly vortexed and then centrifuged at 7500 x g for 5 minutes at 4°C. The wash was then discarded and the RNA pellets were left to dry for 5–10 minutes. The RNA pellets were next redissolved in RNase-free water by pipetting, then incubated in a water bath or heat block set at 55–60°C for 10–15 minutes. RNA samples were then stored at –70°C.

2.4 PCR

2.4.1 Standard PCR

Oligonucleotide primers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm). DNA was amplified by PCR with a final reaction volume of 25µl using 40ng of genomic DNA, 1x PCR buffer (Invitrogen), 0.4mM $MgCl_2$, 200µM each dNTPs (Invitrogen), 10picomoles of each primer, and 1.0 U Taq DNA polymerase (Invitrogen). PCR amplification initiates with a denaturation step at 95°C for 2 minutes followed by 40 cycles of 94°C for 30 seconds, annealing temperature for 45 seconds, 72°C for 45 seconds and ending with a final elongation step at 72°C for 5 minutes.

2.4.2 Hot-shot Master Mix

Reactions using HotShot master mix (Clent, Lifescience) were carried out in a 25µl volume using manufacturer's instructions, of the hot shot master mix

and primers solution using the same primer concentration as section 4.1. PCR amplification initiates with an initial denaturation step at 95°C for 12 minutes followed by 40 cycles of 94°C for 30 seconds, annealing temperature for 30 seconds, 72°C for 45 seconds and ending with a final elongation step at 72°C for 5 minutes.

2.5 Agarose gel electrophoresis

PCR products were visualized under ultraviolet light following agarose gel electrophoresis. Molecular biology grade agarose powder (Bioline) was dissolved in 0.5x TBE and ethidium bromide was added to a final concentration of 0.5 µg/ml. Gels of 1.5% concentration were run on 100Volts for 40 minutes, then visualised on a Bio-Rad gel documentation system with an ultraviolet transilluminator and displayed using Quantity One 1-D analysis software (Bio-Rad, Hercules, USA).

2.6 Sanger Sequencing

PCR product clean-up was carried out using ExoSAP-IT (Affymetrix). Purified PCR products were then sequenced using one of the same primers designed for the PCR (forward or reverse), or an internal nested primer, using dye terminator chemistry with the BigDye terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). A final reaction volume of 10µl was subsequently subjected to an initial denaturation step of 96°C for 1 minute, followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. After the sequencing reaction, the products were precipitated using 5µl of 125mM EDTA and 60µl of absolute ethanol. Sample tubes were centrifuged at 3000 x g for 30 min at room temperature, then spun with 70% ethanol at 800 x g for 15min at 4°C. After being dissolved in Hi-Di formamide (Applied Biosystems), the sequencing reactions were analyzed using an ABI PRISM 3130xl automated capillary sequencer (Applied Biosystems) and

checked for changes using Sequencing Analysis v.5.2 (Applied Biosystems) and SeqScape v.2.5 (Applied Biosystems) software.

2.7 Homozygosity mapping by SNP array analysis

500ng of genomic DNA at a concentration of 50ng/µl extracted from both Saliva and/or blood from affected patients was anonymously analysed on Affymetrix 6.0 high-density SNP microarrays by a commercial company (AROS Biotechnologies Ltd, Aarhus, Denmark). The resulting SNP data reformatted into excel files from their original CEL files, were analysed using IBDfinder software (http://dna.leeds.ac.uk/ibdfinder/) (Carr *et al.*, 2009) to determine the homozygous regions at a genotyping error of 1 in 100, and results were visualised and presented using AutoIdeogram (http://dna.leeds.ac.uk/AutoIdeogram/) and AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/).

First described in 1987 by Lander and Botstein, homozygosity mapping aimed to seek homozygous regions in the genome of affected individuals from inbred consanguineous families, with the purpose of discovering autosomal recessive gene loci (Lander and Botstein, 1987). Called autozygosity mapping later on by Muller and Bishop in 1993, it comprises looking at runs of homozygosity in chromosomal regions thought to be autozygous or Identical-by-descent (IBD) (Figure 2-1), to map the location of disease-causing genes (Mueller and Bishop, 1993). Autozygosity mapping has been very successful in identifying disease causing mutations in autozygous segments of affected consanguineous families members, even in the presence of genetic heterogeneity (Moynihan *et al.*, 1998, White *et al.*, 2007, Khan *et al.*, 2011, Rehman *et al.*, 2011, Sherva *et al.*, 2011, Logan *et al.*, 2014, El-Asrag *et al.*, 2015).

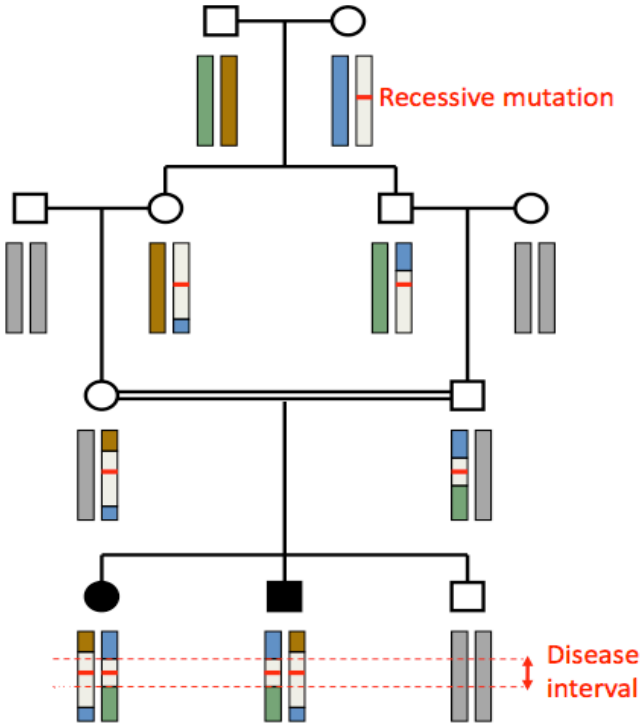


Figure 2-1: Schematic presentation of the concept of autozygosity mapping. The colour coding of haplotypes is depicted by SNPs (or genetic markers in general i.e. microsatellites). The common ancestral haplotype shown in white is harbouring the recessive mutation (red bar), which is transmitted as blocks to the next generations following the crossing over events. The consanguineous union in generation three of two individuals carrying heterozygous haplotypes leads to the segregation of the two copies of the recessive mutation present in the shared haplotype or autozygous segment presented as the disease interval. Taken from (http://autozygosity.org/about/)

2.8 Whole Exome Sequencing (WES)

For whole exome capture and next generation sequencing, 3µg of genomic DNA was either sent to Otogenetics Corporation (Atlanta, Georgia, United States), who provided an average coverage of 30X using SureSelectXT Human All Exon V4 reagent (Agilent), or libraries were prepared in-house. In-house library preparation (Figure 2-2) began with mechanically shearing the DNA. First 3µg of genomic DNA was dissolved in 250µl of 1X Low TE Buffer in a 1.5-mL LoBind tube. This was then run on a Covaris Sample Preparation System Covaris™ S220 for DNA shearing of a size selection of 150-200 bp fragments. Next the samples were purified using Agencourt AMPure XP beads, and their quality was assessed with 2100 Bioanalyzer by loading 1µl of the sample on a Bioanalyzer DNA 1000 chip and visualizing the results on Agilent 2100 Expert Software by generating an

electrophoregram specific for each sample. The electropherogram should show a distribution of a peak size between 150 to 200 nucleotides. In order to repair the ends and achieve blunt-ended fragments with 5'-phosphorylated ends, 48µl of each sample were mixed with the following reagents from the SureSelect Library Prep Kit, ILM (Illumina): 32.2µl of Nuclease-free water, 10µl of 10x end repair buffer, 1.6µl of dNTP mix, 1µl of T4 DNA polymerase, 2µl of Klenow DNA polymerase and 2.2µl of T4 Polynucleotide Kinase. The mix was incubated in a thermal cycler for 30 minutes at 20°C. After the incubation, the samples were purified using Agencourt AMPure XP beads. To 30µl of the purified samples, 11µl of Nuclease-free water, 5µl of 10x Klenow Polymerase Buffer, 1µl of dATP and 3µl of Exo(-) Klenow were added in order to achieve DNA fragments with a 3'-dA overhang, then the mix was incubated in a thermal cycler for 30 minutes at 37°C. Subsequently, samples were purified and indexing-specific paired-end adaptor were ligated by adding to 13µl of the treated DNA samples 15.5µl of Nucleas-free water, 10µl of 5x T4 DNA ligase buffer, 10µl of SureSelect Adaptor Oligo Mix and 1.5µl of T4 DNA Ligase and incubating for 15 minutes at 20°C in a thermal cycler. Next, samples were purified and the 15µl of the adaptor-ligated libraries were amplified by mixing 21µl of Nuclease-free water, 1.25µl of SureSelect Primer from the SureSelect Library Prep Kit, 1.25µl of SureSelect ILM Indexing pre capture PCR reverse primer from the SureSelect target Enrichment Kit ILM Indexing Hy b Module Box#2, 10µl of 5x Herculase II Rxn Buffer, 0.5µl of 100mM dNTP mix and 1µl of Herculase II fusion DNA polymerase from the Herculase II Fusion DNA Polymerase kit (Agilent). The mix was then loaded into a thermal cycler for a denaturation step of 2 minutes at 98°C, amplification step of 30 seconds at 98°C followed by 30 seconds at 65°C and 1 minute at 72°C, repeated for 6 times, and ending with an elongation step of 10 minutes at 72°C. As a quality control step before incurring the cost of a run on the sequencer, samples were purified and assessed for quality and quantity with an Agilent 2100 Bioanalyzer instrument. To attain the quality required to proceed to sequencing, the electropherogram was required to show a single peak in the size range of 250 to 275 bp.

For hybridization, 3.4µl of 750ng of amplified DNA (221ng/µl) was mixed with 40µl of the hybridization buffer prepared by from SureSelect Hyb buffers #1-4, 2µl of SureSelectXT Human All Exon V4 (4Gb of targeted region) (Agilent) capture library mixed with 2µl of 1:3 RNase Block Dilution:water and 5.6µl of SureSelect Block Mix prepared by mixing SureSelect Indexing Block #1-3. The mix was then incubated for 5 minutes at 95°C and 65°C for 24hours in a thermal cycler with a heated lid of 105°C.

After hybridization, hybrid capture selection was achieved by adding the remaining volume of approximately 25-20µl of the hybridization mixture to the previously prepared and washed magnetic bead solution while maintaining the tube at 65°C during the process. Samples of hybrid-capture/bead solution were incubated on a Nutator mixer for 30 minutes at room temperature. Beads and buffer were separated by using a magnetic separator, where the supernatant was discarded and beads were resuspended in 500µl of SureSelect Wash1 Buffer and incubated for 15 minutes a room temperature. Beads were washed for a total of 3 washes with 500µl of SureSelect Wash2 and recuperated by separating the solution from beads on a magnetic separator and dissolved in 30µl of nuclease-free water.While attached to the beads, captured library was amplified to add index tags, by adding to the 14µl of the beads solution 22.5µl of Nucleas-free water, 10µl of of 5x Herculase II Rxn Buffer, 0.5µl of 100mM dNTP mix and 1µl of Herculase II fusion DNA polymerase from the Herculase II Fusion DNA Polymerase kit (Agilent), 1µl of SureSelect ILM Indexing post capture forward PCR primer from the SureSelect target Enrichment Kit ILM Indexing Hy b Module Box#2, and 1µl of PCR primer Index 1 through Index 16 (depends on the number of samples) from SureSelect Library Prep Kit, ILM. The mix was then loaded into a thermal cycler for a denaturation step of 2 minutes at 98°C, followed by amplification achieved by cycling through 30 seconds at 98°C followed by a 30 seconds at 57°C and 1 minute at 72°C, repeated for a total of 12 times, and ending with an elongation step of 10 minutes at 72°C. Next samples were purified and assessed for quality and quantity with the Agilent 2100 Bioanalyzer High Sensitivity DNA assay using

Bioanalyzer High Sensitivity DNA assay chip and reagent kit. The Amplified Capture DNA was required to show an electropherogram with a peak in the size range of 300-400 nucleotides before proceeding to sequencing with the sample. 6 samples per lane were then pooled using the following formula to calculate the volume needed for each sample in order to achieve a 10nanoMolar of final concentration in a 50 µl final volume: Volume = (V(f) x C(f)) / (6*C(i)) where V(f) is the final desired volume of the pool, C(f) is the desired final concentration of the pool which is 10nanoMolar and C(i) is the initial concentration of each index sample. The pooled samples were then passed to the Leeds Sequencing Facility in order to prepare the sample for cluster amplification and load it on an Illumina HiSeq 2000/2500 platform.

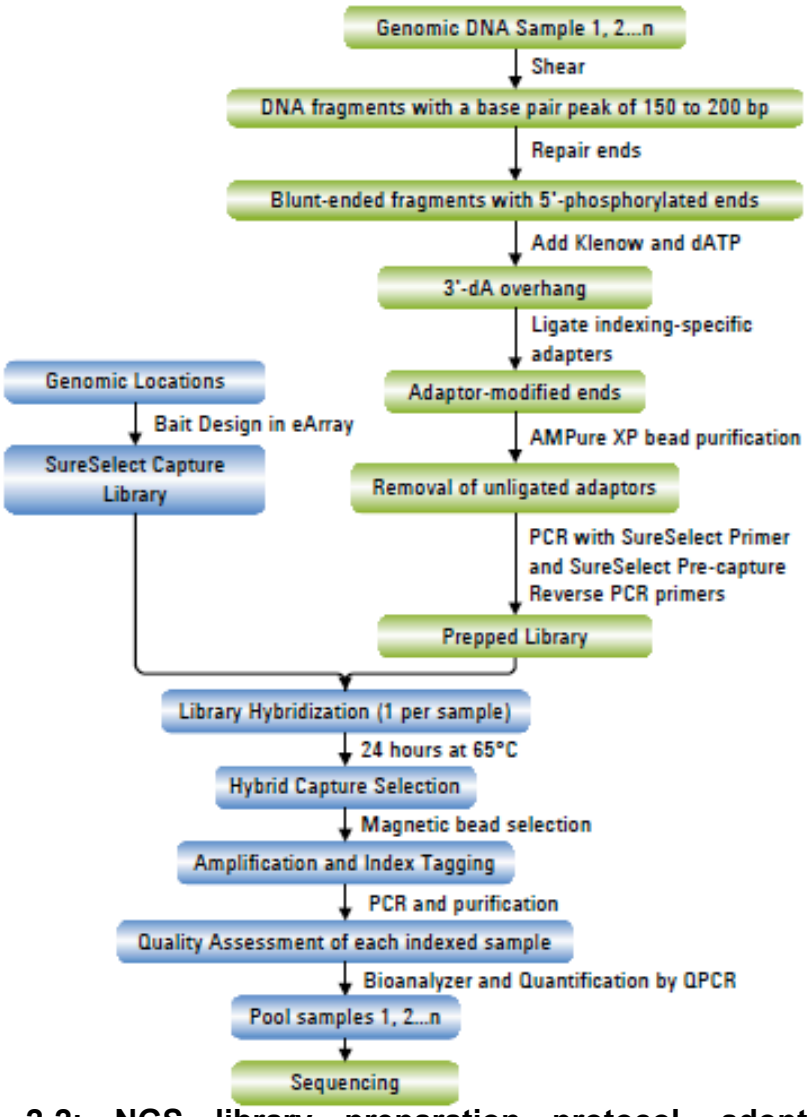


Figure 2-2: NGS library preparation protocol, adopted from SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library protocol.

2.9 RNA sequencing

RNA libraries for each cornea were generated in-house by the Leeds Sequencing Facility using the TruSeq RNA sample preparation Kit v2 (Illumina). The protocol of TruSeq RNA library preparation consists of nine major steps illustrated in figure 2-3 and briefly are: mRNA purification and fragmentation, First strand cDNA synthesis, Second strand cDNA synthesis, Repairing the ends, 3' ends adenylation, Adapters ligation, PCR amplification, Library validation and Libraries normalization and pooling.

The mRNA purification and fragmentation began by diluting 0.1 to 1µg of total RNA diluted in nuclease-free ultra pure water for a final volume of 50ul in a 96-well 0.3 ml PCR plate. 50ul of well vortexed RNA purification beads is then added to each well in order to bind the polyA RNA to the oligo-dT beads and mixed by pipetting up and down before sealing the plate. The mRNA was then denatured by incubating the plate for 5 minutes at 65°C, then at RT for 5min to allow binding of the mRNA to the beads before placing the plate ona magnetic stand at RT for 5 minutes to separate the beads from the solution. The supernatant was then discarded and the unbound mRNA removed by washing the beads by adding 200µl Bead Washing Buffer in each well. The samples were then mixed by pipetting. The plate was then placed on the magnetic stand at RT for 5minutes before discarding the supernatant containing ribosomal and other non-messenger RNA. Add 50 µl of Elution Buffer was then added in each well and samples were mixed by pipetting before sealing the plate and placed on a thermal cycler at 80°C for 2 minutes for mRNA elution. For rebinding the mRNA to the beads, 50µl of the Bead Binding Buffer were added to each well and mixed by pipetting. The mix was then incubated at RT for 5 minutes, and placed on a magnetic stand at RT for 5 minutes before discarding the supernatant. After removing the plate from the magnetic stand, the beads were washed by adding 200µl of Bead Washing Buffer and mixed by pipetting. By replacing the plate on a magnetic stand at RT for 5 minutes and discarding the supernatant, all the contaminants and the residual rRNA that did not bind to the beads were discarded. To the beads, 19.5 µl of Elute

Prime, Fragment Mix that contains random hexamers for RT priming and serves as the first strand cDNA synthesis reaction buffer was added to the samples on the plate after removing it from the magnetic stand. After sealing, the plate was incubated at 94°C for 8 minutes then subsequently briefly centrifuged.

The second stage consists of first incubating the plate at RT for 5minutes while keeping the plate on the magnetic stand while transferring 17µl of the supernatant from each well of the plate to the corresponding well of a new 0.3 ml PCR plate. Then 8µl of the First Strand Master Mix and SuperScript II mixture were added and mixed with the samples. The plate is then sealed and incubated in the thermal cycler with a pre-heated lid of 100°C for 10 minutes at b 25°C, 50 minutes at 42°C and 15 minutes at 70°C.

Straight after incubation for first cDNA strand synthesis, 25µl of Second Strand Master Mix was added to each sample and the plate was sealed and incubated at 16°C for 1 hour, then the plate was left to stand. After bringing the plate to RT, the samples were purified using 90µl AMPure XP beads to each well, mixed by pipetting and incubated at RT for 15 minutes. The plate was then placed on a magnetic stand at RT for 5 minutes and 135µl of the supernatant was removed and discarded from each well of the plate. The beads were then washed twice with 80% ethanol while maintaining the plate on the magnetic stand and the beads were left to dry for 15 minutes after discarding all the ethanol supernatant. The beads were re-suspended in 52.5µl of Resuspension Buffer and incubated at RT for 2 minutes. The plate was then place on a magnetic stand at RT for 5 minutes and 50µl of the supernatant containing the ds cDNA was transferred to a new 96-well 0.3 ml PCR plate.

Stage four of the library preparation consisted of repairing the ends of the mRNA strands into blunt ends by adding to the 50µl of ds cDNA in each well 10µl of Resuspension Buffer and 40µl of End Repair Mix and mixing by

pipetting. After the sealing, the plate was incubated at 30°C for 30 minutes. The same purification step described in the previous paragraph was repeating by adding 160µl well-mixed AMPure XP beads to each sample in the plate and washing twice by 80% ethanol. The beads were re-suspended in 17.5µl of Resuspension Buffer and incubated at RT for 2 minutes. The plate was then place on a magnetic stand at RT for 5 minutes and 15µl of the supernatant were transferred to a new 96-well 0.3 ml PCR plate. For the 3'ends adenylation, 2.5µl of Resuspension Buffer and 12.5µl of A-Tailing Mix was added to each well and mixed by pipetting. After sealing, the plate was incubated at 37°C for 30 minutes then at 70°C for 5 minutes by choosing the pre-heated lid option set to 100°C.

Then the process of ligating the adapters is performed in order to prepare the ds cDNA, for hybridization onto a flow cell by adding 2.5µl of the Resuspension Buffer, 2.5µl of Ligation Mix and 2.5µl of RNA Adapter Index to each well of the plate. Gently pipette and mixing by pipetting, then the samples were incubated at 30°C for 10 minutes after sealing the plate. To inactivate the ligation, 5µl of the Stop Ligation Buffer was added. By adding 42µl of AMPure XP beads to each well the first clean-up step previously described was initiated. After purification, the beads were re-suspended in 52.5µl of Resuspension Buffer and incubated at RT for 2 minutes. The plate was then place on a magnetic stand at RT for 5 minutes and 50µl of the supernatant were transferred to a new 96-well 0.3 ml PCR plate. Subsequently another purification was initiated by adding 50µl of AMPure XP beads to each well. After purification, the beads were re-suspended in 22.5µl of Resuspension Buffer and incubated at RT for 2 minutes. The plate was then place on a magnetic stand at RT for 5 minutes and 20µl of the supernatant were transferred to a new 96-well 0.3 ml PCR plate. For PCR amplification, 5µl of PCR Primer Cocktail to each well of the PCR and 25µl of PCR Master Mix were added to each well of the PCR plate. The PCR amplification step was done on the thermal cycler by incubating first the sealed plate at 98°C for 30 seconds and 15 cycles of 98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds, followed by 72°C for 5

minutes, all done with a pre-heated lid set to 100°C. By adding 50µl of AMPure XP beads to each well the first clean-up step previously described was initiated.

After purification, the beads were re-suspended in 32.5µl of Resuspension Buffer and incubated at RT for 2 minutes. The plate was then place on a magnetic stand at RT for 5 minutes and 30µl of the supernatant were transferred to a new 96-well 0.3 ml PCR plate. For validating the libraries, one quality control analysis was performed, by loading 1µl of resuspended construct on an Agilent Technologies 2100 Bioanalyzer using an Agilent DNA 1000 chip. The final product should be a band at approximately 260 bp. Finally, the concentration of each library was normalized to 10nanoMolar by using Tris-HCl 10mM, pH 8.5 with 0.1% Tween 20. For pooling, 10µl of each normalized sample library was transferred to a new 0.3ml PCR plate to a total volume in each well of 10X the number of combined sample libraries and 20–240µl (20µl is the volume of 2 libraries and 240µl the volume of 24 libraries). Libraries were sequenced on an Illumina HiSeq 2500 instrument.

2.10 Online resources

Oligonucleotide primers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm). The reference sequence for the genes, against which primers were designed, was downloaded from the Genome Browser (http://genome.ucsc.edu/). Literature searches for candidate genes were done using Pubmed (http://www.ncbi.nlm.nih.gov/pubmed). The NCBI gene database (http://www.ncbi.nlm.nih.gov/gene) was also used to investigate information about the gene, the transcript and the corresponding protein. Pubmatrix is a tool used to apply a pairwise comparison of specified keywords belonging to two different lists, in order to mine the literature in Pubmed and extract all the frequency of co-occurrence of both lists (Becker *et al.*, 2003) (http://pubmatrix.grc.nia.nih.gov/).

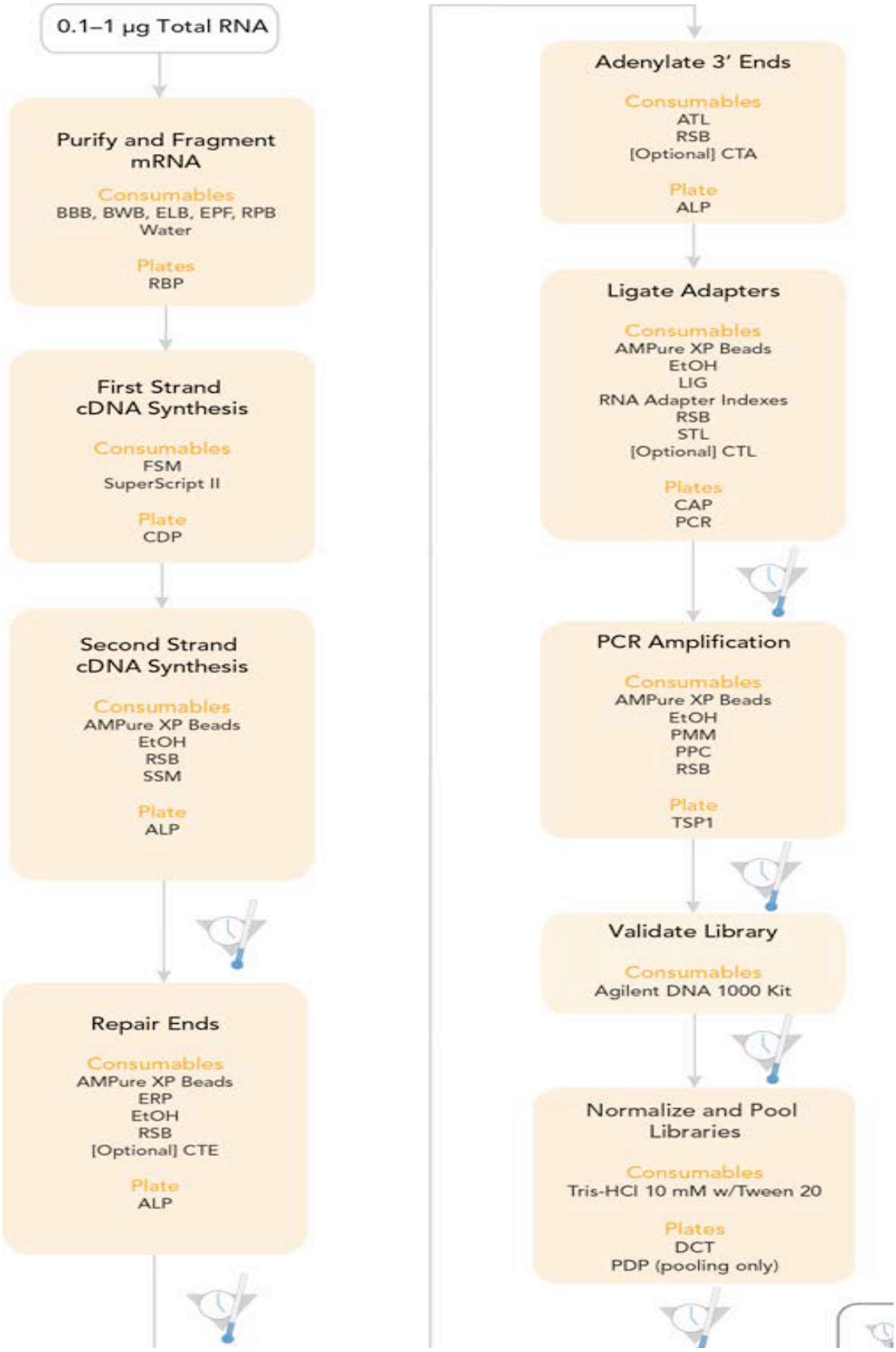


Figure 2-3: RNA seq library preparation protocol. Adopted from TruSeq® RNA Sample Preparation v2 Guide.

2.11 Quality Control of NGS data

The raw Illumina sequencing data reads or fastq files were checked for quality using a java based visual interface software called FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FastQC checks for quality control general criteria, most importantly it checks for quality scores per sequence, GC content, adapter dimer contamination and duplication levels. Qualimap (Garcia-Alcalde *et al.*, 2012) was used to examine the sequencing alignment data in BAM format files by using the default settings. Qualimap (http://qualimap.bioinfo.cipf.es/) is a quality control application written in Java and R languages; it provides both a graphical user interface as well as a command line interface.

2.12 Whole Exome Sequencing

WES analysis was performed using the commands listed in Appendix 1. WES raw data was aligned against the human genome reference sequence (version hg19) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) (http://bio-bwa.sourceforge.net/) in order to create SAM files. The major indexing algorithms mainly used in alignment programs are hash table and suffix tree. Unlike BLAST and Novalign that use the Hash table algorithm, SOAP2, Bowtie and BWA adopt the suffix tree or the FM-indexing algorithm (Yu *et al.*, 2012). The FM-index method (Full-text index in Minute space), is a compressed full-text substring index based on the Burrows-Wheeler Transform (BWT) to index the reference genome. This method searches for inexact matches by using a backtracking strategy (Li and Durbin, 2009). Furthermore, BWA generates a mapping quality score for each read, with higher scores corresponding to more accurate alignment, and indicating Phred-scaled probabilities for incorrect mapping (Yu *et al.*, 2012). In addition to being considered an accurate aligner for short and long sequencing reads, BWA is equally regarded to be time and memory efficient (Li and Durbin, 2009).

Conversion from SAM to BAM format files as well as sorting, indexing and viewing were accomplished using Samtools (Li *et al.*, 2009) (http://samtools.sourceforge.net/). Picard tools (http://broadinstitute.github.io/picard/) was used to remove duplicates. Local realignment, calling variants, depth of coverage and recalibration were achieved using the Genome Analysis ToolKit (GATK) software package (McKenna *et al.*, 2010) (https://www.broadinstitute.org/gatk/). The Integrative Genomics Viewer (IGV) (Robinson *et al.*, 2011) (https://www.broadinstitute.org/igv/) was used as an alignment visualisation tool when needed.

Variants were filtered out from VCF files of all the samples if they were found to be abundant - with a specific Minor Allele Frequency (MAF) greater than 2% - in a set of controls of the same ethnicity as the patient. Then, the VCF files of patients of Pakistani and Indian origin were compared against the Born in Bradford (BiB) (www.borninbradford.nhs.uk/) database using an in-house perl script developed by David Parry (University of Leeds) (http://sourceforge.net/projects/vcfhacks/), all the alleles with MAF > 5% in BiB were filtered out from these particular samples of Indian and Pakistani origin. BiB is a database consisting of a collection of more than 3000 UK exomes to which access was granted following a collaboration on the project with Ian Carr (University of Leeds). On the other hand, from the WES of the 8 Druze controls and 4 Maronites controls collected, a list for each ethnicity was created containing all the alleles carried by 50% of the controls of each set. Subsequently, all the VCF files of Lebanese Druze and Lebanese Maronite origin patients were compared against each corresponding list using a “highlight-similarities” perl script developed specifically for this study (Appendix 2) and all the shared alleles between the 50% alleles list and the patient VCF file were removed. In addition, this same perl script was used throughout this project to highlight shared genes/variants between excel lists.

The remaining genetic variants detected in the VCF file output of GATK were functionally annotated using the ANNOVAR software (Wang *et al.*, 2010, Yang and Wang, 2015) (http://www.openbioinformatics.org/annovar/). ANNOVAR software annotates using the information of several sources described below in order to perform the following tasks:

Gene-based annotation is to identify whether the SNPs or CNVs are functional by determining from several downloaded databases whether the SNP cause protein coding changes and determine which amino acids are affected. The databases downloaded for this analysis are the following:

- refGene hg19 of FASTA sequences for all annotated transcripts in RefSeq Gene.
- knownGene hg19 of FASTA sequences for all annotated transcripts in UCSC known gene
- ensGene hg19 of FASTA sequences for all annotated transcripts in ENSEMBL gene

All the three databases are used to annotate all the variants in the VCF files and to report the regions of occurrence of these variants whether it's in the intronic, exonic or non-coding RNA regions and associating gene names to the appropriate regions.

Filter-based annotation is to identify alleles reported previously in public databases with their corresponding allele frequencies (Kenna *et al.*, 2013), and to calculate based on downloaded algorithms the pathogenicity and conservation scores of every SNP. The hg19 databases and algorithms downloaded for this purpose are the following:

- Whole-exome SIFT: "Sorting tolerant from Intolerant" or SIFT algorithm predicts whether the amino acid substitution affects the protein function by looking at the important conservation positions of the protein sequence (Kumar *et al.*, 2009). SIFT scores range from 0 to 1 with scores < 0.05 are considered damaging.

- Polyphen2 with HDIV and HVAR functions. Polyphen2 HVAR distinguishes mutations with severe effects and used for diagnostics of Mendelian diseases. Polyphen2 HDIV used when evaluating rare alleles at loci possibly involved in complex phenotypes. The scores of “probably damaging” mutations range from 0.9 to 1, “possibly damaging” from 0.4 to 0.9 and “benign” from 0 to 0.4 (Adzhubei *et al.*, 2010, Adzhubei *et al.*, 2013).
- LRT or Likelihood Ratio Test, similar to SIFT and Polyphen, identifies in a protein coding sequence the deleterious mutations affecting highly conserved amino acids and represents three possible predictions: D for deleterious, N for neutral or U for unknown (Chun and Fay, 2009).
- Mutation Taster similar to SIFT, Polyphen and LRT. It presents four different possibilities in predicting the values: A for disease causing automatic, D for disease causing N for neutral and P for polymorphism automatic (Schwarz *et al.*, 2010).
- MutationAssessor it classifies the variants into two groups: predicted functional (H for high and M for medium effects) and predicted non-functional (L for low and N for neutral effects) (Reva *et al.*, 2011).
- FATHMM or Functional Analysis through Hidden Markov Models similar to SIFT and Polyphen. The scores generated by FATHMM less than -1.5 are considered as D for damaging otherwise the score is T for tolerated (Shihab *et al.*, 2014).
- Developed by Coco Dong MetaSVM and MetaLR fill out missing scores from all the previously mentioned annotation algorithms by imputing for whole-exome variants. Their scores range from 0 to 1, with higher scores being more deleterious (http://annovar.openbioinformatics.org/en/latest/user-guide/filter/#-metasvm-annotation).
- GERP++ predicts conservation scores for coding variants, with the higher scores being the most conserved sites (Davydov *et al.*, 2010).

- PhyloP (Cooper *et al.*, 2005) and SiPhy (Garber *et al.*, 2009) similar to GERP++ and based on multiple alignments of 46 genomes and 29 mammals genomes respectively. Higher scores correspond to more conserved.
- SIFT, Polyphen2 HDIV, Polyphen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaSVM, MetaLR, GERP++, PhyloP and SiPhy scores are downloaded through the non-synonymous single-nucleotide variants (nsSNVs) functional prediction and annotation dbNFSP database version 2.6 (Liu *et al.*, 2011, Liu *et al.*, 2013, Liu *et al.*, 2015a).
- esp6500siv2 is a database from the EVS (Exome Variant Server) that contains exome sequencing variants from 6500 exomes for the NHLBI-ESP (NHLBI-Exome Sequencing Project). It includes the indel calls and the chrY calls and is used to identify the variants' allele frequency in the VCF files (http://evs.gs.washington.edu/).
- Exac03 or the Exome Aggregation Consortium version 0.3 contains exome sequencing data of 65000 exome for: AFR (African), AMR (Admixed American), EAS (East Asian), FIN (Finnish), NFE (Non-finnish European), OTH (other), SAS (South Asian)) (http://exac.broadinstitute.org/).
- 1000g2014oct: represents alternative allele frequency data in 1000 Genomes for several populations: ALL, AFR (African), AMR (Admixed American), EAS (East Asian), EUR (European), SAS (South Asian) (http://www.1000genomes.org/).
- dbSNP 138 is a database used to identify the previously reported variants and to report their specific rs number (Sherry *et al.*, 2001) (http://www.ncbi.nlm.nih.gov/SNP/).
- CLINVAR or Clinical Variation database represents information about variants clinical significance with their corresponding variant disease name. The variant clinical significance annotation includes: unknown, untested, non-pathogenic, probable-non-pathogenic, probable-

pathogenic, pathogenic, drug-response, histocompatibility, other (Landrum *et al.*, 2014) (http://www.ncbi.nlm.nih.gov/clinvar/).

- CADD or Combined Annotation Dependent Depletion tool generates a score to each possible mutation in the human genome (except Indels and CNVs) by integrating several annotations into a Cscore (Kircher *et al.*, 2014) (http://cadd.gs.washington.edu/).

The ANNOVAR output file annotated using all these databases and algorithms, is a text file that is opened and modified using Microsoft excel. Through the sort and filter function of Microsoft excel, variants with a minor allele frequency (MAF) of 2% or greater in dbSNP138, were filtered out.

2.13 PLINK/SEQ

One KC aligned BAM file from each family was used in order to carry out an association analysis of variants present in KC familial samples compared to controls, using the open source software PLINK/SEQ (https://atgu.mgh.harvard.edu/plinkseq/). PLINK/SEQ is an open source library for working with human genetic variation data, using the R interface for statistical computing. Exome sequences from 24 unrelated KC cases were compared to 72 controls, giving a 1:3 ratio. BAM files of all the 96 files were pooled together using the UnifiedGenotyper and LiftOverVariants, functions of GATK in order to generate a combined VCF file. The VCF hacks package was then used to translate, annotate and filter the pooled VCF file. The subsequent VCF file was inputted into PLINK/SEQ using a series of commands (Appendix 4). Variants were analysed both pooled by gene and as individual rare variants.

2.14 Transcriptomics

Before alignment, adapter contamination was trimmed from RNA-seq fastq file reads using the python package Cutadapt (https://code.google.com/p/cutadapt/) (Martin, 2011). RNA-seq reads were

aligned to human genome version GRCh38p2 using a bowtie2-based aligner and splice junction mapper, TopHat2 (Kim *et al.*, 2013) (http://ccb.jhu.edu/software/tophat/index.shtml). The reference GTF file used was the gencode GTF –release 22 (GRCH38p2) which contains the comprehensive gene annotation on the reference chromosomes, scaffolds, assembly patches and alternate loci. For novel gene/exon discovery, the RNA-seq Unified Mapper (RUM) (Grant *et al.*, 2011) (http://cbil.upenn.edu/RUM/) was used (List of commands – Appendix 3).

Differential expression analysis was conducted using the Cufflinks software (http://cole-trapnell-lab.github.io/cufflinks/), part of the Tuxedo set (Trapnell *et al.*, 2012), and visualized using the bioinformatics tool CummeRBund (http://compbio.mit.edu/cummeRbund/) (List of commands – Appendix 3). The lists of differentially expressed genes were subjected to pathway analysis using the DAVID bioinformatics suite (http://david.abcc.ncifcrf.gov/) for functional enrichment and to discern the underlying biological processes. KEGG pathway analysis (Ogata *et al.*, 1998) (http://www.genome.jp/kegg/), REACTOME (http://www.reactome.org/) (Joshi-Tope *et al.*, 2003) and WikiPathways (http://www.wikipathways.org/index.php/WikiPathways) (Pico *et al.*, 2008), were similarly used as prediction tools to assign pathways and functions to genes highlighted by RNA-seq analysis.

For differential expression results, EdgeR - a count-based tool - was used for further verification. EdgeR is a freely available bioconductor software package that uses replicated count data to examine differential expression between a set of RNA-seq data (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson *et al.*, 2010). Jointly, Subreads package (Liao *et al.*, 2013) (http://bioinf.wehi.edu.au/subread/) with its specific FeatureCounts program (Liao *et al.*, 2014) was used to count the already mapped RNA-seq reads (http://bioinf.wehi.edu.au/featureCounts/) and create a count matrix before inputting it into EdgeR for differential expression analysis (List of commands – Appendix 3-4). Cummerbund, EdgeR and Subreads were run on R

platform (version R 3.1.1 GUI 1.65 Snow Leopard build) for statistical computing and graphics (https://www.r-project.org/) (Ihaka and Gentleman, 1996).

3 Screening for rare Mendelian variants of major effect in KC families

3.1 Introduction

The pathogenesis of the majority of KC is unknown but genetics is believed to play a significant role (Section 1.4.9). The evidence for a genetic component to KC comes from the increased risk of first-degree relatives and twin studies (Valluri *et al.*, 1999, Tuft *et al.*, 2012) (Wang *et al.*, 2000, Rabinowitz, 2003, Burdon and Vincent, 2013) (section 1.4.9). In addition, the variable incidence of KC in different populations and ethnicities could also be a result of different genetic backgrounds (Section 1.4.2). However, the most striking evidence for a genetic cause of KC is the high rates of this disease in countries and communities that practice consanguinity or endogamy (Pearson *et al.*, 2000). The incidence of KC in the general population is 50 per 100,000 (reviewed by)(Rabinowitz, 1998) but this rises to 2340/100,000 in Jerusalem (Millodot *et al.*, 2011), to 2300/100,000 in India (Jonas *et al.*, 2009), to 830/100,000 in Shahroud-Iran (Hashemi *et al.*, 2014) and to 3333/100,000 in Lebanon (Waked *et al.*, 2012). This data raises the possibility that a recessive Mendelian form of KC exists and investigating this hypothesis is the focus of this chapter.

3.2 Results

3.2.1 Families collection

The centre of this study was a collection of fifteen KC families from a variety of ethnic backgrounds including Indian, Pakistani and Lebanese (Druze and Christian Maronite). The five families of Pakistani origin are all living in the UK (Figures 3-1 and 3-2).

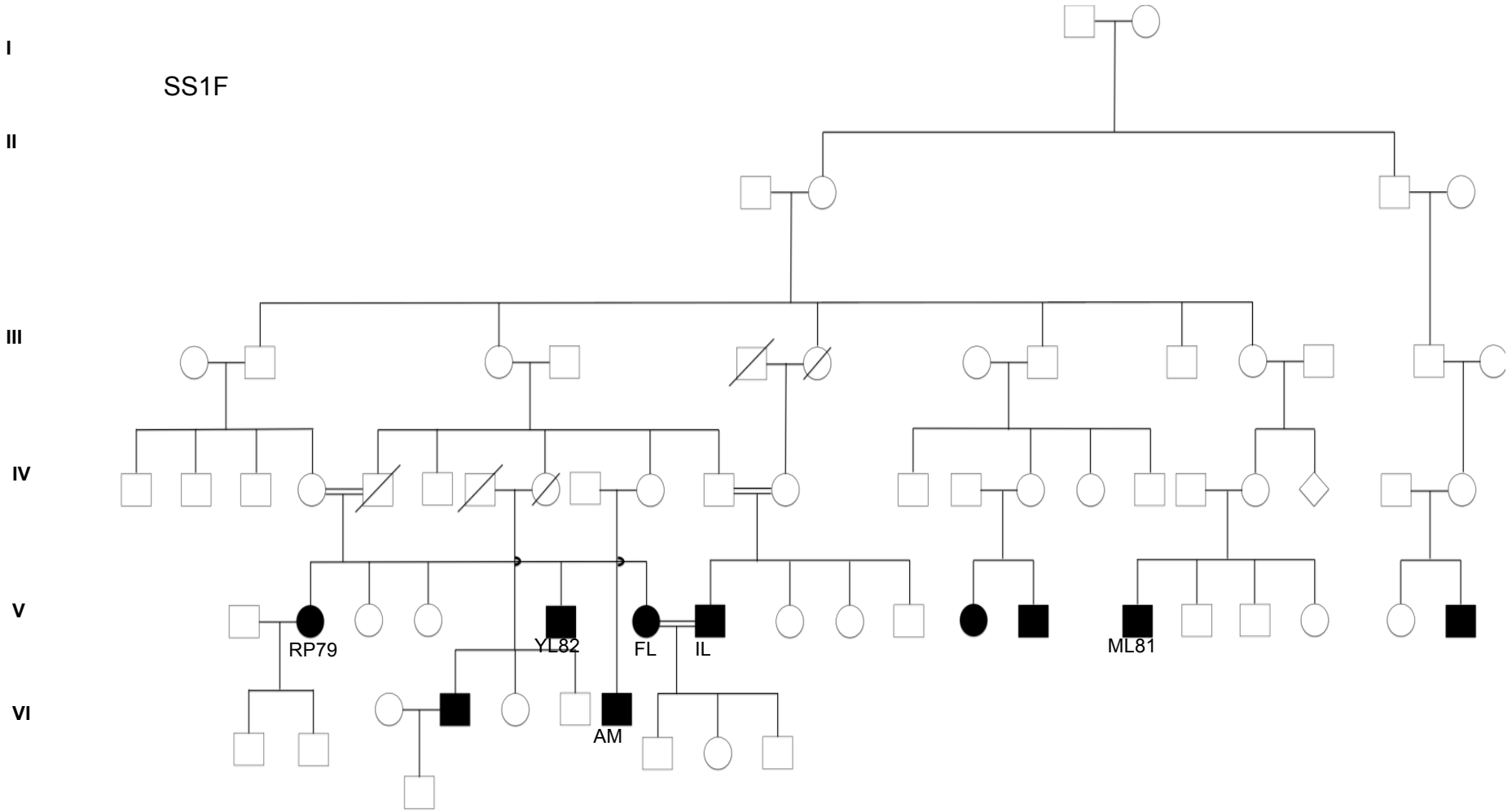


Figure 3-1: Pakistani family pedigrees (1). SS1F is a two generation KC family. The patients that were sampled are: V:2 or RP79; V:9 or YL82; V:10 or AM; V:11 or FL; V:12 or IL and V:18 or ML81.

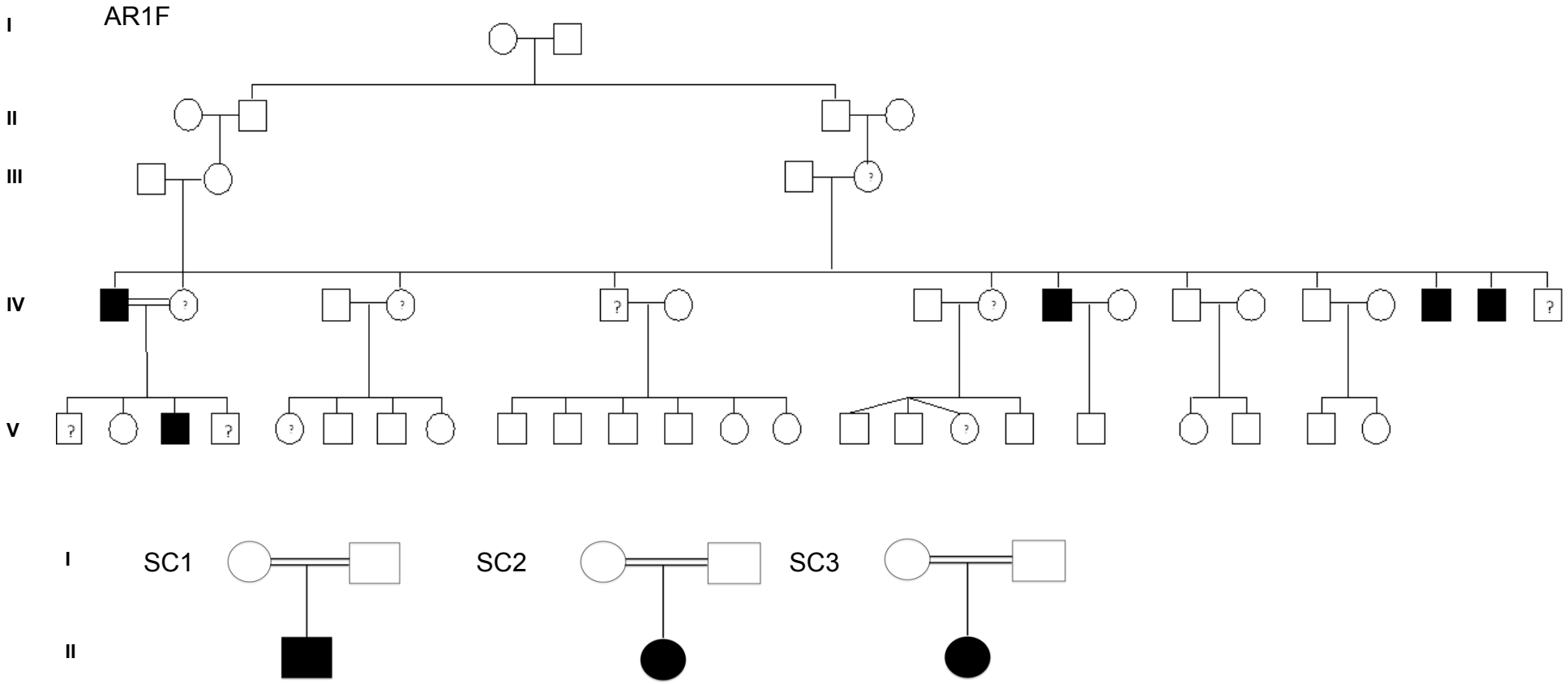


Figure 3-2: Pakistani family pedigrees (2). Pedigrees of four Pakistani families included in this study. AR1F is a two generation KC family and SC1, SC2 and SC3 present single KC cases originating from consanguineous marriages. The question mark in the AR1F pedigree represents the patients with borderline Keratoconus, i.e. they do not show KC symptoms but they harbour some corneal irregularities.

Family SS1F (Figure 3-1) was recruited by Dr Salina Siddiqui (The Eye Department, St. James's University Hospital, Leeds). The pedigree's structure shows 10 affected individuals, four of whom were born from consanguineous marriages (RP79, YL82, FL and IL). In addition, the the children of IL and FL (VI:4, VI:5 and VI:6) patients were below the age of 10 at the time of collection and didn't present any signs of visual impairments. The autosomal recessive mode of inheritance seems to best describe the transmission of the disease in the family.

Family AR1F (Figure 3-2) was recruited by Dr Aine Rice (formerly of the Eye Department, St. James's University Hospital, Leeds). This family has five affected members (IV-1, IV9, IV15, IV16 and V3) but only one is knowingly born from a consanguineous marriage (V:3) and this patient had a very severe case of KC and was grafted as a child. The mode of inheritance in this family is not clear and could be consistent with a recessive, dominant or even an x-linked mode of inheritance given that only males are affected and some of the females have forme-fruste.

The three single case families (SC1, SC2 and SC3) (Figure 3-2) are all from different consanguineous marriages and were recruited by Dr Aine Rice (formerly of the Eye Department, St. James's University Hospital, Leeds).

One Indian family was included in this study, Family KF (Figure 3-3). This family was obtained from Professor Govindasamy Kumaramanickavel (Narayana Nethralaya, Bangalore, India). This family has three affected siblings IV:1, IV:2 and IV:3 originating from consanguineous unaffected parents, suggesting a recessive mode of inheritance.

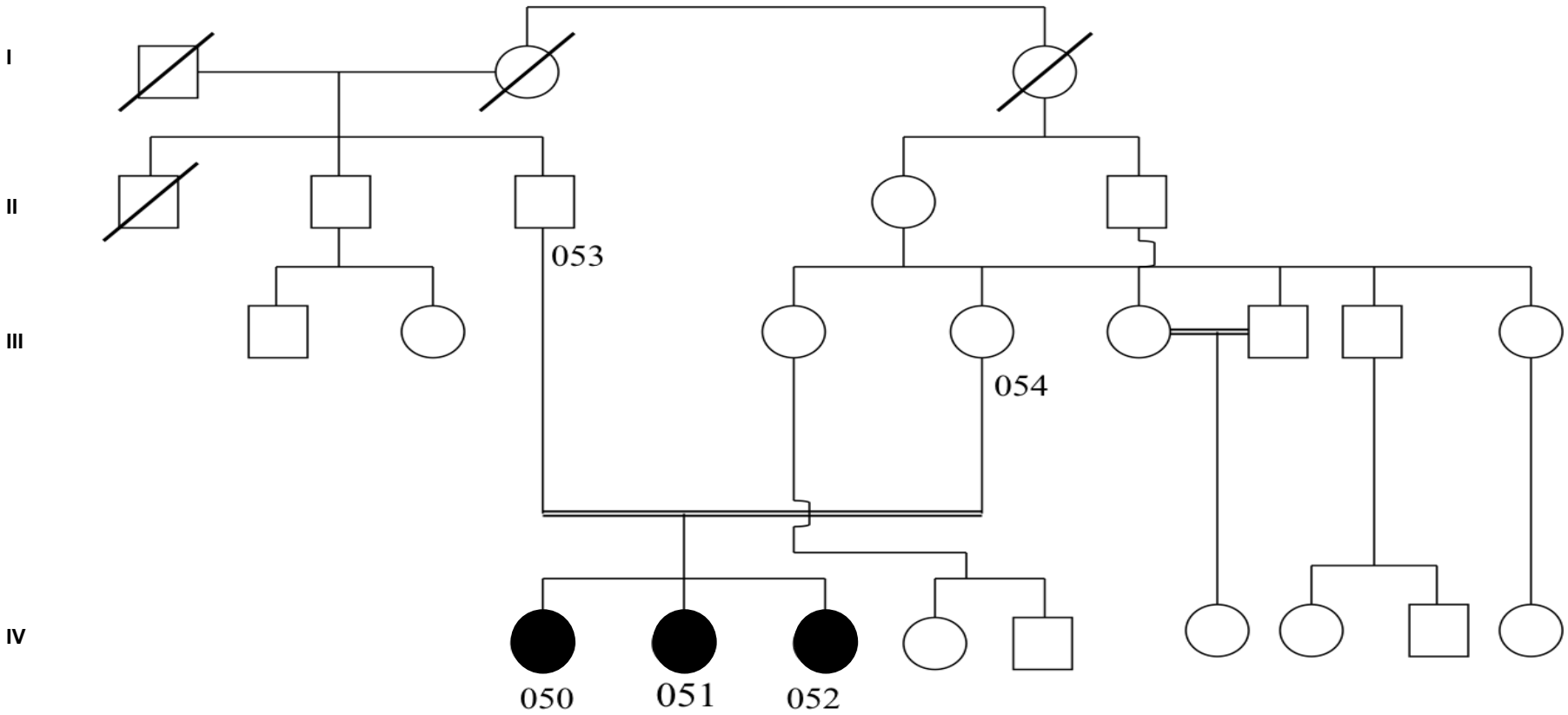
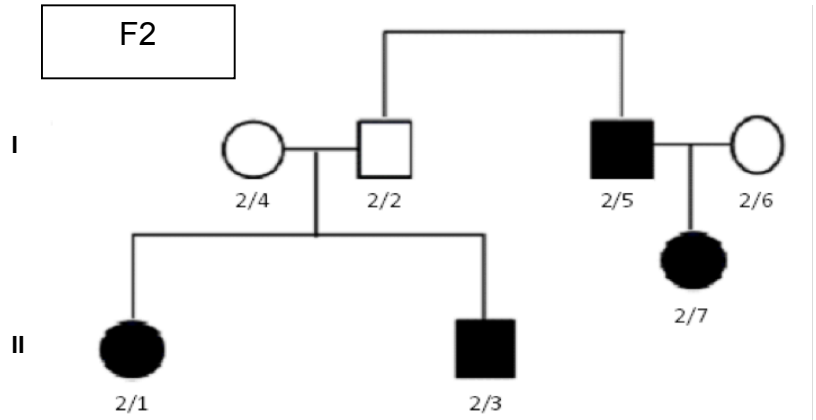


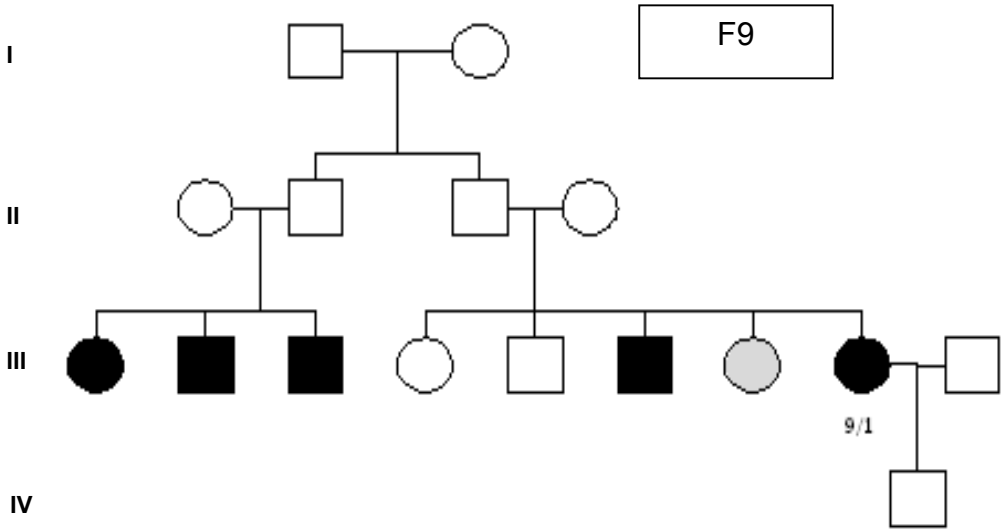
Figure 3-3: KF family pedigree. The pedigree shows three patients in generation IV (IV-1, IV-2 and IV-3) originating from a consanguineous marriage (II-3 and III-4).

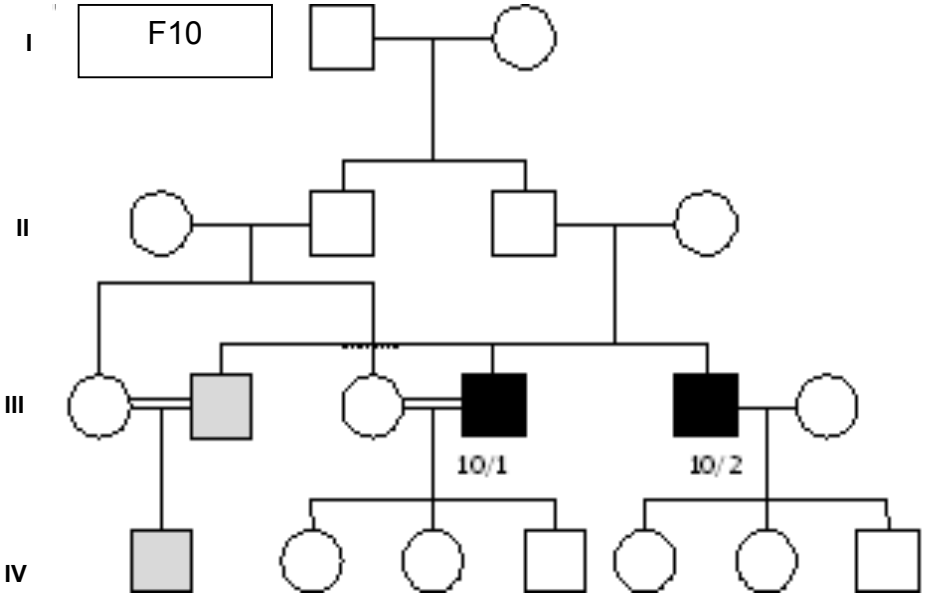
The Lebanese families were recruited by the author, in collaboration with Lebanese ophthalmologists, during a field trip to Lebanon (Figure 3-4). The Druze families were diagnosed by Dr. Bassam Barakat (Consultant Ophthalmologist, Mount Lebanon Hospital, Hazmieh, Lebanon) unless stated otherwise. Six Druze families were recruited: F2, F9, F10, F11, F12 and F13 (Figure 3-4). The Druze families were of particular interest since the Druze population is famous for endogamy. This community follows social customs that turn it into a transnational isolate, favouring marriage from within the same religion, forbidding admixture with other populations and firmly closing the religion to new adherents (Shlush *et al.*, 2008). In fact, the majority of the Druze families collected in this study were not consanguineous; however, most of the parents in the families (or grandparents) share the same surnames, indicating that they could share more distant ancestors in common.

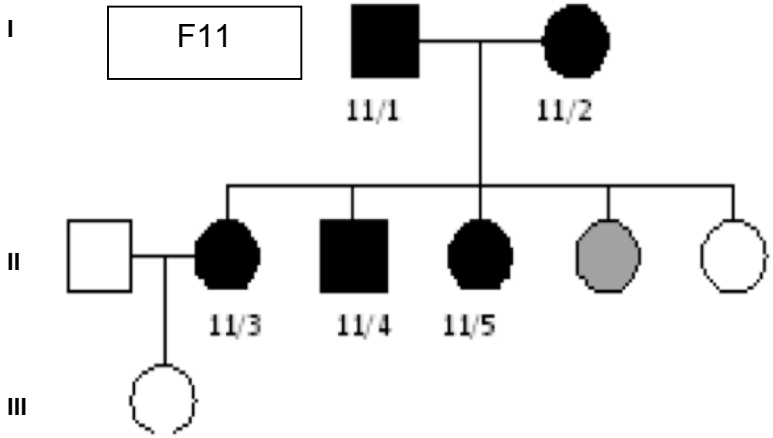
Family F2's diagnosis was ascertained by Dr. Ghassan Chayya (formerly at Department of Ophthalmology, Al Watani Hospital, Aley, Lebanon). F2 is a two-generation family with four affected members and an ill-defined mode of transmission. The family is not known to be consanguineous but it belongs to an endogamous community. The parents (I:1 and I:2) are from the same village and share the same surname.

Family F9 is a four-generation pedigree with five affected members (III:1, III:2, III:3, III:6, III:8) and one borderline KC case (III:7). Three of the younger unaffected individuals in this family (III:3, III:4, and IV:1) are below the age of onset for disease so they may potentially develop KC in the future. Parents of generation I and generation II belong to the same village. The pedigree of family F9 suggests an autosomal recessive mode of inheritance.


F2
I
II
2/4
2/2
2/5
2/6
2/7
2/1
2/3


F9
I
II
III
IV
9/1


F10
I
II
III
IV
10/1
10/2


F11
I
II
III
11/1
11/2
11/3
11/4
11/5

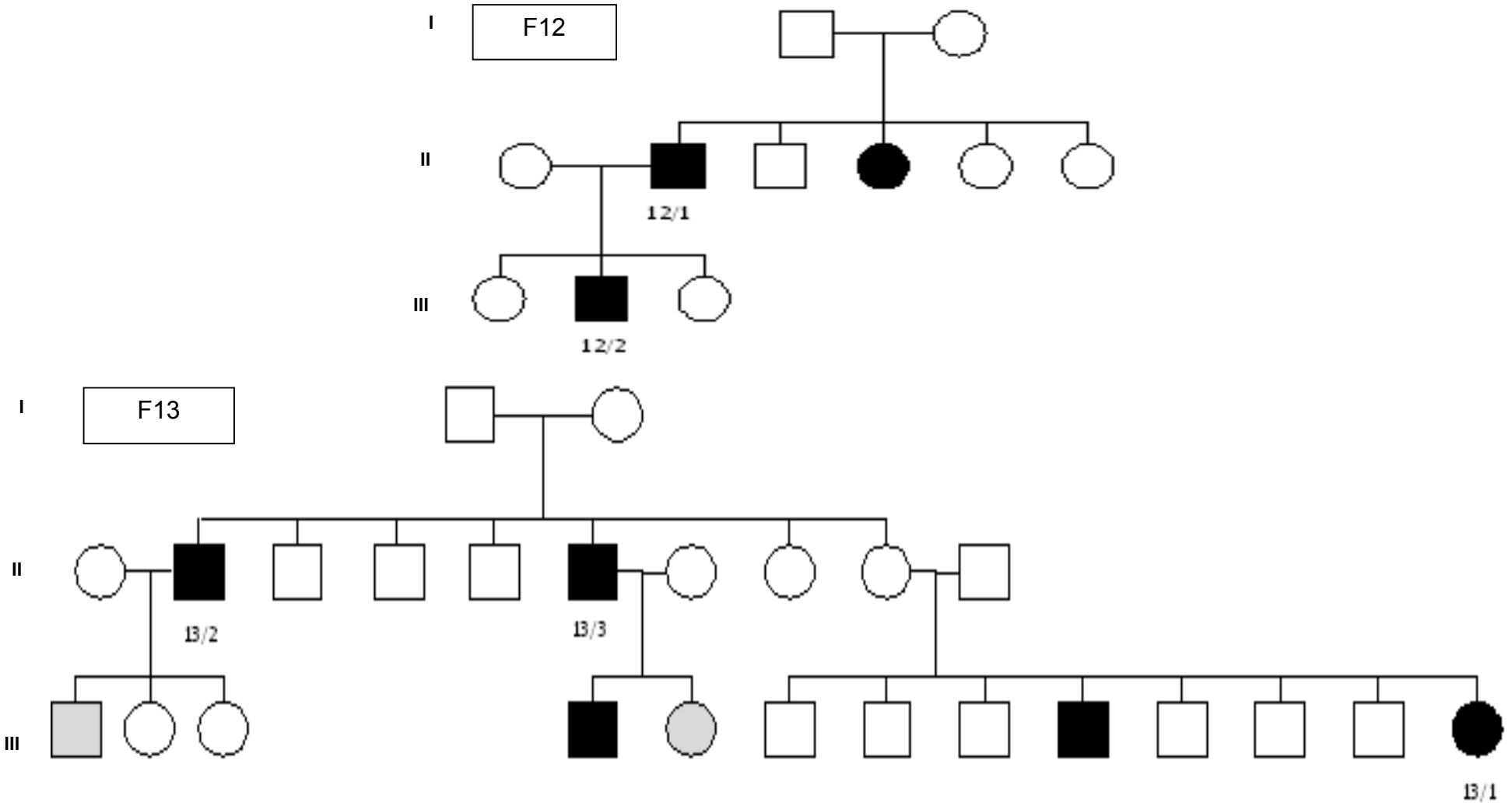


Figure 3-4: Pedigrees of Druze Lebanese KC families. Pedigrees of Lebanese Druze families F2, F9, F10, F11, F12 and F13. The light grey symbol in the pedigrees designates individuals with borderline KC.

F10 (Figure 3-4) is a four-generation pedigree, with two affected members (III:4 and III:5). Members III:2 and IV:1 are reported to have KC but access to these individuals' medical records was not possible. All members of generation IV (except IV:1) were preadolescent at the time of recruitment so their unaffected KC status is not confirmed. The mode of inheritance is not well defined.

The pedigree of family F11 (Figure 3-4) shows five affected individuals in two-generations (I:1, I:2, II:2, II:3 and II:4). Individual II:5 is a borderline KC case and II:6 is too young to diagnose. Parents I:1 and I:2 are distantly related and share one great grandparent. Thus it is possible that the mode of inheritance is autosomal recessive with the appearance of dominant inheritance.

Family F12 (Figure 3-4) is a three-generation pedigree with three affected patients (III:2, II:2 and II:4). The mode of inheritance of family 12 is not well-defined.

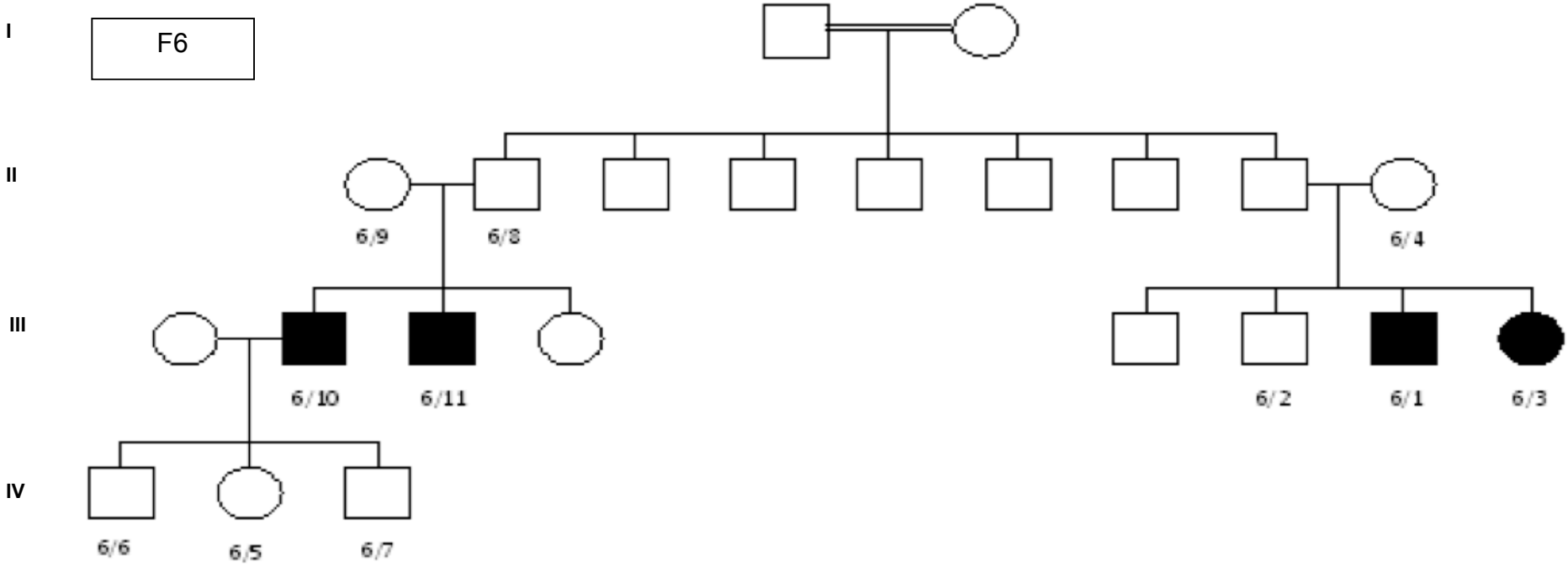
Family F13's pedigree (Figure 3-4) is a three-generation pedigree with 5 affected members (II:2, III:4, II:6, III:9 and III:13) and two borderline KC cases (III:1 and III:5). This pedigree shows a dominant mode of inheritance with reduced penetrance but as this family belongs to an endogamous ethnicity a recessive mode of inheritance is also possible.

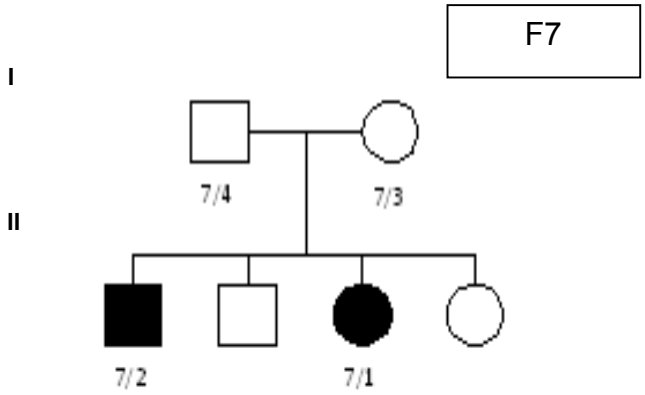
Three Lebanese families of Christian Maronite origins were also collected (Figure 3-5). Family F6's diagnosis was ascertained by Dr Fadi Geagea (Formerly at the Ophthalmology department of Saint Joseph University and the American University of Science and Technology). Family F6's pedigree (Figure 3-5) shows four affected patients (III:2, III:3, III:7 and III:8) and the mode of inheritance is consistent with autosomal recessive.

Family F7's diagnosis was ascertained by Dr. Naji Waked (Department of Ophthalmology, Saint Joseph University, and Hotel-Dieu de France Hospital, Beirut, Lebanon). Family F7 (Figure 3-5) has two affected patients (II:1 and II:3) with unaffected parents indicating an autosomal recessive mode of inheritance.

Family F8's diagnosis was ascertained by Dr. Naji Waked (Department of Ophthalmology, Saint Joseph University, and Hotel-Dieu de France Hospital, Beirut, Lebanon). Family F8 (F 3-5) is a four generation pedigree with three affected individuals (III:2, III:4 and II:6). Unfortunately, a DNA sample for II:6 was not available. The mode of inheritance of this pedigree is consistent with autosomal recessive.

For all the recruited Lebanese families, informed signed consent was obtained and blood was collected. The DNA extraction was performed by the author in the laboratory of Dr. Gretta Abou Sleymane (American University of Science and Technology, Beirut, Lebanon) (Section 2.2.1).


F6
I
II
III
IV
6/9
6/8
6/4
6/10
6/11
6/2
6/1
6/3
6/6
6/5
6/7


F7
I
II
7/4
7/3
7/2
7/1

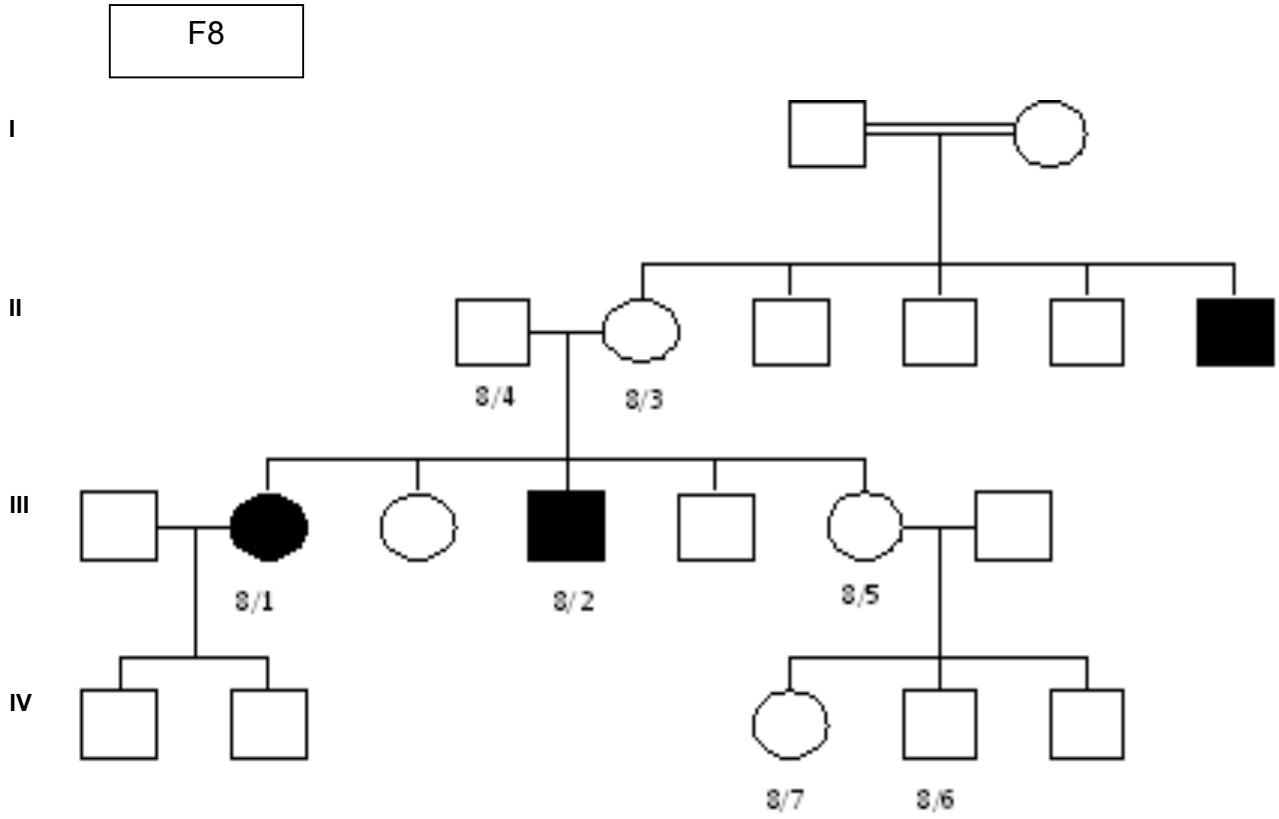


Figure 3-5: Pedigrees of Christian Maronite Lebanese KC families. Pedigrees of the Christian Maronite Lebanese families F6, F7 and F8.

3.2.2 The search for rare recessive alleles causing KC

The families ascertained for this study were all consistent with an autosomal recessive mode of inheritance given their pedigree structure and/or the endogamous population from which they originated. These families were therefore used to try and identify recessive alleles responsible for KC using a combination of autozygosity mapping and WES (section 1.5). Autozygosity mapping was performed using SNP microarray genotyping and the resulting data was analyzed using IBDfinder, MultiIdeogram and AutoIdeogram (Section 2.7). WES was outsourced to Otogenetics Corporation (Atlanta, Georgia, United States) for a few samples but the majority of the samples were processed by the author and run on the Illumina HiSeq 2000/2500 (Section 2.8). WES fastq files were checked for quality using FastQC software (Section 2.11). An example of the quality report is shown in figure 3-6.

Samples were deemed to have passed quality control (QC) if the mean of the phred scores across all bases and sequences is above 30, and all the samples analysed in this study were above this threshold. The fastq files for each sample were then aligned to the reference human genome (Section 2.12). The aligned sequence file is called a bam file and this data was subjected to a further stage of QC using the qualimap application to assess the alignment quality of the bam file (Section 2-11). Figure 3-7 represents one example of a quality control report of the bam file of sample 2-3. The main criteria on which the quality report is based on are: the percentage of mapped reads, the general error rate and the mean mapping quality. There is no specific threshold for these criteria, instead this analysis creates a mapping profile for each sample.

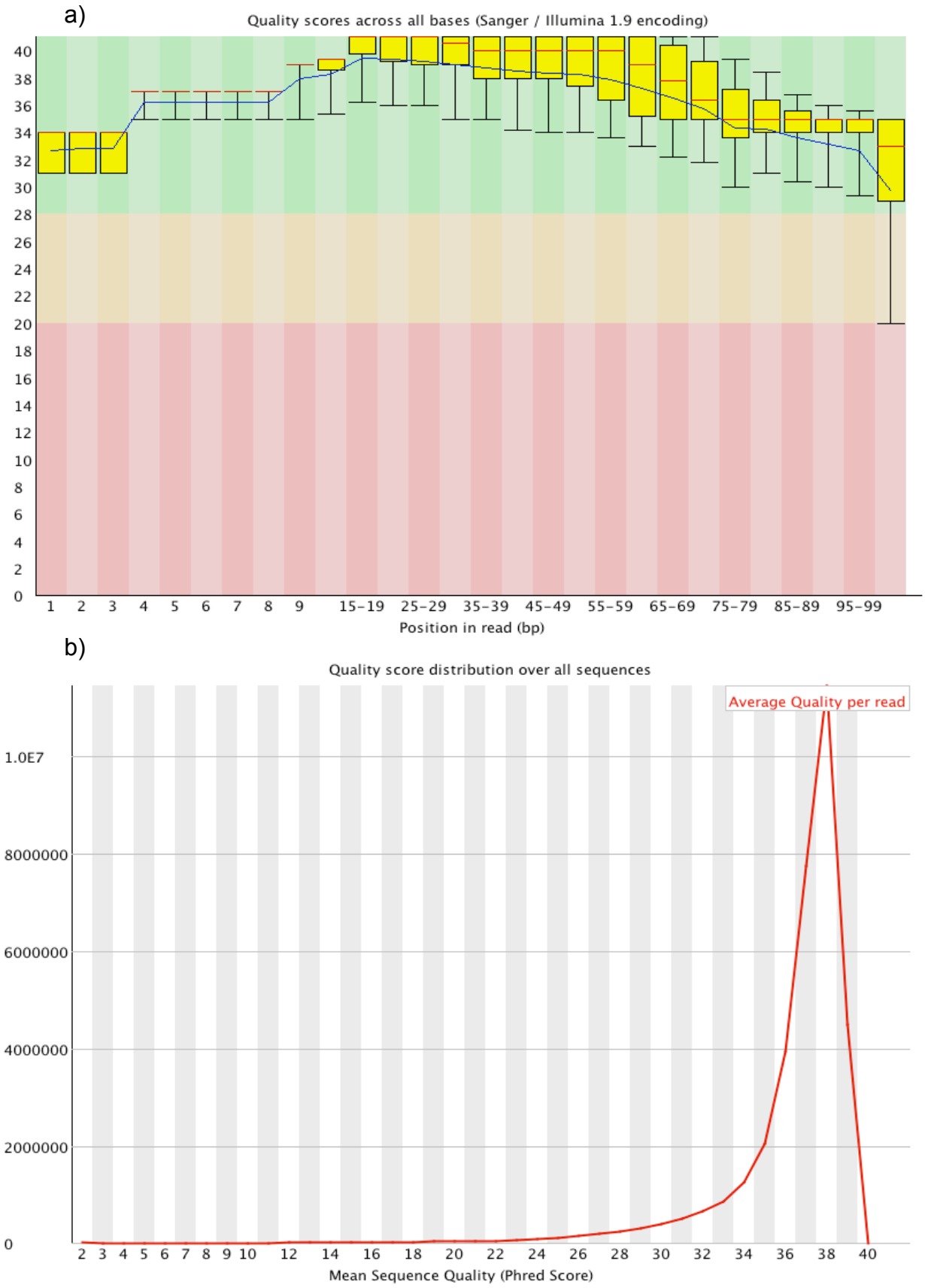


Figure 3-6: Example of the quality control report of the fastq file of sample 2-3. a) A box plot of the quality scores (Phred) obtained across the 100bp of sequence generated during the run. The red line shows the median and the yellow boxes the interquartile range and the whiskers show the range. The (y) axis represents the Phred scores. A Phred score of 30 indicates 99.9% accuracy. b) Represents quality score distribution over all sequences. The (y) axis represents the number of sequence reads.

a)

	Sample 2-3
Number of reads	68,989,763
Mapped reads	67,986,487 (98.55%)
Unmapped reads	1,003,276 (1.45%)
Paired reads	67,986,487 (98.55%)
Mapped reads, only first in pair	34,052,587 (49.36%)
Mapped reads, only second in pair	33,933,900 (49.19%)
Mapped reads, both in pair	67,707,895 (98.14%)
Mapped reads, singletons	278,592 (0.4%)
Read min/max/mean length	100 / 100 / 100
Clipped reads	736,561 / 1.07%
Duplicated mapped reads	15,486,367 / 22.45%
Duplication rate	20.33%
GC Percentage	47.3%
Mean Mapping Quality	48.2
General error rate	0.24%

b)

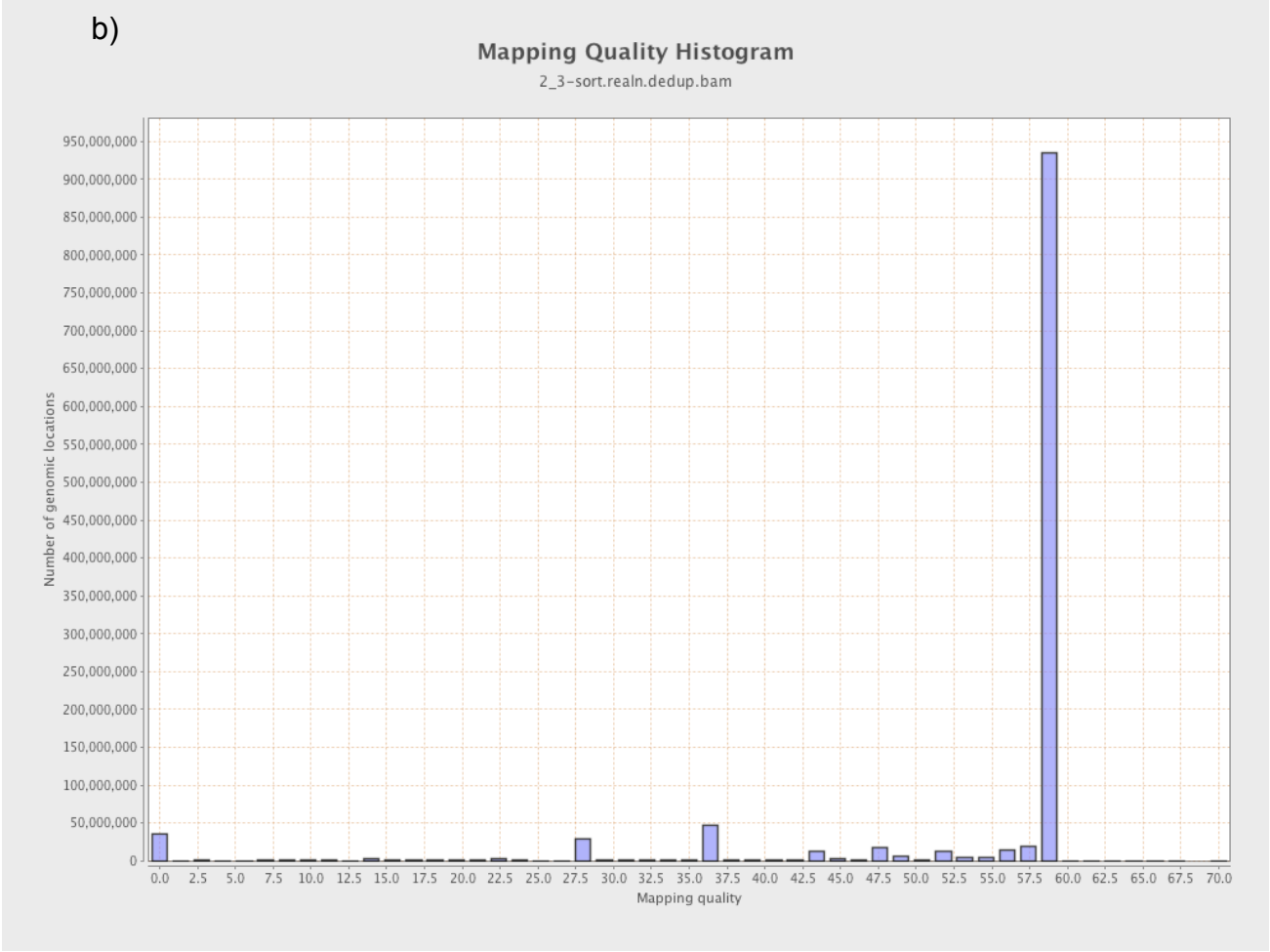


Figure 3-7: Qualimap report of the quality of sample 2-3 aligned data. a) Table presenting the basic quality information and statistics of the aligned sample. b) Plot showing the mapping quality distribution across the reference.

A critical QC step for an aligned sample (Bam file) is the information on the sequencing depth of coverage. This information was acquired using GATK to align the sample file against the genomic regions captured during the preparation of the WES library (Section 2.8). Table 3-1 represents the depth of coverage statistics of sample 2-3 and data for other samples can be found in appendix 5.

Table 3-1: Coverage statistics of sample 2-3 for WES targets generated by GATK. This table represents the percentage of bases targeted by the WES capture reagent covered in the aligned sequence of sample 2-3.

Sample 2-3 (Coverage mean= 63.85X)	Percentage %
Bases coverage above 5	97.8
Bases coverage above 10	95.5
Bases coverage above 15	92.3
Bases coverage above 20	88.1

Finally, the average sequencing coverage of the whole sample was calculated using GATK (command displayed in appendix 1.7). Table 3-2 summarizes the average depth of coverage for all the samples sequenced along with the name of the sequencing facility used to generate the data and the sequencing platform used.

Table 3-2: Sequencing details of all the samples included in this study. This table shows the sample numbers of the samples sequenced along with the identifier of the family they belong to, the sequencing facility responsible for the sequencing, the sequencing platform and the coverage across the sample.

Sample name	Family	Sequencing facility	Platform	Coverage
B1	AR1F	Otogenetics	Illumina 2000	33.30X
B2	AR1F	Otogenetics	Illumina 2000	28.98X
ASD433	AR1F	NGS facility at UOL	Illumina 2500	77.69X
ASD434	AR1F	NGS facility at UOL	Illumina 2500	60.81X
IL	SSF	Otogenetics	Illumina 2000	40.22X
FL	SSF	Otogenetics	Illumina 2000	16.54X
RP79	SSF	NGS facility at UOL*	Illumina 2500	65.52X
YL82	SSF	NGS facility at UOL	Illumina 2500	77.08X
ML81	SSF	NGS facility at UOL	Illumina 2500	60.81X
AM	SSF	NGS facility at UOL	Illumina 2500	77.07X
SC1	SC1	Otogenetics	Illumina 2000	80.50X
SC2	SC2	Otogenetics	Illumina 2000	72.86X
SC3	SC3	Otogenetics	Illumina 2000	58.88X
2-1	F2	NGS facility at UOL	Illumina 2500	64.01X

2-3	F2	NGS facility at UOL	Illumina 2500	63.85X
2-5	F2	NGS facility at UOL	Illumina 2500	57.63X
6-1	F6	NGS facility at UOL	Illumina 2500	82.36X
6-3	F6	NGS facility at UOL	Illumina 2500	65.95X
6-10	F6	NGS facility at UOL	Illumina 2500	82.55X
7-1	F7	NGS facility at UOL	Illumina 2500	61.89X
7-2	F7	NGS facility at UOL	Illumina 2500	82.55X
8-1	F8	NGS facility at UOL	Illumina 2500	67.22X
8-2	F8	NGS facility at UOL	Illumina 2500	84.24X
9-1	F9	NGS facility at UOL	Illumina 2500	60.80X
10-1	F10	NGS facility at UOL	Illumina 2500	79.51X
10-2	F10	NGS facility at UOL	Illumina 2500	58.98X
11-3	F11	NGS facility at UOL	Illumina 2000	79.05X
11-4	F11	NGS facility at UOL	Illumina 2000	64.75X
12-1	F12	NGS facility at UOL	Illumina 2000	56.66X
13-1	F13	Otogenetics	Illumina 2000	35.27X
13-3	F13	Otogenetics	Illumina 2000	36.53X
KF50	KF**	-	-	103.56X
KF51	KF	-	-	83.96X
KF52	KF	-	-	82.28X
KF53	KF	-	-	89.49X
KF54	KF	-	-	91.97X

* NGS facility at UOL is the Next Generation facility at the University of Leeds
** KF family data was sent by collaborators and not generated by the author.

Following all these QC checks, the data from the aligned files were analysed by various tools to enable variant calling and recalibration (pipeline described in Appendix 1). This generates a VCF file for each sample, which was annotated using ANNOVAR software (Section 2.12). The output files of ANNOVAR are excel lists (termed raw lists) which underwent further filtering using excel to generate the final variant lists (termed the working lists) for each sample. First, all the intronic and intergenic variants were removed from the raw excel file. The resulting list was then divided in two. The first list included all the ncRNAs (non-coding RNAs), UTRs (Untranslated regions), upstream and downstream variants for future studies; the second list included all the exonic and splicing variants and represented the working excel sheet for this study. This filtering pipeline is shown in Figure 3-8.

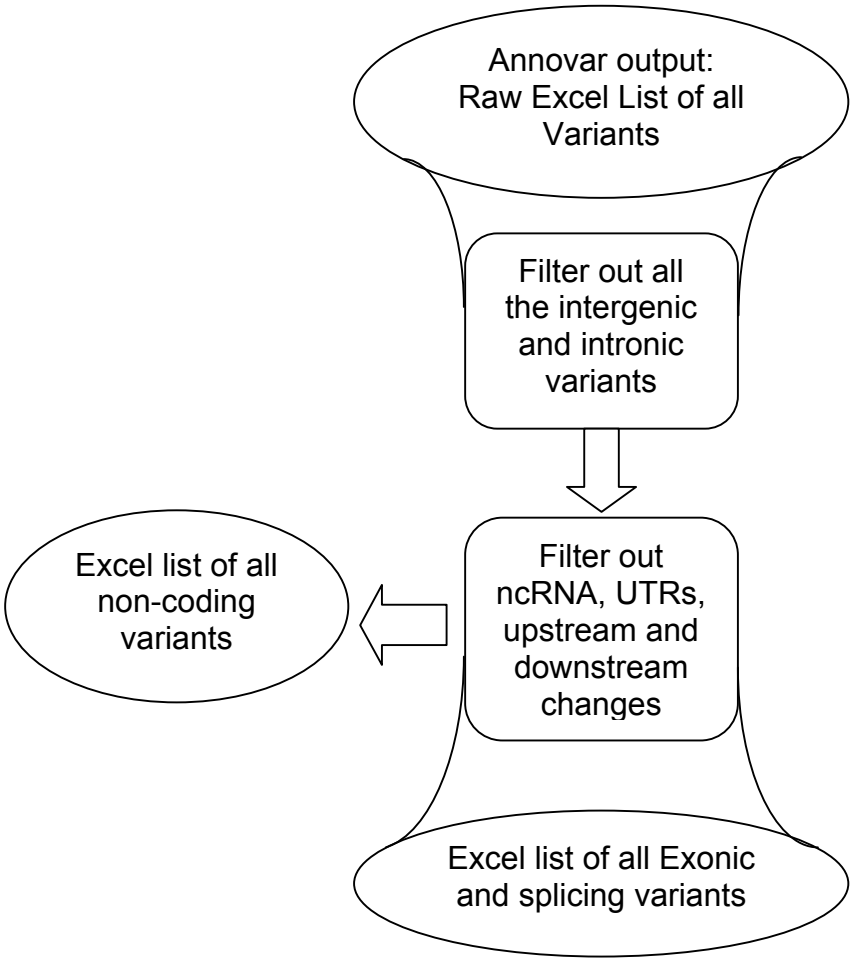


Figure 3-8: Generation of a working excel list of variants. From the ANNOVAR output file at the end of WES sequencing analysis, a working excel list is generated by filtering out all the intergenic, intronic, ncRNAs, UTRs, upstream and downstream variants and leaving only the exonic and splicing variants.

Two additional lists were generated to enable further filtering of the working variant lists. The first list, called sequencing artefacts, combined all the shared variants between non-KC samples sequenced alongside the KC samples on the same Illumina sequencing run. The second list, called the ethnicity specific polymorphism list, comprised all the variants shared between controls from the same ethnic background as the KC samples. Two different ethnicity specific polymorphism lists were generated for the Druze and Maronites Lebanese populations. A similar polymorphism list was generated for the Pakistani samples using data obtained from the BiB database (Section 2.12).

All of the working excel sheets underwent a systematic filtering pipeline combining the autozygosity mapping data and the WES data. Although the original rationale for this study was to map a rare recessive homozygous allele, alternative models were also investigated. Figures 3-9, 3-20 and 3-21 represent the workflows for all the different filtering strategies utilised in this study.

Figure 3-9 describes the filtering strategy applied to the working excel sheets in order to screen for potential alleles with large effect sizes and a homozygous recessive mode of inheritance. For this model, all the variants with a MAF (Minor Allele Frequency) higher than 2% were filtered out, along with all the heterozygous variants, synonymous and non-coding variants. The resulting list was then compared against the list of artefacts and everything shared between the two lists was removed. In addition, all the polymorphisms identified in samples from the same ethnicity were removed. These polymorphisms were considered as the SNPs that appeared in more than 50% of the samples belonging to the same ethnicity.

Although all the data presented below for each family is based on the full dataset, the data was not generated in one go. At the time that this study was initiated, WES was still quite expensive and so samples were sent in a sequential manner as a cost saving exercise. Similarly, SNP data was often generated after the WES on key individuals to try and filter variants and not as a locus mapping tool as is often the case for gene identification studies.

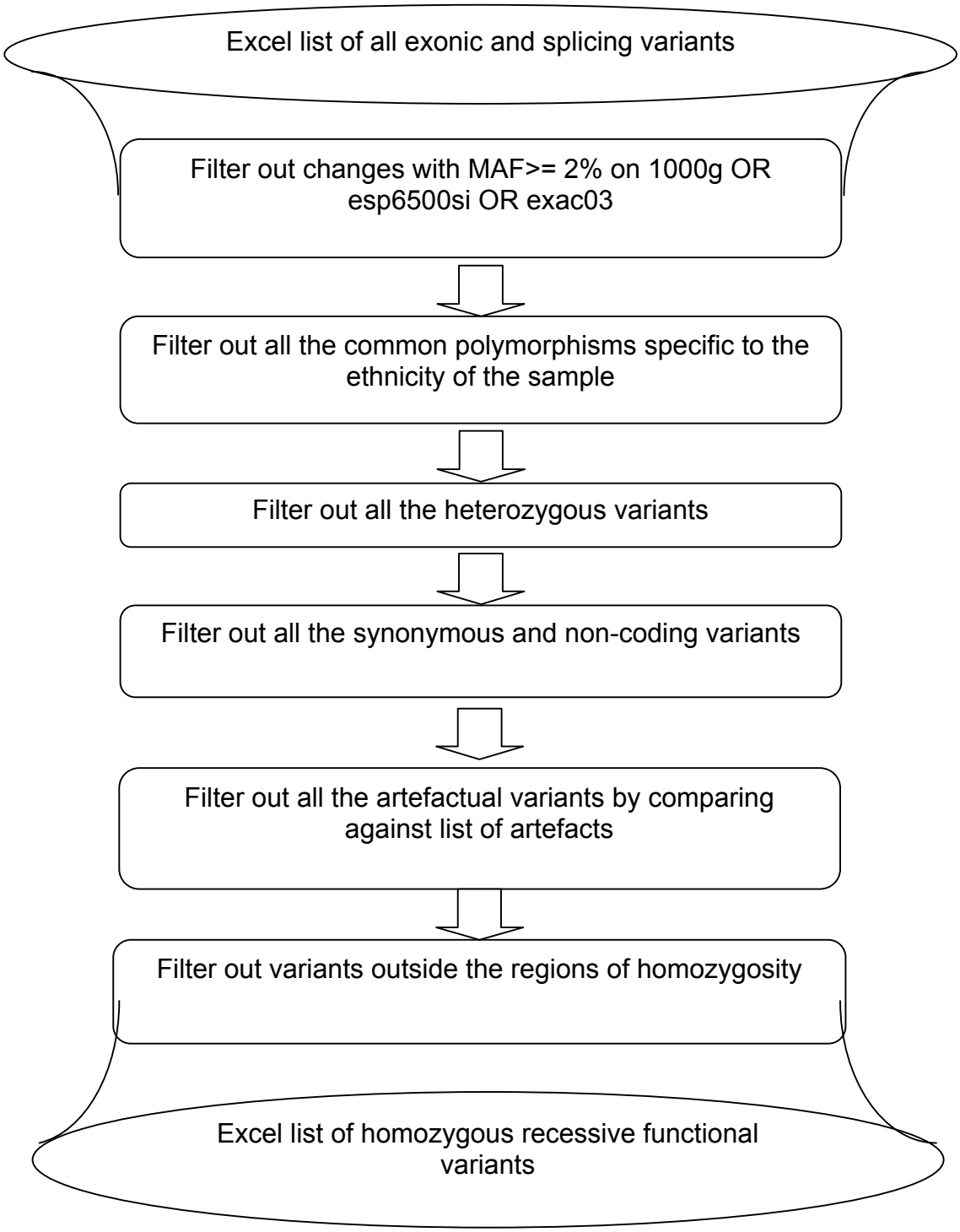


Figure 3-9: Filtering strategy for homozygous recessive hypothesis. The working excel file generated from ANNOVAR undergoes a series of filtering steps by removing any allele with a (Minor Allele Frequency) MAF>=2%, common polymorphisms specific to the same ethnicity, the artefactual variants, the synonymous and non-coding variants and all the heterozygous variants. In addition, all the variants falling outside of the autozygous regions were filtered out where autozygosity mapping data was available. 1000g is the database for the 1000 genome project, esp6500si is the database for the exome variant server and exac03 is the database for the exome aggregation consortium.

3.2.3 Results of the search for rare homozygous variants

3.2.3.1 Families of South Asian origin

3.2.3.1.1 SS1F family

SS1F pedigree is shown in figure 3-1. Samples V:11 and V:12 (FL, IL) were sent to Otogenetics for WES and samples V:2, V:9, V:10 and V:18 (RP79, YL82, AM and ML81) were processed in Leeds. As this family had a likely recessive mode of inheritance, autozygosity mapping was performed on affected family members V:2, V:11 and V:12 (RP79, FL and IL) who are all joined by a clear consanguineous loop. The data was analysed using IBDfinder and the results are presented in Figure 3-10 using the MultiIdeogram software. The software was set to define regions of autozygosity larger than 1Mb.

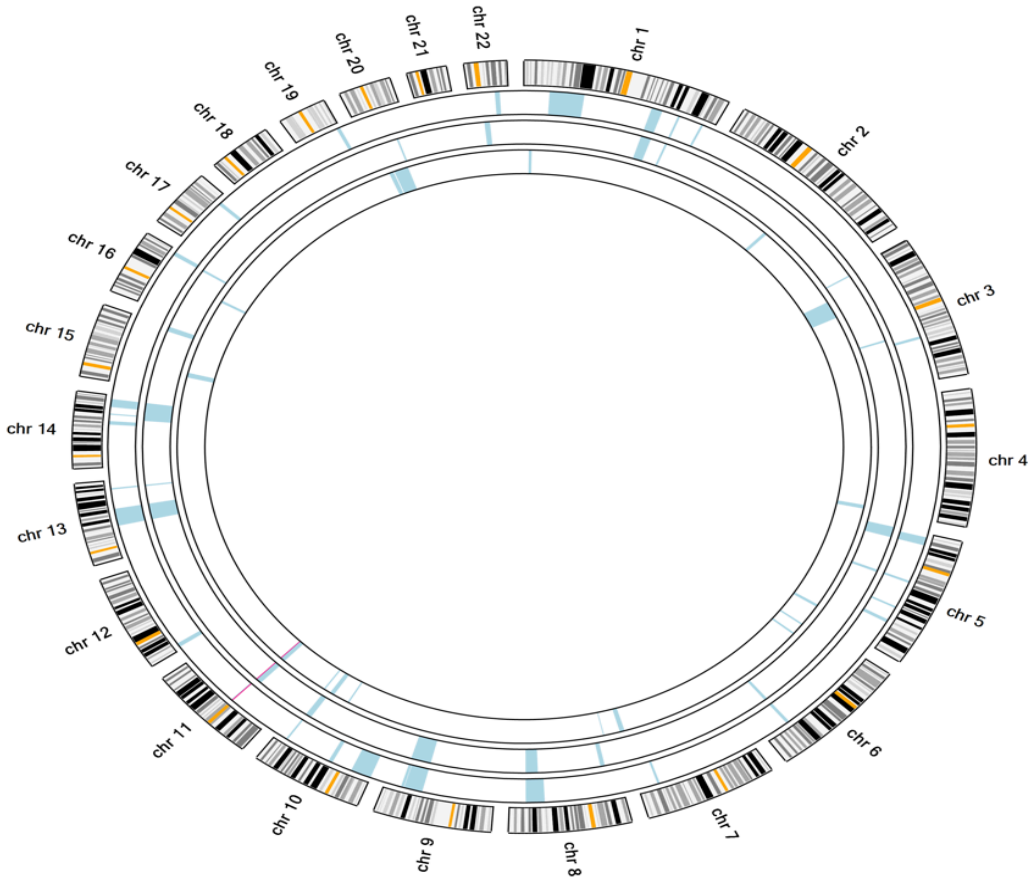


Figure 3-10: SS1F autozygosity mapping. Generated by AgileMultideogram software, this scheme shows in blue the homozygous regions of samples V:11, V:2 and V:12 (FL, RP79 and IL) respectively from the outside to the centre, identified from SNP genotyping data. The only common autozygous region is a 1.2 Mb segment on chromosome 11 (55093883-56295143 hg19) shared between the 3 samples, highlighted in red.

Only one 1.2Mb region of autozygosity was shared between all three affected individuals: chromosome 11: 55,093,883-56,295,143 hg19. SNP data was extracted from the raw non-filtered WES data generated for family members V:9, V:10 and V:18 (YL82, AM and ML81) and combined to the SNP mapping data using AgileMultiIdeogram (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (Section 2.7). The results of this combined data for all six samples is displayed in figure 3-11.

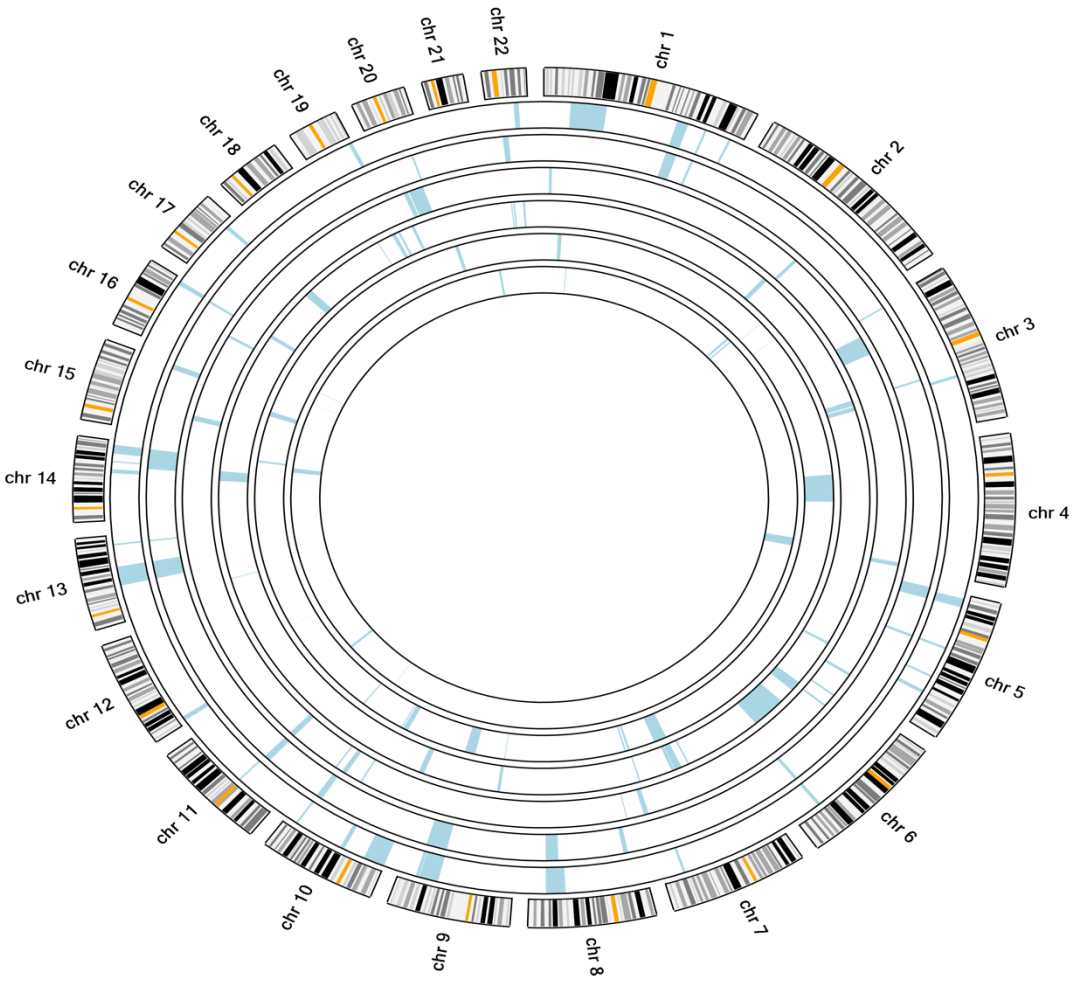


Figure 3-11: SS1F homozygous regions of 6 samples. The multiIdeogram displayed in this picture does not highlight any shared homozygous regions between the 6 SS1F samples, it only highlights in blue the homozygous regions of 6 SSF samples (From outside to the centre: FL, RP, IL, YL82, AM, ML81).

Autozygosity mapping using this combined SNP and WES data for six affected members of SS1F did not reveal any shared homozygous regions. However, this way of determining the shared homozygous regions among the samples with WES and SNP data is not ideal as the WES data is enriched for SNPs in the coding region and will therefore have gaps in the data compared to the data generated from SNP genotyping arrays. Therefore, the original 1.2 Mb candidate region mapped in three individuals was focused upon. This region according to the UCSC genome browser is comprised mainly of a set of olfactory receptor genes (Table 3-3).

Table 3-3: Gene laying in the region of chromosome 11: 55,093,883-56,295,143 hg19. This data was obtained from the genome browser UCSC (https://genome.ucsc.edu/).

Gene name	Strand	Start	End	Exon Count
OR4A16	+	55110676	55111663	1
OR4A15	+	55135359	55136394	1
OR4C15	+	55321782	55322895	1
OR4C16	+	55339603	55340536	1
OR4C11	-	55370829	55371874	1
OR4P4	+	55405833	55406772	1
OR4S2	+	55418379	55419315	1
OR4C6	+	55432642	55433572	1
OR5D13	+	55540913	55541858	1
OR5D14	+	55563031	55563976	1
OR5L1	+	55578942	55579878	1
OR5D18	+	55587105	55588047	1
OR5L2	+	55594694	55595630	1
OR5D16	+	55606227	55607214	1
TRIM51	+	55650772	55659284	7
TRIM51	+	55653455	55659284	5
OR5W2	-	55681125	55682058	1
OR5I1	-	55702931	55703876	1
OR10AG1	-	55735033	55735939	1
OR7E5P	-	55746178	55753881	4
OR5F1	-	55761156	55762101	1
OR5AS1	+	55797894	55798869	1
OR8I2	+	55860783	55861716	1
OR8H2	+	55872518	55873457	1
OR8H3	+	55889848	55890787	1
OR8J3	-	55904246	55905194	1
OR8K5	-	55926869	55927793	1
OR5J2	+	55944093	55945032	1
OR5T2	-	55999581	56000661	1

OR5T3	+	56019675	56020698	1
OR5T1	+	56043114	56044095	1
OR8H1	-	56057602	56058538	1
OR8K3	+	56085782	56086721	1
OR8K1	+	56113514	56114474	1
OR8J1	+	56127690	56128764	1
OR8U8	+	56143099	56508560	2
OR5R1	-	56184733	56185708	1
OR5M9	-	56229944	56230877	1
OR5M3	-	56236963	56238014	1
OR5M8	-	56257910	56258846	1

The 1.2Mb region was closely examined in the unfiltered raw WES data of the SSF family samples. No homozygous variants in this region segregated in all the samples of the SSF family. However, all the patients were manually looked at. Variants in this region were only detected in the three samples where the 1.2 Mb region was highlighted (RP79, FL and IL). However none of the variants in the same gene segregated in the family patients, indicating that there is no single recessive homozygous allele which is the primary cause of KC in this family.

3.2.3.1.2 AR1F family

AR1F family's pedigree is displayed in figure 3-2. Patient V:3 was sent to Otogenetics for WES and samples IV:9, IV:15 and IV:16 were sequenced in Leeds. The mode of inheritance in this pedigree is not well-defined as previously described in Section 3.2.1. SNP microarray genotyping data for patient V:3 was analysed using IBDfinder and the results are presented in table 3-4. Combining The WES data with the SNP data of V:3 patient reveals a long list of variants.

Table 3-4: Autozygosity mapping results of sample AR1F V:3. Physical autozygosity mapping regions with the corresponding coordinates (hg19) for sample V:3 family AR1F.

Chromosome	Start	End	Length (Mb)
1	167779537	234248477	66.46
4	25309685	82249412	56.93
1	105424056	159547992	54.12
2	171109741	212567763	41.45

6	53399535	88679038	35.27
2	80086831	115289261	35.20
4	90335855	125093565	34.75
12	91793711	120736039	28.94
8	101462127	124950627	23.48
3	119376519	141424837	22.04
6	98986070	117807422	18.82
10	72918171	87710065	14.79
2	216471816	230224721	13.75
18	44753290	55774150	11.02
13	85806730	94921466	9.11
18	28426737	36933980	8.50
18	67804296	76116029	8.31
1	235892296	244006957	8.11
9	77355456	85182797	7.82
10	54395856	61517046	7.12
15	85189105	92149266	6.96
11	63959003	69842871	5.88
8	12884163	18640423	5.75
14	89448991	94430979	4.98
6	90686614	95513635	4.82
12	127947706	132287718	4.34
20	5726406	9612854	3.88
19	60708414	63786938	3.07
14	40133556	42402597	2.26
10	14223111	16272890	2.04
2	128713965	130244008	1.53

To test the hypothesis that all affected members of this family have recessive homozygous mutations in the same gene, the SNP data from sample V:3 and raw non-filtered WES data from samples IV:9, IV:15 and IV:16 were combined using the multiIdeogram software in order to check for shared homozygous regions. The result of this combination is shown in figure 3-12, which does not reveal any shared homozygous region between the samples. In order to confirm this result, the raw WES data of all the four samples V:3, IV:9, IV:15 and IV:16 were checked for shared homozygous variants that were subsequently combined with the SNP data. A list of variants was highlighted (Table 3-5). This list of homozygous variants shown to be shared between all the affected members of AR1F and falling into the autozygous regions in sample V:3 can be excluded from harbouring any

variant that can be potentially responsible for KC pathogenesis, since all of these variants have MAF > 2%.

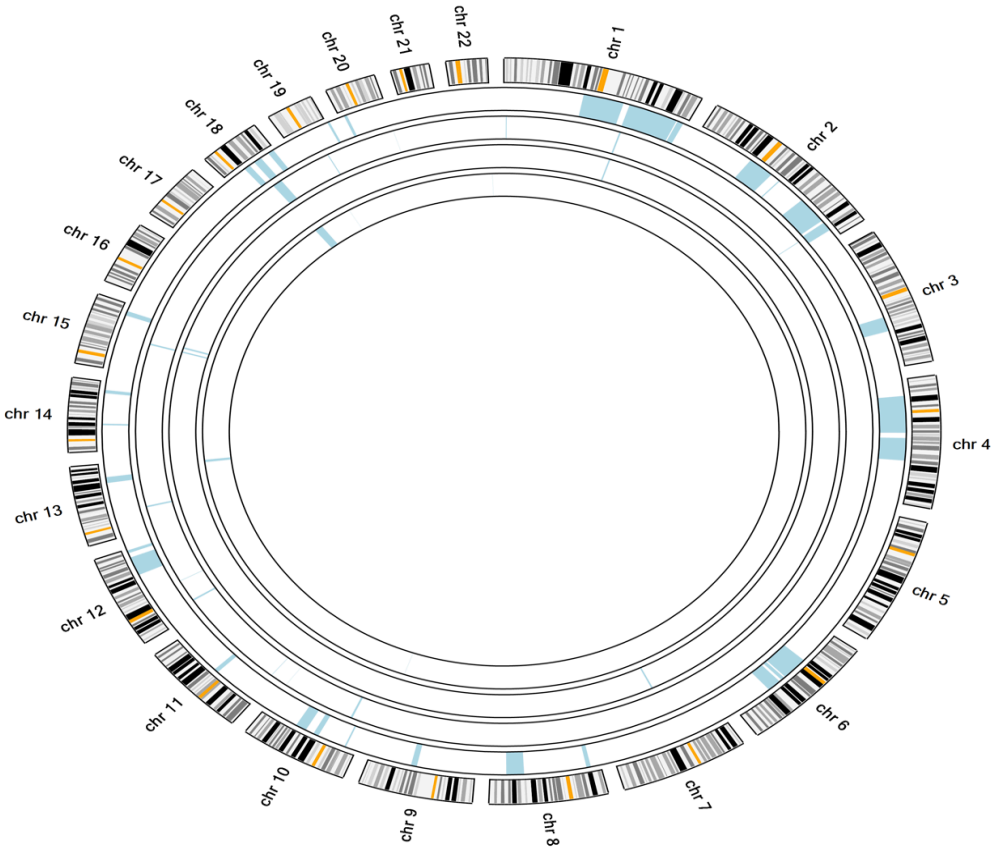


Figure 3-12: Homozygous regions in the four tested samples from the AR1F family. The multiIdeogram displayed in this picture does not highlight any shared homozygous regions between the 4 AR1F samples, it only highlights in blue the homozygous regions of the 4 AR1F samples (From outside to the centre: V:3, V:9, V:15 and V:16).

Table 3-5: Homozygous variants in the AR1F family. This table shows the homozygous variants shared between all the affected members of the AR1F family found in the autozygous regions observed in sample V:3. The MAF of all the variants presented in this table is > 2%.

Gene	Function	AA Change	esp6500si	1000g	exac03	snp138
HSD3BP4	ncRNA_exonic	.	.	0.503794	.	rs12745167
METTL11B	Exonic (synonymous)	exon3:c.C516T:p.Y172Y	0.080815	0.0585064	0.059	rs41272487
METTL11B	Exonic (synonymous)	exon4:c.C741T:p.S247S	0.081034	0.0585064	0.054	rs12075997
TUBA4B	ncRNA_exonic	.	.	0.502396	.	rs10166888
LOC100507291	ncRNA_exonic	.	.	0.25	.	rs11706576
PDGFRL	Splicing	.	.	0.767372	.	rs2705051
KRT19P2	ncRNA_exonic	.	.	0.951877	.	rs2289029

In order to confirm the exclusion of any homozygous autosomal variants from the potential involvement in KC pathogenesis in this family, the WES analysis was explored differently by applying the filtering strategy shown in Figure 3-9 to all the sequenced affected AR1F members. The filtered lists of samples V:3; IV:9; IV:15 and IV:16 were compared using an in house comparison perl script (Section 2.12) and checked for homozygous shared variants. The list of shared variants was then filtered according to the filtering strategy shown in Figure 3-9. The results are displayed in table 3-6, and show only one change on chromosome X, *TENM1*: exon29 c.A6554T p.H2185L, that is shared between all the AR1F affected members. This result is consistent with the pedigree structure and raises the suggestion of an X-linked variant underlying KC.

TENM1 encodes a protein of a heparin binding function and involved in the immune response, negative regulation of cell proliferation and nervous system development processes (Brandau *et al.*, 1999, Minet *et al.*, 1999). In the AR1F family (Table 3-6), this gene harbours a change in exon 29 of the NM_014253 transcript and presents a high conservation score of 669 according to the phastConsElements46way annotation and a MAF < 2% on esp6500 (0.007), 1000g (0.003) and ExAc03 (0.005).

Table 3-6: Homozygous variant shared between the AR1F affected members.
Details of the variant on chromosome X shared between AR1F affected members

Gene	Chr	Function	AA Change	SNP	CADD Phred
TENM1	Xq25	Nonsynonymous	exon29:c.A6554T:p.H2185L	rs144335593	17.35

3.2.3.1.3 Single cases (SCs)

DNA samples of patients SC1, SC2 and SC3 were analysed by SNP microarray genotyping by AROS Biotec (Section 2.7) and WES by Otogenetics (Section 2.8). The genotyping results are presented in figures 3-13, 3-14 and 3-15 and summarized in tables 3-7, 3-9 and 3-11. All three samples were filtered on the basis of harbouring a homozygous recessive

mutation (Figure 3-9) in order to generate lists of homozygous variants. None of the parents were available for sequencing or microarray genotyping, so the lists were not compared to any other family members for filtering purposes. The list of homozygous variants resulting from the combination of the two methods of microarray genotyping and WES were prioritized on CADD Phred scores and Tables 3-8, 3-10 and 3-12 show the variants with scores > 10.

For Sample SC1:

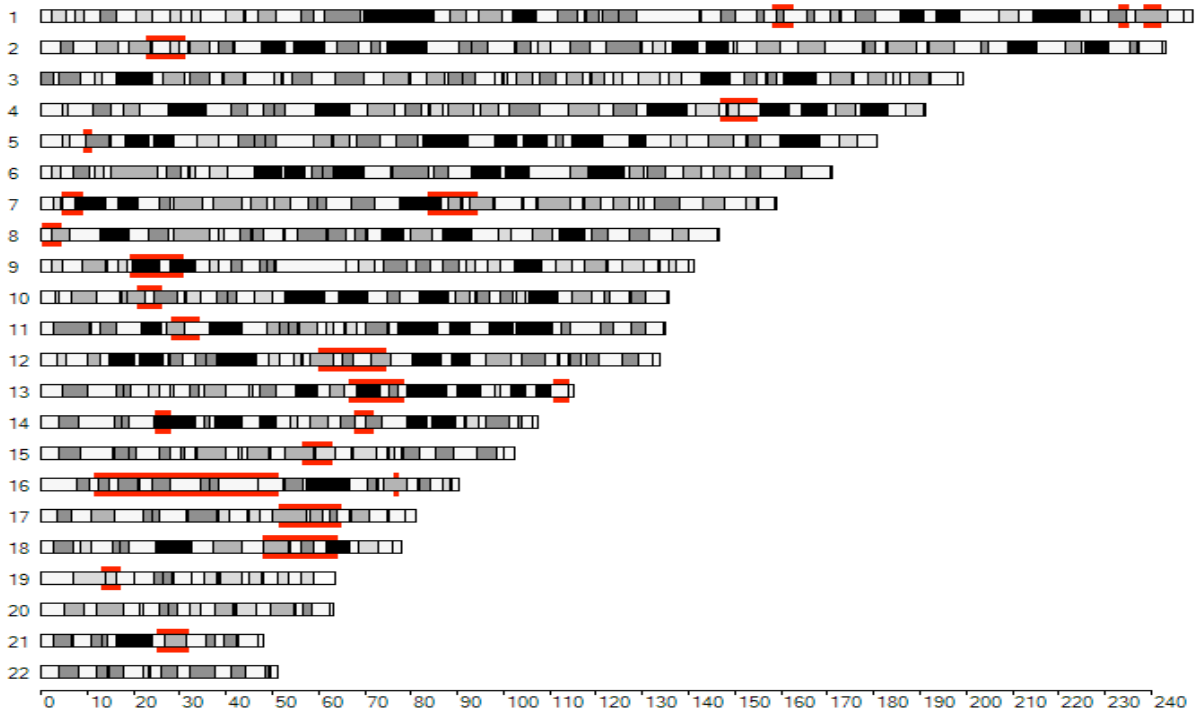


Figure 3-13: Autozygosity results for sample SC1. The autozygous regions found in sample SC1 are highlighted in red. The numbers listed down the left hand side correspond to the numbers of the Chromosomes and the (x) axis gives the genomic location in Mb.

Table 3-7: Autozygosity mapping results of sample SC1. Physical autozygosity mapping regions in patient SC1 with the corresponding coordinates.

Chromosome	Start	End	Length (Mb)
16	11548039	51239640	39.69
18	47992238	64099099	16.10
12	60048676	74643059	14.59
17	51458068	64993508	13.53
13	66535549	78634249	12.09
9	19191738	30928150	11.73
7	83610164	94321588	10.71
2	22830138	31348359	8.51

4	146952104	155217163	8.26
21	24944108	31871469	6.92
15	56553728	63141897	6.58
11	28248131	34238805	5.99
10	20648355	26147284	5.49
1	158023797	162975122	4.95
7	4298964	9196407	4.89
19	12856403	17158095	4.30
8	103565	4373222	4.26
14	67855086	71942461	4.08
1	238595577	242438920	3.84
14	24648557	28153133	3.50
13	110756667	114125098	3.36
1	233185558	235276626	2.09
5	9247630	11091986	1.84
16	76278544	77427046	1.14

Table 3-8: Homozygous variants in sample SC1. Homozygous variants in the autozygous regions observed in sample SC1 and having CADD Phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
CUEDC1	Nonsynonymous	17q22	exon2:c.G110C:p.R37P	.	34
ADAMTS5	Nonsynonymous	21q21.3	exon5:c.G1729A:p.G577S	rs146717299	34
COL28A1	Nonsynonymous	7p21.3	exon19:c.C1544T:p.P515L	rs79817423	32
NWD1	Nonsynonymous	19p13.11	exon5:c.G433A:p.A145T	.	26.3
EPHX3	Nonsynonymous	19p13.12	exon1:c.C191T:p.P64L	.	24.2
CYP4F2	Nonsynonymous	19p13.12	exon9:c.C1021G:p.L341V	rs145174239	24
ARHGEF10	Nonsynonymous	8p23.3	exon27:c.G3256C:p.V1086L	.	23.9
ACE	Nonsynonymous	17q23.3	exon12:c.T1811C:p.L604P	rs140941300	23.7
CYP4F11	Nonsynonymous	19p13.12	exon5:c.C538T:p.R180C	rs148197835	21.6
INTS2	Nonsynonymous	17q23.2	exon25:c.G3605T:p.S1202I	.	20.4
C16orf62	Nonsynonymous	16p12.3	exon27:c.G2635A:p.A879T	rs34342607	19.79
TLN2	Nonsynonymous	15q22.2	exon15:c.G1838C:p.R613T	rs115132378	18.89
KRTAP26-1	Nonsynonymous	21q22.11	exon1:c.C562A:p.P188T	rs12483584	15.91
KRTAP27-1	Nonsynonymous	21q22.11	exon1:c.G284A:p.C95Y	rs117534873	15.36
IFNA17	Frameshift Insertion	9p21.3	exon1:c.170dupA:p.H57fs	rs200400518	15.36
NIT1	Nonsynonymous	1q23.3	exon1:c.A4G:p.T2A	rs138523655	14.46
PTPRB	Nonsynonymous	12q15	exon14:c.G3281A:p.R1094Q	.	14.46
PDILT	Nonsynonymous	16p12.3	exon3:c.A319G:p.K107E	rs139688634	13.5
C2orf71	Nonsynonymous	2p23.2	exon2:c.C3794T:p.P1265L	.	10.69

For sample SC2:

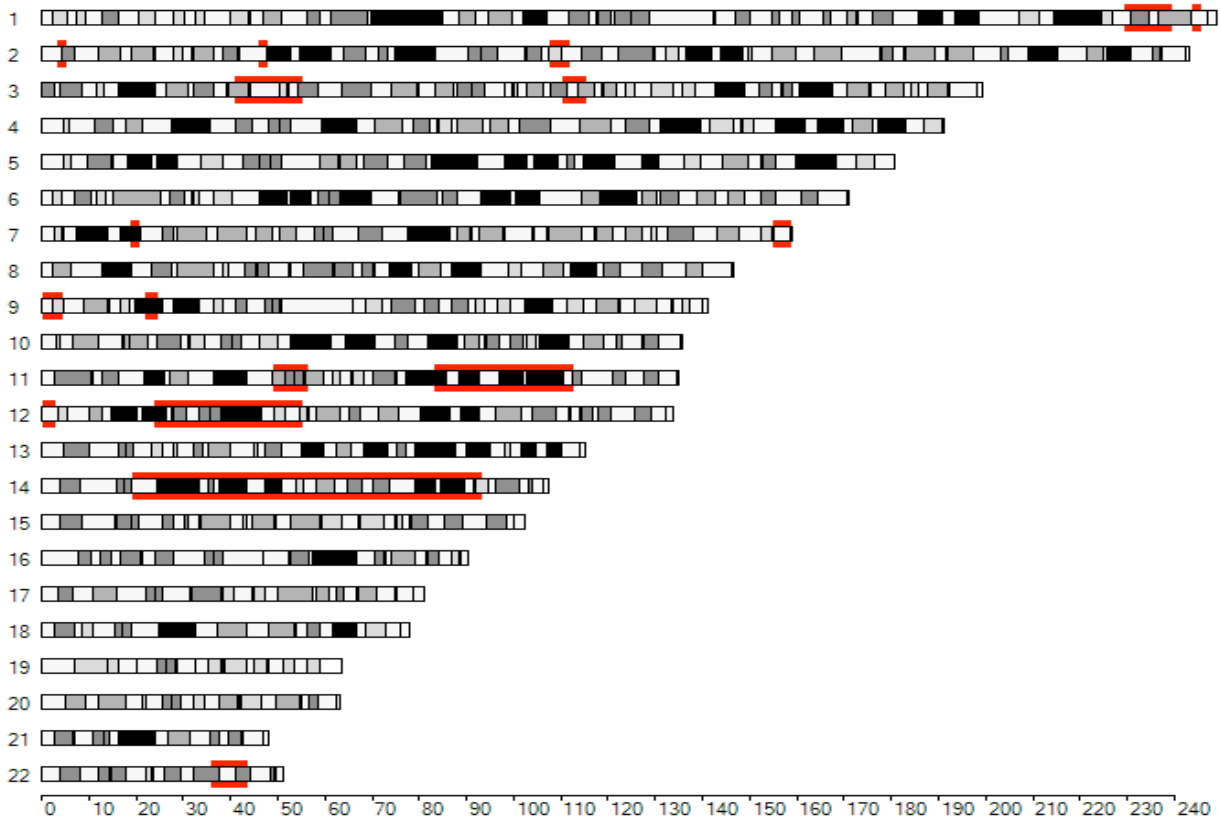


Figure 3-14: Autozygosity results for sample SC2. The autozygous regions of sample SC2 are highlighted in red. The numbers listed to the left hand side correspond to the numbers of the Chromosomes and the (x) axis gives the genomic location in Mb.

Table 3-9: Autozygosity mapping results of sample SC2. Physical autozygosity mapping regions with the corresponding coordinates for sample SC2.

Chromosome	Start	End	Length (Mb)
14	19351304	93409780	74.05
12	23769035	55150037	31.38
11	83424592	112682399	29.25
3	40806123	55481530	14.67
1	229488558	239592241	10.10
22	36085550	43634843	7.54
11	48991209	56340128	7.34
3	110479018	115627393	5.14
2	107746055	112114969	4.36
9	36587	4348391	4.31
7	154941793	158811981	3.87
12	36594	3091411	3.05
9	22122076	24651204	2.52
2	45853323	48065559	2.21
1	243725341	245856184	2.13

7	18760442	20757367	1.99
2	3472592	5152338	1.67

Table 3-10: Homozygous variant of sample SC2. Details of homozygous variants falling in the autozygous regions observed in sample SC2 and having CADD phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
IP6K1	Nonsynonymous	3p21.31	exon4:c.G484T:p.V162F	rs199967442	33
ABHD14B	Nonsynonymous	3p21.2	exon3:c.G424A:p.V142M	.	32
TRAPPC6B	Nonsynonymous	14q21.1	exon4:c.C317G:p.S106C	.	28.5
ABHD10	Nonsynonymous	3q13.2	exon5:c.G737C:p.G246A	.	27
MYH7	Nonsynonymous	14q11.2	exon27:c.G3667C:p.E1223Q	.	24.5
PLEKHH1	Nonsynonymous	14q24.1	exon7:c.C938T:p.P313L	rs200665235	24.3
DPF3	Nonsynonymous	14q24.2	exon9:c.G1190A:p.R397H	.	23.5
LRRC16B	Nonsynonymous	14q11.2	exon33:c.C3385T:p.R1129C	rs116307579	23.4
NGB	Nonsynonymous	14q24.3	exon3:c.G292C:p.A98P	.	23.2
PELI2	Nonsynonymous	14q22.3	exon2:c.A97G:p.N33D	.	22.5
KRT5	Nonsynonymous	12q13.13	exon1:c.G110A:p.R37Q	rs61747181	21.6
PARP3	Nonsynonymous	3p21.2	exon7:c.G943C:p.V315L	.	18.41
BAIAP2L2	Nonsynonymous	22q13.1	exon12:c.T1351C:p.S451P	.	17.71
CWF19L2	Nonframeshift Deletion	11q22.3	exon14:c.2173_2175del:p.725_725del	rs368279169	16.82
HOXC9	Nonframeshift Deletion	12q13.13	exon1:c.66_68del:p.22_23del	.	16.61
DDHD1	Nonsynonymous	14q22.1	exon1:c.G181C:p.G61R	.	15.03
KRT6C	Nonsynonymous	12q13.13	exon1:c.T2C:p.M1T	rs199806927	14.78
KDELC2	Nonsynonymous	11q22.3	exon1:c.C28A:p.L10M	.	14.7
JDP2	Nonsynonymous	14q24.3	exon2:c.C25A:p.P9T	rs201898432	14.49

For sample SC3:

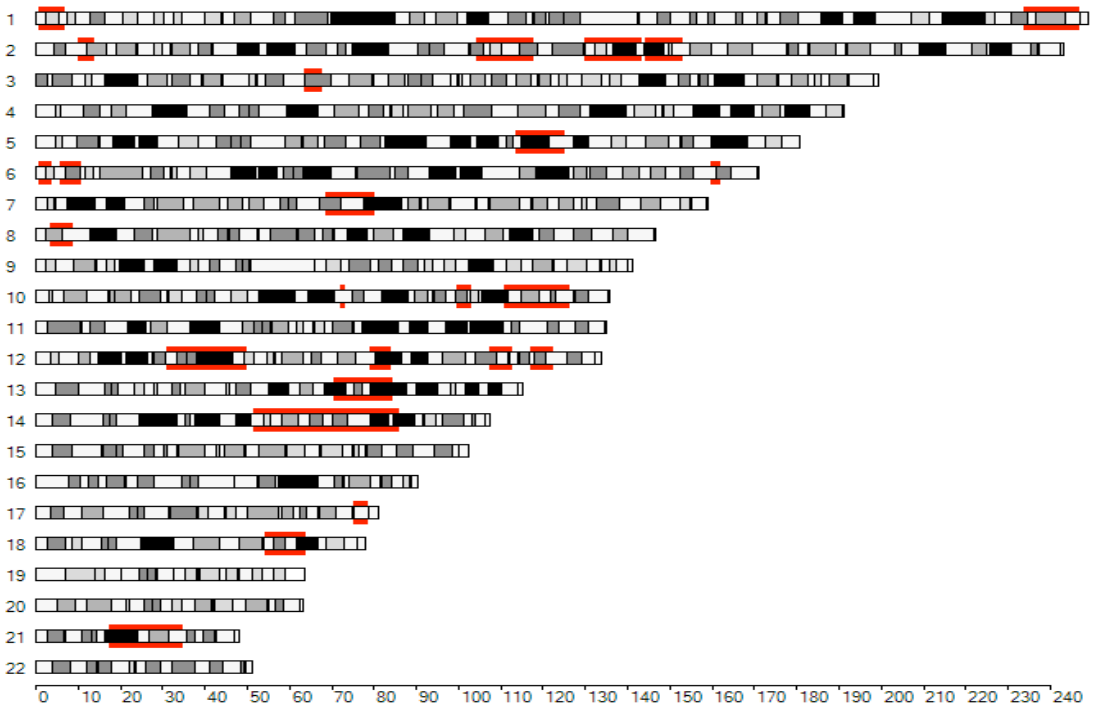


Figure 3-15: Autozygosity results for sample SC3. The autozygous regions observed in sample SC3 are highlighted in red. The numbers listed on the left hand side are the numbers of the Chromosomes and the (x) axis gives the genomic location in Mb.

Table 3-11: Autozygosity mapping results for sample SC3. Physical autozygosity mapping regions with the corresponding coordinates observed in sample SC3.

Chromosome	Start	End	Length (Mb)
14	51491136	85872591	34.38
12	31006813	50032062	19.02
21	17353179	34761104	17.40
10	110848562	126166794	15.31
13	70617166	84514542	13.89
2	129724657	143381900	13.65
2	104191926	117784823	13.59
1	233841459	247165315	13.32
5	113424277	125175740	11.75
7	68410838	80088747	11.67
18	53989956	64016019	10.02
2	144134053	153195011	9.06
1	554484	6632197	6.07
12	116901625	122624505	5.72
8	3139266	8588828	5.44
12	107212660	112653546	5.44
12	79057286	83946721	4.88

6	5657139	10532295	4.87
3	63390235	67877655	4.48
2	9977295	13718817	3.74
17	74942188	78643088	3.70
10	99508023	103004622	3.49
6	775293	3572105	2.79
6	159867886	162088496	2.22
10	71827455	73236902	1.40

Table 3-12: Homozygous variants observed in sample SC3. Details of homozygous variants falling in the autozygous regions observed in sample SC3 and having CADD phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
TAS1R1	Nonsynonymous	1p36.31	exon5:c.1284_1285del:p.F428fs	.	35
ESPN	Nonsynonymous	12q12	exon8:c.C1894T:p.R632C	.	34
PDZRN4	Nonsynonymous	12q13.12	exon1:c.G511A:p.E171K	rs151202383	32
TROAP	Frameshift Deletion	1p36.31	exon13:c.C2036T:p.P679L	.	27
DBX2	Nonsynonymous	12q12	exon2:c.C469T:p.R157W	rs138216785	26.1
NRAP	Splicing(+2)	10q25.3	exon17:c.1635+2T>C,	.	25
P2RX7	Nonsynonymous	12q24.31	exon8:c.G827A:p.R276H	rs7958316	24.2
ACACB	Nonsynonymous	12q24.11	exon36:c.C5057G:p.P1686R	rs61739667	24.1
PLG	Nonsynonymous	6q26	exon11:c.G1335T:p.R445S	rs371661937	23.9
ST6GAL2	Nonsynonymous	2q12.2	exon2:c.T89A:p.F30Y	rs62155507	23.8
PTPRQ	Nonsynonymous	12q21.31	exon6:c.C769G:p.Q257E	rs61729287	23.6
LRP1B	Nonsynonymous	2q22.1	exon41:c.G6436T:p.V2146F	rs145744070	23.1
KLHL17	Nonsynonymous	1p36.33	exon12:c.G1819A:p.A607T	.	22.9
POU3F3	Nonsynonymous	2q12.1	exon1:c.C875A:p.P292Q	.	22.8
PHLPP1	Nonsynonymous	18q21.33	exon1:c.G1066A:p.A356T	rs12957017	22.4
CBX8	Nonsynonymous	17q25.3	exon5:c.G1077C:p.E359D	.	21
KCNH3	Nonsynonymous	12q13.12	exon12:c.C2219T:p.T740M	.	18.07
CCDC177	Nonsynonymous	14q24.1	exon2:c.T1952C:p.L651P	.	17.2
DACH1	Nonsynonymous	13q21.33	exon1:c.G244A:p.G82S	rs202136379	11.44

3.2.3.2 Family KF

Family KF's pedigree (Figure 3-3) is consistent with autosomal recessive inheritance with affected individuals having consanguineous parents, so autozgosity mapping was performed. Affected siblings IV:1 and IV:2 were genotyped by SNP microarray and all three affected siblings (IV:1, IV:2 and

IV:3) along with their parents (II:3 and III:4) underwent WES analysis. The SNP microarray results are presented in figure 3-16 and table 3-13.

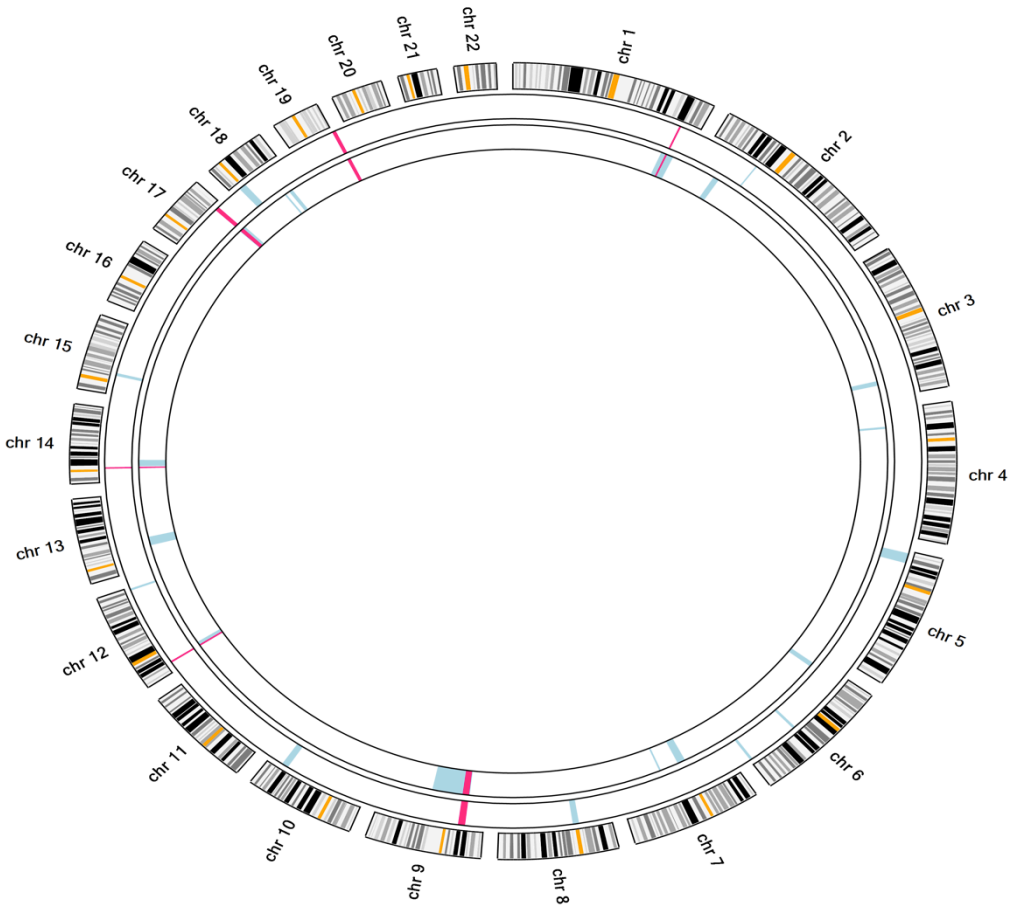


Figure 3-16: Autozygosity mapping in family KF. Visualisation of autozygous regions observed in patients IV:1 and IV:2 respectively, from the outside to the centre. The shared autozygous regions are highlighted in red. The homozygous regions observed in individual samples are shown in blue.

Table 3-13: Autozygosity mapping results in patients IV:1 and IV:2 of family KF. Regions highlighted by autozygosity mapping in patients IV:1 and IV:2 from the family KF with the corresponding coordinates for shared regions.

Chromosome	Start	End	Length (Mb)
9	24026560	33046299	9.01
17	69210898	74785180	5.57
19	58998630	63786938	4.78
1	221701053	223938150	2.23
12	12925252	15159222	2.23
14	20027606	21921082	1.89

Since the WES data from the third sibling IV:3 is available, these regions were checked for homozygosity in the raw unfiltered WES data of sample IV:3. Table 3-14 shows the autozygous regions shared between the three affected siblings of family KF.

Table 3-14: Autozygosity mapping results from analysis of all three affected members of family KF. Regions highlighted by autozygosity mapping in samples IV:1 and IV:2 of KF family, with the corresponding coordinates for shared regions.

Chromosome	Start	End	Length (Mb)
9	25678887	30408448	4.72
12	13370236	15066849	1.69
17	72542827	73062038	0.51
17	74498068	74709434	0.21

In addition, the WES data from the parents exist for this family. A closer look at the data from the unaffected parents (II:2 and III:4) highlighted a homozygous segment of 1.5Mb on Chromosome 9: 25,779,193-27,330,031 in sample II:2 and a homozygous segment of 1.6Mb on Chromosome 9: 25,679,024-27,330,031, which can be excluded from the autozygous regions highlighted in Table 3-14 in the affected siblings as potential regions harbouring potential homozygous pathogenic variants. The final regions of autozygosity > 1Mb highlighted in the affected members of family KF is presented in Table 3-15.

Table 3-15: Autozygosity mapping results observed in analysis of all affected members of family KF and excluded from the unaffected parents. Physical autozygosity mapping regions with the corresponding coordinates, for regions shared in samples IV:1 and IV:2 of KF family, in addition to sample IV:3.

Chromosome	Start	End	Length (Mb)
9	27330031	30408448	3.07
12	13370236	15066849	1.69

In order to detect the homozygous variants falling in these regions of homozygosity, shared between the affected siblings of KF and excluded from the homozygous variants list of the unaffected parents, the filtering strategy shown in Figure 3-8 was applied to all the sequenced affected

siblings (IV:1, IV:2 and IV:3) and unaffected parents (II:3 and III:4) of KF family to generate working excel lists. Subsequently, two major lists were generated: the first list includes all the shared homozygous variants between all the affected siblings (IV:1, IV:2 and IV:3) and the second list represents all the shared homozygous variants between the unaffected parents (II:3 and III:4). Then the two filtered lists were compared using an in house comparison perl script (Section 2.12). The resulting list excluded all the homozygous variants shared between the parents and the siblings and was then filtered according to the filtering strategy shown in Figure 3-9. This strategy did not highlight any homozygous variant shared between the KF affected siblings and absent in the parents in the homozygous form. Thus no potentially pathogenic homozygous alleles of major effect predisposing the KC were observed in the autozygous regions in the KF family.

3.2.3.3 Families of Lebanese origin and Druze ethnicity

3.2.3.3.1 Family F2

Family F2 (Figure 3-4) is not known to be consanguineous but belongs to an endogamous community, where the parents are from the same village and carry the same surname, which means they could possibly share the same great-grandparents. Consequently, autozygosity mapping was performed by SNP genotyping all the affected family members (I:3, II:1, II:2 and II:3) (Figure 3-17). The results show that the affected members of F2 do not share any autozygous regions. In parallel, WES analysis was performed in Leeds for individuals I:3, II:1 and II:2 (2/5, 2/1 and 2/3).

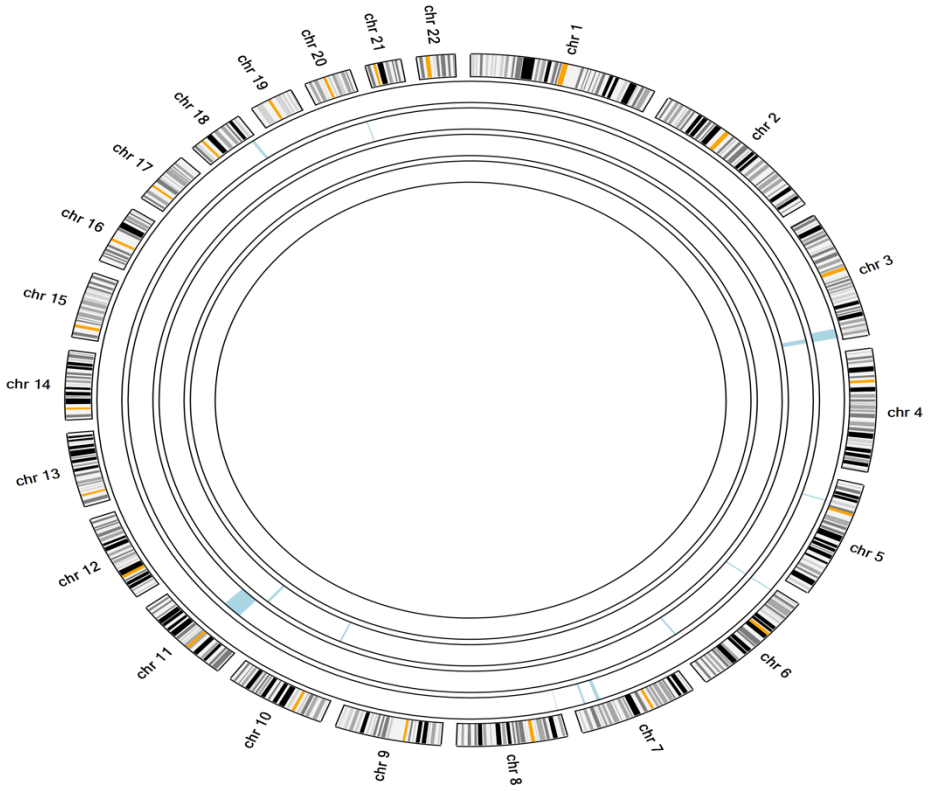


Figure 3-17: Autozygosity mapping in family F2. Visualisation of autozygous regions observed in patients II:1, II:2, I:3 and II:3 respectively, from the outside to the centre. Autozygous regions in individual samples are highlighted in blue.

Although autozygosity mapping failed to highlight a shared autozygous region shared between the affected patients of family F2, homozygous variants were still screened for in the WES data available from affected members of family F2. Since the SNPs in the autozygosity mapping are filtered on locus size (1Mb), double checking the WES data is essential to ensure that mutation(s) in small regions of autozygosity are not missed.

The filtering strategy shown in Figure 3-9 was applied to the working excel lists of the sequenced affected members of F2 (I:3, II:1 and II:2). Next, the filtered lists were compared using an in house comparison perl script (Section 2.12) to check for shared homozygous variants. However, this strategy failed to highlight any homozygous variants shared between the affected F2 members.

3.2.3.3.2 Family F9

Only one patient DNA was available for family F9 (Figure 3-4), this being individual III:8 (9/1). WES analysis was performed and the filtering strategy for homozygous recessive mutations, shown in Figure 3-9, was applied. The resulting list was prioritized based on CADD Phred scores and table 3-16 shows only the homozygous variants having scores > 10.

Table 3-16: Homozygous variants observed in patient III:7 of family F9. Details of homozygous variants identified in family F9 having CADD phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
LMOD3	Nonsynonymous	3p14.1	exon1:c.G248A:p.R83H	rs35740823	26.9
GTF2IRD2	Nonsynonymous	7q11.23	exon16:c.A2105C:p.Y702S	.	24.4
MYO18B	Nonsynonymous	22q12.1	exon43:c.C6901T:p.L2301F	rs182723137	24.4
APPL1	Nonsynonymous	3p14.3	exon20:c.G1888C:p.E630Q	rs142229340	24.3
GPATCH4	Nonsynonymous	1q23.1	exon8:c.G500A:p.R167H	rs199804048	23.7
GALNT12	Nonsynonymous	9q22.33	exon1:c.G179A:p.R60Q	.	20.2
UHRF1	Nonsynonymous	19p13.3	exon4:c.G757C:p.D253H	rs17886098	17.24
POTEF	Nonsynonymous	2q21.1	exon17:c.C2661A:p.D887E	rs201255735	17.16
SPATA31A3	Nonsynonymous	9p13.1	exon4:c.C2783T:p.T928M	rs62565523	16.32
FDFT1	Nonframeshift Deletion	8p23.1	exon1:c.193_198del:p.65_66del	rs71711801	16.2
CHRNA3	Nonframeshift Deletion	15q25.1	exon1:c.67_69del:p.23_23del	rs66793222	15.47
HDGFRP2	Nonframeshift Deletion	19p13.3	exon8:c.840_842del:p.280_281del	.	14.61
TRAK1	Nonframeshift Insertion	3p22.1	exon13:c.1841_1842insGGAGGA:p.T614delinsTEE	.	13.62
PGBD5	Nonsynonymous	1q42.13	exon1:c.G286T:p.A96S	rs113202442	11.99
SCAF1	Nonsynonymous	19q13.33	exon7:c.G1742C:p.R581P	.	11.71
LRP5	Nonsynonymous	11q13.2	exon1:c.G4A:p.E2K	.	10.39

3.2.3.3.3 Family F10

For family F10 (Figure 3-4) WES analysis was undertaken on patients III:4 and III:5 (10/1 and 10/2). The data was filtered as described in figure 3-9 then analysed on the hypothesis of a homozygous recessive mutation being shared by both individuals (Figure 3-9). The resulting list was prioritized on CADD Phred scores and table 3-17 shows only the homozygous variants having scores > 10.

Table 3-17: Homozygous variants shared by the sequenced affected members of F10. Details of homozygous variants shared by III:4 and III:5, KC patients from family F10, and having CADD phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
KCNN3	Nonframeshift Insertion	1q21.3	exon1:c.241_242insAGC :p.P81delinsQP	rs3831942	12.23
ASB18	Nonsynomymous	2q37.2	exon4:c.G722T:p.R241L	rs56247246	11.2

3.2.3.3.4 Family F11

In Family F11 (Figure 3-4), WES analysis and SNP genotyping was performed on samples from patients II:2 and II:3 (11/3 and 11/4). Due to the pedigree structure and the endogamous nature of the community from which the family derived, the data was analysed under a homozygous recessive mutation model. The autozygosity mapping data shows that these patients have no shared autozygous regions (Figure 3-18). Due to the resolution limits of the autozygosity mapping software, the WES data was still analysed under the same shared recessive homozygous allele model but without filtering based on mapping data (Figure 3-9). The resulting list was prioritized on the basis of CADD Phred scores and table 3-18 shows only the homozygous variants having scores > 10.

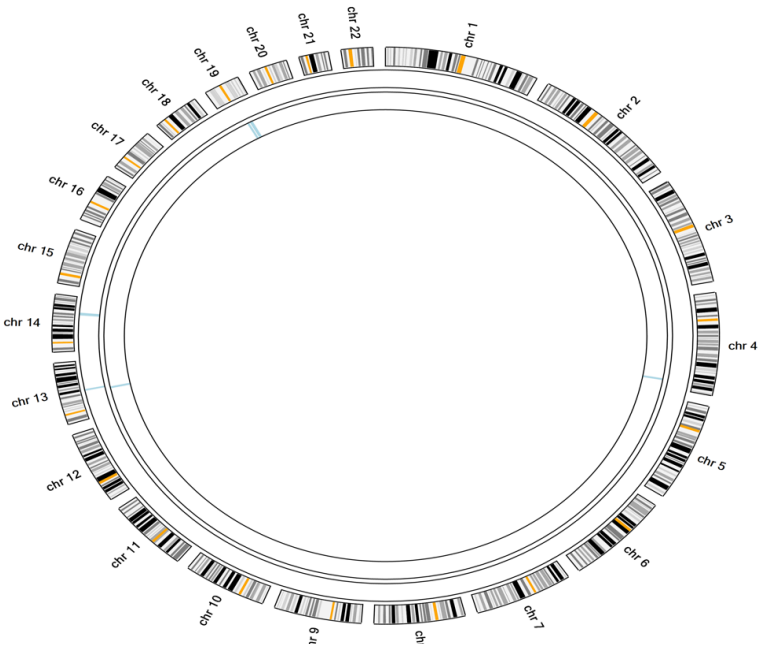


Figure 3-18: Autozygosity mapping in family F11. Visualisation of autozygous regions observed in patients II:2 and II:3 respectively, from the outside to the centre. The shared homozygous regions in each individual are highlighted in blue.

Table 3-18: Homozygous variants shared in common by the sequenced affected members of F11. Variants that are homozygous in both sequenced affected members of family F11 and which have CADD phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
FLG2	Nonsynomymous	1q21.3	exon3:c.G410A:p.G137E	rs6587667	24.1
ZFPM1	Frameshift Deletion	16q24.2	exon10:c.1335_1338del:p.P445fs	.	22.8
FLG2	Nonsynomymous	1q21.3	exon3:c.G2167A:p.E723K	rs16842865	17.82
AGAP3	Frameshift Insertion	7q36.1	exon1:c.91dupG:p.L30fs	.	16.76
FDFT1	Nonframeshift Deletion	8p23.1	exon1:c.193_198del:p.65_66del	rs71711801	16.2
CHRNA3	Nonframeshift Deletion	15q25.1	exon1:c.67_69del:p.23_23del	rs66793222	15.47
DMKN	Nonsynomymous	19q13.12	exon5:c.G839A:p.G280D	rs148799704	14.43
AGAP3	Frameshift Insertion	7q36.1	exon1:c.92_93insG:p.V31fs	.	13.32
AGAP3	Nonsynomymous	7q36.1	exon1:c.T92G:p.V31G	.	13.32
ATXN7	Nonframeshift Insertion	3p14.1	exon2:c.86_87insGCAGCA:p.R29delinsRQQ	.	11.54

3.2.3.3.5 Family F12

For family F12 (Figure 3-4), DNA samples were only available for patients II:2 and III:2. WES analysis was undertaken on patient II:2 (12/1) and filtered using a homozygous recessive model (Figure 3-9). The resulting list was prioritized on CADD Phred scores and table 3-19 shows only the homozygous variants having scores > 10.

Table 3-19: Homozygous variant of sample II:2 member of F12. Details of F12 homozygous variants having CADD phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
GDPD3	Nonsynonymous	16p11.2	exon10:c.G844A:p.E282K	rs148974984	33
CDH7	Nonsynonymous	18q22.1	exon3:c.G505A:p.G169R,	.	28.9
LAMA5	Nonsynonymous	20q13.33	exon32:c.G4037A:p.C1346Y	rs371144423	28.9
PABPC1	Nonsynonymous	8q22.3	exon13:c.A1816G:p.K606E	.	28.1
ZNF366	Splicing	5q13.2	.	.	26.2
PCDHA12	Nonsynonymous	5q31.3	exon1:c.G1528A:p.V510M	rs191376557	26.1
RALGAPA1	Nonsynonymous	14q13.2	exon9:c.T877C:p.C293R	.	25.9
IFT88	Splicing	13q12.11	.	.	24.4
ZNF366	Splicing	5q13.2	.	.	24.3
CFAP46	Nonsynonymous	10q26.3	exon41:c.T5767C:p.W1923R	rs151247898	24.3

PRKRIR	Nonsynonymous	11q13.5	exon5:c.T506A:p.L169Q	.	23.8
USP35	Nonsynonymous	11q14.1	exon6:c.C1088T:p.S363L	rs147270370	23.7
PABPC1	Splicing (+2)	8q22.3	exon14:c.1818+2T>A	.	23.6
RBBP4	Nonsynonymous	1p35.1	exon8:c.C941T:p.S314L	.	23.4
ABI3BP	Splicing (-2)	3q12.2	exon14:c.1160-2A>T	.	23.4
HLTF	Nonsynonymous	3q24	exon8:c.A932G:p.N311S	rs2305868	23.2
TPTE	Nonsynonymous	21p11.1	exon15:c.C811A:p.P271T	.	23.2
ZFPM1	Frameshift Deletion	16q24.2	exon10:c.1335_1338del:p.P445fs	.	23
INF2	Nonsynonymous	14q32.33	exon8:c.C1280T:p.P427L	.	20.2
USP9Y	Nonsynonymous	Yq11.21	exon36:c.A5342T:p.D1781V	.	19.77
ALMS1	Nonframeshift Insertion	2p13.1	exon1:c.35_36insGGA:p.L12delinsLE	.	19.65
ST13	Nonsynonymous	22q13.2	exon1:c.C85A:p.R29S	.	19.32
CHEK2	Nonsynonymous	22q12.1	exon7:c.G799T:p.A267S	.	17.81
TPK1	Nonsynonymous	7q35	exon2:c.C28T:p.P10S	.	17.64
ZNF704	Nonsynonymous	8q21.13	exon4:c.C330A:p.S110R	.	17.03
FDFT1	Nonframeshift Deletion	8p23.1	exon1:c.193_198del:p.65_66del	rs71711801	16.2
CHRNA3	Nonframeshift Deletion	15q25.1	exon1:c.67_69del:p.23_23del	rs66793222	16.27
OBFC1	Nonsynonymous	10q24.33	exon6:c.A572T:p.E191V	.	15.2
THPO	Stopgain	3q27.1	exon2:c.G229T:p.E77X	.	15.16
RFX2	Nonsynonymous	19p13.3	exon5:c.G476T:p.S159I	.	14.84
NPAS3	Nonsynonymous	14q13.1	exon11:c.G2246T:p.S749I	.	14.78
CCT8L2	Nonsynonymous	22q11.1	exon1:c.C1030A:p.P344T	rs41277596	13.62
LAMA5	Nonsynonymous	20q13.33	exon80:c.G10995A:p.M3665I	.	13.41
NFKBIZ	Splicing (+2)	3q12.3	exon11:c.1635+2-&gt	.	10.79
C3AR1	Nonsynonymous	12p13.31	exon2:c.T998C:p.L333P	rs56136109	10.53

3.2.3.3.6 Family F13

For family F13 (Figure 3-4), DNA samples were only available from patients II:2, II:6 and III:13 (13/2, 13/3 and 13/1 respectively). WES analysis was performed on patients II:6 and III:13 (13/3 and 13/1) by Otogenetics, the data was analysed using a shared homozygous recessive mutation model (Figure 3-9) and the resulting list prioritized by CADD Phred scores are shown in table 3-20.

Table 3-20: Homozygous variant shared by sequenced affected members of F13. Details of homozygous variant observed in all members of family F13 and having a CADD phred score > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
FDFT1	Nonframeshift Deletion	14q12	exon1:c.193_198del:p.65_66del	rs71711801	16.2

3.2.3.4 Families of Lebanese origin and Christian Maronite ethnicity

3.2.3.4.1 Family F6

For family F6 (Figure 3-5), the family was analysed using a recessive homozygous mutation model due to the endogamous nature of the population from which the family originates. Samples from affected members III:2, III:7 and III:8 were SNP genotyped and autozygosity mapping was performed (figure 3-19) but no shared autozygous regions were identified.

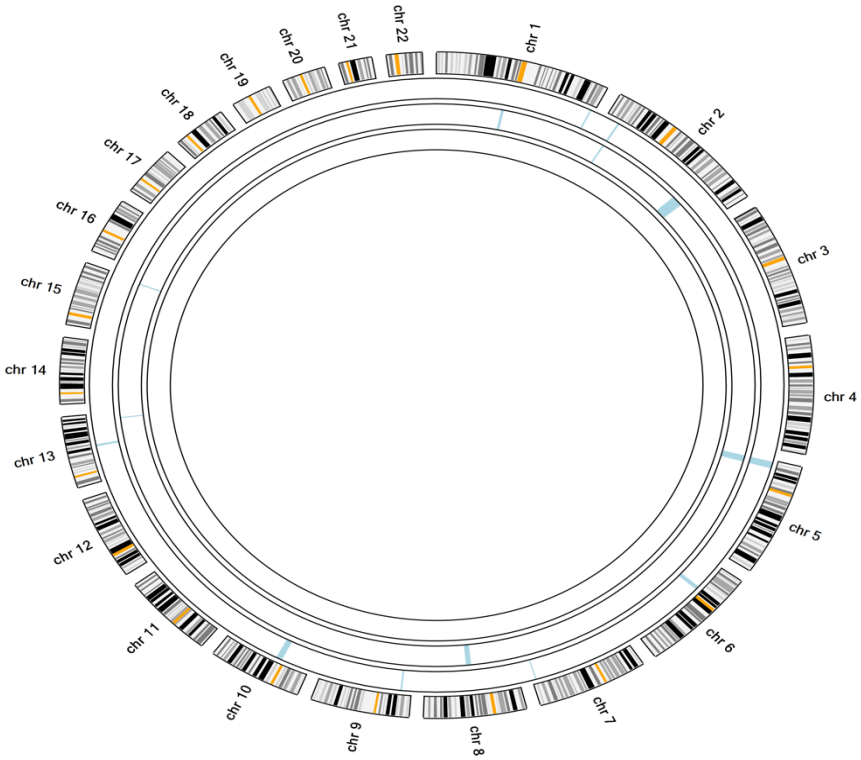


Figure 3-19: F6 autozygosity mapping. Visualisation of autozygous regions shared by genomic DNA from samples from patients III:2, III:7 and III:8 respectively, from the outside to the centre. The homozygous regions are highlighted in blue.

In parallel, WES analysis was performed on III:2, III:7 and III:8 (6/10, 6/1 and 6/3). When this data was analysed with the same homozygous recessive mutation model, two homozygous variants were highlighted (Table 3-21).

Table 3-21: Homozygous variants shared by the sequenced members of F6. Details of homozygous variants shared by both sequenced affected members of family F6 and having CADD Phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
FDFT1	Nonframeshift Deletion	8p23.1	exon1:c.193_198del:p.65_66del	rs71711801	16.2
CHRNA3	Nonframeshift Deletion	15q25.1	exon1:c.67_69del:p.23_23del	rs66793222	16.27

3.2.3.4.2 Family F7

For family F7 (Figure 3-5) WES analysis was performed on samples from patients II:1 and II:3 (7/2 and 7/1) and a filtering strategy for a recessive homozygous mutation was undertaken as detailed in Figure 3-9. This highlighted only two shared homozygous variants (Table 3-22).

Table 3-22: Homozygous variants shared by the sequenced members of F7. Details of homozygous variants shared by both sequenced affected patients in family F7 and which have CADD Phred scores > 10.

Gene	Function	Chr	AA Change	SNP	CADD Phred
FDFT1	Nonframeshift Deletion	8p23.1	exon1:c.193_198del:p.65_66del	rs71711801	16.2
CHRNA3	Nonframeshift Deletion	15q25.1	exon1:c.67_69del:p.23_23del	rs66793222	16.27

3.2.3.4.3 Family F8

For family F8 (Figure 3-5), to check for the presence of a shared recessive homozygous mutation causing the KC phenotype, autozgosity mapping was performed by SNP genotyping samples from patients III:2 and III:4 (8/1 and 8/2). The data shows that these patients do not share any autozygous regions. WES analysis of the same patients using the same model and filtering as described in figure 3-9 highlighted four shared homozygous variants with a CADD Phred scores >10 (Table 3-23).

Table 3-23: List of homozygous recessive variants shared by the affected members of family F8. Details of homozygous variants shared by affected members of family F8 having CADD Phred scores > 10

Gene	Function	Chr	AA Change	SNP	CADD Phred
SCARF2	Nonsynonymous	22q11.21	exon11:c.C1982T:p.P661L	rs9680797	23.6
CHRNA3	Nonframeshift Deletion	15q25.1	exon1:c.67_69del:p.23_23del	rs66793222	16.27
FDFT1	Nonframeshift Deletion	8p23.1	exon1:c.193_198del:p.65_66del	rs71711801	16.2
CFHR1	Stoploss	1q31.3	exon6:c.A992G:p.X331W	rs145057542	11.82

3.2.3.5 Combining the homozygous recessive mutation model data for all families

For all the homozygous variants identified in section 3.2.3, a shared list was generated for the genes implicated in at least two families (Table 3-24). All the shared genes were checked for their association in the literature with terms related to KC using Pubmatrix analysis (Section 2.10) in which all the genes were checked against a list of 10 words associated with KC: Keratoconus, Cornea, Oxidative Stress, Collagen, Eye, Inflammation, Proteinase, Apoptosis, Epithelial cells and Keratocytes.

Table 3-24: List of all genes harbouring homozygous mutations in affected members of at least two KC families. List of genes sharing homozygous variants between the affected family members of KC families and their association in the literature with terms related to KC.

Gene	Location	Families	CADD scores	Associated in the literature with terms related to KC:
CHRNA3	15q25.1	F6, F7, F8, F9, F11, F12	16.27	Epithelial Cells, Inflammation and Apoptosis
FDFT1	8p23.1	F6, F7, F8, F9, F11, F12, F13	16.31	Epithelial Cells, Proteinase, Eye, Apoptosis, Inflammation and collagen
LOC400863	21q22.11	F10, F12	8.965	-
RNF19B	1p35.1	F9, F10	6.34	Apoptosis
ZFPM1	16q24.2	F11, F12	23	Epithelial Cells, Inflammation, Apoptosis, Proteinase and Eye
NOP9	14q12	F9, F13	0.61-0.46	-
AGAP3	7q36.1	F11, F12	16.76-0.002	-
MUC3A	7q22.1	F11, F12, F13	6.38-2.21	Epithelial Cells, Eye, Apoptosis and Inflammation

Out of the 8 genes represented in table 3-24, only 5 genes have been shown by Pubmatrix to be associated to terms related to KC. These genes are the following: Cholinergic Receptor Nicotinic Alpha 3 (*CHRNA3*), Farnesyl-Diphosphate Farnesyltransferase 1 (*FDFT1*), Ring Finger Protein 19B (*RNF19B),* Zinc Finger Protein FOG Family Member 1 (*ZFPM1*) and Mucin 3A Cell Surface Associated (*MUC3A*).

3.2.4 Recessive inheritance: screening for compound heterozygous mutations

In order to investigate a recessive mode of inheritance caused by two different deleterious heterozygous variants in the same gene, the filtering strategy explained in figure 3-20 was applied to the WES data generated for each family. This strategy consisted of filtering out all the variants with a MAF>=2% on 1000g, esp6500si and exac03 databases and removing all the synonymous, non-coding and homozygous variants. From this list, only genes which contained at least two heterozygous variants were highlighted. In addition, similar to the strategy previously used (section 3.2.3), artefacts and ethnicity specific polymorphisms were removed.

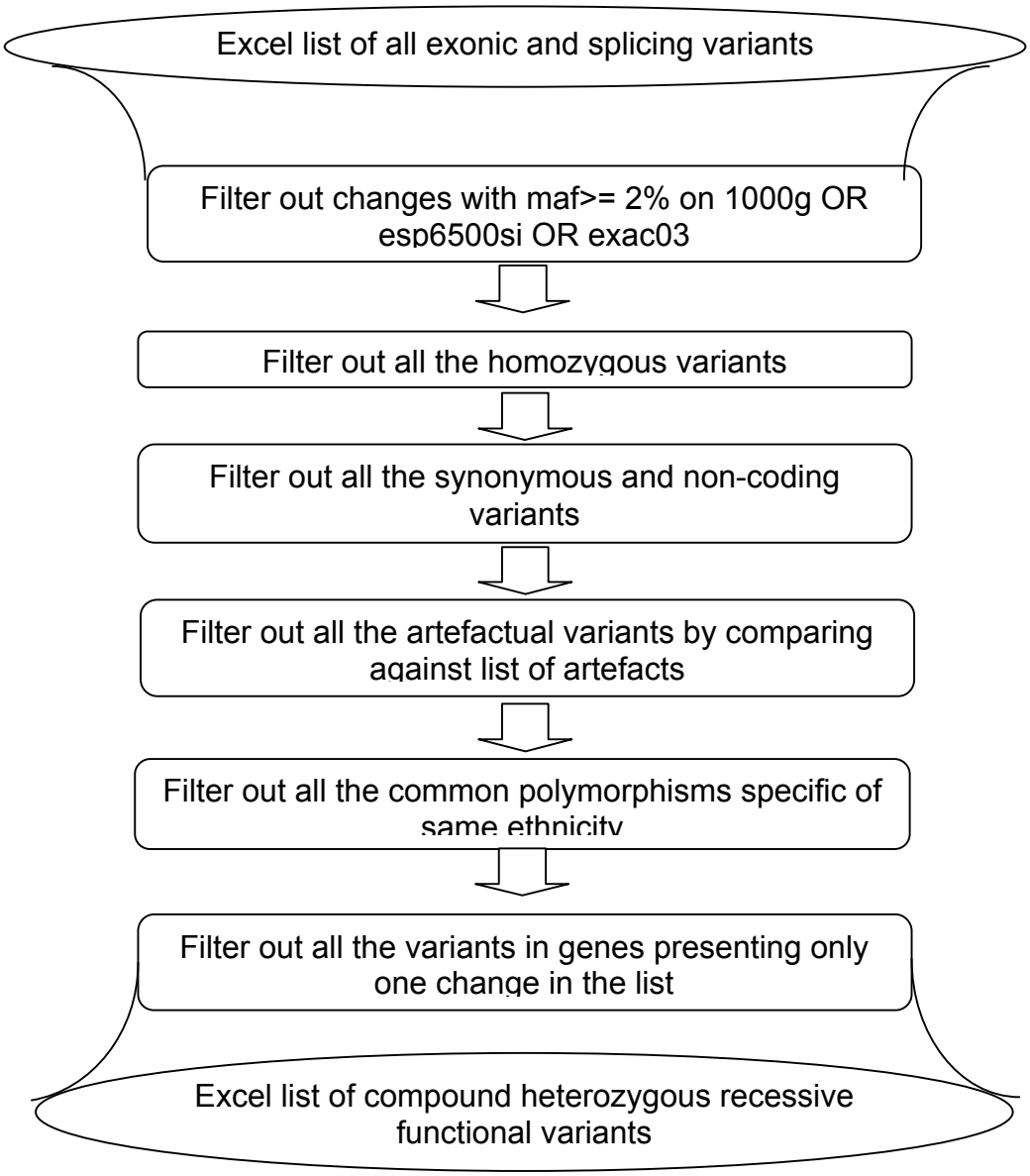


Figure 3-20: Filtering strategy for compound heterozygous recessive hypothesis. The working excel file generated from ANNOVAR undergoes a series of filtering steps by removing any allele with MAF>=2%, common polymorphisms specific to the same ethnicity, the artefactual variants, the synonymous and non-coding variants and all the homozygous variants.

3.2.4.1 Families of South Asian origin

3.2.4.1.1 SS1F family

To search for compound heterozygous mutations in family SS1F (Figure 3-1), the WES generated for individuals V:11 (FL), V:12 (IL), V:2 (RP79), V:9 (YL79), V:10 (AM) and V:18 (ML81) was analysed using this inheritance model (Figure 3-20) and variants prioritized according to their predicted

pathogenicity using CADD Phred scores. The data for each individual was compared to look for shared variants present in all the affected family members using the perl script described in Section 2.12.

This analysis highlighted only one gene on chromosome 12, *GSG1* (Germ Cell Associated 1) (Table 3-25). The first variant is a known SNP (c.C619T, p.R207C, rs376340949) and has a CADD Phred score of 15.19. It is relatively rare, with a frequency of 0.00139776 on the 1000g database and 0.0007578 on ExAc database. The second variant is also a known SNP (c.C514T, p.P172S, rs369468787) but this has a very low CADD Phred score of 6.267 and is present in the databases with a frequency of 0.00419329 on the 1000g database and 0.002125 on ExAc database.

Table 3-25: Compound heterozygous variants shared by the sequenced SS1F KC affected members. Details of the only SS1F compound heterozygous variant pair observed, which was found in GSG1 gene.

Gene	Function	Location	AA Change	SNP	CADD Phred
***GSG1* (Chr12)**	Nonsynonymous	12p13.1	exon4:c.C619T:p.R207C	rs376340949	15.19
	Nonsynonymous	12p13.1	exon4:c.C514T:p.P172S	rs369468787	6.267

3.2.4.1.2 AR1F family

To search for compound heterozygous mutations involved in KC in family AR1F (Figure 3-2), the WES data generated for individuals IV9, IV15, IV16 and V3 was filtered using this inheritance model (Figure 3-20) and variants identified were prioritized according to their predicted pathogenicity using CADD Phred scores. The data for each individual was compared to look for shared variants and/or genes present in all the affected family members using the perl script described in Section 2.12. Table 3-26 shows all the compound heterozygous variants that segregated with KC in family AR1F.

Table 3-26: Compound heterozygous variants shared between AR1F affected members. Details of compound heterozygous variants cosegregating with KC in family AR1F with both variants having at least CADD Phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred
PKD1L1 **(7p12.3)**	Nonsynonymous	exon24:c.C3862T:p.R1288C	rs374546648	24.6
	Nonsynonymous	exon18:c.C2870T:p.S957L	rs141074403	14.44
KLLN **(10q23.31)**	Nonsynonymous	exon1:c.C382G:p.R128G	rs201652303	23.6
	Nonsynonymous	exon1:c.A392G:p.N131S	rs144811392	10.41

3.2.4.1.3 Single cases

To search for compound heterozygous mutations in the single cases, the WES data generated for each was filtereded for genes with two pathogenic variants. Tables 3,27 3-28 and 3-29 display the results of the highlighted genes respectively for SC1, SC2 and SC3 but only for genes where both variants had a CADD Phred score ≥ 20 to reduce the size of the lists. Full lists can be found in Appendix 6.

Table 3-27: Compound heterozygous variants in sample SC1. Details of compound heterozygous variants found in patient SC1 for which both variants have CADD Phred scores ≥ 20.

Gene	Function	AA Change	SNP	CADD Phred
ABI3BP **(3q12.2)**	Nonsynonymous	exon2:c.G167A:p.R56H	rs199998539	34
	Nonsynonymous	exon27:c.G2305C:p.V769L	rs141302553	23.1
AGAP1 **(2q37.2)**	Nonsynonymous	exon12:c.G1390C:p.D464H	.	25.4
	Nonsynonymous	exon12:c.T1402C:p.S468P	.	24.2
C2CD5 **(12p12.1)**	Nonsynonymous	exon8:c.G982T:p.G328W	rs142327845	29.1
	Nonsynonymous	exon8:c.G983T:p.G328V	rs147669786	22.7
CXCR1 **(2q35)**	Nonsynonymous	exon2:c.C1003T:p.R335C	rs16858808	25.4
	Nonsynonymous	exon2:c.A847T:p.I283F	rs142978743	21.8
DMXL1 **(5q23.1)**	Frameshift Deletion	exon18:c.4262delC:p.S1421fs	.	31
	Frameshift Deletion	exon18:c.4256_4260del:p.C1419fs	.	29.7
E4F1 **(16p13.3)**	Nonsynonymous	exon10:c.G1469A:p.R490H	.	26.2
	Nonsynonymous	exon8:c.G1141A:p.V381I	rs145769654	21.8
FAM186B **(12q13.12)**	Nonsynonymous	exon2:c.G321C:p.W107C	rs141974060	31
	Nonsynonymous	exon7:c.G2556C:p.E852D	rs144854212	24.8
FES **(15q26.1)**	Nonsynonymous	exon14:c.A1778G:p.E593G	.	29.7
	Nonsynonymous	exon2:c.G263A:p.R88Q	.	28.1
HCLS1	Nonsynonymous	exon11:c.T1021C:p.Y341H	.	23.5

(3q13.33)	Nonsynonymous	exon11:c.A1015C:p.N339H	.	23.3
MRGPRD	Nonsynonymous	exon1:c.C616T:p.R206W	rs146025023	24.8
(11q13.3)	Nonsynonymous	exon1:c.G641A:p.R214Q	.	23.2
OBSL1	Nonsynonymous	exon5:c.G1850A:p.R617H	rs369705959	25.2
(2q35)	Nonsynonymous	exon11:c.A3730C:p.T1244P	.	21.6
OPLAH	Nonsynonymous	exon13:c.G1750T:p.G584C	rs10107332	27
(8q24.3)	Nonsynonymous	exon10:c.C1331T:p.P444L	.	21.2
OR52N1	Nonsynonymous	exon1:c.G769C:p.A257P	.	26.6
(11p15.4)	Nonsynonymous	exon1:c.C767A:p.P256Q	rs201810252	26.3
RECQL4	Nonsynonymous	exon9:c.C1561T:p.R521W	.	25.6
(8q24.3)	Nonsynonymous	exon10:c.C1625T:p.S542F	rs369852601	24.4
SLC5A4	Nonsynonymous	exon9:c.G1007A:p.R336H	rs62239049	35
(22q12.3)	Stopgain	exon5:c.G415T:p.E139X	rs62239058	31
SPATA20	Nonsynonymous	exon11:c.C1487A:p.P496H	.	33
(17q21.33)	Nonsynonymous	exon11:c.C1486A:p.P496T	.	29.7
TIAM2	Stopgain	exon13:c.T1343A:p.L448X	rs116268446	40
(6q25.3)	Nonsynonymous	exon13:c.G1340C:p.S447T	rs114239511	24.1
TMTC2	Nonsynonymous	exon2:c.C415G:p.R139G	rs200268500	26.8
(12q21.31)	Nonsynonymous	exon6:c.C1730T:p.T577M	rs117020331	21.7
VWA8	Nonsynonymous	exon43:c.A5351C:p.Q1784P	rs200850191	28.9
(13q14.11)	Nonsynonymous	exon43:c.C5350A:p.Q1784K	rs201654359	23
HDGFRP2	Nonsynonymous	exon15:c.G1744C:p.G582R	.	25.8
(19p13.3)	Nonsynonymous	exon15:c.G1750A:p.E584K	.	24.7
	Nonsynonymous	exon15:c.G1745A:p.G582E	.	23.8
	Nonsynonymous	exon15:c.T1739C:p.L580P	.	22
	Nonsynonymous	exon15:c.G1741C:p.A581P	.	20.2
SDK1	Nonsynonymous	exon16:c.T2324C:p.I775T	rs374147324	25.6
(7p22.2)	Nonsynonymous	exon45:c.C6563A:p.T2188K	rs117504728	24.7
	Nonsynonymous	exon21:c.C3080T:p.T1027M	rs61735696	23.4
TTN	Nonsynonymous	exon177:c.G70223A:p.R23408H	.	26.5
(2q31.2)	Nonsynonymous	exon154:c.G49727A:p.R16576H	rs369707906	23.8
	Nonsynonymous	exon154:c.C47839T:p.R15947W	rs368914555	21

Table 3-28: Compound heterozygous variants in sample SC2. Details of compound heterozygous variants found in patient SC2 for which both variants have CADD Phred scores ≥ 20.

Gene	Function	AA Change	SNP	CADD Phred
AGAP1	Nonsynonymous	exon12:c.G1390C:p.D464H	.	25.4
(2q37.2)	Nonsynonymous	exon12:c.T1402C:p.S468P	.	24.2
C17orf96	Nonsynonymous	exon1:c.A326T:p.D109V	.	24.4
(17q12)	Nonsynonymous	exon1:c.G318T:p.E106D	.	20.2
CAPN5	Nonsynonymous	exon2:c.T95C:p.F32S	rs201256547	25.2
(11q13.5)	Nonsynonymous	exon10:c.G1369A:p.V457I	.	22.1
CD24	Frameshift Insertion	exon1:c.68_69insTTTTA:p.Q23fs	.	20.6
(Yq11.222)				

	Nonsynonymous	exon1:c.C65T:p.T22M	.	20.3
DENND2A	Nonsynonymous	exon1:c.C890G:p.P297R	rs374255633	24.1
(7q34)	Nonsynonymous	exon1:c.C54G:p.S18R	rs200642129	20.8
GPR124	Nonsynonymous	exon9:c.C1204T:p.R402W	rs200841231	32
(8p11.23)	Nonsynonymous	exon12:c.C1648T:p.P550S	rs200624144	24.5
KIAA0226	Nonsynonymous	exon13:c.G1954T:p.V652F	rs80094312	28
(3q29)	Nonsynonymous	exon20:c.C2828G:p.A943G	.	24.8
SORBS1	Nonsynonymous	exon28:c.G3292T:p.G1098W	rs372968425	26.9
(10q24.1)	Nonsynonymous	exon28:c.C3095G:p.S1032C	.	24.4
L3MBTL2	Nonsynonymous	exon6:c.G697C:p.A233P	.	32
(22q13.2)	Nonsynonymous	exon8:c.T905G:p.V302G	rs200870155	23.9
MYO1E	Nonsynonymous	exon27:c.C3146A:p.P1049H	rs147579391	31
(15q22.2)	Nonsynonymous	exon7:c.A554G:p.D185G	rs141565214	25
NEB	Nonsynonymous	exon89:c.G13328T:p.S4443I	.	32
(2q23.3)	Nonsynonymous	exon47:c.G6025C:p.D2009H	rs370790324	29.5
	Nonsynonymous	exon143:c.A21317G:p.K7106R	rs375377794	22.9
NFKBIZ	Splicing (+2)	exon11:c.1635+2->TTAATAACC	.	24.9
(3q12.3)	Splicing (+1)	exon11:c.1635+1G>C	.	24.9
	Nonsynonymous	exon10:c.C1931G:p.A644G	.	23.5
P2RY4	Nonsynonymous	exon1:c.G674C:p.R225P	rs202027224	25
(Xq13.1)	Nonsynonymous	exon1:c.G677C:p.R226P	.	23.5
RNF213	Nonsynonymous	exon29:c.G7583A:p.R2528Q	rs202143169	27.8
(17q25.3)	Nonsynonymous	exon57:c.C13913T:p.T4638I	rs141301945	24.6
UNC13D	Nonsynonymous	exon13:c.G1076T:p.R359L	.	25.2
(17q25.1)	Nonsynonymous	exon3:c.G175A:p.A59T	rs9904366	21.6

Table 3-29: Compound heterozygous variants in sample SC3. Details of compound heterozygous variants in patient SC3 for which both variants have CADD Phred scores ≥ 20.

Gene	Function	AA Change	SNP	CADD Phred
COG1	Nonsynonymous	exon1:c.G58C:p.A20P	rs142719529	28.6
(17q25.1)	Nonsynonymous	exon2:c.G547C:p.A183P	rs201454609	25.5
FANCM	Nonsynonymous	exon10:c.G1706A:p.R569H	.	34
(14q21.2)	Nonsynonymous	exon13:c.A2258G:p.D753G	.	23.2
GPAM	Frameshift Insertion	exon15:c.1469_1470insTTTTT:p.C490fs	.	35
(10q25.2)	Nonsynonymous	exon15:c.G1469T:p.C490F	rs199856746	24.4
KCNE1L	Nonsynonymous	exon1:c.G88T:p.G30C	.	24.2
(Xq23)	Nonsynonymous	exon1:c.G87T:p.L29F	rs200894798	22.9
P2RY4	Nonsynonymous	exon11:c.1635+2->TTAATAACC	rs202027224	25
(Xq13.1)	Nonsynonymous	exon11:c.1635+1G>C	.	23.5
SCNN1A	Nonsynonymous	exon10:c.C1931G:p.A644G	rs5742912	29.1
(12p13.31)	Nonsynonymous	exon1:c.G674C:p.R225P	rs201553658	26.9
TMTC2	Nonsynonymous	exon1:c.G677C:p.R226P	rs200268500	26.8

(12q21.31)	Nonsynonymous	exon9:c.T1654C:p.W552R	.	26.4
UTP23	Nonsynonymous	exon3:c.G1018C:p.A340P	.	24.5
(8q24.11)	Nonsynonymous	exon2:c.C415G:p.R139G	.	23.9
NFKBIZ	Splicing (+1)	exon2:c.T410G:p.V137G	.	24.9
(3q12.3)	Splicing (+2)	exon1:c.T65G:p.V22G	.	24.9
	Nonsynonymous	exon1:c.C67G:p.R23G	.	23.5

3.2.4.2 Family KF of Indian origin

To search for compound heterozygous mutations in family KF (Figure 3-3), the WES data generated for affected siblings (IV:1, IV:2 and IV:3) were filtered using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was then compared to look for compound heterozygous variants carried by all affected members using the perl script described in Section 2.12. The resulting list was compared against the heterozygous variants list of the mother (III:4) and the heterozygous variants list of the father (II:3), and the alleles that were both inherited from one parent (ie in the same phase) were excluded. The final list of compound heterozygous variants shared between the siblings is given in Table 3-30.

Table 3-30: Compound heterozygous variants shared by the affected siblings of KF family. Details of compound heterozygous variants found in all three affected members of family KF for which both variants have at least CADD Phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred
ANKRD17	Nonsynonymous	exon13:c.C1841T:p.P614L	.	28.8
(4q13.3)	Nonsynonymous	exon31:c.G6782C:p.G2261A	rs201990436	19.28
ATRN	Nonsynonymous	exon7:c.C802T:p.R268C	rs147164307	24.5
(20p13)	Splicing (-2)	exon16:c.2197-2A>C	rs200054755	24.1
CAND2	Nonsynonymous	exon8:c.C1489T:p.R497W	.	33
(3p25.2)	Nonsynonymous	exon8:c.C2231T:p.T744M	rs370800520	22.5
DNAH11	Nonsynonymous	exon76:c.A12442G:p.I4148V	.	28.3
(7p15.3)	Nonsynonymous	exon66:c.C10768G:p.P3590A	.	19.99
IGFN1	Nonsynonymous	exon14:c.C8981A:p.T2994N	.	25.4
(1q32.1)	Nonsynonymous	exon20:c.C10265G:p.T3422R	rs369378003	23.2
SHROOM3	Nonsynonymous	exon5:c.C2659T:p.R887C	.	34
(4q21.1)	Nonsynonymous	exon4:c.G520A:p.D174N	rs192760386	23.2
TTN	Nonsynonymous	exon24:c.G4184C:p.R1395P	rs72647876	23.9
(2q31.2)	Nonsynonymous	exon3:c.G178T:p.D60Y	rs35683768	23.4
	Nonsynonymous	exon83:c.A21148G:p.R7050G	rs72648982	13.38

3.2.4.3 Families of Lebanese origin and Druze ethnicity

3.2.4.3.1 Family F2

To search for compound heterozygous mutations in family F2 (Figure 3-4), the WES data generated for affected members (II:1, II:2, I:3) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for shared compound heterozygous variants using the perl script described in Section 2.12. The final list of compound heterozygous variants shared by the affected patients is given in Table 3-31

Table 3-31: Compound heterozygous variants shared by the affected siblings of F2 family. Details of compound heterozygous variants found in affected members of family F2 for which both variants have at least CADD Phred scores ≥10.

Gene	Function	AA Change	SNP	CADD Phred	Patients
IGFN1 **(1q32.1)**	Nonsynonymous	exon12:c.G5152A:p.E1718K	rs76484008	19.42	I:3
	Nonsynonymous	exon12:c.G5179A:p.G1727S	rs76698423	22.7	II:1, II:2, I:3
	Frameshift Insertion	exon12:c.G5179A:p.G1727S	.	24.1	II:1, II:2
FAM193A **(4p16.3)**	Nonsynonymous	exon11:c.C1271A:p.P424Q	rs34435530	23.5	II:1, II:2, I:3
	Nonsynonymous	exon6:c.C452A:p.T151N	rs145745319	23.3	II:1, II:2, I:3
SEMA6A **(5q23.1)**	Nonsynonymous	exon2:c.C14G:p.A5G	rs114860630	18.08	II:1, II:2, I:3
	Nonsynonymous	exon19:c.G2383A:p.G795S	rs374035190	22.6	II:1, II:2, I:3

3.2.4.3.2 Family F9

To search for compound heterozygous mutations in sample (III:8) of family F9 (Figure 3-4), the WES data generated for sample (III:8) was analysed using this inheritance model (Figure 3-20) and checked for genes with two pathogenic variants. The variants were prioritized according to their predicted pathogenicity using CADD Phred scores. The final list of compound heterozygous variants is given in Table 3-32.

Table 3-32: Compound Heterozygous variants in sample III:8 from family F9. Details of compound heterozygous variants found in individual III:8 from family F9 for which both variants have at least CADD Phred scores ≥ 20.

Gene	Function	AA Change	SNP	CADD Phred
CUL9	Nonsynonymous	exon12:c.C2882T:p.A961V	rs149800737	24.6
(6p21.1)	Nonsynonymous	exon12:c.G2881T:p.A961S	.	23.9
DYNC2H1	Stopgain	exon26:c.C4018T:p.Q1340X	.	40
(11q22.3)	Nonsynonymous	exon89:c.G12865C:p.G4289R	rs144717489	22.7
HEXDC	Nonsynonymous	exon12:c.G1454A:p.C485Y	rs140743740	24.6
(17q25.3)	Nonsynonymous	exon12:c.C1526A:p.A509E	rs150076334	22.7
LTBP1	Nonsynonymous	exon20:c.T2692A:p.S898T	.	25.9
(2p22.3)	Nonsynonymous	exon21:c.G2885A:p.R962Q	rs141080282	24.2
NEB	Nonsynonymous	exon104:c.T15023C:p.L5008P	rs111517514	28.3
(2q23.3)	Nonsynonymous	exon71:c.G10383C:p.M3461I	rs149025191	20.3
NOD2	Nonsynonymous	exon3:c.C1540T:p.R514W	.	24.9
(16q12.1)	Nonsynonymous	exon3:c.C1969T:p.R657W	rs5743276	23.2
PCNT	Nonsynonymous	exon39:c.T8924C:p.L2975P	rs35881595	26.1
(21q22.3)	Nonsynonymous	exon31:c.C6986G:p.P2329R	rs35848602	23.3
POLR3B	Nonsynonymous	exon13:c.C940T:p.L314F	.	28.4
(12q23.3)	Nonsynonymous	exon13:c.A933T:p.L311F	.	25.1
TAS2R31	Nonsynonymous	exon1:c.G843T:p.W281C	rs139069360	24
(12p13.2)	Nonsynonymous	exon1:c.A711T:p.L237F	rs116926686	21
TTN	Nonsynonymous	exon154:c.G55303A:p.E18435K	rs375422359	22.9
(2q31.2)	Nonsynonymous	exon184:c.G72901A:p.V24301I	rs55675869	22.2
IGSF10	Stopgain	exon4:c.G4804T:p.E1602X	rs79363433	36
(3q25.1)	Stopgain	exon2:c.G220T:p.E74X	.	35
	Nonsynonymous	exon2:c.A1507T:p.R503W	rs3732775	25

3.2.4.3.3 Family F10

To search for compound heterozygous mutations in family F10 (Figure 3-4), the WES data generated for affected members (III:4 and III:5) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for compound heterozygous variants carried by all affected family members using the perl script described in Section 2.12. The final list of compound heterozygous variants shared between the affected patients is given in Table 3-33.

Table 3-33: Compound heterozygous variants shared by the affected siblings in F10 family. Details of compound heterozygous variants cosegregating with KC in affected members of family F10 for which both variants have at least CADD Phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred	Patients
FSIP2 **(2q32.1)**	Nonsynonymous	exon16:c.C4583G:p.T1528R	rs192957612	24.4	III:4, III:5
	Nonsynonymous	exon16:c.C6559G:p.L2187V	.	21.1	III:4, III:5
KLHL8 **(4q22.1)**	Nonsynonymous	exon8:c.G1028A:p.R343Q	rs144074504	29.1	III:4, III:5
	Nonsynonymous	exon7:c.G982A:p.V328I	.	34	III:4, III:5
HK3 **(5q35.2)**	Nonsynonymous	exon13:c.G1800C:p.Q600H	rs61749650	16.27	III:4, III:5
	Nonsynonymous	exon12:c.A1733G:p.Q578R	rs61749651	22.3	III:4, III:5
BCLAF1 **(6q23.3)**	Splicing (+2)	exon11:c.2397+2T>A	.	24.8	III:4
	Nonsynonymous	exon1:c.G250A:p.A84T	rs150210620	25.4	III:5
	Nonsynonymous	exon5:c.C1208G:p.T403R	rs141489829	23	III:4, III:5
	Nonsynonymous	exon4:c.G320A:p.R107H	rs200868536	23.8	III:4, III:5
TNS3 **(7p12.3)**	Nonsynonymous	exon20:c.G2648A:p.R883H	rs200980382	12.19	III:4, III:5
	Nonsynonymous	exon17:c.C1363T:p.R455C	rs148721285	26.4	III:4, III:5
ROBO3 **(11q24.2)**	Nonsynonymous	exon5:c.G850A:p.D284N	rs142090631	29.1	III:4, III:5
	Nonsynonymous	exon21:c.C3150A:p.Y1050X	.	35	III:4, III:5
ABCC4 **(13q32.1)**	Nonsynonymous	exon26:c.C3284T:p.T1095M	rs11568644	26.6	III:4, III:5
	Nonsynonymous	exon1:c.C52A:p.L18I	rs11568681	12.46	III:4, III:5
ADAMTS18 **(16q23.1)**	Nonsynonymous	exon20:c.C3157T:p.R1053W	rs148703569	28.1	III:4, III:5
	Nonsynonymous	exon18:c.A2752G:p.K918E	rs138435590	24.4	III:4, III:5
SSC5D **(19q13.42)**	Nonsynonymous	exon6:c.C615A:p.H205Q	.	15.42	III:4, III:5
	Nonsynonymous	exon6:c.G851A:p.R284Q	.	25.8	III:4, III:5

3.2.4.3.4 Family F11

To search for compound heterozygous mutations in family F11 (Figure 3-4), the WES data generated for affected members (II:2 and II:3) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for shared compound heterozygous variants using the perl script described in Section 2.12. The final list of compound heterozygous variants shared by the affected patients is given in Table 3-34.

Table 3-34: Compound heterozygous variants shared by the affected siblings of F11 family. Details of compound heterozygous variants carried by both affected members of family F11 for which both variants have at least CADD Phred scores ≥ 20.

Gene	Function	AA Change	SNP	CADD Phred
FLG	Nonsynonymous	exon3:c.C5671T:p.R1891W	rs36006086	24.2
(1q21.3)	Nonsynonymous	exon3:c.C4309T:p.R1437C	rs12750571	22.3
XIRP2	Nonsynonymous	exon7:c.T748C:p.S250P	rs143400009	27.3
(2q24.3)	Nonsynonymous	exon7:c.G9464A:p.G3155E	rs16853333	23.2
NFKBIZ	Nonsynonymous	exon3:c.T857C:p.I286T	.	23.5
(3q12.3)	Splicing (+1)	exon2:c.C712T:p.P238S	.	24.9
OR8D4	Nonsynonymous	exon10:c.C1931G:p.A644G	rs79430449	23
(11q24.1)	Frameshift Insertion	exon11:c.1635+1G>C	rs201238608	29.2
CYP1A1	Nonsynonymous	exon1:c.G716A:p.S239N	rs4987133	26.2
(15q24.1)	Nonsynonymous	exon1:c.304dupT:p.L101fs	rs61747605	26.2
USH2A	Nonsynonymous	exon35:c.A6713C:p.E2238A	rs41277212	27.1
(1q41)	Nonsynonymous	exon13:c.G2384A:p.C795Y	.	26.5
	Nonsynonymous	exon8:c.G1434C:p.E478D	rs35730265	23.8

3.2.4.3.5 Family F12

To search for compound heterozygous mutations in sample (II:2) of family F12 (Figure 3-4), the WES data generated for sample (II:2) was analysed using this inheritance model (Figure 3-20) and checked for genes with two pathogenic variants. The variants were prioritized according to their predicted pathogenicity using CADD Phred scores. The final list of compound heterozygous variants is given in Table 3-35.

Table 3-35: Compound Heterozygous variants in sample II:2 of family F12. Details of compound heterozygous variants found in patient II:2 of family F12 for which both variants have at least CADD phred scores ≥ 20.

Gene	Function	AA Change	SNP	CADD Phred
ADRA2B	Nonsynonymous	exon1:c.T1226C:p.F409S	.	27.4
(2q11.1)	Nonsynonymous	exon1:c.C475T:p.R159C	.	23.5
AHNAK	Nonsynonymous	exon5:c.C16168A:p.P5390T	rs77640901	23.2
(11q12.3)	Nonsynonymous	exon5:c.G3292T:p.V1098F	rs79639272	22.9
ANO2	Nonsynonymous	exon22:c.C2371T:p.R791W	.	26.5
(12p13.31)	Nonsynonymous	exon19:c.A1973C:p.Y658S	.	26.2
ATP4A	Nonsynonymous	exon21:c.G3077T:p.G1026V	.	33

(19q13.12)	Nonsynonymous	exon21:c.G3076A:p.G1026R	.	28.3
CEP192	Nonsynonymous	exon20:c.C4180T:p.L1394F	rs61740922	27.7
(18p11.21)	Nonsynonymous	exon10:c.G1375A:p.E459K	rs116020861	21.8
GPRC5C	Nonsynonymous	exon1:c.C40T:p.R14W	.	23.5
(17q25.1)	Nonsynonymous	exon4:c.G1333A:p.V445I	rs139362972	23.4
LRP1B	Nonsynonymous	exon89:c.C13438T:p.P4480S	rs75758215	25.1
(2q22.1)	Nonsynonymous	exon70:c.G10844C:p.G3615A	rs76554185	24.2
MPHOSPH10	Nonsynonymous	exon10:c.G1784A:p.R595Q	rs115361918	33
(2p13.3)	Nonsynonymous	exon3:c.G826T:p.D276Y	.	29.9
NFKBIZ	Splicing (+1)	exon11:c.1635+1G>C	.	24.9
(3q12.3)	Nonsynonymous	exon10:c.C1931G:p.A644G	.	23.5
PGPEP1	Stopgain	exon2:c.C85T:p.Q29X	.	37
(19p13.11)	Nonsynonymous	exon2:c.A56G:p.H19R	.	23.6
PLEKHG2	Nonsynonymous	exon18:c.G2308A:p.A770T	rs111487768	33
(19q13.2)	Nonsynonymous	exon13:c.C1358G:p.P453R	rs34975345	26.4
SYNE1	Nonsynonymous	exon123:c.A22303G:p.S7435G	rs35763277	27.9
(6q25.2)	Nonsynonymous	exon66:c.G10619A:p.R3540H	rs145911138	23.6

3.2.4.3.6 Family 13

To search for compound heterozygous mutations in family F11 (Figure 3-4), the WES data generated for affected members (II:6 and III:13) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for shared compound heterozygous variants using the perl script described in Section 2.12. The final list of compound heterozygous variants shared by the affected patients is given in Table 3-36.

Table 3-36: Compound Heterozygous variants shared by the affected siblings of F13 family. Details of compound heterozygous variants affected individuals II:6 and III:13 for which both variants have at least CADD phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred	Patients
CEP192	Nonsynonymous	exon20:c.C4180T:p.L1394F	rs61740922	27.7	III:13, II:6
(18p11.21)	Nonsynonymous	exon10:c.G1375A:p.E459K	rs116020861	21.8	III:13, II:6
DNAH14	Nonsynonymous	exon55:c.C8410T:p.R2804C	rs140066130	34	III:13, II:6
(1q42.12)	Nonsynonymous	exon46:c.G7022A:p.G2341E	rs41303311	28.2	III:13, II:6
	Nonsynonymous	exon68:c.A10597G:p.M3533V	rs74147200	19.74	III:13, II:6
OBSCN	Nonsynonymous	exon94:c.G20746A:p.E6916K	rs200269110	24.5	III:13
	Nonsynonymous	exon34:c.T9185G:p.V3062G	rs202022683	23.4	III:13

(1q42.13)	Nonsynonymous	exon94:c.C21488T:p.P7163L	rs72762066	22	II:6
	Nonsynonymous	exon22:c.G6373T:p.A2125S	rs75280352	14.82	III:13, II:6
TTN	Nonsynonymous	exon87:c.C22204T:p.R7402C	rs72648987	23.1	III:13, II:6
(2q31.2)	Nonsynonymous	exon102:c.A26080T:p.T8694S	rs72650006	19	III:13, II:6

3.2.4.4 Families of Lebanese origin and Christian Maronite ethnicity

3.2.4.4.1 Family F6

To search for compound heterozygous mutations in family F6 (Figure 3-5), the WES data generated for affected members (III:2, III:7 and III:8) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for shared compound heterozygous variants/genes using the perl script described in Section 2.12. The final list of compound heterozygous variants shared between the affected patients are described in Table 3-37.

Table 3-37: Compound heterozygous variants shared by the affected siblings of F6 family. Details of compound heterozygous variants found in all affected members of family F6 for which both variants have at least CADD Phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred	Patients
ALDH4A1	Nonsynonymous	exon1:c.C47T:p.P16L	rs146450609	17.72	III:7, III:8, III:2
(1p36.13)	Nonsynonymous	exon11:c.T982C:p.F328L	rs41273175	22.1	III:7, III:8, III:2
MUC4	Nonsynonymous	exon2:c.C9358T:p.P3120S	rs77902873	11.66	III:2
(3q29)	Nonsynonymous	exon2:c.C4625T:p.P1542L	rs202029925	13.77	III:7
	Nonsynonymous	exon2:c.G12184A:p.A4062T	rs370995233	14.23	III:7, III:8, III:2
	Nonsynonymous	exon2:c.C5038T:p.P1680S	rs7374593	10.15	III:8

3.2.4.4.2 Family F7

To search for compound heterozygous mutations in family F7 (Figure 3-5), the WES data generated for affected members (II:1 and II:3) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for shared compound heterozygous variants using the perl script described in Section 2.12. The final list of compound

heterozygous variants shared between the affected patients are described in Table 3-38.

Table 3-38: Compound Heterozygous variants shared by the affected siblings of F7 family. Details of compound heterozygous variants found in both affected members of family F7 for which both variants having at least CADD Phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred
MUC16	Nonsynonymous	exon2:c.G9415C:p.V3139L	rs114676657	22.9
(19p13.2)	Nonsynonymous	exon2:c.G12616A:p.A4206T	rs139868374	23.4
MUC4	Nonsynonymous	exon36:c.G38821A:p.E12941K	.	10.62
(3q29)	Nonsynonymous	exon12:c.G36581A:p.R12194Q	rs58500707	10.9

3.2.4.4.3 Family F8

To search for compound heterozygous mutations in family F7 (Figure 3-5), the WES data generated for affected members (III:2 and III:4) were analysed using this inheritance model (Figure 3-20) and the variants prioritized according to their predicted pathogenicity using CADD Phred scores. The data for the affected individuals was compared to look for shared compound heterozygous variants using the perl script described in Section 2.12. The final list of compound heterozygous variants shared by the affected patients is given in Table 3-39.

Table 3-39: Compound Heterozygous variants shared between the affected siblings of F8 family. Details of compound heterozygous variants found in both affected members of family F8 for which both variants have at least CADD Phred scores ≥ 10.

Gene	Function	AA Change	SNP	CADD Phred
SSUH2	Nonsynonymous	exon11:c.G962C:p.R321P	rs149593163	23.4
(3p26.1)	Nonsynonymous	exon4:c.G244A:p.A82T	rs112366230	27.5
SHROOM3	Nonsynonymous	exon5:c.A1691C:p.E564A	rs76656494	17.44
(4q21.1)	Nonsynonymous	exon5:c.C3035A:p.T1012N	rs147732653	13.36
PCDHGB1	Nonsynonymous	exon1:c.G557T:p.S186I	rs62378414	10.98
(5q31.3)	Nonsynonymous	exon1:c.A1195G:p.K399E	rs77250251	18.12
ANKK1	Nonsynonymous	exon2:c.G365A:p.R122H	rs35877321	11.25
(11q23.2)	Nonsynonymous	exon5:c.G766A:p.E256K	.	23
DNAH17	Nonsynonymous	exon2:c.G12073A:p.A4025T	rs146968926	11.85
(17q25.3)	Nonsynonymous	exon2:c.C7727G:p.P2576R	rs61736352	22.8
LAMA5	Nonsynonymous	exon76:c.G12335A:p.R4112Q	rs147290767	23.8
(20q13.33)	Nonsynonymous	exon26:c.G3880A:p.E1294K	rs143551466	25.4
TTN	Nonsynonymous	exon53:c.G7052A:p.R2351Q	rs66961115	20.8

(2q31.2)	Nonsynonymous	exon28:c.G3467A:p.R1156Q	rs67254537	18
	Nonsynonymous	exon186:c.C78593T:p.A26198V	rs56324595	23.6
	Nonsynonymous	exon186:c.G78592T:p.A26198S	rs199506676	22.3
	Nonsynonymous	exon186:c.G78574A:p.E26192K	.	15.3
	Nonsynonymous	exon151:c.G41629A:p.E13877K	rs72955213	18.73
	Nonsynonymous	exon150:c.C31066A:p.R10356S	rs33917087	21.9

3.2.5 Dominant inheritance: heterozygous mutations

Although the majority of the families collected for this study had a pedigree structure consistent with recessive inheritance, they are also consistent with autosomal dominant Mendelian inheritance with reduced penetrance. Indeed, in a few families a vertical inheritance is seen with an affected parent having an affected child (for example families F12, F13 and AR1F). In order to investigate a possible dominant mode of inheritance, the filtering strategy explained in figure 3-21 was applied to the WES data generated from each family. As in the previous models (section 3.2.3), all the variants with a MAF≥2% on 1000g, esp6500si and exac03 databases were removed, all the synonymous, non-coding and homozygous variants were filtered and the ethnically matched variants polymorphisms and artefacts were removed. Where data was available for multiple affected family members, variants lists were compared and those not shared by all affected family members were removed. The data from unaffected family members could not be used to filter the lists because of the possibility of reduced penetrance.

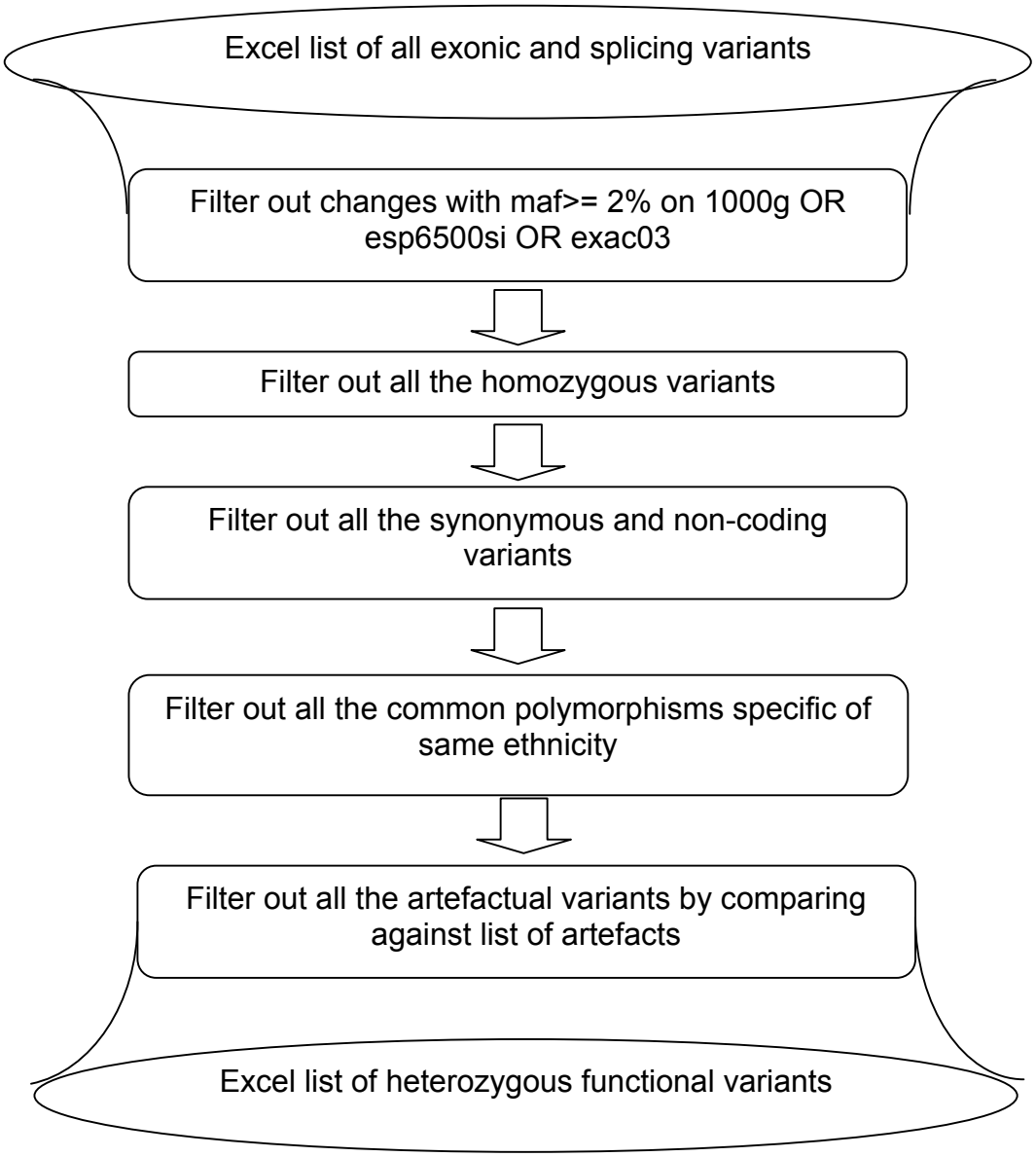


Figure 3-21: Filtering strategy for heterozygous dominant hypothesis

3.2.5.1 Families of South Asian origin

3.2.5.1.1 SS1F family

The results of the shared heterozygous variants present in all six sequenced members of SS1F are listed in Table 3-40. Three genes are highlighted: *MUC3A* (Mucin 3A, Cell Surface Associated), *OR9G9* (Olfactory Receptor, Family 9, Subfamily G, Member 9) and *MAP2K3* (Mitogen-Activated Protein Kinase Kinase

3). Analysis of the highlighted genes in the ExAc databases shows that null alleles in *MUC3A* are very common in the heterozygous state. The CADD phred scores cut off varied from one family to the other in this section, only to keep the size of the lists displayed in this chapter concise. The choice of the cut off was made in order to keep the minimal number of entries in the tables using the lowest CADD phred score cut-off.

Table 3-40: Heterozygous variants shared between the affected siblings of family SS1F. This table details the heterozygous variants found in all affected members of family SSF1 which have CADD Phred scores ≥ 10.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
MUC3A	Stopgain	7q22.1	exon2:c.C3319T:p.Q1107X	rs79874934	28
MUC3A	Splicing (+1)	7q22.1	exon7:c.3934+1A>C	rs73714276	20.6
OR9G9	Nonsynonymous	11q12.1	exon1:c.C505T:p.R169C	rs11228733	22.9
MAP2K3	Nonsynonymous	17p11.2	exon12:c.G928A:p.V310M	rs2363198	26.1

3.2.5.1.2 AR1F family

The shared heterozygous variants found in all five sequenced affected members of family AR1F and which have a CADD Phred score ≥ 25 are listed in Table 3-41.

Table 3-41: Heterozygous variants shared between the affected members of family AR1F. This table details the list of heterozygous variants found in all affected members of family AR1F which have CADD Phred scores ≥ 25.

Gene	Function	Chr	AA Change	SNP	CADD Phred
CLYBL	Stopgain	13q32.3	exon6:c.C775T:p.R259X	rs41281112	35
DOCK5	Nonsynonymous	8p21.2	exon30:c.G3056A:p.R1019H	rs148483229	34
FERMT3	Nonsynonymous	11q13.1	exon12:c.G1393A:p.E465K	rs373999037	30
SLC9A9	Nonsynonymous	3q24	exon1:c.T65A:p.V22E	.	28.7
ITGB6	Nonsynonymous	2q24.2	exon8:c.G1027A:p.V343M	rs61737764	28.2
ULK4	Nonsynonymous	3p22.1	exon33:c.C3379T:p.R1127W	.	27.9
PASK	Nonsynonymous	2q37.3	exon10:c.T2165C:p.L722P	rs201982321	25.3

3.2.5.1.3 Single Cases

Tables 3-42, 3-43 and 3-44 show the heterozygous variants found in patients SC1, SC2 and SC3 which have CADD scores ≥ 35.

Table 3-42: Heterozygous variants in sample SC1. This table lists heterozygous variants found in patient SC1 and having CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
LOR	Stopgain	1q21.3	exon2:c.G592T:p.G198X	.	40
TIAM2	Stopgain	6q25.3	exon13:c.T1343A:p.L448X	rs116268446	40
C15orf65	Stopgain	15q21.3	exon2:c.C263A:p.S88X	rs201241267	39
CHIT1	Stopgain	1q32.1	exon9:c.G1016A:p.W339X	rs201320385	37
ZNF790	Stopgain	19q13.12	exon5:c.C1483T:p.R495X	.	36
ZKSCAN7	Frameshift Deletion	3p21.31	exon6:c.2129_2130del:p.Q710fs	.	35
DMXL1	Frameshift Deletion	5q23.1	exon18:c.4262delC:p.S1421fs	.	35
TMEM222	Nonsynonymous	1p36.11	exon6:c.C622T:p.R208W	rs150461998	35
SCN10A	Nonsynonymous	3p22.2	exon24:c.G4085A:p.R1362Q	rs369399424	35
C3orf67	Nonsynonymous	3p14.2	exon16:c.G1609A:p.E537K	rs139831294	35
CSRP3	Nonsynonymous	11p15.1	exon7:c.G538T:p.G180C	.	35
ITGB7	Nonsynonymous	12q13.13	exon15:c.C2263T:p.R755C	.	35
PLEKHH1	Nonsynonymous	14q24.1	exon13:c.C1852T:p.R618W	.	35
ZNF544	Stopgain	19q13.43	exon7:c.C1843T:p.R615X	rs79258645	35
TGM2	Stopgain	20q11.23	exon9:c.G1303T:p.E435X	.	35
PLA2G3	Stopgain	22q12.2	exon1:c.C436T:p.R146X	rs201767666	35
SLC5A4	Nonsynonymous	22q12.3	exon9:c.G1007A:p.R336H	rs62239049	35

Table 3-43: Heterozygous variants in sample SC2. This table details the heterozygous variants found in patient SC2 which have CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
C15orf65	Stopgain	15q21.3	exon2:c.C263A:p.S88X	rs201241267	39
MAP3K10	Stopgain	19q13.2	exon4:c.G1072T:p.E358X	.	39
TRIM45	Frameshift Deletion	1p13.1	exon10:c.C817T:p.R273C	.	35
GPAM	Frameshift Insertion	10q25.2	exon5:c.1467_1468del:p.V489fs	.	35
CLCN6	Nonsynonymous	1p36.22	exon13:c.G1337A:p.R446Q	rs368875103	35
BBS7	Nonsynonymous	4q27	exon15:c.1468_1469insTTTTTTT:p.C490fs	rs372685495	35
RNF10	Nonsynonymous	12q24.31	exon5:c.C646T:p.R216C	.	35
LAMA1	Nonsynonymous	18p11.31	exon50:c.C7141T:p.R2381C	rs142063208	35

KIAA1683	Stopgain	19p13.11	exon3:c.C3004T:p.Q1002X	rs200745939	35

Table 3-44: Heterozygous variants in sample SC3. This table details heterozygous variants found in patient SC3 and having CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
GCLC	Stopgain	6p12.1	exon10:c.C1139A:p.S380X	.	44
GABRP	Stopgain	5q35.1	exon10:c.C1186T:p.R396X	.	41
C15orf65	Stopgain	15q21.3	exon2:c.C263A:p.S88X	rs201241267	39
DCDC2C	Stopgain	2p25.3	exon5:c.C581G:p.S194X	rs199703744	37
TMEM175	Stopgain	4p16.3	exon6:c.C271T:p.Q91X	.	35
BAIAP2L2	Nonsynonymous	22q13.1	exon12:c.C1273T:p.R425W	rs371417992	35
MRPS18A	Nonsynonymous	6p21.1	exon3:c.G221A:p.R74H	rs377393438	35
COBL	Nonsynonymous	7p12.1	exon2:c.G59C:p.R20P	rs200229807	35
ANXA11	Nonsynonymous	10q22.3	exon14:c.C1354T:p.R452W	rs150671497	35
GPAM	Frameshift Insertion	10q25.2	exon15:c.1469_1470insTTTTT:p.C490fs	.	35
DDX31	Nonsynonymous	9q34.13	exon7:c.C904T:p.R302C	rs374539267	35
MPP4	Nonsynonymous	2q33.1	exon19:c.G1421A:p.R474H	rs190854787	35

3.2.5.2 Family KF of Indian origin

The shared heterozygous variants with CADD Phred scores ≥ 35 present in all three sequenced affected siblings of family KF are listed in Table 3-45.

Table 3-45: Heterozygous variants shared by the affected siblings of family KF. This table details heterozygous variants found in all three affected individuals from family KF and having CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	CADD Phred
WDR90	Stopgain	16p13.3	exon35:c.G4385A:p.W1462X	.	43
EPYC	Stopgain	12q21.33	exon5:c.C562T:p.R188X	rs143768297	37
RTN4	Nonsynonymous	2p16.1	exon7:c.G587A:p.R196H	rs202245854	35
MAGI1	Nonsynonymous	3p14.1	exon22:c.C3592T:p.R1198C	rs201117026	35
FGFBP2	Stopgain	4p15.32	exon1:c.C664T:p.R222X	.	35
KDM7A	Nonsynonymous	7q34	exon3:c.C379T:p.R127C	.	35
LCTL	Nonsynonymous	15q22.31	exon6:c.G227A:p.R76H	.	35
HOXB9	Nonsynonymous	17q21.32	exon2:c.G548A:p.R183H	rs201792149	35
FBXO17	Nonsynonymous	19q13.2	exon6:c.C730T:p.R244C	rs140379005	35

3.2.5.3 Families of Lebanese origin and Druze ethnicity

3.2.5.3.1 Family F2

The shared heterozygous variants with CADD Phred scores ≥ 35 present in all sequenced affected siblings of family F2 are listed in Table 3-46.

Table 3-46: Heterozygous variants shared between the affected siblings of F2 family. This table details heterozygous variants shared by all affected individuals in family F2 and having CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	CADD Phred
ANKEF1	Stopgain	20p12.2	exon9:c.C2045A:p.S682X	rs375202162	36
BCAT1	Nonsynonymous	12p12.1	exon4:c.G341A:p.R114Q	.	35
C14orf105	Stopgain	14q22.3	exon5:c.C544T:p.Q182X	rs34960436	35

3.2.3.3.2 Family F9

The heterozygous variants with CADD Phred scores ≥ 35 present in III:8 from family F9 are listed in Table 3-47.

Table 3-47: Heterozygous variants in Family F9. This table details heterozygous variants with CADD Phred scores ≥ 35 found in individual III:8 from family F9.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
HINT3	Stopgain	6q22.32	exon2:c.A289T:p.R97X	rs201833844	40
DYNC2H1	Stopgain	11q22.3	exon26:c.C4018T:p.Q1340X	.	40
SMAD9	Stopgain	13q13.3	exon4:c.T693A:p.Y231X	.	38
KBTBD13	Stopgain	15q22.31	exon1:c.C1309T:p.Q437X	.	38
FAM71A	Stopgain	1q32.3	exon1:c.A1663T:p.K555X	rs143861665	36
GCKR	Stopgain	2p23.3	exon9:c.C679T:p.R227X	rs149847328	36
IGSF10	Stopgain	3q25.1	exon4:c.G4804T:p.E1602X	rs79363433	36
PCDHGA10	Stopgain	5q31.3	exon1:c.C2188T:p.Q730X	rs62378454	36
OR8K3	Stopgain	11q12.1	exon1:c.C778T:p.Q260X	rs117366703	36
WDR87	Stopgain	19q13.13	exon6:c.G6397T:p.E2133X	.	36
PCK2	Frameshift Insertion	14q11.2	exon10:c.1468dupA:p.P489fs	.	35
NFASC	Nonsynonymous	1q32.1	exon11:c.C1378T:p.R460W	rs184631101	35
CDK18	Nonsynonymous	1q32.1	exon3:c.G260A:p.R87H	rs368734289	35
IGSF10	Stopgain	3q25.1	exon2:c.G220T:p.E74X	.	35
ROPN1L	Stopgain	5p15.2	exon2:c.T135A:p.Y45X	rs41280363	35
OR1J4	Stopgain	9q33.2	exon1:c.C221G:p.S74X	.	35
POLL	Nonsynonymous	10q24.32	exon7:c.G1126A:p.G376S	rs200623399	35
CCDC67	Stopgain	11q21	exon5:c.C364T:p.R122X	.	35
SERPINA10	Stopgain	14q32.13	exon2:c.C262T:p.R88X	rs2232698	35

TRAP1	Nonsynonymous	16p13.3	exon3:c.G224A:p.R75H	rs61758086	35

3.2.3.3.3 Family F10

The heterozygous variants with CADD Phred scores ≥ 35 shared between III:4 and III:5 from family F10 (10/1 and 10/2) are listed in Table 3-48.

Table 3-48: Heterozygous variants shared between the affected siblings of family F10. This table details heterozygous variants found in both affected members of family F10 and having CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
AIM1	Stopgain	6q21	exon2:c.C2812T:p.R938X	rs372779090	37
LAMC3	Frameshift Deletion	9q34.12	exon3:c.695delC:p.T232fs	.	35
ROBO3	Stopgain	11q24.2	exon21:c.C3150A:p.Y1050X	.	35

3.2.3.3.4 Family 11

The heterozygous variants with CADD Phred scores ≥ 35 shared between II:2 and II:3 from family F11 are listed in Table 3-49.

Table 3-49: Heterozygous variants shared by affected siblings of family F11. This table details heterozygous variants found in both affected members of family F11 and having Cadd phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
IRAK3	Stopgain	12q14.3	exon7:c.C616T:p.R206X	rs367600113	39
NT5C3B	Stopgain	17q21.2	exon7:c.C505T:p.Q169X	.	39
SGK494	Stopgain	17q11.2	exon10:c.C840G:p.Y280X	rs185508082	38
COL12A1	Frameshift Insertion	6q13	exon50:c.5585_5586insCC: p.P1862fs	.	36
TAAR2	Stopgain	6q23.2	exon1:c.C604T:p.R202X	rs147900465	36
ARHGEF10L	Nonsynonymous	1p36.13	exon27:c.C3577T:p.R1193C	.	35
MGAT4D	Nonsynonymous	4q31.1	exon10:c.C1024T:p.R342X	rs62346874	35
DKK1	Nonsynonymous	10q21.1	exon2:c.G359T:p.R120L	rs149268042	35
RASSF8	Nonsynonymous	12p12.1	exon5:c.C1183T:p.R395W	rs370277680	35
SDCBP2	Nonsynonymous	20p13	exon3:c.G317A:p.R106Q	rs35367003	35

3.2.3.3.5 Family F12

The heterozygous variants with CADD Phred scores ≥ 35 present in III:2 from family F12 are listed in Table 3-50.

Table 3-50: Heterozygous variants in Family F12. This table details heterozygous variants in individual III:2 from family F12 with Cadd phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
EML6	Stopgain	2p16.1	exon5:c.G531A:p.W177X	.	39
ARL6IP4	Stopgain	12q24.31	exon1:c.G472T:p.E158X	.	37
PGPEP1	Stopgain	19p13.11	exon2:c.C85T:p.Q29X	.	37
CEP290	Frameshift Deletion	12q21.32	exon54:c.7394_7395del:p.E2465fs	.	36
AFF1	Stopgain	4q21.3	exon3:c.C1031G:p.S344X	.	36
PCK2	Frameshift Insertion	14q11.2	exon10:c.1468dupA:p.P489fs	.	35
ANGEL1	Frameshift Deletion	14q24.3	exon5:c.1091delT:p.V364fs	.	35
NFASC	Nonsynonymous	1q32.1	exon11:c.C1378T:p.R460W	rs184631101	35
ZNF445	Stopgain	3p21.31	exon3:c.C58T:p.R20X	.	35
MAP3K5	Nonsynonymous	6q23.3	exon17:c.G2300A:p.R767H	rs114243010	35
LRGUK	Nonsynonymous	7q33	exon10:c.C1196T:p.P399L	rs61749957	35
NYNRIN	Stopgain	14q12	exon4:c.C1810T:p.Q604X	.	35
PYGL	Nonsynonymous	14q22.1	exon5:c.C625T:p.R209C	.	35
DNAH17	Nonsynonymous	17q25.3	exon66:c.G10675A:p.E3559K	.	35
GCAT	Nonsynonymous	22q13.1	exon9:c.C1237T:p.R413W	rs17856459	35

3.2.3.3.6 Family F13

The heterozygous variants with CADD Phred scores ≥ 35 shared between II:6 and III:13 from family F13 are listed in Table 3-51.

Table 3-51: Heterozygous variants shared between the affected siblings of family F13. This table details heterozygous variants in both affected members of family F13 having CADD Phred scores ≥ 35.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
BRCA2	Stopgain	13q13.1	exon27:c.A9976T:p.K3326X	rs11571833	38
OTOR	Frameshift Insertion	20p12.1	exon3:c.300dupC:p.F100fs	.	35
FLNB	Nonsynonymous	3p14.3	exon14:c.C2095T:p.R699W	rs200554477	35
ARID1B	Nonsynonymous	6q25.3	exon7:c.C2405T:p.S802L	rs150140314	35
P2RX5	Nonsynonymous	17p13.2	exon8:c.C856T:p.R286C	rs146166049	35
STARD6	Stopgain	18q21.2	exon1:c.C55T:p.R19X	rs17292725	35
CC2D1A	Nonsynonymous	19p13.12	exon4:c.C314T:p.A105V	rs192358667	35

3.2.5.4 Families of Lebanese origin and Christian Maronite ethnicity

3.2.5.4.1 Family F6

The heterozygous variants with CADD Phred scores ≥ 30 shared between III:2, III:7 and III:8 affected members from family F6 are listed in Table 3-52.

Table 3-52: Heterozygous variants shared between the affected siblings of family F6. This table details heterozygous variants carried by all three affected members of family F6 having CADD Phred scores ≥ 30.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
BMP8B	Nonsynonymous	1p34.2	exon7:c.G1166A:p.R389H	rs41267039	35
PKD2L1	Nonsynonymous	10q24.31	exon8:c.C1273T:p.R425C	rs147426900	33

3.2.5.4.2 Family F7

The heterozygous variants with CADD Phred scores ≥ 30 shared between II:1 and II:3 affected members from family F7 are listed in Table 3-53.

Table 3-53: Heterozygous variants shared between the affected siblings of F7 family. This table details heterozygous variants carried by both affected members of family F7 having CADD Phred scores ≥ 30.

Gene	Function	Chr	AA Change	SNP	Cadd Phred
LRP4	Nonsynonymous	11p11.2	exon10:c.C1117T:p.R373W	rs118009068	33
N4BP3	Nonsynonymous	5q35.3	exon4:c.G950A:p.R317H	rs115186199	32
PPID	Nonsynonymous	4q32.1	exon2:c.C145T:p.R49C	rs2070631	31

3.2.5.4.3 Family F8

The heterozygous variants with CADD Phred scores ≥ 35 shared between III:2 and III:4 affected members from family F8 are listed in Table 3-54.

Table 3-54: Heterozygous variants shared between the affected siblings of family F8. This table details heterozygous variants carried by both affected members of family F8 having CADD Phred scores ≥ 35.

Gene	Chr	Chr	AA Change	SNP	Cadd Phred
TRAF3IP3	Stopgain	1q32.2	exon16:c.C1525T:p.R509X	.	47
FILIP1L	Stopgain	3q12.1	exon2:c.C2542T:p.Q848X	rs202015243	39
BCLAF1	Stopgain	6q23.3	exon5:c.G1201T:p.E401X	rs61731960	39
KRTAP2-4	Stopgain	17q21.2	exon1:c.G282A:p.W94X	rs200049107	38

GSG1	Stopgain	12p13.1	exon4:c.C409T:p.Q137	rs141509089	37
GPNMB	Stopgain	7p15.3	exon11:c.G1699T:p.E567X	rs11537976	36
DNAH14	Frameshift Deletion	1q42.12	exon21:c.3392delA:p.E1131fs	.	35
SLC9A2	Frameshift Deletion	2q12.1	exon7:c.1553delG:p.C518fs	.	35
USP53	Frameshift Deletion	4q26	exon18:c.2947_2950del:p.K983fs	.	35
PPID	Frameshift Deletion	4q32.1	exon4:c.437_438del:p.H146fs	rs368268182	35
KIAA1614	Nonsynonymous	1q25.3	exon2:c.C901T:p.R301C	rs79485039	35
PYGM	Nonsynonymous	11q13.1	exon18:c.G2183A:p.R728H	rs139230055	35

3.2.5.5 Combining the potential compound heterozygous and dominant mutations data for all families

A large number of genes were highlighted in this study with compound heterozygous variants (section 3.2.4) and dominant mutations (section 3.2.5) in each family. Clearly this data set is too big to take forward, so to prioritise candidates, the lists were analysed to look for genes highlighted in more than one family. A shorter version of this list is shown in Table 3-55. This shortened table only shows the genes that were highlighted in two or more families belonging to the different ethnicities studied in this project (South Asian, Lebanese Druze and Lebanese Maronites). In addition, the genes were prioritised by checking for associations in the literature with terms related to KC using pubmatrix: Keratoconus, Cornea, Oxidative Stress, Collagen, Eye, Inflammation, Proteinase, Apoptosis, Epithelial cells and Keratocytes. Finally, the highlighted genes listed in table 3-55 contained variants with CADD Phred scores ≥ 10 and presenting at least 1 hit of association with terms related to KC in the literature. The full list is shown in appendix 8.

Table 3-55: List of genes containing heterozygous variants in multiple KC families. This table presents a list of genes in which heterozygous variants were observed in at least two families out of 15 belonging to at least two different ethnicities, along with their association of terms related to KC in the literature.

Gene	Families	Associated in the literature with terms related to KC:	Total number of hits in the literature
ATM	SC1, F9, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress and Inflammation.	3587
SACS	SC2, KF, F9, F11, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress, Keratocytes and Inflammation.	1179
DST	SC3, KF, F10, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Keratoconus, Oxidative Stress and Inflammation.	987
PRX	SC2, SC3, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress, Keratocytes and Inflammation.	571
MUC4	F6, F7, F9	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress and Inflammation.	441
FES	SC1, SC2, SC3, F11, F13	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress and Inflammation.	302
MUC16	SC1, F7, F9	Proteinase, Epithelial Cells, Collagen, Apoptosis, Eye, Cornea, Oxidative Stress and Inflammation.	255
RPGR	SC1, SC2, F9	Eye, Epithelial Cells, Apoptosis, Cornea and Inflammation.	254
MUC6	SC1, SC2, SC3, F9	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress and Inflammation.	236
NEB	SC1, SC2, KF, F9, F10, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Eye, Oxidative Stress and Inflammation.	206
DSPP	SC1, F9, F10, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Eye, Oxidative Stress and Inflammation.	188
FLG	SC1, SC2, F10, F11	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Keratoconus, Keratocytes, Oxidative Stress and Inflammation.	169
COL6A3	SC2, F9	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye, Oxidative Stress and Inflammation.	129
NCOR2	SC1, SC2, F11	Proteinase, Epithelial Cells, Apoptosis, Eye, Oxidative Stress and Inflammation.	126
TTN	SC1, SC2, SC3, KF, F8, F9, F11, F12, F13	Proteinase, Epithelial Cells, Apoptosis, Eye, Oxidative Stress and Inflammation.	124
PSMB10	SC1, SC2, SC3, F11, F12	Proteinase, Epithelial Cells, Apoptosis, Cornea, Eye, Oxidative Stress and Inflammation.	100
NFKBIZ	SC2, SC3, F11, F12,	Proteinase, Epithelial Cells, Collagen, Apoptosis, Cornea, Eye and Inflammation.	76

LTBP1	SC1, SC3, F9, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Keratocytes, Eye and Inflammation.	72
USH2A	KF, F11, F13	Proteinase, Epithelial Cells, Collagen and Eye.	70
FKRP	F6, F8, F10, F11, F12	Proteinase, Epithelial Cells, Collagen, Apoptosis, Eye and Inflammation	69
ZNF469	AR1F, F2, F6, F11, F12	Keratoconus, Eye, Cornea and Collagen.	59
BCLAF1	F7, F8, F9, F10	Apoptosis, Epithelial Cells, Eye, Inflammation and Proteinase.	42
HCLS1	SC1, F12	Apoptosis, Epithelial Cells, Collagen, Eye, Inflammation and Proteinase.	42
AHNAK	SC1, SC2, F12	Apoptosis, Epithelial Cells, Collagen, Eye, Oxidative Stress, Inflammation and Proteinase.	38
LAMC1	SC1, SC3, F9, F11	Apoptosis, Epithelial Cells, Collagen, Eye, Oxidative Stress, Inflammation and Proteinase.	37
MAP1A	SC3, F9, F12	Apoptosis, Epithelial Cells, Collagen, Eye, Oxidative Stress, Cornea and Proteinase.	31
GPAM	SC1, SC2, SC3, F11, F12, F13	Apoptosis, Epithelial Cells, Oxidative Stress, Inflammation and Proteinase.	26
NFASC	SC2, F9, F12	Epithelial Cells, Collagen, Eye, Inflammation and Proteinase.	26
MUC17	SC1, F2, F9	Epithelial Cells, Collagen, Eye, Inflammation and Apoptosis.	25
OVGP1	SC1, SC2, F9, F11, F12	Apoptosis, Epithelial Cells, Oxidative Stress, Inflammation and Proteinase	25
SYNE1	KF, SC3, F2, F9, F12	Apoptosis, Epithelial Cells, Cornea, Collagen, Eye, Inflammation and Proteinase	23
LAMA5	SC1, SC2, KF, F8, F9	Epithelial Cells, Collagen, Proteinase, Apoptosis and Inflammation.	21
SHROOM3	KF, F8, F9, F12	Epithelial Cells, Eye, Collagen and Inflammation	19
CROCC	F8, F9, F11	Epithelial Cells and Eye	13
NOM1	SC1, SC2, F9	Epithelial Cells, Collagen, Proteinase and Apoptosis	10
ABCD4	SC1, SC3, F10	Epithelial Cells and Apoptosis	8
FLNB	SC1, SC2, SC3, F13	Apoptosis, Epithelial Cells, Collagen, Eye and Proteinase	8
CELA1	KF, F9	Apoptosis, Epithelial Cells, Inflammation and Proteinase	7
SVEP1	SC1, SC2, SC3, F9, F12	Apoptosis, Epithelial Cells, Inflammation and Proteinase	6
OBSCN	SC2, F6, F8, F11, F13	Epithelial Cells and Apoptosis	5
XPO4	SC2, KF, F9, F11, F12	Epithelial Cells	4
EPPK1	SC1, SC2, SC3, F11, F12	Eye and Inflammation	2
UBXN11	SC1, SC2, F8, F9,	Epithelial Cells and Proteinase	2

	F11, F12		
AHNAK2	SC1, KF, F6, F9, F10	Inflammation	1
DNAH1	AR1F, SC1, F11, F12, F13	Collagen	1
DNAJB7	KF, SC3, F12, F13	Inflammation	1
IGFN1	SC3, KF, F2, F9	Inflammation	1
KNDC1	SC1, F12	Epithelial Cells	1
P2RY4	SC1, SC2, SC3, F11, F12, F13	Epithelial Cells	1
SPATA20	SC1, SC2, F11, F12	Epithelial Cells	1

This analysis highlighted nine (p.E679K, p.D1458E, p.Q1623R, p.P2477L, p.P2781L, p.V2803I, p.R3414S, p.R3426Q, p.Q3858H) rare functional ZNF469 variants (MAF < 2%) which were detected in 5 families AR1F, F2, F6, F11 and F12. Mutations in *ZNF469* have been reported to contribute to the pathogenesis of KC (Section 1.4.9.3) (Rohrbach *et al.*, 2013, Sahebjada *et al.*, 2013, De Baere, 2014, Lechner *et al.*, 2014, Vincent *et al.*, 2014, Davidson *et al.*, 2015). Thus PCR primers were designed (Appendix 9), and all identified variants , presented in Table 3-56, were Sanger sequenced in the available family members to check that they segregated with the phenotype in each family. However, none of these families segregated the *ZNF469* mutations in a perfect manner.

Table 3-56: *ZNF469* segregation results. This table shows the variants revealed in *ZNF469* after Sanger sequencing and the families in which they were found.

Mutation	M1	M2	M3	M4	M5	M6	M7	M8	M9
AR1F	E679K	V2803I	-	-	-	-	-	-	-
F2	-	-	-	-	-	-	R3414S	Q1623R	P2477L
F6	-	-	-	D1458E	Q3858H	-	-	-	-
F11	-	-	-	-	-	R3426Q	-	-	-
F12	-	-	P2781L	-	-	-	-	-	-

3.2.6 Enrichment of potentially pathogenic alleles analysis

The families analysed in this study are from communities with a strong tradition of consanguinity and endogamy. While this led to the original hypothesis that homozygous recessive alleles were responsible for the KC phenotype in the pedigrees, it is possible that the consanguineous and endogamous practices in these communities have led to the enrichment of multiple pathogenic alleles. Thus, instead of a single Mendelian allele causing the KC in these families, oligogenic inheritance may be responsible. In this part of the study, the WES data generated was analysed to look for enrichment of multiple pathogenic alleles in this cohort.

For this analysis, the WES data from one affected member of each KC family from this study (Indian, Pakistani, Druze, Maronite) (Section 3.2.2), together with WES data from nine unrelated KC families belonging to the Institute of Ophthalmology/University College London (IoO/UCL) KC cohort were used (Middle Eastern, Caucasians, Asians and Iranians) through collaboration with Professor Alison Hardcastle (IoO/UCL). To undertake the analysis of all these large datasets, the PLINK/SEQ programme was used (section 2.13). Ethnically matched control data was added to the dataset in the ratio of 1:3. The controls used were derived from other genetics projects on-going in the Leeds laboratory, or were specifically sampled and sequenced for this study. While they were selected as individuals not known to have any eye defect, they were not specifically examined to exclude KC. The controls collected were sex matched, and aged above 25 years old to make sure that they had passed the age when they are most prone to develop KC. In total, WES from 72 ethnically and ancestry matched controls were added (Table 3-57). Unfortunately, there was a shortage of Middle Eastern, Druze and Maronite Lebanese controls so additional Caucasian controls downloaded from the 1000 genome project were added for the Lebanese ethnicities and Pakistani controls were added to the Middle Eastern ethnicity group in order to achieve the 1:3 ratio.

Table 3-57: PLINK/SEQ analysis samples. List of KC patients and controls used in the PLINK/SEQ analysis and their ethnicities

Cohort	Patients	Ethnicity	Number of controls
Leeds	5	Pakistani	15
	6	Druze, Lebanese	8 Druze + 10 Caucasians
	3	Maronites, Lebanese	4 Maronites + 5 Caucasians
	1	Indian	3
London	2	Caucasian	6
	4	Middle Eastern	9 Middle Eastern + 3 Pakistanis
	1	Iranian	3
	1	Pakistani	3
	1	Indian	3

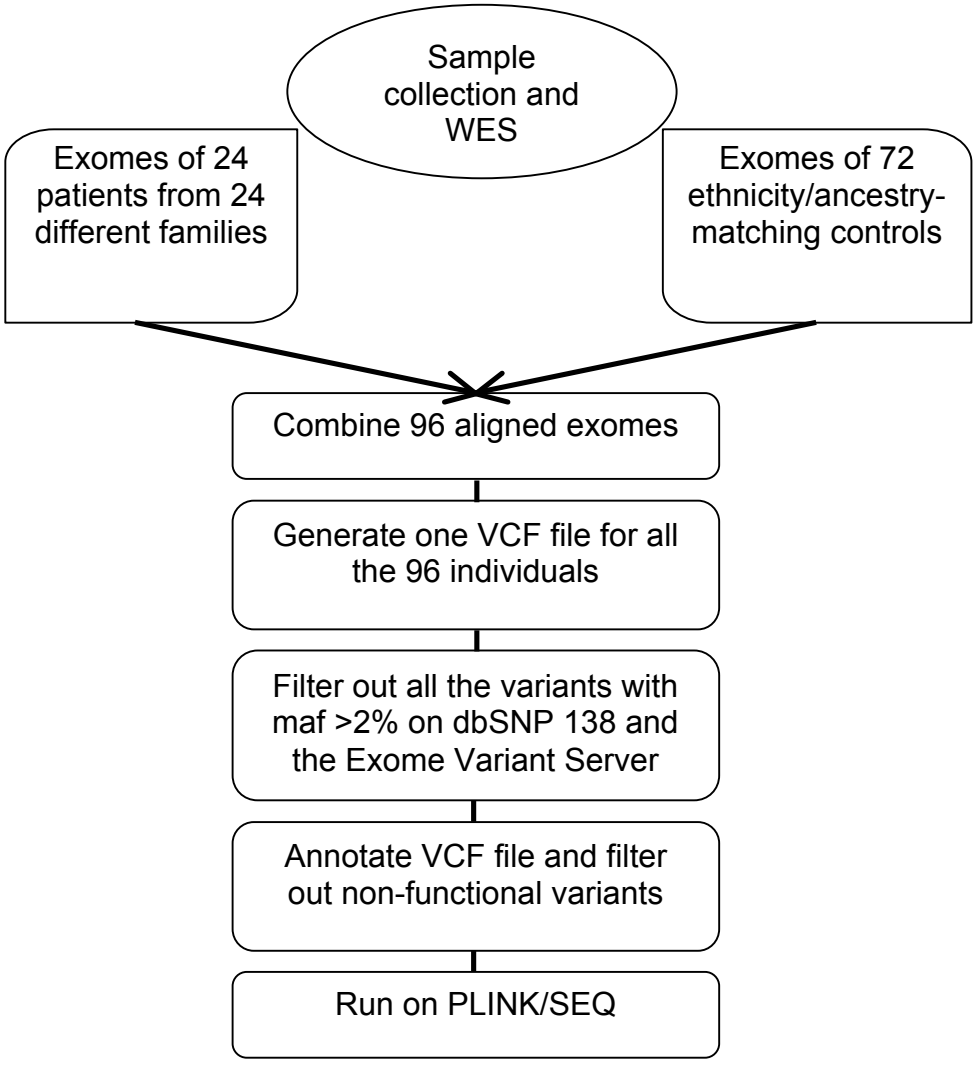


Figure 3-22: PLINK/SEQ results generation. The pipeline described in this figure represents step by step the stages of generating the PLINK/SEQ results, starting from sample collection and WES, to combining the results, annotating the variants and running PLINK/SEQ software.

The strategy used for PLINK/SEQ analysis is described in Figure 3-22. Briefly, the raw sequence data (the fastq files) of the controls and IoO/UCL cases were subjected to the NGS workflow described in Section 3.2.2 previously used to analyse the Leeds cohort. Once aligned to hg19, all of the 96 aligned sequence (all cases and controls) files were combined together and called for variants using GATK (section 2.12). The resulting VCF files were processed using *annotateSnps.pl* and *FilterOnEvsMaf.pl* developed by Dr. David Parry (University of Leeds) (https://github.com/gantzgraf/vcfhacks) to remove all the variants with MAF above 2% in dbSNP138 and Exome Variant Server respectively. Subsequently, the filtered VCF files were translated using the perl script *translateVCF.pl* (developed by Dr. David Parry) and all the synonymous, non-coding, intronic and intergenic variants were filtered out using *filter_by_anno.pl* developed by Dr. James Poulter (University of Leeds). Finally, these filtered VCF files were used as the input files for PLINK/SEQ (section 2.13). Potential enrichments were examined at two levels using PLINK/SEQ: enrichment of variants in a given coding sequence of a gene, and enrichment of single variants.

3.2.6.1 Enrichment of variants in a particular gene

Gene enrichment analysis of gene-based rare-variants (MAF <2%) were identified by running PLINK/SEQ using the combined VCF files. This analysis compared the burden of non-reference genotypes (the variations compared to the reference genome) in cases to controls within a gene. The results of this gene enrichment analysis, where variants in reported genes appeared enriched in cases while no variations were detected in controls (DESC n/0) where DESC is number of case/control minor alleles, are shown in Table 3-58. The full list of enrichments can be found in appendix 10.

Table 3-58: Gene enrichment results. This table represents the genes that have been detected by PLINK/SEQ to carry variants (NVAR) that are totally absent in controls compared to KC samples. This test has been conducted using the Burden statistical test (Lee *et al.*, 2014) and generating respective p-values ($p < 0.05$).

Position (hg19 coordinates)	Gene	NVAR[1]	P-values[2]	I-values[3]	DESC[4]
chr1:220236272-220240666	*BPNT1*	2	0.0056680	0.014175	4/0
chr21:35742799-35742947	*KCNE2*	2	0.00562023	0.012851	3/0
chr3:10146243-10146243	*C3orf24*	1	0.00910865	0.0143229	3/0
chr3:10331519-10331815	*GHRL*	2	0.00507063	0.0112319	3/0
chr7:6438337-6438337	*RAC1*	1	0.006941	0.0143849	3/0
chr9:36651842-36651842	*MELK*	1	0.00619469	0.0146082	3/0

[1]The number of variants in the corresponding gene.

[2]The empirical significance, based on permutations.

[3]The measure of the discreteness of the empirical distribution of the test statistic

[4]The number of case/control minor alleles.

3.2.6.2 Single site enrichment - Specific allele enrichment.

Single variant enrichment in cases compared to controls was also investigated using PLINK/SEQ. This analysis generated a list of specific variants with their respective values, most importantly p-values and Odds ratios. This list was sorted according to p-values and only variants with p-values less than 0.05 were retained. This data was not corrected for multiple testing since this analysis has been used to check for enrichment only but not for association analysis. These constituted 1233 variants summarised in table 3-59.

Since this list of variants is un-annotated with respect to pathogenic prediction status in the PLINK/SEQ output, the variants were annotated using the *variant-effect-predictor.pl* perl script following the command line described in Appendix 4.3. This programme determines the SIFT and Polyphen predictions (section 2.12), in addition to determining the MAF of known variants by using the 1000 genome project data. The list of variants enriched in KC cases with a p-value < 0.05 are listed in table 3-60. The full list can be found in appendix 10.

Table 3-59: General PLINK/SEQ statistics after annotation of the variants. This table represents how many variants were taken through to PLINK/SEQ (Input reads), how many variants were processed for annotation (Variants processed) and the percentage of the novel variants and the existing variants among the overall variants.

Line of input reads	1233
Variants processed	1227
Novel Variants	460 (37.5%)
Existing variants	767 (62.5%)

Table 3-60: List of enriched variants in KC. This list represents the variants enriched in KC patients compared to controls.

Position (hg19)	Ref	Alt	Gene	Function	P-value	OR[1]
chr15:20740399	T	C	*GOLGA6L6*	Nonsynonymous	0.00004567	69.16
chr4:38828828	G	A	*TLR6*	Nonsynonymous	0.000155248	46.3827
chr15:33192236	G	T	*FMN1*	Nonsynonymous	0.000251257	8.48571
chr6:56044835	C	A	*COL21A1*	Nonsynonymous	0.000409741	37.3951
chr3:195514825	G	A	*MUC4*	Nonsynonymous	0.000839195	40.0968
chr14:21502431	G	A	*RNASE13*	Nonsynonymous	0.000884237	21.45
chr3:37370562	A	G	*GOLGA4*	Nonsynonymous	0.00269321	32.1111
chr3:43389767	G	T	*SNRK*	Nonsynonymous	0.00286217	5.69189
chr1:12907400	G	A	*HNRNPCL1*	Nonsynonymous	0.002878	10.65
chr21:38568265	C	A	*TTC3*	Stopgain	0.002878	10.65
chr5:54591272	G	T	*DHX29*	Nonsynonymous	0.00310225	30.6
chrX:117959307	G	A	*ZCCHC12*	Nonsynonymous	0.00310225	30.6
chr20:60904915	C	T	*LAMA5*	Nonsynonymous	0.00323748	30.17650
chr11:119181832	C	G	*MCAM*	Nonsynonymous	0.00342681	17.439
chr12:3129932	C	T	*TEAD4*	Nonsynonymous	0.00342681	17.439
chr17:79414763	C	A	*BAHCC1/RP11_1055B8.7*	Nonsynonymous	0.0046348	26.7882
chr2:152404248	A	G	*NEB*	Nonsynonymous	0.00485881	26.3647
chr1:201180388	G	A	*IGFN1*	Nonsynonymous	0.00495237	15.7317
chr3:195393086	G	A	*SDHAP2*	ncRNA_splicing	0.0052305	35.2545
chr5:74450118	T	C	*ANKRD31*	Nonsynonymous	0.00534894	25.5176
chr19:17731515	G	T	*UNC13A*	Nonsynonymous	0.00534894	25.5176
chr17:41352520	A	C	*NBR1*	Nonsynonymous	0.00552985	15.2439
chr1:206566903	G	A	*SRGAP2*	Nonsynonymous	0.00605368	5.17895
chr4:170476887	G	T	*NEK1*	Nonsynonymous	0.00618972	14.7561
chr2:179595427	A	C	*TTN*	Nonsynonymous	0.00652948	23.82349
chr16:23718102	T	G	*ERN2*	Nonsynonymous	0.00672439	4.77193
chr19:36002386	C	T	*DMKN*	Nonsynonymous	0.00772493	5.30303
chr7:100807151	G	C	*VGF*	Nonsynonymous	0.00869543	9.09091

[1]Odds Ratio

3.2.6.3 Pathway analysis of PLINK/SEQ data

The two levels of potential enrichment (gene and single variant) examined by PLINK/SEQ generated a list of genes for which variants have been found to be enriched in cases compared to controls along with a list of variants that show enrichment with KC in the cohort of familial cases studied. Table 3-61 displays the association of the genes mentioned in both lists (Table 3-58 and 3-60) with terms that are related to KC (Keratoconus, Cornea, Oxidative Stress, Collagen, Eye, Inflammation, Proteinase, Apoptosis, Epithelial cells and Keratocytes) generated using Pubmatrix (Section 2.10), and thus highlights some genes that might be functionally significant in KC pathogenesis.

Table 3-61: Pubmatrix results. This table presents the association of the genes highlighted in this study with terms related to KC in the literature.

Gene	Associated in the literature with terms related to KC:	Total number of hits in the literature
RAC1	Epithelial Cells, Poteinase, Inflammation, Apoptosis, Keratocytes, Collagen and Eye	3,516
MUC4	Cornea, Eye, Epithelial Cells, Inflammation, Proteinase and Apoptosis	437
TLR6	Cornea, Eye, Epithelial Cells, Inflammation, Proteinase and Apoptosis	370
ERN2	Apoptosis, Oxidative Stress, Epithelial Cells, Inflammation Collagen and Eye	304
MCAM	Epithelial Cells, Inflammation, Apoptosis, Proteinase, Collagen, Eye and Cornea	215
NEB	Epithelial Cells, Collagen, Apoptosis, Inflammation and Eye	199
TTN	Proteinase, Epithelial Cells, Collagen, Apoptosis, Oxidative Stress and Eye	120
MELK	Apoptosis, Epithelial Cells, Inflammation and Eye.	68
VGF	Epithelial Cells, Proteinase, Inflammation, Eye and Cornea	64
KCNE2	Epithelial Cells and Eye	56
TEAD4	Epithelial Cells, Apoptosis, Collagen, Eye and Cornea	43
RP11_1055B8.7	Eye, Epithelial Cells, Collagen and cornea	42
NBR1	Epithelial Cells, Oxidative Stress and Inflammation	41
GOLGA4	Epithelial Cells and Eye	33
GHRL	Inflammation and Epithelial Cells	25
NEK1	Epithelial Cells and Eye	20
LAMA5	Epithelial Cells and Collagen	19

DMKN	Epithelial Cells	10
UNC13A	Epithelial Cells	9
SRGAP2	Epithelial Cells and Eye	7
SNRK	Epithelial Cells and Oxidative Stress	6
TTC3	Proteinase	6
COL21A1	Collagen	4
DHX29	-	3
FMN1	-	3
BPNT1	-	2
ZCCHC12	-	1
IGFN1	-	1
SDHAP2	-	0
ANKRD31	-	0
C3orf24	-	0
GOLGA6L6	-	0
RNASE13	-	0
HNRNPCL1	-	0

According to Pubmatrix results table 3-61, the genes that are showing high association with KC terms in the literature are: Ras-related C3 botulinum toxin substrate 1- (*RAC1*), Mucin 4- (*MUC4*), Toll-like receptor 6- (*TLR6*), Endoplasmic reticulum to nucleus signalling 2- (*ERN2*), Melanoma cell adhesion molecule- (*MCAM*), Nebulin- (NEB) and Titin-(*TTN*).

Pathway analysis (Table 3-62) was undertaken using DAVID (Section 2.14) on the proteins encoded by the list of 7 genes presented in the previous paragraph (*RAC1, MUC4, TLR6, ERN2, NEB, MCAM, TTN*). This revealed the involvement of RAC1, TTN and TLR6 in the pathways of regulation of kinase and transferase activities, regulation of phosphate and phosphorus metabolic processes and in the phosphorylation pathway. In addition, MCAM, MUC4 and RAC1 have been shown to be involved in the cell adhesion pathway as well as plasma membrane with the additional involvement of TLR6 in the latter.

Table 3-62: Pathway analysis of six genes (*RAC1, MUC4, TLR6, ERN2, MCAM, TTN)* highlighted as showing high association with KC terms in the literature. Table showing the functional annotation clustering of all the possible pathways/processes in which the 6 highlighted genes are involved. The column genes shows the genes involved in the pathways/processes column.

Cluster Number/ Enrichment score	Pathways/Processes	Genes
1 (E.S: 1.86)	regulation of kinase activity	*RAC1, TTN, TLR6*
	regulation of transferase activity	*RAC1, TTN, TLR6*
	regulation of phosphorylation	*RAC1, TTN, TLR6*
	regulation of phosphate metabolic process	*RAC1, TTN, TLR6*
2 (E.S: 0.97)	Cell adhesion	*MCAM, MUC4, RAC1*
	Plasma membrane	*MCAM, MUC4, RAC1, TLR6*
3 (E.S: 0.78)	Nucleotide binding	*ERN2, RAC1, TTN*
	Phosphoprotein	*ERN2, MCAM, RAC1, TTN*
4 (E.S: 0.87)	Signal	*ERN2, MCAM, MUC4, TLR6*
	Transmembrane	*ERN2, MCAM, MUC4, TLR6*
	Glycoprotein	*MCAM, MUC4, TLR6*
5 (E.S: 0.67)	Cell membrane	*MUC4, RAC1, TLR6*
	Plasma membrane	*MCAM, MUC4, RAC1, TLR6*
6 (E.S: 0.56)	Disulfite bond	*MCAM, MUC4, TTN*

3.3 Discussion

The existence of rare Mendelian forms of complex disease is well established with classic examples including Alzheimer's disease, diabetes and autoimmune disease (Peltonen *et al.*, 2006). While these familial cases only account for a small proportion of the patient population, they allowed gene identification studies to be performed which highlighted disease pathways important in their aetiology. To identify the genetic components underlying the majority of complex disease cases, genome wide association studies have been performed to identify the "common disease-common variants" traditionally believed to underlie complex disease (Risch and Merikangas, 1996). However, with the realisation

that these common risks variants fail to account for the majority of the heritability observed in the majority of complex diseases (Manolio *et al.*, 2009), recently there has been switch to the idea that rare variants with a large risk may underlie a significant proportion of complex disease (Cirulli and Goldstein, 2010, MacArthur *et al.*, 2014). Indeed, for some disorders historically regarded as complex, such as schizophrenia and autism, it has been reported that these actually represent a heterogeneous collection of rare monogenic disorders (McClellan and King, 2010). It was with this background of knowledge that this study set out to identify Mendelian alleles causing KC.

The basis of this study was a collection of KC families which were collected on a field trip to Lebanon, through clinical colleagues at St James's Hospital or through collaboration. The pedigree structure of the KC families, or the endogamous nature of the community they originated from, suggested that an autosomal recessive homozygous mutation was responsible for the KC phenotype. The "endogamous families" were all Lebanese and although these families did show a pedigree structure indicative of recessive inheritance of KC or consanguinity, the choice of studying KC in these families was due to the high levels of consanguinity and endogamy in the Lebanese population. The consanguinity rate in Lebanon is 35.5% (Barbour and Salameh, 2009) and 88% of marriages occur within the same religious communities (Klat and Khudr, 1986). There are high levels of endogamy in the different religious and ethnic groups forming the Lebanese population as there are 17 officially recognized religious communities in Lebanon, despite the country only having a population of 4.5 million (Nakouzi *et al.*, 2015). In particular, the Druze community, which is the religion of many of the endogamous families in this study, follows social customs that turns it into transnational isolates, by favouring marriage from within the same religion and forbidding admixture with other populations by firmly closing the religion to new adherents (Shlush *et al.*, 2008).

The hypothesis of an autosomal recessive form of KC is also backed up by the published literature. The majority of KC cases are sporadic (Patel and McGhee,

2013) (which can be indicative of recessive inheritance in small outbred communities) but familial KC is prominent with a rate ranging from 5 to 27.9% (Wang *et al.*, 2000). The increased prevalence of KC in populations with high levels of consanguineous marriage also suggests the existence of recessive alleles (Jonas *et al.*, 2009, Millodot *et al.*, 2011, Waked *et al.*, 2012, Shneor *et al.*, 2013, Hashemi *et al.*, 2014).

The strategy used to investigate the presence of recessive alleles underlying KC was a combination of autozygosity mapping and WES. Traditional gene identification studies start with a whole genome linkage search to identify a locus, followed by candidate gene screening. However, with the advent of high throughput sequencing technologies, it is now possible to circumvent these steps and identify a candidate gene by WES alone. Miller syndrome was the first example of a Mendelian disorder identified by WES (Ng *et al.*, 2010). Similar to this study, the inheritance pattern for this disorder was unconfirmed at the time of the experiment due to the majority of cases being sporadic with only a few sib pairs reported. However, the combination of analysing WES data from the affected individuals of multiple families under both recessive and dominant models led to the successful identification of recessive mutations in *DHODH* (Dihydroorotate Dehydrogenase) (Ng *et al.*, 2010). Since this first example, many disease genes have been identified using WES for both Mendelian disorders and complex diseases.

In this study, the first experiment focused on sequencing the exome of a single key individual in each family with the hope that convincing homozygous pathogenic mutations in a common gene(s) would be identified. Unfortunately this was not the case so further family members were sequentially sent for WES analysis to try and narrow the number of possible mutations. When this failed to highlight a clear candidate gene, SNP genotyping was undertaken to try and narrow down the list of possible variants further. The data presented in this thesis shows the final dataset by combining all of this data.

Autozygosity mapping was undertaken using a combination of SNP microarray genotyping data and with SNPs derived from the WES data. For many families a clear locus could not be identified. Based on the manual inspection of the autozygosity data, SSF and KF families presented autozygous regions. Family SSF presented a 1.2Mb on chromosome 11, 11q11. Family KF presented two autozygous regions on 9p21.1 and 12p13.1 of 3.07Mb and 1.69Mb respectively. The lack of a shared region of homozygosity in affected family members would indicate that there is no single recessive gene mutated in the family but there are limitations with the methodology used which prevent this conclusion being fully accurate. One important limitation is the rule-based nature of autozygosity mapping. Unlike traditional genetic mapping approaches (Lod scores), for autozygosity mapping it is not possible to factor in variables such as phenocopies and non-penetrance. The non-penetrance cases were avoided by only including affected family members in the mapping analysis, but there was no way to avoid phenocopies and given the frequency of KC this could be a significant risk.

Another reason for the failure to map a locus could be the resolution of the mapping software. The default setting of 1Mb was used in this study, but given that some of the families were endogamous the IBD (Identity By Descent) region in these cases may be small and may be below this resolution limit. However, for the families with a visible consanguineous loop this is unlikely to be problem as it has been shown that the disease associated region of homozygosity in these situations tend to be large (5-70 cM for offspring from first cousin marriages) (Woods *et al.*, 2006).

The SNP data derived from WES was used for autozygosity mapping in 2 families (SSF, AR1F). This data is not as good as the data derived from SNP microarrays. The SNPs on microarrays are chosen for their high levels of heterozygosity and for their even distribution throughout the genome. The SNPs derived from WES data have much lower levels of heterozygosity, for example, one study has shown that ~95% of 0.53 million known polymorphic positions

identified by the 1000 Genomes Project are homozygous by state for the reference allele in six unrelated WES datasets (Carr *et al.*, 2013). Similarly, they only span the targeted exons which are unevenly distributed across the genome so there are significant gaps in the genotyping coverage. This means that smaller regions of autozygosity may be missed in gene poor regions. However, given that the majority of mutations effect the coding portion of a gene (Stenson *et al.*, 2009), this shouldn't be a significant limitation. For all of these reasons, the WES data was still analysed independently of autozygosity mapping data if no autozygous regions were identified.

Despite highlighting autozygous regions in 2 large families (KF and SSF), 1.2Mb on 11q11 of family SSF and 3.07Mb and 1.69Mb on 9p21.1 and 12p13.1 respectively of family KF, no predicted pathogenic mutation was observed in this locus state that this could be due to the limitations of using WES as discussed below.

The majority of the analysis in this study was performed using WES, so the quality of this data was key to the success of this experiment. All the data used in this study was vigorously checked before starting the downstream analysis. This quality control comprises, most importantly, the quality scores across all bases (Example Figure 3-6a). This analysis showed a very good quality scores across the bases across all the samples included in this study with a mean quality score ranging from 28 to 32. Adding to the quality reports, and after NGS analysis the coverage statistics (Example Figure 3-7), which represents the percentage of bases covered above a certain value (n) (n= 5, 10, 15 and 20) was reported for every sample (Appendix 5) and the coverage across the sequenced exome is reported for all the samples in table 3-2. The WES performed in Leeds gave much better coverage than sequencing performed at Otogenetics with a mean average of coverage of 44.78 (±21.27) for Leeds compared to 69.59 (±9.55) for Otogenetics. This is why the method of WES switched mid way through the project.

The results for the homozygous autosomal recessive inheritance model didn't provide a clear causative gene but candidates were identified for most of the families. The list of candidates was too large to take forward so the whole data set was analysed to look for genes identified in multiple families which highlighted eight genes; *CHRNA3*, *FDFT1*, *RNF19B*, *ZFPM1*, *MUC3A*, *NOP9*, *AGAP3* and *LOC400863*. The first five of these have been associated in the literature with terms relating to KC.

The mutations in *MUC3A* (F11, F12 and F13) and *RNF19B* (F9 and F10) (MUC3A: exon2:c.C3319T:p.Q1107X and exon7:c.3934+1A>C and RNF19B: RNF19B: exon1:c.A185C:p.Q62P) have CADD Phred scores < 6, which suggests that they are unlikely to be pathogenic. *ZFPM1* represents an interesting candidate for KC since it is involved in the down regulation of *IL-4* (Interleukine 4) (Kitamura *et al.*, 2011) and the expression of *IL-4* is reported to be altered in KC tear samples. Jun et al. reported *IL-4* expression to be decreased in KC tear samples (Jun *et al.*, 2011), whereas Balasubramanian et al. showed an increase in *IL-4* expression in KC tear samples (Balasubramanian *et al.*, 2012). The *ZFPM1* variant shared between the affected members of F12 and F11 is a 4bp frameshift deletion (exon10:c.1335_1338del:p.P445fs) that is part of a 6bp inframe (exon10:c.1334_1339del:p.445_447del) deletion polymorphisms occurring in more than 50% of the 8 Druze controls sequenced in this study. The change detected in Druze controls is already a known SNP (rs149145771), of 0.951118 MAF on esp 6500si and of 0.001418 on ExAc03. ZFPM1 or zinc finger protein FOG family member 1, is involved in RNA polymerase II transcription factor binding (Mehaffey *et al.*, 2001, Rodriguez *et al.*, 2005). It is involved in the negative regulation of IL-4 biosynthesis, negative regulation of transcription from RNA polymerase II promoter and positive regulation in the interferon-gamma biosynthetic process (Rodriguez *et al.*, 2005, Kitamura *et al.*, 2011).

For *CHRNA3*, the same 3bp in-frame deletion (exon1:c.67_69del:p.23_23del) was identified in families F6, F7, F8, F9, F11 and F12 families. Transcriptome

profiling has identified *CHRNA3* as one of the retinal pigment epithelium (Logan *et al.*) key genes (Strunnikova *et al.*, 2010) and in directional migration of corneal epithelial cells (Chernyavsky *et al.*, 2014). Similarly, the same 6bp in frame deletion (exon1:c.193_198del:p.65_66del) in *FDFT1* has been identified in families F6, F7, F8, F9, F11, F12 and F13 families. FDFT1 or farnesyl-diphosphate farnesyltransferase 1, is involved in the farnesyl-diphosphate farnesyltransferase activity as well as the oxidoreduxtase activity (Robinson *et al.*, 1993).

Given the lack of a clear pathogenic homozygous recessive gene mutated in the families, alternative modes of inheritance were investigated including compound heterozygous recessive and dominant inheritance. Reduced penetrance and variable expression of autosomal dominant keratoconus has been reported in several pedigrees (Rabinowitz *et al.*, 1992, Fullerton *et al.*, 2002, Tyynismaa *et al.*, 2002, Brancati *et al.*, 2004, Tang *et al.*, 2005a, Burdon *et al.*, 2008, Udar *et al.*, 2009, Saee-Rad *et al.*, 2011) and some of the families in this study were consistent with dominant inheritance. As expected, large lists of variants were obtained from this analysis so again, the families were compared to look for commonality in the potentially mutated genes but no clear candidates were identified.

The only genes in Table 3-52 that are associated to KC are: Dystonin (*DST*), FIlaggrin (*FLG*) and Zinc Finger Protein 469 (*ZNF469*). Starting by *DST*, detected in SC3, KF, F10 and F12 families. It has been shown to be a component of the human and KC corneal basement membrane (Millin *et al.*, 1986, Cheng *et al.*, 2001). It is involved in integrin binding, microtubule plus-end binding and protein C-terminus binding (Favre *et al.*, 2001, Camargo *et al.*, 2007, Honnappa *et al.*, 2009, Steiner-Champliaud *et al.*, 2010, Kapur *et al.*, 2012, Blandin *et al.*, 2013, Hein *et al.*, 2015). It is equally involved in the cell motility and cytoskeleton organization processes, hemidesmosome assembly, integrin mediated signalling pathway, maintenance of cell polarity and response

to wounding (Favre *et al.*, 2001, Okumura *et al.*, 2002, Koster *et al.*, 2003, Hamill *et al.*, 2009).

As for *FLG*, detected in families SC1, SC2, F10 and F11, null mutations in this gene has been reported in KC patients with and without Atopic Dermatitis and/or Ichtiyosis Vulgaris (Droitcourt *et al.*, 2011). It is involved in the protein binding and structural molecule activity (McKinley-Grant *et al.*, 1989, Sakabe *et al.*, 2013). It is equally involved in the process of keratinocyte differentiation (Zigrino *et al.*, 2007).

Finally, mutations in *ZNF469* have been thought to contribute to the pathogenesis of KC (Rohrbach *et al.*, 2013, Sahebjada *et al.*, 2013, De Baere, 2014, Lechner *et al.*, 2014, Vincent *et al.*, 2014, Davidson *et al.*, 2015). Thus *ZNF469* was sequenced and checked for segregation in the 5 families where these changes occurred. However, none of these families segregated the *ZNF469* mutations in a perfect manner (Table 3-52). It is thought to be a transcription factor involved in the synthesis of collagen fibres in the cornea and regulating the expression of certain extracellular matrix components (Rohrbach *et al.*, 2013). Mutations in this gene are thought to cause Brittle Cornea Syndrome (Khan *et al.*, 2010).

The original aim of finding a clear major Mendelian gene mutated in the KC families was not achieved. This finding is consistent with other studies (De Bonis *et al.*, 2011, Nowak and Gajecka, 2011, Nielsen *et al.*, 2013, Nowak and Gajecka, 2015), and most importantly the study by Kriszt et al. which suggests that KC results from non-Mendelian alleles of major effect and is a complex disease (Kriszt *et al.*, 2014). Indeed, complexity can be one explanation why the family studies described in this thesis were not able to identify one Mendelian recessive or dominant allele responsible for KC in the large consanguineous families.

The complexity of the disease may lie in polygenic inheritance, indeed digenic inheritance of KC has been suggested in several studies (Bisceglia *et al.*, 2005, Burdon *et al.*, 2008, Steele *et al.*, 2008). Polygenic or oligogenic inheritance is consistent with the findings in this study, since multiple genes were highlighted in the families. Alternatively, the complexity of KC aetiology may reside in the intertwining of multiple genetics and environmental factors making it a heterogeneous, multifactorial disease. KC may therefore be due to the presence of predisposing alleles triggered by environmental factors (Davidson *et al.*, 2014), such as eye rubbing, atopy and ultra violet exposure (Gordon-Shaag *et al.*, 2015).

Clinical heterogeneity in KC may have been another complication preventing the detection of pathogenic alleles. The clinical observation of KC varies between complex pattern of inheritance, inherited (dominant and recessive) and sporadic forms, environmental causes, unilateral and bilateral presentation, different cone shapes, different degrees of severities in different stages of the disease and different speed of development. Consequently, some patients included in the same study might have different phenotypes classified under different clinical subtypes. This might be overcome by the unification and standardization of KC diagnosis, by differentiating all the clinical subtypes under the utilisation of the appropriate diagnostic guidelines and methods.

The methodology used to detect the pathogenic variants may have contributed to the reason why a clear Mendelian gene was not identified. The WES technique used to highlight variants only detects mutations in the coding part of the genome (and non-coding RNAs) including the splice sites and dismisses the majority of non-coding regions. Pathogenic alleles residing in the promoter or UTRs (or even intronic) regions might have been missed. Large deletions or duplications and rearrangements will also have been missed. In addition, all the non-pathogenic variants (synonymous) were filtered out from the final variants lists, which might have missed some variants that could give rise to cryptic splice site events and so might contribute to the genetic predisposition of KC.

Copy number variations (CNV) will also not have been detected using WES. CNVs have been shown to play an important role in many ocular diseases (Chanda *et al.*, 2008, Liu and Allingham, 2011, Liu *et al.*, 2012, Xu *et al.*, 2015) and two studies have investigated CNVs in KC (Abu-Amero *et al.*, 2011a, Rosenfeld *et al.*, 2011). Deletions of locus 5q31.1-q35.3 is associated with abnormal corneal phenotypes abnormalities showing duplications and deletions has been shown identified in an autosomal dominant KC family, however this result was not replicated by any other study (Rosenfeld *et al.*, 2011). New software programmes and computational tools have been developed which allow the analysis of CNV in WES and these can be used to interrogate the data generated in the future (Shi and Majewski, 2013, Zhao *et al.*, 2013, Wang *et al.*, 2014).

Imprinted alleles (Reik *et al.*, 2001) and trinucleotide repeat expansions also weren't investigated in this study. Since KC is a progressive disorder and susceptible to environmental factors, imprinted genes should be investigated since these genes are likely epigenetic targets for environmental interactions with the genome (Waterland, 2006). Similarly, trinucleotide repeat expansions are a possible cause of disease as has recently been shown for Fuchs' endothelial corneal dystrophy (Mootha *et al.*, 2015). This disease mechanism is especially interesting in families where the severity of KC increases through subsequent generations e.g. family AR1F (Marquis Gacy *et al.*, 1995, Mitas, 1997).

Mitochondrial defects were not investigated in this project, as mitochondrial inheritance is not consistent with the pedigree structure (except possibly family KF and F11). However, mitochondrial mutations have been linked to KC pathogenesis (Pathak *et al.*, 2011, Abu-Amero *et al.*, 2014b) as have mitochondrial copy number variation (Abu-Amero *et al.*, 2014b, Hao *et al.*, 2015b) and the accumulation of mitochondrial DNA damage (Atilano *et al.*, 2005).

Gender related differences are not well understood in KC, however the differences in sex hormones in men and women have been suggested to be a possible cause of KC pathogenesis (Bilgihan *et al.*, 2011). However, some studies have found a greater differences in females (Krachmer *et al.*, 1984, Fink *et al.*, 2005) others have found a greater difference in males (Owens and Gamble, 2003, Wagner *et al.*, 2007); while others have not found any evidence of gender bias in KC (Levy *et al.*, 2004, Ertan and Muftuoglu, 2008). This ambiguity could be explained by an X-inactivation mosaicism or lyonisation since it has been already revealed in a mouse model of human filamin A diseased corneas (Douvaras *et al.*, 2012).

As mentioned earlier, KC could be a di-genic, polygenic or a complex disorder. Although this study was not designed to find such genetic variants, it is possible that the high frequency of KC cases in the consanguineous or endogamous families studies in this project may be due to the enrichment of multiple pathogenic alleles in the families. As such, the PLINK/SEQ tool was used in this study to try and tease out multiple variants that didn't segregate in a Mendelian fashion in the families but were enriched in this cohort, as the large datasets were too large to manually do this. Enrichment analysis of rare variants through association studies is very challenging and much larger sample sizes than that used in this study are required (Lee *et al.*, 2014). Therefore, this PLINK/SEQ study is severely underpowered.

Another limitation of this study is using ancestral control data instead of ethnically matched controls. This is suboptimal but is an acceptable alternative since up until recently the Lebanese population had been considered as Caucasian (Baz *et al.*, 2001, Zalloua *et al.*, 2008). However, the major problem in the Lebanese population is the population stratification, which causes false positive results in enrichment and association analysis studies caused by systemic differences between ancestry in the subpopulation rather than an actual alleles association (Jiang *et al.*, 2013). Similarly the lack of Middle Eastern controls was compensated for by use of Pakistani controls due to geographical and population

similarities. However, the future of this project involves being part of a large scale association studies with other research teams and collaborators, where the data generated throughout this project will be included (Section 5.2.3).

Despite the limitations, this PLINK/SEQ study has several advantages. Firstly, it is based on patients selected from multiplex KC families and no sporadic cases are included. This strategy increases the chances of finding genetic variants, since familial KC patients are more likely to have the condition as a result of genetic factors for KC pathogenesis, while sporadic patients might potentially have been subject to other factors (e.g environmental) which may have played a role in their condition.

In spite of these listed PLINK/SEQ advantages, the very small sample size means that this study is underpowered, and findings should therefore be regarded as tentative. Nevertheless this study forms a complementary analysis to be considered alongside the family and transcriptional analyses also conducted throughout this project, that can shed the light on possible genes/rare alleles involved in the genetics of familial KC.

Although KC is not a common disease, filtering variants with MAF <2% may have risked the elimination of some common variants contributing to KC pathogenesis as well.

In conclusion, in this study attempts were made to identify a Mendelian allele in a cohort of KC families. Disappointingly the results are inconclusive. There was no clear stand out candidate mutant genes but long lists of potential candidates. These were too numerous to take forward without further evidence implicating them in the pathogenesis of KC. In the next chapter additional strategies were used to gather more evidence to help highlight candidates for future studies.

4 Transcriptome analysis of keratoconic and healthy human corneas

4.1 Introduction

The aim of this study was to perform transcriptome analysis of normal human cornea and KC corneas using RNAseq to help prioritise candidate genes involved in the pathogenesis of KC. At the time this experiment was initiated there was no corneal RNA-seq data published or in databases, however, a recent report has documented the RNA-seq profile of normal human corneal endothelial tissue (Frausto *et al.*, 2014), as well as the transcriptome of the mouse cornea (Pronin *et al.*, 2014). However, to date, the healthy human anterior corneal transcriptome as well as the KC corneal transcriptome, have not been explored by RNA-seq, though microarray data are available in the literature (Ha *et al.*, 2004, Mace *et al.*, 2011, Ghosh *et al.*, 2013). In this project, RNA-seq was used to study the WT anterior corneal profile along with the differential gene expression profile between KC and wild-type corneas. Thus, this study is the first to use RNA-seq to evaluate differential gene expression between KC and healthy corneas.

The corneas used in this project originated from two different groups: patients and controls (Section 2.3). The KC corneas (n=6) were surgically removed in theatre from patients undergoing a deep anterior lamellar keratoplasty (DALK) where all of the anterior corneal tissue was removed but the endothelial healthy part was left in place. The stroma, Bowman's layer and epithelial tissues were then preserved in RNAlater (Ambion) for a later RNA extraction. Healthy (WT) corneas (n=5) were cadaveric in origin, donated by five different previously healthy individuals, and were used prior to donation in descemet stripping automated endothelial keratoplasty (DSEK), where the inner endothelial layer of the cornea was transplanted, leaving its anterior part. As with KC corneas, the stroma, Bowman's layer and epithelium tissues of the WT corneas were then preserved for a later RNA extraction. Therefore, samples from both groups of

corneas (KC and WT) were anatomically comparable, and consisted of only the anterior corneal part most commonly affected in KC disease.

After collection, the RNA was extracted according to the protocol described in Section 2.3, and RNA libraries for each cornea were generated in-house using the TruSeq RNA sample preparation Kit v2 (Illumina) according to protocol described in Section 2.9 and then sequenced on an Illumina HiSeq 2500 instrument. The RNA-seq paired-end reads generated were then used to study the transcriptomic profiles of KC and WT corneas.

The transcriptomic study carried out in this chapter was divided into three main sections: WT anterior corneal RNA-seq expression gene profiling; Differential expression (DE) analysis between pooled KC and pooled WT corneas; and the combined analysis of WES variations detected in the corresponding genomic DNA of each individual KC cornea, along with its list of RNA-seq differentially expressed (DE) genes by comparison to WT.

The purpose of investigating the DE between pooled KC and pooled WT corneas was to highlight genes and pathways altered in KC diseased corneas compared to healthy-WT corneas. While this approach may potentially also detect primary genetic causes, it is better able to gain insight into the secondary disease processes underway in a KC cornea. Conversely, the purpose of running DE of every KC cornea without its biological replicates was to establish an individual DE profile specific for each KC cornea. By comparing this to the WES variants detected in this particular KC cornea, it was hoped that this might allow correlation between genetic variations and differences in the level of expression of the corresponding genes, potentially highlighting primary causative variants or changes that increase susceptibility to KC.

4.2 Results

4.2.1 Study design and RNA-seq data analysis

The overall pipelines applied in this chapter were developed by the author at the start of the transcriptomic study, by partially adapting available software (Example Figure 4-2). The first steps of RNA-seq analysis processing were the same for the three different analyses conducted in this chapter (Figures 4-1 and 4-2). RNA-seq reads were first checked for quality by running the raw fastq files on the fastQC software (Section 2.11), then they were processed using CutAdapt to trim the ends contaminated by the Illumina adapter sequence (GATCGGAAGAGCG) as well as the bases with low quality scores (<20 phred score) (Section 2.14). The RNA-seq read alignment, the transcriptome assembly, anterior WT cornea profile and DC between the two conditions (KC, WT) were all adapted from the Tuxedo set pipeline (Figure 4-2) (Trapnell *et al.*, 2012). The list of commands used to run all the software is given in appendix 3.

In the analysis workflow, RNA-seq reads of all the samples incorporated in this study were aligned to the reference genome Gencode GRCh38p2 (http://www.gencodegenes.org/releases/20.html) using the Tophat2 function (Section 2.14) of the Tophat software (Kim *et al.*, 2013), which is a gap aligner based around Bowtie2 which allows the alignment of reads around indels with high accuracy. Afterwards, the aligned bam files were assembled and quantified for expression via Cufflinks software using a guided reference transcriptome GRCh38p2, with the purpose of producing a GTF file of annotated transcripts for each sample (Section 2.14).

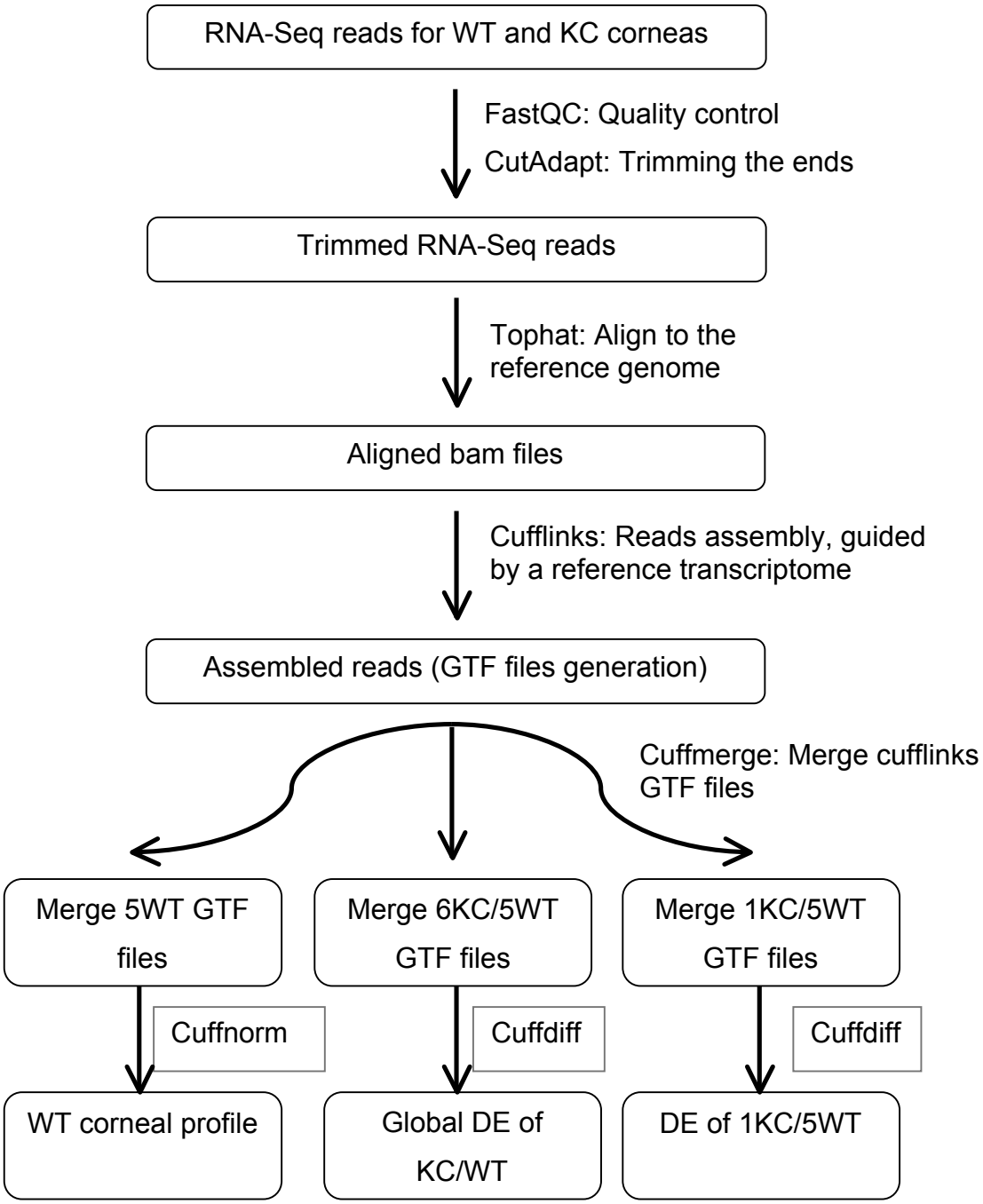


Figure 4-1: Tuxedo suite analysis pipeline. Flowchart describing the analysis pipeline workflow of the three topics described in this chapter, using the Tuxedo suite tool. In addition, it represents the first steps of quality control using the FastQC online software as well as trimming the ends using cutadapt software.

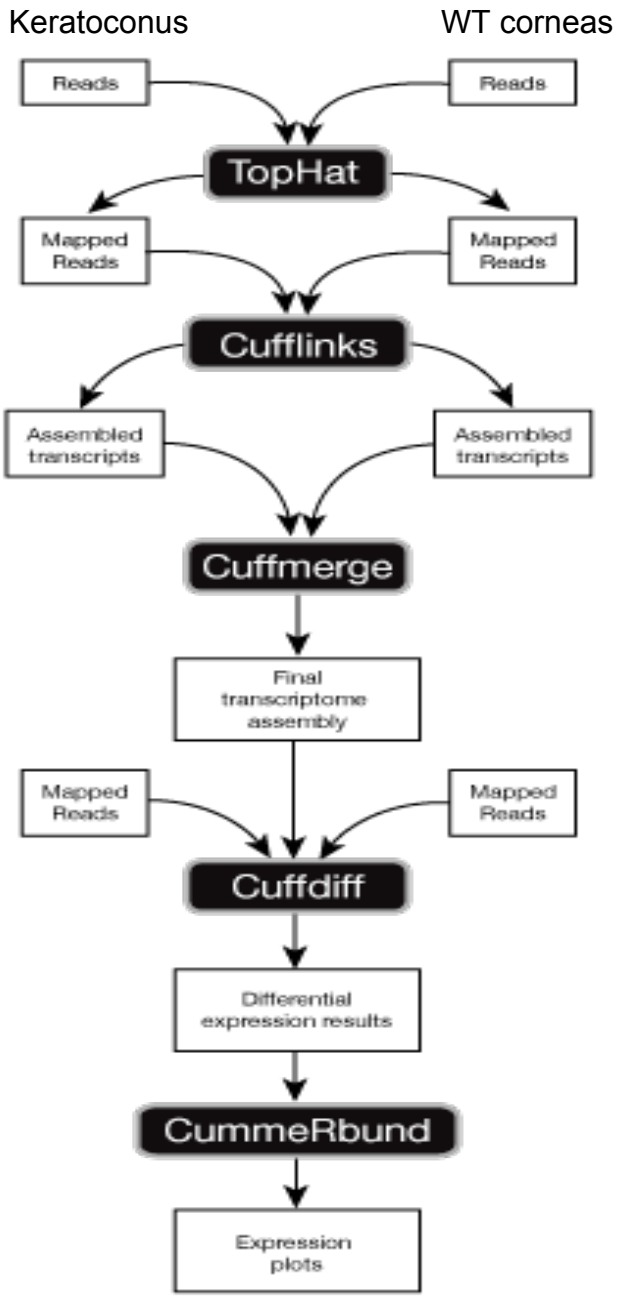


Figure 4-2: Tuxedo suite RNA-seq analysis pipeline. The analysis in this study is based on this flowchart that is the original RNA-seq analysis workflow presented by Trapnell et al. (Trapnell *et al.*, 2012) http://cole-trapnell-lab.github.io/cufflinks/manual/

Depending on the downstream analysis and the end results (Figure 4-1), the GTF file assemblies produced by Cufflinks were merged together using Cuffmerge into a master transcriptome suitable for downstream software analysis with Cuffnorm and Cuffdiff, described further below in Sections 4.2.1.1 and 4.2.1.2. While merging the assemblies and assisted by a supplied genomic DNA reference file (GRCh38p2) supplied from the Gencode website (http://www.gencodegenes.org/releases/22.html), Cuffmerge filtered probable artifacts (e.g. repeats) by automatically running Cuffcompare. In addition, Cuffmerge was supplied by a reference GTF file (GRCh38p2) for the purpose of maximizing the overall assembly quality by merging known and novel isoforms.

4.2.1.1 WT corneal gene expression profiling and novel gene/exon discovery.

The RNA-seq data of the WT corneas were processed in order to generate a healthy corneal gene expression profile. Thus, Cuffnorm was used only on the merged transcripts of WT corneas and tables of normalized FPKM (Fragments Per Kilobase Of Exon Per Million Fragments Mapped) values of expressed genes and isoforms were consequently generated. The annotation of the results generated by Cuffnorm was assisted by the GRCh38p2 reference transcriptome, by running the command presented in Appendix 3.

For the characterization of novel genes and exons, an additional experiment was conducted by following the protocol described by Farkas et al. (Farkas *et al.*, 2013), which required the use of a new alignment algorithm, the RNA-Seq Unified Mapper (RUM) (Section 2.14). The RNA-seq reads checked for quality on FastQC, were then realigned to human genome version hg19 throughout RUM (Section 2.14). RUM generates separate output files containing information on junctions (details presented in Section 4.2.2.2) from which novel junctions delimiting novel exons and/or genes were extracted and further validated by computational means (Section 4.2.2.2).

4.2.1.2 Differential expression with Cuffdiff

What followed the WT anterior corneal transcriptome generation, was the comparison of the pooled WT and pooled KC anterior corneal gene profiles in order to check for significantly differentially expressed genes. As previously described in figure 4-1, in parallel to Cuffnorm and once all the samples' relevant transcripts GTF files were merged, the software Cuffdiff was executed in order to compare the expression levels of genes in both conditions' RNA-seq data. The analysis, visualization and manipulation of the differentially expressed data generated by Cuffdiff were accomplished by CummeRbund (Trapnell *et al.*, 2012). CummeRbund (Section 2.14) is a plotting tool that provides rapid analysis and visualization of Cufflinks analysed high-throughput sequencing RNA-seq data. It is an R package that parses and outputs Cufflinks files and connects them to the R statistical computing environment (http://compbio.mit.edu/cummeRbund/).

4.2.1.3 Differential expression with EdgeR

In order to compare the list of differentially expressed genes following Cufflinks/Cuffdiff analysis, a second analysis approach was adopted (Section 2.14). It is mainly based on using the bioconductor package EdgeR (Robinson *et al.*, 2010) accompanied by Subreads for creating a counts matrix (Liao *et al.*, 2013).

Accordingly, RNA-seq data was run on a different pipeline using different software for quantification, read counting and differential expression analysis (see Figure 4-3). The subread package (Section 2.14) with the featureCounts program was used to assign paired end fragments to genomic features after being aligned using the Tophat software, as well as counting the RNA-seq reads. After counting the reads and creating a matrix of the read counts for each sample replicate, the matrix was then processed through edgeR, which is an R language based software package. The data was processed, analysed and visualised by edgeR on the R platform (Section 2.14).

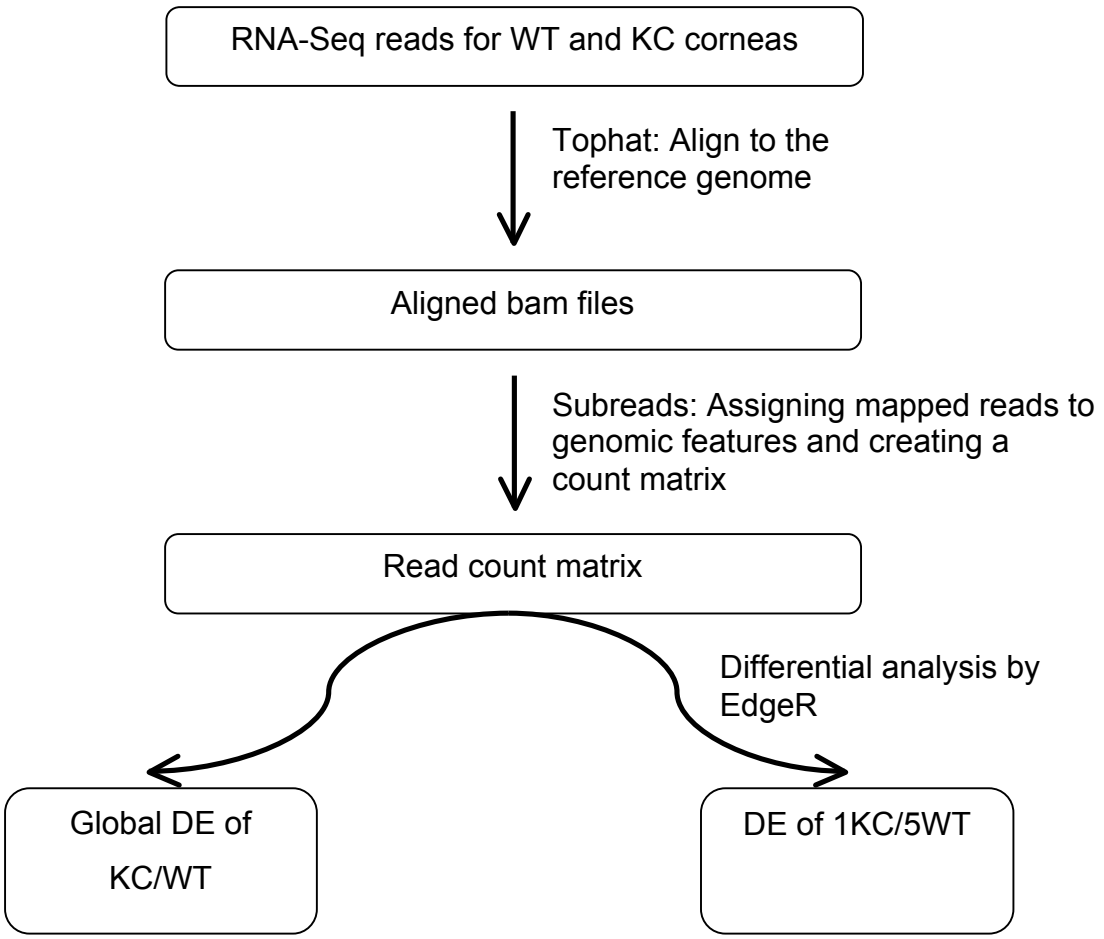


Figure 4-3: Differential expression analysis pipeline carried out using the bioconductor package EdgeR. It represents the alignment by Tophat in the first steps and then the read counts and DE by subreads and edgeR respectively.

4.2.1.4 Pathway analysis

The information gathered in sections 4.2.2 and 4.2.3 (DE and gene expression profile of normal corneas), was uploaded online to DAVID (The Database for Annotation Visualization and Integrated Discovery) pathway analysis (https://david.ncifcrf.gov/) software in order to generate significantly enriched gene clusters, providing information on the pathways up- or down-regulated in normal vs KC corneas (Section 2.14).

4.2.2 The normal anterior corneal gene expression profile

4.2.2.1 Analysis of gene expression using Cufflinks/Cuffnorm

The WT cornea RNA-seq data were run on the Cufflinks' program Cuffnorm, in order to generate normalized expression values for each gene, transcript, TSS (transcript's inferred start site) group and CDS (coding sequence) group, expressed in disease-free corneas. The pipeline of this experiment is illustrated in Figure 4-1. Data from all the 5 WT corneas were inputted into Cuffnorm using the command stated in appendix 3.3.

The output from Cuffnorm consisted of tab delimited text files of normalized FPKM values using the default settings of classic-FPKM, which implies that the library size factor is adjusted to 1, meaning that no scaling is applied to the FPKM values or the fragment counts.

The tab delimited file for FPKM-tracking of genes, which displays the summed FPKM values of transcripts sharing the same gene ID, was opened and analysed using Microsoft excel. The normalized FPKM values for genes expressed in individual WT corneas are averaged for the summed values in the 5 WT corneas, resulting in one global FPKM value for each gene.

Table 4-1 displays the first 30 genes with the highest averaged FPKM values expressed in WT corneas. In addition, the first 2500 protein coding genes with the highest FPKM values were exported to DAVID pathway analysis, and the results of the most significantly enriched pathways are listed in Table 4-2.

Table 4-1: The genes expressed in WT corneas. This table represents only the first 30 genes with the highest averaged FPKM values.

Gene name	Feature	Average FPKM values of 5 WT corneas
AC055736.3	miRNA	13280.036
KRT126P, KRT5	unprocessed_pseudogene, protein_coding	13167.468
FTH1	protein_coding	8611.054
AC055736.1	miRNA	8587.032

CLU, MIR6843	protein_coding, miRNA	7915.372
GAPDH	protein_coding	7040.092
TMSB4XP6	processed_pseudogene	6607.128
SNORA31,TPT1	snoRNA, protein_coding	5656.316
RPL37A	protein_coding	5397.854
S100A4	protein_coding	5244.868
RPS19	protein_coding	4956.726
KRT14	protein_coding	4787.77
ADIRF, AGAP11, BMS1P3, FAM25A, RP11-96C23.11, RP11-96C23.13, RP11-96C23.14	protein_coding, processed_transcript, transcribed_unprocessed_pseudogene, protein_coding, transcribed_unprocessed_pseudogene, processed_transcript, processed_transcript	4453.96
ALDOA	protein_coding	4443.002
AC016739.2	processed_pseudogene	4306.724
RPL35	protein_coding	4178.794
TGFBI	protein_coding	4060.062
TMSB10	protein_coding	3844.912
RP11-234A1.1	processed_pseudogene	3712.04
RPL13AP5	processed_pseudogene	3693.7
RPL30	protein_coding	3646.864
RP11-478C6.4	processed_pseudogene	3538.67
SPDYE2	protein_coding	3493.416
S100A6	protein_coding	3462.404
ALDH3A1	protein_coding	3438.33
AC022154.7, RPL18	antisense, protein_coding	3428.672
RPS11, SNORD35B	protein_coding, snoRNA	3323.018
RPS14	protein_coding	3315.86
S100A9	protein_coding	3257.374
RPL11	protein_coding	3232.048

Table 4-2: The most significantly enriched pathways in WT corneas. Table showing the first 7 mostly enriched clusters of pathway analysis of WT corneas, along with each pathway's p-value, Benjamini corrections and number of genes belonging to every particular pathway.

Cluster Number/ Enrichment score	Pathways	P-values	Benjamini Corrections	Number of genes
1 (ES: 40.06)	Ribonucleoprotein	3.6E-64	1.2E-61	139
	Translational elongation	5.7E-53	2.1E-49	79
	Ribosome	7.9E-53	1.8E-50	64
	Protein biosynthesis	1.7E-51	2.8E-49	102
	Ribonucleoprotein complex	2.2 E-51	7.0E-49	192
	Translation	3.0E-45	5.4E-42	139
	Cytosolic part	2.5E-39	3.2E-37	86

	Structural molecule activity	6.9E-31	4.2E-28	174
2 (ES: 16.54)	Respiratory chain	7.0E-28	5.9E-26	46
	Oxidative phosphorylation	6.1E-26	5.7E-24	72
	Mitochondrial membrane	1.2E-25	1.1E-23	125
	Organelle inner membrane	2.8E-25	2.3E-23	111
	Organelle envelope	9.6E-25	6.8E-23	166
	Generation of precursor metabolites and energy	1.1E-24	1.3E-21	105
	Mitochondrion	8.2E-24	6.2E-22	227
	Electron transport	6.5E-18	3.6E-16	44
	Transit peptide	3.5E-13	1.6E-11	101
	NADH dehydrogenase	1.2E-12	2.1E-10	25
	Mitochondrial ATP synthesis coupled electron transport	1.7E-12	2.6E-10	29
	Oxidoreductase activity, acting on NADH and NADPH	8.8E-11	1.3E-8	33
3 (ES: 14.15)	Intracellular non-membrane bounded organelle	6.2E-20	2.0E-18	451
	Cytoskeleton	9.3E-5	1.1E-3	211
4 (ES: 12.43)	Ubl conjugation	4.1E-13	1.7E-11	117
	Isopeptide bond	1.1E-12	4.2E-11	76
5 (ES: 11.75)	Melanosome	9.4E-24	3.7E-22	51
	Vesicle	1.1E-7	1.9E-6	127
	Membrane-bounded vesicle	9.6E-7	1.6E-5	108
	Cytoplasmic vesicle	1.7E-6	2.6E-5	118
	Cytoplasmic membrane-bounded vesicle	2.1E-6	3.0E-5	104
6 (ES: 10.89)	RNA processing	1.4E-14	3.3E-12	127
	RNA splicing	5.0E-14	9.6E-12	80
	RNA splicing via spliceosome	5.8E-14	1.1E-11	54
	Spliceosome	3.3E-9	7.0E-8	40
7 (ES: 9.31)	Membrane enclosed lumen	1.6E-14	4.0E-13	327
	Nuclear lumen	7.6E-8	1.3E-6	239
	Nucleolus	8.5E-6	1.2E-4	123
	Nucleoplasm	1.7E-4	1.9E-3	142

The results represented in Table 4-2, show the combined expression of miRNA genes, ribonucloprotein genes, protein coding genes and pseudogenes. The 20 most highly expressed protein coding genes are represented in Table 4-3.

Table 4-3: Most highly expressed protein coding mRNAs. This table presents the top 20 protein coding mRNA expressed in normal anterior human corneas.

Gene name	Gene symbol	Average FPKM values
Keratin 5, type II	*KRT5*	13167.468
Ferritin, heavy polypeptide 1	*FTH1*	8611.054
Clusterin	*CLU*	7915.372
Glyceraldehyde-3-phosphate dehydrogenase	*GAPDH*	7040.092

Tumor protein, translationally-controlled 1	*TPT1*	5656.316
Ribosomal protein L37a	*RPL37A*	5397.854
S100 calcium binding protein A4	*S100A4*	5244.868
Ribosomal protein S19	*RPS19*	4956.726
Keratin 14, type I	*KRT14*	4787.77
Adipogenesis regulatory factor **Family with sequence similarity 25, member A**	*ADIRF* *FAM25A*	4453.96
Aldolase A, fructose-bisphosphate	*ALDOA*	4443.002
Ribosomal protein L35	*RPL35*	4178.794
Transforming growth factor, beta-induced	*TGFBI*	4060.062
Thymosin beta 10	*TMSB10*	3844.912
Ribosomal protein L30	*RPL30*	3646.864
Speedy/RINGO cell cycle regulator family member E2	*SPDYE2*	3493.416
S100 calcium binding protein A6	*S100A6*	3462.404
Aldehyde dehydrogenase 3 family, member A1	*ALDH3A1*	3438.33
Ribosomal protein L18	*RPL18*	3428.672
Ribosomal protein S11	*RPS11*	3323.018

4.2.2.2 Identification of previously un-annotated coding exons and genes in the normal cornea

For novel gene and exon discovery in the WT corneas, the RNA-seq reads were realigned to human genome version hg19 using the RNA-seq Unified Mapper (RUM) (Section 2.14). In addition to RUM being an aligner implemented in Perl, it is also capable of junction calling and feature quantification, which makes it able to accurately identify novel splice variants, including splicing events and novel junctions with low read depth (Grant *et al.*, 2011). Combining information on novel junctions with the genome browser UCSC facilitated the locating of new junctions, which could potentially highlight a new exon or gene.

The alignment pipeline of RUM maps reads in three phases. First, it uses Bowtie, based on the Burrows–Wheeler algorithm, to map the reads against the hg19 version of the genome. In the second step it maps the reads against the transcriptome database. Finally, in the third step it maps against the genome using BLAT. Afterwards, Bowtie and BLAT alignments are merged after processing the aligned reads for RUM. This takes advantage of both algorithms by combining the speed of Bowtie with the sensitivity of BLAT (Grant *et al.*,

2011). Table 4-4 shows statistics on the mapping data generated by RUM in each WT cornea.

Table 4-4: RUM mapping statistics for every individual WT cornea. This table shows the statistics from RUM mapping. It represents the mapping percentage of bases covered in the exomes of the WT corneas across the whole genome. The percentages shown in column number two indicate the number of bases covered in the whole genome by reads that map uniquely to one genomic location. Column number three represents the number of bases covered in the whole genome by reads that map to multiple locations. The consistency between forward and reverse mapping is reported in terms of percentage and displayed in the last column.

Sample	No. bases covered by unique reads	No. bases covered by non-unique reads	Forward and reverse mapped consistently
MA28	139,146,189 (4.48%)	20.726,871 (0.66%)	27,935,376 (85.3%)
MA35	113,146,533 (3.64%)	17,463,784 (0.56%)	35,807,738 (87.7%)
MA51	111,306,258 (3.58%)	15,411,901 (0.49%)	6,405,133 (85%)
MA67	63,168,897 (2.03%)	10,911,345 (0.35%)	26,386,559 (86.1%)
MA82	67,280,262 (2.16%)	11,218,305 (0.36%)	19,039,207 (80.3%)

RUM generated basic alignment information for each paired end read in 2 files, the first being a file of unique alignments which mapped to only one genomic location and the second a file of non-unique alignments mapping to multiple genomic locations. The SAM file with all the alignments of unique and non-unique mapped reads is also an output of RUM, along with a Depth-of-coverage file which gives the number of reads mapping to each genome location. Quantified values of genes, exons, introns and junctions are generated and written in a feature quantification file. In addition, three output files are generated containing information on junctions that span gaps of 15 bases or more, which are determined by the default settings of RUM to be introns. The first file is a tab delimited spread sheet file containing expanded information on each junction. The second file is a bed file containing all the junctions, within which the colour blue is assigned to all the high quality junctions and all the other junctions are coloured red. The high quality junctions are listed separately in a third bed file, reporting all of the junctions that have the known splice signals and are uniquely mapping reads with at least eight bases on each side of the junction. In this file the junctions that are not known are shown in green, while the colour blue is assigned to the known junctions (http://cbil.upenn.edu/RUM/).

In this section of the results, only the high quality junctions bed file was of interest. The high quality bed files of all the 5WT corneas were uploaded individually into genome browser through the “mydata, custom track” function. The file was then exported to “Table Browser” under the “Tools” function. In order to only pick up novel genes/exons, the option that allows the online tool to extract the entries in the uploaded file that have no overlap with UCSC genes was selected under the intersection tab. Otherwise all the other setting were kept to default. This allowed the filtering out of all the junctions that overlapped with known genes in the UCSC records, leaving only the novel ones.

The bed files of all the 5 WT corneas were then checked for shared junctions among all the 5 corneas assigned with a green colour in the bed file (RGB code 0,255,127), using the perl script described in Section 2.12.

The result is a list of potentially novel junctions harbouring novel exons and/or genes (Table 4-5) appearing in the RNA-seq data for all 5WT corneas. The annotation of these junctions was carried out by inputting their coordinates to the UCSC browser and checking where these coordinates fall in the genome. Results are shown in the description column in Table 4-5.

Table 4-5: Table presenting the coordinates and description of potentially novel exons and genes shared between all 5 WT corneas. The chromosome and position represent the location of the new gene/exon on the genome, the average reads represent the average of the read coverage in all the 5 corneas and the description gives the gene within which this feature lies where known.

Chr	Position (hg19)	Average reads	Description
4	22728688…22728804	28.2	*GBA3 gene*
11	16996210…16996433	26.6	*PLEKHA7 gene*
X	100110437…100110693	18.8	*NOX1*
2	171608929…171609086	12.6	*AK023515*
19	45952452…45954783	11.2	*ERCC1*
3	169200949…169201452	6.6	*MECOM*
1	12053433…12053848	6.0	*MFN2*
2	27939549…27939763	5.2	-
9	94767588…94768059	5.2	-

8	56963001…56963225	5.0	-
2	127314138…127314258	4.8	-
3	169200948…169201452	3.6	*MECOM*
10	98510015…98510163	3.2	-
15	41849322…41849490	2.8	-
11	3991910…3992032	2.2	*STIM1*
3	169200951…169201452	2.0	*MECOM*
1	12053433…12053849	1.6	*MFN2*
6	166999922…167000071	1.4	*RPS6KA2*

4.2.3 Differential expression (DE)

4.2.3.1 Global statistics and quality control of processed RNA-reads prior to DE

This section describes the quality control and global analysis of cufflinks results after running the pipeline on data from all the 6 KC corneas pooled together and compared with the 5 pooled WT corneas. The results of Cuffdiff analysis following the Cufflinks pipeline (Figure 4-1 and 4-2) were reported in a set of tab-delimited text files for 4 classes: genes, isoforms, coding sequence (cds) and primary transcript (tss). The gene classes were explored by CummeRbund using the R commands described in appendix 3.4. Through CummeRbund, dendrograms, box plots and density plots were generated, which aimed to assess the variability for the class “genes” between replicates for each condition (WT and KC) and identify outliers in order to exclude them in the downstream analysis. The outliers are presented in the dendrogram shown in figure 4-4.a. The dendrogram provided insight into the relationships between conditions for the gene classes and revealed that sample KC_3, WT_3 and WT_4 were outliers compared to all the other replicates included in this study. Therefore, the analysis was repeated after removing these outliers from the study, in order to homogenise the statistical distribution in each set of conditions. The results of the new analyses that included 5 KC corneal replicates and 3 WT corneal replicates are presented in figures 4-5.b.

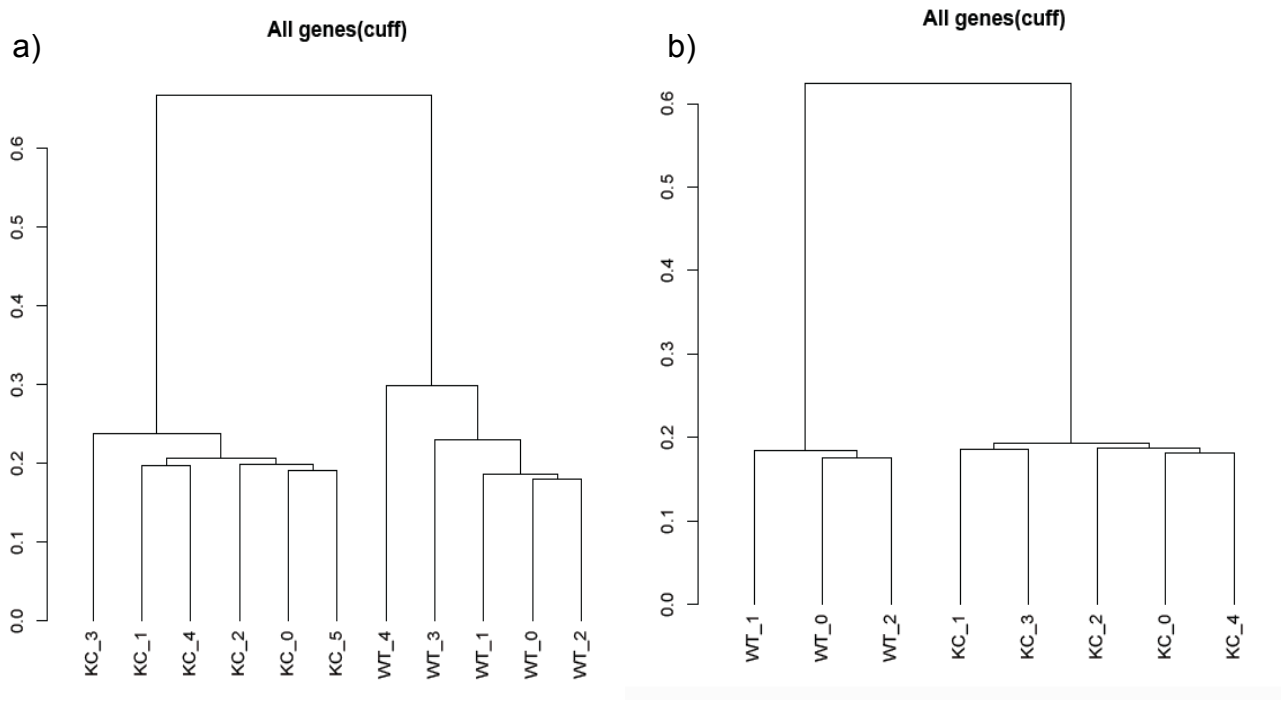


Figure 4-4: Dendrograms identifying the outlier replicates for each condition (KC or WT). a) Shows all the condition replicates, while b) shows results after removing the outliers in each condition, namely KC_3 for Keratoconus and WT_3 and WT_4 for Wild-type Corneas. The (y) axis represents the Height.

In addition, the same conclusion on outliers can be drawn from the density plots and box plots displayed in figure 4-5. The density plot was generated in order to assess the distributions of FPKM scores across WT and KC corneas (figure 4-5.a) as well as across each and every individual replicate (figure 4-5.b). Figure 4-5.c and figure 4-5.d show the box plots of the FPKM distributions for KC and WT conditions together with individual replicates. The box plot of figure 4-5.d in particular shows that the statistical distribution of the FPKM of the class “genes” is different between the sample replicate KC_3 and all the other KC replicates (KC_0, KC_1, KC_2, KC_4 and KC_5). The same is true for the replicates WT_3 and WT_4, for which the statistical distribution of the FPKM values does not constitute a box plot homogenous and similar to the box plot of the other WT replicates (WT_3, WT_4 and WT_5). In addition to finding replicate outliers, the WT and KC sets showed differences in terms of the FPKM distribution and that might be due to a number of factors discussed in Section 4.3.

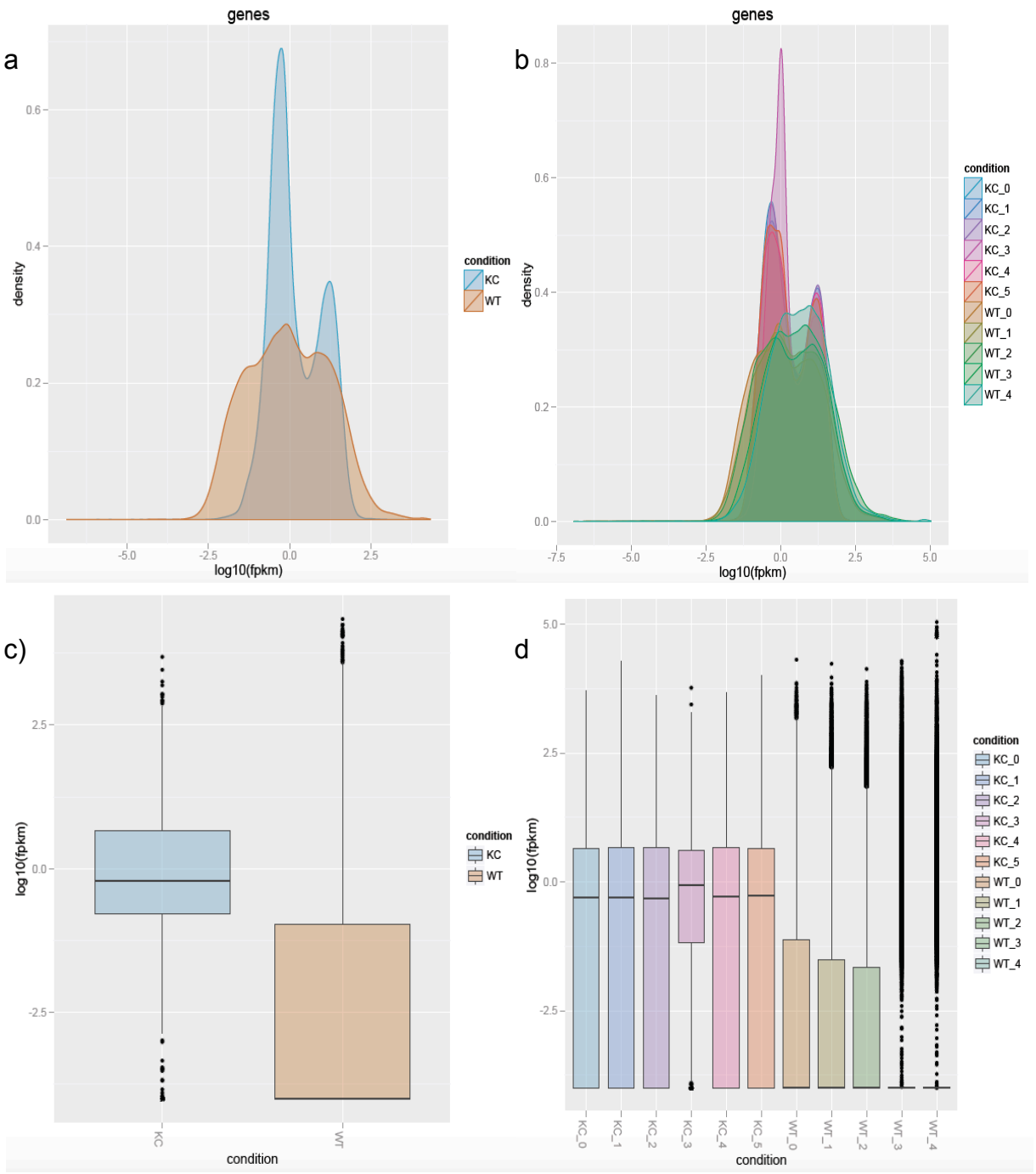


Figure 4-5: CummeRbund analysis exploration of 6 KC corneas compared to 5 WT corneas. a) Density plot shows the FPKM distribution for KC and WT conditions. b) Density plot exposes each individual replicate's FPKM distributions for all the samples included in the study and belonging to both conditions. c) Box plot of FPKM distributions for individual conditions (KC and WT). d) Box plot showing individual replicate's FPKM distributions for all the samples included in the study and belonging to both conditions.

Figure 4-6 displays the same statistics and plots shown in figure 4-5 after elimination of the replicate outliers. This resulted in a more homogenous FPKM distribution within the different condition sets (Figure 4-6.a and figure 4-6.b). In addition, the box plots of the replicates belonging to the same condition have a similar statistical FPKM distribution (Figure 4-6.c and figure 4-6.d).

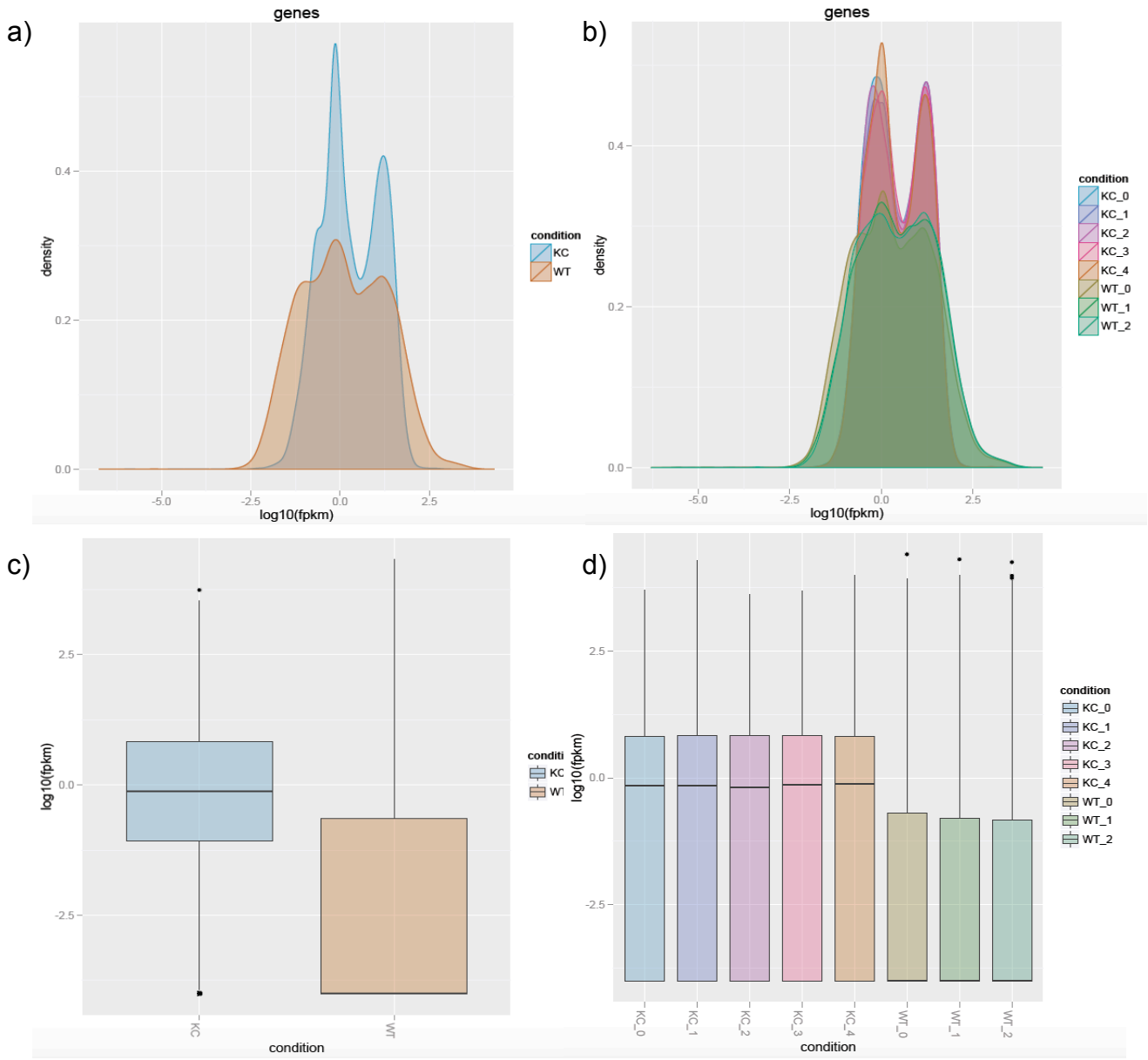


Figure 4-6: CummeRbund analysis of RNA-seq data from 5 KC corneas compared to 3 WT corneas after removing the outliers. a) Density plot shows the FPKM distribution for pooled KC and WT RNAseq data. b) Density plot of each individual WT vs KC RNA-seq profile exposes the individual replicate FPKM distributions for each sample included in the study. c) Box plot of FPKM distributions for individual conditions (5KC and 3WT). d) Box plot showing FPKM distributions for each individual WT and KC replicate included in the study. The black line in the box plots shows the median and the boxes the interquartile range and the whiskers show the range and the dots show the outliers.

After removing the two WT and one KC replicate outliers, the estimated dispersion for each sample was calculated by CummeRbund and visualised as a quality control measure by using the Cuffdiff replicate-dispersion method default settings. The dispersion modelling by Cuffdiff used pooled data within each condition to build a model, then provided a single global model for both conditions by averaging all the models. The plot in figure 4-7 shows some overdispersion in the data. However this does not illustrate poor data, since Cuffdiff takes into account the overdispersion while analysing differential expression by following a negative binominal distribution.

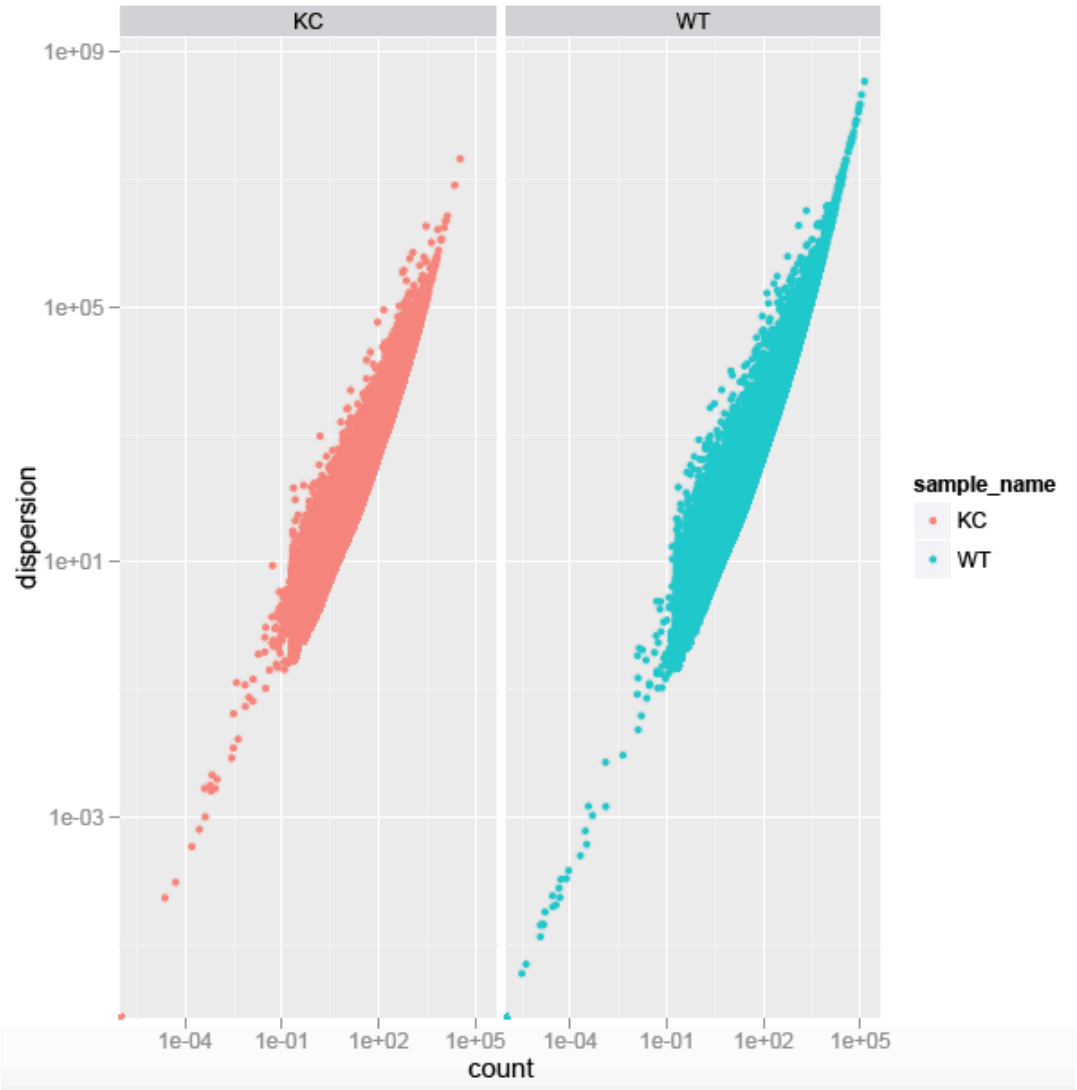


Figure 4-7: Count vs Dispersion plot by conditions (KC and WT corneas) for all the genes. The graphs of dispersion of KC (left) and WT (right) are skewed to the left which represents an overdispersion in the data.

4.2.3.2 Differential gene expression analysis of KC and WT corneas using the Cufflinks/Cuffdiff pipeline

Changes in expression at the level of transcripts, primary transcripts and genes are produced by Cuffdiff in a number of tab-delimited output files. Cuffdiff additionally features changes in splicing by tracking changes in the relative abundance of transcripts sharing a common transcription start site using a tss_ID (the ID of the transcript's inferred start site). It provides as well a list of changes in relative promoter use within a gene by tracing changes in the relative abundances of the primary transcripts of each gene by giving a p_ID for the coding sequence.

For each of the four classes (genes, isoforms, coding sequence-cds and primary transcript-tss), Cuffdiff outputs one of the following files per class:

- FPKM tracking file, which constitutes the FPKM calculation of each transcript, primary transcript, and gene in each sample.

- Count-tracking file, which constitutes the estimation of the number of fragments that originated from each transcript, primary transcript, and gene in each sample.

- Read Group tracking file, which comprises the expression calculations and fragment count for each transcript, primary transcript, and gene in each replicate.

- Differential expression test tab-delimited file, which lists the differential expression testing between samples for spliced transcripts, primary transcripts, genes, and coding sequences.

In addition, Cuffdiff calculates the differential splicing between isoforms processed from a single primary transcript, differential CDS and differential promoter use between samples, and outputs this data into tab-delimited files.

Table 4-6 represents the number of significantly and non-significantly differentially expressed genes, isoforms, coding sequence and primary transcripts, as well as the number of upregulated and dowregulated genes,

isoforms, cds and tss in pooled 5 KC corneas compared to 3 WT corneas. However, in this project, only the differential gene expression files were studied.

CummeRbund allowed visualization of the Cuffdiff results of the differential gene expression data. Figure 4-8 displays the volcano plot matrix for both KC and WT conditions.

Table 4-6: Cufflinks/Cuffdiff statistics. Table representing the numbers of Cufflinks/Cuffdiff significantly and non-significantly differentially expressed genes, isoforms, cds and tss along with their status of up or down regulation in 5 pooled KC corneas compared to the 3 pooled WT corneas.

Property	Non Significantly Different	Significantly Different	Upregulated KC/WT	Downregulated KC/WT
Genes	52005	28815	23346	5472
Isoforms	289592	21409	18381	3030
Cds	81744	5841	3502	2341
Tss	137146	32675	24217	8460

In this section, the top 10 mostly significant hits of upregulated and downregulated genes from the Cufflinks results are shown in Table 4-7. In addition, the top genes that are totally absent in the KC corneas compared to WT, and in WT compared to KC corneas, are presented in Table 4-8.

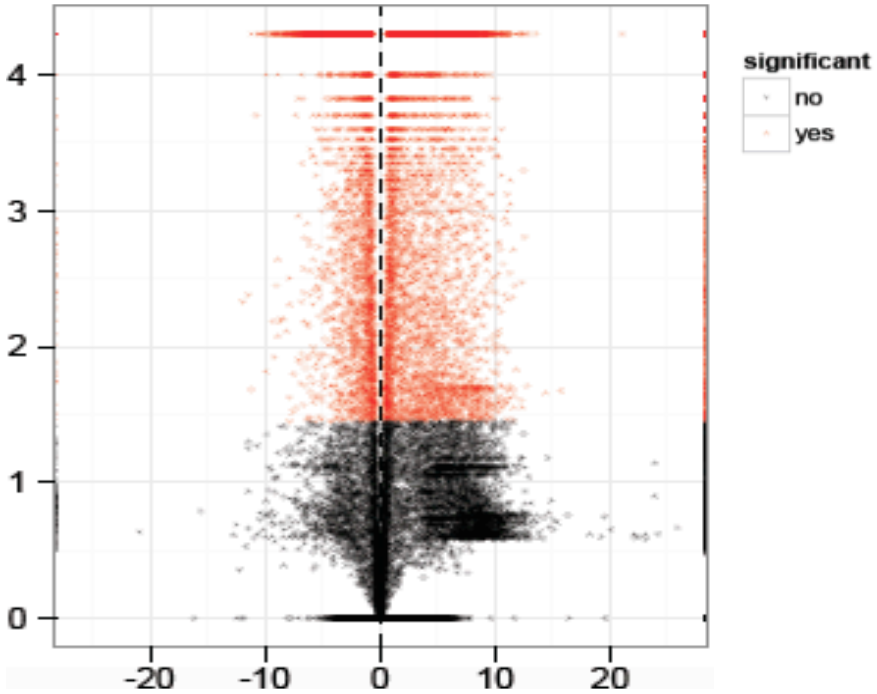


Figure 4-8: Volcano plot matrix of DE between KC and WT. Volcano plot matrix of genes that are differentially expressed between normal and keratoconic corneas. The red dots correspond to the significant hits, the black dots correspond to non significant differentially expressed genes. The x axis corresponds to the log2(Fold change) and the y axis corresponds to the –log10(p value).

Table 4-7: Differentially expressed genes generated by Cufflinks. The 10 most differentially expressed genes downregulated in the data of pooled KC corneas, and the 10 most upreglated genes, as generated by Cufflinks and sorted by their corresponding FDR values. The positive log2FC values correspond to the downregulated genes in pooled KC compared to WT, while the negative log2FC values correspond to the upregulated genes in pooled KC corneas compared to WT. The q-values are the corrected p-values calculated by Cufflinks using the Benjamini-Hochberg method.

Status	Symbol	Feature	Log2FC	q-value
Downregulated in KC compared to WT	*B2M*	protein_coding	11.2169	8.89E-05
	ALDOA	protein_coding	10.7145	8.89E-05
	RPLP1	Pseudogene	10.5149	8.89E-05
	FTH1	protein_coding	10.4474	8.89E-05
	GAPDHP1	Pseudogene	10.1332	8.89E-05
	NEAT1, mascRNA-menRNA	lincRNA, sRNA	10.0876	8.89E-05
	KRT14	protein_coding	9.96943	8.89E-05
	RPL13AP5	Pseudogene	9.82667	8.89E-05
	POLR2J2, POLR2J3, RASA4, RASA4B, RP11-	protein_coding, protein_coding, protein_coding, protein_coding,	9.52429	8.89E-05

	514P8.6, RP11-514P8.7, UPK3BL	protein_coding, protein_coding, protein_coding.		
Upregulated in KC compared to WT	*CST3*	protein_coding	9.51004	8.89E-05
	LEF1-AS1	processed_transcript	-21.0433	8.89E-05
	NRXN1	protein_coding	-13.5154	8.89E-05
	CA1	protein_coding	-12.9627	8.89E-05
	ADORA3	protein_coding	-12.6881	8.89E-05
	SLC14A1	protein_coding	-12.4454	8.89E-05
	CCDC38	protein_coding	-12.2557	8.89E-05
	ITK	protein_coding	-12.2356	8.89E-05
	IL5RA	protein_coding	-12.1758	8.89E-05
	TTC6	protein_coding	-11.8669	8.89E-05
	MYOCD	protein_coding	-11.8069	8.89E-05

Table 4-8: List of genes present in KC and absent in WT corneas or vice versa. The 10 most highly expressed genes that are present in WT corneas and totally absent in KC cornea, followed by the 10 genes most highly expressed in KC corneas that are totally absent in WT corneas. The q-values are the corrected p-values calculated by Cufflinks using the Benjamini-Hochberg method.

Status	Gene	Feature	FPKM in KC	FPKM in WT	q-value
Genes absent in KC corneas compared to WT	*HLA-C*	protein_coding	0	255.713	8.89E-05
	RP3-340B19.2	pseudogene	0	108.34	8.89E-05
	RPS18	protein_coding	0	88.5531	8.89E-05
	SEPT2P1	pseudogene	0	53.507	8.89E-05
	RP11-367G18.2	pseudogene	0	20.2826	8.89E-05
	CIDEB	protein_coding	0	18.6623	8.89E-05
	DEFB4A	protein_coding	0	16.0385	8.89E-05
	APOBEC3A	protein_coding	0	14.7865	8.89E-05
	DDR1	protein_coding	0	13.0465	8.89E-05
	RP11-10C24.1	lincRNA	0	8.6025	8.89E-05
Genes absent in WT corneas compared to KC	*WI2-2373I1.1*	pseudogene	251.369	0	8.89E-05
	AC003101.1, MIR365B	lincRNA, miRNA	217.97	0	8.89E-05
	MIR299, MIR411	miRNA	194.478	0	8.89E-05
	OR4M1	protein_coding	126.335	0	8.89E-05
	OR4K5	protein_coding	110.402	0	8.89E-05
	OR4F15, OR4F6	protein_coding	109.251	0	8.89E-05
	OR14A2, OR6F1	protein_coding	101.487	0	8.89E-05
	AC006539.6	pseudogene	99.6426	0	8.89E-05
	OR4K2	protein_coding	98.2316	0	8.89E-05
	OR4K1	protein_coding	8.89E-05	0	8.89E-05

4.2.3.3 Differential gene expression analysis of pooled KC and WT corneas using the edgeR pipeline

In parallel to Cufflinks, the same RNA-seq data was run on a different pipeline using different software for quantification and counting of the reads, and for DE analysis. The subread package (Section 2.14) with the featureCounts program was used in order to assign paired end fragments to genomic features, as well as counting the RNA-seq reads. After counting the reads and creating a matrix of the read counts for each sample replicate, the matrix is then processed through edgeR, an R language based software package (Section 2.14).

EdgeR is specifically designed to detect the relative changes in expression levels between two conditions i.e pairwise comparison, using empirical Bayes estimation and exact tests based on a negative binomial model, but is not concerned with quantification of expression levels (Robinson *et al.*, 2010). In other words, EdgeR does not estimate the absolute expression levels of a certain transcript, it rather looks at the relative changes in expression of this particular transcript between the two conditions in question (Chen *et al.*, 2008, revised 2015). The result of edgeR is an excel list of differentially expressed genes with their specific log fold change (log FC), the log count per million (logCPM) which can be converted to RPKM (Read Per Kilobase per Million mapped reads) for single end RNA-seq data or FPKM values for paired-end data, along with their corresponding p-values and FDR (False Discovery Rate) values (Li *et al.*, 2012a). The differentially expressed gene list is filtered for significant genes (FDR > 0.05) and a minimum fold change of 2 (-1 < logFC < +1 filtered out).

Table 4-9 represents the number figures of significantly and non-significantly differentially expressed genes along with the upregulated and downregulated genes.

Table 4-9: EdgeR statistics. Table detailing the numbers of significantly differentially expressed genes detected by EdgeR analysis, grouped according to whether they are up or down regulation.

EdgeR result	Non Significant Difference	Significant Difference	Upregulated KC/WT	Downregulated KC/WT
Genes	30028	35220	26309	8911

In addition to the list of differentially expressed genes, visualization of data can be performed using R, by plotting a smear graph to show the distribution of differentially expressed genes by counts per million. Figure 4-9 shows the smear graph of differentially expressed genes between the 5 KC corneas and 3 WT corneas (KC_0, KC_1,KC_2, KC_4, KC_5 and WT_3, WT_4, WT_5).

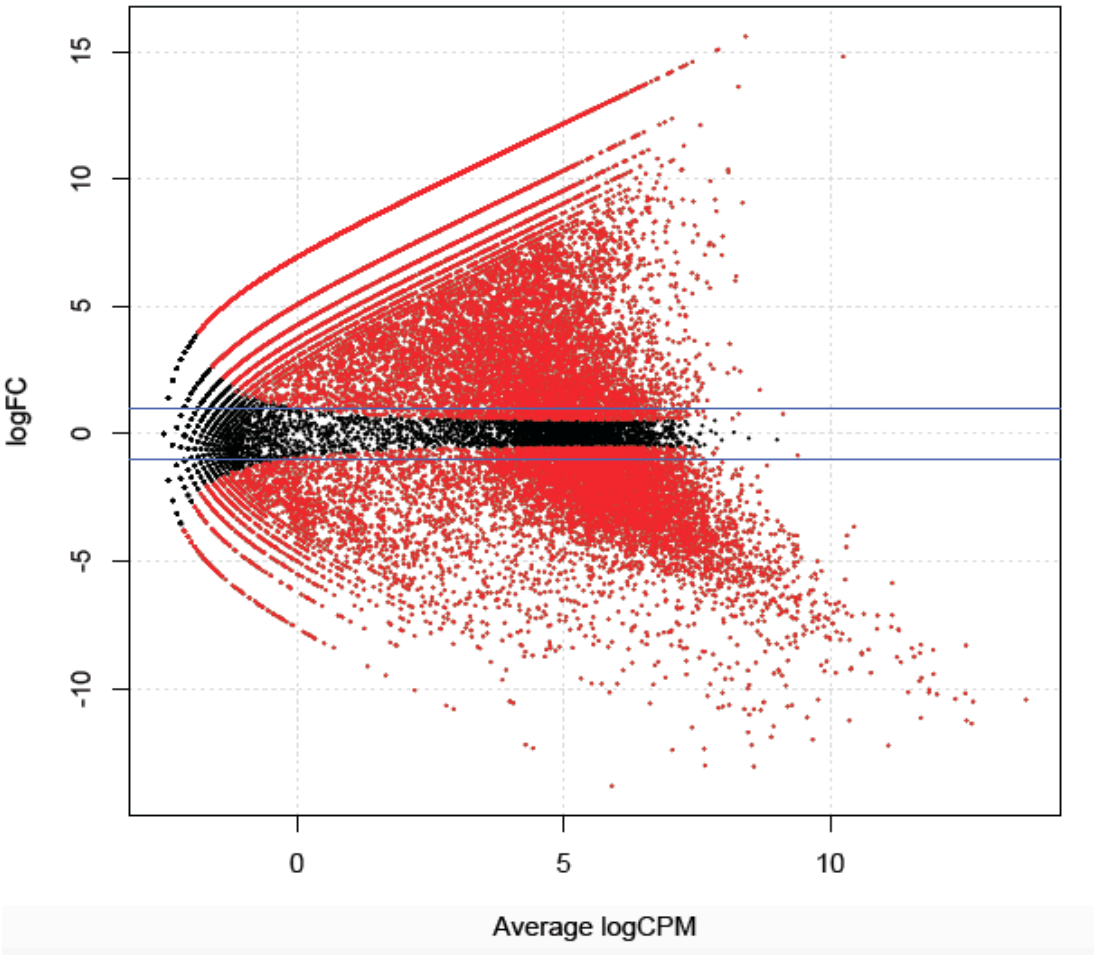


Figure 4-9: EdgeR DE plot. A smear graph showing the distribution of the differentially expressed genes detected in RNA-seq data analysed by edgeR, displayed by plotting the average logCPM (log2 counts per million) for each gene against its logFC (log2 Fold change). The blue lines delimit the threshold of a 2 fold-change (logFC range= -1, +1). All the genes falling in the area delimited between the blue lines have a 0-2 range of fold change, and are excluded from the final analysis. The genes falling outside the blue line range are considered for further analysis.

From the edgeR results, the 10 most significant upregulated and 10 most significant downregulated genes in KC corneas compared to WT are presented in Table 4-10.

Table 4-10: Differentially expressed genes between pooled WT and pooled KC corneas detected in RNAseq data by edgeR. The 10 most significantly upregulated and 10 most significantly downregulated genes in KC vs WT corneas, generated by edgeR and sorted by their corresponding FDR values. The negative logFC values correspond to the downregulated genes in KC compared to WT, and in contrast the positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Status	Symbol	Feature	logFC	FDR
Downregulated in KC compared to WT	*EEF1A1P5*	processed_pseudogene	-12.22586709	1.4867E-277
	MT-CO2	protein_coding	-11.36332101	5.9962E-269
	MT-CO3	protein_coding	-11.242322	1.4741E-265
	MT-ATP6	protein_coding	-11.14214094	2.147E-259
	RP11-234A1.1	processed_pseudogene	-11.99410887	6.2517E-254
	RPL13AP5	processed_pseudogene	-11.24404978	2.5852E-251
	KRT5	protein_coding	-10.43234188	8.4098E-247
	MT-ND4	protein_coding	-10.50619684	5.7367E-246
	MT-RNR2	protein_coding	-10.39711274	3.6219E-242
	RPS28P7	processed_pseudogene	-13.05276699	9.3026E-240
Upregulated in KC compared to WT	*TTN*	protein_coding	14.80004336	7.3408E-172
	MUC17	protein_coding	15.59338905	4.6454E-142
	MUC19	processed_transcript	13.62085338	1.2085E-139
	DNAH8	protein_coding	15.08070662	8.1949E-133
	XIRP2	protein_coding	15.05058135	2.8313E-132
	PKHD1L1	protein_coding	14.5968029	3.938E-124
	RYR3	protein_coding	10.26310516	7.354E-124
	RYR2	protein_coding	10.37602748	8.6621E-124
	CCDC168	protein_coding	12.11510674	1.5279E-123
	LRP2	protein_coding	14.48650695	3.6734E-122

4.2.3.4 Differential gene expression analysis of KC and WT corneas using the Cufflinks/Cuffdiff pipeline

In order to minimize the error and get confirmation from two different software packages on the results presented in this chapter, only the genes deemed to be differentially expressed in both Cufflinks/Cuffdiff and edgeR analyses were taken forward for Pathway analysis. Table 4-11 shows the number of genes that are classed as having DE by both edgeR and Cufflinks/Cuffdiff analyses. The difference between the outcomes of both analyses is explained in Section 4.3.3. Table 4-12 shows the 10 most significantly upregulated and 10 most significantly downregulated genes that appeared to overlap between the two previously mentioned software packages. The significance was reported equally by both software results (q-values by Cufflinks/Cuffdiff and FDR values by edgeR). In

addition, the overlapped results, in particular results of Table 4-12, were reported uniformly by both software to go in the same direction of upregulation or downregulation.

Table 4-11: EdgeR and Cufflinks/Cuffdiff overlap. Table representing the edgeR numbers of genes that are significantly differentially expressed genes overlapping between edgeR and Cufflinks/Cuffdiff results along with their status of up or down regulation.

EdgeR result	Significant Difference	Upregulated KC/WT	Downregulated KC/WT
Genes	18006	12846	5160

Table 4-12: Genes analysed by RNAseq and found to be differentially expressed between pooled KC and pooled WT corneas by both the Cufflinks/Cuffdiff and edgeR programmes. The edgeR results of the 10 most significantly upregulated and 10 most significantly downregulated genes between pooled KC and WT corneas, detected in RNAseq data by both Cufflinks/Cuffdiff and edgeR analyses, and sorted by their corresponding FDR values. The negative logFC values correspond to the downregulated genes in KC compared to WT, and in contrast the positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Status	Symbol	Feature	logFC	FDR
Downregulated in pooled KC compared to WT	*EEF1A1P5*	processed_pseudogene	-12.22586709	1.4867E-277
	RP11-234A1.1	processed_pseudogene	-11.99410887	6.2517E-254
	RPL13AP5	processed_pseudogene	-11.24404978	2.5852E-251
	KRT14	protein_coding	-10.21555574	1.3102E-236
	AC016739.2	processed_pseudogene	-11.47024012	9.7337E-235
	RP11-51O6.1	transcribed_processed_pseudogene	-12.20271809	1.5682E-233
	FTH1	protein_coding	-10.03571653	2.682E-231
	RPL3P4	processed_pseudogene	-10.69199626	7.0413E-228
	MTATP6P1	unprocessed_pseudogene	-10.38725886	1.6355E-228
	PERP	protein_coding	-9.699775015	1.3316E-221
Upregulated in pooled KC compared to WT	*TTN*	protein_coding	14.80004336	7.3408E-172
	MUC17	protein_coding	15.59338905	4.6454E-142
	MUC19	processed_transcript	13.62085338	1.2085E-139
	DNAH8	protein_coding	15.08070662	8.1949E-133
	XIRP2	protein_coding	15.05058135	2.8313E-132
	RYR3	protein_coding	10.26310516	7.354E-124
	RYR2	protein_coding	10.37602748	8.6621E-124
	LRP2	protein_coding	14.48650695	3.6734E-122
	COL6A5	protein_coding	14.36963531	4.5616E-120
	PTPRQ	protein_coding	14.21170612	3.1612E-117
	FAT3	protein_coding	9.920727171	5.0790E-117

4.2.3.5 Pathway analysis

4.2.3.5.1 Pathway analysis of DE genes detected by both Cufflinks/Cuffdiff and edgeR

The top 3000 protein coding genes, appearing on edgeR FDR values -regardless of their status of up or downregulation- to be the most significantly differentially expressed genes and detected as such by the results of Cufflinks/Cuffdiff program, were uploaded online for DAVID (The Database for Annotation Visualization and Integrated Discovery) pathway analysis (Section 2.14). Only the genes that were shown to be significantly differentially expressed in the same manner (up-regulated in both analyses and down-regulated in both analyses) were considered in this analysis. The functional annotation clustering tool was used in order to determine computationally the major biological functions specifically enriched in the differentially expressed gene list.

The results of DAVID functional clustering of the top 3000 most significantly differentially expressed genes detected in Cufflinks/Cuffdiff and edgeR analysis, are summarized in Table 4-13. In addition, Tables 4-14 and 4-15 show respectively the pathways that are enriched in the upregulated and downregulated genes.

Table 4-13: Pathways enriched in the DE expressed genes in pooled KC compared to WT corneas. Table showing the functional annotation clustering of the top 3000 significantly differentially expressed genes reported by FDR values of edgeR and detected by both Cufflinks/Cuffdiff and edgeR analyses. Only the clustering pathways with significant corrected p-values (Benjamini value < 0.05) are presented in this table.

Cluster Number/ Enrichment score	Pathways	P_Value	Benjamini Corrections	Number of genes
1 (ES: 44.83)	Glycoprotein	8.70E-72	6.90E-69	1000
	Glycosylation site: N-linked (GlcNAc...)	2.60E-69	1.70E-65	964
	Topological domain: Extracellular	3.50E-59	1.10E-55	686
	Topological domain: Cytoplasmic	1.30E-57	2.90E-54	799
	Membrane	5.30E-48	2.10E-45	1232
2 (ES: 32.24)	Disulfide bond	1.70E-32	2.70E-30	636
	Signal	1.10E-21	7.50E-20	645
	Secreted	6.20E-04	7.30E-03	289
	Extracellular region	2.10E-01	6.40E-01	352

3 (ES: 26.2)	Plasma membrane part	4.00E-30	1.30E-27	561
	Intrinsic to plasma membrane	2.40E-27	5.30E-25	348
	Integral to plasma membrane	1.30E-26	2.10E-24	340
4 (ES: 16.38)	cell adhesion	2.60E-22	2.10E-20	139
	biological adhesion	1.20E-17	9.90E-15	200
	cell-cell adhesion	9.70E-11	3.10E-08	88
5 (ES: 12.18)	Ribosome	5.90E-27	5.90E-25	52
	translational elongation	4.60E-24	2.00E-20	62
	protein biosynthesis	7.20E-11	2.10E-09	63
	Translation	2.50E-05	4.00E-03	83
	ribonucleoprotein complex	1.90E-03	1.80E-02	113
6 (ES: 11.37)	synaptic transmission	1.80E-14	1.30E-11	102
	transmission of nerve impulse	3.90E-13	2.00E-10	111
	cell-cell signaling	9.90E-10	2.30E-07	154
7 (ES: 10.79)	cell junction	8.30E-16	3.60E-14	120
	Synapse	4.90E-11	2.10E-09	110
	postsynaptic cell membrane	2.50E-08	5.90E-07	40
8 (ES: 10.65)	Receptor	3.00E-26	2.60E-24	378
	neurological system process	8.90E-24	1.90E-20	325
	Transducer	2.00E-19	1.20E-17	224
	Cognition	2.80E-14	1.70E-11	233
	GPCR, rhodopsin-like superfamily	8.80E-11	7.40E-08	177
	Olfaction	1.80E-10	5.00E-09	108
	Olfactory receptor	6.00E-07	1.70E-04	105
9 (ES: 10.03)	insoluble fraction	1.60E-12	1.00E-10	220
	membrane fraction	1.50E-11	7.70E-10	210
	cell fraction	3.30E-08	1.10E-06	250
10 (ES: 9.76)	Sodium transport	1.00E-12	4.30E-11	49
	metal ion transport	3.20E-11	1.20E-08	131
	monovalent inorganic cation transport	1.50E-10	4.60E-08	97
	Receptor	3.00E-26	2.60E-24	378
15 (ES: 5.09)	Na+ channel	3.90E-07	1.20E-04	10
	Sodium ion transport-associated	3.90E-07	1.20E-04	10
	voltage-gated sodium channel complex	1.50E-05	2.90E-04	10
16 (ES: 4.55)	Dynein	2.90E-07	6.30E-06	18
	Axoneme	1.70E-04	2.20E-03	18
	microtubule-based movement	3.10E-04	2.60E-02	34
	Cilium	2.20E-03	1.90E-02	36

Table 4-14: Pathways enriched among genes upregulated in pooled KC corneas compared to WT. Table showing the functional annotation clustering of the top 3000 significantly differentially expressed genes reported by FDR values of edgeR and detected by both Cufflinks/Cuffdiff and edgeR analyses. Only the clustering pathways with significant corrected p values (Benjamini value < 0.05) are presented in this table.

Cluster Number/ Enrichment score	Pathways	P-values	Benjamini Corrections	Number of genes
1 (ES: 76.25)	glycosylation site: N-linked (GlcNAc)	3.10E-153	1.90E-149	1147
	disulfide bond	4.30E-97	1.60E-94	807
	membrane	3.20E-64	2.60E-62	1278
	signal peptide	8.50E-43	7.50E-40	723
	Secreted	3.80E-14	1.30E-12	348
	extracellular region	2.90E-09	1.30E-07	407
2 (ES: 59.14)	receptor	7.70E-81	1.90E-78	507
	cell membrane	2.90E-80	5.20E-78	630
	neurological system process	4.20E-71	1.70E-67	416
	GPCR, rhodopsin-like superfamily	2.50E-65	2.60E-62	292
	G-protein coupled receptor protein signaling pathway	8.50E-63	1.70E-59	381
	cognition	1.80E-53	2.30E-50	315
	sensory perception	7.00E-53	6.90E-50	291
	Olfactory receptor	2.50E-48	1.80E-45	188
3 (ES: 32.37)	integral to plasma membrane	6.70E-35	8.90E-33	346
	intrinsic to plasma membrane	1.20E-34	1.30E-32	351
	plasma membrane part	9.50E-30	8.50E-28	530
4 (ES: 18.03)	synaptic transmission	6.10E-21	2.70E-18	110
	transmission of nerve impulse	9.20E-20	3.60E-17	120
	cell-cell signaling	1.50E-15	4.10E-13	164
5 (ES: 14.92)	ion transport	5.30E-37	3.00E-35	202
	substrate specific channel activity	4.70E-28	2.00E-25	150
	ion channel activity	4.80E-28	1.60E-25	147
	channel activity	2.80E-27	7.30E-25	152
	voltage-gated channel activity	3.40E-12	3.00E-10	70
	potassium ion transport	4.30E-08	6.50E-06	52
	transport	7.80E-08	1.60E-06	312
6 (ES: 12.1)	metal ion transport	1.50E-18	5.30E-16	143
	Sodium transport	1.40E-13	4.50E-12	50
7 (ES: 11.56)	cell adhesion	1.40E-15	4.40E-13	184
	biological adhesion	1.70E-15	4.40E-13	184
	Cadherin	2.70E-12	9.40E-10	47
8 (ES: 9.25)	synapse	1.80E-15	1.10E-13	116
	neurotransmitter binding	1.10E-14	1.10E-12	50
	postsynaptic cell membrane	5.70E-14	1.90E-12	49
	neurotransmitter receptor activity	8.10E-13	7.60E-11	45

	postsynaptic membrane	7.80E-12	4.20E-10	55
	cell junction	4.60E-10	2.30E-08	136
	chloride channel	7.50E-07	1.30E-05	24
	anion binding	7.50E-05	2.20E-03	30
9 (ES: 7.28)	peptide receptor activity	9.10E-11	6.60E-09	47
	neuropeptide receptor activity	1.60E-06	7.90E-05	20
10 (ES: 6.6)	extracellular matrix	1.20E-07	2.40E-06	66
	proteinaceous extracellular matrix	3.40E-07	1.10E-05	86
	extracellular matrix	3.70E-07	1.10E-05	91

Table 4-15: Pathways enriched in genes downregulated in pooled KC corneas compared to WT. Table showing the functional annotation clustering of the top 3000 significantly differentially expressed genes reported by FDR values of edgeR and detected by both Cufflinks/Cuffdiff and edgeR analyses. Only the clustering pathways with corrected p values (Benjamini value < 0.05) are presented in this table.

Cluster Number/ Enrichment score	**Pathways**	**P-values**	**Benjamini Corrections**	**Number of genes**
1 (ES: 33.15)	ribonucleoprotein complex	9.70E-49	3.40E-46	222
	translation	1.30E-40	5.40E-37	157
	translational elongation	1.40E-39	2.90E-36	77
	protein biosynthesis	5.90E-39	6.10E-37	104
2 (ES: 12.53)	pigment granule	3.40E-23	3.40E-21	57
	melanosome	3.40E-23	3.40E-21	57
	vesicle	1.40E-08	2.90E-07	165
	cytoplasmic vesicle	1.60E-07	2.80E-06	155
3 (ES: 12.31)	ubl conjugation	4.80E-16	2.30E-14	159
	isopeptide bond	1.20E-14	5.30E-13	100
4 (ES: 12.3)	Apoptosis	9.30E-14	3.70E-12	111
	programmed cell death	1.80E-13	3.20E-11	170
5 (ES: 12.03)	membrane-enclosed lumen	6.60E-18	3.10E-16	433
	organelle lumen	2.10E-16	8.20E-15	421
	nuclear lumen	9.60E-12	2.40E-10	330
	nucleoplasm	8.00E-08	1.40E-06	203
6 (ES: 11.91)	RNA splicing	6.60E-17	9.30E-14	104
	mRNA metabolic process	7.80E-14	1.60E-11	118
	RNA processing	5.00E-13	7.30E-11	155
	spliceosome	4.90E-10	1.10E-08	52
7 (ES: 11.11)	organelle membrane	1.60E-22	1.40E-20	300
	mitochondrion	1.30E-20	1.20E-18	219
	organelle envelope	4.20E-19	2.40E-17	188
	respiratory chain	1.80E-17	1.30E-15	42
	Oxidative phosphorylation	6.60E-15	4.10E-13	66
	electron transport chain	4.10E-11	4.60E-09	49

	transit peptide	4.60E-07	7.80E-06	109
	NADH dehydrogenase activity	8.70E-06	5.30E-04	20
8 (ES: 10.87)	non-membrane-bounded organelle	3.40E-16	1.10E-14	564
	cytoskeleton	2.10E-02	1.20E-01	253
9 (ES: 10.06)	protein localization	4.80E-16	2.10E-13	235
	protein transport	3.60E-14	1.00E-11	204
	transport	2.10E-08	4.40E-07	318
	intracellular protein transport	2.90E-08	1.90E-06	103
	protein targeting	4.20E-07	2.40E-05	65
10 (ES: 10.05)	protein catabolic process	2.10E-10	1.90E-08	162
	macromolecule catabolic process	2.30E-10	2.00E-08	194
	proteolysis	4.00E-04	1.30E-02	212

4.2.3.5.2 Pathway analysis of genes detected as differentially expressed by Cufflinks/Cuffdiff and totally absent in either pooled KC or pooled WT corneas

In this sub-section, genes found to be differentially expressed in analyses with Cufflinks/Cuffdiff and totally absent in KC corneas compared to WT corneas are subjected to DAVID functional clustering, the results of which are displayed in Table 4-16. DAVID analyses of genes found to be differentially expressed with Cufflinks/Cuffdiff and totally absent in WT corneas compared to KC corneas are presented in Table 4-17.

Table 4-16: Pathways enriched in the genes expressed in pooled KC and absent in pooled WT corneas. Table showing the functional annotation clustering of the 800 genes of significant q-values that appear on Cufflinks/Cuffdiff to be totally absent in WT corneas compared to KC. Only the clustering pathways with significant corrected p values (Benjamini value < 0.05) are presented in this table.

Cluster Number/ Enrichment score	Pathways	P-values	Benjamini Corrections	Number of genes
1 (ES: 160.16)	Olfactory receptor	2.2E-259	27.4E-257	241
	G-protein coupled receptor	6.9E-236	1.2E-233	287
	GPCR Rhodopsin-like superfamily	4.4E-235	1.2E-232	280
	Transducer	2.9E-227	2.6E-225	288
	Sensory transduction	2.9E-217	4.2E-215	245
	Olfactory transduction	1.6E-202	1.4E-200	224
	Sensory perception	6.4E-180	2.3E-177	254
	Cognition	1.3E-169	3.7E-167	257
	Receptor	7.0E-160	4.8E-158	299
	Neurological system process	3.4E-148	8.1E-146	268
	Cell surface receptor linked signal	5.9E-138	1.2E-135	308

	transduction			
	Cell membrane	3.7E-127	2.1E-125	308
	Disulfite bond	8.5E-125	4.1E-123	347
	Glycoprotein	1.9E-104	8.1E-103	389
	Plasma membrane	5.6E-63	1.3E-60	343
	Integral to membrane	1.4E-46	1.6E-44	384
	Intrinsic to membrane	1.8E-44	1.3E-42	388
2 (ES: 2.8)	Four helical cytokine, core	3.5E-6	3.9E-4	11
	Sematotropin hormone	8.0E-4	4.8E-2	4
4 (ES: 1.97)	IFabd	6.8E-7	8.0E-5	7
	Interferon alpha/beta/delta	1.7E-5	1.5E-3	7
	Interferon alpha	5.4E-5	4.3E-3	6
	Leucocyte	8.1E-5	2.1E-3	6
	Antiviral	8.5E-4	1.9E-2	5
5 (ES: 1.65)	Keratin	4.4E-6	1.3E-4	18
	Intermediate filament cytoskeleton	7.0E-4	3.9E-2	19
10 (ES: 1.35)	Secreted	9.1E-4	1.8E-2	79
11 (ES: 1.09)	LRRNT	9.0E-4	5.2E-2	9

Table 4-17: Pathways enriched in the genes expressed in pooled WT and absent in pooled KC corneas. Table showing the functional annotation clustering of the 26 genes of significant q-values that appear on Cufflinks/Cuffdiff analysis to be totally absent in KC corneas compared to WT. Only the clustering pathways with significant corrected p value (Benjamini value < 0.05) are presented in this table.

Cluster Number/ Enrichment score	Pathways	P-values	Benjamini Corrections	Number of genes
1 (ES: 1.81)	MHC class I protein complex	5.1E-5	4.4E-3	3
	Antigen processing and presentation	2.5E-3	3.8E-2	3

4.2.4 Differential expression profile for each cornea and its WES

4.2.4.1 WES and RNA-Seq combination for each individual KC

The RNA-seq data for each of the 6 KC corneas were checked individually for differential expression, by running the edgeR protocol (Section 2.14) and the Cufflinks/Cuffdiff protocol (Section 2.14) on each individual cornea and compared with the pooled 3 WT corneal replicates. Using the pipeline described in Figure 4-10, the differential gene expression profiles generated by both software packages were checked for significant changes (Cufflinks/Cuffdiff q-values and edgeR FDR values < 0.05), then a common list of genes between both files was generated. In parallel, WES libraries were prepared from the DNA extracted from

blood of the KC cornea donors according to the protocol described in Section 2.9. The WES reads were checked for quality, aligned, processed using GATK and annotated by ANNOVAR following the pipeline described in appendix 1.6. The result is a raw, non-filtered, annotated list of changes specific to every individual who donated their KC corneas.

Next, the list of significant differentially expressed genes identified in analyses by both software packages was compared against the list of changes from WES, in order to check for genomic mutations in these particular differentially expressed genes. The results displayed in Tables 4-18 to 4-23 show, patient by patient, the exonic and splicing variants detected in WES in genes that have been shown to be differentially expressed in the cornea of that KC patient compared to WT, and having the highest CADD scores (≥ 30) along with MAF<2%.

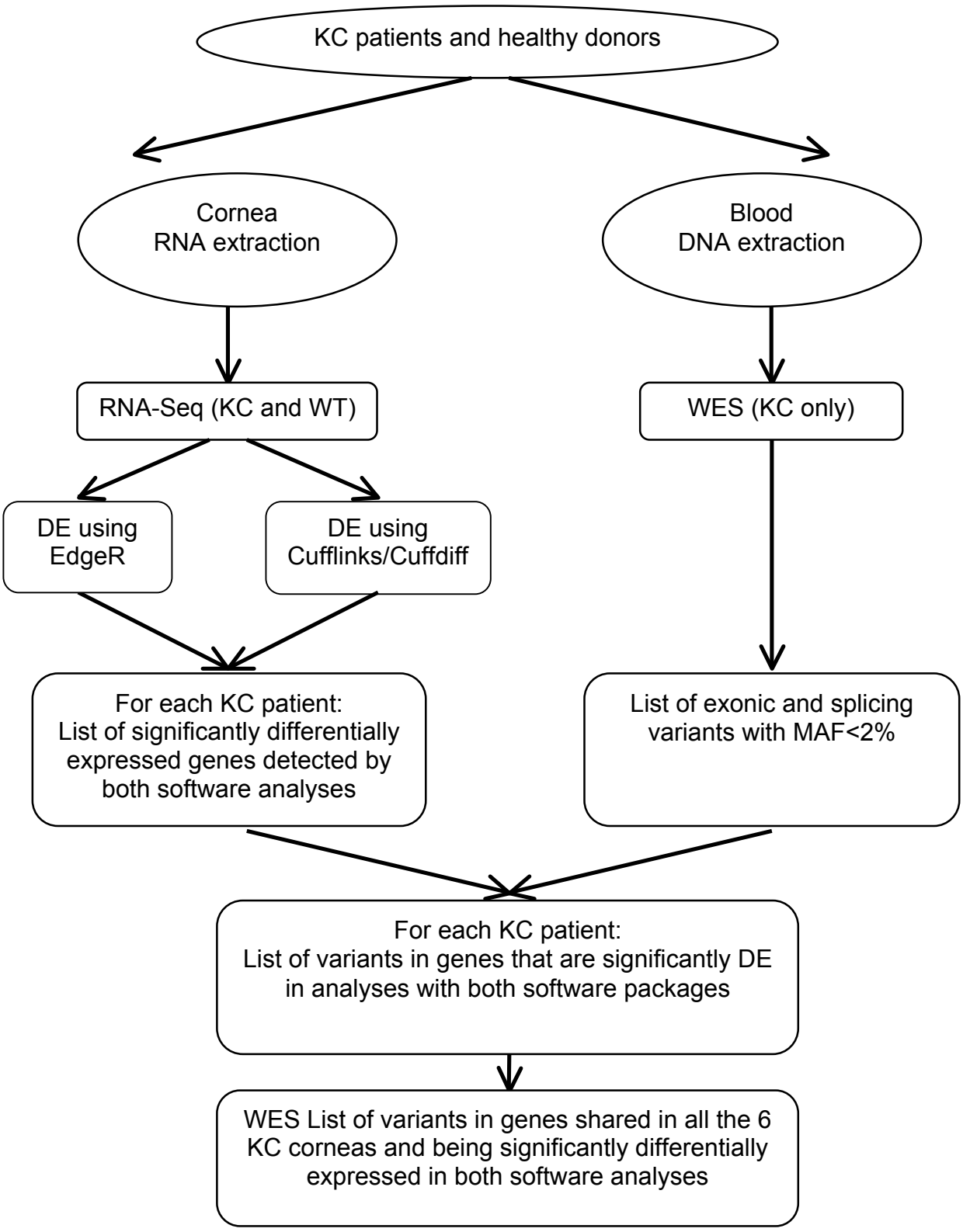


Figure 4-10: Combining RNA-seq and WES data. The pipeline presented in this chart combines the analysis of DE in KC vs WT corneal RNA-seq data with WES data of from the patients from whom the KC corneas were taken, to produce a list of genes differentially expressed and containing potentially biologically significant genetic variants

Table 4-18: MA18-KC (KC-0) WES and DE results combined. Genes with coding variants detected in WES of MA18-KC (KC-0) cornea, with CADD Phred score > 30, differentially expressed (FDR < 0.05) by comparison with WT in analyses with both edgeR and Cufflinks, and listed in descending order of CADD score. logFC and FDR values outputted by edgeR are also given. The negative logFC values correspond to the downregulated genes in KC compared to WT. The positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the significant corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Gene	Location	Cadd Phred	AA Change	Function	esp6500	1000g	exac03	snp138	Status	logFC	FDR
GJD3	17q21.2	36	exon1:c.C237A:p.L79L	synonymous	0.0148	0.0058	0.0120	rs141158879	het	4.6270	8.607E-0
CAPN2	1q41	35	exon1:c.233_234insCACGGTAGGAAGCG:p.P78fs	frameshift insertion	.	0.0012	0.0040	.	het	-3.7916	7.750E-0
ERCC6	10q11.23	35	exon9:c.1918dupT:p.W640fs	frameshift insertion	0.0001	.	.	.	het	1.8119	5.714E-0
KIAA1919	6q21	35	exon4:c.T545A:p.L182X	stopgain	.	0.0070	0.0180	rs117505745	het	2.9732	2.303E-0
CDC6	17q21.2	34	exon6:c.G883A:p.D295N	nonsynonymous	0.0156	0.0056	0.0130	rs4135012	het	3.6574	1.122E-0
GLP1R	6p21.2	34	exon13:c.C1261T:p.R421W	nonsynonymous	0.0007	.	0.0004	rs146868158	het	11.5732	7.380E-2
ITGA7	12q13.2	34	exon15:c.C2191T:p.R731W	nonsynonymous	.	.	0.0000	.	het	5.4918	3.726E-1
LCA5	6q14.1	34	exon5:c.G902A:p.R301H	nonsynonymous	0.0004	0.0006	0.0005	rs139142572	het	2.9487	1.959E-0
MINK1	17p13.2	34	exon24:c.C2830T:p.R944W	nonsynonymous	.	.	.	.	het	-3.5925	5.617E-1
RYR3	15q14	34	exon50:c.C7613T:p.T2538M	nonsynonymous	.	.	0.0001	.	het	10.2694	9.010E-5
KRT76	12q13.13	33	exon6:c.C1234T:p.R412W	nonsynonymous	0.0028	0.0004	0.0018	rs147080851	het	4.5511	2.354E-0
MUC6	11p15.5	33	exon31:c.C5760A:p.Y1920X	stopgain	.	.	0.0002	rs75482640	het	10.9099	6.703E-1
NEB	2q23.3	33	exon118:c.G16546C:p.D5516H	nonsynonymous	0.0020	0.0002	0.0019	rs201979610	het	7.7729	1.472E-4
SLC14A2	18q12.3	33	exon20:c.G2687A:p.R896H	nonsynonymous	0.0115	0.0052	0.0082	rs41301139	het	8.6663	2.573E-2
CLEC18B	16q23.1	32	exon12:c.G1288A:p.D430N	nonsynonymous	.	.	0.0001	rs616457	het	3.8842	1.363E-0
IQGAP2	5q13.3	32	exon11:c.C1169T:p.T390I	nonsynonymous	0.0153	0.0054	0.0170	rs34950321	het	8.3978	2.174E-3
PCDHA3	5q31.3	31	exon1:c.525_528del:p.T175fs	frameshift deletion	0.0084	0.0052	0.0095	rs141677026	het	9.1078	3.641E-2
CLIC5	6p21.1	31	exon3:c.C770A:p.P257H	nonsynonymous	0.0198	0.0080	0.0160	rs35822882	het	3.8150	2.159E-0

CYP4F12	19p13.12	31	exon7:c.C820T:p.R274W	nonsynonymous	0.0008	0.0032	0.0038	rs140750774	het	1.9135	7.211E-0
SGK2	20q13.12	30	exon11:c.C1045T:p.H349Y	nonsynonymous	0.0012	0.0002	0.0015	rs35793869	het	2.6070	4.301E-0

Table 4-19: MA26-KC (KC-1) WES and RNA results combination. Genes with coding variants detected in WES of MA26-KC (KC-1) cornea, with CADD Phred score > 30, differentially expressed (FDR < 0.05) by comparison with WT in analyses with both edgeR and Cufflinks, and listed in descending order of CADD score. logFC and FDR values outputted by edgeR are also given. The negative logFC values correspond to the downregulated genes in KC compared to WT. The positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the significant corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Gene	Location	Cadd Phred	AA change	Function	esp6500	1000g	exac03	snp138	Status	logFC	FDR
EXOC3L1	16q22.1	40	exon9:c.C1405T:p.R469X	Stopgain	0.002	0.005	0.003	rs141772271	het	1.3891	3.741E-02
ARF3	12q13.12	37	exon4:c.C295T:p.R99X	Stopgain	.	.	.	.	het	1.8946	1.907E-04
CLTCL1	22q11.21	36	exon15:c.C2409G:p.Y803X	Stopgain	.	0.0001	0.00007	.	het	11.5639	6.262E-24
ADAMTS14	10q22.1	35	exon6:c.C988T:p.R330C	nonsynonymous	0.002	0.002	0.003	rs41307534	het	5.5164	7.262E-19
CHRNA1	2q31.1	35	exon6:c.G686A:p.R229H	nonsynonymous	.	.	0.000008	.	hom	1.7616	7.652E-04
DNAH10	12q24.31	35	exon71:c.C12227T:p.P4076L	nonsynonymous	.	.	0.00001	.	het	6.6654	8.088E-25
EPHA2	1p36.13	35	exon15:c.G2627A:p.R876H	nonsynonymous	0.01	0.007	0.01	rs35903225	het	-6.3393	2.231E-21
HOXB9	17q21.32	35	exon2:c.G548A:p.R183H	nonsynonymous	.	0.01	0.005	rs201792149	het	11.2620	5.211E-21
ITGB5	3q21.2	35	exon14:c.G2290A:p.A764T	nonsynonymous	.	.	0.00001	.	het	2.7821	7.123E-05
QARS	3p21.31	35	exon21:c.G2036A:p.R679H	nonsynonymous	.	0.0001	0.0001	rs201781920	het	6.7922	1.345E-27
PCDHB4	8q23.1	35	exon59:c.9880delG:p.G3294fs	frameshift deletion	.	.	.	.	het	14.982	2.676E-65
CELA1	12q13.13	34	exon2:c.G53A:p.R18H	nonsynonymous	0.006	0.0027	0.003	rs74336876	het	10.2626	1.060E-13
ENO3	17p13.2	34	exon9:c.G1079A:p.R360H	nonsynonymous	.	0.001	0.001	.	het	10.7091	1.543E-16
FAT4	4q28.1	34	exon9:c.C7466T:p.A2489V	nonsynonymous	.	0.002	0.001	.	het	11.3575	7.591E-22
FGF6	12p13.32	34	exon3:c.G563A:p.R188Q	nonsynonymous	0.002	0.0003	0.002	rs71583765	het	3.2039	8.573E-07
MYOF	10q23.33	34	exon24:c.G2483T:p.G828V	nonsynonymous	.	.	.	.	het	-2.5609	1.929E-05

PLEKHG5	1p36.31	34	exon20:c.G2458A:p.G820S	nonsynonymous	.	0.001	0.001	rs202191898	het	10.6194	7.579E-16
SEMA5B	3q21.1	34	exon8:c.G761A:p.R254H	nonsynonymous	0.0003	0.0003	0.0002	rs184302192	het	-3.0496	2.429E-02
SULT6B1	2p22.2	34	exon6:c.G662T:p.R221L	nonsynonymous	0.01	0.01	0.005	rs11569761	hom	2.1783	8.268E-03
SVOPL	7q34	34	exon5:c.G224A:p.R75Q	nonsynonymous	0.001	0.002	0.001	rs144481050	het	-4.1790	5.034E-12
TUFM	16p11.2	34	exon8:c.G1001A:p.R334Q	nonsynonymous	.	.	0.00001	.	het	9.2492	2.816E-38
ABCA4	1p22.1	33	exon42:c.G5882A:p.G1961E	nonsynonymous	0.003	0.003	0.005	rs1800553	het	9.2492	2.816E-38
HPS5	11p15.1	33	exon6:c.C452T:p.S151L	nonsynonymous	.	.	0.000008	.	het	7.8064	6.103E-32
LYST	1q42.3	33	exon23:c.G6524T:p.C2175F	nonsynonymous	.	.	0.000008	.	het	5.4436	2.223E-19
MUC6	11p15.5	33	exon31:c.C5760A:p.Y1920X	stopgain	.	.	0.0001	rs75482640	het	-3.7400	6.860E-06
SIM1	6q16.3	33	exon10:c.G1427T:p.R476M	nonsynonymous	.	.	0.000008	.	het	5.0471	1.011E-14
FAAH	1p33	32	exon8:c.C1067T:p.A356V	nonsynonymous	0.0009	0.008	0.003	rs77101686	het	1.5615	1.431E-02
GRIK5	19q13.2	32	exon8:c.A935T:p.H312L	nonsynonymous	.	.	.	.	het	2.9500	3.285E-06
MAP2	2q34	32	exon4:c.G163A:p.E55K	nonsynonymous	0.009	0.004	0.009	rs139310749	het	4.7500	8.979E-12
QRICH2	17q25.1	32	exon15:c.C4618T:p.R1540W	nonsynonymous	0.001	0.0009	0.001	rs143270448	het	6.9976	5.273E-35
LIPH	3q27.2	31	exon4:c.C539G:p.A180G	nonsynonymous	0.00007	.	0.0002	rs138900382	het	2.2893	1.253E-05
NRG3	10q23.1	31	exon9:c.G1951A:p.E651K	nonsynonymous	0.002	0.002	0.004	rs138878772	het	10.3015	7.012E-14
BEST3	12q15	30	exon2:c.G34A:p.D12N	nonsynonymous	0.00007	.	0.0003	rs143236582	het	11.9206	8.038E-28
CORO2A	9q22.33	30	exon7:c.G862A:p.V288M	nonsynonymous	.	0.0003	0.00009	.	het	5.2176	8.361E-09

Table 4-20: MA46-KC (KC-2) WES and RNA results combination. Genes with coding variants detected in WES of MA46-KC (KC-2) cornea, with CADD Phred score > 30, differentially expressed (FDR < 0.05) by comparison with WT in analyses with both edgeR and Cufflinks, and listed in descending order of CADD score. logFC and FDR values outputted by edgeR are also given. The negative logFC values correspond to the downregulated genes in KC compared to WT. The positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the significant corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Gene	Location	Cadd Phred	AA change	Function	esp6500	1000g	exac03	snp138	Status	logFC	FDR
EXOC3L1	16q22.1	40	exon9:c.C1405T:p.R469X	stopgain	0.002	0.005	0.003	rs141772271	het	4.5691	3.308E-1

ARF3	12q13.12	37	exon4:c.C295T:p.R99X	stopgain	.	.	.	.	het	-7.5381	1.084E-2
CLTCL1	22q11.21	36	exon15:c.C2409G:p.Y803X	stopgain	.	0.0001	0.00007	.	het	2.6967	6.811E-0
ADAMTS14	10q22.1	35	exon6:c.C988T:p.R330C	nonsynonymous	0.002	0.002	0.003	rs41307534	het	2.1090	8.203E-0
CHRNA1	2q31.1	35	exon6:c.G686A:p.R229H	nonsynonymous	.	.	0.000008	.	hom	10.4161	6.120E-1
DNAH10	12q24.31	35	exon71:c.C12227T:p.P4076L	nonsynonymous	.	.	0.00001	.	het	9.4395	9.365E-4
EPHA2	1p36.13	35	exon15:c.G2627A:p.R876H	nonsynonymous	0.01	0.007	0.01	rs35903225	het	-4.1715	9.710E-0
HOXB9	17q21.32	35	exon2:c.G548A:p.R183H	nonsynonymous	.	0.01	0.005	rs201792149	het	9.5784	1.717E-1
ITGB5	3q21.2	35	exon14:c.G2290A:p.A764T	nonsynonymous	.	.	0.00001	.	het	-3.0929	2.665E-0
PCDHB4	8q23.1	35	exon59:c.9880delG:p.G3294fs	frameshift deletion	.	.	.	.	het	14.982	2.676E-6
QARS	3p21.31	35	exon21:c.G2036A:p.R679H	nonsynonymous	.	0.0001	0.0001	rs201781920	het	-3.8537	2.098E-1
RFX3	9p24.2	35	exon2:c.C107G:p.S36X	Stopgain	.	.	.	.	hom	4.3896	1.625E-1
CELA1	12q13.13	34	exon2:c.G53A:p.R18H	nonsynonymous	0.006	0.002	0.003	rs74336876	het	9.9630	2.421E-1
ENO3	17p13.2	34	exon9:c.G1079A:p.R360H	nonsynonymous	.	0.001	0.001	.	het	5.5094	5.202E-2
FAT4	4q28.1	34	exon9:c.C7466T:p.A2489V	nonsynonymous	.	0.002	0.001	.	het	7.1999	3.523E-3
FGF6	12p13.32	34	exon3:c.G563A:p.R188Q	nonsynonymous	0.002	0.0003	0.002	rs71583765	het	9.8691	6.936E-1
MYOF	10q23.33	34	exon24:c.G2483T:p.G828V	nonsynonymous	.	.	.	.	het	-1.7159	1.805E-0
PLEKHG5	1p36.31	34	exon20:c.G2458A:p.G820S	nonsynonymous	.	0.001	0.001	rs202191898	het	-3.1372	1.471E-0
SEMA5B	3q21.1	34	exon8:c.G761A:p.R254H	nonsynonymous	0.0003	0.0003	0.0002	rs184302192	het	4.0324	2.880E-0
SULT6B1	2p22.2	34	exon6:c.G662T:p.R221L	nonsynonymous	0.01	0.01	0.005	rs11569761	hom	10.0014	1.590E-1
SVOPL	7q34	34	exon5:c.G224A:p.R75Q	nonsynonymous	0.001	0.002	0.001	rs144481050	het	5.8807	8.508E-1
TUFM	16p11.2	34	exon8:c.G1001A:p.R334Q	nonsynonymous	.	.	0.00001	.	het	-3.8814	1.576E-0
VPS13D	1p36.22	34	exon18:c.C2225T:p.T742M	nonsynonymous	0.00007	0.0001	0.00009	rs370784596	het	0.9944	4.361E-0
ABCA4	1p22.1	33	exon42:c.G5882A:p.G1961E	nonsynonymous	0.003	0.003	0.005	rs1800553	het	7.3277	6.235E-3
HPS5	11p15.1	33	exon6:c.C452T:p.S151L	nonsynonymous	.	.	0.000008	.	het	1.9566	4.309E-0
LYST	1q42.3	33	exon23:c.G6524T:p.C2175F	nonsynonymous	.	.	0.000008	.	het	2.8186	9.247E-0
MUC6	11p15.5	33	exon31:c.C5760A:p.Y1920X	stopgain	.	.	0.0001	rs75482640	het	10.6192	3.118E-1
PLEKHG2	19q13.2	33	exon18:c.G2308A:p.A770T	nonsynonymous	0	0.004	0.006	rs111487768	het	1.2152	2.979E-0
SIM1	6q16.3	33	exon10:c.G1427T:p.R476M	nonsynonymous	.	.	0.000008	.	het	11.2230	5.050E-2

FAAH	1p33	32	exon8:c.C1067T:p.A356V	nonsynonymous	0.0009	0.008	0.003	rs77101686	het	1.6595	8.508E-0
GRIK5	19q13.2	32	exon8:c.A935T:p.H312L	nonsynonymous	.	.	.	.	het	5.1611	5.996E-1
MAP2	2q34	32	exon4:c.G163A:p.E55K	nonsynonymous	0.009	0.004	0.009	rs139310749	het	4.0142	1.450E-1
QRICH2	17q25.1	32	exon15:c.C4618T:p.R1540W	nonsynonymous	0.001	0.0009	0.001	rs143270448	het	4.7309	6.931E-1
LIPH	3q27.2	31	exon4:c.C539G:p.A180G	nonsynonymous	0.00007	.	0.0002	rs138900382	het	-2.0507	1.185E-0
NRG3	10q23.1	31	exon9:c.G1951A:p.E651K	nonsynonymous	0.002	0.002	0.004	rs138878772	het	12.5978	9.644E-3
BEST3	12q15	30	exon2:c.G34A:p.D12N	nonsynonymous	0.00007	.	0.0003	rs143236582	het	11.3476	3.287E-2
CORO2A	9q22.33	30	exon7:c.G862A:p.V288M	nonsynonymous	.	0.0003	0.00009	.	het	-2.7944	4.694E-0

Table 4-21: MA47-KC (KC-3) WES and RNA results combination. Genes with coding variants detected in WES of MA47-KC (KC-3) cornea, with CADD Phred score > 30, differentially expressed (FDR < 0.05) by comparison with WT in analyses with both edgeR and Cufflinks, and listed in descending order of CADD score. logFC and FDR values outputted by edgeR are also given. The negative logFC values correspond to the downregulated genes in KC compared to WT. The positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the significant corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Gene	Chr	Cadd Phred	AA change	Function	esp6500	1000g	exac03	snp138	Status	logFC	FDR
MELK	9p13.2	37	exon15:c.T1389A:p.Y463X	Stopgain	.	.	.	.	het	3.0609	1.135E-0
DNAH14	1q42.12	35	exon21:c.3392delA:p.E1131fs	frameshift deletion	0.00316	0.003	0.00540	.	het	6.9016	4.035E-3
MYOF	10q23.33	35	exon43:c.4915delC:p.H1639fs	frameshift deletion	.	.	.	.	het	-1.3893	1.252E-0
RAPGEFL1	17q21.1	35	exon15:c.C1297T:p.R433C	Nonsynonymous	.	.	0.00001	.	het	-2.4690	4.338E-0
ABCC8	11p15.1	34	exon32:c.G3938A:p.R1313H	Nonsynonymous	0.00008	.	0.00002	rs372153432	het	12.8512	6.087E-3
CBLC	19q13.32	34	exon5:c.C790T:p.R264W	Nonsynonymous	.	.	.	.	het	-2.0362	2.011E-0
DIXDC1	11q23.1	34	exon7:c.G644T:p.R215L	Nonsynonymous	.	0.0014	0.00069	.	het	1.2164	3.380E-0
FAT4	4q28.1	34	exon9:c.C7466T:p.A2489V	Nonsynonymous	.	0.0022	0.00178	.	het	6.9778	2.017E-3
OBSCN	1q42.13	34	exon52:c.G13765A:p.G4589R	Nonsynonymous	.	0.0002	0.00016	.	het	2.4176	1.923E-0

PRB2	12p13.2	34	exon3:c.C511T:p.R171X	Stopgain	0.00008	.	0.00036	rs201606820	het	6.0079	2.441E-0
LAMB1	7q31.1	33	exon32:c.G4940A:p.R1647H	Nonsynonymous	.	0.0006	0.00015	.	het	-1.3644	2.971E-0
MUC6	11p15.5	33	exon31:c.C5760A:p.Y1920X	Stopgain	.	.	0.00018	rs75482640	het	10.6229	9.366E-1
ZNF407	18q22.3	33	exon8:c.G5702A:p.R1901Q	Nonsynonymous	0.00008	0.0004	0.00018	rs201662777	het	2.9746	1.021E-0
EXOC3L1	16q22.1	32	exon5:c.C640T:p.R214W	Nonsynonymous	.	.	0.00007	.	het	4.9289	1.092E-1
MAP7D1	1p34.3	32	exon2:c.C310T:p.R104W	Nonsynonymous	0.00046	0.0144	0.00629	rs2296266	het	-3.3991	9.654E-0
TJP1	15q13.1	32	exon5:c.G353T:p.R118L	Nonsynonymous	0.00017	0.0008	0.00055	rs148793257	het	-2.4484	4.939E-0
PDIA2	16p13.3	31	exon9:c.1388_1389del:p.T463fs	frameshift deletion	0.00637	0.004	0.00645	rs201624048	het	7.6983	1.135E-1
ABCA12	2q35	31	exon8:c.C1013T:p.P338L	Nonsynonymous	0.00008	0.0004	0.00024	rs374169066	het	2.0576	2.614E-0
ATP12A	13q12.12	31	exon17:c.C2373A:p.D791E	Nonsynonymous	.	0.0010	0.00044	rs369210106	het	9.7798	1.695E-3
MRPL57	13q12.11	31	exon2:c.C109T:p.R37C	Nonsynonymous	0.00823	0.0040	0.00822	rs117575190	het	-3.3080	1.449E-0
SH3BP1	22q13.1	31	exon14:c.G1243A:p.V415M	Nonsynonymous	.	.	0.00002	.	het	-2.6176	1.612E-0
PREPL	2p21	30	exon13:c.T1967C:p.F656S	Nonsynonymous	.	.	.	.	het	1.7669	1.773E-0

Table 4-22: MA52-KC (KC-4) WES and RNA results combination. Genes with coding variants detected in WES of MA52-KC (KC-4) cornea, with CADD Phred score > 30, differentially expressed (FDR < 0.05) by comparison with WT in analyses with both edgeR and Cufflinks, and listed in descending order of CADD score. logFC and FDR values outputted by edgeR are also given. The negative logFC values correspond to the downregulated genes in KC compared to WT. The positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the significant corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Gene	Location	Cadd Phred	AA change	Function	esp6500	1000g	exac03	snp138	Status	logFC	FDR
MAP3K15	Xp22.12	40	exon24:c.C3364T:p.R1122X	Stopgain	0.001	.	0.001	rs140104197	het	10.505	2.416E-31
ADAP2	17q11.2	37	exon6:c.C606A:p.Y202X	Stopgain	.	.	.	.	het	3.1543	2.513E-08
TMEM184A	7p22.3	36	exon2:c.C166T:p.R56X	Stopgain	.	.	0.000	.	het	-4.206	6.460E-12
ISLR2	15q24.1	35	exon3:c.1821_1834del:p.P607fs	frameshift deletion	.	.	.	.	het	8.2956	5.038E-20

MYO18B	22q12.1	35	exon43:c.7625delC:p.S2542fs	frameshift deletion	.	.	.	.	het	13.699	4.673E-47
LCNL1	9q34.3	35	exon3:c.C217T:p.R73X	Stopgain	0.002	0.0004	0.001	rs201231448	het	4.7379	4.028E-07
NINL	20p11.21	35	exon6:c.A562T:p.K188X	Stopgain	.	.	.	.	het	2.6532	4.719E-05
ZNF341	20q11.22	35	exon14:c.C1810T:p.R604C	nonsynonymous	.	0.0002	0.000	.	het	2.5897	1.731E-06
PCSK6	15q26.3	34	exon11:c.G1381A:p.D461N	nonsynonymous	0.0002	.	0.000	rs369879412	het	1.7383	2.874E-03
PDE6A	5q32	34	exon5:c.C878T:p.P293L	nonsynonymous	0.0036	0.0018	0.003	rs114973968	het	2.9627	4.899E-08
PGA5	11q12.2	34	exon9:c.G1066T:p.E356X	Stopgain	.	.	.	.	hom	9.6352	1.892E-10
XDH	2p23.1	34	exon35:c.C3886T:p.R1296W	nonsynonymous	0.0131	0.0068	0.014	rs45564939	het	-3.209	2.097E-08
CSRNP3	2q24.3	33	exon4:c.C671T:p.T224M	nonsynonymous	0.0001	.	0.000	rs371859963	het	5.7549	2.549E-11
FLNB	3p14.3	33	exon35:c.G6001A:p.V2001M	nonsynonymous	.	.	.	.	het	-1.558	6.565E-03
KIF20A	5q31.2	33	exon19:c.C2516T:p.P839L	nonsynonymous	0.0113	0.0054	0.011	rs3172747	het	5.4441	5.522E-18
MUC6	11p15.5	33	exon31:c.C5760A:p.Y1920X	Stopgain	.	.	0.000	rs75482640	het	10.850	1.823E-17
ABCG4	11q23.3	32	exon11:c.G1184A:p.R395Q	nonsynonymous	0.0002	.	0.000	rs149710791	het	5.8140	2.679E-21
ANKRD45	1q25.1	32	exon5:c.G634A:p.E212K	nonsynonymous	0.0048	0.0016	0.003	rs34387166	het	5.0854	1.890E-11
LAMC3	9q34.12	32	exon19:c.C3371T:p.S1124F	nonsynonymous	0.0095	0.0032	0.007	rs113259170	het	7.4532	4.396E-29
NOL6	9p13.3	32	exon6:c.C787T:p.R263C	nonsynonymous	0.0002	.	0.000	rs116746143	het	-1.749	4.317E-03
PRSS12	4q26	32	exon8:c.G1583A:p.G528E	nonsynonymous	.	.	.	.	het	3.5565	4.068E-09
SH3BP2	4p16.3	32	exon11:c.C1429T:p.R477W	nonsynonymous	0.0032	0.0042	0.005	rs148761331	het	-3.779	3.431E-10
RTN4RL1	17p13.3	31	exon2:c.G439A:p.G147S	nonsynonymous	0.0111	0.0048	0.008	rs181444163	het	4.1425	2.787E-06

Table 4-23: MA58-KC (KC-5) WES and RNA results combination. Genes with coding variants detected in WES of MA58-KC (KC-5) cornea, with CADD Phred score > 30, differentially expressed (FDR < 0.05) by comparison with WT in analyses with both edgeR and Cufflinks, and listed in descending order of CADD score. logFC and FDR values outputted by edgeR are also given. The negative logFC values correspond to the downregulated genes in KC compared to WT. The positive logFC values correspond to the upregulated genes in KC corneas compared to WT. The FDR (false discovery rate) values are the significant corrected p-values calculated by edgeR using the Benjamini-Hochberg method.

Gene	Location	Cadd Phred	AA change	Function	esp6500	1000g	exac03	snp138	Status	logFC	FDR
SERPINA12	14q32.13	40	exon12:c.C1027T:p.R343C	Stopgain	0.0055	0.0016	0.00526	rs61757459	het	11.0433	1.105E-18
STAP2	19p13.3	38	exon4:c.1566delG:p.E522fs	Stopgain	0.0082	0.0040	0.00988	rs79657645	het	-2.8014	3.394E-06
NAA38	17p13.1	37	exon12:c.C1984T:p.R662C	Stopgain	0.0095	0.0040	0.00895	rs144163075	het	-1.8149	5.299E-03
OR52L1	11p15.4	36	exon5:c.G674A:p.R225H	Stopgain	0.0003	0.0026	0.00499	rs186343423	het	9.9348	7.873E-12
ITGB3	17q21.32	35	exon3:c.C220T:p.R74W	Nonsynonymous	.	0.0002	0.00001	rs151219882	het	6.4788	1.627E-05
LAMA1	18p11.31	35	exon31:c.C5760A:p.Y1920X	Nonsynonymous	0.0072	0.0022	0.00642	rs140718292	het	5.9761	1.032E-16
MTHFSD	16q24.1	35	exon1:c.G168A:p.W56X	Nonsynonymous	.	0.0002	0.00002	rs36053000	het	1.5067	1.611E-02
SLX4IP	20p12.2	35	exon1:c.819delG:p.L273fs	Nonsynonymous	0.0003	0.0056	0.00327	rs78689061	het	2.6782	3.555E-04
SYT6	1p13.2	35	exon1:c.G890A:p.W297X	Nonsynonymous	.	.	0.00001	.	het	11.5635	2.029E-23
TTI2	8p12	35	exon12:c.G1891A:p.V631M	Nonsynonymous	0.0038	0.0016	0.00308	rs138108276	het	1.5734	9.210E-03
ARGFX	3q13.33	34	exon3:c.C631T:p.R211X	Stopgain	.	.	.	.	het	10.5894	1.620E-15
ARMC9	2q37.1	34	exon5:c.C289T:p.R97C	Nonsynonymous	.	.	0.00001	.	het	4.5931	1.433E-13
ADAM19	5q33.3	33	exon6:c.C507G:p.Y169X	Nonsynonymous	0.0003	0.0004	0.00027	rs146453088	het	3.8431	1.051E-10
MUC6	11p15.5	33	exon5:c.G974A:p.R325Q	Stopgain	.	.	0.00018	rs75482640	het	11.3245	3.328E-21
TFR2	7q22.1	33	exon15:c.G1742A:p.R581H	Nonsynonymous	0.0201	0.0082	0.01900	rs41295942	het	4.0148	1.280E-12
ANGPTL6	19p13.2	32	exon4:c.C1063T:p.R355C	Nonsynonymous	.	0.0038	0.00132	.	het	6.4644	1.923E-08
UGGT2	13q32.1	32	exon33:c.A3802T:p.N1268Y	Nonsynonymous	0.0229	0.0082	0.01900	rs45473699	het	1.2701	3.601E-02
PLA2R1	2q24.2	31	exon2:c.33_34del:p.V11fs	Nonsynonymous	0.0010	0.0008	0.00054	rs146287694	het	1.4065	1.317E-02

4.2.4.2 Significant DE genes harbouring changes in all 6 individual KC corneas

In addition to checking for variants in genes that are significantly differentially expressed in each individual KC cornea, the lists of variants in DE genes for all 6 KC corneas were compared to look for shared DE genes containing variants in the WES data profiles generated from blood DNA of all KC corneas. The results of this analysis are presented in Table 4-24.

Table 4-24: Genes containing variants in the WES all of the 6 KC corneas and being significantly differentially expressed. This table shows the genes, which were significantly differentially expressed between KC and WT corneas and in which WES analysis showed variants in all 6 KC corneas.

Gene symbol	Gene name	Function/Related pathway
HRNR	Hornerin	Calcium ion binding
MUC6	Mucin 6, oligomeric mucus/gel-forming	Extracellular matrix structural constituent
MUC2	Mucin 2, oligomeric mucus/gel-forming	Protein binding
ATXN8OS	ATXN8 opposite strand(ncRNA)	Antisense transcript to KLHL1 gene
LINC00910	Long intergenic non-protein coding RNA910 (ncRNA)	-
ZNF738	zinc finger protein 738	Ubiquitin-protein transferase activity and DNA binding
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	mRNA splicing process, transcription regulatory region DNA binding
ANKRD36BP2	Ankyrin repeat domain 36B pseudogene 2	-
MCM3AP-AS1	MCM3AP antisense RNA 1	-
LINC00955	Long intergenic non-protein coding RNA 955	-
FAT4	FAT atypical cadherin 4	Calcium ion binding and protein binding, cell-cell adhesion pathway
FDFT1	Farnesyl-diphosphate farnesyltransferase 1	Oxidoreductase activity, oxidation-reduction and steroid biosynthetic process
GXYLT1P3	Glucoside xylosyltransferase 1 pseudogene 3	-

4.3 Discussion

Over the years, human genetic research has enabled the significant advancement of the understanding of many molecular aspects underlying human inherited disorders. However, most of the human genetic disease studies, and in particular, the studies investigating complex diseases i.e. GWAS variants do not generally provide relevant clues to the contribution of the associated genetic variations or the functional impact of the DNA variation to disease onset. In addition finding the causative variations for these complex diseases is not an easy task in the first place (Costa *et al.*, 2013, Pouladi *et al.*, 2015). Thus, global gene expression-transcriptomics studies were adopted as a powerful tool for providing functional insights into the secondary disease process of many inherited diseases, as well as investigating the primary pathogenesis of complex and common genetic conditions (Morley *et al.*, 2004, Twine *et al.*, 2011).

Transciptome studies have progressed from hybridization methods including microarrays, to tag methods (Horan, 2009, Wang *et al.*, 2009). However, RNA-seq presents many advantages over other methods (reviewed by (Wang *et al.*, 2009, Garber *et al.*, 2011)), consequently in the recent years, RNA-Seq has become the method of choice for transcriptomic studies. Moreover, RNA-seq gives researchers the opportunity to characterize the sequences and quantify all the RNAs present in a sample as well as the discovery of new transcripts and splicing events (Nagalakshmi *et al.*, 2008, Wang *et al.*, 2009, Trapnell *et al.*, 2012, Finotello and Di Camillo, 2015).

In this study, RNA was extracted from the anterior section (Epithelium, Bowman's Layer, and Stroma) of 5 WT and 6 KC corneas, and the transcriptome data for each cornea was generated. Using the RNA-seq data generated, three main areas of study were explored: 1- the RNA profile of healthy human corneas; 2- the differential expression between the RNA profile of KC vs. WT corneas; and 3- the individual RNA profile of each KC

cornea, considered in combination with the WES data of the same KC cornea investigated.

For all the mentioned strategies, RNA-seq paired-end reads were mapped using the Tophat software (Kim *et al.*, 2013) and aligned to a reference genome (Gencode GRCh38p2), but not a reference transcriptome, allowing the discovery of novel transcripts; however the reference transcriptome (GRCh38p2) was used as an annotation guide. Tophat was the aligner of choice in this study since one of the advantages it offers over other available software packages is the gap-alignment function, which allows the alignment of reads containing indels (Kim *et al.*, 2013). In addition, Tophat is a spliced-read mapper, which, after mapping the reads to the genome, performs the analyses of the mapping results, in order to build all possible exon-exon junctions for splice junction identification (Trapnell *et al.*, 2012).

4.3.1 The RNA profile of healthy human corneas.

As previously described in Figure 4-1, the RNA-seq data of the WT corneas was generated in order to observe a healthy anterior corneal gene expression profile. Thus, Cuffnorm was used only on the merged transcripts of WT corneas, in order to normalize the expression levels among all the replicates by applying an additional step of normalization for library size. Cuffnorm reports FPKM and normalized estimates of the number of fragments for each class (genes, TSS group, transcript and CDS group) which basically accounts for differences in library size (Trapnell *et al.*, 2012). In this Cuffnorm analysis, the library size factor used was 1 by applying the default cufflinks settings, which suggest multiplying the gene read counts for each sample by 1 and multiplying it by the mean read counts of the same gene across all samples (Aleksic *et al.*, 2014). FPKM and library size present challenges and introduce some biases to the data (Zheng *et al.*, 2011, Aleksic *et al.*, 2014). FPKM can be biased in reporting expression values for genes expressed at low levels and presenting small transcripts (Dillies *et al.*, 2013). This is because small transcripts can present low read counts but when accounting for length biases

the library normalized value would be multiplied by the exon length of each transcript, leading to an overrepresentation of the actual biological significance of the expression (Aleksic *et al.*, 2014). The library size normalization may be a source of bias only in case of the presence of large differences in a particular set of mRNAs between samples.(Robinson and Oshlack, 2010).

4.3.1.1 The RNA profile results

After running Cuffnorm, the tab-delimited “gene.fpkm_tracking” was the file of interest for this study which represented the normalised FPKM of transcripts sharing the same gene-id. The averaged FPKM values were sorted according to expression levels from high to low and mitochondrial genes (mt-genes) were removed from the final list. The exclusion of the mt-genes risks missing biologically important observations, since their expression might be prominent in corneal tissues and removing these genes from the expression data might be introducing bias. However their high FPKM values do not necessarily correlate with their high expression. Indeed, it is not really clear whether their high FPKM values are due to the biological overexpression of these particular genes in corneal tissues or to the overrepresentation of mitochondrial DNA in the cells. This will be overcome in the future work of this project, by running RNA-seq analysis on several other different tissues than cornea and comparing mitochondrial gene expression levels in different tissues.

The pathways and functional structures that are most enriched in normal human corneas, as presented in Table 4-2, are ribonucleoproteins, translational elongation, protein biosynthesis, oxidative phosphorylation, mitochondrial structure/function, the cytoskeleton, mRNA processing and RNA splicing.

There are several considerations to take into account while interpreting this data. First, the pathway analysis can be biased due to the fact that only the first 3000 protein coding genes can be uploaded to DAVID, risking the

omission/overrepresentation/misrepresentation of crucial pathways involved in human corneal function or structure. Second, none of these data have been confirmed in the lab by RT-PCR, and that specific point is one of the future directions of this project (Section 5.2.3). Third and most importantly, the corneal profiles from these donors might not be totally representative of the *in-vivo* corneal mRNA profile. The enucleation procedure from the donor's eye takes place as soon as possible to a maximum of 24 hours after death, so the time between enucleation and preservation may vary quite significantly between samples and may extend to a full 24 hours. In addition, after enucleation, the tissue from the donor is preserved in an EAGLES 2% FBS medium (Section 2.3) at 34 degrees a period of time that can go no longer than 28 days. Then in the theatre, while performing DSEK (Section 2.3) the tissue remains in open air for at least 20 minutes before disposing it into RNAlater (Section 2.3). While undergoing all of this procedure, the mRNA is prone to degradation. However this degradation is expected to be minimal, since this same procedure is followed for the tissues that are extracted for transplantation. Thus, the difference in the RNA profiles of KC and WT corneas in Figures 4-4 and 4-5 might be due to a combination of difference in freshness in both tissues as well as the manifestation of the disease.

4.3.1.2 Discovery of potential novel genes/exons

Note that replication and confirmation of these results in the lab by RT-PCR was not carried out before writing this thesis, but is an essential next step in the future direction of this project.

In addition to annotating transcripts expressed in the normal cornea, these data facilitated a search for potential novel cornea-specific isoforms of existing genes or novel genes. Following the protocol described by Farkas et al. (Farkas *et al.*, 2013), a new alignment algorithm, the RNA-Seq Unified Mapper (RUM), was used for novel gene and exon discovery.

In order to pursue that goal, the RNA-seq data was aligned on RUM, which is a different aligner to Tophat (Section 4.2.2.2). In addition to RUM being an aligner implemented in Perl, it is a junction calling and feature quantification software which makes it able to accurately identify novel splice variants, including splicing events and novel junctions with low read depth (Grant *et al.*, 2011). Combining information on novel junctions with the genome browser UCSC, leads to the identification of the location of the new junctions, which sometimes consists in the identification of a new exon or a new gene. The alignment pipeline of RUM consists of mapping reads in three phases. First, it uses Bowtie -based on the Burrows–Wheeler algorithm- to map the reads against the hg19 version of the genome, then in the second step it maps the reads against the transcriptome database and finally in the third step the non-mapped reads by Bowtie are mapped against the hg19 version of the genome by BLAT. Afterwards, the results from the three mappings are merged together. Thus, the RUM alignment approach improves the quality of mapping by combining the speed of Bowtie with the sensitivity of BLAT.

The list of the novel findings in genes or exons represented in Table 4-4 represents the canonical novel junctions (GT-AG splice site pairs). The novel junctions mapping within existing genes can be divided into four clusters: 3 novel junctions genes (*MECOM*), 2 novel junctions genes (*MFN2*), 1 novel junction genes (*GBA3, PLEKA7, NOX1, AK023515, ERCC1, SITM1* and *RPS6KA2*) and novel junctions in specified regions of the genome that don't fall into known genes regions (6 novel junctions). In the known genes, these novel junctions might be indicative of novel exons. However, the novel junctions that fall in un-annotated regions of the genome might be indicative of novel genes. According to the Table 4-4 results, the novel genes might be composed of two exons only (1 junction), unless there are other non-canonical junctions (GC-AG splice site) contributing to the splice sites of other exons composing the genes. Further analyses are in progress in order to answer for the speculations postulated in this section and section 4.2.2.2.

The results of Table 4-4 need to be verified in the lab by RT-PCR and looked at in different tissues in order to confirm the existence of these features and to determine whether these are novel findings unique to the cornea only or whether similar findings occur in different tissues.

4.3.2 Sample assessment and differential expression analysis

DE analyses were first conducted by running the Cufflinks package (4.2.3.2) on all the 6 KC and 5 WT corneas, and visualized using the plotting tool, part of the Tuxedo suite pipeline CummeRbund (Trapnell *et al.*, 2012). The initial visualization of the data (Figures 4-4, 4-5) served as an evaluation of the quality of the RNA-seq data of the samples used, both WT and KC, and as a quality control step which revealed a heterogeneous distribution of FPKM values among samples: samples KC_3 for the Keratoconus condition and WT_3 and WT_4 of the wild-type corneas condition were tagged as outliers in the biological replicates of each condition.

Since this is a new project and working out how to deal better with data of different quality status is still an on-going task, it has been (perhaps naively) decided to remove the samples that were revealed to be outliers and did not uniformly conform with the profiles of the other samples. It is known with RNA-seq data that the more replicates used, the better the outcome of the data is. However, a conservative approach was taken in this respect while conducting this project. 2 samples have been removed from the WT samples (WT_3 and WT_4) to drop the number of samples from 5 to 3 which is still acceptable in terms of number of biological replicates, and one KC sample (KC_3) has been removed as well to drop the number of samples from 6 to 5.

Another subject that catches the attention while doing the quality control assessment is the differences in the profiles between the KC and WT corneas. The RNA seq data profiles emerging from the two different

conditions, KC and WT, are very different even though they derive from the same type of tissue (anterior cornea).

The differences might be due to: (1) the method of handling the tissues during collection. For KC corneas, the tissues were surgically removed in theatre and directly preserved in RNA later Ambion solution. As for the WT corneas, the tissues were extracted as soon as possible but up to no longer than 24 hours after death. Hence, the low and larger range of FPKM values (presented in a wider base in Figure 4-5 a,c) in the RNA profile of WT corneas might be due to the degradation of RNA caused by the delay in the tissue extraction of the WT corneas. However, this degradation is thought to be minimal since this same procedure of extraction of WT tissues is used to extract tissues for research as well as for transplantation (Figueiredo, 2013). Thus the WT tissues for donation, extracted through the same protocol, are still biologically viable. (2) The difference in tissue organisation between KC and WT caused by the disease progression. It is well known that KC corneas show keratocyte apoptosis and major changes in the structure of the cornea (Described in details in Section 1.4.5). Thus, these changes might be produced by the actual differences in molecules composition between KC and WT corneas, which are basically reflected by the changes in mRNA levels presented in Section 4.2.3.2.

All things considered, these changes in profiles between KC and WT corneas, are most likely due to the a mixture of both factors: slight RNA degradation due to the handling of the corneal tissues in addition, it is due to the actual differences between both KC and WT tissues caused by the development of the disease.

4.3.3 Differential expression analyses

Since there is no clear consensus about the best pipeline or method for DE yet (Seyednasrollah *et al.*, 2015), two methods have been used in this project

for DE: these are the Cuffdiff2 part of the Cufflinks package (Trapnell *et al.*, 2013) and the bioconductor package edgeR (Robinson *et al.*, 2010) preceded by Subreads (Liao *et al.*, 2013) for creating a counts matrix.

The previously mentioned software, Cufflinks/Cuffdiff on one side and Subreads/edgeR on the other side, present differences in the processing of the RNA mapped reads at many levels: read counts, normalization and differential expression analysis.

In general, the most widely adopted approach for calculating read counts is summing up the total number of reads overlapping the exons of a gene (Finotello and Di Camillo, 2015). For Cufflinks, counting the fragments or paired end reads is modelled in order to account for the abundance of different splicing isoforms of a gene (Trapnell *et al.*, 2010). The RNA counts are normalized for library size as well as gene length bias by using the FPKM (Fragments Per Kilobase of exon per Million Fragments mapped) method. According to Mortazavi et al., FPKM values are computed by “dividing the number of fragments aligned to gene exons by the total number of mapped reads and by the sum of exonic bases” (Mortazavi *et al.*, 2008) it also estimates fragment abundances in relation to the estimated number of fragments (Trapnell *et al.*, 2010).

As for the edgeR approach, the counting of the reads was done by the command FeatureCount of the Rsubread package (Liao *et al.*, 2013) where only the fragments that have both ends mapped were considered for counting. The read counts are then normalised by edgeR using the TMM (Trimmed Mean of M-values) normalization method, which takes into account the high count genes in terms of sample-specific effect, a bias induced by the sequencing depth of a sample, and accounts for library composition differences (i.e. sequencing depth) between samples (Robinson and Oshlack, 2010). It has been shown to remove up to 30% of the genes of extreme M-

values i.e. fold-change between samples to correct for library size differences (Robinson *et al.*, 2010).

As for differential expression analysis, both tools use a negative binominal parametric model for their ability to capture technical and biological variation (Robles *et al.*, 2012, Seyednasrollah *et al.*, 2015). Cuffdiff by default uses a beta negative binomial model for fragment counts in order to control for variability and read mapping ambiguity and minimise the count data overdispersion (Trapnell *et al.*, 2012). As for edgeR, the negative binominal model is again used in the gene-wise dispersion estimates shrinking towards a “trend” estimated value across genes instead of a single value in order to minimise data overdispersion (Robinson *et al.*, 2010). The negative binominal based methods have been shown to present higher performance in dealing with differential expression than other tools (i.e. Poisson-based models) (Zhang *et al.,* Di *et al.*, 2011, Kvam *et al.*, 2012, Zhang *et al.*, 2014).

The previously presented technical differences between the two software packages can be interpreted by the ability of Cufflinks to identify the abundance of different transcripts within a sample and the difference in transcript expression, in addition to identifying novel transcripts through assembly. In addition, it provides information on differentially spliced genes or genes undergoing other types of isoform-level regulations (Trapnell *et al.*, 2012). As for edgeR, it is interpreted by the ability to identify if the observed differences in total expression of known transcripts between the two conditions (KC and WT corneas) can be attributed significantly to the experimental condition alone and not due to the biological variability (Robinson *et al.*, 2010). This method represents an advantage over the other method of differential expression analysis, since it simplifies the technical influences that need to be taken into account during the analysis course (e.g. the length bias) (Robinson *et al.*, 2010).

After briefly overviewing the differences between edgeR and Cuffdiff it is not surprising to observe differences in the DE results outputted by the two software packages. However since no particular RNA seq method can be optimal under all circumstances (Seyednasrollah *et al.*, 2015), and knowing the advantages and disadvantages of both methods used in this project (Cuffdiff and edgeR), it was decided to only take into consideration the genes that appeared to be differentially expressed in both analyses. In parallel, the Cuffdiff method was used to detect genes that were totally absent in KC corneas compared to WT and vice versa, and later on in the future of the project it can be used for the detection of novel isoforms and splicing events.

4.3.4 Differential expression results

It is important to note that the transcriptomics project discussed in this chapter remains an on-going project and the results presented are preliminary and need more exploration and further validation.

The main focus in this section was to study the genes that were highly significantly downregulated in pooled KC corneas (Table 4-12) as well as the pathway analysis of the downregulated genes and the overall differentially expressed genes in pooled KC corneas. EdgeR was able to pick up mitochondrial genes that Cufflinks wasn't able to detect (Tables 4-10 and 4-25) (MT-CO2, MT-CO3, MT-ATP6, MT-ND4, MT-RNR2). The downregulation of mt-genes may be read in two ways.

Firstly it is plausible that mitochondria may be involved in KC pathogenesis. The downregulation of mt-genes in KC corneas might be due to mtDNA mutations, which have previously been reported in KC corneas (Abu-Amero *et al.*, 2014b, Hao *et al.*, 2015b) and might be the result of mitochondrial dysfunction (Chwa *et al.*, 2008). Both mtDNA damage – which has been shown to be increased in KC (Atilano *et al.*, 2005) - and mt-dysfunction in KC corneas (Chwa *et al.*, 2008, Ishii *et al.*, 2013) might be related to the elevation

of oxidative stress in KC corneas (Chwa *et al.*, 2006, Arnal *et al.*, 2011, Wojcik *et al.*, 2013a, Wojcik *et al.*, 2013b, Karamichos *et al.*, 2014, Toprak *et al.*, 2014, Goncu *et al.*, 2015). In fact, the downregulation of MT-RNR2 detected by edgeR in this study (Tables 4-10 and 4-25), if confirmed by RT-PCR, would be a confirmation of the elevation of oxidative stress in KC corneas, since it has been reported by Crawford and co-workers that the downregulation of 16S rRNA (product of MT-RNR2) is specific to mitochondria in response to stress and is a protection mechanism of cells against oxidative stress (Crawford *et al.*, 1997).

An alternative explanation for the apparent downregulation of mt-genes in KC corneas is that it might be due to the decreased number of mitochondria in KC corneas compared to WT, which may in turn be the result of the decreased number of cells present in the tissue due to keratocyte apoptosis in KC (Kim *et al.*, 1999, Kaldawy *et al.*, 2002, Matthews *et al.*, 2007, Synowiec *et al.*, 2014).

However, either phenomenon, if proven, could be accounted for by elevation of oxidative stress, causing mitochondrial DNA damage and the apoptosis of KC fibroblasts (Section 1.4.7.2) (Chwa *et al.*, 2006).

Table 4-25: Mitochondrial genes found in edgeR analysis of RNAseq data to be differentially expressed between pooled KC and pooled WT corneas.

Symbol	Feature	logFC	FDR
MT-CO2	protein_coding	-11.36332101	5.9962E-269
MT-CO3	protein_coding	-11.242322	1.4741E-265
MT-ATP6	protein_coding	-11.14214094	2.147E-259
MT-ND4	protein_coding	-10.50619685	5.7367E-246
MT-RNR2	Mt_rRNA	-10.39711274	3.6219E-242
MT-CO1	protein_coding	-10.17871168	1.3102E-236
MT-ND1	protein_coding	-10.14822195	3.0305E-234
MT-ND2	protein_coding	-10.13405602	1.0652E-232
MT-ND4L	protein_coding	-10.42501571	2.2202E-229
MT-ATP8	protein_coding	-10.64143114	1.0031E-227
MT-CYB	protein_coding	-9.569938616	4.0729E-218
MT-RNR1	Mt_rRNA	-9.371262465	3.8672E-210

MT-ND3	protein_coding	-9.482230403	1.097E-203
MT-ND5	protein_coding	-8.742280986	2.5622E-196
MT-TT	Mt_tRNA	-6.488225382	2.57429E-85
MT-TL1	Mt_tRNA	-6.129903761	9.8316E-66
MT-ND6	protein_coding	-4.077143136	5.9182E-57
MT-TV	Mt_tRNA	-4.41231946	2.17498E-32

DAVID pathway analysis of genes found in both Cufflinks/Cuffdiff and EdgeR analyses to be dowregulated in pooled KC corneas highlighted several pathways (Table 4-15). The main pathways/structures/functional groups reported computationally by DAVID, to which the encoded proteins belong are ribonucleoproteins, protein biosynthesis, translational elongation, mitochondrial structure/function, the respiratory chain, oxidative phosphorylation, apoptosis, membrane-bounded vesicle trafficking, protein ubiquitination and the proteasome, regulation of ligase activity, RNA splicing and the spliceosome, protein folding, GTPase activity and response to oxidative stress.

The genes that were taken further for pathway analysis were those which appeared to be significantly differentially expressed in both DE analyses. The global DE analysis between healthy and disease corneas revealed the enrichment of several pathways (Table 4-13); among these pathways the glycosylation, glycoprotein and membrane components were the most enriched, followed mainly by proteins related to the extracellular domain, the intrinsic and integral to plasma membrane, the neurological system process and most particularly sensory perception, cell adhesion and signalling, translational elongation, cytosolic ribosomes, protein synthesis, ion transport, neurotransmitter receptor activity, cilium and many more pathways. These pathways have shown significant enrichment in the differentially expressed genes between pooled KC and pooled WT corneas, suggesting that these pathways may be the downstream secondary changes in KC corneas compared to WT and the end result of pathways to be altered in KC pathogenesis.

In addition, the subset of genes that were found to be totally absent in pooled KC corneas compared to WT, and vice versa, were checked for enriched pathways (Table 4-16 and 4-17). Surprisingly, the genes that were shown to be totally absent in WT corneas compared to KC belong to the olfactory pathway which was enriched in KC corneas compared to other pathways, along with the G-protein coupled receptor, sensory transduction, disulfide bond, cell membrane and glycoprotein functional groups. The olfactory signalling genes have previously been reported to be expressed in the eye by Pronin et al., who suggested that they may play a role in sensing chemicals in the ocular environment (Pronin *et al.*, 2014). However, the reason behind the upregulation or the activation of the genes involved in ocular signalling ought to be further studied in details in KC corneas.

As for the genes that are absent in KC corneas compared to WT, these clustered in only two significant pathways, the MHC class I protein complex and the antigen processing and presentation grouping. These pathways need to be closely looked at and further investigated.

4.3.5 Differential expression profile for each cornea and its WES

Table 4-24 shows 13 genes shown to harbour high CADD Phred score variants in the WES data generated from blood DNA of 6 KC patients who donated their KC corneas. Among these genes, 5 have no known function: 2 long intergenic non-coding RNA (*LINC00910 and LINC00955*), 2 pseudogenes (*ANKRD36BP2* and *GXYLT1P3*) and 1 antisense gene (*MCM3AP-AS1*). The functions and pathways which belong to the rest of the genes are mainly: Calcium ion binding (*HRNR* and *FAT4*), protein and DNA binding (*MUC2, ZNF738, HNRNPL* and *FAT4*), extracellular matrix structural constituent (*MUC6*), cell-cell adhesion pathway (*FAT4*), mRNA splicing process (*HNRNPL*), Ubiquitin-protein tranferase activity (*ZNF738*) and oxidoreductase activity (*FDFT1*).

Along with harbouring potential pathogenic variants, these genes were picked according to the criteria of being significantly differentially expressed in these particular individual KC corneas. Therefore, since these variants appear equally in WES data from all 6 individuals who donated KC corneas, they might be or artefacts and data noise or variants in genes that present some susceptibility to KC development. Thus, further validation of the reported variants is crucial in order to make the differentiation between the two hypotheses.

4.3.6 DE of KC genes in the literature and comparison with data obtained in this project

So far, KC transcriptomics studies were mainly limited to microarrays (Bochert *et al.*, 2002, Bochert *et al.*, 2003, Ha *et al.*, 2003, Nielsen *et al.*, 2003, Mootha *et al.*, 2009, Sutton *et al.*, 2010, Mace *et al.*, 2011, Ghosh *et al.*, 2013, Lackner *et al.*, 2014). These studies have highlighted a number of molecules that have been reported by several groups to be significantly differentially expressed in KC corneal epithelium.

Lysyl Oxidase (LOX) mRNA has been reported by Nielsen et al. to be downregulated in KC corneal epithelium (Nielsen *et al.*, 2003). Indeed, a correlation has been found between the downregulation of LOX mRNA levels (Nielsen *et al.*, 2003, Shetty *et al.*, 2015c) and the downregulation of LOX enzymes and activity in KC corneal tissues (Dudakova *et al.*, 2012, Dudakova *et al.*, 2015a, Shetty *et al.*, 2015b). By immunohistochemistry Dudakova et al. found that the LOX enzyme distribution in KC corneas is decreased about 63% and LOX activity is reduced by 38% in KC corneas compared to control tissue (Dudakova *et al.*, 2012, Dudakova and Jirsova, 2013). In addition to LOX, Dudakova et al. investigated the presence and activity of lysyl oxidase-like enzymes in KC corneas (LOXL1-4) by IHC and found LOXL2, 3 and 4 enzymes are present at decreased levels in KC corneas compared to normal corneas (Dudakova *et al.*, 2015b).

Tissue inhibitor of metalloproteinase 1 and 3 (TIMP1 and TIMP3) mRNA levels were shown to be downregulated in cultured keratoconic corneal stromal fibroblasts (Lee *et al.*, 2009) and the alterations in expression of these two transcripts have been shown to have an effect on keratocyte apoptosis in Keratoconus (Matthews *et al.*, 2007). However, TIMP3 mRNA has been shown to be upregulated in KC corneal epithelium (Nielsen *et al.*, 2003) but Kenny et al. also reported a TIMP1 mRNA downregulation in KC corneas (Kenney *et al.*, 2005).

The Wnt/beta-catenin signalling pathway has been shown to be involved in KC pathogenesis by Katoh et al. who found the angiopoietin isoform (*ANGPTL7*) mRNA to be upregulated in KC corneas (Katoh and Katoh, 2006). In addition to *ANGPTL7*, the mRNA for another protein belonging to the Wnt signalling pathway, Secreted Frizzled-related protein 1 (*SFRP1*), has been shown to be significantly upregulated in KC corneal epitheliums compared to controls (Sutton *et al.*, 2010). This correlated with an upregulation of SFRP1 protein levels in keratoconic tear samples (You *et al.*, 2013a) and SFRP1 expression in KC corneas detected by immunostaining (Iqbal *et al.*, 2013, You *et al.*, 2013b). Furthermore, Iqbal et al. established a link between the expression of SFRP1 and microtubule-associated protein light chain 3 (MAPLC3), and showed correlation of expressivity between the two molecules in KC epithelium (Iqbal *et al.*, 2013).

In addition, inflammatory molecules and abnormal levels of inflammatory enzymes have been shown to be present in subjects with KC (Lema and Duran, 2005, Lema *et al.*, 2009, Lema *et al.*, 2010, Balasubramanian *et al.*, 2012, Balasubramanian *et al.*, 2013, Sorkhabi *et al.*, 2015, Wisse *et al.*, 2015) and several research groups have suggested that KC may be an inflammatory disorder (Galvis *et al.*, 2015a, Galvis *et al.*, 2015b, McMonnies, 2015, Sorkhabi *et al.*, 2015, Wisse *et al.*, 2015).

Tears of KC patients showed increased protein levels for Interleukin 4, 5, 6 and 8 (IL-4, -5, -6 and -8), Tumor necrosis factor alpha and beta (TNF-α and -β) and matrix metalloproteinases 1, 3, 7, 9 and 13 (MMP-1, -3, -7, -9 and -13) (Lema and Duran, 2005, Lema *et al.*, 2009, Lema *et al.*, 2010, Balasubramanian *et al.*, 2012, Balasubramanian *et al.*, 2013). Indeed, Shetty et al. demonstrated increased MMP9 transcript levels in KC corneal epithelium (Shetty *et al.*, 2015b). Membrane-Type-1 Matrix Metalloproteinase (MT1-MMP also known as MMP14) and MMP1 proteins have been shown to be present at elevated levels in KC corneas compared to healthy corneas (Collier *et al.*, 2000). Along with MMP1, Seppälä et al. showed elevated protein levels of extracellular matrix metalloproteinase inducer (EMMPRIN-also known as BASIGIN or BSG) in KC corneas compared to controls (Seppala *et al.*, 2006). In a recent study, Sorkhabi et al. by applying ELISA (enzyme-linked immune-adsorbent assay) reported the elevation levels of IL6, IL-1b and interferon gamma (IFN-γ) in the tear film of patients with KC compared to controls and reduced levels of the anti-inflammatory molecule IL-10 (Sorkhabi *et al.*, 2015).

Table 4-26: **The expression status in the transcriptomic study of this project of molecules reported to be up- or down-regulated in KC in the literature, either as transcripts or proteins.** The negative fold change (logFC) values in the edgeR results represent the genes downregulated in KC compared to WT and the positive logFC values represent the upregulated genes. The opposite is displayed by Cufflinks/Cuffdiff results, where negative values indicate upregulated genes and positive values indicate downregulated genes in KC compared to WT. The EdgeR false discovery rate (FDR) and Cufflinks q-value are measures of significance for these findings.

Molecule	In the literature	Location	In this project's data			
			EdgeR		Cufflinks/Cuffdiff	
			logFC	FDR	Log2FC	q-value
LOX	Downregulated	Epithelium	-2.078918	2.162E-18	1.34894	0.000088
LOXL2	Downregulated	Epithelium, keratocytes and ECM	1.599106	2.930E-09	-2.5646	0.000088
LOXL3	Downregulated	Corneal Stromal ECM, keratocytes	4.211806	1.793E-40	-3.55227	0.000175
LOXL4	Downregulated	Cornea	5.079647	9.267E-48	-5.90001	0.000088
TIMP1	Downregulated	Cultured stromal fibroblasts	-3.678936	1.295E-51	2.91972	0.000088
TIMP3	Downregulated	Cultured stromal fibroblasts	-1.307778	3.673E-06	4.60986	0.000088
	Upregulated	Epithelium				
ANGPTL7	Upregulated	Cornea	-1.438655	4.138E-09	0.572331	Not significant
SFRP1	Upregulated	Epithelium	0.725255	0.010711	-1.78529	0.000088
LC3 (MAP1LC3A)	Upregulated	Epithelium	-0.864908	0.001321	2.33208	0.000088
(MAP1LC3B)			-4.249841	4.152E-66	2.94531	0.000088
IL-1b	Upregulated	Tear film	3.430826	1.060E-25	-4.88189	0.000088
IL-4	Upregulated/Downregulated	Tear	10.38126	5.261E-52	Absent in WT	0.000088

IL-8	Upregulated	Tear	-	-	-	-
IL-10	Downregulated	Tear film	9.507738	4.852E-39	Absent in WT	0.000088
IL-12	Downregulated	Tear	-	-	-	-
IL-13	Downregulated	Tear	3.937115	1.307E-27	-5.63766	0.00026
IL-17	Downregulated	Tear	-	-	-	-
TNF-α (TNF)	Upregulated/Downregulated	Tear	0	Not significant	0	Not significant
TNF-β (LTA)	Upregulated	Tear	0	Not significant	0	Not significant
MMP-1	Upregulated	Tear and epithelium	-3.751623	1.072E-53	4.38037	0.000088
MMP-3	Upregulated	Tear	-4.151740	9.757E-64	3.57976	0.000088
MMP-7	Upregulated	Tear	3.151356	2.723E-25	-4.99721	0.000088
MMP-9	Upregulated	Tear and epithelium	-0.630913	0.014244	0.989254	0.018816
MMP-13	Upregulated	Tear	8.780041	1.938E-65	-8.19363	Not significant
MT1-MMP (MMP14)	Upregulated	Epithelium, stroma	-3.662756	1.071E-51	2.91814	0.000088
IFNG	Upregulated/Downregulated	Tear film	9.831951	1.001E-43	Absent in WT	0.000088
EMMPRIN (BSG)	Upregulated	Epithelium	-4.486983	1.186E-72	3.86172	0.000088
CCL5	Downregulated	Tear	0	Not significant	0	Not significant

Indeed, Jun et al. reported an increase in IL6 in the tear samples of KC patients (Jun *et al.*, 2011). However, contrary to what was reported by other studies (Lema and Duran, 2005, Lema *et al.*, 2009, Balasubramanian *et al.*, 2012, Sorkhabi *et al.*, 2015), Jun et al. reported a decrease in TNF-α, IFN-γ and IL4 along with a decrease in IL12, IL13, IL17 and chemokine C-C motif ligand 5 (CCL5) protein levels (Jun *et al.*, 2011).

As displayed in Table 4-26, the concordance between the published transcripts and proteins expression results in KC corneal tissues and tear samples, and the transcriptomic results of this project was not met. What is worth mentioning however is the concordance between the edgeR and the Cufflinks/Cuffdiff results (Table 4-26). All the molecules that were shown to be upregulated by edgeR (positive values) were shown similarly to be upregulated by Cufflinks/Cuffdiff (negative values), and likewise for downregulated genes. The consistency between both software packages is reassuring and suggests reproducibility.

The upregulation of SFRP1, INFG, IL1B, IL4, IL5 and IL6 proteins in the tears of KC patients, as reported the literature (Lema and Duran, 2005, Lema *et al.*, 2009, Lema *et al.*, 2010, Balasubramanian *et al.*, 2012, Balasubramanian *et al.*, 2013, Sorkhabi *et al.*, 2015, Wisse *et al.*, 2015), was reproduced at the mRNA level in this study in the anterior cornea (Epithelium and stroma). However, in contrast to what was previously reported in the literature, Jun et al. reported the downregulation of INFG and IL4 by tears of KC patients (Jun *et al.*, 2011). This is refuted in the anterior corneal tissues of KC patients belonging to this study, data from which confirm the prevalent view that these molecules are upregulated in KC.

IL10 and IL13 showed upregulation in this study, contrary to what was reported in tear film by Sorkhabi et al. and Jun et al. (Jun *et al.*, 2011, Sorkhabi *et al.*, 2015). IL8 mRNA, also reported to be upregulated, was entirely absent in this transcriptomics data. As for IL12 and IL17, Table 4-3

shows all the different levels of expression of the Interleukins belonging to the family of both molecules. The data shown in Table 4-27, in contrast to that reported in the literature, shows an overall overexpression of IL12 and IL17 molecules in KC corneas compared to WT ones, with IL17A, IL17B, IL17C, IL17F and IL17REL totally absent in WT corneas compared to KC corneas.

Table 4-27: IL12 and IL17 mRNA expression in KC vs WT corneas. This table presents the relative mRNA expression levels in KC vs WT corneas of the different molecules of the IL12 and IL17 gene families in this study data, generated by edgeR and Cufflinks/Cuffdiff software. The negative values in the edgeR results represent the downregulated genes in KC compared to WT and the positive values represent the upregulated genes. The opposite is displayed by Cufflinks/Cuffdiff results, where negative values indicate upregulated genes in KC compared to WT and positive values indicate downregulated genes in KC compared to WT. The grey highlighted rows indicate the non-significant expression at least in one software out of two of the respective genes under the gene name.

Gene name	**EdgeR**		**Cufflinks/Cuffdiff**	
	LogFC	FDR	Log2(FC)	q-value
IL12A	7.337325317	2.22472E-61	-8.28369	0.00017521
IL12A-AS1	7.364415973	4.37403E-15	Absent in WT	1
IL12B	8.167424774	1.11908E-44	-9.11893	0.00741423
IL12RB1	6.995871078	3.13826E-63	-6.41503	8.89E-05
IL12RB2	3.622758594	3.86764E-25	-4.74611	8.89E-05
IL17A	10.02478206	1.39228E-46	Absent in WT	8.89E-05
IL17B	10.07562421	2.44396E-47	Absent in WT	8.89E-05
IL17C	9.13894429	4.38385E-34	Absent in WT	8.89E-05
IL17D	2.789039895	1.56606E-15	4.11254	0.00949148
IL17F	8.973326452	6.98790E-32	Absent in WT	8.89E-05
IL17RA	-0.765742346	0.002171404	-0.447973	0.165678
IL17RB	5.522428596	6.43264E-54	-4.23767	0.0052161
IL17RC	-1.168080397	1.32906E-06	-0.214326	0.324508
IL17RD	3.053661123	9.33650E-26	-4.44741	8.89E-05
IL17RE	0.673512442	0.013778942	-1.71264	8.89E-05
IL17REL	10.21871012	1.64510E-49	Absent in WT	8.89E-05

The notion of concordance in expressivity between SFRP1 and LC3 proteins (represented here by MAP1LC3A and MAP1LC3B) that was proposed by Iqbal et al. was not supported by this study at the mRNA level. SFRP1 transcript was shown to have low expression in KC corneas compared to WT. However, although the LC3 proteins, MAP1LC3A and MAP1LC3B, were not concordant with this expression pattern, they both showed high expression in KC corneas compared to WT, with MAP1LC3B showing higher expression then MAP1LC3A (Iqbal *et al.*, 2013).

The downregulation of LOX in the epithelium and TIMP1, TIMP3 transcripts in the cultured stromal fibroblasts reported in the literature by Lee et al. (Kenney *et al.*, 2005, Lee *et al.*, 2009) was replicated in this study; however the reported upregulation of TIMP3 in the epithelium (Nielsen et al., 2003) and the downregulation of the LOX-like molecules (LOXL2, LOXL3 and LOXL4) (Dudakova *et al.*, 2015b) is not supported by the data presented here; all the LOXL mRNA levels have been shown to be upregulated by both analysis pipelines in this study.

The upregulation of matrix metalloproteinases -1, -3, -9 and -14 (MMP1, MMP3, MMP9 and MT1-MMP or MMP14) proteins in the literature (Collier *et al.*, 2000, Mackiewicz *et al.*, 2006, Rohini *et al.*, 2007, Predovic *et al.*, 2008, Honda *et al.*, 2010, Balasubramanian *et al.*, 2012, Ortak *et al.*, 2012) was not replicated by this study at the mRNA level. *MMP1*, *MMP3* and *MMP14* mRNAs showed an obvious downregulation (logFC(MMP1)=-3.75 by edgeR and 4.38 by Cufflinks/Cuffdiff; logFC(MMP3)=-4.15 by edgeR and 3.57 by Cufflinks/Cuffdiff; logFC(MMP14)=-4.48 by edgeR and 3.86 by Cufflinks/Cuffdiff) and *MMP9* showed a slight downregulation with a logFC of -0.63 by edgeR and 0.9 by Cufflinks/Cuffdiff. On the other hand, the upregulation of proteins MMP7 and MMP13 reported in the literature (Lema and Duran, 2005, Lema *et al.*, 2009, Lema *et al.*, 2010, Balasubramanian *et al.*, 2012, Balasubramanian *et al.*, 2013) was replicated at the mRNA level in this study. The upregulation of *MMP7* mRNA was reproduced by both

software pipelines, but upregulation of *MMP13* mRNA was only significant as detected using the EdgeR software.

ANGPTL7 and *EMMPRIN (BSG)* mRNAs were shown to be downregulated in this study, contrary to what was reported in the literature for *ANGPTL7* mRNA and EMMPRIN protein in KC corneas (Katoh and Katoh, 2006, Seppala *et al.*, 2006). The *EMMPRIN (BSG)* downregulation was reproduced in both software pipelines, but *ANGPTL7* downregulation was not significant as detected by Cufflinks/Cuffdiff.

The tumor necrosis factor alpha and beta (*TNF* and *LTA*) and *CCL5* were not shown to be expressed at all in this study in KC or WT samples; however Table 4-28 displays the expression of all the molecules belonging to the TNF superfamily.

Table 4-28: Tumor Necrosis Superfamily (TNFSF) molecule relative expression in KC and WT corneas. This Table presents the relative mRNA expression levels of the different molecules of the TNF superfamily in the RNAseq data obtained in this study, generated by edgeR and Cufflinks/Cuffdiff software. The negative values in the edgeR results represent the downregulated genes in KC compared to WT and the positive values represent the upregulated genes. The opposite is displayed by Cufflinks/Cuffdiff results, where negative values indicate upregulated genes and positive values indicate downregulated genes in KC compared to WT. The grey highlighted rows indicate the non-significant expression of the respective genes under the gene name.

Gene name	EdgeR		Cufflinks/Cuffdiff	
	LogFC	FDR	Log2(FC)	q-value
TNFSF1 (LTA)	2.650751436	0.041229871	Absent in KC	1
TNFSF2 (TNF)	0	1	0	1
TNFSF3 (LTB)	0	1	0	1
TNFSF4	4.992927245	2.18387E-31	-7.53996	0.000756584
TNFSF5 (CD40LG)	8.887583083	7.95553E-56	-9.2253	0.257105
TNFSF6 (FASLG)	8.649235383	4.58784E-52	-10.8772	0.254251
TNFSF7 (CD70)	7.583993723	2.89235E-46	-6.56671	0.117423
TNFSF8	10.44992465	4.62908E-53	-10.2661	0.269479
TNFSF9	-0.609435043	0.023303107	0.142394	0.477428
TNFSF10	-3.894716008	5.09477E-56	2.55883	0.000088859

TNFSF11	11.23240834	1.2093E-65	Absent in WT	0.000088859
TNFSF12	-0.073982827	1	-5.94882	0.125906
TNFSF13	-3.105703745	7.14221E-32	-3.96242	0.0336818
TNFSF13B (TNFSF20)	3.185847597	2.37689E-24	-5.0593	0.000088859
TNFSF14	4.353447709	1.08589E-32	-7.72964	0.0504677
TNFSF15	1.133696812	6.80362E-05	-3.59291	0.000088859
TNFSF18	6.516246397	3.66526E-31	-8.46529	0.200658
EDA	0.643243885	0.022177547	-2.3049	0.00306173
TFAIP1	-3.463682655	3.03821E-46	0.997899	0.0122231
TNFAIP2	-3.595703891	6.12697E-50	2.69266	8.89E-05
TNFAIP3	-0.786289	0.001617244	-1.51856	8.89E-05
TNFAIP6	2.613333204	5.74200E-20	-3.10054	8.89E-05
TNFAIP8	-0.469464215	0.085289828	-1.4191	8.89E-05
TNFAIP8L1	0.617835524	0.057100961	-2.06762	0.0305103
TNFAIP8L2	9.606790967	1.78336E-40	-	-
TNFAIP8L3	-0.138344396	0.811626743	-1.75773	8.89E-05

The reason of the obvious discordance in the majority of cases between the literature and the data generated in this project might be largely due to the differences in tissues where these molecules were investigated. Most of the literature investigated the differential expression in tears or in the general cornea and not in anterior corneas. In addition, the differences in methods (microarray, RT-PCR, western blots, immunohistochemistry and RNA-seq) might present a major difference between studies especially if different protocols were adopted between studies.

In addition, the data presented in the literature and not replicated by other researchers is rather preliminary and requires more validation through replication. The data reported in this project were computationally replicated by two different software packages suggesting a certain level of reliability. However, reproducibility by RT-PCR and immunohistochemistry of the same results is crucial in order to validate these findings.

4.3.7 Limitations of this study

In addition to the major limitations of RNA-seq method (Discussed in details in Section 5.1.2.2), one of the limitations of this study lies in the experimental design at the level of pooling the samples in the Illumina sequencing machine, thus creating an undesirable batch effect (Auer and Doerge, 2010, Leek *et al.*, 2010). The batch effect trap in this study consisted of pooling the KC corneas together on one lane of the Illumina machine on one run, and then pooling WT corneas on a different run on a different day. This potentially created differences between conditions that are purely related to technical management rather than to biological variables. This specific technical behaviour risked the validity of the biological conclusions (Baggerly *et al.*, 2004, Akey *et al.*, 2007) from this study however the valid surrogate variable analysis (sva) statistical analysis (Leek, 2014) can help undo the batch effect. This limitation will be overcome in the future of this study by repeating in one run and at low coverage all the samples included in this study and eliminate all the variations that were due mainly to the batch effect. The effect of this issue on the results of this project is not yet known however only replication of the results reported is going to define how accurate the findings are.

A second limitation lies in sampling routes for KC corneas and WT corneas, as the donation process was different. In the case of WT corneas, after removal from a deceased donor (as soon as possible and no longer than 24 hours after death) the tissue has gone into a process of incubation at 34 degrees for an unknown time, which could potentially have been up to but not exceeded 28 days. In contrast the cornea from a KC patient has been obtained fresh from theatre, from a living donor and with no incubation or long preservation in anything but RNAlater. This accounts for the technical difference that might slightly affect the profiles of KC and WT corneas. However, this should not present a major difference since the WT corneas follow a protocol in order to be suitable for transplantation, thus they are biologically viable.

In addition to the limitations, 3 main points are worth bearing in mind while interpreting the results and discussing the analysis: first point is that the RNA extracted from the corneas originates from different cell types (epithelial, stromal keratocytes and other cells) which implies that the RNA extracted is not cell specific but rather tissue specific. This presents a limitation when it comes to KC since in one condition (KC per say) one type of cells (epithelial) will be overrepresented due to the apoptosis of the other type of cells (keratocytes) (Section 1.4.5), which will lead theoretically to an overrepresentation of markers that are cell specific instead of molecules that are disease specific. This only affects the results of the pooled KC corneas against WT, since it looks for secondary effects highlighting the pathways of the end effects. Looking at individual KC corneas' RNA-seq profiles combined with the WES data of the same individual corneas might highlight the primary cause and overcome this issue.

Second, KC is a progressive disease and the stage of getting the KC corneas from patients is a stage where the disease has reached a severe end-point that requires corneal transplantation, the RNA-seq data generated from these KC corneas are not displaying the changes that occur at the beginning of the KC disease process, they are rather secondary effects. This data is therefore more representative of the transcriptomic profile of the end point in the disease development. However, this point was overcome by looking at the profiles of individual corneas.

Third, the reduction of the number of replicates in the DE analysis from 6 KC corneas to 5 KC corneas and the WT from 5 WT corneas to 3 WT corneas because of the odd profiles that they have shown on cummeRbund (Figure 4-5 and 4-6). However, using 3 replicates per condition is still biologically valid. New corneas will be recruited to continue the work on this project, which will improve the quality of the results by enhancing reproducibility.

Since this is an ongoing project and the time to generate this set of data and present it was limited, the investigation was limited to one class of output files generated by Cufflinks software. In the future work of this project, all the other output files will be studied in order to pull out all the new isoforms specific to corneas and in particular KC corneas. Additionally, miRNAs and ncRNAs are to be studied in order to explore the role of these molecules in KC pathogenesis.

5 General Summary and Concluding Remarks

5.1 General Summary

KC presents with a variable prevalence across populations and ethnicities (Section 1.4.5) ranging from 0.3 to 3,333 per 100,000 (Gorskova and Sevost'ianov, 1998, Waked *et al.*, 2012). Indeed, KC is the most common corneal ectasia. It appears around puberty and progresses until it is stabilised by the fourth decade of life. KC is thought to be a multifactorial disease with both genetic and environmental contributions (Gordon-Shaag *et al.*, 2015), making the discovery of the mechanisms underlying the disease a challenging task. Despite efforts over the years to determine the aetiology of keratoconus, the underlying pathology remains poorly understood.

5.1.1 Candidate gene study replication

Candidate gene studies, linkage analysis and association studies have been the main approaches used to date in an attempt to identify genetic variants contributing to susceptibility to KC (Section 1.4.9). Findings from these studies have identified a number of genes (Appendix 11) potentially involved in KC pathogenesis, described in Section 1.4.9. To determine whether variants in these genes might be involved in causing KC in the cohort described in this study, coding variants and splice sites in these genes were identified and are listed below. Only changes that segregated in the sequenced patients of the families are reported in Table 5-1.

Table 5-1: Variants in known KC candidate genes found in the Leeds families. Genetic variants found in the KC patients deriving from families in the Leeds cohort, in genes previously reported to be involved with KC development in the literature.

Genes	Family	Amino Acid Change	snp138	CADD Phred
COL4A1 (13q34)	F11	exon45:c.G3997A:p.D1333N	rs141395813	28.4
	F8	exon32:c.C2624G:p.P875R	rs201964644	15.08
COL4A3 (2q36.3)	SC2	exon49:c.A4484G:p.Q1495R	rs77964815	13.09
COL8A2 (1p34.3)	F11	exon4:c.G1504A:p.G502S	rs142307403	18.89
FLG (1q21.3)	F11	exon3:c.C5671T:p.R1891W	rs36006086	24.2
	F10	exon3:c.G3095C:p.R1032T	.	22.6

	F11	exon3:c.C4309T:p.R1437C	rs12750571	22.3
	F11, SC1	exon3:c.C9808T:p.R3270C	rs35621145	17.89
	SC2	exon3:c.G5332T:p.A1778S	.	15.1
	F11	exon3:c.C4568T:p.T1523I	rs12750081	12.76
	F11, SC1	exon3:c.C6045A:p.D2015E	rs71626704	12.11
	SC2	exon3:c.G3896A:p.R1299Q	rs150251062	12.02
	KF	exon3:c.G5212A:p.G1738R	rs376237195	9.808
	SC1	exon3:c.C7618T:p.R2540W	.	8.219
FOXO1 (13q14.11)	F10	exon1:c.C251T:p.P84L	.	7.596
HGF (7q21.11)	KF	exon16:c.G1765A:p.V589I	.	8.201
MPDZ (9p23)	F8	exon22:c.C3203A:p.A1068D	.	32
	F9	exon14:c.C1739G:p.S580C	rs149265684	28.7
NEIL1 (15q24.2)	SC1	exon2:c.C404G:p.A135G	rs184655220	23
PARP1 (1q42.12)	F12	exon6:c.G727A:p.D243N	rs200470832	23
	SC3	exon7:c.G943A:p.V315I	.	11.61
RAB3GAP1 (2q21.3)	SC1	exon22:c.G2540A:p.R847Q	.	23
RXRA (9q34.2)	F10	exon2:c.G10A:p.V4I	rs137938878	17.48
XRCC1 (19q13.31)	F9	exon16:c.A1727C:p.N576T	rs2307177	24.1
	SC2	exon14:c.A1526C:p.E509A	.	23.1
ZNF469 (16q24.2)	F11	exon2:c.G10277A:p.R3426Q	rs75288466	23.7
	SC3	exon2:c.C8594T:p.P2865L	.	20.9
	F12, F9	exon2:c.9010_9024del:p.3004_3008del	.	12.58
	F9	exon1:c.A3062C:p.E1021A	.	12.24
	F12, F9	exon2:c.C8342T:p.P2781L	.	9.066

The changes presented in Table 5-1 are exonic/splice site non-synonymous variants with CADD Phred scores > 7, a MAF of < 2% in the 1000 genome database, the exome server database (esp6500si) and Exac (Exac03) and were excluded from controls of the same ethnicities They have all been shown to segregate in the affected patients of the corresponding family. The full list of all the variants (including variants having lower CADD Phred scores) in the genes associated with KC in the literature in the Leeds cohort KC families is displayed in Appendix 11. Of the 43 genes reported in the literature to be potentially involved in KC pathogenesis (Genes listed in Appendix 11 and described in details in Section 1.4.9) (Wilson *et al.*, 1996a, Heon *et al.*, 2002a, Aldave, 2005, Aldave *et al.*, 2006, Udar *et al.*, 2006, Aldave *et al.*, 2007, Kim *et al.*, 2008, Stabuc-Silih *et al.*, 2009, Liskova *et al.*, 2010a, Romero-Jimenez *et al.*, 2010, Burdon *et al.*, 2011, De Bonis *et al.*, 2011, Karolak *et al.*, 2011, Muszynska *et al.*, 2011, Bykhovskaya *et al.*,

2012, Czugala *et al.*, 2012b, Guan *et al.*, 2012, Kok *et al.*, 2012, Li *et al.*, 2012a, Lu *et al.*, 2013, Mikami *et al.*, 2013, Wang *et al.*, 2013a, Kokolakis *et al.*, 2014, Lechner *et al.*, 2014, Sahebjada *et al.*, 2014, Wojcik *et al.*, 2014a, Hao *et al.*, 2015a, Karolak *et al.*, 2015a, Saravani *et al.*, 2015, Synowiec *et al.*, 2015, Wang *et al.*, 2015a), 13 have single base pair (bp) substitutions in the coding region leading to missense variants (Table 5-1) and one of these also has a nonsense variant, a nonframeshift 15bp deletion in *ZNF469* (Table 5-1).

The variants in the genes *COL4A1*, *COL8A2*, *FLG*, *MPDZ*, *NEIL1*, *PARP1*, *RAB3GAP1*, *RXRA*, *XRCC1*, *ZNF469* are more highly ranked by CADD Phred scores, having scores of > 15. These relatively high pathogenicity scores in rare variants segregating with KC in families, together with the fact that these genes have already been implicated in KC pathogenesis in previous studies, suggest that these variants have the potential to contribute to KC pathogenesis in the corresponding families. However, a number of further tests need to be carried out to determine whether this is indeed the case.

Firstly, these variants have not yet been checked for segregation in all of the unaffected members of the families. Testing the segregation of these variants in the complete families may help to confirm or refute their contribution to KC development. Secondly, while the changes appearing in Table 5-1 and Appendix 11 were all excluded from ethnically matched controls (BiB database, Lebanese Druze and Lebanese Maronites), the numbers for Lebanese Druze and Maronites were low and more controls need to be tested. Thirdly, more common variants, both functional and synonymous or non-coding, detected in these genes, should ideally be further investigated in KC patients in the future work of this project. This is because the involvement of these genes in KC might be the result of variants laying in important functional elements belonging to the non-coding regions of the genes (Lomelin *et al.*, 2010). Such a detailed examination of all changes will not be possible for whole exomes or genomes as the

number of variants would be too great to analyse, but specific genes implicated by multiple lines of evidence could be examined in this way.

Indeed, looking only at the frequency of specific variants in KC cases and controls may lead to bias because by definition rare variants have been selected and these are less likely to be found in controls than common variants. Ideally the entire genes should be screened for all variants in well phenotyped “KC-free” individuals, in order to determine the range of variants found, or “mutational load” in these genes in healthy individuals. Therefore, looking at the mutational load of the 13 genes reported in Table 5-1 in the KC cases and ethnically matched controls is a crucial analysis to follow up on this study.

Furthermore, specific genes and variants for which there is already strong evidence of involvement in KC from multiple different lines of evidence should also be looked at in sporadic KC cohorts, in order to establish a clear association with KC in the general population and not only in families. To achieve this on a larger scale, a multiplex assay which involves screening all the coding and non-coding regions of the corresponding genes as well as the promoter, upstream and downstream regions should be considered in order to make sure that all these parts of the genes are covered in the screened patients.

Interestingly, several families segregated variants in more than one of the genes listed in Table 5-1. For example, members of family F11 carried variants in four different candidate genes (*COL4A1*, *COLA8A2*, *FLG*, *ZNF469*). It is possible that all of these variants contribute to KC pathogenesis, which, if proven, would confirm the high level of genetic complexity for KC that has been postulated by many groups working on the condition.

In spite of the weaknesses described above and the incomplete nature of this study, the fact that 10 genes (*COL4A1*, *COL8A2*, *FLG*, *MPDZ*, *NEIL1*, *PARP1*, *RAB3GAP1*, *RXRA*, *XRCC1*, *ZNF469*) previously reported to carry variants contributing to KC also have potentially pathogenic changes in this KC cohort suggests that these variants are strong candidates for involvement in KC pathogenesis in the families investigated by the author. The remaining 30 genes were not replicated in this study. However this conclusion is based on the assumption that rare alleles of major effect are the main contributors to KC pathogenesis. Common variants with small effect might also contribute to the disease, and genes harbouring such changes will have been missed.

5.1.2 WES and RNA-seq combined analysis in this study

5.1.2.1 WES in families

Linkage analysis reported by researchers studying large KC families identified potential candidate genes for KC (Fullerton *et al.*, 2002, Brancati *et al.*, 2004, Hutchings *et al.*, 2005, Tang *et al.*, 2005b, Li *et al.*, 2006, Burdon *et al.*, 2008, Bisceglia *et al.*, 2009, Czugala *et al.*, 2012b). However, these studies only detected rare variants which were rarely reproduced by other researchers. Similarly, with genome-wide association analysis in large KC cohorts (Burdon *et al.*, 2011, Li *et al.*, 2012b, Lu *et al.*, 2013), common variants have been found to be associated, and some have been replicated in multiple studies, suggesting a potential involvement in KC pathogenesis (Section 1.4.9.3) (Bae *et al.*, 2013, Lechner *et al.*, 2014, Sahebjada *et al.*, 2014, Hao *et al.*, 2015a). However, missing heritability is still a major problem in KC. Thus in this study, several approaches have been followed in order to implicate key genes via multiple lines of evidence. Familial KC patients' coding regions have been sequenced by WES (Chapter 3), and in addition, the transcriptome of WT corneas and differentially expressed profiles of KC corneas have been characterised (Chapter 4).

5.1.2.1.1 Advantages of the approach used

GWAS addresses common variants of modest effect, testing the common disease common variant hypothesis. However, this current study tests the alternative hypothesis of common disease caused by rare variants, by looking for rare alleles of major effect causing a significant portion of KC in patients. Thus, the family based WES, where at least two KC patients from each family were subject to WES, has been a method of choice in this study to determine the genetic basis of KC for several reasons. First, it is a global approach which allows the screening of coding regions of all genes in KC patients (and controls/unaffected familial members). It is of course also possible that non-coding variants lying outside the coding regions screened by WES might be contributing to the disease pathogenesis and several examples in the literature confirm this statement (Cui *et al.*, 2010, Herdewyn *et al.*, 2012, Roessler *et al.*, 2012, Cavalieri *et al.*, 2013, Makrythanasis and Antonarakis, 2013, van de Ven *et al.*, 2013, Liu *et al.*, 2015b). However, the exome represents an enriched subset of the genome where protein-altering variants are more likely to have functional consequences (MacArthur *et al.*, 2014).

Second, WES is a valid approach for novel gene discoveries, as distinct from the candidate gene approach which aims to discover novel variants and thus replicate previous studies rather than identifying novel KC implicated genes. Thus it is a hypothesis independent approach with the potential to identify new pathways and mechanisms in corneal development and maintenance. Similarly, it has an advantage over the GWAS studies since it looks for rare variants with major effect instead of common variants with minimal effects, meaning that it can be applied to a relatively small cohort of families with the possibility of success. In order to achieve significance in GWAS studies, a large number of cases (WES or SNP data) should be incorporated in the study design, making it very costly.

Third, WES is a tool to search for rare variants laying in the coding region. The idea of rare variants involved in causing common genetically complex

diseases is still somewhat controversial. However several examples in the literature support the occurrence of rare variants in complex traits and the contribution of these variants to complex disease aetiology (Schork *et al.*, 2009, Helbig *et al.*, 2013, Lettre, 2014, Moutsianas *et al.*, 2015). Indeed, Hirschsprung disease (Alves *et al.*, 2013) and age-related macular degeneration (AMD) (van de Ven *et al.*, 2013) represent model for the involvement of common and rare variants in disease pathogenesis.

The results of performing WES on KC patients in this study have proved inconclusive so far, but analysis is as yet incomplete. The starting hypothesis of this study was that one or a few recessive KC alleles would be enriched in endogamous consanguineous populations and would thus be easier to detect. 15 families belonging to such populations were therefore sampled. Families were tested by both WES and autozygosity mapping. However, unfortunately the 15 families incorporated in this study did not present any prominent shared autozygous regions to focus on, and no single gene proved to be commonly mutated in KC families. The results obtained so far therefore suggest that, if recessive alleles of large effect exist, they are rare and can occur in multiple different genes, there is no common recessive locus or loci shared between the families of this study, and the condition seems more genetically complex than anticipated.

5.1.2.1.2 Limitations of the approach used

Following the approach described in the previous paragraph was justifiable at the start of this project, since there was less evidence for genetic complexity in KC and it remained possible that relatively common alleles of major effect underlay a proportion of KC cases. Further evidence of the complexity of KC emerged during the course of this project, while data from families was being generated (reported by (Kriszt *et al.*, 2014) and reviewed by (Jeyabalan *et al.*, 2013, Abu-Amero *et al.*, 2014a, Davidson *et al.*, 2014, McMonnies, 2014)), which corroborated the main conclusion of family

studies, namely that no single recessive allele or mutations in a single gene accounted for a large number of cases/families.

In addition, some technical and analytical limitations to WES should be taken into consideration when interpreting the WES data. The technical limitations derive from the fact that some genes or part of genes, do not fall into the WES targeted region, or there is insufficient coverage of certain regions where possibly the causative gene/variants reside. Additionally, mismapped reads and/or alignment errors, may lead to false calling of a certain variant that could be mistaken for a real one. Indeed, in some occurrences, a variant could be truly covered but not accurately called. Moreover, genetic heterogeneity may create an analytical bias in case discontinuous filtering strategies are adopted (Wang *et al.*, 2013b, Tetreault *et al.*, 2015). Indeed, the use of stringent filtering strategies could lead to exclusion of variants important in the disease aetiology. These technical and analytical limitations can be overcome for false-positive variants, by the verification of supposedly causative variants by directed sequencing (i.e. Sanger sequencing), as well as checking in case-control cohorts. However, the false-negative variants are the most problematic, since they are usually missed from the final shortlist of variants.

The strategy followed in this study was to combine the data from WES with the differential expression data of KC corneas against WT corneas in order to shortlist the WES variants according to the expression study. A recent study focusing only on WES case studies in order to detect rare variants implicated in complex diseases used a similar approach (Wu *et al.*, 2015). The study design proposed by Wu et al. closely resembled the study design developed in this project, endorsing the work done in this project. It comprised calling variants from WES data of complex disease cases then filtering according to allele frequency first then according to functionality and disease phenotype spectrum, leading to a gene prioritization process. Following up comes the combination of the information from WES with information from other data (i.e. mRNA data, somatic DNA variations data,

epigenetics data, open source available databases) then confirming the variants and gene involvement through targeted deep sequencing.

5.1.2.2 RNA-seq in KC and WT corneas

The RNA-seq analyses presented in this study are preliminary results, with further analysis still ongoing. There were three aims behind conducting this analysis; firstly to characterise the normal anterior corneal RNA profile; secondly to characterise the transcriptome during the keratoconus disease process, and finally, by combining with WES data, to look for potential primary causative variants.

Variants that could constitute the primary cause of, or major contributors to the susceptibility for, KC were screened for by first extracting the RNA differential gene expression profiles of each individual KC cornea against the pooled WT RNA-seq data. The genes highlighted as having significantly differential gene expression profiles in RNA were then checked for deleterious variants in the WES data of the corresponding individual KC cornea. The rationale behind this approach was to look for variants that might be directly responsible for the changes in the RNA differential expression profiles, in genes expressed in the anterior cornea, at the level of each individual cornea. Knowing that the causative variants KC might vary between different individuals, this method aimed to identify potential KC associated variants unique to each one of the six individual KC corneas analysed. In addition it was hoped that combined analysis of the genes implicated from each cornea could potentially detection of common pathways involved in the development of KC.

The secondary consequences of KC were looked for by comparing the RNA-seq data from the pooled KC corneas against the RNA-seq data from the pooled WT corneas. This analysis aimed to highlight transcripts that are differentially expressed in all KC corneas and might therefore be components of pathways that are altered during the KC disease process.

Both approaches are equally important in the study of KC, since the aetiology of KC has remained a mystery until now, and much of the current understanding of the development of KC pathology is speculative and remains to be confirmed by replication (Section 1.4.7). Insights into the cause of KC at the level of each individual KC cornea, together with data on shared altered pathways in all the KC corneas, provide a useful tool to bridge the gap in understanding between causes and effects, and may help to unravel some of the complexities of KC at the level of defective pathways.

To our knowledge, this is the first study to establish an RNA-seq based gene profile for the normal anterior corneal tissue. It is also the first study to perform RNA-seq in the analysis of primary and secondary alterations in KC corneas, and to combine WES and RNA-seq analysis from the same KC patients to extract variants from the WES profile in genes that are significantly differentially expressed in KC corneas.

RNA-seq technology is now the best tool to study transcriptomics in healthy and diseased tissues. However as with any other technique it has a number of limitations at the technical and analytical levels (Wang *et al.*, 2009). The first limitation of all, resides in comparing fresh KC corneas, collected on the spot, to healthy corneas taken from cadavers up to 28 days after death (Sections 4.3.1.1 and 4.3.2), which will mean loss of many short half-life RNA species. However, this will always be an issue while working with human corneal tissues, since the procedure of enucleation is standard. The technical limitations start when the RNA is sheared for fragmentation purposes, which leads to depleted ends. The cDNA synthesis, amplification and sequencing steps can lead to non-uniform coverage of certain transcripts (i.e. PCR artefacts), introducing bias in the overrepresentation of some transcripts compared to others (Ozsolak and Milos, 2011, Yendrek *et al.*, 2012, Kratz and Carninci, 2014). In addition, during the library construction, the cDNA synthesis and RNA manipulations potentially introduce further errors and artefacts, including the possibility of dissociation of the nascent cDNA from the template and annealing to a similar template;

self-priming arising from RNA secondary structure; and the low fidelity and lack of proofreading of the RNA transcriptase when compared to other polymerases.

As for the bioinformatic analytical limitations, they reside mainly in the alignment of small reads to the whole genome, since these reads can often be non-uniquely aligned to several similar locations in the genome (Ozsolak and Milos, 2011). RNA-seq technologies that lead to the generation of longer read fragments, such as the PacBio platform, are a way to overcome this limitation. Mapping reads that span splice junctions also present a challenging task since for complex transcriptomes alternative splicing events are prominent (Kim *et al.*, 2013). Length bias must usually be corrected for in RNA-seq analysis, since a bias is introduced during the fragmentation and RNA size selection steps, which might lead to an overrepresentation of longer transcripts (Trapnell *et al.*, 2012).

5.1.2.3 Combining results of RNA-seq and WES in families

The power of this research resides in combining all the approaches followed so far in the course of this study. The combined analysis focused on genes that are expressed in corneas, significantly differentially expressed in individual KC corneas compared to pooled WT corneas, and that have potentially pathogenic variants in the corresponding WES data. The resulting list consists of WES variants from all the individual KC corneas that appear to be functional (missense, splicing and Indels) and have MAF < 2%. The list of WES identified variants in differentially expressed genes from every KC cornea was then compared to the PLINK/SEQ lists of genes enriched for variants in KC cases in Chapter 3, in order to look for genes in common. Subsequently, all of the genes highlighted by both the PLINK/SEQ and differentially expressed genes with “functional” WES variants analyses, were checked for in the families described in Chapter 3, to look for variants that segregate in the families, in genes that are common to all lists. The result is a list of genes harbouring functional variants that segregate in families, appear enriched in variants on PLINK/SEQ analysis, are significantly

differentially expressed in individual KC corneas compared to pooled WT, and that harbour changes in the WES data of the corresponding individual KC corneas. The corresponding list is called “combined list”.

To further filter and prioritize genes that might be directly related to the primary defect in KC, the previously generated list of genes or “combined list”, was subject to two further filtering steps. Only the genes that appeared in at least two families were included, and only the genes that appeared to be down-regulated in all the individual KC corneas were kept in the final list. This last step risks excluding mutations that cause KC through up-regulation of transcription. However, given that nonsense mutations are likely to be subject to nonsense-mediated decay, and that mutations that reduce or abolish function are more common as a cause of disease than increased function mutations, it seems reasonable to suppose that most mutations affecting transcription will manifest as downregulation. This strategy led to a narrowing down of the list of genes highlighted to 5 genes: *FLNB* (Filamin B, beta), *ITGB4* (Integrin beta 4), *KIAA0100, LAMA5* (Laminin subunit alpha 5) and *PCDH1* (Protocadherin 1) (Table 5-2). The differential expression levels of Table 5-2 genes in the individual KC corneas compared to pooled WT corneas, are displayed in Table 5-3.

The strategy followed in this section served to prioritize the list of genes in the WES families data, according to their occurrence in PLINK/SEQ as well as to their downregulation in KC corneas compared to WT. This strategy does not assume that the missense variants of Table 5-2 are causal for the down-regulation of the genes presented in Table 5-3. Instead, it serves to highlight the overlap between the WES families’ data and the differential expression analysis of KC/WT corneas. However, in the future, searching for splice variants and mutations in promoter regions of the downregulated genes would be the orthodox way to check for the direct effect of these particular mutations on the expression of candidate genes.

Table 5-2: Genetic variants segregating in KC families, in genes shown to be downregulated in individual KC corneas. This table reports the different variants in *FLNB, ITGB4, KIAA0100, LAMA5* and *PCDH1* in different KC families' WES data.

Gene	Family	CADD Phred	Amino Acid Change	snp138	Esp6500si	1000g	Exac03	Status
FLNB (3p14.3)	F13	35	exon14:c.C2095T:p.R699W	rs200554477	0.000077	0.000399	0.000173	het
	F7	25	exon22:c.G3785C:p.G1262A	rs111330368	0.004229	0.002396	0.005461	het
	F13	23.8	exon26:c.T4447C:p.F1483L	rs142631042	.	0.000998	0.0008167	het
	F13, SC1, SC2, SC3	20.1	exon41:c.A6899C:p.H2300P	rs202222289	.	.	0.003068	het
	F11	9.944	exon38:c.A6313G:p.M2105V	.	.	.	.	het
ITGB4 (17q25.1)	F11	34	exon20:c.G2477A:p.R826H	rs200694443	.	0.000199	0.0002389	het
	SC1	31	exon6:c.C670T:p.R224W	.	.	.	0.0000082	het
KIAA0100 (17q11.2)	F13, F9	25.3	exon35:c.C6037G:p.L2013V	rs201896448	0.000384	.	0.0003789	het
	SC3	22.3	exon16:c.C2474T:p.S825L	.	.	0.000199	0.000313	het
LAMA5 (20q13.33)	F9, F12, F13	28.9	exon32:c.G4037A:p.C1346Y	rs371144423	0.000077	.	0.0000663	het/hom/het
	F8	25.4	exon28:c.G3467A:p.R1156Q	rs143551466	0.000154	0.000399	0.00047	het
	SC2	25.1	exon29:c.G3587T:p.S1196I	.	.	.	.	het
	F8, KF	23.8	exon53:c.G7052A:p.R2351Q	rs147290767	0.002335	0.009584	0.008156	het
	F9, F12	13.41	exon80:c.G10995A:p.M3665I	.	.	.	.	het/hom
	KF	8.749	exon53:c.G7138A:p.G2380S	rs150998056	0.001161	0.006190	0.003745	het
PCDH1 (5q31.3)	F10	25.2	exon2:c.G829A:p.G277S	.	.	.	0.0000247	het
	KF	11.92	exon5:c.A3688T:p.T1230S	rs373225365	0.000077	0.002396	0.00155	het

Table 5-3: LogFC and FDR values of *FLNB, ITGB4, KIA0100, LAMA5* and *PCDH1* in all the six individual KC corneas. This table represents the result output of EdgeR, with logFC being the log of the fold change and FDR the corrected p-value.

KC cornea	MA18		MA26		MA46		MA47		MA52		MA58	
Gene	logFC	FDR value	logFC	FDR value	logFC	FDR value	logFC	FDR value	logFC	FDR value	logFC	FDR value
FLNB	-1.637	0.0044	-	-	-1.505	0.00921	-1.715	0.00284	-1.559	0.0065	-	-
ITGB4	-3.575	2.84E-09	-3.625	1.58E-09	-3.635	1.53E-09	-3.435	9.68E-09	-3.542	3.22E-09	-3.728	8.30E-10
KIAA0100	-1.327	0.02895	-1.332	0.02724	-1.617	0.006782	-1.319	0.03030	-1.412	0.01846	-	-
LAMA5	-2.703	1.10E-05	-2.678	1.19E-05	-3.312	1.52E-07	-2.753	7.99E-06	-2.765	6.80E-06	-2.391	8.73E-05
PCDH1	-4.184	1.99E-11	-4.100	3.26E-11	-4.409	2.95E-12	-4.028	8.26E-11	-4.235	1.15E-11	-3.828	4.59E-10

FLNB is an actin binding protein (actin binding protein 278), mRNA splice variants of which control the organization of cytoskeleton actin and specifically bind to integrin (beta) subunits (van der Flier *et al.*, 2002). FLNB has been reported to be involved in keratinocyte development, epithelial cell morphogenesis and skeletal muscle tissue development (http://www.ncbi.nlm.nih.gov/gene/2317). More precisely, FLNB has been reported in the literature to be involved in the cytoskeletal anchoring at plasma membrane (Xu *et al.*, 1998) and signal transduction (Leedman *et al.*, 1993). Moreover, FLNB is involved in the MAPK signalling pathway, suggesting a role in processing cellular environmental information, and in the cell community of the focal adhesion pathway (http://www.genome.jp/dbget-bin/www_bget?hsa:2317).

The *FLNB* locus has been shown to be associated in meta-analysis GWAS to Primary Open-Angle Glaucoma (POAG) (Springelkamp *et al.*, 2015). Association between POAG and KC has been established in the literature (Goel *et al.*, 2015). In the families WES data of this project, missense heterozygous mutations of *FLNB* have been found (Table 5-2) in six families (F7, F11, F13, SC1, SC2 and SC3). Four variants (c.C2095T:p.R699W; c.G3785C:p.G1262A; c.T4447C:p.F1483L and exon41:c.A6899C:p.H2300P) in *FLNB* have CADD Phred scores > 20, in exons 14, 22, 26 and 41 respectively.

ITGB4 is a none-covalently associated transmembrane glycoprotein receptor, which tends to bind to the alpha 6 integrin subunit. It is involved in the formation of stable cell adhesions and in G-protein coupled receptor binding (Ni *et al.*, 2014). ITGB4 is a receptor for the laminins and particularly for Laminin-5, an essential component of basement membranes. In keratinocytes the ITGB4 and laminin-5 play an important role in cell migration and cell-cell adhesion (Geuijen and Sonnenberg, 2002, Hintermann *et al.*, 2005). Involved in the integrin mediated signalling pathway, ITGB4 is equally involved in amelogenesis (Schumann *et al.*, 2013), autophagy (Li *et al.*, 2013a), cell adhesion (Suzuki and Naitoh, 1990),

cell motility (Hamill *et al.*, 2009), cell matrix adhesion (Schumann *et al.*, 2013), hemidesmosome assembly (Koster *et al.*, 2003, Schumann *et al.*, 2013), mesodermal cell differentiation (Brafman *et al.*, 2013), nail and skin development (Schumann *et al.*, 2013) and response to wound healing (Hamill *et al.*, 2009).

According to KEGG (Section 2.14), ITGB4 is involved in the PI3K-akt (phosphatidylinositol 3' kinase) cell signalling pathway and in the ECM-receptor pathway. Moreover, it plays a role in the regulation of the actin skeleton pathway in cell motility, as well as in the focal adhesion pathway (http://www.genome.jp/dbget-bin/www_bget?hsa:3691). Two heterozygous *ITGB4* variants (c.G2477A:p.R826H; c.C670T:p.R224W) were identified in two families in this project's cohort (F11, SC1), these having CADD Phred scores > 30.

KIAA0100 is a protein coding gene the function of which is still not well understood. By in-silico modelling, KIAA0100 has been proposed to be a component in the extracellular region (http://www.ncbi.nlm.nih.gov/gene/9703) involved in membrane trafficking (Gene Ontology). Nothing else is yet known about this gene or protein. In this project, two heterozygous variants (c.C2474T:p.S825L and c.C6037G:p.L2013V;) were identified in exons 16 and 35 respectively of *KIAA0100*, segregating in three different families belonging to the KC cohort of this study (F9, F13, SC3) and having CADD Phred scores > 20.

LAMA5 is a non-collageneous component of the basement membrane. It functions in integrin binding and is a structural molecule (Doi *et al.*, 2002). According to the literature, *LAMA5* is involved in cell differentiation, migration, proliferation and recognition (Doi *et al.*, 2002). Additionally, it is involved in angiogenesis, cytoskeleton organization, embryo development, endothelial cell differentiation, focal adhesion assembly and the integrin-mediated signalling pathway (Doi *et al.*, 2002). From the inferred electronic

annotation, *LAMA5* has been thought to play a role in cilium assembly, establishment of protein localization to plasma membrane and regulation of cell adhesion and cell migration (http://www.ncbi.nlm.nih.gov/gene/3911). According to KEGG, LAMA5 is involved in the PI3K-Akt cell signalling, ECM-receptor and focal adhesion pathways (http://www.genome.jp/dbget-bin/www_bget?hsa:3911). Moreover, according to WikiPathways and the REACTOME (Section 2.14), LAMA5 is involved in the inflammatory response pathway (http://www.wikipathways.org/index.php/Pathway:WP453) as well as in the extracellular matrix degradation pathway (http://www.reactome.org/content/detail/1474228). In this study, two homozygous variants (c.G4037A:p.C1346Y; c.G10995A:p.M3665I) were identified in *LAMA5* in one family (F12) in exons 32 and 80 respectively, with corresponding CADD Phred scores of 28.9 and 13.41. In addition, five heterozygous *LAMA5* variants (c.G4037A:p.C1346Y; c.G3467A:p.R1156Q; c.G3587T:p.S1196I; c.G7052A:p.R2351Q and c.G7138A:p.G2380S) were identified in the families WES component of this study in five families (F8, F9, F13, KF, SC2), with CADD scores of > 20.

PCDH1 is a calcium ion binding, protocadherin membrane protein found at cell-cell boundaries. It is involved in the process of cell-cell signalling and nervous system development (Sano *et al.*, 1993). According to KEGG, PCDH1 is a member of the beta integrin and cadherin protein families, and is involved in centrosome formation and ciliogenesis (http://www.genome.jp/dbget-bin/www_bget?hsa:56141). In this study, *PCDH1* mutations presented in two families (F10, KF), these being variants (c.G829A:p.G277S and c.A3688T:p.T1230S) in exons 2 and 5, having respective CADD Phred scores of 25.2 and 11.92.

The data presented highlight the possible involvement of these five genes, *FLNB, ITGB4, KIAA0100, LAMA5, PCDH1,* and corresponding variant proteins in KC pathogenesis. The fact that these genes appear to be highlighted in all three sets of analysis conducted in this study (WES data of KC families, PLINK/SEQ data and differential RNA-seq expression data),

make them strong candidate genes in which variants may contribute to the susceptibility to KC development. While it is known that some of the genes reported in this section and their corresponding proteins (ITGB4, LAMA5) are normally involved in keratinocyte formation and function (Geuijen and Sonnenberg, 2002, Hintermann *et al.*, 2005), the present study has highlighted that they may also be involved in KC pathogenesis.

The literature associated with three of the five genes suggests that defects in the focal adhesion pathway (FLNB, ITGB4, LAMA5), as well as the MAPK signalling, PI3-akt signalling and ECM-receptor pathways, may therefore be involved in KC pathogenesis. However, many of the keratoconus pathways implicated by other researchers do not overlap with the pathways that have been highlighted in this study.

The lack of overlap between the five most strongly supported candidate genes identified by this study and those reported literature could imply that the reports of pathways associated with KC in the literature, which are often speculative and based on un-replicated studies (Section 1.4.7), are largely incorrect. Alternatively, much of the data reported in this study might be artefactual. However, the combined approach conducted in this project, using results from different sets of data, attempts to minimise the amount of noise generated by the different analyses. Hence further analysis of the candidate genes reported in this project are warranted. Only confirmation of the results of this project by replicating the outcomes in larger datasets can confirm these findings and elucidate the potential role of sequence variation in these proteins in KC development.

The combination of the RNA-seq data and the WES data in this section focused mainly on the downregulated genes. However this data may highlight more genes in the upregulated category, and indeed other strong candidates excluded at various stages of filtering. These analyses were not

completed due to time restrictions. Another next objective for this project will therefore be to explore all the aspects of this data more rigorously.

5.2 Future directions and concluding remarks

5.2.1 Summary of key findings

Four years were spent in developing and carrying out the work of this project, during which families were collected, WES data and RNA-seq data were generated and analysis strategies were developed. The project started with the aim of searching for recessive alleles of large effect by autozygosity mapping in consanguineous, and to some extent endogamous, populations. However during years one and two it became evident that, if such alleles exist, they are relatively rare in this cohort at least, and no one gene is a common site for a significant proportion of them.

The family approach yielded long lists of potential pathogenic variants segregating in the families. Subsequently, the data generated in the family approach was checked for association in an underpowered association analysis which also generated long lists of genes apparently enriched in variants in KC cases from the families studied. However, no one gene was consistently affected in either analyses. These findings therefore do not support the initial hypothesis of this study that “one or a few recessive KC alleles will be enriched in endogamous consanguineous populations, but rather suggest that KC is indeed a genetically highly complex disorder and that a small family cohort is not sufficiently powerful to identify the few alleles of large effect, found in multiple different genes, that might exist (Chapter 3). It was therefore necessary to switch approaches half way through the study to a more complex association based approach, combined with transcriptome differential analyses, to identify candidate KC susceptibility genes. Although large amounts of data were generated, the time restriction led to the under-exploration of all aspects of the data.

The RNA-seq analysis first profiled the transcriptome of the normal anterior human cornea in addition to identifying a number of putative new cornea-specific transcripts. Moreover, the RNA-seq analysis generated long lists of differentially expressed genes both in single and pooled KC corneas against the pooled WT corneas. Differential expression in pooled KC vs WT corneas serves as a tool to overview the secondary consequences of KC, while examination of the differentially expressed gene lists in individual corneas, as well as detecting secondary changes, is more likely to highlight primary causative changes.

Furthermore, through the combination of WES and RNA-seq datasets, this study yielded a list of 5 genes implicated by both approaches, namely *FLNB, ITGB4, KIAA0100, LAMA5* and *PCDH1* (Section 5.1.2.3). The fact that these 5 genes encode structural proteins, thought to be involved in maintaining tissue structure and shape as well as cell-cell adhesion, is consistent with their involvement in KC pathogenesis and highlights them as strong KC candidates (Section 5.1.2.3). However, further validation of the highlighted variants is required. This can be obtained by Sanger sequencing in families to confirm and segregate variants. In addition, these genes should now be screened for these and other variants in a large cohort of sporadic KC cases, in order to obtain further proof of the involvement of these genes in KC in the general population.

Interestingly, this is the first study to highlight genes implicated in KC pathogenesis by combining WES and RNA-seq data. This data - if confirmed by further validation tests - will in effect have been replicated within this study since it is highlighted by multiple different analyses. However, replication in other data sets is required in order to corroborate a conclusion on the involvement of the corresponding genes in KC pathogenesis.

Moreover, since there is a fine line between considering a result as real or as an artefact due to the amount of noise generated by any applied method (especially bioinformatics), replication is essential in order to achieve robust

results. With this in mind, the 43 KC candidate genes described in the literature were checked for variants in this study and 13 of these were found to carry highlighted variants. Although this result is subject to confirmation by further analyses, the fact that 10 of the 13 genes (*COL4A1*, *COL8A2*, *FLG*, *MPDZ*, *NEIL1*, *PARP1*, *RAB3GAP1*, *RXRA*, *XRCC1*, *ZNF469*) contain potentially pathogenic variants with CADD scores of greater than 15 again suggests that these 10 are strong candidates for involvement in KC pathogenesis, warranting further investigation. However, the 30 candidate genes that did not contain segregating variants in this cohort are unlikely to have a role in KC development in these patients, though this does not exclude their involvement in other KC cases and families.

Nonetheless, this project is still an ongoing project where all the data generated during the course of four years, in addition to new data, will be more fully investigated for the purpose of finding new insights into the genetic basis of KC.

5.2.2 Retrospective insights

Notwithstanding the decades of study into the genetics of KC, its aetiology remains a challenge that researchers are still trying to unravel. This may in part be due to the diagnostic as well as genetic complexity of KC. In order for researchers to achieve a better understanding of KC, the diagnostic components of this disease must first be addressed by researchers. It is important to differentiate between secondary KC associated with other diseases and primary KC. This strategy has the potential to determine whether other diseases that are associated with secondary KC involve genes/proteins that contribute directly to genetic predisposition for KC, or whether these associations are caused by secondary consequences such as eye rubbing. Once this differentiation is established, the figure for prevalence of KC can be more accurately established. In addition, the pathways involved in diseases associated with secondary KC may give clues as to pathways involved in primary KC. It is also necessary to define the different

phenotypes involved in KC manifestations, and categorise the patients according to the different sub-types and the forme-fruste of the disease. Once that categorisation is clear, any genetic variants involved may be easier to detect in well-defined homogeneous cohorts of patients with each subtype. Accordingly, and looking retrospectively at this project, better categorising of KC phenotype during patients' recruitment might have increased the power to identify genetic variants contributing to susceptibility to the different KC types.

As for the family study, in order to address its lack of power in the search of rare alleles of major effect, a larger and more powerful cohort might potentially have increased the chances of positive findings. This large cohort would have been more powerful if all patients belong in origin to one ethnicity, where all the patients and the ethnically matched controls have undergone thorough phenotyping. Furthermore, given the weakness of the family study, the emphasis might have been more productively directed towards sampling a large cohort of KC cases rather than sampling families.

To strengthen the RNA-seq study, a larger number of KC corneas would have been at the core of the analysis. In addition, the technical batch effect generated by the design of the RNA-seq experiment would have been avoided by ensuring a better experimental design. Moreover, the WES analysis would have been done in parallel to the RNA-seq analysis, avoiding the time lag between producing the two sets of data.

5.2.3 Future directions

The next task in this project is to validate the results generated so far, in the families and in a larger cohort of familial and sporadic patients, by Sanger or multiplex sequencing. For these purposes a sample of 250 Leeds KC patients has already been obtained, and a further 2000 KC patient DNAs are available through collaboration with colleagues at Moorfields Eye Hospital.

In addition, only protein coding genes and coding variants were explored in WES data. However, microRNA genetic variations have been associated with KC (Lechner *et al.*, 2013b) as well as non-coding variants being associated with eye diseases (Braun *et al.*, 2013). Thus in the future, the search for KC associated variants will be inclusive of all the variants identified in the lists, applying less stringent filtering (including UTRs and synonymous variants). Similarly when combining the genomic WES data and the RNA-seq data in individual KC corneas, all the coding and non-coding variants in the WES data will be checked for a corresponding effect in the RNA-seq data, since functional elements reside in non-coding regions and intronic variations could result in pre-mRNA splicing (Baralle and Baralle, 2005, Lomelin *et al.*, 2010, Ward and Cooper, 2010). However, expanding the search to non-coding variants in genomic sequence will vastly increase the complexity of the analysis required and is therefore a huge challenge

In the near future, Whole Genome Sequencing (WGS) will be considered as an alternative replacement for WES technology, especially with the continuous improvement in software analysis to handle large data, reducing time consumption and cost. The switch to WGS in this project will allow not only the wider coverage of genomic sequence, increasing the possibility to detect KC causative variants, but it will also be a useful and global tool to explore copy number variations (CNVs) in case these also contribute to KC causation. CNVs have been studied in KC but only in mitochondria (Hao *et al.*, 2015b), thus this has yet to be investigated further.

Since the main purpose is to look for KC causative variants, the collection of more sporadic KC patients in addition to families belonging to endogamous and consanguineous ethnicities will be a continuous task for this project. In the process of collecting families, forme-fruste patients ought to be identified and looked at separately. Studying the forme-fruste KC patients will be key to finding the molecular pathogenesis and will help characterize the main genetic components that lead to the development of the disease. Obtaining

data from further families and cases and the better characterisation of forme-fruste KC in those families and cases will greatly increase the power of genetic studies.

In order to increase the power of the transcriptomic data, KC and WT cornea collection will also be an ongoing process. Although the RNA-seq data generated in this project was focussed on mRNA transcripts, in the future the analysis will be broadened to study microRNAs and other non-coding RNAs.

From the analytical point of view, the focus of the RNA-seq analysis in this project was mainly directed towards the primary transcripts of protein coding genes. In the future of this project, another aspect of the analysis that will be further advanced is in the identification of novel isoforms and splice variants.

Once a clear specification of elements involved in KC pathogenesis is confirmed, functional analysis is the next step to be achieved.

Collectively, in the era of collaboration between research groups, the WES data generated in this study can also serve as a relatively small dataset that can be included with other similar studies of WES data in order to run a more appropriately powered GWAS on KC patients and thus possibly highlight rare and common susceptibility alleles for KC.

5.2.4 Implications for KC research and patients

In addition to the scientific interest, a better understanding of the pathophysiology of KC could also have clinical implications. It could help clarify the diagnosis, allowing better characterization of forme-fruste KC, and distinguishing this from severe astigmatism through screening of patients for mutations in genes known to be implicated in KC. Thus diagnosis will be strengthened through a blood test, analysed alongside the corneal topographies.

A better understanding of the pathophysiology of the disease could have substantial implications for KC treatment, opening up new therapies far from the conventional treatment approaches that are used currently. Molecular characterising of the different KC subtypes could also facilitate a personalised approach to treatment, improving patient outcomes. It could in some cases clarify whether KC in any given patient is syndromic, which might have wider implications for treatment beyond the eye. For example this approach could clarify whether KC is an inflammatory disorder in some cases.

In the future treatment of KC, for cases where the molecular pathogenesis is determined, slowing down the progression of KC and minimizing the need for corneal grafting ought to be a priority. In the light of personalised medicine, the well phenotyped groups of patients based on genotype, might have customised treatments that are proven to work best in their subtype of KC, hence improving outcomes for KC treatment. Advances in therapeutic strategies, drug specifications and gene therapy might make this a feasible option for the future.

References

ABEDINIA, M., PAIN, T., ALGAR, E. M. & HOLMES, R. S. 1990. Bovine corneal aldehyde dehydrogenase: the major soluble corneal protein with a possible dual protective role for the eye. *Exp Eye Res,* 51**,** 419-26.

ABU-AMERO, K. K., AL-MUAMMAR, A. M. & KONDKAR, A. A. 2014a. Genetics of Keratoconus: Where Do We Stand? *Journal of Ophthalmology*.

ABU-AMERO, K. K., AZAD, T. A., KALANTAN, H., SULTAN, T. & AL-MUAMMAR, A. M. 2014b. Mitochondrial Sequence Changes in Keratoconus Patients. *Investigative Ophthalmology & Visual Science,* 55**,** 1706-1710.

ABU-AMERO, K. K., HELLANI, A. M., AL MANSOURI, S. M., KALANTAN, H. & AL-MUAMMAR, A. M. 2011a. High-resolution analysis of DNA copy number alterations in patients with isolated sporadic keratoconus. *Molecular Vision,* 17**,** 822-826.

ABU-AMERO, K. K., KALANTAN, H. & AL-MUAMMAR, A. M. 2011b. Analysis of the VSX1 gene in keratoconus patients from Saudi Arabia. *Molecular Vision,* 17**,** 667-672.

ADZHUBEI, I., JORDAN, D. M. & SUNYAEV, S. R. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet,* Chapter 7**,** Unit7 20.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

AKEY, J. M., BISWAS, S., LEEK, J. T. & STOREY, J. D. 2007. On the design and analysis of gene expression studies in human populations. *Nat Genet,* 39**,** 807-8; author reply 808-9.

AKHTAR, S., BRON, A. J., SALVI, S. M., HAWKSWORTH, N. R., TUFT, S. J. & MEEK, K. M. 2008. Ultrastructural analysis of collagen fibrils and proteoglycans in keratoconus. *Acta Ophthalmol,* 86**,** 764-72.

AL-MUAMMAR, A. M., KALANTAN, H., AZAD, T. A., SULTAN, T. & ABU-AMERO, K. K. 2015. Analysis of the SOD1 Gene in Keratoconus Patients from Saudi Arabia. *Ophthalmic Genet***,** 1-3.

ALDAVE, A. J. 2005. VSX1 mutation and corneal dystrophies. *Ophthalmology,* 112**,** 170-1; author reply 171-2.

ALDAVE, A. J., BOURLA, N., YELLORE, V. S., RAYNER, S. A., KHAN, M. A., SALEM, A. K. & SONMEZ, B. 2007. Keratoconus is not associated with mutations in COL8A1 and COL8A2. *Cornea,* 26**,** 963-965.

ALDAVE, A. J., YELLORE, V. S., SALEM, A. K., YOO, G. L., RAYNER, S. A., YANG, H. Y., TANG, G. Y., PICONELL, Y. & RABINOWITZ, Y. S. 2006. NoVSX1 gene mutations associated with keratoconus. *Investigative Ophthalmology & Visual Science,* 47**,** 2820-2822.

ALEKSIC, J., CARL, S. H. & FRYE, M. 2014. Beyond library size: a field guide to NGS normalization. *bioRxiv*.

ALTMULLER, J., PALMER, L. J., FISCHER, G., SCHERB, H. & WJST, M. 2001. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet,* 69**,** 936-50.

ALTSHULER, D., DALY, M. J. & LANDER, E. S. 2008. Genetic mapping in human disease. *Science,* 322**,** 881-8.

ALVES, M. M., SRIBUDIANI, Y., BROUWER, R. W., AMIEL, J., ANTINOLO, G., BORREGO, S., CECCHERINI, I., CHAKRAVARTI, A., FERNANDEZ, R. M., GARCIA-BARCELO, M. M., GRISERI, P., LYONNET, S., TAM, P. K., VAN IJCKEN, W. F., EGGEN, B. J., TE MEERMAN, G. J. & HOFSTRA, R. M. 2013. Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev Biol,* 382**,** 320-9.

ANWAR, M. & TEICHMANN, K. D. 2002. Deep lamellar keratoplasty: surgical techniques for anterior lamellar keratoplasty with and without baring of Descemet's membrane. *Cornea,* 21**,** 374-83.

ARNAL, E., PERIS-MARTINEZ, C., MENEZO, J. L., JOHNSEN-SORIANO, S. & ROMERO, F. J. 2011. Oxidative stress in keratoconus? *Invest Ophthalmol Vis Sci,* 52**,** 8592-7.

ASCASO, F. J., DEL BUEY, M. A., HUERVA, V., LATRE, B. & PALOMAR, A. 1993. Noonan's syndrome with keratoconus and optic disc coloboma. *Eur J Ophthalmol,* 3**,** 101-3.

ATILANO, S. R., COSKUN, P., CHWA, M., JORDAN, N., REDDY, V., LE, K., WALLACE, D. C. & KENNEY, M. C. 2005. Accumulation of mitochondrial DNA damage in keratoconus corneas. *Invest Ophthalmol Vis Sci,* 46**,** 1256-63.

AUER, P. L. & DOERGE, R. W. 2010. Statistical design and analysis of RNA sequencing data. *Genetics,* 185**,** 405-16.

BAE, H. A., MILLS, R. A., LINDSAY, R. G., PHILLIPS, T., COSTER, D. J., MITCHELL, P., WANG, J. J., CRAIG, J. E. & BURDON, K. P. 2013. Replication and meta-analysis of candidate loci identified variation at RAB3GAP1 associated with keratoconus. *Invest Ophthalmol Vis Sci,* 54**,** 5132-5.

BAGGERLY, K. A., EDMONSON, S. R., MORRIS, J. S. & COOMBES, K. R. 2004. High-resolution serum proteomic patterns for ovarian cancer detection. *Endocr Relat Cancer,* 11**,** 583-4; author reply 585-7.

BALASUBRAMANIAN, S. A., MOHAN, S., PYE, D. C. & WILLCOX, M. D. 2012. Proteases, proteolysis and inflammatory molecules in the tears of people with keratoconus. *Acta Ophthalmol,* 90**,** e303-9.

BALASUBRAMANIAN, S. A., PYE, D. C. & WILLCOX, M. D. 2010. Are proteinases the reason for keratoconus? *Curr Eye Res,* 35**,** 185-91.

BALASUBRAMANIAN, S. A., PYE, D. C. & WILLCOX, M. D. 2013. Effects of eye rubbing on the levels of protease, protease activity and cytokines in tears: relevance in keratoconus. *Clin Exp Optom,* 96**,** 214-8.

BARALLE, D. & BARALLE, M. 2005. Splicing in action: assessing disease causing sequence changes. *J Med Genet,* 42**,** 737-48.

BARBOUR, B. & SALAMEH, P. 2009. Consanguinity in Lebanon: prevalence, distribution and determinants. *J Biosoc Sci,* 41**,** 505-17.

BARRAQUER, J. I. 1972. Lamellar keratoplasty. (Special techniques). *Ann Ophthalmol,* 4**,** 437-69.

BASAK, S. K., BASAK, S., MOHANTA, A. & BHOWMICK, A. 2005. Epidemiological and microbiological diagnosis of suppurative keratitis in Gangetic West Bengal, eastern India. *Indian J Ophthalmol,* 53**,** 17-22.

BAWAZEER, A. M., HODGE, W. G. & LORIMER, B. 2000. Atopy and keratoconus: a multivariate analysis. *Br J Ophthalmol,* 84**,** 834-6.

BAZ, E. M., MAHFOUZ, R. A., TABARANI, R. S., AYOUB, T. H. & ZAATARI, G. S. 2001. Lebanese population: prevalence of the erythrocyte phenotypes. *J Med Liban,* 49**,** 140-2.

BECKER, K. G., HOSACK, D. A., DENNIS, G., JR., LEMPICKI, R. A., BRIGHT, T. J., CHEADLE, C. & ENGEL, J. 2003. PubMatrix: a tool for multiplex literature mining. *BMC Bioinformatics,* 4**,** 61.

BECKH, U., SCHONHERR, U. & NAUMANN, G. O. 1995. [Autosomal dominant keratoconus as the chief ocular symptom in Lobstein osteogenesis imperfecta tarda]. *Klin Monbl Augenheilkd,* 206**,** 268-72.

BEHNDIG, A., KARLSSON, K., JOHANSSON, B. O., BRANNSTROM, T. & MARKLUND, S. L. 2001. Superoxide dismutase isoenzymes in the normal and diseased human cornea. *Invest Ophthalmol Vis Sci,* 42**,** 2293-6.

BILGIHAN, K., HONDUR, A., SUL, S. & OZTURK, S. 2011. Pregnancy-induced progression of keratoconus. *Cornea,* 30**,** 991-4.

BILGIN, B., UNAL, B., UNAL, M., DOGAN, E., CETINKAYA, A., AKYOL, M., YUCEL, I., AKAR, Y., APAYDIN, C. & ILHAN, D. 2013. Keratoconus presenting with bilateral simultaneous acute corneal hydrops. *Cont Lens Anterior Eye,* 36**,** 98-100.

BISCEGLIA, L., CIASCHETTI, M., DE BONIS, P., CAMPO, P. A. P., PIZZICOLI, C., SCALA, C., GRIFA, M., CIAVARELLA, P., DELLE NOCI, N., VAIRA, F., MACALUSO, C. & ZELANTE, L. 2005. VSX1 mutational analysis in a series of Italian patients affected by keratoconus: Detection of a novel mutation. *Investigative Ophthalmology & Visual Science,* 46**,** 39-45.

BISCEGLIA, L., DE BONIS, P., PIZZICOLI, C., FISCHETTI, L., LABORANTE, A., DI PERNA, M., GIULIANI, F., DELLE NOCI, N., BUZZONETTI, L. & ZELANTE, L. 2009. Linkage Analysis in Keratoconus: Replication of Locus 5q21.2 and Identification of Other Suggestive Loci. *Investigative Ophthalmology & Visual Science,* 50**,** 1081-1086.

BLANDIN, G., MARCHAND, S., CHARTON, K., DANIELE, N., GICQUEL, E., BOUCHETEIL, J. B., BENTAIB, A., BARRAULT, L., STOCKHOLM, D., BARTOLI, M. & RICHARD, I. 2013. A human skeletal muscle interactome centered on proteins involved in muscular dystrophies: LGMD interactome. *Skelet Muscle,* 3**,** 3.

BOCHERT, A., BERLAU, J., KOCZAN, D., SEITZ, B., THIESSEN, H. J. & GUTHOFF, R. F. 2003. Gene expression in keratoconus. Initial results using DNA microarrays. *Ophthalmologe,* 100**,** 545-549.

BOCHERT, A., GERBER, T., KOCZAN, D., SOMMER, U., STACHS, O., THIESSEN, H. J. & GUTHOFF, R. F. 2002. Gene expression and electron-scattering in normal and keratoconus cornea. *Investigative Ophthalmology & Visual Science,* 43**,** U891-U891.

BONNEL, S., BERGUIGA, M., DE RIVOYRE, B., BEDUBOURG, G., SENDON, D., FROUSSART-MAILLE, F. & RIGAL-SASTOURNE, J. C. 2015. Demarcation line evaluation of iontophoresis-assisted transepithelial corneal collagen cross-linking for keratoconus. *J Refract Surg,* 31**,** 36-40.

BRAFMAN, D. A., PHUNG, C., KUMAR, N. & WILLERT, K. 2013. Regulation of endodermal differentiation of human embryonic stem cells through integrin-ECM interactions. *Cell Death Differ,* 20**,** 369-81.

BRANCATI, F., VALENTE, E. M., SARKOZY, A., FEHER, J., CASTORI, M., DEL DUCA, P., MINGARELLI, R., PIZZUTI, A. & DALLAPICCOLA, B.

2004. A locus for autosomal dominant keratoconus maps to human chromosome 3p14-q13. *Journal of Medical Genetics,* 41**,** 188-192.

BRANDAU, O., SCHUSTER, V., WEISS, M., HELLEBRAND, H., FINK, F. M., KRECZY, A., FRIEDRICH, W., STRAHM, B., NIEMEYER, C., BELOHRADSKY, B. H. & MEINDL, A. 1999. Epstein-Barr virus-negative boys with non-Hodgkin lymphoma are mutated in the SH2D1A gene, as are patients with X-linked lymphoproliferative disease (XLP). *Hum Mol Genet,* 8**,** 2407-13.

BRAUN, T. A., MULLINS, R. F., WAGNER, A. H., ANDORF, J. L., JOHNSTON, R. M., BAKALL, B. B., DELUCA, A. P., FISHMAN, G. A., LAM, B. L., WELEBER, R. G., CIDECIYAN, A. V., JACOBSON, S. G., SHEFFIELD, V. C., TUCKER, B. A. & STONE, E. M. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet,* 22**,** 5136-45.

BRON, A. J. 2001. The architecture of the corneal stroma. *Br J Ophthalmol,* 85**,** 379-81.

BROOKES, N. H., LOH, I. P., CLOVER, G. M., POOLE, C. A. & SHERWIN, T. 2003. Involvement of corneal nerves in the progression of keratoconus. *Exp Eye Res,* 77**,** 515-24.

BROWN, D., CHWA, M. M., OPBROEK, A. & KENNEY, M. C. 1993. Keratoconus corneas: increased gelatinolytic activity appears after modification of inhibitors. *Curr Eye Res,* 12**,** 571-81.

BROWN, D. J., LIN, B., CHWA, M., ATILANO, S. R., KIM, D. W. & KENNEY, M. C. 2004. Elements of the nitric oxide pathway can degrade TIMP-1 and increase gelatinase activity. *Mol Vis,* 10**,** 281-8.

BRUNA, F. 1954. [Association of keratoconus with retinitis pigmentosa]. *Boll Ocul,* 33**,** 145-57.

BUDDI, R., LIN, B., ATILANO, S. R., ZORAPAPEL, N. C., KENNEY, M. C. & BROWN, D. J. 2002. Evidence of oxidative stress in human corneal diseases. *J Histochem Cytochem,* 50**,** 341-51.

BURDON, K. P., COSTER, D. J., CHARLESWORTH, J. C., MILLS, R. A., LAURIE, K. J., GIUNTA, C., HEWITT, A. W., LATIMER, P. & CRAIG, J. E. 2008. Apparent autosomal dominant keratoconus in a large Australian pedigree accounted for by digenic inheritance of two novel loci. *Human Genetics,* 124**,** 379-386.

BURDON, K. P., MACGREGOR, S., BYKHOVSKAYA, Y., JAVADIYAN, S., LI, X., LAURIE, K. J., MUSZYNSKA, D., LINDSAY, R., LECHNER, J., HARITUNIANS, T., HENDERS, A. K., DASH, D., SISCOVICK, D., ANAND, S., ALDAVE, A., COSTER, D. J., SZCZOTKA-FLYNN, L., MILLS, R. A., IYENGAR, S. K., TAYLOR, K. D., PHILLIPS, T., MONTGOMERY, G. W., ROTTER, J. I., HEWITT, A. W., SHARMA, S., RABINOWITZ, Y. S., WILLOUGHBY, C. & CRAIG, J. E. 2011. Association of Polymorphisms in the Hepatocyte Growth Factor Gene Promoter with Keratoconus. *Investigative Ophthalmology & Visual Science,* 52**,** 8514-8519.

BURDON, K. P. & VINCENT, A. L. 2013. Insights into keratoconus from a genetic perspective. *Clinical and Experimental Optometry,* 96**,** 146-154.

BUREAU, J., POULIQUEN, Y. & LORANS, G. 1993. FIBROCYTE RESPONSE TO INTERLEUKIN-1 STIMULATION IN KERATOCONUS. *Klinische Monatsblatter Fur Augenheilkunde,* 203**,** 269-274.

BURRIS, T. E., AYER, C. T., EVENSEN, D. A. & DAVENPORT, J. M. 1991. Effects of intrastromal corneal ring size and thickness on corneal flattening in human eyes. *Refract Corneal Surg,* 7**,** 46-50.

BYKHOVSKAYA, Y., LI, X., EPIFANTSEVA, I., HARITUNIANS, T., SISCOVICK, D., ALDAVE, A., SZCZOTKA-FLYNN, L., IYENGAR, S. K., TAYLOR, K. D., ROTTER, J. I. & RABINOWITZ, Y. S. 2012. Variation in the Lysyl Oxidase (LOX) Gene Is Associated with Keratoconus in Family-Based and Case-Control Studies. *Investigative Ophthalmology & Visual Science,* 53**,** 4152-4157.

CALHOUN, F. P., JR. 1951. Degenerations and dystrophies. *Trans Am Acad Ophthalmol Otolaryngol,* 55**,** 366-81.

CAMARGO, L. M., COLLURA, V., RAIN, J. C., MIZUGUCHI, K., HERMJAKOB, H., KERRIEN, S., BONNERT, T. P., WHITING, P. J. & BRANDON, N. J. 2007. Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psychiatry,* 12**,** 74-86.

CARR, I. M., BHASKAR, S., O'SULLIVAN, J., ALDAHMESH, M. A., SHAMSELDIN, H. E., MARKHAM, A. F., BONTHRON, D. T., BLACK, G. & ALKURAYA, F. S. 2013. Autozygosity mapping with exome sequence data. *Hum Mutat,* 34**,** 50-6.

CARR, I. M., SHERIDAN, E., HAYWARD, B. E., MARKHAM, A. F. & BONTHRON, D. T. 2009. IBDfinder and SNPsetter: tools for pedigree-independent identification of autozygous regions in individuals with recessive inherited disease. *Hum Mutat,* 30**,** 960-7.

CASSIDY, D., BELTZ, J., JHANJI, V. & LOUGHNAN, M. S. 2013. Recent advances in corneal transplantation for keratoconus. *Clin Exp Optom,* 96**,** 165-72.

CAVALIERI, S., POZZI, E., GATTI, R. A. & BRUSCO, A. 2013. Deep-intronic ATM mutation detected by genomic resequencing and corrected in vitro by antisense morpholino oligonucleotide (AMO). *Eur J Hum Genet,* 21**,** 774-8.

CEJKOVA, J., STIPEK, S., CRKOVSKA, J., ARDAN, T., PLATENIK, J., CEJKA, C. & MIDELFART, A. 2004a. UV Rays, the prooxidant/antioxidant imbalance in the cornea and oxidative eye damage. *Physiol Res,* 53**,** 1-10.

CEJKOVA, J., VEJRAZKA, M., PLATENIK, J. & STIPEK, S. 2004b. Age-related changes in superoxide dismutase, glutathione peroxidase, catalase and xanthine oxidoreductase/xanthine oxidase activities in the rabbit cornea. *Exp Gerontol,* 39**,** 1537-43.

CHAERKADY, R., SHAO, H., SCOTT, S.-G., PANDEY, A., JUN, A. S. & CHAKRAVARTI, S. 2013. The keratoconus corneal proteome: Loss of epithelial integrity and stromal degeneration. *Journal of Proteomics,* 87**,** 122-131.

CHANDA, B., ASAI-COAKWELL, M., YE, M., MUNGALL, A. J., BARROW, M., DOBYNS, W. B., BEHESTI, H., SOWDEN, J. C., CARTER, N. P., WALTER, M. A. & LEHMANN, O. J. 2008. A novel mechanistic spectrum underlies glaucoma-associated chromosome 6p25 copy number variation. *Hum Mol Genet,* 17**,** 3446-58.

CHEN, C. J. & FURR, P. 1983. Bilateral keratoconus in a patient with gyrate atrophy and hyperornithinemia. *Am J Ophthalmol,* 95**,** 705-6.

CHEN, Y., MCCARTHY, D., ROBINSON, M. & SMYTH, G. K. 2008, revised 2015. edgeR: differential expression analysis of digital gene expression data (User's Guide). http://www.bioconductor.org.

CHENG, E. L., MARUYAMA, I., SUNDARRAJ, N., SUGAR, J., FEDER, R. S. & YUE, B. Y. 2001. Expression of type XII collagen and hemidesmosome-associated proteins in keratoconus corneas. *Curr Eye Res,* 22**,** 333-40.

CHERNYAVSKY, A. I., GALITOVSKIY, V., SHCHEPOTIN, I. B., JESTER, J. V. & GRANDO, S. A. 2014. The acetylcholine signaling network of corneal epithelium and its role in regulation of random and directional migration of corneal epithelial cells. *Invest Ophthalmol Vis Sci,* 55**,** 6921-33.

CHEUNG, I. M. Y., MCGHEE, C. N. J. & SHERWIN, T. 2013. A new perspective on the pathobiology of keratoconus: interplay of stromal wound healing and reactive species-associated processes. *Clinical and Experimental Optometry,* 96**,** 188-196.

CHEUNG, I. M. Y., MCGHEE, C. N. J. & SHERWIN, T. 2014. Deficient repair regulatory response to injury in keratoconic stromal cells. *Clinical and Experimental Optometry,* 97**,** 234-239.

CHUN, S. & FAY, J. C. 2009. Identification of deleterious mutations within three human genomes. *Genome Res,* 19**,** 1553-61.

CHWA, M., ATILANO, S. R., HERTZOG, D., ZHENG, H., LANGBERG, J., KIM, D. W. & KENNEY, M. C. 2008. Hypersensitive response to oxidative stress in keratoconus corneal fibroblasts. *Investigative Ophthalmology & Visual Science,* 49**,** 4361-4369.

CHWA, M., ATILANO, S. R., REDDY, V., JORDAN, N., KIM, D. W. & KENNEY, M. C. 2006. Increased stress-induced generation of reactive oxygen species and apoptosis in human keratoconus fibroblasts. *Invest Ophthalmol Vis Sci,* 47**,** 1902-10.

CHWIERALSKI, C. E., WELTE, T. & BUHLING, F. 2006. Cathepsin-regulated apoptosis. *Apoptosis,* 11**,** 143-9.

CINAR, Y., CINGU, A. K., TURKCU, F. M., CINAR, T., YUKSEL, H., OZKURT, Z. G. & CACA, I. 2014. Comparison of accelerated and conventional corneal collagen cross-linking for progressive keratoconus. *Cutan Ocul Toxicol,* 33**,** 218-22.

CIRULLI, E. T. & GOLDSTEIN, D. B. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet,* 11**,** 415-25.

COLIN, J., COCHENER, B., SAVARY, G. & MALET, F. 2000. Correcting keratoconus with intracorneal rings. *J Cataract Refract Surg,* 26**,** 1117-22.

COLLIER, S. A. 2001. Is the corneal degradation in keratoconus caused by matrix-metalloproteinases? *Clin Experiment Ophthalmol,* 29**,** 340-4.

COLLIER, S. A., MADIGAN, M. C. & PENFOLD, P. L. 2000. Expression of membrane-type 1 matrix metalloproteinase (MT1-MMP) and MMP-2 in normal and keratoconus corneas. *Curr Eye Res,* 21**,** 662-8.

COOPER, G. M., STONE, E. A., ASIMENOS, G., GREEN, E. D., BATZOGLOU, S. & SIDOW, A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res,* 15**,** 901-13.

COSTA, V., APRILE, M., ESPOSITO, R. & CICCODICOLA, A. 2013. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *Eur J Hum Genet,* 21**,** 134-42.

COYLE, J. T. 1984. Keratoconus and eye rubbing. *Am J Ophthalmol,* 97**,** 527-8.

CRAWFORD, D. R., WANG, Y., SCHOOLS, G. P., KOCHHEISER, J. & DAVIES, K. J. 1997. Down-regulation of mammalian mitochondrial RNAs during oxidative stress. *Free Radic Biol Med,* 22**,** 551-9.

CREWS, M. J., DRIEBE, W. T., JR. & STERN, G. A. 1994. The clinical management of keratoconus: a 6 year retrospective study. *CLAO J,* 20**,** 194-7.

CRISTINA KENNEY, M. & BROWN, D. J. 2003. The cascade hypothesis of keratoconus. *Cont Lens Anterior Eye,* 26**,** 139-46.

CUI, L., ZHOU, H., YU, J., SUN, E., ZHANG, Y., JIA, W., JIAO, Y., SNELLINGEN, T., LIU, X., LIM, A., WANG, N. & LIU, N. 2010. Noncoding variant in the complement factor H gene and risk of exudative age-related macular degeneration in a Chinese population. *Invest Ophthalmol Vis Sci,* 51**,** 1116-20.

CZUGALA, M., KAROLAK, J. A., NOWAK, D. M., POLAKOWSKI, P., PITARQUE, J., MOLINARI, A., RYDZANICZ, M., BEJJANI, B. A., YUE, B. Y., SZAFLIK, J. P. & GAJECKA, M. 2012a. Novel mutation and three other sequence variants segregating with phenotype at keratoconus 13q32 susceptibility locus. *Eur J Hum Genet,* 20**,** 389-97.

CZUGALA, M., KAROLAK, J. A., NOWAK, D. M., POLAKOWSKI, P., PITARQUE, J., MOLINARI, A., RYDZANICZ, M., BEJJANI, B. A., YUE, B. Y. J. T., SZAFLIK, J. P. & GAJECKA, M. 2012b. Novel mutation and three other sequence variants segregating with phenotype at keratoconus 13q32 susceptibility locus. *European Journal of Human Genetics,* 20**,** 389-397.

DAMJI, K. F., SOHOCKI, M. M., KHAN, R., GUPTA, S. K., RAHIM, M., LOYER, M., HUSSEIN, N., KARIM, N., LADAK, S. S., JAMAL, A., BULMAN, D. & KOENEKOOP, R. K. 2001. Leber's congenital amaurosis with anterior keratoconus in Pakistani families is caused by the Trp278X mutation in the AIPL1 gene on 17p. *Canadian Journal of Ophthalmology-Journal Canadien D Ophtalmologie,* 36**,** 252-259.

DART, J. K., RADFORD, C. F., MINASSIAN, D., VERMA, S. & STAPLETON, F. 2008. Risk factors for microbial keratitis with contemporary contact lenses: a case-control study. *Ophthalmology,* 115**,** 1647-54, 1654 e1-3.

DART, J. K., STAPLETON, F. & MINASSIAN, D. 1991. Contact lenses and other risk factors in microbial keratitis. *Lancet,* 338**,** 650-3.

DASH, D. P., GEORGE, S., O'PREY, D., BURNS, D., NABILI, S., DONNELLY, U., HUGHES, A. E., SILVESTRI, G., JACKSON, J., FRAZER, D., HEON, E. & WILLOUGHBY, C. E. 2010. Mutational screening of VSX1 in keratoconus patients from the European population. *Eye,* 24**,** 1085-1092.

DAVIDSON, A. E., BORASIO, E., LISKOVA, P., KHAN, A. O., HASSAN, H., CHEETHAM, M. E., PLAGNOL, V., ALKURAYA, F. S., TUFT, S. J. & HARDCASTLE, A. J. 2015. Brittle Cornea Syndrome ZNF469 Mutation Carrier Phenotype and Segregation Analysis of Rare ZNF469 Variants in Familial Keratoconus. *Investigative ophthalmology & visual science,* 56**,** 578-86.

DAVIDSON, A. E., HAYES, S., HARDCASTLE, A. J. & TUFT, S. J. 2014. The pathogenesis of keratoconus. *Eye,* 28**,** 189-195.

DAVYDOV, E. V., GOODE, D. L., SIROTA, M., COOPER, G. M., SIDOW, A. & BATZOGLOU, S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol,* 6**,** e1001025.

DE BAERE, E. 2014. Heterozygous Coding ZNF469 Variants Enriched in New Zealand Patients With Isolated Keratoconus. *Investigative Ophthalmology & Visual Science,* 55**,** 5636-5636.

DE BONIS, P., LABORANTE, A., PIZZICOLI, C., STALLONE, R., BARBANO, R., LONGO, C., MAZZILLI, E., ZELANTE, L. & BISCEGLIA, L. 2011. Mutational screening of VSX1, SPARC, SOD1, LOX, and TIMP3 in keratoconus. *Molecular Vision,* 17**,** 2482-2494.

DEHKORDI, F. A., RASHKI, A., BAGHERI, N., CHALESHTORI, M. H., MEMARZADEH, E., SALEHI, A., GHATREH, H., ZANDI, F., YAZDANPANAHI, N., TABATABAIEFAR, M. A. & CHALESHTORI, M. H. 2013. Study of VSX1 Mutations in Patients with Keratoconus in Southwest Iran Using PCR-Single-Strand Conformation Polymorphism/Heteroduplex Analysis and Sequencing Method. *Acta Cytologica,* 57**,** 646-651.

DELMONTE, D. W. & KIM, T. 2011. Anatomy and physiology of the cornea. *J Cataract Refract Surg,* 37**,** 588-98.

DI, Y., SCHAFER DANIEL, W., CUMBIE JASON, S. & CHANG JEFF, H. 2011. The NBP Negative Binomial Model for Assessing Differential Gene Expression from RNA-Seq. *Statistical Applications in Genetics and Molecular Biology.*

DILLIES, M.-A., RAU, A., AUBERT, J., HENNEQUET-ANTIER, C., JEANMOUGIN, M., SERVANT, N., KEIME, C., MAROT, G., CASTEL, D., ESTELLE, J., GUERNEC, G., JAGLA, B., JOUNEAU, L., LALOË, D., LE GALL, C., SCHAËFFER, B., LE CROM, S., GUEDJ, M. & JAFFRÉZIC, F. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics,* 14**,** 671-683.

DOI, M., THYBOLL, J., KORTESMAA, J., JANSSON, K., IIVANAINEN, A., PARVARDEH, M., TIMPL, R., HEDIN, U., SWEDENBORG, J. & TRYGGVASON, K. 2002. Recombinant human laminin-10 (alpha5beta1gamma1). Production, purification, and migration-promoting activity on vascular endothelial cells. *J Biol Chem,* 277**,** 12741-8.

DOUVARAS, P., LIU, W., MORT, R. L., MCKIE, L., WEST, K. M., CROSS, S. H., MORLEY, S. D. & WEST, J. D. 2012. Normal X-inactivation mosaicism in corneas of heterozygous Flna(Dilp2/+)female mice--a model of human Filamin A (FLNA) diseases. *BMC Research Notes,* 5**,** 122-122.

DROITCOURT, C., TOUBOUL, D., GED, C., EZZEDINE, K., CARIO-ANDRE, M., DE VERNEUIL, H., COLIN, J. & TAIEB, A. 2011. A Prospective Study of Filaggrin Null Mutations in Keratoconus Patients with or without Atopic Disorders. *Dermatology,* 222**,** 336-341.

DUA, H. S., FARAJ, L. A., SAID, D. G., GRAY, T. & LOWE, J. 2013. Human corneal anatomy redefined: a novel pre-Descemet's layer (Dua's layer). *Ophthalmology,* 120**,** 1778-85.

DUDAKOVA, L. & JIRSOVA, K. 2013. The impairment of lysyl oxidase in keratoconus and in keratoconus-associated disorders. *J Neural Transm,* 120**,** 977-82.

DUDAKOVA, L., LISKOVA, P., TROJEK, T., PALOS, M., KALASOVA, S. & JIRSOVA, K. 2012. Changes in lysyl oxidase (LOX) distribution and its decreased activity in keratoconus corneas. *Experimental Eye Research,* 104**,** 74-81.

DUDAKOVA, L., PALOS, M., JIRSOVA, K., STRANECKY, V., KREPELOVA, A., HYSI, P. G. & LISKOVA, P. 2015a. Validation of rs2956540:G>C and rs3735520:G>A association with keratoconus in a population of European descent. *Eur J Hum Genet*.

DUDAKOVA, L., SASAKI, T., LISKOVA, P., PALOS, M. & JIRSOVA, K. 2015b. The presence of lysyl oxidase-like enzymes in human control and keratoconic corneas. *Histol Histopathol***,** 11649.

DURAN, J. A. & LEMA, I. 2003. Inflammatory Markers in Keratoconus. *Investigative Ophthalmology & Visual Science,* 44**,** 1314-1314.

EDWARDS, M., MCGHEE, C. N. & DEAN, S. 2001. The genetics of keratoconus. *Clin Experiment Ophthalmol,* 29**,** 345-51.

EICHLER, E. E., FLINT, J., GIBSON, G., KONG, A., LEAL, S. M., MOORE, J. H. & NADEAU, J. H. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet,* 11**,** 446-50.

EL-ASRAG, M. E., SERGOUNIOTIS, P. I., MCKIBBIN, M., PLAGNOL, V., SHERIDAN, E., WASEEM, N., ABDELHAMED, Z., MCKEEFRY, D., VAN SCHIL, K., POULTER, J. A., JOHNSON, C. A., CARR, I. M., LEROY, B. P., DE BAERE, E., INGLEHEARN, C. F., WEBSTER, A. R., TOOMES, C. & ALI, M. 2015. Biallelic mutations in the autophagy regulator DRAM2 cause retinal dystrophy with early macular involvement. *Am J Hum Genet,* 96**,** 948-54.

ERAN, P., ALMOGIT, A., DAVID, Z., WOLF, H. R., HANA, G., YANIV, B., ELON, P. & ISAAC, A. 2008. The D144E substitution in the VSX1 gene: A non-pathogenic variant or a disease causing mutation? *Ophthalmic Genetics,* 29**,** 53-59.

ERTAN, A. & MUFTUOGLU, O. 2008. Keratoconus clinical findings according to different age and gender groups. *Cornea,* 27**,** 1109-13.

ETZINE, S. 1954. Keratoconus and mental defect. *S Afr Med J,* 28**,** 520-1.

FARKAS, M. H., GRANT, G. R., WHITE, J. A., SOUSA, M. E., CONSUGAR, M. B. & PIERCE, E. A. 2013. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics,* 14**,** 486.

FAVRE, B., FONTAO, L., KOSTER, J., SHAFAATIAN, R., JAUNIN, F., SAURAT, J. H., SONNENBERG, A. & BORRADORI, L. 2001. The hemidesmosomal protein bullous pemphigoid antigen 1 and the integrin beta 4 subunit bind to ERBIN. Molecular cloning of multiple alternative splice variants of ERBIN and analysis of their tissue expression. *J Biol Chem,* 276**,** 32427-36.

FIGUEIREDO, F. 2013. STANDARDS FOR THE RETRIEVAL OF HUMAN OCULAR TISSUE USED IN TRANSPLANTATION, RESEARCH AND TRAINING. *Royal College of Ophthalmologists*.

FILENIUS, S., TERVO, T. & VIRTANEN, I. 2003. Production of fibronectin and tenascin isoforms and their role in the adhesion of human immortalized corneal epithelial cells. *Invest Ophthalmol Vis Sci,* 44**,** 3317-25.

FINI, M. E. & STRAMER, B. M. 2005. How the cornea heals: cornea-specific repair mechanisms affecting surgical outcomes. *Cornea,* 24**,** S2-S11.

FINK, B. A., WAGNER, H., STEGER-MAY, K., ROSENSTIEL, C., ROEDIGER, T., MCMAHON, T. T., GORDON, M. O. & ZADNIK, K. 2005. Differences in keratoconus as a function of gender. *Am J Ophthalmol,* 140**,** 459-68.

FINOTELLO, F. & DI CAMILLO, B. 2015. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Brief Funct Genomics,* 14**,** 130-42.

FOGLA, R. & IYER, G. K. 2002. Keratoconus associated with cone-rod dystrophy: a case report. *Cornea,* 21**,** 331-2.

FOROUTAN, A. R. & DASTJERDI, M. H. 2007. Shifting-bubble sign in big-bubble technique in deep anterior lamellar keratoplasty. *Cornea,* 26**,** 117; author reply 117-8.

FRANCESCHETTI, A. 1954. Classification and treatment of hereditary corneal dystrophies. *AMA Arch Ophthalmol,* 52**,** 1-12.

FRANCOIS, J., NEETENS, A. & SMETS, R. M. 1982. Bardet-Biedl syndrome and keratoconus. *Bull Soc Belge Ophtalmol,* 203**,** 117-21.

FRAUSTO, R. F., WANG, C. & ALDAVE, A. J. 2014. Transcriptome Analysis of the Human Corneal Endothelium. *Investigative Ophthalmology & Visual Science,* 55**,** 7821-7830.

FULLERTON, J., PAPROCKI, P., FOOTE, S., MACKEY, D. A., WILLIAMSON, R. & FORREST, S. 2002. Identity-by-descent approach to gene localisation in eight individuals affected by keratoconus from north-west Tasmania, Australia. *Human Genetics,* 110**,** 462-470.

GAJECKA, M., RADHAKRISHNA, U., WINTERS, D., NATH, S. K., RYDZANICZ, M., RATNAMALA, U., EWING, K., MOLINARI, A., PITARQUE, J. A., LEE, K., LEAL, S. M. & BEJJANI, B. A. 2009. Localization of a Gene for Keratoconus to a 5.6-Mb Interval on 13q32. *Investigative Ophthalmology & Visual Science,* 50**,** 1531-1539.

GALVIS, V., SHERWIN, T., TELLO, A., MERAYO, J., BARRERA, R. & ACERA, A. 2015a. Keratoconus: an inflammatory disorder? *Eye (Lond),* 29**,** 843-59.

GALVIS, V., TELLO, A., BARRERA, R. & NINO, C. A. 2015b. Inflammation in Keratoconus. *Cornea,* 34**,** e22-3.

GARBER, M., GRABHERR, M. G., GUTTMAN, M. & TRAPNELL, C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods,* 8**,** 469-77.

GARBER, M., GUTTMAN, M., CLAMP, M., ZODY, M. C., FRIEDMAN, N. & XIE, X. 2009. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics,* 25**,** i54-i62.

GARCIA-ALCALDE, F., OKONECHNIKOV, K., CARBONELL, J., CRUZ, L. M., GOTZ, S., TARAZONA, S., DOPAZO, J., MEYER, T. F. & CONESA, A. 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics,* 28**,** 2678-9.

GARCIA-LLEDO, M., FEINBAUM, C. & ALIO, J. L. 2006. Contact lens fitting in keratoconus. *Compr Ophthalmol Update,* 7**,** 47-52.

GASSET, A. R., HINSON, W. A. & FRIAS, J. L. 1978. KERATOCONUS AND ATOPIC DISEASES. *Annals of Ophthalmology,* 10**,** 991-994.

GEORGIOU, T., FUNNELL, C. L., CASSELS-BROWN, A. & O'CONOR, R. 2004. Influence of ethnic origin on the incidence of keratoconus and associated atopic disease in Asians and white patients. *Eye (Lond),* 18**,** 379-83.

GEUIJEN, C. A. & SONNENBERG, A. 2002. Dynamics of the alpha6beta4 integrin in keratinocytes. *Mol Biol Cell,* 13**,** 3845-58.

GHARAEE, H., ABRISHAMI, M., SHAFIEE, M. & EHSAEI, A. 2014. White-to-white corneal diameter: normal values in healthy Iranian population obtained with the Orbscan II. *Int J Ophthalmol,* 7**,** 309-12.

GHOSH, A., ZHOU, L., GHOSH, A., SHETTY, R. & BEUERMAN, R. 2013. Proteomic and gene expression patterns of keratoconus. *Indian Journal of Ophthalmology,* 61**,** 389-391.

GHOSHEH, F. R., CREMONA, F. A., RAPUANO, C. J., COHEN, E. J., AYRES, B. D., HAMMERSMITH, K. M., RABER, I. M. & LAIBSON, P. R. 2008. Trends in penetrating keratoplasty in the United States 1980-2005. *Int Ophthalmol,* 28**,** 147-53.

GOEL, S., GANGER, A. & GUPTA, V. 2015. Bilateral juvenile onset primary open-angle glaucoma among keratoconus patients. *J Glaucoma,* 24**,** e25-7.

GOMES, J. A., TAN, D., RAPUANO, C. J., BELIN, M. W., AMBROSIO, R., JR., GUELL, J. L., MALECAZE, F., NISHIDA, K. & SANGWAN, V. S. 2015. Global consensus on keratoconus and ectatic diseases. *Cornea,* 34**,** 359-69.

GONCU, T., AKAL, A., ADBELLI, F. M., CAKMAK, S., SEZEN, H. & YLMAZ, O. F. 2015. Tear Film and Serum Prolidase Activity and Oxidative Stress in Patients With Keratoconus. *Cornea,* 34**,** 1019-23.

GONDHOWIARDJO, T. D., VAN HAERINGEN, N. J., HOEKZEMA, R., PELS, L. & KIJLSTRA, A. 1991. Detection of aldehyde dehydrogenase activity in human corneal extracts. *Curr Eye Res,* 10**,** 1001-7.

GORDON-SHAAG, A., MILLODOT, M., ESSA, M., GARTH, J., GHARA, M. & SHNEOR, E. 2013. Is Consanguinity a Risk Factor for Keratoconus? *Optometry and Vision Science,* 90**,** 448-454.

GORDON-SHAAG, A., MILLODOT, M., SHNEOR, E. & LIU, Y. 2015. The genetic and environmental factors for keratoconus. *Biomed Res Int,* 2015**,** 795738.

GORSKOVA, E. N. & SEVOST'IANOV, E. N. 1998. [Epidemiology of keratoconus in the Urals]. *Vestn Oftalmol,* 114**,** 38-40.

GRANT, G. R., FARKAS, M. H., PIZARRO, A. D., LAHENS, N. F., SCHUG, J., BRUNK, B. P., STOECKERT, C. J., HOGENESCH, J. B. & PIERCE, E. A. 2011. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics,* 27**,** 2518-28.

GRITZ, D. C. & MCDONNELL, P. J. 1988. Keratoconus and ocular massage. *Am J Ophthalmol,* 106**,** 757-8.

GRZYBOWSKI, A. & MCGHEE, C. N. 2013. The early history of keratoconus prior to Nottingham's landmark 1854 treatise on conical cornea: a review. *Clin Exp Optom,* 96**,** 140-5.

GUAN, T., LIU, C., MA, Z. & DING, S. 2012. The point mutation and polymorphism in keratoconus candidate gene TGFBI in Chinese population. *Gene,* 503**,** 137-139.

GUEDEZ, L., STETLER-STEVENSON, W. G., WOLFF, L., WANG, J., FUKUSHIMA, P., MANSOOR, A. & STETLER-STEVENSON, M. 1998. In vitro suppression of programmed cell death of B cells by tissue inhibitor of metalloproteinases-1. *J Clin Invest,* 102**,** 2002-10.

GUERIN, M., E, O. C., WALSH, C. & FULCHER, T. 2008. Visual outcomes and graft survival following corneal transplants: the need for an Irish National Corneal Transplant Registry. *Ir J Med Sci,* 177**,** 107-10.

HA, N. T., NAKAYASU, K., MURAKAMI, A., ISHIDOH, K. & KANAI, A. 2003. Identification of differentially expressed genes in keratoconus keratocyte by microarray analysis. *Investigative Ophthalmology & Visual Science,* 44**,** U238-U238.

HA, N. T., NAKAYASU, K., MURAKAMI, A., ISHIDOH, K. & KANAI, A. 2004. Microarray analysis identified differentially expressed genes in keratocytes from keratoconus patients. *Current Eye Research,* 28**,** 373-379.

HAIMOVICI, R. & CULBERTSON, W. W. 1991. Optical lamellar keratoplasty using the barraquer microkeratome. *Refract Corneal Surg,* 7**,** 42-5.

HAMEED, A., KHALIQ, S., ISMAIL, M., EBENEZER, N. D., JORDAN, T., MEHDI, S. Q., PAYNE, A. M. & BHATTACHARYA, S. S. 1999. A novel locus for Leber congential amaurosis (LCA) with anterior keratoconus mapping to chromosome 17p13. *American Journal of Human Genetics,* 65**,** A253-A253.

HAMILL, K. J., HOPKINSON, S. B., DEBIASE, P. & JONES, J. C. 2009. BPAG1e maintains keratinocyte polarity through beta4 integrin-mediated modulation of Rac1 and cofilin activities. *Mol Biol Cell,* 20**,** 2954-62.

HAMMERSTEIN, W. 1971. [Significance of the sex ratio in the determination of x-chromosomal heredity demonstrated on the clinical picture of keratoconus]. *Klin Monbl Augenheilkd,* 159**,** 602-8.

HAO, X. D., CHEN, P., CHEN, Z. L., LI, S. X. & WANG, Y. 2015a. Evaluating the Association between Keratoconus and Reported Genetic Loci in a Han Chinese Population. *Ophthalmic Genet,* 36**,** 132-6.

HAO, X. D., CHEN, P., WANG, Y., LI, S. X. & XIE, L. X. 2015b. Mitochondrial DNA copy number, but not haplogroup is associated with keratoconus in Han Chinese population. *Experimental eye research,* 132**,** 59-63.

HAROLD A. STEIN, R. M. S., MELVIN I. FREEMAN 2013. *The Ophthalmic Assistant, 9th Edition*, Canada: Elsevier Mosby.

HASHEMI, H., KHABAZKHOOB, M., YAZDANI, N., OSTADIMOGHADDAM, H., NOROUZIRAD, R., AMANZADEH, K., MIRAFTAB, M., DERAKHSHAN, A. & YEKTA, A. 2014. The prevalence of keratoconus in a young population in Mashhad, Iran. *Ophthalmic Physiol Opt,* 34**,** 519-27.

HASSELL, J. R. & BIRK, D. E. 2010. The molecular basis of corneal transparency. *Exp Eye Res,* 91**,** 326-35.

HEAVEN, C. J., LALLOO, F. & MCHALE, E. 2000. Keratoconus associated with chromosome 13 ring abnormality. *Br J Ophthalmol,* 84**,** 1079.

HEIN, M. Y., HUBNER, N. C., POSER, I., COX, J., NAGARAJ, N., TOYODA, Y., GAK, I. A., WEISSWANGE, I., MANSFELD, J., BUCHHOLZ, F., HYMAN, A. A. & MANN, M. 2015. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell,* 163**,** 712-23.

HELBIG, I., HODGE, S. E. & OTTMAN, R. 2013. Familial cosegregation of rare genetic variants with disease in complex disorders. *Eur J Hum Genet,* 21**,** 444-50.

HELENA, M. C., BAERVELDT, F., KIM, W. J. & WILSON, S. E. 1998. Keratocyte apoptosis after corneal surgery. *Invest Ophthalmol Vis Sci,* 39**,** 276-83.

HEON, E., GREENBERG, A., KOPP, K. K., ROOTMAN, D., VINCENT, A. L., BILLINGSLEY, G., PRISTON, M., DORVAL, K. M., CHOW, R. L., MCINNES, R. R., HEATHCOTE, G., WESTALL, C., SUTPHIN, J. E., SEMINA, E., BREMNER, R. & STONE, E. M. 2002a. VSX1: A gene for posterior polymorphous dystrophy and keratoconus. *Human Molecular Genetics,* 11**,** 1029-1036.

HEON, E., KOPP, K. K., BREMNER, R., ROOTMAN, D., CHOW, R. L., DORVAL, D., VINCENT, A. L., WESTALL, C., SUTPHIN, J. E. & STONE, E. M. 2002b. VSX1: A gene for posterior polymorphous dystrophy and keratoconus. *Investigative Ophthalmology & Visual Science,* 43**,** U809-U809.

HERDEWYN, S., ZHAO, H., MOISSE, M., RACE, V., MATTHIJS, G., REUMERS, J., KUSTERS, B., SCHELHAAS, H. J., VAN DEN BERG, L.

H., GORIS, A., ROBBERECHT, W., LAMBRECHTS, D. & VAN DAMME, P. 2012. Whole-genome sequencing reveals a coding non-pathogenic variant tagging a non-coding pathogenic hexanucleotide repeat expansion in C9orf72 as cause of amyotrophic lateral sclerosis. *Hum Mol Genet,* 21**,** 2412-9.

HINTERMANN, E., YANG, N., O'SULLIVAN, D., HIGGINS, J. M. & QUARANTA, V. 2005. Integrin alpha6beta4-erbB2 complex inhibits haptotaxis by up-regulating E-cadherin cell-cell junctions in keratinocytes. *J Biol Chem,* 280**,** 8004-15.

HOANG-XUAN, T., ELMALEH, C., DHERMY, P., SAVOLDELLI, M., D'HERMIES, F. & POULIQUEN, Y. 1989. [Association of a lattice dystrophy and keratoconus: anatomo-clinical study apropos of a case]. *Bull Soc Ophtalmol Fr,* 89**,** 35-8.

HONDA, N., MIYAI, T., NEJIMA, R., MIYATA, K., MIMURA, T., USUI, T., AIHARA, M., ARAIE, M. & AMANO, S. 2010. Effect of latanoprost on the expression of matrix metalloproteinases and tissue inhibitor of metalloproteinase 1 on the ocular surface. *Arch Ophthalmol,* 128**,** 466-71.

HONNAPPA, S., GOUVEIA, S. M., WEISBRICH, A., DAMBERGER, F. F., BHAVESH, N. S., JAWHARI, H., GRIGORIEV, I., VAN RIJSSEL, F. J., BUEY, R. M., LAWERA, A., JELESAROV, I., WINKLER, F. K., WUTHRICH, K., AKHMANOVA, A. & STEINMETZ, M. O. 2009. An EB1-binding motif acts as a microtubule tip localization signal. *Cell,* 138**,** 366-76.

HORAN, M. P. 2009. Application of serial analysis of gene expression to the study of human genetic disease. *Hum Genet,* 126**,** 605-14.

HUGHES, A. E., BRADLEY, D. T., CAMPBELL, M., LECHNER, J., DASH, D. P., SIMPSON, D. A. & WILLOUGHBY, C. E. 2011. Mutation Altering the miR-184 Seed Region Causes Familial Keratoconus with Cataract. *American Journal of Human Genetics,* 89**,** 628-633.

HUGHES, A. E., DASH, D. P., JACKSON, A. J., FRAZER, D. G. & SILVESTRI, G. 2003. Familial keratoconus with cataract: Linkage to the long arm of chromosome 15 and exclusion of candidate genes. *Investigative Ophthalmology & Visual Science,* 44**,** 5063-5066.

HUTCHINGS, H., GINISTY, H., LE GALLO, M., LEVY, D., STOESSER, F., ROULAND, J. F., ARNE, J. L., LALAUX, M. H., CALVAS, P., ROTH, M. P., HOVNANIAN, A. & MALECAZE, F. 2005. Identification of a new locus for isolated familial keratoconus at 2p24. *Journal of Medical Genetics,* 42**,** 88-94.

IHAKA, R. & GENTLEMAN, R. 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics,* 5**,** 299-314.

IHALAINEN, A. 1986. Clinical and epidemiological features of keratoconus genetic and external factors in the pathogenesis of the disease. *Acta ophthalmologica. Supplement,* 178**,** 1-64.

IHANAMAKI, T., PELLINIEMI, L. J. & VUORIO, E. 2004. Collagens and collagen-related matrix components in the human and mouse eye. *Prog Retin Eye Res,* 23**,** 403-34.

ILIFF, B. W., RIAZUDDIN, S. A. & GOTTSCH, J. D. 2012a. Documenting the corneal phenotype associated with the MIR184 c.57C>T mutation. *Am J Hum Genet,* 90**,** 934; author reply 934-5.

ILIFF, B. W., RIAZUDDIN, S. A. & GOTTSCH, J. D. 2012b. A Single-Base Substitution in the Seed Region of miR-184 Causes EDICT Syndrome. *Investigative Ophthalmology & Visual Science,* 53**,** 348-353.

IQBAL, O., FISHER, G., VIRA, S., SYED, D., SADEGHI, N., FREEMAN, D., CAMPBELL, E., SUGAR, J., FEDER, R., FAREED, J. & BOUCHARD, C. 2013. Increased expression of secreted frizzled-related protein-1 and microtubule-associated protein light chain 3 in keratoconus. *Cornea,* 32**,** 702-7.

ISHII, T., MIYAZAWA, M., ONOUCHI, H., YASUDA, K., HARTMAN, P. S. & ISHII, N. 2013. Model animals for the study of oxidative stress from complex II. *Biochimica Et Biophysica Acta-Bioenergetics,* 1827**,** 588-597.

IWAMOTO, T. & DEVOE, A. G. 1976. Electron microscopical study of the Fleisher ring. *Arch Ophthalmol,* 94**,** 1579-84.

JAFRI, B., LICHTER, H. & STULTING, R. D. 2004. Asymmetric keratoconus attributed to eye rubbing. *Cornea,* 23**,** 560-4.

JEOUNG, J. W., KIM, M. K., PARK, S. S., KIM, S. Y., KO, H. S., WEE, W. R. & LEE, J. H. 2012. VSX1 Gene and Keratoconus: Genetic Analysis in Korean Patients. *Cornea,* 31**,** 746-750.

JESTER, J. V., MOLLER-PEDERSEN, T., HUANG, J., SAX, C. M., KAYS, W. T., CAVANGH, H. D., PETROLL, W. M. & PIATIGORSKY, J. 1999. The cellular basis of corneal transparency: evidence for 'corneal crystallins'. *J Cell Sci,* 112 (Pt 5)**,** 613-22.

JEYABALAN, N., SHETTY, R., GHOSH, A., ANANDULA, V. R., GHOSH, A. S. & KUMARAMANICKAVEL, G. 2013. Genetic and genomic perspective to understand the molecular pathogenesis of keratoconus. *Indian Journal of Ophthalmology,* 61**,** 384-388.

JHANJI, V., SHARMA, N. & VAJPAYEE, R. B. 2011. Management of keratoconus: current scenario. *Br J Ophthalmol,* 95**,** 1044-50.

JIANG, Y., EPSTEIN, M. P. & CONNEELY, K. N. 2013. Assessing the impact of population stratification on association studies of rare variation. *Hum Hered,* 76**,** 28-35.

JONAS, J. B., NANGIA, V., MATIN, A., KULKARNI, M. & BHOJWANI, K. 2009. Prevalence and associations of keratoconus in rural maharashtra in central India: the central India eye and medical study. *Am J Ophthalmol,* 148**,** 760-5.

JOSHI-TOPE, G., VASTRIK, I., GOPINATH, G. R., MATTHEWS, L., SCHMIDT, E., GILLESPIE, M., D'EUSTACHIO, P., JASSAL, B., LEWIS, S., WU, G., BIRNEY, E. & STEIN, L. 2003. The Genome Knowledgebase: a resource for biologists and bioinformaticists. *Cold Spring Harb Symp Quant Biol,* 68**,** 237-43.

JUN, A. S., COPE, L., SPECK, C., FENG, X., LEE, S., MENG, H., HAMAD, A. & CHAKRAVARTI, S. 2011. Subnormal cytokine profile in the tear fluid of keratoconus patients. *PLoS One,* 6**,** e16437.

JUSTICE, H. H. 1946. The human eye in anatomical transparencies. *Educ Focus,* 17**,** 26.

KALDAWY, R. M., WAGNER, J., CHING, S. & SEIGEL, G. M. 2002. Evidence of apoptotic cell death in keratoconus. *Cornea,* 21**,** 206-9.

KAMOLMATYAKUL, S., CHEN, W., YANG, S., ABE, Y., MOROI, R., ASHIQUE, A. M. & LI, Y. P. 2004. IL-1alpha stimulates cathepsin K expression in osteoclasts via the tyrosine kinase-NF-kappaB pathway. *J Dent Res,* 83**,** 791-6.

KAPUR, M., WANG, W., MALONEY, M. T., MILLAN, I., LUNDIN, V. F., TRAN, T. A. & YANG, Y. 2012. Calcium tips the balance: a microtubule plus end to lattice binding switch operates in the carboxyl terminus of BPAG1n4. *EMBO Rep,* 13**,** 1021-9.

KARAMICHOS, D., HUTCHEON, A. E., RICH, C. B., TRINKAUS-RANDALL, V., ASARA, J. M. & ZIESKE, J. D. 2014. In vitro model suggests oxidative stress involved in keratoconus disease. *Sci Rep,* 4**,** 4608.

KAROLAK, J. A., KULINSKA, K., NOWAK, D. M., PITARQUE, J. A., MOLINARI, A., RYDZANICZ, M., BEJJANI, B. A. & GAJECKA, M. 2011. Sequence variants in COL4A1 and COL4A2 genes in Ecuadorian families with keratoconus. *Molecular Vision,* 17**,** 827-843.

KAROLAK, J. A., POLAKOWSKI, P., SZAFLIK, J., SZAFLIK, J. P. & GAJECKA, M. 2015a. Molecular Screening of Keratoconus Susceptibility Sequence Variants in VSX1, TGFBI, DOCK9, STK24, and IPO5 Genes in Polish Patients and Novel TGFBI Variant Identification. *Ophthalmic Genet***,** 1-7.

KAROLAK, J. A., RYDZANICZ, M., GINTER-MATUSZEWSKA, B., PITARQUE, J. A., MOLINARI, A., BEJJANI, B. A. & GAJECKA, M. 2015b. Variant c.2262A>C in DOCK9 Leads to Exon Skipping in Keratoconus Family. *Invest Ophthalmol Vis Sci,* 56**,** 7687-90.

KATOH, Y. & KATOH, M. 2006. Comparative integromics on Angiopoietin family members. *International Journal of Molecular Medicine,* 17**,** 1145-1149.

KAUSHAL, S., JHANJI, V., SHARMA, N., TANDON, R., TITIYAL, J. S. & VAJPAYEE, R. B. 2008. "Tuck In" Lamellar Keratoplasty (TILK) for corneal ectasias involving corneal periphery. *Br J Ophthalmol,* 92**,** 286-90.

KAYA, V., KARAKAYA, M., UTINE, C. A., ALBAYRAK, S., OGE, O. F. & YILMAZ, O. F. 2007. Evaluation of the corneal topographic characteristics of keratoconus with orbscan II in patients with and without atopy. *Cornea,* 26**,** 945-8.

KAYS, W. T. & PIATIGORSKY, J. 1997. Aldehyde dehydrogenase class 3 expression: identification of a cornea-preferred gene promoter in transgenic mice. *Proc Natl Acad Sci U S A,* 94**,** 13594-9.

KEAY, L., EDWARDS, K., NADUVILATH, T., TAYLOR, H. R., SNIBSON, G. R., FORDE, K. & STAPLETON, F. 2006. Microbial keratitis predisposing factors and morbidity. *Ophthalmology,* 113**,** 109-16.

KENNA, K. P., MCLAUGHLIN, R. L., HARDIMAN, O. & BRADLEY, D. G. 2013. Using reference databases of genetic variation to evaluate the potential pathogenicity of candidate disease variants. *Hum Mutat,* 34**,** 836-41.

KENNEDY, R. H., BOURNE, W. M. & DYER, J. A. 1986. A 48-year clinical and epidemiologic study of keratoconus. *Am J Ophthalmol,* 101**,** 267-73.

KENNEY, M. C., BROWN, D. J. & RAJEEV, B. 2000. Everett Kinsey lecture. The elusive causes of keratoconus: a working hypothesis. *CLAO J,* 26**,** 10-3.

KENNEY, M. C., CHWA, M., ATILANO, S. R., TRAN, A., CARBALLO, M., SAGHIZADEH, M., VASILIOU, V., ADACHI, W. & BROWN, D. J. 2005. Increased levels of catalase and cathepsin V/L2 but decreased TIMP-1 in keratoconus corneas: Evidence that oxidative stress plays a role in this disorder. *Investigative Ophthalmology & Visual Science,* 46**,** 823-832.

KHAN, A. O., ALDAHMESH, M. A., MOHAMED, J. N. & ALKURAYA, F. S. 2010. Blue sclera with and without corneal fragility (brittle cornea syndrome) in a consanguineous family harboring ZNF469 mutation (p.E1392X). *Arch Ophthalmol,* 128**,** 1376-9.

KHAN, K., RUDKIN, A., PARRY, D. A., BURDON, K. P., MCKIBBIN, M., LOGAN, C. V., ABDELHAMED, Z. I., MUECKE, J. S., FERNANDEZ-FUENTES, N., LAURIE, K. J., SHIRES, M., FOGARTY, R., CARR, I. M., POULTER, J. A., MORGAN, J. E., MOHAMED, M. D., JAFRI, H., RAASHID, Y., MENG, N., PISETH, H., TOOMES, C., CASSON, R. J., TAYLOR, G. R., HAMMERTON, M., SHERIDAN, E., JOHNSON, C. A., INGLEHEARN, C. F., CRAIG, J. E. & ALI, M. 2011. Homozygous mutations in PXDN cause congenital cataract, corneal opacity, and developmental glaucoma. *Am J Hum Genet,* 89**,** 464-73.

KHAN, W. A., ZAHEER, N. & KHAN, S. 2015. Corneal collagen cross-linking for keratoconus: results of 3-year follow-up in Pakistani population. *Can J Ophthalmol,* 50**,** 143-50.

KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol,* 14**,** R36.

KIM, H. & JOO, C. K. 2008. Measure of keratoconus progression using Orbscan II. *J Refract Surg,* 24**,** 600-5.

KIM, J. & NETTO, M. V. 2004. Keratoconus associated with hyperimmunoglobulin E syndrome. *Cornea,* 23**,** 93-6.

KIM, S.-H., MOK, J.-W., KIM, H.-S. & JOO, C. K. 2008. Association of-31T > C and-511 C > T polymorphisms in the interleukin 1 beta (IL1B) promoter in Korean keratoconus patients. *Molecular Vision,* 14**,** 2109-2116.

KIM, W. J., RABINOWITZ, Y. S., MEISLER, D. M. & WILSON, S. E. 1999. Keratocyte apoptosis associated with keratoconus. *Exp Eye Res,* 69**,** 475-81.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet,* 46**,** 310-5.

KITAMURA, N., MORI, A., TATSUMI, H., NEMOTO, S., HIROI, T. & KAMINUMA, O. 2011. Zinc finger protein, multitype 1, suppresses human Th2 development via downregulation of IL-4. *Int Arch Allergy Immunol,* 155 Suppl 1**,** 53-6.

KLAT, M. & KHUDR, A. 1986. Religious endogamy and consanguinity in marriage patterns in Beirut, Lebanon. *Soc Biol,* 33**,** 138-45.

KNOLL, A. 1955. [Simultaneous occurrence of degenerative retinitis pigmentosa and keratoconus]. *Szemeszet,* 92**,** 165-8.

KOENIG, S. B. 2008. Bilateral recurrent self-induced keratoconus. *Eye Contact Lens,* 34**,** 343-4.

KOK, Y. O., TAN, G. F. L. & LOON, S. C. 2012. Review: Keratoconus in Asia. *Cornea,* 31**,** 581-593.

KOKOLAKIS, N. S., GAZOULI, M., CHATZIRALLI, I. P., KOUTSANDREA, C., GATZIOUFAS, Z., PEPONIS, V. G., DROUTSAS, K. D., KALOGEROPOULOS, C., ANAGNOU, N., MILTSAKAKIS, D. & MOSCHOS, M. M. 2014. Polymorphism Analysis of COL4A3 and COL4A4 Genes in Greek Patients with Keratoconus. *Ophthalmic Genetics,* 35**,** 226-228.

KOMAI, Y. & USHIKI, T. 1991. The three-dimensional organization of collagen fibrils in the human cornea and sclera. *Invest Ophthalmol Vis Sci,* 32**,** 2244-58.

KOPPEN, C., LEYSEN, I. & TASSIGNON, M. J. 2010. Riboflavin/UVA cross-linking for keratoconus in Down syndrome. *J Refract Surg,* 26**,** 623-4.

KOSTER, J., GEERTS, D., FAVRE, B., BORRADORI, L. & SONNENBERG, A. 2003. Analysis of the interactions between BP180, BP230, plectin and the integrin alpha6beta4 important for hemidesmosome assembly. *J Cell Sci,* 116**,** 387-99.

KRACHMER, J. H., FEDER, R. S. & BELIN, M. W. 1984. Keratoconus and related noninflammatory corneal thinning disorders. *Surv Ophthalmol,* 28**,** 293-322.

KRATZ, A. & CARNINCI, P. 2014. The devil in the details of RNA-seq. *Nat Biotechnol,* 32**,** 882-4.

KRISZT, A., LOSONCZY, G., BERTA, A., VEREB, G. & TAKACS, L. 2014. Segregation analysis suggests that keratoconus is a complex non-mendelian disease. *Acta Ophthalmologica,* 92**,** e562-e568.

KUMAR, P., HENIKOFF, S. & NG, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc,* 4**,** 1073-81.

KVAM, V. M., LIU P FAU - SI, Y. & SI, Y. 2012. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.,* 99.

LACKNER, E. M., MATTHAEI, M., MENG, H., ARDJOMAND, N., EBERHART, C. G. & JUN, A. S. 2014. Design and analysis of keratoconus tissue microarrays. *Cornea,* 33**,** 49-55.

LAFOND, G., BAZIN, R. & LAJOIE, C. 2001. Bilateral severe keratoconus after laser in situ keratomileusis in a patient with forme fruste keratoconus. *J Cataract Refract Surg,* 27**,** 1115-8.

LANDER, E. S. & BOTSTEIN, D. 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science,* 236**,** 1567-70.

LANDRUM, M. J., LEE, J. M., RILEY, G. R., JANG, W., RUBINSTEIN, W. S., CHURCH, D. M. & MAGLOTT, D. R. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res,* 42**,** D980-5.

LAWLESS, M., COSTER, D. J., PHILLIPS, A. J. & LOANE, M. 1989. Keratoconus: diagnosis and management. *Aust N Z J Ophthalmol,* 17**,** 33-60.

LECHNER, J., BAE, H. A., GUDURIC-FUCHS, J., RICE, A., GOVINDARAJAN, G., SIDDIQUI, S., ABI FARRAJ, L., YIP, S. P., YAP, M., DAS, M., SOUZEAU, E., COSTER, D., MILLS, R. A., LINDSAY, R., PHILLIPS, T., MITCHELL, P., ALI, M., INGLEHEARN, C. F., SUNDARESAN, P., CRAIG, J. E., SIMPSON, D. A., BURDON, K. P. & WILLOUGHBY, C. E. 2013a. Mutational analysis of MIR184 in sporadic keratoconus and myopia. *Invest Ophthalmol Vis Sci,* 54**,** 5266-72.

LECHNER, J., BAE, H. A., GUDURIC-FUCHS, J., RICE, A., GOVINDARAJAN, G., SIDDIQUI, S., FARRAJ, L. A., YIP, S. P., YAP, M., DAS, M., SOUZEAU, E., COSTER, D., MILLS, R. A., LINDSAY, R., PHILLIPS, T., MITCHELL, P., ALI, M., INGLEHEARN, C. F., SUNDARESAN, P., CRAIG, J. E., SIMPSON, D. A., BURDON, K. P. & WILLOUGHBY, C. E. 2013b. Mutational Analysis of MIR184 in Sporadic Keratoconus and Myopia. *Investigative Ophthalmology & Visual Science,* 54**,** 5266-5272.

LECHNER, J., DASH, D. P., MUSZYNSKA, D., HOSSEINI, M., SEGEV, F., GEORGE, S., FRAZER, D. G., MOORE, J. E., KAYE, S. B., YOUNG, T., SIMPSON, D. A., CHURCHILL, A. J., HEON, E. & WILLOUGHBY, C. E. 2013c. Mutational Spectrum of the ZEB1 Gene in Corneal Dystrophies

Supports a Genotype-Phenotype Correlation. *Investigative Ophthalmology & Visual Science,* 54**,** 3215-3223.

LECHNER, J., PORTER, L. F., RICE, A., VITART, V., ARMSTRONG, D. J., SCHORDERET, D. F., MUNIER, F. L., WRIGHT, A. F., INGLEHEARN, C. F., BLACK, G. C., SIMPSON, D. A., MANSON, F. & WILLOUGHBY, C. E. 2014. Enrichment of pathogenic alleles in the brittle cornea gene, ZNF469, in keratoconus. *Human Molecular Genetics,* 23**,** 5527-5535.

LEE, A. & SAKHALKAR, M. V. 2014. Ocular manifestations of Noonan syndrome in twin siblings: a case report of keratoconus with acute corneal hydrops. *Indian J Ophthalmol,* 62**,** 1171-3.

LEE, J.-E., OUM, B. S., CHOI, H. Y., LEE, S. U. & LEE, J. S. 2009. Evaluation of differentially expressed genes identified in keratoconus. *Molecular Vision,* 15**,** 2480-2487.

LEE, L. R., HIRST, L. W. & READSHAW, G. 1995. Clinical detection of unilateral keratoconus. *Aust N Z J Ophthalmol,* 23**,** 129-33.

LEE, S., ABECASIS, G. R., BOEHNKE, M. & LIN, X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet,* 95**,** 5-23.

LEEDMAN, P. J., FAULKNER-JONES, B., CRAM, D. S., HARRISON, P. J., WEST, J., O'BRIEN, E., SIMPSON, R., COPPEL, R. L. & HARRISON, L. C. 1993. Cloning from the thyroid of a protein related to actin binding protein that is recognized by Graves disease immunoglobulins. *Proc Natl Acad Sci U S A,* 90**,** 5994-8.

LEEK, J. T. 2014. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Research*.

LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. & IRIZARRY, R. A. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet,* 11**,** 733-739.

LEGEAIS, J. M., PARC, C., D'HERMIES, F., POULIQUEN, Y. & RENARD, G. 2001. Nineteen years of penetrating keratoplasty in the Hotel-Dieu Hospital in Paris. *Cornea,* 20**,** 603-6.

LEMA, I., BREA, D., RODRIGUEZ-GONZALEZ, R., DIEZ-FEIJOO, E. & SOBRINO, T. 2010. Proteomic analysis of the tear film in patients with keratoconus. *Mol Vis,* 16**,** 2055-61.

LEMA, I. & DURAN, J. A. 2005. Inflammatory molecules in the tears of patients with keratoconus. *Ophthalmology,* 112**,** 654-9.

LEMA, I., SOBRINO, T., DURAN, J. A., BREA, D. & DIEZ-FEIJOO, E. 2009. Subclinical keratoconus and inflammatory molecules from tears. *Br J Ophthalmol,* 93**,** 820-4.

LETTRE, G. 2014. Rare and low-frequency variants in human common diseases and other complex traits. *J Med Genet,* 51**,** 705-14.

LEVY, D., HUTCHINGS, H., ROULAND, J. F., GUELL, J., BURILLON, C., ARNE, J. L., COLIN, J., LAROCHE, L., MONTARD, M., DELBOSC, B., APTEL, I., GINISTY, H., GRANDJEAN, H. & MALECAZE, F. 2004. Videokeratographic anomalies in familial keratoconus. *Ophthalmology,* 111**,** 867-874.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, H., HUANG, S., WANG, S., WANG, L., QI, L., ZHANG, Y., ZHANG, S., ZHAO, B. & MIAO, J. 2013a. Relationship between annexin A7 and integrin beta4 in autophagy. *Int J Biochem Cell Biol,* 45**,** 2605-11.

LI, J., WITTEN, D. M., JOHNSTONE, I. M. & TIBSHIRANI, R. 2012a. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics,* 13**,** 523-38.

LI, N., FAN, Z., PENG, X., PANG, X. & TIAN, C. 2014. Clinical observation of transepithelial corneal collagen cross-linking by lontophoresis of riboflavin in treatment of keratoconus. *Eye Sci,* 29**,** 160-4.

LI, X., BYKHOVSKAYA, Y., HARITUNIANS, T., SISCOVICK, D., ALDAVE, A., SZCZOTKA-FLYNN, L., IYENGAR, S. K., ROTTER, J. I., TAYLOR, K. D. & RABINOWITZ, Y. S. 2012b. A genome-wide association study identifies a potential novel gene locus for keratoconus, one of the commonest causes for corneal transplantation in developed countries. *Human Molecular Genetics,* 21**,** 421-429.

LI, X., BYKHOVSKAYA, Y., TANG, Y. G., PICORNELL, Y., HARITUNIANS, T., ALDAVE, A. J., SZCZOTKA-FLYNN, L., IYENGAR, S. K., ROTTER, J. I., TAYLOR, K. D. & RABINOWITZ, Y. S. 2013b. An Association Between the Calpastatin (CAST) Gene and Keratoconus. *Cornea,* 32**,** 696-701.

LI, X., RABINOWITZ, Y. S., RASHEED, K. & YANG, H. 2004. Longitudinal study of the normal eyes in unilateral keratoconus patients. *Ophthalmology,* 111**,** 440-6.

LI, X., RABINOWITZ, Y. S., TANG, Y. G., PICORNELL, Y., TAYLOR, K. D., HU, M. & YANG, H. 2006. Two-stage genome-wide linkage scan in keratoconus sib pair families. *Investigative Ophthalmology & Visual Science,* 47**,** 3791-3795.

LIANG, Y., CHANG, C. & LU, Q. 2015. The Genetics and Epigenetics of Atopic Dermatitis-Filaggrin and Other Polymorphisms. *Clin Rev Allergy Immunol*.

LIAO, Y., SMYTH, G. K. & SHI, W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res,* 41**,** e108.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LISKOVA, P., EBENEZER, N. D., HYSI, P. G., GWILLIAM, R., EL-ASHRY, M. F., MOODALEY, L. C., HAU, S., TWA, M., TUFT, S. J. & BHATACHARYA, S. S. 2007a. Molecular analysis of the VSX1 gene in familial keratoconus. *Molecular Vision,* 13**,** 1887-1891.

LISKOVA, P., FILIPEC, M., MERJAVA, S., JIRSOVA, K. & TUFT, S. J. 2010a. Variable ocular phenotypes of posterior polymorphous corneal dystrophy caused by mutations in the ZEB1 gene. *Ophthalmic Genetics,* 31**,** 230-234.

LISKOVA, P., HYSI, P. G., WASEEM, N., EBENEZER, N. D., BHATTACHARYA, S. S. & TUFT, S. J. 2010b. Evidence for Keratoconus Susceptibility Locus on Chromosome 14 A Genome-wide Linkage Screen Using Single-Nucleotide Polymorphism Markers. *Archives of Ophthalmology,* 128**,** 1191-1195.

LISKOVA, P., TUFT, S. J., GWILLIAM, R., EBENEZER, N. D., JIRSOVA, K., PRESCOTT, Q., MARTINCOVA, R., PRETORIUS, M., SINCLAIR, N., BOASE, D. L., JEFFREY, M. J., DELOUKAS, P., HARDCASTLE, A. J., FILIPEC, M. & BHATTACHARYA, S. S. 2007b. Novel mutations in the

ZEB1 gene identified in Czech and British patients with posterior polymorphous corneal dystrophy. *Hum Mutat,* 28**,** 638.

LIU, M. M., CHAN, C. C. & TUO, J. 2012. Genetic mechanisms and age-related macular degeneration: common variants, rare variants, copy number variations, epigenetics, and mitochondrial genetics. *Hum Genomics,* 6**,** 13.

LIU, X., JIAN, X. & BOERWINKLE, E. 2011. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat,* 32**,** 894-9.

LIU, X., JIAN, X. & BOERWINKLE, E. 2013. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat,* 34**,** E2393-402.

LIU, X., WU, C., LI, C. & BOERWINKLE, E. 2015a. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. *Hum Mutat*.

LIU, Y. & ALLINGHAM, R. R. 2011. Molecular genetics in glaucoma. *Exp Eye Res,* 93**,** 331-9.

LIU, Y., SIDHU, A., BEAN, L. H., CONWAY, R. L. & FRIDOVICH-KEIL, J. L. 2015b. Genetic and functional studies reveal a novel noncoding variant in GALT associated with a false positive newborn screening result for galactosemia. *Clin Chim Acta,* 446**,** 171-4.

LOGAN, C. V., SZABADKAI, G., SHARPE, J. A., PARRY, D. A., TORELLI, S., CHILDS, A. M., KRIEK, M., PHADKE, R., JOHNSON, C. A., ROBERTS, N. Y., BONTHRON, D. T., PYSDEN, K. A., WHYTE, T., MUNTEANU, I., FOLEY, A. R., WHEWAY, G., SZYMANSKA, K., NATARAJAN, S., ABDELHAMED, Z. A., MORGAN, J. E., ROPER, H., SANTEN, G. W., NIKS, E. H., VAN DER POL, W. L., LINDHOUT, D., RAFFAELLO, A., DE STEFANI, D., DEN DUNNEN, J. T., SUN, Y., GINJAAR, I., SEWRY, C. A., HURLES, M., RIZZUTO, R., DUCHEN, M. R., MUNTONI, F. & SHERIDAN, E. 2014. Loss-of-function mutations in MICU1 cause a brain and muscle disorder linked to primary alterations in mitochondrial calcium signaling. *Nat Genet,* 46**,** 188-93.

LOH, R. S., CHAN, C. M., TI, S. E., LIM, L., CHAN, K. S. & TAN, D. T. 2009. Emerging prevalence of microsporidial keratitis in Singapore: epidemiology, clinical features, and management. *Ophthalmology,* 116**,** 2348-53.

LOMELIN, D., JORGENSON, E. & RISCH, N. 2010. Human genetic variation recognizes functional elements in noncoding sequence. *Genome Res,* 20**,** 311-9.

LOWELL, F. C. & CARROLL, J. M. 1970. A study of the occurrence of atopic traits in patients with keratoconus. *J Allergy,* 46**,** 32-9.

LU, Y., VITART, V., BURDON, K. P., KHOR, C. C., BYKHOVSKAYA, Y., MIRSHAHI, A., HEWITT, A. W., KOEHN, D., HYSI, P. G., RAMDAS, W. D., ZELLER, T., VITHANA, E. N., CORNES, B. K., TAY, W.-T., TAI, E. S., CHENG, C.-Y., LIU, J., FOO, J.-N., SAW, S. M., THORLEIFSSON, G., STEFANSSON, K., DIMASI, D. P., MILLS, R. A., MOUNTAIN, J., ANG, W., HOEHN, R., VERHOEVEN, V. J. M., GRUS, F., WOLFS, R., CASTAGNE, R., LACKNER, K. J., SPRINGELKAMP, H., YANG, J., JONASSON, F., LEUNG, D. Y. L., CHEN, L. J., THAM, C. C. Y., RUDAN, I., VATAVUK, Z., HAYWARD, C., GIBSON, J., CREE, A. J., MACLEOD, A., ENNIS, S., POLASEK, O., CAMPBELL, H., WILSON, J. F., VISWANATHAN, A. C., FLECK, B., LI, X., SISCOVICK, D., TAYLOR, K. D., ROTTER, J. I., YAZAR, S., ULMER, M., LI, J.,

YASPAN, B. L., OZEL, A. B., RICHARDS, J. E., MOROI, S. E., HAINES, J. L., KANG, J. H., PASQUALE, L. R., ALLINGHAM, R. R., ASHLEY-KOCH, A., MITCHELL, P., WANG, J. J., WRIGHT, A. F., PENNELL, C., SPECTOR, T. D., YOUNG, T. L., KLAVER, C. C. W., MARTIN, N. G., MONTGOMERY, G. W., ANDERSON, M. G., AUNG, T., WILLOUGHBY, C. E., WIGGS, J. L., PANG, C. P., THORSTEINSDOTTIR, U., LOTERY, A. J., HAMMOND, C. J., VAN DUIJN, C. M., HAUSER, M. A., RABINOWITZ, Y. S., PFEIFFER, N., MACKEY, D. A., CRAIG, J. E., MACGREGOR, S., WONG, T. Y. & CONSORTIUM, N. 2013. Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus. *Nature Genetics,* 45**,** 155-163.

MACARTHUR, D. G., MANOLIO, T. A., DIMMOCK, D. P., REHM, H. L., SHENDURE, J., ABECASIS, G. R., ADAMS, D. R., ALTMAN, R. B., ANTONARAKIS, S. E., ASHLEY, E. A., BARRETT, J. C., BIESECKER, L. G., CONRAD, D. F., COOPER, G. M., COX, N. J., DALY, M. J., GERSTEIN, M. B., GOLDSTEIN, D. B., HIRSCHHORN, J. N., LEAL, S. M., PENNACCHIO, L. A., STAMATOYANNOPOULOS, J. A., SUNYAEV, S. R., VALLE, D., VOIGHT, B. F., WINCKLER, W. & GUNTER, C. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature,* 508**,** 469-76.

MACE, M., GALIACY, S. D., ERRAUD, A., MEJIA, J. E., ETCHEVERS, H., ALLOUCHE, M., DESJARDINS, L., CALVAS, P. & MALECAZE, F. 2011. Comparative Transcriptome and Network Biology Analyses Demonstrate Antiproliferative and Hyperapoptotic Phenotypes in Human Keratoconus Corneas. *Investigative Ophthalmology & Visual Science,* 52**,** 6181-6191.

MACKIEWICZ, Z., MAATTA, M., STENMAN, M., KONTTINEN, L., TERVO, T. & KONTTINEN, Y. T. 2006. Collagenolytic proteinases in keratoconus. *Cornea,* 25**,** 603-10.

MACSAI, M., MAGUEN, E. & NUCCI, P. 1997. Keratoconus and Turner's syndrome. *Cornea,* 16**,** 534-6.

MAEDA, N., KLYCE, S. D., SMOLEK, M. K. & THOMPSON, H. W. 1994. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci,* 35**,** 2749-57.

MAGUIRE, L. J. & BOURNE, W. M. 1989. Corneal topography of early keratoconus. *Am J Ophthalmol,* 108**,** 107-12.

MAKRYTHANASIS, P. & ANTONARAKIS, S. E. 2013. Pathogenic variants in non-protein-coding sequences. *Clin Genet,* 84**,** 422-8.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R., CHAKRAVARTI, A., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G., HAINES, J. L., MACKAY, T. F., MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MARQUIS GACY, A., GOELLNER, G., JURANIĆ, N., MACURA, S. & MCMURRAY, C. T. 1995. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell,* 81**,** 533-540.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011,* 17.

MARUYAMA, Y., WANG, X., LI, Y., SUGAR, J. & YUE, B. 2001. Involvement of Sp1 elements in the promoter activity of genes affected in keratoconus. *Investigative Ophthalmology & Visual Science,* 42**,** S587-S587.

MATHEW, J. H., BERGMANSON, J. P. & DOUGHTY, M. J. 2008. Fine structure of the interface between the anterior limiting lamina and the anterior stromal fibrils of the human cornea. *Invest Ophthalmol Vis Sci,* 49**,** 3914-8.

MATTHEWS, F. J., COOK, S. D., MAJID, M. A., DICK, A. D. & SMITH, V. A. 2007. Changes in the balance of the tissue inhibitor of matrix metalloproteinases (TIMPs)-1 and -3 may promote keratocyte apoptosis in keratoconus. *Exp Eye Res,* 84**,** 1125-34.

MAURICE, D. M. 1957. The structure and transparency of the cornea. *J Physiol,* 136**,** 263-86.

MAYCOCK, N. J. & MARSHALL, J. 2014. Genomics of corneal wound healing: a review of the literature. *Acta Ophthalmol,* 92**,** e170-84.

MAZZOTTA, C., TRAVERSI, C., RAISKUP, F., RIZZO, C. L. & RENIERI, A. 2014. First identification of a triple corneal dystrophy association: keratoconus, epithelial basement membrane corneal dystrophy and fuchs' endothelial corneal dystrophy. *Case reports in ophthalmology,* 5**,** 281-8.

MCCLELLAN, J. & KING, M. C. 2010. Genetic heterogeneity in human disease. *Cell,* 141**,** 210-7.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.

MCKINLEY-GRANT, L. J., IDLER, W. W., BERNSTEIN, I. A., PARRY, D. A., CANNIZZARO, L., CROCE, C. M., HUEBNER, K., LESSIN, S. R. & STEINERT, P. M. 1989. Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21. *Proc Natl Acad Sci U S A,* 86**,** 4848-52.

MCMAHON, T. T., KIM, L. S., FISHMAN, G. A., STONE, E. M., ZHAO, X. C., YEE, R. W. & MALICKI, J. 2009. CRB1 Gene Mutations Are Associated with Keratoconus in Patients with Leber Congenital Amaurosis. *Investigative Ophthalmology & Visual Science,* 50**,** 3185-3187.

MCMONNIES, C. W. 2007. Abnormal rubbing and keratectasia. *Eye Contact Lens,* 33**,** 265-71.

MCMONNIES, C. W. 2014. Epigenetic Mechanisms Might Help Explain Environmental Contributions to the Pathogenesis of Keratoconus. *Eye & Contact Lens-Science and Clinical Practice,* 40**,** 371-375.

MCMONNIES, C. W. 2015. Inflammation and keratoconus. *Optom Vis Sci,* 92**,** e35-41.

MEEK, K. M., TUFT, S. J., HUANG, Y., GILL, P. S., HAYES, S., NEWTON, R. H. & BRON, A. J. 2005. Changes in collagen orientation and distribution in keratoconus corneas. *Invest Ophthalmol Vis Sci,* 46**,** 1948-56.

MEHAFFEY, M. G., NEWTON, A. L., GANDHI, M. J., CROSSLEY, M. & DRACHMAN, J. G. 2001. X-linked thrombocytopenia caused by a novel mutation of GATA-1. *Blood,* 98**,** 2681-8.

MEIJERS, M. H. M., AISA, C. M., BILLINGHAM, M. E. J., RUSSELL, R. G. G. & BUNNING, R. A. D. 1994. The effect of interleukin-1β and transforming

growth factor β on cathepsin B activity in human articular chondrocytes. *Agents and Actions,* 41**,** C198-C200.

MELLES, G. R., LANDER, F., RIETVELD, F. J., REMEIJER, L., BEEKHUIS, W. H. & BINDER, P. S. 1999. A new surgical technique for deep stromal, anterior lamellar keratoplasty. *Br J Ophthalmol,* 83**,** 327-33.

MELLES, G. R., REMEIJER, L., GEERARDS, A. J. & BEEKHUIS, W. H. 2000. A quick surgical technique for deep, anterior lamellar keratoplasty using visco-dissection. *Cornea,* 19**,** 427-32.

MIKAMI, T., MEGURO, A., TESHIGAWARA, T., TAKEUCHI, M., UEMOTO, R., KAWAGOE, T., NOMURA, E., ASUKATA, Y., ISHIOKA, M., IWASAKI, M., FUKAGAWA, K., KONOMI, K., SHIMAZAKI, J., NISHIDA, T. & MIZUKI, N. 2013. Interleukin 1 beta promoter polymorphism is associated with keratoconus in a Japanese population. *Molecular Vision,* 19**,** 845-851.

MILLIN, J. A., GOLUB, B. M. & FOSTER, C. S. 1986. Human basement membrane components of keratoconus and normal corneas. *Invest Ophthalmol Vis Sci,* 27**,** 604-7.

MILLODOT, M., SHNEOR, E., ALBOU, S., ATLANI, E. & GORDON-SHAAG, A. 2011. Prevalence and associated factors of keratoconus in Jerusalem: a cross-sectional study. *Ophthalmic Epidemiol,* 18**,** 91-7.

MINET, A. D., RUBIN, B. P., TUCKER, R. P., BAUMGARTNER, S. & CHIQUET-EHRISMANN, R. 1999. Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci,* 112 (Pt 12)**,** 2019-32.

MITAS, M. 1997. Trinucleotide repeats associated with human disease. *Nucleic Acids Research,* 25**,** 2245-2253.

MOK, J.-W., BAEK, S.-J. & JOO, C.-K. 2008. VSX1 gene variants are associated with keratoconus in unrelated Korean patients. *Journal of Human Genetics,* 53**,** 842-849.

MOOTHA, V. V., HUSSAIN, I., CUNNUSAMY, K., GRAHAM, E., GONG, X., NEELAM, S., XING, C., KITTLER, R. & PETROLL, W. M. 2015. TCF4 Triplet Repeat Expansion and Nuclear RNA Foci in Fuchs' Endothelial Corneal Dystrophy. *Invest Ophthalmol Vis Sci,* 56**,** 2003-11.

MOOTHA, V. V., KANOFF, J. M., SHANKARDAS, J. & DIMITRIJEVICH, S. 2009. Marked reduction of alcohol dehydrogenase in keratoconus corneal fibroblasts. *Molecular Vision,* 15**,** 706-712.

MORISHIGE, N., WAHLERT, A. J., KENNEY, M. C., BROWN, D. J., KAWAMOTO, K., CHIKAMA, T., NISHIDA, T. & JESTER, J. V. 2007. Second-harmonic imaging microscopy of normal human and keratoconus cornea. *Invest Ophthalmol Vis Sci,* 48**,** 1087-94.

MORLEY, M., MOLONY, C. M., WEBER, T. M., DEVLIN, J. L., EWENS, K. G., SPIELMAN, R. S. & CHEUNG, V. G. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature,* 430**,** 743-7.

MORRISON, D. A., ROSSER, E. M. & CLAOUE, C. 2001. Keratoconus associated with a chromosome 7,11 translocation. *Eye (Lond),* 15**,** 556-7.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MOSCHOS, M. M., KOKOLAKIS, N., GAZOULI, M., CHATZIRALLI, I. P., DROUTSAS, D., ANAGNOU, N. P. & LADAS, I. D. 2015. Polymorphism

Analysis of VSX1 and SOD1 Genes in Greek Patients with Keratoconus. *Ophthalmic Genet,* 36**,** 213-7.

MOUTSIANAS, L., AGARWALA, V., FUCHSBERGER, C., FLANNICK, J., RIVAS, M. A., GAULTON, K. J., ALBERS, P. K., MCVEAN, G., BOEHNKE, M., ALTSHULER, D. & MCCARTHY, M. I. 2015. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet,* 11**,** e1005165.

MOYNIHAN, L. M., BUNDEY, S. E., HEATH, D., JONES, E. L., MCHALE, D. P., MUELLER, R. F., MARKHAM, A. F. & LENCH, N. J. 1998. Autozygosity mapping, to chromosome 11q25, of a rare autosomal recessive syndrome causing histiocytosis, joint contractures, and sensorineural deafness. *Am J Hum Genet,* 62**,** 1123-8.

MUELLER, R. F. & BISHOP, D. T. 1993. Autozygosity mapping, complex consanguinity, and autosomal recessive disorders. *J Med Genet,* 30**,** 798-9.

MULLER, L. J., PELS, E. & VRENSEN, G. F. 2001. The effects of organ-culture on the density of keratocytes and collagen fibers in human corneas. *Cornea,* 20**,** 86-95.

MUSZYNSKA, D., LECHNER, J., DASH, D. P., HEON, E., HUGHES, A. E. & WILLOUGHBY, C. E. 2011. Identification and Characterisation of a Novel Missense Homeodomain Mutation in ZEB1 Resulting in Keratoconus. *ARVO Annual Meeting Abstract Search and Program Planner,* 2011**,** 1077-1077.

NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science,* 320**,** 1344-9.

NAKOUZI, G., KREIDIEH, K. & YAZBEK, S. 2015. A review of the diverse genetic disorders in the Lebanese population: highlighting the urgency for community genetic services. *J Community Genet,* 6**,** 83-105.

NEMET, A. Y., VINKER, S., BAHAR, I. & KAISERMAN, I. 2010. The association of keratoconus with immune disorders. *Cornea,* 29**,** 1261-4.

NEWKIRK, K. M., CHANDLER, H. L., PARENT, A. E., YOUNG, D. C., COLITZ, C. M., WILKIE, D. A. & KUSEWITT, D. F. 2007. Ultraviolet radiation-induced corneal degeneration in 129 mice. *Toxicol Pathol,* 35**,** 819-26.

NEWSOME, D. A., GROSS, J. & HASSELL, J. R. 1982. Human corneal stroma contains three distinct collagens. *Invest Ophthalmol Vis Sci,* 22**,** 376-81.

NG, S. B., BUCKINGHAM, K. J., LEE, C., BIGHAM, A. W., TABOR, H. K., DENT, K. M., HUFF, C. D., SHANNON, P. T., JABS, E. W., NICKERSON, D. A., SHENDURE, J. & BAMSHAD, M. J. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet,* 42**,** 30-5.

NI, X., EPSHTEIN, Y., CHEN, W., ZHOU, T., XIE, L., GARCIA, J. G. & JACOBSON, J. R. 2014. Interaction of integrin beta4 with S1P receptors in S1P- and HGF-induced endothelial barrier enhancement. *J Cell Biochem,* 115**,** 1187-95.

NIEDERER, R. L., PERUMAL, D., SHERWIN, T. & MCGHEE, C. N. 2008. Laser scanning in vivo confocal microscopy reveals reduced innervation and reduction in cell density in all layers of the keratoconic cornea. *Invest Ophthalmol Vis Sci,* 49**,** 2964-70.

NIELSEN, K., BIRKENKAMP-DEMTRODER, K., EHLERS, N. & ORNTOFT, T. F. 2003. Identification of differentially expressed genes in keratoconus epithelium analyzed on microarrays. *Investigative Ophthalmology & Visual Science,* 44**,** 2466-2476.

NIELSEN, K., HJORTDAL, J., PIHLMANN, M. & CORYDON, T. J. 2013. Update on the keratoconus genetics. *Acta Ophthalmologica,* 91**,** 106-113.

NOBLE, B. A., AGRAWAL, A., COLLINS, C., SALDANA, M., BROGDEN, P. R. & ZUBERBUHLER, B. 2007. Deep Anterior Lamellar Keratoplasty (DALK): visual outcome and complications for a heterogeneous group of corneal pathologies. *Cornea,* 26**,** 59-64.

NOWAK, D. M. & GAJECKA, M. 2011. The genetics of keratoconus. *Middle East African journal of ophthalmology,* 18**,** 2-6.

NOWAK, D. M. & GAJECKA, M. 2015. Nonrandom Distribution of miRNAs Genes and Single Nucleotide Variants in Keratoconus Loci. *PLoS One,* 10**,** e0132143.

NOWAK, D. M., KAROLAK, J. A., KUBIAK, J., GUT, M., PITARQUE, J. A., MOLINARI, A., BEJJANI, B. A. & GAJECKA, M. 2013. Substitution at IL1RN and Deletion at SLC4A11 Segregating with Phenotype in Familial Keratoconus. *Investigative Ophthalmology & Visual Science,* 54**,** 2207-2215.

NUCCI, P. & BRANCATO, R. 1991. Keratoconus and congenital hip dysplasia. *Am J Ophthalmol,* 111**,** 775-6.

OGATA, H., GOTO, S., FUJIBUCHI, W. & KANEHISA, M. 1998. Computation with the KEGG pathway database. *Biosystems,* 47**,** 119-28.

OKUMURA, M., YAMAKAWA, H., OHARA, O. & OWARIBE, K. 2002. Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1) including the domain structure closely related to MACF (microtubule actin cross-linking factor). *J Biol Chem,* 277**,** 6682-7.

OLIVEIRA-SOTO, L. & EFRON, N. 2001. Morphology of corneal nerves using confocal microscopy. *Cornea,* 20**,** 374-84.

ONG, H. S. & CORBETT, M. C. 2015. Corneal infections in the 21st century. *Postgrad Med J,* 91**,** 565-71.

OPBROEK, A., KENNEY, M. C. & BROWN, D. 1993. Characterization of a human corneal metalloproteinase inhibitor (TIMP-1). *Curr Eye Res,* 12**,** 877-83.

OPHTHALMOLOGY, A. A. O. 2013. Corneal Ectasia. *Preferred Practice Pattern Guidelines*.

ORTAK, H., SOGUT, E., TAS, U., MESCI, C. & MENDIL, D. 2012. The relation between keratoconus and plasma levels of MMP-2, zinc, and SOD. *Cornea,* 31**,** 1048-51.

OWENS, H. & GAMBLE, G. 2003. A profile of keratoconus in New Zealand. *Cornea,* 22**,** 122-5.

OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet,* 12**,** 87-98.

PALAMAR, M., ONAY, H., OZDEMIR, T. R., ARSLAN, E., EGRILMEZ, S., OZKINAY, F. & YAGCI, A. 2014. Relationship Between IL1 beta-511C > T and ILRN VNTR Polymorphisms and Keratoconus. *Cornea,* 33**,** 145-147.

PALIWAL, P., SINGH, A., TANDON, R., TITIYAL, J. S. & SHARMA, A. 2009. A novel VSX1 mutation identified in an individual with keratoconus in India. *Molecular Vision,* 15**,** 2475-2479.

PALIWAL, P., TANDON, R., DUBE, D., KAUR, P. & SHARMA, A. 2011. Familial segregation of a VSX1 mutation adds a new dimension to its role in the causation of keratoconus. *Molecular Vision,* 17**,** 481-485.

PANNEBAKER, C., CHANDLER, H. L. & NICHOLS, J. J. 2010. Tear proteomics in keratoconus. *Mol Vis,* 16**,** 1949-57.

PARKER, J. S., VAN DIJK, K. & MELLES, G. R. 2015. Treatment options for advanced keratoconus: A review. *Surv Ophthalmol,* 60**,** 459-80.

PARTHASARATHY, A., POR, Y. M. & TAN, D. T. 2008. Using a "small bubble technique" to aid in success in Anwar's "big bubble technique" of deep lamellar keratoplasty with complete baring of Descemet's membrane. *Br J Ophthalmol,* 92**,** 422.

PATEL, D. & MCGHEE, C. 2013. Understanding keratoconus: what have we learned from the New Zealand perspective? *Clin Exp Optom,* 96**,** 183-7.

PATEL, H. Y., ORMONDE, S., BROOKES, N. H., MOFFATT, L. S. & MCGHEE, C. N. 2005. The indications and outcome of paediatric corneal transplantation in New Zealand: 1991-2003. *Br J Ophthalmol,* 89**,** 404-8.

PATEL, S. & MCLAUGHLIN, J. M. 1999. Effects of central corneal thickness on measurement of intra-ocular pressure in keratoconus and post-keratoplasty. *Ophthalmic Physiol Opt,* 19**,** 236-41.

PATHAK, D., NAYAK, B., SINGH, M., SHARMA, N., TITIYAL, J. S. & DADA, R. 2011. Mitochondrial complex 1 gene analysis in keratoconus. *Molecular Vision,* 17**,** 1514-1525.

PEARSON, A. R., SONEJI, B., SARVANANTHAN, N. & SANDFORD-SMITH, J. H. 2000. Does ethnic origin influence the incidence or severity of keratoconus? *Eye (Lond),* 14 (Pt 4)**,** 625-8.

PELTONEN, L., PEROLA, M., NAUKKARINEN, J. & PALOTIE, A. 2006. Lessons from studying monogenic disease for common disease. *Hum Mol Genet,* 15 Spec No 1**,** R67-74.

PERLMAN, J. M. & ZAIDMAN, G. W. 1994. Bilateral keratoconus in Crouzon's syndrome. *Cornea,* 13**,** 80-1.

PICO, A. R., KELDER, T., VAN IERSEL, M. P., HANSPERS, K., CONKLIN, B. R. & EVELO, C. 2008. WikiPathways: pathway editing for the people. *PLoS Biol,* 6**,** e184.

PINNA, A., SALVO, M., DORE, S. & CARTA, F. 2005. Corneal graft rejection after penetrating keratoplasty for keratoconus in Turner's syndrome. *Eur J Ophthalmol,* 15**,** 271-3.

PODSKOCHY, A., GAN, L. & FAGERHOLM, P. 2000. Apoptosis in UV-exposed rabbit corneas. *Cornea,* 19**,** 99-103.

POULADI, N., BIME, C., GARCIA, J. G. & LUSSIER, Y. A. 2015. Complex genetics of pulmonary diseases: lessons from genome-wide association studies and next-generation sequencing. *Transl Res*.

POYALES-GALAN, F., FERNANDEZ-AITOR-GARCIA, A., GARZON-JIMENEZ, N., ORTIZ-DE-ZARATE, B. & ELIPE-GOSALVEZ, V. 2009. [Management of Descemet's membrane rupture by intracameral injection of SF6 in acute hydrops]. *Arch Soc Esp Oftalmol,* 84**,** 533-6.

PREDOVIC, J., BALOG, T., MAROTTI, T., GABRIC, N., BOHAC, M., ROMAC, I. & DEKARIS, I. 2008. The expression of human corneal MMP-2, MMP-9, proMMP-13 and TIMP-1 in bullous keratopathy and keratoconus. *Coll Antropol,* 32 Suppl 2**,** 15-9.

PRONIN, A., LEVAY, K., VELMESHEV, D., FAGHIHI, M., SHESTOPALOV, V. I. & SLEPAK, V. Z. 2014. Expression of olfactory signaling genes in the eye. *PLoS One,* 9**,** e96435.

RABINOWITZ, Y. S. 1998. Keratoconus. *Survey of Ophthalmology,* 42**,** 297-319.

RABINOWITZ, Y. S. 2003. The genetics of keratoconus. *Ophthalmol Clin North Am,* 16**,** 607-20, vii.

RABINOWITZ, Y. S., LI, X., CANEDO, A. L., AMBROSIO, R., JR. & BYKHOVSKAYA, Y. 2014. Optical coherence tomography combined with videokeratography to differentiate mild keratoconus subtypes. *J Refract Surg,* 30**,** 80-7.

RABINOWITZ, Y. S., MAUMENEE, I. H., LUNDERGAN, M. K., PUFFENBERGER, E., ZHU, D., ANTONARAKIS, S. & FRANCOMANO, C. A. 1992. Molecular genetic analysis in autosomal dominant keratoconus. *Cornea,* 11**,** 302-8.

RADNER, W., ZEHETMAYER, M., AUFREITER, R. & MALLINGER, R. 1998. Interlacing and cross-angle distribution of collagen lamellae in the human cornea. *Cornea,* 17**,** 537-43.

RAHI, A., DAVIES, P., RUBEN, M., LOBASCHER, D. & MENON, J. 1977. Keratoconus and coexisting atopic disease. *Br J Ophthalmol,* 61**,** 761-4.

RAO, N. A., FERNANDEZ, M. A., CID, L. L., ROMERO, J. L. & SEVANIAN, A. 1987. Retinal lipid peroxidation in experimental uveitis. *Arch Ophthalmol,* 105**,** 1712-6.

RAO, V. A., SWATHI, P. & THAPPA, D. M. 2008. Bilateral keratoconus with oculocutaneous albinism. *Indian J Dermatol Venereol Leprol,* 74**,** 407-9.

RATHI, V. M., MANDATHARA, P. S. & DUMPATI, S. 2013. Contact lens in keratoconus. *Indian J Ophthalmol,* 61**,** 410-5.

RAU, S. & DUNCKER, G. I. 1994. [Keratoconus in Mulvihill-Smith syndrome]. *Klin Monbl Augenheilkd,* 205**,** 44-6.

REGISTRY, T. A. C. G. 1993. The Australian Corneal Graft Registry. 1990 to 1992 report. *Aust N Z J Ophthalmol,* 21**,** 1-48.

REHMAN, S., BAIG, S. M., EIBERG, H., REHMAN, S., AHMAD, I., MALIK, N. A., TOMMERUP, N. & HANSEN, L. 2011. Autozygosity mapping of a large consanguineous Pakistani family reveals a novel non-syndromic autosomal recessive mental retardation locus on 11p15-tel. *Neurogenetics,* 12**,** 247-51.

REIK, W., DAVIES, K., DEAN, W., KELSEY, G. & CONSTANCIA, M. 2001. Imprinted genes and the coordination of fetal and postnatal growth in mammals. *Novartis Found Symp,* 237**,** 19-31; discussion 31-42.

REVA, B., ANTIPIN, Y. & SANDER, C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res,* 39**,** e118.

RHO, C. R., PARK, J. H., JUNG, Y. H. & KIM, M. S. 2014. A case of concomitant keratoconus and granular corneal dystrophy type II. *Cont Lens Anterior Eye,* 37**,** 314-6.

RINGVOLD, A. 1998. Corneal epithelium and UV-protection of the eye. *Acta Ophthalmol Scand,* 76**,** 149-53.

RISCH, N. & MERIKANGAS, K. 1996. The future of genetic studies of complex human diseases. *Science,* 273**,** 1516-7.

RIZZUTI, A. B. 1970. Diagnostic illumination test for keratoconus. *Am J Ophthalmol,* 70**,** 141-3.

ROBINSON, G. W., TSAY, Y. H., KIENZLE, B. K., SMITH-MONROY, C. A. & BISHOP, R. W. 1993. Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation. *Mol Cell Biol,* 13**,** 2706-17.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotech,* 29**,** 24-26.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.

ROBINSON, M. D. & OSHLACK, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol,* 11**,** R25.

ROBLES, J. A., QURESHI, S. E., STEPHEN, S. J., WILSON, S. R., BURDEN, C. J. & TAYLOR, J. M. 2012. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics,* 13**,** 484.

RODRIGUEZ, P., BONTE, E., KRIJGSVELD, J., KOLODZIEJ, K. E., GUYOT, B., HECK, A. J., VYAS, P., DE BOER, E., GROSVELD, F. & STROUBOULIS, J. 2005. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo j,* 24**,** 2354-66.

ROESSLER, E., HU, P., HONG, S. K., SRIVASTAVA, K., CARRINGTON, B., SOOD, R., PETRYKOWSKA, H., ELNITSKI, L., RIBEIRO, L. A., RICHIERI-COSTA, A., FELDMAN, B., ODENWALD, W. F. & MUENKE, M. 2012. Unique alterations of an ultraconserved non-coding element in the 3'UTR of ZIC2 in holoprosencephaly. *PLoS One,* 7**,** e39026.

ROGERS, G. J. & ATTENBOROUGH, M. 2014. Bilateral superior keratoconus: two case reports. *Eye (Lond),* 28**,** 1254-7.

ROHINI, G., MURUGESWARI, P., PRAJNA, N. V., LALITHA, P. & MUTHUKKARUPPAN, V. 2007. Matrix metalloproteinases (MMP-8, MMP-9) and the tissue inhibitors of metalloproteinases (TIMP-1, TIMP-2) in patients with fungal keratitis. *Cornea,* 26**,** 207-11.

ROHRBACH, M., SPENCER, H. L., PORTER, L. F., BURKITT-WRIGHT, E. M. M., BUERER, C., JANECKE, A., BAKSHI, M., SILLENCE, D., AL-HUSSAIN, H., BAUMGARTNER, M., STEINMANN, B., BLACK, G. C. M., MANSON, F. D. C. & GIUNTA, C. 2013. ZNF469 frequently mutated in the brittle cornea syndrome (BCS) is a single exon gene possibly regulating the expression of several extracellular matrix components. *Molecular Genetics and Metabolism,* 109**,** 289-295.

ROMERO-JIMENEZ, M., SANTODOMINGO-RUBIDO, J., GONZALEZ-MEIJOME, J.-M., FLORES-RODRIGUEZ, P. & VILLA-COLLAR, C. 2015. Which soft lens power is better for piggyback in keratoconus? Part II. *Contact lens & anterior eye : the journal of the British Contact Lens Association,* 38**,** 48-53.

ROMERO-JIMENEZ, M., SANTODOMINGO-RUBIDO, J. & WOLFFSOHN, J. S. 2010. Keratoconus: a review. *Cont Lens Anterior Eye,* 33**,** 157-66; quiz 205.

ROSENFELD, J. A., DRAUTZ, J. M., CLERICUZIO, C. L., CUSHING, T., RASKIN, S., MARTIN, J., TERVO, R. C., PITARQUE, J. A., NOWAK, D. M., KAROLAK, J. A., LAMB, A. N., SCHULTZ, R. A., BALLIF, B. C., BEJJANI, B. A., GAJECKA, M. & SHAFFER, L. G. 2011. Deletions and Duplications of Developmental Pathway Genes in 5q31 Contribute to Abnormal Phenotypes. *American Journal of Medical Genetics Part A,* 155A**,** 1906-1916.

SABTI, S., TAPPEINER, C. & FRUEH, B. E. 2015. Corneal Cross-Linking in a 4-Year-Old Child With Keratoconus and Down Syndrome. *Cornea,* 34**,** 1157-60.

SACCHETTI, M., MACCHI, I., TIEZZI, A., LA CAVA, M., MASSARO-GIORDANO, G. & LAMBIASE, A. 2016. Pathophysiology of Corneal

Dystrophies: From Cellular Genetic Alteration to Clinical Findings. *J Cell Physiol,* 231**,** 261-9.

SAEE-RAD, S., HASHEMI, H., MIRAFTAB, M., NOORI-DALOII, M. R., CHALESHTORI, M. H., RAOOFIAN, R., JAFARI, F., GREENE, W., FAKHRAIE, G., REZVAN, F. & HEIDARI, M. 2011. Mutation analysis of VSX1 and SOD1 in Iranian patients with keratoconus. *Molecular Vision,* 17**,** 3128-3136.

SAFFRA, N. & REINHERZ, B. 2015. Keratoconus in an adult with 22q11.2 deletion syndrome. *BMJ case reports,* 2015.

SAHEBJADA, S., SCHACHE, M., RICHARDSON, A. J., SNIBSON, G., DANIELL, M. & BAIRD, P. N. 2014. Association of the Hepatocyte Growth Factor Gene with Keratoconus in an Australian Population. *Plos One,* 9.

SAHEBJADA, S., SCHACHE, M., RICHARDSON, A. J., SNIBSON, G., MACGREGOR, S., DANIELL, M. & BAIRD, P. N. 2013. Evaluating the Association Between Keratoconus and the Corneal Thickness Genes in an Independent Australian Population. *Investigative Ophthalmology & Visual Science,* 54**,** 8224-8228.

SAKABE, J., YAMAMOTO, M., HIRAKAWA, S., MOTOYAMA, A., OHTA, I., TATSUNO, K., ITO, T., KABASHIMA, K., HIBINO, T. & TOKURA, Y. 2013. Kallikrein-related peptidase 5 functions in proteolytic processing of profilaggrin in cultured human keratinocytes. *J Biol Chem,* 288**,** 17179-89.

SALOUTI, R., NOWROOZZADEH, M. H., ZAMANI, M. & GHOREYSHI, M. 2010. Combined anterior keratoconus and Fuchs' endothelial dystrophy: a report of two cases. *Clinical and Experimental Optometry,* 93**,** 268-270.

SANO, K., TANIHARA, H., HEIMARK, R. L., OBATA, S., DAVIDSON, M., ST JOHN, T., TAKETANI, S. & SUZUKI, S. 1993. Protocadherins: a large family of cadherin-related molecules in central nervous system. *EMBO J,* 12**,** 2249-56.

SARAVANI, R., HASANIAN-LANGROUDI, F., VALIDAD, M. H., YARI, D., BAHARI, G., FARAMARZI, M., KHATERI, M. & BAHADORAM, S. 2015. Evaluation of possible relationship between COL4A4 gene polymorphisms and risk of keratoconus. *Cornea,* 34**,** 318-22.

SAWAGUCHI, S., TWINING, S. S., YUE, B. Y., WILSON, P. M., SUGAR, J. & CHAN, S. K. 1990. Alpha-1 proteinase inhibitor levels in keratoconus. *Exp Eye Res,* 50**,** 549-54.

SAWAGUCHI, S., YUE, B. Y., SUGAR, J. & GILBOY, J. E. 1989. Lysosomal enzyme abnormalities in keratoconus. *Arch Ophthalmol,* 107**,** 1507-10.

SCHAUMBERG, D. A., SNOW, K. K. & DANA, M. R. 1998. The epidemic of Acanthamoeba keratitis: where do we stand? *Cornea,* 17**,** 3-10.

SCHORK, N. J., MURRAY, S. S., FRAZER, K. A. & TOPOL, E. J. 2009. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev,* 19**,** 212-9.

SCHUMANN, H., KIRITSI, D., PIGORS, M., HAUSSER, I., KOHLHASE, J., PETERS, J., OTT, H., HYLA-KLEKOT, L., GACKA, E., SIERON, A. L., VALARI, M., BRUCKNER-TUDERMAN, L. & HAS, C. 2013. Phenotypic spectrum of epidermolysis bullosa associated with alpha6beta4 integrin mutations. *Br J Dermatol,* 169**,** 115-24.

SCHWARZ, J. M., RODELSPERGER, C., SCHUELKE, M. & SEELOW, D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods,* 7**,** 575-6.

SEDAGHAT, M., BAGHERI, M., GHAVAMI, S. & BAMDAD, S. 2015. Changes in corneal topography and biomechanical properties after collagen cross linking for keratoconus: 1-year results. *Middle East Afr J Ophthalmol,* 22**,** 212-9.

SEGAL, O., BARKANA, Y., HOUROVITZ, D., BEHRMAN, S., KAMUN, Y., AVNI, I. & ZADOK, D. 2003. Scleral contact lenses may help where other modalities fail. *Cornea,* 22**,** 308-10.

SEPPALA, H. P., MAATTA, M., RAUTIA, M., MACKIEWICZ, Z., TUISKU, I., TERVO, T. & KONTTINEN, Y. T. 2006. EMMPRIN and MMP-1 in keratoconus. *Cornea,* 25**,** 325-30.

SEVOST'IANOV, E. N., GINIATULLIN, R. U., GORSKOVA, E. N. & TEPLOVA, S. N. 2002. [Keratocyte apoptosis in keratoconus]. *Vestn Oftalmol,* 118**,** 36-8.

SEYEDIAN, M. A., ALIAKBARI, S., MIRAFTAB, M., HASHEMI, H., ASGARI, S. & KHABAZKHOOB, M. 2015. Corneal Collagen Cross-Linking in the Treatment of Progressive Keratoconus: A Randomized Controlled Contralateral Eye Study. *Middle East Afr J Ophthalmol,* 22**,** 340-5.

SEYEDNASROLLAH, F., LAIHO, A. & ELO, L. L. 2015. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform,* 16**,** 59-70.

SHAH, A., SACHDEV, A., COGGON, D. & HOSSAIN, P. 2011. Geographic variations in microbial keratitis: an analysis of the peer-reviewed literature. *Br J Ophthalmol,* 95**,** 762-7.

SHAIKH, S. & TA, C. N. 2002. Evaluation and management of herpes zoster ophthalmicus. *Am Fam Physician,* 66**,** 1723-30.

SHARMA, S., TANEJA, M., GUPTA, R., UPPONI, A., GOPINATHAN, U., NUTHETI, R. & GARG, P. 2007. Comparison of clinical and microbiological profiles in smear-positive and smear-negative cases of suspected microbial keratitis. *Indian J Ophthalmol,* 55**,** 21-5.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res,* 29**,** 308-11.

SHERVA, R., BALDWIN, C. T., INZELBERG, R., VARDARAJAN, B., CUPPLES, L. A., LUNETTA, K., BOWIRRAT, A., NAJ, A., PERICAK-VANCE, M., FRIEDLAND, R. P. & FARRER, L. A. 2011. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. *J Alzheimers Dis,* 23**,** 349-59.

SHETTY, R., NUIJTS, R. M., NANAIAH, S. G., ANANDULA, V. R., GHOSH, A., JAYADEV, C., PAHUJA, N., KUMARAMANICKAVEL, G. & NALLATHAMBI, J. 2015a. Two novel missense substitutions in the VSX1 gene: clinical and genetic analysis of families with Keratoconus from India. *BMC Med Genet,* 16**,** 33.

SHETTY, R., SATHYANARAYANAMOORTHY, A., RAMACHANDRA, R. A., ARORA, V., GHOSH, A., SRIVATSA, P. R., PAHUJA, N., NUIJTS, R. M., SINHA-ROY, A., MOHAN, R. R. & GHOSH, A. 2015b. Attenuation of lysyl oxidase and collagen gene expression in keratoconus patient corneal epithelium corresponds to disease severity. *Mol Vis,* 21**,** 12-25.

SHETTY, R., SATHYANARAYANAMOORTHY, A., RAMACHANDRA, R. A., ARORA, V., GHOSH, A., SRIVATSA, P. R., PAHUJA, N., NUIJTS, R. M. M. A., SINHA-ROY, A., MOHAN, R. R. & GHOSH, A. 2015c. Attenuation

of lysyl oxidase and collagen gene expression in keratoconus patient corneal epithelium corresponds to disease severity. *Molecular vision,* 21**,** 12-25.

SHI, Y. & MAJEWSKI, J. 2013. FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data. *Bioinformatics,* 29**,** 1461-2.

SHIHAB, H. A., GOUGH, J., MORT, M., COOPER, D. N., DAY, I. N. & GAUNT, T. R. 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics,* 8**,** 11.

SHIMAZAKI, J., SHIMMURA, S., ISHIOKA, M. & TSUBOTA, K. 2002. Randomized clinical trial of deep lamellar keratoplasty vs penetrating keratoplasty. *Am J Ophthalmol,* 134**,** 159-65.

SHLUSH, L. I., BEHAR, D. M., YUDKOVSKY, G., TEMPLETON, A., HADID, Y., BASIS, F., HAMMER, M., ITZKOVITZ, S. & SKORECKI, K. 2008. The Druze: a population genetic refugium of the Near East. *PLoS One,* 3**,** e2105.

SHNEOR, E., MILLODOT, M., BLUMBERG, S., ORTENBERG, I., BEHRMAN, S. & GORDON-SHAAG, A. 2013. Characteristics of 244 patients with keratoconus seen in an optometric contact lens practice. *Clin Exp Optom,* 96**,** 219-24.

SINJAB, M. 2012. *Quick guide to the management of keratoconus*, Springer.

SLADE, S. G. 2007. Applications for the femtosecond laser in corneal surgery. *Curr Opin Ophthalmol,* 18**,** 338-41.

SMITH, V. A. & EASTY, D. L. 2000. Matrix metalloproteinase 2: involvement in keratoconus. *Eur J Ophthalmol,* 10**,** 215-26.

SMITH, V. A., HOH, H. B., LITTLETON, M. & EASTY, D. L. 1995. Over-expression of a gelatinase A activity in keratoconus. *Eye (Lond),* 9 (Pt 4)**,** 429-33.

SORBARA, L., MARAM, J., FONN, D., WOODS, C. & SIMPSON, T. 2010. Metrics of the normal cornea: anterior segment imaging with the Visante OCT. *Clin Exp Optom,* 93**,** 150-6.

SORKHABI, R., GHORBANIHAGHJO, A., TAHERI, N. & AHOOR, M. H. 2015. Tear film inflammatory mediators in patients with keratoconus. *Int Ophthalmol,* 35**,** 467-72.

SPRINGELKAMP, H., MISHRA, A., HYSI, P. G., GHARAHKHANI, P., HOHN, R., KHOR, C. C., COOKE BAILEY, J. N., LUO, X., RAMDAS, W. D., VITHANA, E., KOH, V., YAZAR, S., XU, L., FORWARD, H., KEARNS, L. S., AMIN, N., IGLESIAS, A. I., SIM, K. S., VAN LEEUWEN, E. M., DEMIRKAN, A., VAN DER LEE, S., LOON, S. C., RIVADENEIRA, F., NAG, A., SANFILIPPO, P. G., SCHILLERT, A., DE JONG, P. T., OOSTRA, B. A., UITTERLINDEN, A. G., HOFMAN, A., ZHOU, T., BURDON, K. P., SPECTOR, T. D., LACKNER, K. J., SAW, S. M., VINGERLING, J. R., TEO, Y. Y., PASQUALE, L. R., WOLFS, R. C., LEMIJ, H. G., TAI, E. S., JONAS, J. B., CHENG, C. Y., AUNG, T., JANSONIUS, N. M., KLAVER, C. C., CRAIG, J. E., YOUNG, T. L., HAINES, J. L., MACGREGOR, S., MACKEY, D. A., PFEIFFER, N., WONG, T. Y., WIGGS, J. L., HEWITT, A. W., VAN DUIJN, C. M. & HAMMOND, C. J. 2015. Meta-analysis of Genome-Wide Association Studies Identifies Novel Loci Associated With Optic Disc Morphology. *Genet Epidemiol,* 39**,** 207-16.

SPRINGS, C. L., JOSEPH, M. A., ODOM, J. V. & WILEY, L. A. 2002. Predictability of donor lamellar graft diameter and thickness in an artificial anterior chamber system. *Cornea,* 21**,** 696-9.

SRINIVASAN, M., GONZALES, C. A., GEORGE, C., CEVALLOS, V., MASCARENHAS, J. M., ASOKAN, B., WILKINS, J., SMOLIN, G. & WHITCHER, J. P. 1997. Epidemiology and aetiological diagnosis of corneal ulceration in Madurai, south India. *Br J Ophthalmol,* 81**,** 965-71.

STABUC-SILIH, M., RAVNIK-GLAVAC, M., GLAVAC, D., HAWLINA, M. & STRAZISAR, M. 2009. Polymorphisms in COL4A3 and COL4A4 genes associated with keratoconus. *Molecular Vision,* 15**,** 2848-2859.

STABUC-SILIH, M., STRAZISAR, M., HAWLINA, M. & GLAVAC, D. 2010. Absence of Pathogenic Mutations in VSX1 and SOD1 Genes in Patients With Keratoconus. *Cornea,* 29**,** 172-176.

STEELE, T. M., FABINYI, D. C., COUPER, T. A. & LOUGHNAN, M. S. 2008. Prevalence of Orbscan II corneal abnormalities in relatives of patients with keratoconus. *Clin Experiment Ophthalmol,* 36**,** 824-30.

STEHR-GREEN, J. K., BAILEY, T. M. & VISVESVARA, G. S. 1989. The epidemiology of Acanthamoeba keratitis in the United States. *Am J Ophthalmol,* 107**,** 331-6.

STEINER-CHAMPLIAUD, M. F., SCHNEIDER, Y., FAVRE, B., PAULHE, F., PRAETZEL-WUNDER, S., FAULKNER, G., KONIECZNY, P., RAITH, M., WICHE, G., ADEBOLA, A., LIEM, R. K., LANGBEIN, L., SONNENBERG, A., FONTAO, L. & BORRADORI, L. 2010. BPAG1 isoform-b: complex distribution pattern in striated and heart muscle and association with plectin and alpha-actinin. *Exp Cell Res,* 316**,** 297-313.

STENSON, P. D., MORT, M., BALL, E. V., HOWELLS, K., PHILLIPS, A. D., THOMAS, N. S. & COOPER, D. N. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med,* 1**,** 13.

STRUNNIKOVA, N. V., MAMINISHKIS, A., BARB, J. J., WANG, F., ZHI, C., SERGEEV, Y., CHEN, W., EDWARDS, A. O., STAMBOLIAN, D., ABECASIS, G., SWAROOP, A., MUNSON, P. J. & MILLER, S. S. 2010. Transcriptome analysis and molecular signature of human retinal pigment epithelium. *Hum Mol Genet,* 19**,** 2468-86.

SUGAR, J. & MACSAI, M. S. 2012. What Causes Keratoconus? *Cornea,* 31**,** 716-719.

SUTTON, G., MADIGAN, M., ROUFAS, A. & MCAVOY, J. 2010. Secreted frizzled-related protein 1 (SFRP1) is highly upregulated in keratoconus epithelium: a novel finding highlighting a new potential focus for keratoconus research and treatment. *Clinical and Experimental Ophthalmology,* 38**,** 43-48.

SUZUKI, S. & NAITOH, Y. 1990. Amino acid sequence of a novel integrin beta 4 subunit and primary expression of the mRNA in epithelial cells. *EMBO J,* 9**,** 757-63.

SYKAKIS, E., KARIM, R., EVANS, J. R., BUNCE, C., AMISSAH-ARTHUR, K. N., PATWARY, S., MCDONNELL, P. J. & HAMADA, S. 2015. Corneal collagen cross-linking for treating keratoconus. *Cochrane Database Syst Rev,* 3**,** CD010621.

SYNOWIEC, E., WOJCIK, K. A., CZUBATKA, A., POLAKOWSKI, P., IZDEBSKA, J., SZAFLIK, J., BLASIAK, J. & SZAFLIK, J. P. 2015. Lack of association between polymorphisms of the DNA base excision repair

genes MUTYH and hOGG1 and keratoconus in a Polish subpopulation. *Arch Med Sci,* 11**,** 1101-10.

SYNOWIEC, E., WOJCIK, K. A., IZDEBSKA, J., BLASIAK, J., SZAFLIK, J. & SZAFLIK, J. P. 2014. Polymorphisms of the Apoptosis-Related FAS and FAS Ligand Genes in Keratoconus and Fuchs Endothelial Corneal Dystrophy. *Tohoku Journal of Experimental Medicine,* 234**,** 17-27.

TAKAHASHI, A., NAKAYASU, K., OKISAKA, S. & KANAI, A. 1990. [Quantitative analysis of collagen fiber in keratoconus]. *Nippon Ganka Gakkai Zasshi,* 94**,** 1068-73.

TANG, Y. G., PICORNELL, Y., SU, X., LI, X., YANG, H. & RABINOWITZ, Y. S. 2008. Three VSX1 gene mutations, L159M, R166W, and H244R, are not associated with keratoconus. *Cornea,* 27**,** 189-192.

TANG, Y. G., RABINOWITZ, Y. S., TAYLOR, K. D., LI, X., HU, M., PICORNELL, Y. & YANG, H. 2005a. Genomewide linkage scan in a multigeneration Caucasian pedigree identifies a novel locus for keratoconus on chromosome 5q14.3-q21.1. *Genet Med,* 7**,** 397-405.

TANG, Y. M. G., RABINOWITZ, Y. S., TAYLOR, K. D., LI, X. H., HU, M. S., PICORNELL, Y. & YANG, H. Y. 2005b. Genomewide linkage scan in a multigeneration Caucasian pedigree identifies a novel locus for keratoconus on chromosome 5q14.3-q21.1. *Genetics in Medicine,* 7**,** 397-405.

TANWAR, M., KUMAR, M., NAYAK, B., PATHAK, D., SHARMA, N., TITIYAL, J. S. & DADA, R. 2010. VSX1 gene analysis in keratoconus. *Molecular Vision,* 16**,** 2395-2401.

TEMSTET, C., SANDALI, O., BOUHERAOUA, N., HAMICHE, T., GALAN, A., EL SANHARAWI, M., BASLI, E., LAROCHE, L. & BORDERIE, V. 2015. Corneal epithelial thickness mapping using Fourier-domain optical coherence tomography for detection of form fruste keratoconus. *J Cataract Refract Surg,* 41**,** 812-20.

TETREAULT, M., BAREKE, E., NADAF, J., ALIREZAIE, N. & MAJEWSKI, J. 2015. Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. *Expert Rev Mol Diagn,* 15**,** 749-60.

THOTA, S., MILLER, W. L. & BERGMANSON, J. P. 2006. Acute corneal hydrops: a case report including confocal and histopathological considerations. *Cont Lens Anterior Eye,* 29**,** 69-73.

TOH, T., LIEW, S. H., MACKINNON, J. R., HEWITT, A. W., POULSEN, J. L., SPECTOR, T. D., GILBERT, C. E., CRAIG, J. E., HAMMOND, C. J. & MACKEY, D. A. 2005. Central corneal thickness is highly heritable: the twin eye studies. *Invest Ophthalmol Vis Sci,* 46**,** 3718-22.

TOPRAK, I., KUCUKATAY, V., YILDIRIM, C., KILIC-TOPRAK, E. & KILIC-ERKEK, O. 2014. Increased systemic oxidative stress in patients with keratoconus. *Eye (Lond),* 28**,** 285-9.

TOSHIDA, H., KOGURE, N., INOUE, N. & MURAKAMI, A. 2007. Trends in microbial keratitis in Japan. *Eye Contact Lens,* 33**,** 70-3.

TRAPNELL, C., HENDRICKSON, D. G., SAUVAGEAU, M., GOFF, L., RINN, J. L. & PACHTER, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol,* 31**,** 46-53.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols,* 7**,** 562-578.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol,* 28**,** 511-5.

TUFT, S. J., HASSAN, H., GEORGE, S., FRAZER, D. G., WILLOUGHBY, C. E. & LISKOVA, P. 2012. Keratoconus in 18 pairs of twins. *Acta Ophthalmologica,* 90**,** e482-e486.

TUORI, A., VIRTANEN, I., AINE, E. & UUSITALO, H. 1997. The expression of tenascin and fibronectin in keratoconus, scarred and normal human cornea. *Graefes Arch Clin Exp Ophthalmol,* 235**,** 222-9.

TWINE, N. A., JANITZ, K., WILKINS, M. R. & JANITZ, M. 2011. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One,* 6**,** e16266.

TYYNISMAA, H., SISTONEN, P., TUUPANEN, S., TERVO, T., DAMMERT, A., LATVALA, T. & ALITALO, T. 2002. A locus for autosomal dominant keratoconus: Linkage to 16q22.3-q23.1 in Finnish families. *Investigative Ophthalmology & Visual Science,* 43**,** 3160-3164.

UDAR, N., ATILANO, S. R., BROWN, D. J., HOLGUIN, B., SMALL, K., NESBURN, A. B. & KENNEY, M. C. 2006. SOD1: A candidate gene for keratoconus. *Investigative Ophthalmology & Visual Science,* 47**,** 3345-3351.

UDAR, N., ATILANO, S. R., SMALL, K., NESBURN, A. B. & KENNEY, M. C. 2009. SOD1 Haplotypes in Familial Keratoconus. *Cornea,* 28**,** 902-907.

VAJPAYEE, R. B., BHARTIYA, P. & SHARMA, N. 2002. Central lamellar keratoplasty with peripheral intralamellar tuck: a new surgical technique for keratoglobus. *Cornea,* 21**,** 657-60.

VALLURI, S., MINKOVITZ, J. B., BUDAK, K., ESSARY, L. R., WALKER, R. S., CHANSUE, E., CABRERA, G. M., KOCH, D. D. & PEPOSE, J. S. 1999. Comparative corneal topography and refractive variables in monozygotic and dizygotic twins. *Am J Ophthalmol,* 127**,** 158-63.

VAN DE VEN, J. P., NILSSON, S. C., TAN, P. L., BUITENDIJK, G. H., RISTAU, T., MOHLIN, F. C., NABUURS, S. B., SCHOENMAKER-KOLLER, F. E., SMAILHODZIC, D., CAMPOCHIARO, P. A., ZACK, D. J., DUVVARI, M. R., BAKKER, B., PAUN, C. C., BOON, C. J., UITTERLINDEN, A. G., LIAKOPOULOS, S., KLEVERING, B. J., FAUSER, S., DAHA, M. R., KATSANIS, N., KLAVER, C. C., BLOM, A. M., HOYNG, C. B. & DEN HOLLANDER, A. I. 2013. A functional variant in the CFI gene confers a high risk of age-related macular degeneration. *Nat Genet,* 45**,** 813-7.

VAN DER FLIER, A., KUIKMAN, I., KRAMER, D., GEERTS, D., KREFT, M., TAKAFUTA, T., SHAPIRO, S. S. & SONNENBERG, A. 2002. Different splice variants of filamin-B affect myogenesis, subcellular distribution, and determine binding to integrin β subunits. *The Journal of Cell Biology,* 156**,** 361-376.

VAN DIJK, K., LIARAKOS, V. S., PARKER, J., HAM, L., LIE, J. T., GROENEVELD-VAN BEEK, E. A. & MELLES, G. R. 2015. Bowman layer transplantation to reduce and stabilize progressive, advanced keratoconus. *Ophthalmology,* 122**,** 909-17.

VASILEVA, V. A. 1957. [Epithelial development of human cornea in embryogenesis]. *Arkh Anat Gistol Embriol,* 34**,** 59-63.

VERMA, A., DAS, M., SRINIVASAN, M., PRAJNA, N. V. & SUNDARESAN, P. 2013. Investigation of VSX1 sequence variants in South Indian patients with sporadic cases of keratoconus. *BMC research notes,* 6**,** 103-103.

VINCENT, A. L. 2014. Corneal dystrophies and genetics in the International Committee for Classification of Corneal Dystrophies era: a review. *Clin Experiment Ophthalmol,* 42**,** 4-12.

VINCENT, A. L., JORDAN, C., SHECK, L., NIEDERER, R., PATEL, D. V. & MCGHEE, C. N. 2013a. Screening the visual system homeobox 1 gene in keratoconus and posterior polymorphous dystrophy cohorts identifies a novel variant. *Mol Vis,* 19**,** 852-60.

VINCENT, A. L., JORDAN, C., SHECK, L., NIEDERER, R., PATEL, D. V. & MCGHEE, C. N. J. 2013b. Screening the visual system homeobox 1 gene in keratoconus and posterior polymorphous dystrophy cohorts identifies a novel variant. *Molecular Vision,* 19**,** 852-860.

VINCENT, A. L., JORDAN, C. A., CADZOW, M. J., MERRIMAN, T. R. & MCGHEE, C. N. 2014. Mutations in the Zinc Finger Protein Gene, ZNF469, Contribute to the Pathogenesis of Keratoconus. *Investigative Ophthalmology & Visual Science,* 55**,** 5629-5635.

VINCENT, A. L., NIEDERER, R. L., RICHARDS, A., KAROLYI, B., PATEL, D. V. & MCGHEE, C. N. 2009. Phenotypic characterisation and ZEB1 mutational analysis in posterior polymorphous corneal dystrophy in a New Zealand population. *Mol Vis,* 15**,** 2544-53.

VINCIGUERRA, P., RANDLEMAN, J. B., ROMANO, V., LEGROTTAGLIE, E. F., ROSETTA, P., CAMESASCA, F. I., PISCOPO, R., AZZOLINI, C. & VINCIGUERRA, R. 2014. Transepithelial iontophoresis corneal collagen cross-linking for progressive keratoconus: initial clinical outcomes. *J Refract Surg,* 30**,** 746-53.

VISSER, E. S., VISSER, R., VAN LIER, H. J. & OTTEN, H. M. 2007a. Modern scleral lenses part I: clinical features. *Eye Contact Lens,* 33**,** 13-20.

VISSER, E. S., VISSER, R., VAN LIER, H. J. & OTTEN, H. M. 2007b. Modern scleral lenses part II: patient satisfaction. *Eye Contact Lens,* 33**,** 21-5.

VON DER HEYDT, R. 1930. Slit-Lamp Observations on Keratoconus. *Trans Am Ophthalmol Soc,* 28**,** 352-61.

WACHTMEISTER, L., INGEMANSSON, S. O. & MOLLER, E. 1982. Atopy and HLA antigens in patients with keratoconus. *Acta Ophthalmol (Copenh),* 60**,** 113-22.

WAGNER, H., BARR, J. T. & ZADNIK, K. 2007. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study: methods and findings to date. *Cont Lens Anterior Eye,* 30**,** 223-32.

WAKED, N., FAYAD, A. M., FADLALLAH, A. & EL RAMI, H. 2012. [Keratoconus screening in a Lebanese students' population]. *J Fr Ophtalmol,* 35**,** 23-9.

WANG, H., NETTLETON, D. & YING, K. 2014. Copy number variation detection using next generation sequencing read counts. *BMC Bioinformatics,* 15**,** 109.

WANG, K., LI, M. & HAKONARSON, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research,* 38**,** e164.

WANG, L., MA, R., DU, G., GUO, H. & HUANG, Y. 2015a. Biocompatibility of helicoidal multilamellar arginine-glycine-aspartic acid-functionalized silk biomaterials in a rabbit corneal model. *Journal of Biomedical Materials Research Part B-Applied Biomaterials,* 103**,** 204-211.

WANG, Y., JIN, T., ZHANG, X., WEI, W., CUI, Y., GENG, T., LIU, Q., GAO, J., LIU, M., CHEN, C., ZHANG, C. & ZHU, X. 2013a. Common Single Nucleotide Polymorphisms and Keratoconus in the Han Chinese Population. *Ophthalmic Genetics,* 34**,** 160-166.

WANG, Y., RABINOWITZ, Y. S., ROTTER, J. I. & YANG, H. 2000. Genetic epidemiological study of keratoconus: Evidence for major gene determination. *American Journal of Medical Genetics,* 93**,** 403-409.

WANG, Y., WEI, W., ZHANG, C., ZHANG, X., LIU, M., ZHU, X. & XU, K. 2015b. Association of Interleukin-1 Gene Single Nucleotide Polymorphisms with Keratoconus in Chinese Han Population. *Curr Eye Res***,** 1-6.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet,* 10**,** 57-63.

WANG, Z., LIU, X., YANG, B. Z. & GELERNTER, J. 2013b. The role and challenges of exome sequencing in studies of human diseases. *Front Genet,* 4**,** 160.

WARD, A. J. & COOPER, T. A. 2010. The Pathobiology of Splicing. *The Journal of pathology,* 220**,** 152-163.

WATERLAND, R. A. 2006. Epigenetic mechanisms and gastrointestinal development. *J Pediatr,* 149**,** S137-42.

WATSKY, M. A., MCDERMOTT, M. L. & EDELHAUSER, H. F. 1989. In vitro corneal endothelial permeability in rabbit and human: the effects of age, cataract surgery and diabetes. *Exp Eye Res,* 49**,** 751-67.

WEED, K. H., MACEWEN, C. J., GILES, T., LOW, J. & MCGHEE, C. N. 2008. The Dundee University Scottish Keratoconus study: demographics, corneal signs, associated diseases, and eye rubbing. *Eye (Lond),* 22**,** 534-41.

WEED, K. H. & MCGHEE, C. N. 1998. Referral patterns, treatment management and visual outcome in keratoconus. *Eye (Lond),* 12 (Pt 4)**,** 663-8.

WEISS, J. S., MOLLER, H. U., ALDAVE, A. J., SEITZ, B., BREDRUP, C., KIVELA, T., MUNIER, F. L., RAPUANO, C. J., NISCHAL, K. K., KIM, E. K., SUTPHIN, J., BUSIN, M., LABBE, A., KENYON, K. R., KINOSHITA, S. & LISCH, W. 2015. IC3D Classification of Corneal Dystrophies-Edition 2. *Cornea,* 34**,** 117-159.

WEISS, J. S., MOLLER, H. U., LISCH, W., KINOSHITA, S., ALDAVE, A. J., BELIN, M. W., KIVELA, T., BUSIN, M., MUNIER, F. L., SEITZ, B., SUTPHIN, J., BREDRUP, C., MANNIS, M. J., RAPUANO, C. J., VAN RIJ, G., KIM, E. K. & KLINTWORTH, G. K. 2008. The IC3D classification of the corneal dystrophies. *Cornea,* 27 Suppl 2**,** S1-83.

WHEELER, J., HAUSER, M. A., AFSHARI, N. A., ALLINGHAM, R. R. & LIU, Y. 2012. The Genetics of Keratoconus: A Review. *Reprod Syst Sex Disord*.

WHITE, D. R., GANESH, A., NISHIMURA, D., RATTENBERRY, E., AHMED, S., SMITH, U. M., PASHA, S., RAEBURN, S., TREMBATH, R. C., RAJAB, A., MACDONALD, F., BANIN, E., STONE, E. M., JOHNSON, C. A., SHEFFIELD, V. C. & MAHER, E. R. 2007. Autozygosity mapping of Bardet-Biedl syndrome to 12q21.2 and confirmation of FLJ23560 as BBS10. *Eur J Hum Genet,* 15**,** 173-8.

WHITELOCK, R. B., FUKUCHI, T., ZHOU, L., TWINING, S. S., SUGAR, J., FEDER, R. S. & YUE, B. Y. 1997a. Cathepsin G, acid phosphatase, and alpha 1-proteinase inhibitor messenger RNA levels in keratoconus corneas. *Invest Ophthalmol Vis Sci,* 38**,** 529-34.

WHITELOCK, R. B., LI, Y., ZHOU, L. L., SUGAR, J. & YUE, B. Y. 1997b. Expression of transcription factors in keratoconus, a cornea-thinning disease. *Biochem Biophys Res Commun,* 235**,** 253-8.

WILLOUGHBY, C. E., PONZIN, D., FERRARI, S., LOBO, A., LANDAU, K. & OMIDI, Y. 2010. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review. *Clinical & Experimental Ophthalmology,* 38**,** 2-11.

WILSON, C. M., D'ATH, P. J., PARMAR, D. N. & SYKAKIS, E. 2014. Keratoconus and granular dystrophy. *BMJ Case Rep,* 2014.

WILSON, S. E., HE, Y. G., WENG, J., LI, Q., MCDOWALL, A. W., VITAL, M. & CHWANG, E. L. 1996a. Epithelial injury induces keratocyte apoptosis: hypothesized role for the interleukin-1 system in the modulation of corneal tissue organization and wound healing. *Exp Eye Res,* 62**,** 325-7.

WILSON, S. E., LI, Q., WENG, J., BARRY-LANE, P. A., JESTER, J. V., LIANG, Q. & WORDINGER, R. J. 1996b. The Fas-Fas ligand system and other modulators of apoptosis in the cornea. *Invest Ophthalmol Vis Sci,* 37**,** 1582-92.

WISSE, R. P., KUIPER, J. J., GANS, R., IMHOF, S., RADSTAKE, T. R. & VAN DER LELIJ, A. 2015. Cytokine Expression in Keratoconus and its Corneal Microenvironment: A Systematic Review. *Ocul Surf*.

WOJCIK, K. A., BLASIAK, J., KUROWSKA, A. K., SZAFLIK, J. & SZAFLIK, J. P. 2013a. [Oxidative stress in the pathogenesis of keratoconus]. *Klin Oczna,* 115**,** 311-6.

WOJCIK, K. A., BLASIAK, J., SZAFLIK, J. & SZAFLIK, J. P. 2014a. Role of biochemical factors in the pathogenesis of keratoconus. *Acta Biochimica Polonica,* 61**,** 55-62.

WOJCIK, K. A., KAMINSKA, A., BLASIAK, J., SZAFLIK, J. & SZAFLIK, J. P. 2013b. Oxidative Stress in the Pathogenesis of Keratoconus and Fuchs Endothelial Corneal Dystrophy. *International Journal of Molecular Sciences,* 14**,** 19294-19308.

WOJCIK, K. A., SYNOWIEC, E., SOBIERAJCZYK, K., IZDEBSKA, J., BLASIAK, J., SZAFLIK, J. & SZAFLIK, J. P. 2014b. Polymorphism of the DNA Base Excision Repair Genes in Keratoconus. *International Journal of Molecular Sciences,* 15**,** 19682-19699.

WOLLENSAK, G., SPORL, E. & SEILER, T. 2003. [Treatment of keratoconus by collagen cross linking]. *Ophthalmologe,* 100**,** 44-9.

WOODS, C. G., COX, J., SPRINGELL, K., HAMPSHIRE, D. J., MOHAMED, M. D., MCKIBBIN, M., STERN, R., RAYMOND, F. L., SANDFORD, R., MALIK SHARIF, S., KARBANI, G., AHMED, M., BOND, J., CLAYTON, D. & INGLEHEARN, C. F. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet,* 78**,** 889-96.

WOODWARD, E. G. & MORRIS, M. T. 1990. Joint hypermobility in keratoconus. *Ophthalmic Physiol Opt,* 10**,** 360-2.

WU, L., SCHAID, D. J., SICOTTE, H., WIEBEN, E. D., LI, H. & PETERSEN, G. M. 2015. Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. *J Med Genet,* 52**,** 10-6.

XU, D., HOU, S., ZHANG, J., JIANG, Y., KIJLSTRA, A. & YANG, P. 2015. Copy number variations and gene polymorphisms of complement components in ocular Behcet's disease and Vogt-Koyanagi-Harada syndrome. *Sci Rep,* 5**,** 12989.

XU, W., XIE, Z., CHUNG, D. W. & DAVIE, E. W. 1998. A novel human actin-binding protein homologue that binds to platelet glycoprotein Ibalpha. *Blood,* 92**,** 1268-76.

YAKES, F. M. & VAN HOUTEN, B. 1997. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. *Proc Natl Acad Sci U S A,* 94**,** 514-9.

YANG, H. & WANG, K. 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protocols,* 10**,** 1556-1566.

YEH, S. & SMITH, J. A. 2008. Management of acute hydrops with perforation in a patient with keratoconus and cone dystrophy: Case report and literature review. *Cornea,* 27**,** 1062-1065.

YENDREK, C. R., AINSWORTH, E. A. & THIMMAPURAM, J. 2012. The bench scientist's guide to statistical analysis of RNA-Seq data. *BMC Res Notes,* 5**,** 506.

YOU, J., HODGE, C., WEN, L., MCAVOY, J. W., MADIGAN, M. C. & SUTTON, G. 2013a. Tear levels of SFRP1 are significantly reduced in keratoconus patients. *Mol Vis,* 19**,** 509-xxx.

YOU, J., WEN, L., ROUFAS, A., MADIGAN, M. C. & SUTTON, G. 2013b. Expression of SFRP Family Proteins in Human Keratoconus Corneas. *PLoS One,* 8**,** e66770.

YOU, Z., CAO, X., TAYLOR, A. B., HART, P. J. & LEVINE, R. L. 2010. Characterization of a covalent polysulfane bridge in copper-zinc superoxide dismutase. *Biochemistry,* 49**,** 1191-8.

YU, X., GUDA, K., WILLIS, J., VEIGL, M., WANG, Z., MARKOWITZ, S., ADAMS, M. D. & SUN, S. 2012. How do alignment programs perform on sequencing data with varying qualities and from repetitive regions? *BioData Min,* 5**,** 6.

ZADNIK, K., BARR, J. T., GORDON, M. O. & EDRINGTON, T. B. 1996. Biomicroscopic signs and disease severity in keratoconus. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study Group. *Cornea,* 15**,** 139-46.

ZADNIK, K., STEGER-MAY, K., FINK, B. A., JOSLIN, C. E., NICHOLS, J. J., ROSENSTIEL, C. E., TYLER, J. A., YU, J. A., RAASCH, T. W. & SCHECHTMAN, K. B. 2002. Between-eye asymmetry in keratoconus. *Cornea,* 21**,** 671-9.

ZALLOUA, P. A., XUE, Y., KHALIFE, J., MAKHOUL, N., DEBIANE, L., PLATT, D. E., ROYYURU, A. K., HERRERA, R. J., HERNANZ, D. F., BLUE-SMITH, J., WELLS, R. S., COMAS, D., BERTRANPETIT, J. & TYLER-SMITH, C. 2008. Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am J Hum Genet,* 82**,** 873-82.

ZHANG, Z. H., JHAVERI, D. J., MARSHALL, V. M., BAUER, D. C., EDSON, J., NARAYANAN, R. K., ROBINSON, G. J., LUNDBERG, A. E., BARTLETT, P. F., WRAY, N. R. & ZHAO, Q.-Y. 2014. A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. *PLoS ONE,* 9**,** e103207.

ZHANG, Z. H., JHAVERI, D. J., MARSHALL, V. M., BAUER, D. C., EDSON, J., NARAYANAN, R. K., ROBINSON, G. J., LUNDBERG, A. E., BARTLETT, P. F., WRAY, N. R. & ZHAO, Q. Y. A comparative study of techniques for differential expression analysis on RNA-Seq data.

ZHAO, M., WANG, Q., WANG, Q., JIA, P. & ZHAO, Z. 2013. Computational tools for copy number variation (CNV) detection using next-generation

sequencing data: features and perspectives. *BMC Bioinformatics,* 14 Suppl 11**,** S1.

ZHENG, W., CHUNG, L. M. & ZHAO, H. 2011. Bias detection and correction in RNA-Sequencing data. *BMC Bioinformatics,* 12**,** 290-290.

ZHOU, L., SAWAGUCHI, S., TWINING, S. S., SUGAR, J., FEDER, R. S. & YUE, B. Y. 1998. Expression of degradative enzymes and protease inhibitors in corneas with keratoconus. *Invest Ophthalmol Vis Sci,* 39**,** 1117-24.

ZHU, L. & ZHU, H. 2014. Ocular herpes: the pathophysiology, management and treatment of herpetic eye diseases. *Virol Sin,* 29**,** 327-42.

ZIGRINO, P., STEIGER, J., FOX, J. W., LOFFEK, S., SCHILD, A., NISCHT, R. & MAUCH, C. 2007. Role of ADAM-9 disintegrin-cysteine-rich domains in human keratinocyte migration. *J Biol Chem,* 282**,** 30785-93.

ZUK, O., HECHTER, E., SUNYAEV, S. R. & LANDER, E. S. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A,* 109**,** 1193-8.

Appendix

1. WES commands

1.1 Alignment

Reference file (fasta format) indexing

$ **bwa index** -p hg_19.bwaidx –a bwtsw hg_19.fasta

Sample fastq files alignment to the reference file

$ **bwa aln** –t 4 hg_19.fa R1-sequence.fq > R1-sequence.sai

$ **bwa aln** –t 4 hg_19.fa R2-sequence.fq > R2-sequence.sai

$ **bwa sampe** –r "@RG\tID:ID\tSM:ID\tPL:Illumina\tPU:HiSeq" hg_19.fa R1-sequence.sai R2-sequence.sai R1-sequence.fq R2-sequence.fq > output.sam

Sorting and indexing of the output.sam file according to the reference file using samtools.

Create a reference index file

$ **samtools faidx** hg_19.fa

Index the output.sam to the reference index and sort

$ **samtools view –bt** hg_19.fa.fai output.sam > output.bam

$ **samtools sort** output.bam output-sort

$ **samtools index** output-sort.bam

1.2 Local realignment

Local realignments around indels using GATK.

List of dubious alignments generation

$ java –Xmx50g –jar GenomeAnalysisTK.jar –T **RealignerTargetCreator** –R hg_19.fasta –I output-sort.bam –known hg19_indels.vcf –log output.intervals.log –o output.intervals

Realignment over the dubious regions

$ java – Xmx50g –jar GenomeAnalysisTK.jar –T **IndelRealigner** –R hg19.fasta –I output-sort.bam –targetIntervals output.intervals –log output.realn.log –o output-sort.realn.bam

1.3 Remove duplicates

Removal of duplicate reads using Picard tools

$ java –Xmx20g –Dsnappy.disable=true –jar **MarkDuplicates.jar** INPUT=output-sort.realn.bam REMOVE_DUPLICATES=true VALIDATION_STRINGENCY=LENIENT AS=true METRICS_FILE=output-sort.realn.metrics.dups OUTPUT=output-sort.realn.dedup.bam

Re-indexing of the file without the duplicates using samtools

$ **samtools index** output-sort.realn.dedup.bam

1.4 Variant calling

Calling variants of the *output-sort.realn.dedup.bam* is done using GATK:

$ java –Xmx30g –jar GenomeAnalysisTK.jar –T **UnifiedGenotyper** –R hg_19.fa –I output-sort.realn.dedup.bam –D dbsnp.vcf –L intervals.interval_list –nt 8 –stand_call_conf 50.0 –stand_emit_conf 10.0 –dcov 200 –I INFO –A AlleleBalance –A DepthOfCoverage –A FisherStrand –log output.variants.log –o output.variants.SNP.vcf -mbq 15 -glm SNP

$ java –Xmx30g –jar GenomeAnalysisTK.jar –T **UnifiedGenotyper** –R hg_19.fa –I output-sort.realn.dedup.bam –D dbsnp.vcf –L intervals.interval_list –nt 8 –stand_call_conf 50.0 –stand_emit_conf 10.0 –dcov 200 –I INFO –A AlleleBalance –A DepthOfCoverage –A FisherStrand –log output.variants.log –o output.variants.INDEL.vcf -mbq 15 -glm INDEL

1.5 Variant recalibration

$ java –Xmx30g –jar GATK.jar –T **VariantRecalibrator** –R hg_19.fa –input output.variants.SNP.vcf resource:hapmap,VCF,known=false,training=true,truth=true,prior=15.0 hapmap.vars.vcf –resource:omni,VCF,known=false,training=true,truth=false,prior=12.0 hg19.omnivars.vcf –resource:dbsnp,vcf,known=true,training=false,truth=false,prior=8.0 hg_19.dbsnp.vcf –an QD –an HaplotypeScore –an MQRankSum –an ReadPosRankSum –an FS –an MQ –-maxGaussians 6 –nt 8 -mode BOTH –log output.recal.log –recalFile output.recal.recal –tranchesFile output.recal.tranches –rscriptFile output.plot.R

Apply recalibration:

$ java –Xmx30g –jar GATK.jar –T **ApplyRecalibration** –R hg_19.fa –input output.variants.SNP.vcf –recalFile output.recal.recal –tranchesFile output.recal.tranches –ts_filter_level 99.0 –log output.apply_recal.log –o output.variants.recal.SNP.vcf

Combine variants:

$ java –Xmx30g –jar GATK.jar –T **CombineVariants** –R hg_19.fa --variant output.variants.recal.SNP.vcf --variant output.variants.INDEL.vcf –o combined.vcf

1.6 Annotate

Convert the *combined.vcf* to a file compatible with ANNOVAR.

$ Perl **convert2annovar.pl** combine.vcf –format vcf4 –withzyg -includeinfo –o combine.annovar.vcf

Annotating the *combined.vcf* file using ANNOVAR

$ perl **table_annovar.pl** combine.annovar.vcf /humandb/ -buildver hg19 -protocol refGene,ensGene,knownGene,genomicSuperDups,phastConsElements46way,cytoBand,esp6500si_all,1000g2014oct_all,snp138,ljb26_all,exac03,cadd,clinvar_20140929,avsift -otherinfo -remove -operation g,g,g,r,r,r,f,f,f,f,f,f,f,f -nastring . –outfile combine.annovar.annotated -argument "-splicing_threshold 5",,,,,,,,,,,,,

1.7 Depth of coverage

$java -Xmx30g -jar GenomeAnalysisTK.jar -T **DepthOfCoverage** -R hg_19.fasta -mbq 17 -mmq 10 -I Sample.fastq -o SamplesStats.txt -L V5-Capturekit_Regions.bed -ct 5 -ct 10 -ct 15 -ct 20

2. Highlight-similarities – perl script

The below script is developed specifically for this study.

highlight-similarities-xlsx

Copyright (C) 2014 Qais Yousef

highlight-similarities-xlsx is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 2 of the License, or (at your option) any later version.

highlight-similarities-xlsx is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with highlight-similarities-xlsx. If not, see <http://www.gnu.org/licenses/>.

```
#!/usr/bin/perl -w


use strict;
use POSIX;
use Spreadsheet::XLSX;
use Excel::Writer::XLSX;

# auto flush the output after every print
$| = 1;
```

```
# ignore some annoying warnings from Spreadsheet::XLSX
$SIG{'__WARN__'} = sub { warn $_[0] unless (caller eq
"Spreadsheet::XLSX"); };

my $destname = shift @ARGV;
my $sourcename = shift @ARGV
        or die "invocation: $0 <destination file> <source1 file> <source2 file>
.. <sourceN file>";

print "Opening <$sourcename> (this might take a while if the file is large)...";
my $source_book = Spreadsheet::XLSX->new($sourcename)
        or die "Could not open source Excel file $sourcename: $!";
my $dest_book = Excel::Writer::XLSX->new($destname)
        or die "Could not open source Excel file $destname: $!";

# Create an array of the result of the source files to compare with
my @cmpname;
my $cmp_book;
my $cmp_sheet;
foreach my $arg (0 .. $#ARGV)
{
        $cmpname[$arg] = shift @ARGV
                or die "invocation: $0 <destination file> <source1 file>
<source2 file> .. <sourceN file>";
        if (!-e $cmpname[$arg])
        {
                die "invalid file name: <$cmpname[$arg]>";
        }
}

my $matches = 0;
my $format = $dest_book->add_format();
$format->set_bg_color('yellow');

# which colums to compare?
my $cell0 = 1;
my $cell1 = 4;
my $cell2 = 6;

# final result
my @result = ();


# copy the source_sheet to dest_book
my $dest_sheet = $dest_book->add_worksheet();
my $source_sheet = $source_book->{Worksheet}[0];
foreach my $src_row_index ($source_sheet->{MinRow} .. $source_sheet-
>{MaxRow})
{
        next unless defined $source_sheet->{MaxRow};
```

```
	next unless $source_sheet->{MinRow} <= $source_sheet-
>{MaxRow};

	foreach my $src_col_index ($source_sheet->{MinCol} ..
$source_sheet->{MaxCol})
	{
		next unless defined $source_sheet->{MaxCol};
		next unless $source_sheet->{MinCol} <= $source_sheet-
>{MaxCol};

		my $source_cell = $source_sheet-
>{Cells}[$src_row_index][$src_col_index];
		$dest_sheet->write($src_row_index, $src_col_index,
$source_cell->{Val});
	}
}

print "[2K\rComparing <$sourcename> ", $source_sheet->{MaxRow}+1, "
rows\nwith:\n";

print "Opening <$cmpname[0]> (this might take a while if the file is large)...";
$cmp_book = Spreadsheet::XLSX->new($cmpname[0])
	or die "Could not open source Excel file $cmpname[0]: $!";
$cmp_sheet = $cmp_book->{Worksheet}[0];
print "[2K\r - <$cmpname[0]>\t", $cmp_sheet->{MaxRow}+1, " rows\n";

foreach my $src_row_index ($source_sheet->{MinRow} .. $source_sheet-
>{MaxRow})
{
	foreach my $cmp_row_index ($cmp_sheet->{MinRow} .. $cmp_sheet-
>{MaxRow})
	{
		my $source_cell0 = $source_sheet-
>{Cells}[$src_row_index][$cell0];
		my $cmp_cell0 = $cmp_sheet-
>{Cells}[$cmp_row_index][$cell0];
		my $source_cell1 = $source_sheet-
>{Cells}[$src_row_index][$cell1];
		my $cmp_cell1 = $cmp_sheet-
>{Cells}[$cmp_row_index][$cell1];
		my $source_cell2 = $source_sheet-
>{Cells}[$src_row_index][$cell2];
		my $cmp_cell2 = $cmp_sheet-
>{Cells}[$cmp_row_index][$cell2];
		if ($source_cell0 && $cmp_cell0 &&
			$source_cell1 && $cmp_cell1 &&
			$source_cell2 && $cmp_cell2 &&
			$source_cell0->Value eq $cmp_cell0->Value &&
			$source_cell1->Value eq $cmp_cell1->Value &&
			$source_cell2->Value eq $cmp_cell2->Value)
		{
```

```
                        push(@result, $src_row_index);
                        last;
                } # end of source_cell check
        } # foreach cmp_row_index
        if ($source_sheet->{MaxRow})
        {
                print floor($src_row_index/$source_sheet->{MaxRow}*100),
"%\r";
        }
} # foreach src_row_index

# compare with the result of the files using the result variable
if ($#cmpname > 0)
{
        foreach my $i (1 .. $#cmpname)
        {
                print "Opening <$cmpname[$i]> (this might take a while if the
file is large)...";
                $cmp_book = Spreadsheet::XLSX->new($cmpname[$i])
                        or die "Could not open source Excel file $cmpname[$i]:
$!";
                $cmp_sheet = $cmp_book->{Worksheet}[0];
                print "[2K\r - <$cmpname[$i]>\t", $cmp_sheet->{MaxRow}+1, "
rows\n";

                foreach my $j (0 .. $#result)
                {
                        my $src_row_index = $result[$j];
                        if ($src_row_index < 0)
                        {
                                next;
                        }

                        my $found = 0;

                        foreach my $cmp_row_index ($cmp_sheet->{MinRow} ..
$cmp_sheet->{MaxRow})
                        {
                                my $source_cell0 = $source_sheet
>{Cells}[$src_row_index][$cell0];
                                my $cmp_cell0 = $cmp_sheet-
>{Cells}[$cmp_row_index][$cell0];
                                my $source_cell1 = $source_sheet-
>{Cells}[$src_row_index][$cell1];
                                my $cmp_cell1 = $cmp_sheet-
>{Cells}[$cmp_row_index][$cell1];
                                my $source_cell2 = $source_sheet-
>{Cells}[$src_row_index][$cell2];
                                my $cmp_cell2 = $cmp_sheet-
>{Cells}[$cmp_row_index][$cell2];
                                if ($source_cell0 && $cmp_cell0 &&
```

```
                                        $source_cell1 && $cmp_cell1 &&
                                        $source_cell2 && $cmp_cell2 &&
                                        $source_cell0->Value eq $cmp_cell0-
>Value &&
                                        $source_cell1->Value eq $cmp_cell1-
>Value &&
                                        $source_cell2->Value eq $cmp_cell2-
>Value)
                                {
                                        $found = 1;
                                        last;
                                } # end of source_cell check
                        } # foreach cmp_row_index
                        if ($source_sheet->{MaxRow})
                        {
                                print floor($src_row_index/$source_sheet-
>{MaxRow}*100), "%\r";
                        }
                        if (!$found)
                        {
                                $result[$j] = -1;
                        }
                } # foreach src_row_index
        } # foreach cmp_sheet_number
}

# Finally highlight the rseult in the output
foreach my $i (0 .. $#result)
{
        if ($result[$i] >= 0)
        {
                $matches += 1;
                $dest_sheet->set_row($result[$i], undef, $format);
        }
}

print  $matches, " matches found\n";
print  "Saved changes to ", $destname, "\n";

print "done!\n";
```

3. RNA-seq commands

3.1 Prepare the fastq files

Trimming the ends of RNA-seq reads using Cut-adapt.

$ cutadapt -q20 -aGATCGGAAGAGCG –o output-trimmed.fastq.gz input.fastq.gz > log.txt

3.2 Align the RNA-seq reads

Tophat alignment:

Index the reference file (GRCh38p2) and align according to the indexed file

$ tophat index GRCh38p2.fasta

$ tophat --library-type fr-firststrand -p 8 --no-coverage-search GRCh38p2/ Sample-R1.fastq Sample-R2.fastq -o Sample/

RUM alignment:

Indexed hg19 is downloaded from online RUM library

$ Perl rum_runner align -i /indexes/hg19 --name Sample -o Sample/ --chunks 12 Sample-R1.fastq Sample-R2.fastq

3.3 Cufflinks pipeline

Assemble the RNA reads with the guidance of a gtf file:

$ /cufflinks/cufflinks -p16 –g GRCh38p2.gtf –o /tophat/Sample-accepted_hits.bam

Create an assemblies.txt file with the output of cufflinks:

KC-01/cufflinks/KC-01/transcripts.gtf
KC-02/cufflinks/KC-02/transcripts.gtf
KC-03/cufflinks/KC-03/transcripts.gtf
KC-04/cufflinks/KC-04/transcripts.gtf
KC-05/cufflinks/KC-05/transcripts.gtf
KC-06/cufflinks/KC-06/transcripts.gtf
WT-01/cufflinks/WT-01/transcripts.gtf
WT-02/cufflinks/WT-02/transcripts.gtf
WT-03/cufflinks/WT-03/transcripts.gtf
WT-04/cufflinks/WT-04/transcripts.gtf
WT-05/cufflinks/WT-05/transcripts.gtf

Merge the gtf files of all the samples with the reference gtf file using Cuffmerge:

```
$ cufflinks/cuffmerge -p 16 -o output-merged/ -g GRCh38p2.gtf -s GRCh38p2.fa assemblies.txt
```

Perform the differential expression analysis using Cuffdiff:

```
$ cufflinks/cuffdiff -p16 –o output-cuffdiff/ -L KC,WT merged/merged.gtf -b GRCh38p2.fa -p 16 -L KC,WT -u output-merged/merged.gtf tophat/KC-01/accepted_hits.bam,tophat/KC-02/accepted_hits.bam,/tophat/KC-03/accepted_hits.bam,/tophat/KC-04/accepted_hits.bam,/tophat/KC-05/accepted_hits.bam,/tophat/KC-06/accepted_hits.bam /tophat/WT-01/accepted_hits.bam,/tophat/WT-02/accepted_hits.bam,/tophat/WT-03/accepted_hits.bam,/tophat/WT-04/accepted_hits.bam,/tophat/WT-05/accepted_hits.bam
```

For transcriptome profile generation of the WT anterior human cornea:

Create an assembliesWT.txt file with the output of cufflinks:

```
WT-01/cufflinks/WT-01/transcripts.gtf
WT-02/cufflinks/WT-02/transcripts.gtf
WT-03/cufflinks/WT-03/transcripts.gtf
WT-04/cufflinks/WT-04/transcripts.gtf
WT-05/cufflinks/WT-05/transcripts.gtf
```

Merge the gtf files of all the WT samples with the reference gtf file using Cuffmerge:

```
$ cufflinks/cuffmerge -p 16 -o output-mergedWT/ -g GRCh38p2.gtf -s GRCh38p2.fa assembliesWT.txt
```

Generate normalised reads for all the WT RNA reads:

```
$ cufflinks/cuffnorm -p 16 -L WT28,WT35,WT51,WT67,WT82 -o output-cuffnormWT/ MergedWT/merged.gtf /tophat/WT-01/accepted_hits.bam,/tophat/WT-02/accepted_hits.bam,/tophat/WT-03/accepted_hits.bam,/tophat/WT-04/accepted_hits.bam,/tophat/WT-05/accepted_hits.bam --output-format cuffdiff
```

Visualize using CummeRbund:

```
> library(cummeRbund)
> cuff<-readCufflinks()
> cuff
```

Evaluate the overdispersion in the data:

```
> disp<-dispersionPlot(genes(cuff))
> disp
```

Assess the distributions of FPKM scores across samples:

```
> dens<-csDensity(genes(cuff))
> dens
> densRep<-csDensity(genes(cuff),replicates=T)
> densRep
> dend<-csDendro(genes(cuff))
> dend.rep<-csDendro(genes(cuff),replicates=T)
```

Visualize boxplots:

```
 > b<-csBoxplot(genes(cuff))
> b
> brep<-csBoxplot(genes(cuff),replicates=T)
> brep
```

A matrix of pairwise scatterplots can be drawn using the csScatterMatrix() method.

```
> s<-csScatterMatrix(genes(cuff))
> s
```

Generate volcano plots:

```
> v<-csVolcanoMatrix(genes(cuff))
> v
```

For individual pairwise comparisons:

```
> v<-csVolcano(genes(cuff),"KC","WT")
> v
```

3.4 EdgeR pipeline

Count the reads using the featureCounts option of subread software:

```
$ subread/featureCounts -a GRCh38p2.gtf –T 12 -o Counts.txt –s 2 -p
-B /WT28/tophat_out/accepted_hits.bam tophat/KC-
01/accepted_hits.bam tophat/KC-02/accepted_hits.bam /tophat/KC-
03/accepted_hits.bam /tophat/KC-04/accepted_hits.bam /tophat/KC-
05/accepted_hits.bam /tophat/KC-06/accepted_hits.bam /tophat/WT-
```

01/accepted_hits.bam /tophat/WT-02/accepted_hits.bam /tophat/WT-03/accepted_hits.bam /tophat/WT-04/accepted_hits.bam /tophat/WT-05/accepted_hits.bam

Differential analysis using EdgeR is done on R platform. If using for the first time, packages need to be installed.

Generate a .txt file containing all the conditions “Conditions.txt”:

Sample	Condition	Label
MA18	KC	KC01
MA26	KC	KC02
MA46	KC	KC03
MA47	KC	KC04
MA52	KC	KC05
MA58	KC	KC06
WT28	Control	WT01
WT35	Control	WT02
WT51	Control	WT03
WT67	Control	WT04
WT82	Control	WT05

Load R and download bioconductor, Limma, edgeR and Excel packages:

```
$ R
> source(“http://bioconductor.org/biocLite.R”)
> biocLite(“limma”)
> biocLite(“edgeR”)
> biocLite(“xlsx”)
```

Open edgeR in the terminal:

```
> library(“edgeR”)
> library(“xlsx”)
```

Import “Conditions.txt” file:

```
> targets <- readTargets(“Conditions.txt”)
> targets
```

Import the Table of counts “Counts.txt” generated by Subreads to EdgeR:

```
> rawdata <- read.delim(“Counts.txt”, row.names=1, stringsAsFactors=FALSE)
> head(rawdata)
```

Convert raw data into a DGEList:

```
> y <- DGEList(counts=rawdata[,6:16], group=targets$Condition, genes=data.frame(Length=rawdata[,5]))
```

```
> colnames(y) <- targets$Label
```

Count the number of genes:

```
> dim(y)
```

Add gene annotation using annotation.gtf file:

```
> anno <- read.delim(file=" GRCh38p2.gtf")
> head(anno)
```

Match the ENSG number in the data to the ENSG number, Gene Symbol and Feature info

```
> m <- match(rownames(y), anno[,9])
> y$genes <- anno[m, c(9,13,11)]
> colnames(y$genes) <- c("GeneID", "Symbol", "Feature")
> head(y$genes)
```

Calculate Normalisation factors for each of the samples:

```
>  y <- calcNormFactors(y)
> y$samples
```

Plot an MDS plot to show how consistent the replicates are with each other:

```
> plotMDS(y)
```

Calculate the BCV (square root of common dispersion) for the dataset averaged over all genes:

```
> y <- estimateCommonDisp(y, verbose=TRUE)
```

Estimate gene specific dispersions and plot the estimated dispersions:

```
> y <- estimateTagwiseDisp(y)
> plotBCV(y)
```

Compute exact gene wise tests for differential expression between conditions:

```
> et <- exactTest(y)
```

Check out the top 10 results:

```
> top <- topTags(et)
> top
```

Generate a summary of the number of DE genes in the data:

```
> summary(de <- decideTestsDGE(et))
```

Plot a smear graph to show over DE gene expression:

```
> detags <- rownames(y)[as.logical(de)]
> plotSmear (et, de.tags=detags)
> abline(h=c(-1,1), col="blue")
```

The blue lines indicate a 2-fold change.

Write DE.genes to a table:

> DE.genes <- as.data.frame(topTags(et, n=insert no. of DE genes here+1and -1 genes numbers))

Write to excel:
> library("xlsx")
> write.xlsx(DE.genes, file = "DE.genes.xls")

4. PLINK/SEQ commands

4.1 Prepare the vcf file

Calling variants of the *output-sort.realn.dedup.bam* is done using GATK:

$ java –Xmx30g –jar GenomeAnalysisTK.jar –T **UnifiedGenotyper** –R hg_19.fa –I output1-sort.realn.dedup.bam –I output2-sort.realn.dedup.bam … –I outputn-sort.realn.dedup.bam –D dbsnp.vcf –L intervals.interval_list –nt 8 –stand_call_conf 50.0 –stand_emit_conf 10.0 –dcov 200 –l INFO –A AlleleBalance –A DepthOfCoverage –A FisherStrand –log outputPlinkSeq.variants.log –o outputPlinkSeq.variants.SNP.vcf -mbq 15 -glm SNP

$ java –Xmx30g –jar GenomeAnalysisTK.jar –T **UnifiedGenotyper** –R hg_19.fa –I output1-sort.realn.dedup.bam –I output2-sort.realn.dedup.bam … –I outputn-sort.realn.dedup.bam –D dbsnp.vcf –L intervals.interval_list –nt 8 –stand_call_conf 50.0 –stand_emit_conf 10.0 –dcov 200 –l INFO –A AlleleBalance –A DepthOfCoverage –A FisherStrand –log outputPlinkSeq.variants.log –o outputPlinkSeq.variants.INDEL.vcf -mbq 15 -glm INDEL

$ java –Xmx30g –jar GATK.jar –T **VariantRecalibrator** –R hg_19.fa –input outputPlinkSeq.variants.SNP.vcf resource:hapmap,VCF,known=false,training=true,truth=true,prior=15.0 hapmap.vars.vcf –resource:omni,VCF,known=false,training=true,truth=false,prior=12.0 hg19.omnivars.vcf –resource:dbsnp,vcf,known=true,training=false,truth=false,prior=8.0 hg_19.dbsnp.vcf –an QD –an HaplotypeScore –an MQRankSum –an ReadPosRankSum –an FS –an MQ –-maxGaussians 6 –nt 8 -mode BOTH –log outputPlinkSeq.recal.log –recalFile outputPlinkSeq.recal.recal –tranchesFile outputPlinkSeq.recal.tranches –rscriptFile outputPlinkSeq.plot.R

Apply recalibration:

$ java –Xmx30g –jar GATK.jar –T **ApplyRecalibration** –R hg_19.fa –input outputPlinkSeq.variants.raw.vcf –recalFile outputPlinkSeq.recal.recal –tranchesFile

outputPlinkSeq.recal.tranches –ts_filter_level 99.0 –log outputPlinkSeq.apply_recal.log –o outputPlinkSeq.variants.recal.vcf

Combine variants:

$ java –Xmx30g –jar GATK.jar –T **CombineVariants** –R hg_19.fa --variant outputPlinkSeq.variants.SNP.vcf --variant outputPlinkSeq.variants.INDEL.vcf –o combinedPlinkSeq.vcf

Remove chr from the contig lines:

$ sed s/##contig=\<ID=chr/##contig=\<ID=/ combined.vcf > combined.chrcontig.vcf

$ Bgzip combined.chrcontig.vcf

$ Tabix –p VCF combined.chrcontig.vcf.gz

Liftover variants using the GATK to convert the vcf file from hg19 to b37 to be able to use dbsnp 141.b37:

$ java -Xmx4g -jar GATK.jar -T LiftoverVariants -V combinedPlinkSeq.vcf -o combinedPlinkSeq.converted.vcf -R hg_19.fasta -chain liftover_chains/hg19tob37.chain -dict /b37/human_g1k_v37.dict

$ Bgzip combinedPlinkSeq.converted.vcf

$ Tabix –p combinedPlinkSeq.converted.vcf

Remove all the variants with MAF > 2% in dbSNP141 and below:

$ perl annotateSnps.pl -i combinedPlinkSeq.converted.vcf.gz -o combinedPlinkSeq.converted.vcf.dbSNP2.vcf -f 2 --progress -forks 8 -d /b37/dbSnp141.b37.vcf.gz

Remove all the variants with MAF > 2% in EVS (esp6500si):

$ perl filterOnEvsMaf.pl -i combinedPlinkSeq.converted.vcf.dbSNP2.vcf -o combinedPlinkSeq.converted.vcf.dbSNP2.EVS2.vcf -f 2 -p -forks 8 -d EVS/files/

Annotate VCF file using: TrasnlateVCF.pl (Dave’s)

```
$ Perl translateVCF.pl -i combinedPlinkSeq.converted.vcf.dbSNP2.EVS2.vcf -o combinedPlinkSeq.converted.vcf.dbSNP2.EVS2.annotated.vcf -b hg19 –p
```

Filter out all the non functional variants:

```
$ Perl filter_by_anno.pl combinedPlinkSeq.converted.vcf.dbSNP2.EVS2.annotated.vcf
```

The output file is called PlinkSeq.vcf

4.2 Run PLINK/SEQ

Create project

```
$ pseq EnrichementProject new-project --resources PlinkSeq/hg19
```

Load vcf into project

```
$ pseq EnrichementProject load-vcf -- PlinkSeq.vcf
```

Load phenotype information

```
$ pseq EnrichementProject load-pheno --file EnrichementProject.phe
```

Perform enrichment per gene:

```
$ pseq EnrichementProject assoc --phenotype kc --mask loc.group=refseq > EnrichementProject-Association.txt
```

Perform enrichment per variant:

```
$ pseq EnrichementProject v-assoc --phenotype kc
```

4.3 Post PLINK/SEQ the vcf changes

Open the vcf file on excel in order to check the variants and filter out all the variants with p-value < 0.05 then use it in this step and an input file in order to convert it back to a vcf file:

```
$ perl gwas_to_vcf.pl -i PlinkSeq-Variant.xlsx -o PlinkSeq-Variant.xlsx
```

Add info to the SNPs in the file:

```
$ perl annotateSnps.pl -d hg19/dbSnp141.hg19liftover.vcf.gz -i PlinkSeq-Variant-snp2.vcf -o PlinkSeq-Variant.vcf --progress
```

Annotate functional information by using VEP (Variant Effect Predictor):

```
$ perl variant_effect_predictor.pl --offline --everything --dir variant_effect_predictor_76/ensembl-tools-release-76/scripts/variant_effect_predictor/vep_cache/ --assembly GRCh37 --vcf -i PlinkSeq-Variant-snp2.vcf -o vep.PlinkSeq-Variant-snp2.vcf
```

convert back from .vcf to .xlsx formats:

```
$ perl vcf_to_gwas.pl -i vep.PlinkSeq-Variant-snp2.vcf -v --fields symbol consequence hgvsc hgvsc -canonical
```

5. Coverage Statistics per sample

sample_id	total	mean	%_bases_above_5	%_bases_above_10	%_bases_above_15	%_bases_above_20
10-1	3176830071	63.04	97.7	96	93.3	89.4
2_1	3440255711	68.27	97.7	96.3	94.2	91
2-3	3399997683	67.47	97.8	96.5	94.4	91.3
12_1	2633409626	52.26	92.8	88	82.4	76.4
10-1	4297713470	85.29	97.8	96.5	94.5	91.9
13-3	1708370434	33.9	91.3	86.4	79.9	72.1
2_5	3080542049	61.13	97.8	96.1	93.4	89.4
11_3	3678424737	73	94.2	91.3	87.8	83.8
13-1	1648396247	32.71	91	85.9	79.2	71.2
9-1	3265253057	64.8	97.5	96	93.6	90.1
11_4	3014746101	59.83	93.7	89.8	85.2	80.1
KF54	4924475810	97.73	98	97.4	96.4	95
KF52	4402986990	87.38	97.9	97	95.7	93.9
KF53	4789108194	95.04	98.1	97.4	96.3	94.7
KF50	5541449339	109.97	98.1	97.5	96.7	95.7
KF51	4490698135	89.12	98	97.2	96.1	94.4
6-1	4418144022	87.68	98	97.1	95.8	94
8-1	4519915855	89.7	98	97.1	95.7	93.9
8-2	3609948124	71.64	97.7	96.5	94.6	91.8
6-10	8830126388	175.23	98.2	98	97.5	97.1
6-3	3532484205	70.1	97.7	96.4	94.4	91.5
7-1	3317951079	65.84	97.7	96.1	93.3	89.6
IL	1883409306	37.38	91.5	87.1	81.5	74.6
ASD433	4181485035	82.98	98	97	95.6	93.5
ASD434	3261048630	64.72	97.8	96.4	94.1	90.7

RP79	3530172618	70.06	95.3	93.5	91.4	88.6
B1	1553327374	30.83	89.1	79.3	68.9	58.8
SC1	3409526344	67.66	93.8	90.1	85.7	80.7
SC2	2748579236	54.55	93.4	89.3	84.3	78.7
SC3	3764586050	74.71	94.1	91.2	87.6	83.5
B2	1356163327	26.91	90.5	83.6	74.2	63.3
FL	774774852	15.38	86	68.7	48.8	31.1
YL82	4143528276	82.23	98	97	95.5	93.3
ML81	3276060921	65.01	97.8	96.3	93.8	90.3
AM	4150478090	82.37	97.8	96.7	95	92.7

6. Compound heterozygous full lists of Chapter 3- Section 3.2.4

SC1 Compound heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AAChange.refGene
34	ABI3BP	nonsynonymous	3q12.2	.	0.000199681	rs199998539	0.000455	ABI3BP:NM_015429:exon2:c.G167A:p.R56H
23.1	ABI3BP	nonsynonymous	3q12.2	0.002419	0.00778754	rs141302553	0.005041	ABI3BP:NM_015429:exon27:c.G2305C:p.V769L
25.4	AGAP1	nonsynonymous	2q37.2	.	.	.	0.002459	AGAP1:NM_014914:exon12:c.G1390C:p.D464H
24.2	AGAP1	nonsynonymous	2q37.2	.	.	.	0.0008237	AGAP1:NM_014914:exon12:c.T1402C:p.S468P
23.4	APOB	nonsynonymous	2p24.1	.	.	.	.	APOB:NM_000384:exon26:c.G6313T:p.D2105Y
19.98	APOB	nonsynonymous	2p24.1	0.002615	0.0061901	rs72653095	0.004374	APOB:NM_000384:exon26:c.C8462T:p.P2821L
14.64	APOB	nonsynonymous	2p24.1	.	.	.	0.00008236	APOB:NM_000384:exon26:c.C8719T:p.R2907C
22.6	C2CD2	nonsynonymous	21q22.3	0.000615	0.00339457	rs140022230	0.003286	C2CD2:NM_199050:exon10:c.G937A:p.V313I
17.16	C2CD2	nonsynonymous	21q22.3	.	.	rs201984634	0.0002387	C2CD2:NM_015500:exon1:c.C236G:p.A79G
29.1	C2CD5	nonsynonymous	12p12.1	.	0.000199681	rs142327845	0.000453	C2CD5:NM_001286174:exon8:c.G982T:p.G328W
22.7	C2CD5	nonsynonymous	12p12.1	.	0.000199681	rs147669786	0.000453	C2CD5:NM_001286174:exon8:c.G983T:p.G328V
25.3	COL18A1	nonsynonymous	21q22.3	0.012614	0.0175719	rs61736805	0.007476	COL18A1:NM_030582:exon42:c.G4354T:p.V1452L

13.14	COL18A1	nonsynonymous	21q22.3	.	.	.	.	COL18A1:NM_030582:exon34:c.C3371G:p.P1124R
25.4	CXCR1	nonsynonymous	2q35	0.022682	0.0229633	rs16858808	0.027	CXCR1:NM_000634:exon2:c.C1003T:p.R335C
21.8	CXCR1	nonsynonymous	2q35	.	0.00579073	rs142978743	0.003599	CXCR1:NM_000634:exon2:c.A847T:p.I283F
29.7	DMXL1	frameshift deletion	5q23.1	.	.	.	.	DMXL1:NM_001290321:exon18:c.4256_4260del:p.C1419fs
31	DMXL1	frameshift deletion	5q23.1	.	.	.	.	DMXL1:NM_001290321:exon18:c.4262delC:p.S1421fs
19.39	DMXL1	nonsynonymous	5q23.1	.	.	.	.	DMXL1:NM_001290321:exon11:c.G1327T:p.G443C
24.6	DNAH6	nonsynonymous	2p11.2	.	.	.	0.0001209	DNAH6:NM_001370:exon16:c.C2572T:p.R858C
14.2	DNAH6	nonsynonymous	2p11.2	.	.	.	0.0002803	DNAH6:NM_001370:exon36:c.A5851G:p.I1951V
26.2	E4F1	nonsynonymous	16p13.3	.	.	.	0.00004123	E4F1:NM_001288776:exon10:c.G1469A:p.R490H
21.8	E4F1	nonsynonymous	16p13.3	0.000077	0.0071885	rs145769654	0.002949	E4F1:NM_001288776:exon8:c.G1141A:p.V381I
17.98	ENAH	nonsynonymous	1q42.12	.	.	.	.	ENAH:NM_001008493:exon5:c.C661A:p.R221S
15.18	ENAH	nonsynonymous	1q42.12	.	.	.	.	ENAH:NM_001008493:exon5:c.G651T:p.E217D
31	FAM186B	nonsynonymous	12q13.12	0.001076	0.00259585	rs141974060	0.003089	FAM186B:NM_032130:exon2:c.G321C:p.W107C
24.8	FAM186B	nonsynonymous	12q13.12	0.001153	0.00259585	rs144854212	0.003122	FAM186B:NM_032130:exon7:c.G2556C:p.E852D
15.12	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.G2344C:p.E782Q
14.53	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.A2345C:p.E782A
10.89	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.A2333C:p.Q778P
21.4	FBXO7	nonsynonymous	22q12.3	.	0.00119808	.	0.0004365	FBXO7:NM_012179:exon2:c.A155G:p.Y52C
13.27	FBXO7	nonsynonymous	22q12.3	.	.	.	0.000008236	FBXO7:NM_012179:exon2:c.T260C:p.I87T
29.7	FES	nonsynonymous	15q26.1	.	.	.	0.0002418	FES:NM_001143783:exon14:c.A1778G:p.E593G
28.1	FES	nonsynonymous	15q26.1	.	.	.	0.000008239	FES:NM_001143784:exon2:c.G263A:p.R88Q
17.89	FLG	nonsynonymous	1q21.3	0.01753	0.0249601	rs35621145	0.023	FLG:NM_002016:exon3:c.C9808T:p.R3270C
12.11	FLG	nonsynonymous	1q21.3	0.017607	0.0247604	rs71626704	0.023	FLG:NM_002016:exon3:c.C6045A:p.D2015E
15.29	FNDC1	nonsynonymous	6q25.3	.	.	.	.	FNDC1:NM_032532:exon14:c.G4442C:p.R1481P
13.15	FNDC1	nonsynonymous	6q25.3	.	.	rs200598839	.	FNDC1:NM_032532:exon14:c.C4436G:p.T1479S
22.9	GPRIN1	nonsynonymous	5q35.2	.	.	rs201635586	.	GPRIN1:NM_052899:exon2:c.A698T:p.E233V
19.14	GPRIN1	nonsynonymous	5q35.2	.	.	rs199714570	.	GPRIN1:NM_052899:exon2:c.G707T:p.G236V
13.71	GPRIN1	nonsynonymous	5q35.2	.	.	.	.	GPRIN1:NM_052899:exon2:c.T713C:p.L238S
23.5	HCLS1	nonsynonymous	3q13.33	.	.	.	.	HCLS1:NM_001292041:exon11:c.T1021C:p.Y341H

23.3	HCLS1	nonsynonymous	3q13.33	.	.	.	.	HCLS1:NM_001292041:exon11:c.A1015C:p.N339H
16.09	HCLS1	nonsynonymous	3q13.33	.	.	.	.	HCLS1:NM_001292041:exon11:c.C1020G:p.D340E
15.94	HCLS1	nonsynonymous	3q13.33	.	.	.	.	HCLS1:NM_001292041:exon11:c.A1016C:p.N339T
25.8	HDGFRP2	nonsynonymous	19p13.3	.	.	.	.	HDGFRP2:NM_001001520:exon15:c.G1744C:p.G582R
24.7	HDGFRP2	nonsynonymous	19p13.3	.	.	.	.	HDGFRP2:NM_001001520:exon15:c.G1750A:p.E584K
23.8	HDGFRP2	nonsynonymous	19p13.3	.	.	.	.	HDGFRP2:NM_001001520:exon15:c.G1745A:p.G582E
22	HDGFRP2	nonsynonymous	19p13.3	.	.	.	.	HDGFRP2:NM_001001520:exon15:c.T1739C:p.L580P
20.2	HDGFRP2	nonsynonymous	19p13.3	.	.	.	.	HDGFRP2:NM_001001520:exon15:c.G1741C:p.A581P
18.54	HDGFRP2	nonsynonymous	19p13.3	.	.	rs200976909	.	HDGFRP2:NM_001001520:exon15:c.C1738G:p.L580V
21.3	KMT2D	nonsynonymous	12q13.12	0.003701	0.0259585	rs3741626	0.012	KMT2D:NM_003482:exon31:c.C7144T:p.P2382S
18.39	KMT2D	nonsynonymous	12q13.12	.	0.00319489	.	0.001091	KMT2D:NM_003482:exon24:c.G5477T:p.G1826V
23.7	KNDC1	nonsynonymous	10q26.3	0.020734	0.0131789	rs11101618	0.02	KNDC1:NM_152643:exon4:c.G382T:p.A128S
13.44	KNDC1	nonsynonymous	10q26.3	0.02202	0.0131789	rs11101620	0.022	KNDC1:NM_152643:exon4:c.G467A:p.R156Q
11.15	KNDC1	nonsynonymous	10q26.3	0.021706	0.0131789	rs11101619	0.022	KNDC1:NM_152643:exon4:c.A463G:p.S155G
10.92	KNDC1	nonsynonymous	10q26.3	0.022989	0.0133786	rs41317288	0.024	KNDC1:NM_152643:exon11:c.G1808A:p.C603Y
13.42	KRTAP10-6	nonsynonymous	21q22.3	.	.	.	0.00001712	KRTAP10-6:NM_198688:exon1:c.C149G:p.A50G
12.51	KRTAP10-6	nonsynonymous	21q22.3	.	.	.	.	KRTAP10-6:NM_198688:exon1:c.G148A:p.A50T
12.05	LCE4A	nonsynonymous	1q21.3	.	.	rs118176527	.	LCE4A:NM_178356:exon1:c.A161G:p.E54G
11.85	LCE4A	nonsynonymous	1q21.3	.	.	.	.	LCE4A:NM_178356:exon1:c.G166T:p.G56C
40	LOR	stopgain	1q21.3	.	.	.	.	LOR:NM_000427:exon2:c.G592T:p.G198X
10.87	LOR	nonsynonymous	1q21.3	.	.	.	.	LOR:NM_000427:exon2:c.T565G:p.Y189D
18.57	MAP1S	nonsynonymous	19p13.11	.	.	rs199703392	0.001036	MAP1S:NM_018174:exon5:c.C1405G:p.R469G
14.98	MAP1S	nonsynonymous	19p13.11	.	.	rs201564716	0.0007222	MAP1S:NM_018174:exon5:c.C1400G:p.P467R
24.8	MRGPRD	nonsynonymous	11q13.3	0.003542	0.00399361	rs146025023	0.001161	MRGPRD:NM_198923:exon1:c.C616T:p.R206W
23.2	MRGPRD	nonsynonymous	11q13.3	.	.	.	0.0001977	MRGPRD:NM_198923:exon1:c.G641A:p.R214Q
12.94	MRPL19	nonsynonymous	2p12	.	.	.	0.00007513	MRPL19:NM_014763:exon1:c.T77C:p.L26P
12.58	MRPL19	nonsynonymous	2p12	0.006561	0.00698882	rs41382847	0.006452	MRPL19:NM_014763:exon3:c.G313A:p.V105I
11.71	MUC5B	nonsynonymous	11p15.5	0.023567	0.0117812	rs55657020	0.027	MUC5B:NM_002458:exon41:c.G16306A:p.V5436M
11	MUC5B	nonsynonymous	11p15.5	0.025274	0.0147764	rs56220864	0.021	MUC5B:NM_002458:exon41:c.G16422C:p.E5474D

23.2	NLRX1	nonsynonymous	11q23.3	.	0.0163738	rs200846476	0.00832	NLRX1:NM_001282143:exon7:c.C2035G:p.L679V
12.58	NLRX1	nonsynonymous	11q23.3	.	.	.	0.00004942	NLRX1:NM_001282143:exon6:c.G1129A:p.V377I
11.96	NLRX1	nonsynonymous	11q23.3	.	0.000998403	rs151306288	0.000338	NLRX1:NM_001282143:exon7:c.G2228A:p.R743H
25.2	OBSL1	nonsynonymous	2q35	.	0.000399361	rs369705959	0.0003222	OBSL1:NM_001173408:exon5:c.G1850A:p.R617H
21.6	OBSL1	nonsynonymous	2q35	.	.	.	0.00604	OBSL1:NM_001173431:exon11:c.A3730C:p.T1244P
27	OPLAH	nonsynonymous	8q24.3	0.01469	0.00638978	rs10107332	0.014	OPLAH:NM_017570:exon13:c.G1750T:p.G584C
21.2	OPLAH	nonsynonymous	8q24.3	.	.	.	0.00002496	OPLAH:NM_017570:exon10:c.C1331T:p.P444L
26.6	OR52N1	nonsynonymous	11p15.4	.	.	.	.	OR52N1:NM_001001913:exon1:c.G769C:p.A257P
26.3	OR52N1	nonsynonymous	11p15.4	.	.	rs201810252	.	OR52N1:NM_001001913:exon1:c.C767A:p.P256Q
17.24	OVGP1	nonsynonymous	1p13.2	.	.	.	.	OVGP1:NM_002557:exon11:c.G1594C:p.V532L
13.21	OVGP1	nonsynonymous	1p13.2	.	.	rs74322126	.	OVGP1:NM_002557:exon11:c.A1565C:p.K522T
10.96	OVGP1	nonsynonymous	1p13.2	.	.	.	.	OVGP1:NM_002557:exon11:c.T1587G:p.H529Q
24.8	PCM1	nonsynonymous	8p22	.	.	.	0.00002483	PCM1:NM_006197:exon18:c.A2656G:p.T886A
12.81	PCM1	nonsynonymous	8p22	.	0.0071885	rs373403157	0.003683	PCM1:NM_006197:exon5:c.C467T:p.A156V
25.6	RECQL4	nonsynonymous	8q24.3	.	0.000199681	.	0.0002572	RECQL4:NM_004260:exon9:c.C1561T:p.R521W
24.4	RECQL4	nonsynonymous	8q24.3	0.000159	0.00279553	rs369852601	0.001683	RECQL4:NM_004260:exon10:c.C1625T:p.S542F
25.6	SDK1	nonsynonymous	7p22.2	0.000077	.	rs374147324	0.00005766	SDK1:NM_152744:exon16:c.T2324C:p.I775T
24.7	SDK1	nonsynonymous	7p22.2	0.000242	0.00958466	rs117504728	0.002929	SDK1:NM_152744:exon45:c.C6563A:p.T2188K
23.4	SDK1	nonsynonymous	7p22.2	0.016454	0.0179712	rs61735696	0.018	SDK1:NM_152744:exon21:c.C3080T:p.T1027M
35	SLC5A4	nonsynonymous	22q12.3	0.010303	0.0113818	rs62239049	0.012	SLC5A4:NM_014227:exon9:c.G1007A:p.R336H
31	SLC5A4	nonsynonymous	22q12.3	0.010764	0.0113818	rs62239058	0.012	SLC5A4:NM_014227:exon5:c.G415T:p.E139X
33	SPATA20	nonsynonymous	17q21.33	.	.	.	0.00008776	SPATA20:NM_001258372:exon11:c.C1487A:p.P496H
29.7	SPATA20	nonsynonymous	17q21.33	.	.	.	0.00006033	SPATA20:NM_001258372:exon11:c.C1486A:p.P496T
40	TIAM2	stopgain	6q25.3	0.016767	0.028754	rs116268446	0.014	TIAM2:NM_001010927:exon13:c.T1343A:p.L448X
24.1	TIAM2	nonsynonymous	6q25.3	0.016456	0.028754	rs114239511	0.014	TIAM2:NM_001010927:exon13:c.G1340C:p.S447T
26.8	TMTC2	nonsynonymous	12q21.31	.	.	rs200268500	0.013	TMTC2:NM_152588:exon2:c.C415G:p.R139G
21.7	TMTC2	nonsynonymous	12q21.31	0.000692	0.000798722	rs117020331	0.000906	TMTC2:NM_152588:exon6:c.C1730T:p.T577M
26.5	TTN	nonsynonymous	2q31.2	.	0.000599042	.	0.0002315	TTN:NM_003319:exon177:c.G70223A:p.R23408H
23.8	TTN	nonsynonymous	2q31.2	.	0.00199681	rs369707906	0.0008361	TTN:NM_003319:exon154:c.G49727A:p.R16576H
21	TTN	nonsynonymous	2q31.2	0.000081	0.00199681	rs368914555	0.0008436	TTN:NM_003319:exon154:c.C47839T:p.R15947W
18.85	TTN	nonsynonymous	2q31.2	.	0.00139776	.	0.0003395	TTN:NM_003319:exon154:c.A47354G:p.D15785G

16.45	TTN	nonsynonymous	2q31.2	.	0.00199681	rs375627540	0.000803	TTN:NM_133378:exon72:c.G18068A:p.G6023D
26.3	USP43	nonsynonymous	17p13.1	0.000744	0.000998403	rs191605843	0.001804	USP43:NM_001267576:exon15:c.C3196T:p.R1066C
12.24	USP43	nonsynonymous	17p13.1	.	.	rs75472522	0.008175	USP43:NM_001267576:exon14:c.G2200C:p.G734R
28.9	VWA8	nonsynonymous	13q14.11	.	0.00399361	rs200850191	0.00135	VWA8:NM_015058:exon43:c.A5351C:p.Q1784P
23	VWA8	nonsynonymous	13q14.11	.	0.00399361	rs201654359	0.00135	VWA8:NM_015058:exon43:c.C5350A:p.Q1784K
24.9	ZNF880	nonsynonymous	19q13.41	.	.	rs12975097	0.023	ZNF880:NM_001145434:exon4:c.G1238T:p.C413F
17.36	ZNF880	nonsynonymous	19q13.41	.	.	rs76053634	0.005047	ZNF880:NM_001145434:exon4:c.A1217G:p.Q406R

SC2 Compound heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
21.8	ADCY8	nonsynonymous	8q24.22	0.022085	0.0179712	rs2228949	0.02	ADCY8:NM_001115:exon1:c.G238A:p.A80T
17.85	ADCY8	nonsynonymous	8q24.22	.	.	rs201402635	0.002825	ADCY8:NM_001115:exon3:c.G1120C:p.V374L
25.4	AGAP1	nonsynonymous	2q37.2	.	.	.	0.002459	AGAP1:NM_014914:exon12:c.G1390C:p.D464H
24.2	AGAP1	nonsynonymous	2q37.2	.	.	.	0.0008237	AGAP1:NM_014914:exon12:c.T1402C:p.S468P
24.4	C17orf96	nonsynonymous	17q12	.	0.00119808	.	0.002137	C17orf96:NM_001130677:exon1:c.A326T:p.D109V
20.2	C17orf96	nonsynonymous	17q12	.	0.00119808	.	0.002174	C17orf96:NM_001130677:exon1:c.G318T:p.E106D
25.2	CAPN5	nonsynonymous	11q13.5	.	.	rs201256547	0.005429	CAPN5:NM_004055:exon2:c.T95C:p.F32S
22.1	CAPN5	nonsynonymous	11q13.5	.	0.000399361	.	0.00003295	CAPN5:NM_004055:exon10:c.G1369A:p.V457I
20.6	CD24	frameshift insertion	Yq11.222	.	.	.	.	CD24:NM_001291737:exon1:c.68_69insTTTTA:p.Q23fs
20.3	CD24	nonsynonymous	Yq11.222	.	.	.	0.022	CD24:NM_001291737:exon1:c.C65T:p.T22M
19.96	CD24	stopgain	Yq11.222	.	.	.	.	CD24:NM_001291737:exon1:c.C67T:p.Q23X
14.32	COL6A3	nonsynonymous	2q37.3	0.001153	0.0265575	rs77181645	0.023	COL6A3:NM_057164:exon4:c.G1267T:p.A423S
13.37	COL6A3	nonsynonymous	2q37.3	0.001153	0.0265575	rs113155945	0.023	COL6A3:NM_057164:exon4:c.G1198A:p.A400T
24.1	DENND2A	nonsynonymous	7q34	.	0.00698882	rs374255633	0.004011	DENND2A:NM_015689:exon1:c.C890G:p.P297R
20.8	DENND2A	nonsynonymous	7q34	.	0.00698882	rs200642129	0.003995	DENND2A:NM_015689:exon1:c.C54G:p.S18R
29.1	DIEXF	nonsynonymous	1q32.2	.	0.000599042	.	0.0002553	DIEXF:NM_014388:exon7:c.G1135T:p.D379Y

18.17	DIEXF	nonsynonymous	1q32.2	.	0.000599042	.	0.0002718	DIEXF:NM_014388:exon4:c.G463A:p.E155K
24.4	DMKN	nonsynonymous	19q13.12	.	.	.	0.000008236	DMKN:NM_001126056:exon2:c.T551C:p.F184S
14.43	DMKN	nonsynonymous	19q13.12	.	.	rs148799704	.	DMKN:NM_001126057:exon5:c.G839A:p.G280D
15.16	FAM170A	nonsynonymous	5q23.1	.	0.00259585	rs199769553	0.001464	FAM170A:NM_001163991:exon2:c.C167G:p.S56C
10.53	FAM170A	nonsynonymous	5q23.1	.	0.00259585	rs201150712	0.001464	FAM170A:NM_001163991:exon2:c.C119T:p.A40V
15.71	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.G2346C:p.E782D
15.12	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.G2344C:p.E782Q
10.89	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.A2333C:p.Q778P
15.1	FLG	nonsynonymous	1q21.3	.	0.000199681	.	0.0003295	FLG:NM_002016:exon3:c.G5332T:p.A1778S
12.02	FLG	nonsynonymous	1q21.3	0.001384	0.00199681	rs150251062	0.001911	FLG:NM_002016:exon3:c.G3896A:p.R1299Q
32	GPR124	nonsynonymous	8p11.23	0.000615	.	rs200841231	0.0007335	GPR124:NM_032777:exon9:c.C1204T:p.R402W
24.5	GPR124	nonsynonymous	8p11.23	0.000077	0.00579073	rs200624144	0.002908	GPR124:NM_032777:exon12:c.C1648T:p.P550S
22.9	GPRIN1	nonsynonymous	5q35.2	.	.	rs201635586	.	GPRIN1:NM_052899:exon2:c.A698T:p.E233V
19.14	GPRIN1	nonsynonymous	5q35.2	.	.	rs199714570	.	GPRIN1:NM_052899:exon2:c.G707T:p.G236V
13.71	GPRIN1	nonsynonymous	5q35.2	.	.	.	.	GPRIN1:NM_052899:exon2:c.T713C:p.L238S
18.1	IPP	nonsynonymous	1p34.1	.	0.000199681	.	0.0002224	IPP:NM_001145349:exon4:c.G802C:p.E268Q
11.25	IPP	nonsynonymous	1p34.1	.	0.000199681	.	0.0002224	IPP:NM_001145349:exon4:c.G788A:p.C263Y
15.94	KCNN3	nonsynonymous	1q21.3	.	.	.	0.025	KCNN3:NM_001204087:exon1:c.T197A:p.L66H
11.71	KCNN3	nonsynonymous	1q21.3	.	.	.	0.023	KCNN3:NM_001204087:exon1:c.C191A:p.P64Q
28	KIAA0226	nonsynonymous	3q29	.	.	rs80094312	0.027	KIAA0226:NM_014687:exon13:c.G1954T:p.V652F
24.8	KIAA0226	nonsynonymous	3q29	.	.	.	0.0001574	KIAA0226:NM_014687:exon20:c.C2828G:p.A943G
32	L3MBTL2	nonsynonymous	22q13.2	.	.	.	0.007265	L3MBTL2:NM_031488:exon6:c.G697C:p.A233P
23.9	L3MBTL2	nonsynonymous	22q13.2	.	.	rs200870155	0.026	L3MBTL2:NM_031488:exon8:c.T905G:p.V302G
23	MTUS2	nonsynonymous	13q12.3	0.028401	0.0155751	rs61999321	0.026	MTUS2:NM_001033602:exon5:c.C2870T:p.T957M
16.08	MTUS2	nonsynonymous	13q12.3	0.00008	0.00179712	rs376426921	0.0004836	MTUS2:NM_015233:exon3:c.C316G:p.R106G
31	MYO1E	nonsynonymous	15q22.2	0.001388	0.00179712	rs147579391	0.002413	MYO1E:NM_004998:exon27:c.C3146A:p.P1049H
25	MYO1E	nonsynonymous	15q22.2	0.001157	0.00199681	rs141565214	0.002496	MYO1E:NM_004998:exon7:c.A554G:p.D185G
24.1	MYOCD	nonsynonymous	17p12	.	0.00159744	rs369225309	0.002557	MYOCD:NM_001146312:exon11:c.G2074A:p.D692N

19.18	MYOCD	nonsynonymous	17p12	.	0.00159744	rs374404083	0.002216	MYOCD:NM_153604:exon11:c.A2081G:p.D694G
10.14	MYOCD	nonsynonymous	17p12	.	0.00239617	rs193131989	0.002034	MYOCD:NM_153604:exon13:c.C2376A:p.S792R
32	NEB	nonsynonymous	2q23.3	.	.	.	.	NEB:NM_004543:exon89:c.G13328T:p.S4443I
29.5	NEB	nonsynonymous	2q23.3	.	.	rs370790324	0.000008276	NEB:NM_001164507:exon47:c.G6025C:p.D2009H
22.9	NEB	nonsynonymous	2q23.3	.	.	rs375377794	.	NEB:NM_001164508:exon143:c.A21317G:p.K7106R
24.9	NFKBIZ	splicing	3q12.3	.	.	.	.	NM_001005474:exon11:c.1635+1G>C
23.5	NFKBIZ	nonsynonymous	3q12.3	.	.	.	.	NFKBIZ:NM_031419:exon10:c.C1931G:p.A644G
19.97	NFKBIZ	nonsynonymous	3q12.3	.	.	.	.	NFKBIZ:NM_031419:exon10:c.G1930A:p.A644T
10.79	NFKBIZ	splicing	3q12.3	.	.	.	.	NM_001005474:exon11:c.1635+2->TTAATAACC
23.2	NLRX1	nonsynonymous	11q23.3	.	0.0163738	rs200846476	0.00832	NLRX1:NM_001282143:exon7:c.C2035G:p.L679V
12.58	NLRX1	nonsynonymous	11q23.3	.	.	.	0.00004942	NLRX1:NM_001282143:exon6:c.G1129A:p.V377I
17.24	OVGP1	nonsynonymous	1p13.2	.	.	.	.	OVGP1:NM_002557:exon11:c.G1594C:p.V532L
13.21	OVGP1	nonsynonymous	1p13.2	.	.	rs74322126	.	OVGP1:NM_002557:exon11:c.A1565C:p.K522T
25	P2RY4	nonsynonymous	Xq13.1	.	.	rs202027224	0.002463	P2RY4:NM_002565:exon1:c.G674C:p.R225P
23.5	P2RY4	nonsynonymous	Xq13.1	.	.	.	0.0003464	P2RY4:NM_002565:exon1:c.G677C:p.R226P
17.86	RHOT2	nonsynonymous	16p13.3	.	0.00559105	rs117897697	0.001481	RHOT2:NM_138769:exon13:c.G1060A:p.G354S
12.16	RHOT2	nonframeshift deletion	16p13.3	.	0.00559105	rs145581089	0.001483	RHOT2:NM_138769:exon18:c.1689_1691del:p.563_564del
27.8	RNF213	nonsynonymous	17q25.3	.	0.000199681	rs202143169	0.0002636	RNF213:NM_001256071:exon29:c.G7583A:p.R2528Q
24.6	RNF213	nonsynonymous	17q25.3	0.001	0.000399361	rs141301945	0.000626	RNF213:NM_001256071:exon57:c.C13913T:p.T4638I
19.68	SACS	nonsynonymous	13q12.12	0.000154	0.000798722	rs376680832	0.0008813	SACS:NM_001278055:exon8:c.T5020C:p.C1674R
11.49	SACS	nonsynonymous	13q12.12	0.000461	0.000199681	rs149638449	0.0006095	SACS:NM_001278055:exon6:c.G1444A:p.A482T
23.3	SCMH1	nonsynonymous	1p34.2	.	.	.	.	SCMH1:NM_001172222:exon6:c.T854C:p.I285T
15.8	SCMH1	nonsynonymous	1p34.2	.	0.000998403	.	0.0009637	SCMH1:NM_001172221:exon13:c.A1496G:p.D499G
29	SDK1	nonsynonymous	7p22.2	0.008535	0.0125799	rs34775958	0.00775	SDK1:NM_152744:exon6:c.C950T:p.A317V
13.08	SDK1	nonsynonymous	7p22.2	0.002153	0.00878594	rs117824452	0.006244	SDK1:NM_152744:exon7:c.C1100T:p.P367L
24.4	SLFN5	nonsynonymous	17q12	0.008227	0.0251597	rs117695297	0.017	SLFN5:NM_144975:exon5:c.C1963G:p.R655G
13.19	SLFN5	nonsynonymous	17q12	.	0.00259585	.	0.001787	SLFN5:NM_144975:exon5:c.G2138A:p.R713H
26.9	SORBS1	nonsynonymous	10q24.1	0.000077	.	rs372968425	0.00000823	SORBS1:NM_001034954:exon28:c.G3292T:p.G1098W

							6	
24.4	SORBS1	nonsynonymous	10q24.1	.	0.00159744	.	0.00104	SORBS1:NM_001034954:exon28:c.C3095G:p.S1032C
16.9	SSPO	frameshift deletion	7q36.1	.	.	.	0.0003453	SSPO:NM_198455:exon69:c.9782delG:p.W3261fs
13.01	SSPO	nonsynonymous	7q36.1	0.014379	0.00858626	rs189214453	0.017	SSPO:NM_198455:exon50:c.G7442A:p.R2481Q
13.88	TRAK1	nonframeshift insertion	3p22.1	.	.	rs10634555	.	TRAK1:NM_001265609:exon13:c.1841_1842insGGA:p.T614delinsTE
13.62	TRAK1	nonframeshift insertion	3p22.1	.	.	.	.	TRAK1:NM_001265609:exon13:c.1841_1842insGGAGGA:p.T614delinsTEE
23.1	TTN	nonsynonymous	2q31.2	0.017343	0.00938498	rs72648987	0.016	TTN:NM_133378:exon87:c.C22204T:p.R7402C
19.93	TTN	nonsynonymous	2q31.2	0.015037	0.00738818	rs56307213	0.015	TTN:NM_003319:exon154:c.G52588C:p.D17530H
19	TTN	nonsynonymous	2q31.2	0.02046	0.0105831	rs72650006	0.019	TTN:NM_133378:exon102:c.A26080T:p.T8694S
25.2	UNC13D	nonsynonymous	17q25.1	.	.	.	0.00002471	UNC13D:NM_199242:exon13:c.G1076T:p.R359L
21.6	UNC13D	nonsynonymous	17q25.1	0.016146	0.0155751	rs9904366	0.016	UNC13D:NM_199242:exon3:c.G175A:p.A59T
16.45	YBX3	nonsynonymous	12p13.2	.	0.000199681	.	.	YBX3:NM_001145426:exon1:c.G208C:p.A70P
10.38	YBX3	nonsynonymous	12p13.2	.	.	.	.	YBX3:NM_001145426:exon1:c.C200A:p.T67N

SC3 Compound heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
32	ATG2B	nonsynonymous	14q32.2	0.00025	0.00299521	rs78100888	0.001696	ATG2B:NM_018036:exon2:c.G320A:p.R107H
16.34	ATG2B	nonsynonymous	14q32.2	0.000231	0.000599042	rs146341899	0.0007907	ATG2B:NM_018036:exon30:c.G4397T:p.G1466V
14.12	CASP6	nonsynonymous	4q25	.	.	.	.	CASP6:NM_001226:exon1:c.C22G:p.R8G
13.98	CASP6	nonsynonymous	4q25	.	.	.	.	CASP6:NM_001226:exon1:c.G17A:p.G6E
11.51	CASP6	frameshift insertion	4q25	.	.	.	.	CASP6:NM_001226:exon1:c.16_17insC:p.G6fs
28.6	COG1	nonsynonymous	17q25.1	0.013247	0.00479233	rs142719529	0.011	COG1:NM_018714:exon1:c.G58C:p.A20P
25.5	COG1	nonsynonymous	17q25.1	.	.	rs201454609	0.00256	COG1:NM_018714:exon2:c.G547C:p.A183P

24	DST	nonsynonymous	6p12.1	.	0.000599042	rs200345187	0.0004633	DST:NM_015548:exon43:c.A8784C:p.Q2928H
10.93	DST	nonsynonymous	6p12.1	.	.	.	.	DST:NM_015548:exon84:c.G15424A:p.G5142S
23.8	EDEM3	splicing	1q25.3	0.000879	0.0119808	rs200358232	0.005758	NM_025191:exon18:c.1846-5->T
15.42	EDEM3	nonsynonymous	1q25.3	.	0.0119808	rs193034365	0.005701	EDEM3:NM_025191:exon1:c.A60C:p.R20S
24.2	EPPK1	nonsynonymous	8q24.3	0.013834	0.00738818	rs116925616	0.014	EPPK1:NM_031308:exon2:c.C3196T:p.R1066C
14.12	EPPK1	nonsynonymous	8q24.3	.	.	rs191805316	0.0001246	EPPK1:NM_031308:exon2:c.G6644C:p.R2215P
21	FAM71E2	frameshift insertion	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.2336_2337insCTGGGACCCC:p.P779fs
14.88	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.T2338A:p.W780R
10.89	FAM71E2	nonsynonymous	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.A2333C:p.Q778P
34	FANCM	nonsynonymous	14q21.2	.	.	.	0.00002471	FANCM:NM_020937:exon10:c.G1706A:p.R569H
23.2	FANCM	nonsynonymous	14q21.2	.	.	.	0.00009884	FANCM:NM_020937:exon13:c.A2258G:p.D753G
24.4	GPAM	nonsynonymous	10q25.2	.	.	rs199856746	0.00924	GPAM:NM_001244949:exon15:c.G1469T:p.C490F
23.3	GPAM	frameshift insertion	10q25.2	.	.	.	.	GPAM:NM_001244949:exon15:c.1469_1470insTTTTT:p.C490fs
22.9	GPRIN1	nonsynonymous	5q35.2	.	.	rs201635586	.	GPRIN1:NM_052899:exon2:c.A698T:p.E233V
19.14	GPRIN1	nonsynonymous	5q35.2	.	.	rs199714570	.	GPRIN1:NM_052899:exon2:c.G707T:p.G236V
13.71	GPRIN1	nonsynonymous	5q35.2	.	.	.	.	GPRIN1:NM_052899:exon2:c.T713C:p.L238S
24.7	ITPR1	nonsynonymous	3p26.1	.	0.00119808	.	0.0007689	ITPR1:NM_001168272:exon29:c.G3565A:p.V1189M
15.83	ITPR1	nonsynonymous	3p26.1	.	.	.	0.007936	ITPR1:NM_001168272:exon28:c.G3458A:p.G1153E
24.2	KCNE1L	nonsynonymous	Xq23	.	.	.	0.00008777	KCNE1L:NM_012282:exon1:c.G88T:p.G30C
22.9	KCNE1L	nonsynonymous	Xq23	.	.	rs200894798	0.009642	KCNE1L:NM_012282:exon1:c.G87T:p.L29F
16.51	MUC6	nonsynonymous	11p15.5	.	.	rs75538227	0.003782	MUC6:NM_005961:exon31:c.C5765T:p.T1922I
15.35	MUC6	nonsynonymous	11p15.5	.	.	.	0.0002386	MUC6:NM_005961:exon7:c.G854C:p.G285A
24.9	NFKBIZ	splicing	3q12.3	.	.	.	.	NM_001005474:exon11:c.1635+1G>C,NM_031419:exon10:c.1935+1G>C
23.5	NFKBIZ	nonsynonymous	3q12.3	.	.	.	.	NFKBIZ:NM_031419:exon10:c.C1931G:p.A644G
19.97	NFKBIZ	nonsynonymous	3q12.3	.	.	.	.	NFKBIZ:NM_031419:exon10:c.G1930A:p.A644T
24.2	OR9A2	nonsynonymous	7q34	.	0.000199681	rs56540628	0.00014	OR9A2:NM_001001658:exon1:c.G349A:p.D117N
18.78	OR9A2	nonframeshift deletion	7q34	.	.	.	0.00001647	OR9A2:NM_001001658:exon1:c.70_72del:p.24_24del
25	P2RY4	nonsynonymous	Xq13.1	.	.	rs202027224	0.002463	P2RY4:NM_002565:exon1:c.G674C:p.R225P

23.5	P2RY4	nonsynonymous	Xq13.1	.	.	.	0.0003464	P2RY4:NM_002565:exon1:c.G677C:p.R226P
29.1	SCNN1A	nonsynonymous	12p13.31	0.017453	0.0133786	rs5742912	0.018	SCNN1A:NM_001159576:exon9:c.T1654C:p.W552R
26.9	SCNN1A	nonsynonymous	12p13.31	.	.	rs201553658	0.0009566	SCNN1A:NM_001159576:exon3:c.G1018C:p.A340P
26.8	TMTC2	nonsynonymous	12q21.31	.	.	rs200268500	0.013	TMTC2:NM_152588:exon2:c.C415G:p.R139G
26.4	TMTC2	nonsynonymous	12q21.31	.	.	.	0.001648	TMTC2:NM_152588:exon2:c.T410G:p.V137G
18.07	TTN	nonsynonymous	2q31.2	.	0.000998403	rs141142920	0.0003542	TTN:NM_003319:exon27:c.G6221T:p.R2074L
15.77	TTN	nonsynonymous	2q31.2	.	0.000998403	rs372782502	0.0003641	TTN:NM_003319:exon154:c.T42902C:p.V14301A
14.26	TTN	nonsynonymous	2q31.2	0.006707	0.00439297	rs72648927	0.007397	TTN:NM_133378:exon48:c.C11252G:p.P3751R
24.5	UTP23	nonsynonymous	8q24.11	.	.	.	.	UTP23:NM_032334:exon1:c.T65G:p.V22G
23.9	UTP23	nonsynonymous	8q24.11	.	.	.	0.00001934	UTP23:NM_032334:exon1:c.C67G:p.R23G

F9 Compound heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
24.6	CUL9	nonsynonymous	6p21.1	0.000077	.	rs149800737	0.00002471	CUL9:NM_015089:exon12:c.C2882T:p.A961V
23.9	CUL9	nonsynonymous	6p21.1	.	.	.	.	CUL9:NM_015089:exon12:c.G2881T:p.A961S
17.01	CUL9	nonsynonymous	6p21.1	0.006612	0.00219649	rs145911635	0.006383	CUL9:NM_015089:exon7:c.G1849A:p.E617K
33	DCAF15	frameshift insertion	19p13.12	.	.	.	.	DCAF15:NM_138353:exon9:c.1440dupG:p.E480fs
15.78	DCAF15	nonsynonymous	19p13.12	.	.	.	.	DCAF15:NM_138353:exon9:c.G1438C:p.E480Q
40	DYNC2H1	stopgain	11q22.3	.	.	.	.	DYNC2H1:NM_001080463:exon26:c.C4018T:p.Q1340X
22.7	DYNC2H1	nonsynonymous	11q22.3	0.011624	0.00439297	rs144717489	0.013	DYNC2H1:NM_001377:exon89:c.G12865C:p.G4289R
24.6	HEXDC	nonsynonymous	17q25.3	0.003336	0.00539137	rs140743740	0.001441	HEXDC:NM_173620:exon12:c.G1454A:p.C485Y
22.7	HEXDC	nonsynonymous	17q25.3	0.003179	0.00539137	rs150076334	0.001425	HEXDC:NM_173620:exon12:c.C1526A:p.A509E
36	IGSF10	stopgain	3q25.1	0.002768	0.00279553	rs79363433	0.002627	IGSF10:NM_178822:exon4:c.G4804T:p.E1602X
35	IGSF10	stopgain	3q25.1	.	.	.	.	IGSF10:NM_178822:exon2:c.G220T:p.E74X
25	IGSF10	nonsynonymous	3q25.1	0.007996	0.0121805	rs3732775	0.01	IGSF10:NM_001178145:exon2:c.A1507T:p.R503W
28.9	LAMA5	nonsynonymous	20q13.33	0.000077	.	rs371144423	0.00006636	LAMA5:NM_005560:exon32:c.G4037A:p.C1346Y
13.41	LAMA5	nonsynonymous	20q13.33	.	.	.	.	LAMA5:NM_005560:exon80:c.G10995A:p.M3665I

25.9	LTBP1	nonsynonymous	2p22.3	.	.	.	.	LTBP1:NM_000627:exon20:c.T2692A:p.S898T
24.2	LTBP1	nonsynonymous	2p22.3	0.004844	0.00239617	rs141080282	0.004596	LTBP1:NM_001166264:exon21:c.G2885A:p.R962Q
22.9	MUC16	nonsynonymous	19p13.2	0.021355	0.0131789	rs114676657	0.022	MUC16:NM_024690:exon36:c.G38821A:p.E12941K
19.34	MUC16	nonsynonymous	19p13.2	0.004124	0.00559105	rs114920126	0.002628	MUC16:NM_024690:exon70:c.G41972A:p.R13991Q
28.3	NEB	nonsynonymous	2q23.3	0.000583	0.000199681	rs111517514	0.000447	NEB:NM_004543:exon104:c.T15023C:p.L5008P
20.3	NEB	nonsynonymous	2q23.3	0.013294	0.00878594	rs149025191	0.009322	NEB:NM_004543:exon71:c.G10383C:p.M3461I
24.9	NOD2	nonsynonymous	16q12.1	.	0.000399361	.	0.0001647	NOD2:NM_001293557:exon3:c.C1540T:p.R514W
23.2	NOD2	nonsynonymous	16q12.1	0.000385	0.000199681	rs5743276	0.0003542	NOD2:NM_001293557:exon3:c.C1969T:p.R657W
21	PAQR6	nonsynonymous	1q22	0.007689	0.00539137	rs34514408	0.007333	PAQR6:NM_001272109:exon4:c.G4T:p.A2S
16.77	PAQR6	nonsynonymous	1q22	0.007843	0.00539137	rs62001925	0.007392	PAQR6:NM_001272109:exon4:c.C44T:p.S15L
18.12	PCDHGB1	nonsynonymous	5q31.3	0.029647	0.0235623	rs77250251	0.028	PCDHGB1:NM_018922:exon1:c.A1195G:p.K399E
10.98	PCDHGB1	nonsynonymous	5q31.3	0.029952	0.0235623	rs62378414	0.028	PCDHGB1:NM_018922:exon1:c.G557T:p.S186I
26.1	PCNT	nonsynonymous	21q22.3	0.006075	0.00459265	rs35881595	0.00277	PCNT:NM_006031:exon39:c.T8924C:p.L2975P
23.3	PCNT	nonsynonymous	21q22.3	0.006766	0.00459265	rs35848602	0.002817	PCNT:NM_006031:exon31:c.C6986G:p.P2329R
28.4	POLR3B	nonsynonymous	12q23.3	.	.	.	0.001602	POLR3B:NM_001160708:exon13:c.C940T:p.L314F
25.1	POLR3B	nonsynonymous	12q23.3	.	.	.	0.0003229	POLR3B:NM_001160708:exon13:c.A933T:p.L311F
15.77	RPGR	nonsynonymous	Xp11.4	.	.	.	0.0004837	RPGR:NM_001034853:exon15:c.A2801G:p.E934G
12.23	RPGR	nonsynonymous	Xp11.4	.	.	.	.	RPGR:NM_001034853:exon15:c.G2795A:p.G932E
17.44	SHROOM3	nonsynonymous	4q21.1	0.012225	0.00658946	rs76656494	0.012	SHROOM3:NM_020859:exon5:c.A1691C:p.E564A
13.84	SHROOM3	nonsynonymous	4q21.1	.	.	.	0.00005026	SHROOM3:NM_020859:exon5:c.G2651T:p.R884L
13.36	SHROOM3	nonsynonymous	4q21.1	0.005297	0.00399361	rs147732653	0.006311	SHROOM3:NM_020859:exon5:c.C3035A:p.T1012N
15.83	SPINT2	nonsynonymous	19q13.2	0.001538	0.000798722	rs143567268	0.001631	SPINT2:NM_001166103:exon6:c.G544A:p.G182R
14.54	SPINT2	nonsynonymous	19q13.2	0.024527	0.0177716	rs11548457	0.028	SPINT2:NM_001166103:exon6:c.G427C:p.V143L
16.49	SSH1	nonsynonymous	12q24.11	0.000077	0.00898562	rs117900986	0.00229	SSH1:NM_001161331:exon13:c.A1895G:p.N632S
15.93	SSH1	nonsynonymous	12q24.11	.	.	.	.	SSH1:NM_018984:exon15:c.C1901G:p.A634G
16.09	SSPO	nonsynonymous	7q36.1	0.029224	0.0161741	rs117048984	0.017	SSPO:NM_198455:exon2:c.A230G:p.Y77C
14.51	SSPO	nonsynonymous	7q36.1	0.021425	0.0249601	rs73727627	0.009719	SSPO:NM_198455:exon63:c.C9187T:p.R3063C
12.42	SSPO	nonsynonymous	7q36.1	0.018129	0.0285543	rs58369703	0.012	SSPO:NM_198455:exon91:c.G13139C:p.R4380P
10.31	SSPO	nonsynonymous	7q36.1	0.016326	0.0249601	rs55857423	0.00791	SSPO:NM_198455:exon91:c.G13195A:p.A4399T
24	TAS2R31	nonsynonymous	12p13.2	0.014555	0.014377	rs139069360	0.015	TAS2R31:NM_176885:exon1:c.G843T:p.W281C
21	TAS2R31	nonsynonymous	12p13.2	0.01161	0.0101837	rs116926686	0.014	TAS2R31:NM_176885:exon1:c.A711T:p.L237F

22.9	TTN	nonsynonymous	2q31.2	0.000082	.	rs375422359	0.00001655	TTN:NM_003319:exon154:c.G55303A:p.E18435K
22.2	TTN	nonsynonymous	2q31.2	0.015796	0.0231629	rs55675869	0.028	TTN:NM_003319:exon184:c.G72901A:p.V24301I
15.09	TTN	nonsynonymous	2q31.2	0.015783	0.0223642	rs3813245	0.026	TTN:NM_003319:exon154:c.T49525C:p.Y16509H
15	TTN	nonsynonymous	2q31.2	0.015959	0.0223642	rs3731745	0.026	TTN:NM_003319:exon154:c.T52070C:p.I17357T
14.22	TTN	nonsynonymous	2q31.2	0.015579	0.0223642	rs72648257	0.026	TTN:NM_003319:exon170:c.A67852G:p.S22618G
13.59	TTN	nonsynonymous	2q31.2	0.015659	0.0295527	rs2303832	0.029	TTN:NM_003319:exon105:c.A25928T:p.K8643I
10.65	TTN	nonsynonymous	2q31.2	0.015832	0.0223642	rs3813246	0.026	TTN:NM_003319:exon154:c.A50084G:p.N16695S
10.46	TTN	nonsynonymous	2q31.2	.	.	.	.	TTN:NM_133379:exon46:c.T10975G:p.F3659V
12.58	ZNF469	nonframeshift deletion	16q24.2	.	0.00179712	.	0.003504	ZNF469:NM_001127464:exon2:c.9010_9024del:p.3004_3008del
12.24	ZNF469	nonsynonymous	16q24.2	.	.	.	.	ZNF469:NM_001127464:exon1:c.A3062C:p.E1021A

F11 Compound heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
12.36	CNBD1	nonframeshift deletion	8q21.3	0.013658	0.00519169	rs201174674	0.009674	CNBD1:NM_173538:exon9:c.1116_1118del:p.372_373del
15.72	CNBD1	nonsynonymous	8q21.3	0.021681	0.0215655	rs16894901	0.006598	CNBD1:NM_173538:exon3:c.C205A:p.Q69K
26.2	CYP1A1	nonsynonymous	15q24.1	0.001155	0.000798722	rs4987133	0.001252	CYP1A1:NM_000499:exon3:c.T857C:p.I286T
26.2	CYP1A1	nonsynonymous	15q24.1	0.001155	0.000599042	rs61747605	0.001186	CYP1A1:NM_000499:exon2:c.C712T:p.P238S
24.2	DNAH2	nonsynonymous	17p13.1	0.021913	0.0127796	rs79350244	0.023	DNAH2:NM_020877:exon78:c.A12184C:p.I4062L
19.64	DNAH2	nonsynonymous	17p13.1	0.021759	0.0127796	rs117465420	0.023	DNAH2:NM_020877:exon79:c.A12303T:p.L4101F
15.76	DNAH2	nonsynonymous	17p13.1	0.008304	0.00319489	rs117487916	0.008475	DNAH2:NM_020877:exon37:c.G5986A:p.D1996N
32	EPPK1	nonsynonymous	8q24.3	0.006663	0.00359425	rs140364895	0.006588	EPPK1:NM_031308:exon2:c.C1234T:p.R412W
14.12	EPPK1	nonsynonymous	8q24.3	.	.	rs191805316	0.0001246	EPPK1:NM_031308:exon2:c.G6644C:p.R2215P
24.2	FLG	nonsynonymous	1q21.3	0.017684	0.0249601	rs36006086	0.023	FLG:NM_002016:exon3:c.C5671T:p.R1891W
22.3	FLG	nonsynonymous	1q21.3	0.004306	0.00479233	rs12750571	0.004958	FLG:NM_002016:exon3:c.C4309T:p.R1437C
17.89	FLG	nonsynonymous	1q21.3	0.01753	0.0249601	rs35621145	0.023	FLG:NM_002016:exon3:c.C9808T:p.R3270C
12.76	FLG	nonsynonymous	1q21.3	0.017684	0.0249601	rs12750081	0.023	FLG:NM_002016:exon3:c.C4568T:p.T1523I
12.11	FLG	nonsynonymous	1q21.3	0.017607	0.0247604	rs71626704	0.023	FLG:NM_002016:exon3:c.C6045A:p.D2015E

24.9	NFKBIZ	splicing	3q12.3	.	.	.	.	NM_001005474:exon11:c.1635+1G>C
23.5	NFKBIZ	nonsynonymous	3q12.3	.	.	.	.	NFKBIZ:NM_031419:exon10:c.C1931G:p.A644G
19.97	NFKBIZ	nonsynonymous	3q12.3	.	.	.	.	NFKBIZ:NM_031419:exon10:c.G1930A:p.A644T
29.2	OR8D4	frameshift insertion	11q24.1	0.017891	0.0231629	rs201238608	0.005518	OR8D4:NM_001005197:exon1:c.304dupT:p.L101fs
23	OR8D4	nonsynonymous	11q24.1	0.01669	0.0227636	rs79430449	0.005222	OR8D4:NM_001005197:exon1:c.G716A:p.S239N
21.1	RIMS2	nonsynonymous	8q22.3	0.003423	0.00199681	rs182266368	0.004304	RIMS2:NM_014677:exon15:c.G2479A:p.D827N
16.12	RIMS2	nonsynonymous	8q22.3	.	.	.	.	RIMS2:NM_001282881:exon1:c.T17G:p.L6W
28.4	TCTN3	nonsynonymous	10q24.1	.	.	.	.	TCTN3:NM_001143973:exon2:c.C368G:p.P123R
11.38	TCTN3	nonsynonymous	10q24.1	0.004921	0.00359425	rs141088838	0.007182	TCTN3:NM_015631:exon9:c.G1030A:p.G344R
27.1	USH2A	nonsynonymous	1q41	0.021529	0.00758786	rs41277212	0.022	USH2A:NM_206933:exon35:c.A6713C:p.E2238A
26.5	USH2A	nonsynonymous	1q41	.	.	.	0.000008237	USH2A:NM_007123:exon13:c.G2384A:p.C795Y
27.3	XIRP2	nonsynonymous	2q24.3	0.002507	0.00439297	rs143400009	0.006258	XIRP2:NM_001199144:exon7:c.T748C:p.S250P
23.2	XIRP2	nonsynonymous	2q24.3	0.017556	0.0235623	rs16853333	0.022	XIRP2:NM_001199144:exon7:c.G9464A:p.G3155E

F12 Compound heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
11.73	ABCA13	nonframeshift deletion	7p12.3	0.012644	0.00579073	rs200644776	0.01	ABCA13:NM_152701:exon18:c.7911_7913del:p.2637_2638del
36	AFF1	stopgain	4q21.3	.	.	.	.	AFF1:NM_005935:exon3:c.C1031G:p.S344X
35	ANGEL1	frameshift deletion	14q24.3	.	.	.	.	ANGEL1:NM_015305:exon5:c.1091delT:p.V364fs
16.85	ANGPT1	splicing	8q23.1	.	.	rs149628829	0.007953	NM_001146:exon6:c.809-4->T
37	ARL6IP4	stopgain	12q24.31	.	.	.	0.0008843	ARL6IP4:NM_001002251:exon1:c.G472T:p.E158X
13.3	CBL	nonframeshift insertion	11q23.3	.	0.00559105	.	0.0005325	CBL:NM_005188:exon1:c.105_106insCAC:p.P35delinsPH

20.6	CD24	frameshift insertion	Yq11.222	.	.	.	.	CD24:NM_001291737:exon1:c.68_69insTTTTA:p.Q23fs
36	CEP290	frameshift deletion	12q21.32	.	0.000599042	.	0.0007287	CEP290:NM_025114:exon54:c.7394_7395del:p.E2465fs
12.36	CNBD1	nonframeshift deletion	8q21.3	0.013658	0.00519169	rs201174674	0.009674	CNBD1:NM_173538:exon9:c.1116_1118del:p.372_373del
12.26	CXXC4	nonframeshift insertion	4q24	.	.	.	0.0007379	CXXC4:NM_025212:exon2:c.495_496insGGC:p.S166delinsGS
35	DNAH17	nonsynonymous	17q25.3	.	.	.	.	DNAH17:NM_173628:exon66:c.G10675A:p.E3559K
26.2	DPYSL4	.	10q26.3	0.00137	0.00359425	.	0.0005386	.
21.7	EFNA3	nonframeshift deletion	1q22	0.013709	0.014976	rs199552063	0.017	EFNA3:NM_004952:exon2:c.208_219del:p.70_73del
39	EML6	stopgain	2p16.1	.	.	.	.	EML6:NM_001039753:exon5:c.G531A:p.W177X
16.97	FAM111B	frameshift deletion	11q12.1	0.020451	0.0185703	rs199851144	0.026	FAM111B:NM_001142704:exon2:c.302delT:p.V101fs
23.4	FAM71E2	frameshift insertion	19q13.42	.	.	.	.	FAM71E2:NM_001145402:exon9:c.2336_2337insCTGGGACCCC:p.P779fs
15.82	FAM83D	nonframeshift insertion	20q11.23	.	0.0257588	.	0.014	FAM83D:NM_030919:exon1:c.327_328insGCG:p.G109delinsGA
12.6	FIGNL2	nonframeshift deletion	12q13.13	0.000247	.	.	0.0001734	FIGNL2:NM_001013690:exon2:c.67_69del:p.23_23del
35	GCAT	nonsynonymous	22q13.1	0.007458	0.00439297	rs17856459	0.00864	GCAT:NM_001171690:exon9:c.C1237T:p.R413W
17.1	GTF2F1	nonframeshift deletion	19p13.3	0.005112	.	.	0.0007083	GTF2F1:NM_002096:exon7:c.751_753del:p.251_251del
11.49	GUSB	splicing	7q11.21	.	.	.	0.015	NM_001293104:exon12:c.1220-5T>-
11.18	HORMAD1	splicing	1q21.3	.	.	rs367543855	0.0005871	NM_032132:exon13:c.872-5T>-
10.02	HRC	nonframeshift insertion	19q13.33	.	.	.	.	HRC:NM_002152:exon1:c.781_782insATGATG:p.D261delinsDDD
19.87	KIAA2018	nonframeshift deletion	3q13.2	.	.	.	.	KIAA2018:NM_001009899:exon7:c.4410_4412del:p.1470_1471del
21.2	LOC100996693	nonframeshift deletion	2q35	.	0.00319489	.	0.008202	LOC100996693:NM_001286811:exon2:c.380_385del:p.127_129del
35	LRGUK	nonsynonymous	7q33	0.003844	0.00159744	rs61749957	0.00411	LRGUK:NM_144648:exon10:c.C1196T:p.P399L

35	MAP3K5	nonsynonymous	6q23.3	.	0.000199681	rs114243010	0.00008236	MAP3K5:NM_005923:exon17:c.G2300A:p.R767H
12.67	NCL	nonframeshift deletion	2q37.1	.	0.00539137	.	0.005898	NCL:NM_005381:exon4:c.783_785del:p.261_262del
35	NFASC	nonsynonymous	1q32.1	0.000077	0.000199681	rs184631101	0.00008236	NFASC:NM_001160331:exon11:c.C1378T:p.R460W
15.27	NRG2	nonframeshift deletion	5q31.2	.	.	.	0.00008997	NRG2:NM_001184935:exon8:c.2150_2158del:p.717_720del
35	NYNRIN	stopgain	14q12	.	.	.	.	NYNRIN:NM_025081:exon4:c.C1810T:p.Q604X
35	PCK2	frameshift insertion	14q11.2	.	.	.	.	PCK2:NM_001291556:exon10:c.1468dupA:p.P489fs
37	PGPEP1	stopgain	19p13.11	.	.	.	.	PGPEP1:NM_001300927:exon2:c.C85T:p.Q29X
11.09	PPM1F	nonframeshift deletion	22q11.22	0.007514	.	.	0.002028	PPM1F:NM_014634:exon3:c.312_314del:p.104_105del
12.58	PRUNE2	nonframeshift deletion	9q21.2	0.014067	0.0131789	rs144828766	0.004293	PRUNE2:NM_015225:exon8:c.5871_5882del:p.1957_1961del
35	PYGL	nonsynonymous	14q22.1	.	.	.	.	PYGL:NM_001163940:exon5:c.C625T:p.R209C
22.2	TLDC2	nonframeshift deletion	20q11.23	0.002636	0.00139776	.	0.0008319	TLDC2:NM_080628:exon4:c.417_419del:p.139_140del
26.6	ZC2HC1C	frameshift insertion	14q24.3	0.003254	0.000998403	.	0.002442	ZC2HC1C:NM_024643:exon2:c.983dupG:p.R328fs
10.99	ZIC5	nonframeshift deletion	13q32.3	.	.	.	0.006557	ZIC5:NM_033132:exon1:c.1260_1262del:p.420_421del
35	ZNF445	stopgain	3p21.31	.	.	.	.	ZNF445:NM_181489:exon3:c.C58T:p.R20X
12.58	ZNF469	nonframeshift deletion	16q24.2	.	0.00179712	.	0.003504	ZNF469:NM_001127464:exon2:c.9010_9024del:p.3004_3008del

7. **Heterozygous full lists of Chapter 3- Section 3.2.5** (CADD Phred scores ≥ 30)

AR1F heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
35	CLYBL	stopgain	13q32.3	0.020529	0.0219649	rs41281112	0.027	CLYBL:NM_206808:exon6:c.C775T:p.R259X
34	DOCK5	nonsynonymous	8p21.2	0.000231	0.000798722	rs148483229	0.0001154	DOCK5:NM_024940:exon30:c.G3056A:p.R1019H
30	FERMT3	nonsynonymous	11q13.1	0.000233	0.00638978	rs373999037	0.003478	FERMT3:NM_031471:exon12:c.G1393A:p.E465K

SC1 heterozygous variants full list:

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
40	LOR	stopgain	1q21.3	.	.	.	.	LOR:NM_000427:exon2:c.G592T:p.G198X
40	TIAM2	stopgain	6q25.3	0.016767	0.028754	rs116268446	0.014	TIAM2:NM_001010927:exon13:c.T1343A:p.L448X
39	C15orf65	stopgain	15q21.3	.	.	rs201241267	0.014	C15orf65:NM_001198784:exon2:c.C263A:p.S88X
37	CHIT1	stopgain	1q32.1	.	.	rs201320385	.	CHIT1:NM_001256125:exon9:c.G1016A:p.W339X
36	ZNF790	stopgain	19q13.12	.	.	.	0.00004118	ZNF790:NM_001242800:exon5:c.C1483T:p.R495X
35	ZKSCAN7	frameshift deletion	3p21.31	.	.	.	.	ZKSCAN7:NM_001288590:exon6:c.2129_2130del:p.Q710fs
35	DMXL1	frameshift deletion	5q23.1	.	.	.	.	DMXL1:NM_001290321:exon18:c.4262delC:p.S1421fs
35	TMEM222	nonsynonymous	1p36.11	0.009303	0.0081869	rs150461998	0.009398	TMEM222:NM_032125:exon6:c.C622T:p.R208W
35	SCN10A	nonsynonymous	3p22.2	0.000077	0.00199681	rs369399424	0.0009884	SCN10A:NM_001293307:exon24:c.G4085A:p.R1362Q
35	C3orf67	nonsynonymous	3p14.2	0.000769	0.000599042	rs139831294	0.0001812	C3orf67:NM_198463:exon16:c.G1609A:p.E537K
35	CSRP3	nonsynonymous	11p15.1	.	0.000199681	.	0.0001483	CSRP3:NM_003476:exon7:c.G538T:p.G180C
35	ITGB7	nonsynonymous	12q13.13	.	.	.	0.000008236	ITGB7:NM_000889:exon15:c.C2263T:p.R755C
35	PLEKHH1	nonsynonymous	14q24.1	.	0.000199681	.	0.0001737	PLEKHH1:NM_020715:exon13:c.C1852T:p.R618W
35	ZNF544	stopgain	19q13.43	0.007996	0.00898562	rs79258645	0.015	ZNF544:NM_014480:exon7:c.C1843T:p.R615X
35	TGM2	stopgain	20q11.23	.	.	.	0.0001565	TGM2:NM_004613:exon9:c.G1303T:p.E435X
35	PLA2G3	stopgain	22q12.2	.	.	rs201767666	0.00003295	PLA2G3:NM_015715:exon1:c.C436T:p.R146X
35	SLC5A4	nonsynonymous	22q12.3	0.010303	0.0113818	rs62239049	0.012	SLC5A4:NM_014227:exon9:c.G1007A:p.R336H
34	MMS19	frameshift insertion	10q24.1	.	.	.	.	MMS19:NM_001289403:exon23:c.2248_2249insTCTGGGC:p.P750fs

34	TOP2A	frameshift insertion	17q21.2	.	.	.	.	TOP2A:NM_001067:exon3:c.231dupT:p.V78fs
34	PTPRU	nonsynonymous	1p35.3	.	.	.	0.0003709	PTPRU:NM_001195001:exon27:c.G3917C:p.R1306P
34	ABI3BP	nonsynonymous	3q12.2	.	0.000199681	rs199998539	0.000455	ABI3BP:NM_015429:exon2:c.G167A:p.R56H
34	MRPS30	nonsynonymous	5p12	0.008996	0.00519169	rs35601455	0.011	MRPS30:NM_016640:exon1:c.C305T:p.A102V
34	MYL10	nonsynonymous	7q22.1	0.000923	0.00279553	rs146674429	0.00126	MYL10:NM_138403:exon2:c.C94T:p.R32W
34	NPDC1	nonsynonymous	9q34.3	.	.	.	0.00001946	NPDC1:NM_015392:exon7:c.G736A:p.A246T
34	SPPL3	nonsynonymous	12q24.31	.	.	.	0.00008242	SPPL3:NM_139015:exon7:c.G545C:p.R182P
34	ABCD4	nonsynonymous	14q24.3	.	.	rs77210030	0.0005518	ABCD4:NM_005050:exon13:c.T1373C:p.F458S
34	FUK	nonsynonymous	16q22.1	0.000402	0.000599042	rs186275161	0.0003663	FUK:NM_145059:exon16:c.G1874A:p.R625Q
34	CCL5	nonsynonymous	17q12	.	0.000798722	rs200852773	0.0009801	CCL5:NM_002985:exon3:c.G200A:p.R67Q
34	MRPS7	nonsynonymous	17q25.1	.	.	.	0.00004118	MRPS7:NM_015971:exon5:c.C538G:p.R180G
34	UPB1	nonsynonymous	22q11.23	.	.	.	.	UPB1:NM_016327:exon7:c.C827T:p.A276V
33	PRIMPOL	frameshift deletion	4q35.1	.	.	.	0.0006177	PRIMPOL:NM_001300767:exon13:c.1039_1039del:p.E347fs
33	HSPG2	nonsynonymous	1p36.12	.	.	.	0.00006595	HSPG2:NM_001291860:exon26:c.G3349A:p.G1117S
33	GPX7	nonsynonymous	1p32.3	.	.	.	0.00001647	GPX7:NM_015696:exon2:c.C316T:p.R106C
33	GHRL	stopgain	3p25.3	.	.	.	0.00001647	GHRL:NM_001134941:exon2:c.G50A:p.W17X
33	ABCC5	nonsynonymous	3q27.1	.	.	.	.	ABCC5:NM_005688:exon23:c.G3394A:p.A1132T
33	CPED1	nonsynonymous	7q31.31	.	.	.	.	CPED1:NM_001105533:exon16:c.G2161A:p.G721R
33	KDELC2	nonsynonymous	11q22.3	0.016593	0.00898562	rs74911261	0.016	KDELC2:NM_153705:exon3:c.C431T:p.P144L
33	VPS18	nonsynonymous	15q15.1	.	.	.	.	VPS18:NM_020857:exon4:c.C1162T:p.R388C
33	MYLK3	nonsynonymous	16q11.2	.	.	rs200890698	0.017	MYLK3:NM_182493:exon11:c.C2129T:p.S710F
33	SPATA20	nonsynonymous	17q21.33	.	.	.	0.00008776	SPATA20:NM_001258372:exon11:c.C1487A:p.P496H
33	GPATCH1	nonsynonymous	19q13.11	0.000154	.	rs374963833	0.00006589	GPATCH1:NM_018025:exon16:c.C2242T:p.R748C
33	VPS16	nonsynonymous	20p13	.	0.00139776	rs199753911	0.00159	VPS16:NM_080413:exon16:c.C1495T:p.R499C
32	REV1	nonsynonymous	2q11.2	.	0.00139776	rs199850949	0.0006507	REV1:NM_001037872:exon17:c.T2774C:p.L925P
32	NXPE3	nonsynonymous	3q12.3	0.000231	0.000199681	rs140128269	0.000173	NXPE3:NM_001134456:exon8:c.C1448T:p.T483M
32	WFS1	nonsynonymous	4p16.1	0.004767	0.00379393	rs35932623	0.004789	WFS1:NM_001145853:exon8:c.C2452T:p.R818C
32	RGS7BP	nonsynonymous	5q12.3	.	.	.	0.0004779	RGS7BP:NM_001029875:exon1:c.G32C:p.R11P
32	TMEM194A	nonsynonymous	12q13.3	.	.	.	.	TMEM194A:NM_001130963:exon1:c.G11A:p.G4E

32	ACTR10	nonsynonymous	14q23.1	.	.	rs76669080	0.001371	ACTR10:NM_018477:exon1:c.A35G:p.E12G
32	PIF1	nonsynonymous	15q22.31	.	.	.	0.01	PIF1:NM_001286496:exon4:c.G778C:p.A260P
32	LPIN3	nonsynonymous	20q12	0.029064	0.0299521	rs41277016	0.026	LPIN3:NM_001301860:exon2:c.C181T:p.R61W
32	NOL12	nonsynonymous	22q13.1	.	.	rs373277967	0.00003295	NOL12:NM_024313:exon6:c.C583T:p.R195C
32	L3MBTL2	nonsynonymous	22q13.2	.	.	.	0.007265	L3MBTL2:NM_031488:exon6:c.G697C:p.A233P
31	NUP210	nonsynonymous	3p25.1	.	.	.	0.000008243	NUP210:NM_024923:exon2:c.C178T:p.R60W
31	PPID	nonsynonymous	4q32.1	0.02891	0.0277556	rs2070631	0.026	PPID:NM_005038:exon2:c.C145T:p.R49C
31	KIAA1244	nonsynonymous	6q23.3	.	.	rs372977543	0.011	KIAA1244:NM_020340:exon34:c.T6506G:p.V2169G
31	CNTRL	nonsynonymous	9q33.2	0.000231	0.00119808	rs138015820	0.001046	CNTRL:NM_007018:exon7:c.G1069A:p.A357T
31	RELT	nonsynonymous	11q13.4	0.006777	0.0103834	rs141813904	0.009275	RELT:NM_032871:exon8:c.C767T:p.T256M
31	FAM186B	nonsynonymous	12q13.12	0.001076	0.00259585	rs141974060	0.003089	FAM186B:NM_032130:exon2:c.G321C:p.W107C
31	NCOR2	nonsynonymous	12q24.31	.	.	rs112797765	.	NCOR2:NM_001077261:exon38:c.G5488A:p.G1830R
31	FBN1	nonsynonymous	15q21.1	.	.	.	0.000008237	FBN1:NM_000138:exon46:c.C5548T:p.R1850C
31	ITGB4	nonsynonymous	17q25.1	.	.	.	0.000008237	ITGB4:NM_001005619:exon6:c.C670T:p.R224W
31	TMEM259	nonsynonymous	19p13.3	.	.	rs199888865	0.0008705	TMEM259:NM_001033026:exon9:c.G1159C:p.A387P
31	SLC5A4	stopgain	22q12.3	0.010764	0.0113818	rs62239058	0.012	SLC5A4:NM_014227:exon5:c.G415T:p.E139X
30	ZNF76	nonsynonymous	6p21.31	0.012456	0.00638978	rs33959228	0.016	ZNF76:NM_001292032:exon9:c.C814T:p.R272C

SC2 heterozygous variants full list (>20):

CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
39	C15orf65	stopgain	15q21.3	.	.	rs201241267	0.014	C15orf65:NM_001198784:exon2:c.C263A:p.S88X
39	MAP3K10	stopgain	19q13.2	.	.	.	0.0008072	MAP3K10:NM_002446:exon4:c.G1072T:p.E358X
35	TRIM45	frameshift deletion	1p13.1	.	.	.	0.0001071	TRIM45:NM_001145635:exon5:c.1467_1468del:p.I489fs
35	GPAM	frameshift insertion	10q25.2	.	.	.	.	GPAM:NM_001244949:exon15:c.1468_1469insTTTTTT:p.C490fs
35	BBS7	nonsynonymous	4q27	0.000077	0.000798722	rs372685495	0.0004201	BBS7:NM_018190:exon13:c.G1337A:p.R446Q
35	RNF10	nonsynonymous	12q24.31	.	.	.	0.00003295	RNF10:NM_014868:exon5:c.C646T:p.R216C

35	KIAA1683	stopgain	19p13.11	0.000438	0.000399361	rs200745939	0.002004	KIAA1683:NM_001145304:exon3:c.C3004T:p.Q100X
35	CLCN6	nonsynonymous	1p36.22	0.000077	.	rs368875103	0.00004942	CLCN6:NM_001256959:exon10:c.C817T:p.R273C
35	LAMA1	nonsynonymous	18p11.31	0.00123	0.00279553	rs142063208	0.002405	LAMA1:NM_005559:exon50:c.C7141T:p.R2381C
34	PLD4	frameshift deletion	14q32.33	.	0.000798722	.	0.000853	PLD4:NM_138790:exon8:c.937_944del:p.C313fs
34	SPATA20	frameshift insertion	17q21.33	.	.	rs149572077	.	SPATA20:NM_001258372:exon11:c.1488_1489insAAAA:p.P496fs
34	TMEM79	nonsynonymous	1q22	.	0.000399361	rs181268580	0.00002471	TMEM79:NM_032323:exon3:c.C841T:p.R281W
34	LHPP	nonsynonymous	10q26.13	0.000077	0.000199681	rs372122309	0.00002472	LHPP:NM_022126:exon6:c.G664A:p.G222S
34	SPPL3	nonsynonymous	12q24.31	.	.	.	0.00008242	SPPL3:NM_139015:exon7:c.G545C:p.R182P
34	FBLN7	nonsynonymous	2q13	.	0.000199681	rs199980872	0.00003295	FBLN7:NM_001128165:exon7:c.G1034A:p.R345Q
34	CLYBL	nonsynonymous	13q32.3	.	.	.	0.00009884	CLYBL:NM_206808:exon2:c.G217A:p.D73N
34	PAG1	nonsynonymous	8q21.13	.	0.00139776	.	0.001211	PAG1:NM_018440:exon7:c.G796A:p.D266N
34	PDE8B	nonsynonymous	5q13.3	.	0.00239617	.	0.0009307	PDE8B:NM_001029851:exon15:c.G1630A:p.D544N
34	SLC9A5	nonsynonymous	16q22.1	.	.	.	0.013	SLC9A5:NM_004594:exon15:c.G2180C:p.R727P
34	INADL	nonsynonymous	1p31.3	0.011379	0.00499201	rs41289428	0.009843	INADL:NM_176877:exon35:c.G4495A:p.E1499K
34	XPO1	nonsynonymous	2p15	.	.	rs79074863	0.012	XPO1:NM_003400:exon16:c.G1754T:p.C585F
34	CYLD	nonsynonymous	16q12.1	.	.	rs200154154	0.00934	CYLD:NM_001042355:exon17:c.C2597A:p.T866K
34	PLA2G4D	nonsynonymous	15q15.1	0.007613	0.00838658	rs56985825	0.004176	PLA2G4D:NM_178034:exon12:c.C997T:p.R333W
34	GJD4	stopgain	10p11.21	0.000169	0.00339457	rs370286081	0.001934	GJD4:NM_153368:exon2:c.C946T:p.R316X
34	EPHA8	nonsynonymous	1p36.12	0.000077	0.00199681	rs146856523	0.001548	EPHA8:NM_020526:exon15:c.G2636A:p.R879Q
33	NPHP3	nonsynonymous	3q22.1	0.000769	0.00419329	rs142663818	0.001771	NPHP3:NM_153240:exon2:c.C449T:p.A150V
33	SNCAIP	nonsynonymous	5q23.2	0.010918	0.00638978	rs140850272	0.008953	SNCAIP:NM_001242935:exon8:c.G1319A:p.R440H
33	BIVM-ERCC5,ERCC5	nonsynonymous	13q33.1	.	0.000998403	.	0.0002553	ERCC5:NM_000123:exon12:c.C2600T:p.P867L
33	PON3	nonsynonymous	7q21.3	0.002307	0.00159744	rs139856535	0.002792	PON3:NM_000940:exon9:c.G971A:p.G324D
33	RUNX3	nonsynonymous	1p36.11	.	.	.	0.00001647	RUNX3:NM_004350:exon4:c.G590A:p.R197H
33	FGL1	nonsynonymous	8p22	0.011994	0.00938498	rs2653414	0.015	FGL1:NM_004467:exon7:c.G767T:p.W256L
33	CCZ1	nonsynonymous	7p22.1	.	.	.	0.0001079	CCZ1:NM_015622:exon7:c.C533T:p.T178M
32	CACNA1B	frameshift insertion	9q34.3	.	.	.	.	CACNA1B:NM_000718:exon1:c.284_285insTCCATTGAGTAT:p.P95fs

32	NFASC	nonsynonymous	1q32.1	0.005151	0.0061901	rs3795564	0.009406	NFASC:NM_001160331:exon4:c.C458T:p.T153M
32	SERPINI1	nonsynonymous	3q26.1	.	.	.	.	SERPINI1:NM_001122752:exon2:c.C89T:p.S30L
32	CRTC2	nonsynonymous	1q21.3	.	0.00159744	rs199770709	0.001112	CRTC2:NM_181715:exon6:c.G602A:p.R201H
32	NEB	nonsynonymous	2q23.3	.	.	.	.	NEB:NM_004543:exon89:c.G13328T:p.S4443I
32	KPNA7	nonsynonymous	7q22.1	.	0.00219649	rs202200015	0.002697	KPNA7:NM_001145715:exon4:c.C536T:p.A179V
32	SGCE	nonsynonymous	7q21.3	.	0.000798722	.	0.0003625	SGCE:NM_001301139:exon3:c.G287A:p.R96H
32	HEXA	nonsynonymous	15q23	.	.	rs121907956	0.00007413	HEXA:NM_000520:exon13:c.G1496A:p.R499H
32	RGS7BP	nonsynonymous	5q12.3	.	.	.	0.0004779	RGS7BP:NM_001029875:exon1:c.G32C:p.R11P
32	AMBRA1	nonsynonymous	11p11.2	.	.	.	0.000008236	AMBRA1:NM_001267782:exon7:c.G707A:p.R236Q
32	EPSTI1	nonsynonymous	13q14.11	.	.	.	.	EPSTI1:NM_001002264:exon6:c.G502A:p.E168K
32	MYBL2	nonsynonymous	20q13.12	.	.	rs199623276	0.00485	MYBL2:NM_002466:exon3:c.G172C:p.A58P
32	L3MBTL2	nonsynonymous	22q13.2	.	.	.	0.007265	L3MBTL2:NM_031488:exon6:c.G697C:p.A233P
32	GPR124	nonsynonymous	8p11.23	0.000615	.	rs200841231	0.0007335	GPR124:NM_032777:exon9:c.C1204T:p.R402W
32	DCAF12	nonsynonymous	9p13.3	0.000231	0.00399361	rs149115505	0.002183	DCAF12:NM_015397:exon7:c.C908T:p.A303V
32	SEPSECS	nonsynonymous	4p15.2	.	.	.	.	SEPSECS:NM_016955:exon5:c.C671T:p.T224I
32	SH3BP2	nonsynonymous	4p16.3	0.003152	0.00419329	rs148761331	0.004761	SH3BP2:NM_001122681:exon11:c.C1429T:p.R477V
31	SAMD9L	nonsynonymous	7q21.2	.	0.00279553	.	0.0008484	SAMD9L:NM_152703:exon5:c.G695A:p.R232H
31	TBX18	nonsynonymous	6q14.3	.	0.0139776	rs200742037	0.008105	TBX18:NM_001080508:exon5:c.G868A:p.G290R
31	CCDC125	nonsynonymous	5q13.2	0.000384	0.000599042	rs115866581	0.00008237	CCDC125:NM_001297696:exon5:c.C479T:p.T160M
31	BEAN1	stopgain	16q21	.	.	.	0.00003758	BEAN1:NM_001136106:exon4:c.C247T:p.R83X
31	MYO1E	nonsynonymous	15q22.2	0.001388	0.00179712	rs147579391	0.002413	MYO1E:NM_004998:exon27:c.C3146A:p.P1049H
31	TNS3	nonsynonymous	7p12.3	.	0.000199681	.	0.0001738	TNS3:NM_022748:exon23:c.G3538C:p.E1180Q
31	PIP5K1C	nonsynonymous	19p13.3	.	.	.	0.00001647	PIP5K1C:NM_001195733:exon11:c.G1334A:p.R445Q
31	TMEM259	nonsynonymous	19p13.3	.	.	rs199888865	0.0008705	TMEM259:NM_001033026:exon9:c.G1159C:p.A38P
30	SMAD6	nonsynonymous	15q22.31	0.002465	0.00279553	rs188799901	0.002231	SMAD6:NM_005585:exon1:c.G61A:p.D21N

8. Pubmatrix results

Pubmatrix results of all the heterozygous variants occurring in at least two families belonging to 2 different ethnicities

PubMatrix	Keratoconus	Cornea	Oxidative Stress	Collagen	Eye	Inflammation	Proteinase	Apoptosis	Epithelial cells	Keratocytes
ABCC4	0	1	9	7	11	20	4	24	105	0
ABCD4	0	0	0	0	0	0	0	2	6	0
AHNAK	0	0	2	3	1	4	3	4	21	0
AHNAK2	0	0	0	0	0	1	0	0	0	0
ATM	0	12	401	40	106	243	255	1646	884	0
BCLAF1	0	0	0	0	1	1	5	23	12	0
CELA1	0	0	0	0	0	1	3	1	2	0
COL6A3	0	2	1	102	1	6	5	6	6	0
CROCC	0	0	0	0	7	0	0	0	6	0
DNAH1	0	0	0	1	0	0	0	0	0	0
DNAH10	0	0	0	0	0	0	0	0	0	0
DNAH17	0	0	0	0	0	0	0	0	0	0
DNAJB7	0	0	0	0	0	1	0	0	0	0
DSPP	0	0	1	86	2	9	34	12	44	0
DST	2	23	48	508	80	72	42	34	178	0
EPPK1	0	0	0	0	1	1	0	0	0	0
FAM186B	0	0	0	0	0	0	0	0	0	0
FAM188B	0	0	0	0	0	0	0	0	0	0
FAM71E2	0	0	0	0	0	0	0	0	0	0
FES	2	3	35	12	51	48	42	32	77	0
FKRP	0	0	0	4	51	4	5	4	1	0
FLG	1	6	6	6	14	40	18	15	62	1

FLNB	0	0	0	2	1	0	1	2	2	0
GPAM	0	0	3	0	0	8	1	5	9	0
HCLS1	0	0	0	2	1	2	8	20	9	0
HRCT1	0	0	0	0	0	0	0	0	0	0
IGFN1	0	0	0	0	0	1	0	0	0	0
KIAA1683	0	0	0	0	0	0	0	0	0	0
KNDC1	0	0	0	0	0	0	0	0	1	0
KRTAP5-7	0	0	0	0	0	0	0	0	0	0
LAMA5	0	0	0	3	0	2	1	4	11	0
LAMC1	0	0	1	6	7	2	2	7	12	0
LTBP1	0	0	0	15	5	3	18	4	26	1
MAP1A	0	1	2	1	6	0	5	5	11	0
MUC16	0	37	5	2	56	18	20	17	100	0
MUC17	0	0	0	1	2	3	0	2	17	0
MUC4	0	38	3	7	71	40	21	43	218	0
MUC6	0	1	1	2	10	33	41	7	141	0
NCOR2	0	0	3	0	3	14	10	24	72	0
NEB	0	0	9	9	6	20	15	18	129	0
NFASC	0	0	0	2	6	4	2	0	12	0
NFKBIZ	0	3	0	1	5	31	2	6	28	0
NOM1	0	0	0	2	0	0	5	1	2	0
OBSCN	0	0	0	0	0	0	0	2	3	0
OVGP1	0	0	1	0	0	1	3	2	18	0
P2RY4	0	0	0	0	0	0	0	0	1	0
PCDHGB1	0	0	0	0	0	0	0	0	0	0

PKHD1L1	0	0	0	0	0	0	0	0	0	0
PLEKHA4	0	0	0	0	0	0	0	0	0	0
PRX	0	3	313	8	12	24	25	101	84	1
PSMB10	0	1	1	0	5	7	70	6	10	0
RBMXL3	0	0	0	0	0	0	0	0	0	0
RPGR	0	1	0	0	239	0	0	5	9	0
SACS	0	22	38	102	280	181	68	67	419	2
SHROOM3	0	0	0	1	5	1	0	0	12	0
SPATA20	0	0	0	0	0	0	0	0	1	0
SSPO	0	0	0	0	0	0	0	0	0	0
SVEP1	0	0	0	0	0	3	1	1	1	0
SYNE1	0	1	0	1	4	1	1	3	12	0
TTN	0	0	7	20	2	7	47	14	27	0
UBXN11	0	0	0	0	0	0	1	0	1	0
USH2A	0	0	0	1	60	0	1	0	8	0
XPO4	0	0	0	0	0	0	0	0	4	0
ZNF469	11	23	0	9	16	0	0	0	0	0
ZNF880	0	0	0	0	0	0	0	0	0	0

9. ZNF469 primers

Product size	Forward	Melting Temperature
304	CTCCCCACCCACTACCAG CCAGGAGCAGGCCACAGA	58.27 61.32

227	CATCCCTCTTCCCAGACCTG ACACGTCTTACTGAGCACCT	59.16 58.67
250	GAGAATCCACTGCACATCGG CTGTCTCCTTCTGCCTGGC	58.71 60.08
260	GAAGGGCCAAGCTCCACAT TGACGTGGCTGGGGTCTC	57.89 61.31
250	CAGCCGATCAGAGGAAGGTG TGGAGGGAAGAGCTGGGA	60.18 59.13
269	TGCGCATCATCAAGAAGTCC CAGGATGGGGCTGCTTCC	58.63 60.12
239	CTCCCGACCAAGCCCAAG CCGAAGAGCTGGCTGAGG	59.81 66.67

10. PLINK/SEQ results

Gene enrichment results:

LOCUS	POS	ALIAS	NVAR	TEST	P	ALIAS	I	Symbol	DESC
NM_024011	chr1:1635973..1647893	CDK11A	3	BURDEN	0.0386905	CDK11A	0.00298507	CDK11A	23/60
NM_033529	chr1:1635973..1647893	.	3	BURDEN	0.0511811	.	0.00395257	.	23/60
NM_002923	chr1:192778206..192778269	RGS2	2	BURDEN	0.0356164	RGS2	0.0741758	RGS2	2/0
NM_006085	chr1:220236272..220240666	BPNT1	2	BURDEN	0.00566802	BPNT1	0.0141758	BPNT1	4/0
NM_003679	chr1:241725615..241755283	KMO	4	BURDEN	0.0599078	KMO	0.00925926	KMO	3/1
NM_001004686	chr1:248201939..248202474	OR2L2	4	BURDEN	0.0383481	OR2L2	0.00591716	OR2L2	4/4
NM_014023	chr10:1132232..1151117	WDR37	4	BURDEN	0.0577778	WDR37	0.0178571	WDR37	4/4
NM_022761	chr11:111753245..111753246	C11orf1	2	BURDEN	0.0118343	C11orf1	0.0296108	C11orf1	3/1
NM_003986	chr11:27076985..27076985	BBOX1	1	BURDEN	0.0346667	BBOX1	0.0106952	BBOX1	4/3
NM_001003750	chr11:55861103..55861523	OR8I2	2	BURDEN	0.0351351	OR8I2	0.0704607	OR8I2	2/0
NM_001004458	chr11:57982524..57982635	OR1S1	3	BURDEN	0.0572687	OR1S1	0.0132743	OR1S1	2/1

NM_001005167	chr11:5862797..5862797	OR52E6	1	BURDEN	0.0439189	OR52E6	0.00677966	OR52E6	3/3
NM_024099	chr11:62432404..62434492	C11orf48	9	BURDEN	0.0333333	C11orf48	0.00257069	C11orf48	19/36
NM_001043229	chr11:62433781..62434492	METTL12	6	BURDEN	0.0093147	METTL12	0.00133156	METTL12	19/32
NM_005959	chr11:92702929..92715404	MTNR1B	3	BURDEN	0.0402477	MTNR1B	0.00310559	MTNR1B	3/2
NM_001097643	chr12:11285909..11286841	TAS2R30	7	BURDEN	0.0577778	TAS2R30	0.00892857	TAS2R30	1/10
NM_001759	chr12:4388038..4409107	CCND2	2	BURDEN	0.0277228	CCND2	0.0555556	CCND2	2/0
NM_173086	chr12:52863031..52867520	KRT6C	4	BURDEN	0.0173697	KRT6C	0.00372671	KRT6C	4/2
NM_031290	chr13:52439662..52440202	CCDC70	2	BURDEN	0.0287474	CCDC70	0.0596708	CCDC70	2/0
NM_001004479	chr14:20711524..20711858	OR11H4	3	BURDEN	0.0496183	OR11H4	0.00766284	OR11H4	3/2
NM_001012264	chr14:21502431..21502431	RNASE13	1	BURDEN	0.000649984	RNASE13	9.29E-05	RNASE13	6/1
NM_001193635	chr14:24459428..24464283	DHRS4L2	3	BURDEN	0.0530612	DHRS4L2	0.0204918	DHRS4L2	3/3
NM_003813	chr14:70924643..70925559	ADAM21	6	BURDEN	0.0575221	ADAM21	0.0222222	ADAM21	4/3
NM_003814	chr14:70989377..70990774	ADAM20	3	BURDEN	0.0490566	ADAM20	0.00757576	ADAM20	2/1
NM_001001413	chr15:22742709..22743398	GOLGA6L1	14	BURDEN	0.013766	GOLGA6L1	0.000984252	GOLGA6L1	31/50
NM_024793	chr16:3569909..3580663	CLUAP1	2	BURDEN	0.0368272	CLUAP1	0.0539773	CLUAP1	2/0
NM_002153	chr16:82069137..82131943	HSD17B2	4	BURDEN	0.0245614	HSD17B2	0.00351494	HSD17B2	3/1
NM_033184	chr17:39221816..39221816	KRTAP2-4	1	BURDEN	0.0345745	KRTAP2-4	0.00266667	KRTAP2-4	4/3
NM_033188	chr17:39305619..39305910	KRTAP4-5	9	BURDEN	0.0117155	KRTAP4-5	0.000837521	KRTAP4-5	52/77
NM_145041	chr17:41365129..41368562	TMEM106A	4	BURDEN	0.0401235	TMEM106A	0.0123839	TMEM106A	4/2
NM_014804	chr17:6493855..6524149	KIAA0753	5	BURDEN	0.0572687	KIAA0753	0.00442478	KIAA0753	3/2
NM_001146036	chr18:43328350..43329744	.	2	BURDEN	0.0342105	.	0.0633245	.	2/0
NM_015865	chr18:43328350..43329744	SLC14A1	2	BURDEN	0.0341207	SLC14A1	0.0657895	SLC14A1	2/0
NM_001080821	chr19:12501584..12502493	ZNF799	6	BURDEN	0.0101229	ZNF799	0.00144718	ZNF799	9/8
NM_001145404	chr19:19905458..19916911	ZNF506	4	BURDEN	0.0440678	ZNF506	0.0102041	ZNF506	4/2
NM_001099269	chr19:19905458..19916911	.	4	BURDEN	0.0306346	.	0.0131579	.	4/2
NM_199285	chr19:42814708..42814812	PRR19	2	BURDEN	0.0181818	PRR19	0.00390117	PRR19	3/1
NM_002783	chr19:43430027..43433676	PSG7	4	BURDEN	0.00656969	PSG7	0.0112676	PSG7	4/1
NM_001206650	chr19:43430027..43433676	.	4	BURDEN	0.00645161	.	0.0119871	.	4/1
NM_003445	chr19:44500677..44501296	ZNF155	4	BURDEN	0.0341463	ZNF155	0.00488998	ZNF155	3/1

NM_198089	chr19:44500677..44501296	.	4	BURDEN	0.0254083	.	0.00545455	.	3/1
NM_002257	chr19:51322450..51323581	KLK1	2	BURDEN	0.0398773	KLK1	0.0523077	KLK1	2/0
NM_033341	chr19:53793083..53793083	BIRC8	1	BURDEN	0.0212766	BIRC8	0.0410959	BIRC8	2/0
NM_001004301	chr19:53993629..53995115	ZNF813	5	BURDEN	0.0431894	ZNF813	0.0133333	ZNF813	4/4
NM_018328	chr2:149220153..149247720	MBD5	12	BURDEN	0.01566	MBD5	0.00223964	MBD5	7/7
NM_015702	chr2:150427649..150438708	MMADHC	3	BURDEN	0.0490566	MMADHC	0.00378788	MMADHC	3/2
NM_001127391	chr2:202172256..202216069	.	8	BURDEN	0.0153173	.	0.00219058	.	5/4
NM_139163	chr2:202172256..202216069	ALS2CR12	8	BURDEN	0.0185185	ALS2CR12	0.00264901	ALS2CR12	5/4
NM_018848	chr20:10389422..10393680	.	2	BURDEN	0.0290456	.	0.00623701	.	3/1
NM_170784	chr20:10389422..10393680	MKKS	2	BURDEN	0.0290456	MKKS	0.00623701	MKKS	3/1
NM_020157	chr20:16729640..16730591	OTOR	2	BURDEN	0.0295359	OTOR	0.0486258	OTOR	2/0
NM_001037732	chr20:168558..168664	DEFB128	2	BURDEN	0.0183246	DEFB128	0.00393185	DEFB128	3/1
NM_015511	chr20:34827827..34843656	C20orf4	6	BURDEN	0.0336788	C20orf4	0.00779221	C20orf4	4/3
NM_172201	chr21:35742799..35742947	KCNE2	2	BURDEN	0.00562023	KCNE2	0.0128514	KCNE2	3/0
NM_145641	chr22:36537775..36537775	APOL3	1	BURDEN	0.047619	APOL3	0.00735294	APOL3	4/3
NM_145642	chr22:36537775..36537775	.	1	BURDEN	0.040625	.	0.00940439	.	4/3
NM_145640	chr22:36537775..36556836	.	3	BURDEN	0.0565217	.	0.0131004	.	4/4
NM_173472	chr3:10146243..10146243	C3orf24	1	BURDEN	0.00910865	C3orf24	0.0143229	C3orf24	3/0
NM_001164839	chr3:10146243..10146243	.	1	BURDEN	0.00872274	.	0.015586	.	3/0
NM_001134941	chr3:10331519..10331815	.	2	BURDEN	0.00507063	.	0.0112319	.	3/0
NM_001134946	chr3:10331519..10331815	.	2	BURDEN	0.00483425	.	0.0117444	.	3/0
NM_016362	chr3:10331519..10331815	GHRL	2	BURDEN	0.00920447	GHRL	0.0164474	GHRL	3/0
NM_001134944	chr3:10331519..10331815	.	2	BURDEN	0.0104556	.	0.0171898	.	3/0
NM_001134945	chr3:10331519..10331815	.	2	BURDEN	0.0100937	.	0.017316	.	3/0
NM_001184720	chr3:148711967..148727061	.	4	BURDEN	0.00775623	.	0.000554324	.	4/1
NM_004130	chr3:148711967..148727061	GYG1	4	BURDEN	0.00800457	GYG1	0.000572082	GYG1	4/1
NM_001184721	chr3:148711967..148744773	.	5	BURDEN	0.0169492	.	0.00363636	.	4/2
NM_001178146	chr3:151154509..151155879	.	6	BURDEN	0.013712	.	0.00196078	.	4/2
NM_001178145	chr3:151154509..151156366	.	7	BURDEN	0.0335917	.	0.00259067	.	4/2

NM_018029	chr3:73111624..73111747	EBLN2	2	BURDEN	0.015891	EBLN2	0.00227273	EBLN2	4/2
NM_018309	chr3:99998531..100037987	TBC1D23	3	BURDEN	0.0555556	TBC1D23	0.103004	TBC1D23	3/2
NM_001199198	chr3:99998531..100037987	.	3	BURDEN	0.0485075	.	0.104869	.	3/2
NM_000253	chr4:100503136..100543977	MTTP	6	BURDEN	0.0562771	MTTP	0.00869565	MTTP	8/12
NM_183075	chr4:108852804..108866739	CYP2U1	3	BURDEN	0.0192308	CYP2U1	0.00275103	CYP2U1	3/1
NM_006583	chr4:110754408..110757197	RRH	2	BURDEN	0.0401235	RRH	0.00928793	RRH	3/1
NM_003263	chr4:38799283..38800360	TLR1	3	BURDEN	0.0332542	TLR1	0.0619048	TLR1	2/0
NM_006068	chr4:38828828..38830124	TLR6	3	BURDEN	0.0171149	TLR6	0.00611995	TLR6	5/4
NM_032927	chr4:4239589..4247938	TMEM128	2	BURDEN	0.0100874	TMEM128	0.00201884	TMEM128	4/2
NM_001038603	chr5:68715712..68728852	.	4	BURDEN	0.0290456	.	0.002079	.	3/1
NM_001244734	chr5:68715712..68728852	.	4	BURDEN	0.0290456	.	0.002079	.	3/1
NM_030908	chr6:132022285..132022286	OR2A4	2	BURDEN	0.0494297	OR2A4	0.00381679	OR2A4	7/5
NM_012328	chr7:108212243..108213739	DNAJB9	2	BURDEN	0.0204082	DNAJB9	0.0525547	DNAJB9	2/0
NM_001005281	chr7:143701257..143701257	OR6B1	1	BURDEN	0.046595	OR6B1	0.0791367	OR6B1	2/0
NM_001005480	chr7:143807139..143807598	OR2A2	2	BURDEN	0.025974	OR2A2	0.0557621	OR2A2	2/0
NM_003014	chr7:37947150..37953987	SFRP4	2	BURDEN	0.0577778	SFRP4	0.0178571	SFRP4	3/1
NM_006908	chr7:6438337..6438337	.	1	BURDEN	0.006941	.	0.0143849	.	3/0
NM_018890	chr7:6438337..6438337	RAC1	1	BURDEN	0.00842866	RAC1	0.0156627	RAC1	3/0
NM_052937	chr8:52732961..52733242	PCMTD1	14	BURDEN	0.00427742	PCMTD1	0.000305623	PCMTD1	101/25
NM_184086	chr8:67062103..67062103	.	1	BURDEN	0.0194715	.	0.0389972	.	2/0
NM_033058	chr8:67062103..67062103	.	1	BURDEN	0.0244328	.	0.041958	.	2/0
NM_184085	chr8:67062103..67062103	TRIM55	1	BURDEN	0.0190217	TRIM55	0.0421769	TRIM55	2/0
NM_184087	chr8:67062103..67062103	.	1	BURDEN	0.0189702	.	0.0434193	.	2/0
NM_144650	chr8:67344758..67380507	ADHFE1	8	BURDEN	0.0208644	ADHFE1	0.00149254	ADHFE1	8/9
NM_001080526	chr8:82370643..82371545	FABP9	6	BURDEN	0.0139303	FABP9	0.0059761	FABP9	6/3
NM_014581	chr9:136083528..136083528	OBP2B	1	BURDEN	0.000173531	OBP2B	4.96E-05	OBP2B	10/6
NM_001256693	chr9:36651842..36651842	.	1	BURDEN	0.00619469	.	0.0146082	.	3/0
NM_018465	chr9:5361143..5361758	C9orf46	2	BURDEN	0.0408805	C9orf46	0.00630915	C9orf46	3/2

NM_001098412	chrX:49189238..49189246	GAGE13	4	BURDEN	0.0408805	GAGE13	0.022082	GAGE13	5/3
NM_001127200_dup1	chrX:49189238..49189246	.	4	BURDEN	0.0436242	.	0.023569	.	5/3
NM_007268	chrX:65242157..65253634	VSIG4	6	BURDEN	0.0445205	VSIG4	0.00687285	VSIG4	7/8
NM_001100431	chrX:65242157..65253634	.	6	BURDEN	0.0522088	.	0.00806452	.	7/8
NM_001184831	chrX:65253454..65253634	.	4	BURDEN	0.0445205	.	0.00687285	.	6/7
NM_001184830	chrX:65253454..65253634	.	4	BURDEN	0.0467626	.	0.00722022	.	6/7
NM_000044	chrX:66765155..66766412	AR	10	BURDEN	0.027027	AR	0.00193424	AR	22/42

Variant enrichment results:

VAR	REF	ALT	P	OR	ID	symbol	consequence
chr15:82934933	T	C	2.63408E-06	126.818	.	GOLGA6L18	missense_variant
chr3:195508787	G	T	1.93718E-05	33.3333	.	MUC4	missense_variant
chr2:240981681	G	C	2.82076e-05	66	rs201306820	PRR21	missense_variant
chr9:135946801	C	A	3.64897E-05	477	.	CEL	missense_variant
chr15:20740399	T	C	0.00004567	69.16	.	GOLGA6L6	missense_variant
chr1:12907385	G	A	0.000137135	9.72222	rs150590256	HNRNPCL1	missense_variant
chr4:38828828	G	A	0.000155248	46.3827	rs386599647;rs5743820	TLR6	missense_variant
chr14:106053342	T	A	0.000231739	65	rs202028307	IGHA2	missense_variant
chr15:33192236	G	T	0.000251257	8.4857099999999992	rs75033201	FMN1	missense_variant
chr3:195508790	T	C	0.000277934	9.37037	.	MUC4	missense_variant
chr6:56044835	C	A	0.000409741	37.3951	rs199696799	COL21A1	missense_variant
chr14:19378395	A	G	0.000515808	11	rs201551285	OR11H12	missense_variant
chr3:195508789	C	T	0.000574645	9.85714	.	MUC4	missense_variant
chr19:53384111	A	C	0.000700508	38.3012	.	ZNF320	missense_variant
chr19:53384114	T	A	0.000700508	38.3012	.	ZNF320	missense_variant
chr3:195514825	G	A	0.000839195	40.0968	rs200800552	MUC4	missense_variant

chr3:195514846	A	G	0.000853006	39.2857	rs200225290	MUC4	missense_variant
chr14:21502431	G	A	0.000884237	21.45	rs113995906	RNASE13	missense_variant
chr19:16630967	T	G	0.00142853	7.0540500000000002	.	CHERP	missense_variant
chrX:8433535	A	C	0.00161576	10.1429	rs201965035	VCX3B	missense_variant
chr16:22546163	A	C	0.00185599	99	.	NPIPB5	missense_variant
chr16:22546169	C	G	0.00185599	99	rs370469653	NPIPB5	missense_variant
chrX:49189238	C	T	0.0020202	153	.	GAGE13	missense_variant
chr15:34823738	G	A	0.00202224	42.6	rs200544945	GOLGA8B	missense_variant
chr1:12907358	T	C	0.002179	8.4359	rs74587302	HNRNPCL1	missense_variant
chr1:12907380	C	A	0.002179	8.4359	rs141207681	HNRNPCL1	missense_variant
chr3:37370562	A	G	0.00269321	32.1111	.	GOLGA4	missense_variant
chr3:43389767	G	T	0.00286217	5.6918899999999999	rs73831264	SNRK	missense_variant
chr1:12907400	G	A	0.002878	10.65	rs149302457	HNRNPCL1	missense_variant
chr21:38568265	C	A	0.002878	10.65	rs78583828	TTC3	stop_gained
chr5:54591272	G	T	0.00310225	30.6	rs76212559	DHX29	missense_variant
chrX:117959307	G	A	0.00310225	30.6	rs190906444	ZCCHC12	missense_variant
chr22:19711964	C	T	0.00316581	31.1053	.	GP1BB	stop_gained
chr20:60904915	C	T	0.00323748	30.176500000000001	rs371144423	LAMA5	missense_variant
chr1:12907350	C	T	0.00330257	5.85263	rs2359486	HNRNPCL1	missense_variant
chr11:119181832	C	G	0.00342681	17.439	rs370970803	MCAM	missense_variant
chr12:3129932	C	T	0.00342681	17.439	rs71583736	TEAD4	missense_variant
chr7:100645525	A	C	0.00370918	10.5455	rs140153344	MUC12	missense_variant
chr8:52733003	C	T	0.00454955	4.1666699999999999	rs79195845	PCMTD1	missense_variant
chr17:79414763	C	A	0.0046348	26.7882	rs374971809	RP11-1055B8.7	missense_variant
chr2:62066566	G	T	0.00484355	5.89474	rs200331923	FAM161A	missense_variant
chr2:152404248	A	G	0.00485881	26.3647	rs111517514	NEB	missense_variant
chr4:88036182	C	A	0.00492001	4.21143	rs201122599	AFF1	missense_variant
chr1:201180388	G	A	0.00495237	15.7317	.	IGFN1	missense_variant
chr3:195393086	G	A	0.0052305	35.2545	rs201125791	SDHAP2	splice_donor_variant&nc_tra script_variant

chr5:74450118	T	C	0.00534894	25.5176	.	ANKRD31	missense_variant
chr19:17731515	G	T	0.00534894	25.5176	rs201764059	UNC13A	missense_variant
chr17:41352520	A	C	0.00552985	15.2439	.	NBR1	missense_variant
chr1:206566903	G	A	0.00605368	5.17895	rs2987927	SRGAP2	missense_variant&splice_reg n_variant
chr4:170476887	G	T	0.00618972	14.7561	rs201059497	NEK1	missense_variant
chr4:1388497..1388499	TCA	T	0.00620063	3.8	.	CRIPAK	frameshift_variant&feature_ uncation
chr2:179595427	A	C	0.00652948	23.823499999999999	.	TTN	missense_variant
chr16:23718102	T	G	0.00672439	4.77193	.	ERN2	missense_variant
chr17:16675944	C	A	0.00726744	31.127700000000001	rs201462518	CCDC144A	missense_variant
chr19:36002386	C	T	0.00772493	5.30303	rs117522133	DMKN	missense_variant
chrX:49189241	A	C	0.00824176	37.799999999999997	.	GAGE13	missense_variant
chr7:100807151	G	C	0.00869543	9.0909099999999992	.	VGF	missense_variant
chr16:22546136	T	C	0.00887354	31	.	NPIPB5	missense_variant
chr3:195513413	G	A	0.00903632	4.75789	rs7374593	MUC4	missense_variant
chr6:32549481	C	T	0.00985746	4.12088	.	HLA-DRB1	missense_variant
chr3:195506185	G	A	0.00991377	6.69231	rs149026852	MUC4	missense_variant

11. Candidate genes list

Genes
BANP
CAST
CDH11
COL27A1, COL4A1, COL4A2, COL4A3, COL4A4, COL5A1, COL8A1, COL8A2
DOCK9
FLG
FNDC3B

FOXO1
HGF
hOGG1
IL1A, IL1B, IL1RN
IPO5, SPARC, STK24, LOX
MIR184, ZNF469
MPDZ
MUTYH
NEIL1
NF1B
NUB1
PARP1
POLG
RAB3GAP1
RAD51
RXRA
SLC4A11
VSX1, VSX2, TGFBI, SOD1, ZEB1
XRCC1

Variants in KC candidate genes detected in Leeds cohort families:

Family	CADD	Gene	Exonic Function	cytoBand	esp6500si	1000g2014	snp138	exac03	AA Change
F2	23.8	FLG	nonsynonymous	1q21.3	0.10772	0.28774	rs12405278	0.229	FLG:NM_002016:exon3:c.C5095T:p.R1699C
F2	22.6	FLG	nonsynonymous	1q21.3	0.077095	.	rs75448155	0.191	FLG:NM_002016:exon3:c.G10903A:p.D3635N
F2	22.6	FLG	nonsynonymous	1q21.3	0.139243	0.327676	rs74129461	0.256	FLG:NM_002016:exon3:c.G2263A:p.E755K
F2	22.5	IL1A	nonsynonymous	2q13	0.261648	0.217452	rs17561	0.272	IL1A:NM_000575:exon5:c.G340T:p.A114S

F2	20.7	FLG	nonsynonymous	1q21.3	0.10772	0.287141	rs11204978	0.229	FLG:NM_002016:exon3:c.C4445A:p.S1482Y
F2	20.3	COL4A3	nonsynonymous	2q36.3	0.063256	0.100639	rs34505188	0.091	COL4A3:NM_000091:exon21:c.G1223A:p.R40
F2	17.66	FLG	nonsynonymous	1q21.3	0.118714	0.296725	rs12407807	0.232	FLG:NM_002016:exon3:c.G5051A:p.R1684H
F2	15.37	FLG	nonsynonymous	1q21.3	0.107676	0.28774	rs11586631	0.229	FLG:NM_002016:exon3:c.G4079A:p.R1360H
F2	15.37	FLG	nonsynonymous	1q21.3	0.155774	0.340655	rs41267154	0.26	FLG:NM_002016:exon3:c.G995T:p.G332V
F2	14.66	FLG	nonsynonymous	1q21.3	0.099721	.	rs71625202	0.206	FLG:NM_002016:exon3:c.G7097C:p.S2366T
F2	13.01	FLG	nonsynonymous	1q21.3	0.084081	.	rs11582087	0.188	FLG:NM_002016:exon3:c.A8506C:p.S2836R
F2	11.35	FLG	nonsynonymous	1q21.3	0.10795	0.28754	rs12407748	0.229	FLG:NM_002016:exon3:c.G5672A:p.R1891Q
F2	10.76	FLG	nonsynonymous	1q21.3	0.107873	0.286941	rs77422831	0.229	FLG:NM_002016:exon3:c.G11213A:p.R3738H
F2	10.19	FLG	nonsynonymous	1q21.3	0.107753	0.28754	rs12405241	0.229	FLG:NM_002016:exon3:c.C5414T:p.A1805V
F2	10.05	FLG	nonsynonymous	1q21.3	0.209519	.	rs7522925	0.221	FLG:NM_002016:exon3:c.C6323T:p.A2108V
F2	9.817	COL4A2	synonymous	13q34	0.241984	0.186302	rs4773199	0.198	COL4A2:NM_001846:exon43:c.G4089A:p.A13
F2	8.403	FLG	nonsynonymous	1q21.3	0.331462	0.514177	rs2065955	0.299	FLG:NM_002016:exon3:c.G10307C:p.G3436A
F2	8.221	COL27A1	synonymous	9q32	0.198524	0.234225	rs1687410	0.157	COL27A1:NM_032888:exon43:c.A4083G:p.G1
F2	7.041	FLG	synonymous	1q21.3	0.180231	0.396366	rs57672167	0.269	FLG:NM_002016:exon3:c.G8673T:p.V2891V
F2	6.882	FLG	nonsynonymous	1q21.3	0.107796	0.287141	rs11588170	0.229	FLG:NM_002016:exon3:c.G1330A:p.G444R
F2	6.855	FLG	synonymous	1q21.3	0.251771	0.469649	rs9436066	0.299	FLG:NM_002016:exon3:c.G9645T:p.V3215V
F2	6.267	FLG	synonymous	1q21.3	0.108027	0.28754	rs2065956	0.229	FLG:NM_002016:exon3:c.G10017A:p.Q3339Q
F2	4.804	FLG	nonsynonymous	1q21.3	0.197846	.	rs7546186	0.219	FLG:NM_002016:exon3:c.G6134C:p.S2045T
F2	3.844	FLG	synonymous	1q21.3	0.057488	.	rs57670307	0.182	FLG:NM_002016:exon3:c.C8343T:p.D2781D
F2	3.616	COL4A3	synonymous	2q36.3	0.068087	0.105232	rs34019152	0.092	COL4A3:NM_000091:exon23:c.G1452A:p.G48
F2	1.414	FLG	nonsynonymous	1q21.3	0.270567	0.464058	rs2011331	0.295	FLG:NM_002016:exon3:c.A1360G:p.T454A
F2	1.143	FLG	nonsynonymous	1q21.3	0.109565	0.290535	rs72697000	0.234	FLG:NM_002016:exon3:c.G10590T:p.R3530S
F2	1.027	FLG	nonsynonymous	1q21.3	0.281154	0.47484	rs3126072	0.297	FLG:NM_002016:exon3:c.G7633A:p.G2545R
F2	0.93	FLG	synonymous	1q21.3	.	0.453474	rs6664985	0.278	FLG:NM_002016:exon3:c.C6990T:p.H2330H
F2	0.819	FLG	nonsynonymous	1q21.3	0.198601	.	rs7512857	0.218	FLG:NM_002016:exon3:c.T6058G:p.S2020A
F2	0.384	FLG	nonsynonymous	1q21.3	0.089428	.	rs80221306	0.208	FLG:NM_002016:exon3:c.A8807G:p.D2936G
F2	0.264	FLG	synonymous	1q21.3	0.26984	0.463658	rs2184954	0.294	FLG:NM_002016:exon3:c.T6498C:p.S2166S
F2	0.147	FLG	synonymous	1q21.3	0.333385	0.522564	rs3126067	0.301	FLG:NM_002016:exon3:c.T10473C:p.N3491N

F2	0.138	FLG	synonymous	1q21.3	0.262729	0.473642	rs3091276	0.284	FLG:NM_002016:exon3:c.T10194C:p.S3398S
F2	0.135	FLG	synonymous	1q21.3	0.282177	0.473243	rs3120653	0.298	FLG:NM_002016:exon3:c.T2508C:p.D836D
F2	0.113	FLG	synonymous	1q21.3	0.263611	0.463059	rs66831674	0.294	FLG:NM_002016:exon3:c.T3387C:p.S1129S
F2	0.085	FLG	nonsynonymous	1q21.3	.	.	rs78974621	0.235	FLG:NM_002016:exon3:c.A6626G:p.H2209R
F2	0.046	FLG	synonymous	1q21.3	0.194449	0.327676	rs7512554	0.221	FLG:NM_002016:exon3:c.T6354C:p.H2118H
F2	0.034	FLG	nonsynonymous	1q21.3	0.143626	.	rs66954353	0.237	FLG:NM_002016:exon3:c.A6574C:p.K2192Q
F2	0.025	FLG	synonymous	1q21.3	0.255077	0.447085	rs3126066	0.278	FLG:NM_002016:exon3:c.T10491C:p.D3497D
F2	0.017	FLG	nonsynonymous	1q21.3	0.154475	0.330471	rs71625201	0.258	FLG:NM_002016:exon3:c.G7192C:p.E2398Q
F2	0.016	FLG	synonymous	1q21.3	0.203829	0.367212	rs2338554	0.241	FLG:NM_002016:exon3:c.T6603C:p.D2201D
F2	0.011	FLG	nonsynonymous	1q21.3	.	0.25639	rs7518080	0.193	FLG:NM_002016:exon3:c.G10691A:p.R3564H
F2	0.006	FLG	nonsynonymous	1q21.3	0.15426	0.343051	rs71625200	0.261	FLG:NM_002016:exon3:c.A7330G:p.K2444E
F2	0.006	FLG	nonsynonymous	1q21.3	.	0.539736	rs2184953	0.317	FLG:NM_002016:exon3:c.T6580C:p.Y2194H
F2	0.004	FLG	synonymous	1q21.3	0.156005	0.341054	rs6681433	0.261	FLG:NM_002016:exon3:c.A9966G:p.Q3322Q
F2	0.004	FLG	synonymous	1q21.3	0.172305	0.328075	rs80353812	0.22	FLG:NM_002016:exon3:c.T6072C:p.H2024H
F2	0.002	FLG	nonsynonymous	1q21.3	0.195294	0.327676	rs7512553	0.221	FLG:NM_002016:exon3:c.T6355C:p.Y2119H
F2	0.002	FLG	nonsynonymous	1q21.3	0.15713	0.343051	rs11581433	0.261	FLG:NM_002016:exon3:c.A4126G:p.R1376G
F2	0.001	FLG	nonsynonymous	1q21.3	0.269799	0.46226	rs3126074	0.294	FLG:NM_002016:exon3:c.C7521G:p.H2507Q
F2	0.001	FLG	nonsynonymous	1q21.3	0.15762	0.342452	rs55650366	0.261	FLG:NM_002016:exon3:c.T7442C:p.L2481S
F2	0.001	FLG	nonsynonymous	1q21.3	0.101799	.	rs78179835	0.208	FLG:NM_002016:exon3:c.G6891C:p.E2297D
F2	0.001	FLG	nonsynonymous	1q21.3	.	0.34385	rs74129452	0.237	FLG:NM_002016:exon3:c.A6462C:p.Q2154H
F2	0.001	FLG	nonsynonymous	1q21.3	0.346917	0.538139	rs3126079	0.317	FLG:NM_002016:exon3:c.C5883A:p.H1961Q
F2	0.001	FLG	nonsynonymous	1q21.3	0.310165	0.534145	rs58001094	0.306	FLG:NM_002016:exon3:c.C3500G:p.A1167G
F2	0.001	FLG	nonsynonymous	1q21.3	0.158312	0.344649	rs11584340	0.261	FLG:NM_002016:exon3:c.C1432T:p.P478S
F2	0.001	COL4A3	nonsynonymous	2q36.3	0.073534	0.111422	rs11677877	0.094	COL4A3:NM_000091:exon22:c.A1352G:p.H45
F7	6.039	MUTYH	nonsynonymous	1p34.1	0.254115	0.313498	rs3219489	0.268	MUTYH:NM_001048171:exon12:c.G972C:p.Q
F7	8.811	COL4A2	nonsynonymous	13q34	0.231056	0.190895	rs9583500	0.171	COL4A2:NM_001846:exon28:c.C2152T:p.P718
F7	32	XRCC1	nonsynonymous	19q13.31	0.061664	0.123802	rs1799782	0.093	XRCC1:NM_006297:exon6:c.C580T:p.R194W
F8	34	DOCK9	nonsynonymous	13q32.3	0.071235	0.0788738	rs117633128	0.07	DOCK9:NM_001130049:exon1:c.C52T:p.R18W
F8	32	PARP1	nonsynonymous	1q42.12	0.123328	0.196885	rs1136410	0.202	PARP1:NM_001618:exon17:c.T2285C:p.V762A

F8	32	MPDZ	nonsynonymous	9p23	.	.	.	0.00001	MPDZ:NM_001261406:exon22:c.C3203A:p.A1
F8	25.1	COL4A4	nonsynonymous	2q36.3	0.034696	0.0205671	rs1800516	0.028	COL4A4:NM_000092:exon23:c.G1634C:p.G54
F8	24.3	FLG	nonsynonymous	1q21.3	0.027953	.	rs77166567	0.141	FLG:NM_002016:exon3:c.G7015A:p.D2339N
F8	23	COL4A3	nonsynonymous	2q36.3	0.16577	0.115815	rs55703767	0.167	COL4A3:NM_000091:exon17:c.G976T:p.D326
F8	17.73	ZNF469	nonsynonymous	16q24.2	0.055848	0.0435304	rs3812954	0.041	ZNF469:NM_001127464:exon2:c.A8246T:p.D2
F8	16.83	COL5A1	synonymous	9q34.3	0.081578	0.101837	rs41306391	0.044	COL5A1:NM_000093:exon7:c.C1092T:p.P364
F8	15.08	COL4A1	nonsynonymous	13q34	0.000154	0.0001996	rs201964644	0.00024	COL4A1:NM_001845:exon32:c.C2624G:p.P87
F8	12.45	NUB1	synonymous	7q36.1	0.000818	0.0003993	rs182317005	0.00019	NUB1:NM_001243351:exon13:c.C1521T:p.T5
F8	11.22	COL4A2	synonymous	13q34	0.087228	0.100839	rs391859	0.1	COL4A2:NM_001846:exon47:c.G4617A:p.A15
F8	9.907	COL4A4	splicing	2q36.3	0.074102	0.103035	rs3769641	0.087	NM_000092:exon30:c.2384-5T>
F8	6.039	MUTYH	nonsynonymous	1p34.1	0.254115	0.313498	rs3219489	0.268	MUTYH:NM_001048171:exon12:c.G972C:p.Q
F8	3.614	COL8A2	synonymous	1p34.3	0.003537	0.0017971	rs149090218	0.00303	COL8A2:NM_001294347:exon4:c.C1551T:p.T5
F8	3.252	ZNF469	nonsynonymous	16q24.2	.	0.028754	rs3812951	0.016	ZNF469:NM_001127464:exon2:c.G10891A:p. R
F8	0.93	FLG	synonymous	1q21.3	.	0.453474	rs6664985	0.278	FLG:NM_002016:exon3:c.C6990T:p.H2330H
F8	0.085	FLG	nonsynonymous	1q21.3	.	.	rs78974621	0.235	FLG:NM_002016:exon3:c.A6626G:p.H2209R
F8	0.039	COL4A1	synonymous	13q34	0.252499	0.20647	rs995224	0.258	COL4A1:NM_001845:exon21:c.T1257C:p.P419
F9	28.7	MPDZ	nonsynonymous	9p23	0.000165	0.0005990	rs149265684	0.00028	MPDZ:NM_001261406:exon14:c.C1739G:p.S5
F9	24.1	ZNF469	nonframeshift deletion	16q24.2	.	0.0017971	.	0.00350	ZNF469:NM_001127464:exon2:c.9010_9024d 004_3008del
F9	23.8	FLG	nonsynonymous	1q21.3	0.10772	0.28774	rs12405278	0.229	FLG:NM_002016:exon3:c.C5095T:p.R1699C
F9	22.6	FLG	nonsynonymous	1q21.3	0.077095	.	rs75448155	0.191	FLG:NM_002016:exon3:c.G10903A:p.D3635N
F9	22.6	FLG	nonsynonymous	1q21.3	0.139243	0.327676	rs74129461	0.256	FLG:NM_002016:exon3:c.G2263A:p.E755K
F9	22.5	IL1A	nonsynonymous	2q13	0.261648	0.217452	rs17561	0.272	IL1A:NM_000575:exon5:c.G340T:p.A114S
F9	20.7	FLG	nonsynonymous	1q21.3	0.10772	0.287141	rs11204978	0.229	FLG:NM_002016:exon3:c.C4445A:p.S1482Y
F9	20.3	COL4A3	nonsynonymous	2q36.3	0.063256	0.100639	rs34505188	0.091	COL4A3:NM_000091:exon21:c.G1223A:p.R40
F9	19.32	MPDZ	nonsynonymous	9p23	0.029927	0.0135783	rs34605667	0.033	MPDZ:NM_001261407:exon41:c.G5537A:p.R1
F9	17.66	FLG	nonsynonymous	1q21.3	0.118714	0.296725	rs12407807	0.232	FLG:NM_002016:exon3:c.G5051A:p.R1684H
F9	15.37	FLG	nonsynonymous	1q21.3	0.107676	0.28774	rs11586631	0.229	FLG:NM_002016:exon3:c.G4079A:p.R1360H
F9	15.37	FLG	nonsynonymous	1q21.3	0.155774	0.340655	rs41267154	0.26	FLG:NM_002016:exon3:c.G995T:p.G332V

F9	14.66	FLG	nonsynonymous	1q21.3	0.099721	.	rs71625202	0.206	FLG:NM_002016:exon3:c.G7097C:p.S2366T
F9	13.89	COL4A2	synonymous	13q34	0.262849	0.238618	rs76425569	0.231	COL4A2:NM_001846:exon19:c.G1095A:p.P365
F9	13.5	ZNF469	nonsynonymous	16q24.2	.	.	.	.	ZNF469:NM_001127464:exon1:c.A3062C:p.E1
F9	13.01	FLG	nonsynonymous	1q21.3	0.084081	.	rs11582087	0.188	FLG:NM_002016:exon3:c.A8506C:p.S2836R
F9	12.89	COL4A2	synonymous	13q34	0.284764	0.246805	rs74941798	0.295	COL4A2:NM_001846:exon19:c.C1179T:p.I393
F9	12.45	ZNF469	nonsynonymous	16q24.2	.	0.028754	rs3812951	0.016	ZNF469:NM_001127464:exon2:c.G10891A:p.G R
F9	12.41	COL5A1	synonymous	9q34.3	0.425188	0.408147	rs2228560	0.375	COL5A1:NM_000093:exon58:c.G4482C:p.P149
F9	12.24	ZNF469	synonymous	16q24.2	.	.	.	.	ZNF469:NM_001127464:exon1:c.G2808T:p.A9
F9	12.19	LOX	nonsynonymous	5q23.2	0.175339	0.156749	rs1800449	0.17	LOX:NM_002317:exon1:c.G473A:p.R158Q
F9	11.35	FLG	nonsynonymous	1q21.3	0.10795	0.28754	rs12407748	0.229	FLG:NM_002016:exon3:c.G5672A:p.R1891Q
F9	11.22	COL4A2	synonymous	13q34	0.087228	0.100839	rs391859	0.1	COL4A2:NM_001846:exon47:c.G4617A:p.A15
F9	11.04	PARP1	synonymous	1q42.12	0.147701	0.218251	rs1805404	0.208	PARP1:NM_001618:exon2:c.C243T:p.D81D
F9	10.76	FLG	nonsynonymous	1q21.3	0.107873	0.286941	rs77422831	0.229	FLG:NM_002016:exon3:c.G11213A:p.R3738H
F9	10.23	ZNF469	nonsynonymous	16q24.2	.	0.0471246	rs113937803	0.019	ZNF469:NM_001127464:exon1:c.G1285A:p.A4
F9	10.22	FLG	nonsynonymous	1q21.3	0.060298	.	rs2184952	0.085	FLG:NM_002016:exon3:c.G8548A:p.G2850S
F9	10.19	FLG	nonsynonymous	1q21.3	0.107753	0.28754	rs12405241	0.229	FLG:NM_002016:exon3:c.C5414T:p.A1805V
F9	10.05	FLG	nonsynonymous	1q21.3	0.209519	.	rs7522925	0.221	FLG:NM_002016:exon3:c.C6323T:p.A2108V
F9	9.922	ZEB1	synonymous	10p11.22	0.064278	0.0858626	rs7918614	0.032	ZEB1:NM_001128128:exon2:c.C141T:p.D47D
F9	9.85	COL4A2	Splicing	13q34	0.192017	0.190895	rs3803228	0.175	NM_001846:exon32:c.2759-5T>C
F9	9.817	COL4A2	synonymous	13q34	0.241984	0.186302	rs4773199	0.198	COL4A2:NM_001846:exon43:c.G4089A:p.A13
F9	9.066	ZNF469	nonsynonymous	16q24.2	0.025843	0.0365415	rs3812955	0.016	ZNF469:NM_001127464:exon2:c.G8128A:p.A2
F9	8.403	FLG	nonsynonymous	1q21.3	0.331462	0.514177	rs2065955	0.299	FLG:NM_002016:exon3:c.G10307C:p.G3436A
F9	8.221	COL27A1	synonymous	9q32	0.198524	0.234225	rs1687410	0.157	COL27A1:NM_032888:exon43:c.A4083G:p.G1
F9	8.181	ZNF469	synonymous	16q24.2	.	.	.	0.00003	ZNF469:NM_001127464:exon2:c.G4836A:p.A1
F9	8.17	ZNF469	synonymous	16q24.2	0.022996	0.0339457	rs74032865	0.017	ZNF469:NM_001127464:exon2:c.C4350T:p.S1
F9	7.45	ZNF469	synonymous	16q24.2	.	0.0163738	rs117310292	0.011	ZNF469:NM_001127464:exon2:c.G4842A:p.S1
F9	7.041	FLG	synonymous	1q21.3	0.180231	0.396366	rs57672167	0.269	FLG:NM_002016:exon3:c.G8673T:p.V2891V
F9	6.916	IL1RN	synonymous	2q13	0.333154	0.35004	rs315952	0.308	IL1RN:NM_173842:exon4:c.T390C:p.S130S

F9	6.882	FLG	nonsynonymous	1q21.3	0.107796	0.287141	rs11588170	0.229	FLG:NM_002016:exon3:c.G1330A:p.G444R
F9	6.855	FLG	synonymous	1q21.3	0.251771	0.469649	rs9436066	0.299	FLG:NM_002016:exon3:c.G9645T:p.V3215V
F9	6.267	FLG	synonymous	1q21.3	0.108027	0.28754	rs2065956	0.229	FLG:NM_002016:exon3:c.G10017A:p.Q3339Q
F9	4.804	FLG	nonsynonymous	1q21.3	0.197846	.	rs7546186	0.219	FLG:NM_002016:exon3:c.G6134C:p.S2045T
F9	3.616	COL4A3	synonymous	2q36.3	0.068087	0.105232	rs34019152	0.092	COL4A3:NM_000091:exon23:c.G1452A:p.G48
F9	3.252	ZNF469	nonsynonymous	16q24.2	.	.	.	.	ZNF469:NM_001127464:exon2:c.C8342T:p.P2
F9	2.925	ZNF469	synonymous	16q24.2	.	0.113618	rs61472141	0.045	ZNF469:NM_001127464:exon2:c.G6378A:p.P2
F9	1.976	FLG	nonsynonymous	1q21.3	.	0.313498	rs113544881	0.215	FLG:NM_002016:exon3:c.T5828A:p.L1943H
F9	1.414	FLG	nonsynonymous	1q21.3	0.270567	0.464058	rs2011331	0.295	FLG:NM_002016:exon3:c.A1360G:p.T454A
F9	1.296	HGF	synonymous	7q21.11	0.104061	0.103435	rs5745635	0.042	HGF:NM_000601:exon3:c.A333G:p.E111E
F9	1.143	FLG	nonsynonymous	1q21.3	0.109565	0.290535	rs72697000	0.234	FLG:NM_002016:exon3:c.G10590T:p.R3530S
F9	1.027	FLG	nonsynonymous	1q21.3	0.281154	0.47484	rs3126072	0.297	FLG:NM_002016:exon3:c.G7633A:p.G2545R
F9	0.819	FLG	nonsynonymous	1q21.3	0.198601	.	rs7512857	0.218	FLG:NM_002016:exon3:c.T6058G:p.S2020A
F9	0.449	FLG	nonsynonymous	1q21.3	.	.	rs80059102	0.22	FLG:NM_002016:exon3:c.T5839G:p.W1947G
F9	0.268	FLG	nonsynonymous	1q21.3	.	0.206669	rs7540123	0.187	FLG:NM_002016:exon3:c.C10702G:p.Q3568E
F9	0.264	FLG	synonymous	1q21.3	0.26984	0.463658	rs2184954	0.294	FLG:NM_002016:exon3:c.T6498C:p.S2166S
F9	0.147	FLG	synonymous	1q21.3	0.333385	0.522564	rs3126067	0.301	FLG:NM_002016:exon3:c.T10473C:p.N3491N
F9	0.138	FLG	synonymous	1q21.3	0.262729	0.473642	rs3091276	0.284	FLG:NM_002016:exon3:c.T10194C:p.S3398S
F9	0.135	FLG	synonymous	1q21.3	0.282177	0.473243	rs3120653	0.298	FLG:NM_002016:exon3:c.T2508C:p.D836D
F9	0.113	FLG	synonymous	1q21.3	0.263611	0.463059	rs66831674	0.294	FLG:NM_002016:exon3:c.T3387C:p.S1129S
F9	0.097	FLG	synonymous	1q21.3	.	.	rs201652079	0.00017	FLG:NM_002016:exon3:c.T7842C:p.A2614A
F9	0.097	XRCC1	nonsynonymous	19q13.31	0.011379	0.0213658	rs2307177	0.012	XRCC1:NM_006297:exon16:c.A1727C:p.N576T
F9	0.085	FLG	nonsynonymous	1q21.3	.	.	rs78974621	0.235	FLG:NM_002016:exon3:c.A6626G:p.H2209R
F9	0.046	FLG	synonymous	1q21.3	0.194449	0.327676	rs7512554	0.221	FLG:NM_002016:exon3:c.T6354C:p.H2118H
F9	0.039	COL4A1	synonymous	13q34	0.252499	0.20647	rs995224	0.258	COL4A1:NM_001845:exon21:c.T1257C:p.P419
F9	0.034	FLG	nonsynonymous	1q21.3	0.143626	.	rs66954353	0.237	FLG:NM_002016:exon3:c.A6574C:p.K2192Q
F9	0.025	FLG	synonymous	1q21.3	0.255077	0.447085	rs3126066	0.278	FLG:NM_002016:exon3:c.T10491C:p.D3497D
F9	0.017	FLG	nonsynonymous	1q21.3	0.154475	0.330471	rs71625201	0.258	FLG:NM_002016:exon3:c.G7192C:p.E2398Q
F9	0.016	FLG	synonymous	1q21.3	0.203829	0.367212	rs2338554	0.241	FLG:NM_002016:exon3:c.T6603C:p.D2201D

F9	0.011	FLG	nonsynonymous	1q21.3	.	0.25639	rs7518080	0.193	FLG:NM_002016:exon3:c.G10691A:p.R3564H
F9	0.006	FLG	nonsynonymous	1q21.3	0.15426	0.343051	rs71625200	0.261	FLG:NM_002016:exon3:c.A7330G:p.K2444E
F9	0.006	FLG	nonsynonymous	1q21.3	.	0.539736	rs2184953	0.317	FLG:NM_002016:exon3:c.T6580C:p.Y2194H
F9	0.004	FLG	synonymous	1q21.3	0.156005	0.341054	rs6681433	0.261	FLG:NM_002016:exon3:c.A9966G:p.Q3322Q
F9	0.004	FLG	synonymous	1q21.3	0.172305	0.328075	rs80353812	0.22	FLG:NM_002016:exon3:c.T6072C:p.H2024H
F9	0.002	FLG	nonsynonymous	1q21.3	0.195294	0.327676	rs7512553	0.221	FLG:NM_002016:exon3:c.T6355C:p.Y2119H
F9	0.002	FLG	nonsynonymous	1q21.3	0.15713	0.343051	rs11581433	0.261	FLG:NM_002016:exon3:c.A4126G:p.R1376G
F9	0.002	POLG	nonframeshift deletion	15q26.1	.	.	.	0.003212	POLG:NM_001126131:exon2:c.153_158del:p.del
F9	0.001	FLG	nonsynonymous	1q21.3	.	0.206669	rs7532285	0.187	FLG:NM_002016:exon3:c.A10703G:p.Q3568R
F9	0.001	FLG	nonsynonymous	1q21.3	.	.	.	0.000107 1	FLG:NM_002016:exon3:c.C10087T:p.P3363S
F9	0.001	FLG	nonsynonymous	1q21.3	.	0.0027955	rs138143041	0.00016	FLG:NM_002016:exon3:c.T7687C:p.W2563R
F9	0.001	FLG	nonsynonymous	1q21.3	0.269799	0.46226	rs3126074	0.294	FLG:NM_002016:exon3:c.C7521G:p.H2507Q
F9	0.001	FLG	nonsynonymous	1q21.3	0.15762	0.342452	rs55650366	0.261	FLG:NM_002016:exon3:c.T7442C:p.L2481S
F9	0.001	FLG	nonsynonymous	1q21.3	0.101799	.	rs78179835	0.208	FLG:NM_002016:exon3:c.G6891C:p.E2297D
F9	0.001	FLG	nonsynonymous	1q21.3	.	0.34385	rs74129452	0.237	FLG:NM_002016:exon3:c.A6462C:p.Q2154H
F9	0.001	FLG	nonsynonymous	1q21.3	0.346917	0.538139	rs3126079	0.317	FLG:NM_002016:exon3:c.C5883A:p.H1961Q
F9	0.001	FLG	nonsynonymous	1q21.3	0.310165	0.534145	rs58001094	0.306	FLG:NM_002016:exon3:c.C3500G:p.A1167G
F9	0.001	FLG	nonsynonymous	1q21.3	0.158312	0.344649	rs11584340	0.261	FLG:NM_002016:exon3:c.C1432T:p.P478S
F9	0.001	COL4A3	nonsynonymous	2q36.3	0.073534	0.111422	rs11677877	0.094	COL4A3:NM_000091:exon22:c.A1352G:p.H45
F9		XRCC1	nonsynonymous	19q13.31	0.042776	0.0670927	rs25489	0.066	XRCC1:NM_006297:exon9:c.G839A:p.R280H
F9		XRCC1	synonymous	19q13.31	0.42032	0.321286	rs915927	0.335	XRCC1:NM_006297:exon7:c.A618G:p.P206P
F10	28.2	CDH11	nonsynonymous	16q21	0.260111	0.18151	rs35195	0.273	CDH11:NM_001797:exon6:c.C764T:p.T255M
F10	23.8	FLG	nonsynonymous	1q21.3	0.10772	0.28774	rs12405278	0.229	FLG:NM_002016:exon3:c.C5095T:p.R1699C
F10	22.6	FLG	nonsynonymous	1q21.3	0.077095	.	rs75448155	0.191	FLG:NM_002016:exon3:c.G10903A:p.D3635N
F10	22.6	FLG	nonsynonymous	1q21.3	.	.	.	0.000008236	FLG:NM_002016:exon3:c.G3095C:p.R1032T
F10	22.6	FLG	nonsynonymous	1q21.3	0.139243	0.327676	rs74129461	0.256	FLG:NM_002016:exon3:c.G2263A:p.E755K
F10	20.7	FLG	nonsynonymous	1q21.3	0.10772	0.287141	rs11204978	0.229	FLG:NM_002016:exon3:c.C4445A:p.S1482Y

F10	20.3	COL4A3	nonsynonymous	2q36.3	0.063256	0.100639	rs34505188	0.091	COL4A3:NM_000091:exon21:c.G1223A:p.R40
F10	19.86	CDH11	synonymous	16q21	0.322159	0.223642	rs28216	0.33	CDH11:NM_001797:exon7:c.G945A:p.S315S
F10	17.73	ZNF469	nonsynonymous	16q24.2	0.055848	0.0435304	rs3812954	0.041	ZNF469:NM_001127464:exon2:c.A8246T:p.D2
F10	17.66	FLG	nonsynonymous	1q21.3	0.118714	0.296725	rs12407807	0.232	FLG:NM_002016:exon3:c.G5051A:p.R1684H
F10	17.48	RXRA	nonsynonymous	9q34.2	0.000077	.	rs137938878	0.00011	RXRA:NM_001291921:exon2:c.G10A:p.V4I
F10	15.37	FLG	nonsynonymous	1q21.3	0.107676	0.28774	rs11586631	0.229	FLG:NM_002016:exon3:c.G4079A:p.R1360H
F10	15.37	FLG	nonsynonymous	1q21.3	0.155774	0.340655	rs41267154	0.26	FLG:NM_002016:exon3:c.G995T:p.G332V
F10	14.66	FLG	nonsynonymous	1q21.3	0.099721	.	rs71625202	0.206	FLG:NM_002016:exon3:c.G7097C:p.S2366T
F10	14.3	DOCK9	synonymous	13q32.3	0.13827	0.167532	rs12428661	0.181	DOCK9:NM_001130049:exon5:c.C477T:p.V15
F10	13.01	FLG	nonsynonymous	1q21.3	0.084081	.	rs11582087	0.188	FLG:NM_002016:exon3:c.A8506C:p.S2836R
F10	12.72	COL4A2	nonsynonymous	13q34	0.103497	0.103834	rs3803230	0.106	COL4A2:NM_001846:exon27:c.G2048C:p.G68
F10	12.45	XRCC1	nonsynonymous	19q13.31	0.042776	0.0670927	rs25489	0.066	XRCC1:NM_006297:exon9:c.G839A:p.R280H
F10	12.41	COL5A1	synonymous	9q34.3	0.425188	0.408147	rs2228560	0.375	COL5A1:NM_000093:exon58:c.G4482C:p.P14
F10	11.35	FLG	nonsynonymous	1q21.3	0.10795	0.28754	rs12407748	0.229	FLG:NM_002016:exon3:c.G5672A:p.R1891Q
F10	10.76	FLG	nonsynonymous	1q21.3	0.107873	0.286941	rs77422831	0.229	FLG:NM_002016:exon3:c.G11213A:p.R3738H
F10	10.51	DOCK9	Synonymous	13q32.3	0.004942	0.0037939	rs41279128	0.00443	DOCK9:uc001vns.2:exon15:c.C1686T:p.C562C
F10	10.22	FLG	nonsynonymous	1q21.3	0.060298	.	rs2184952	0.085	FLG:NM_002016:exon3:c.G8548A:p.G2850S
F10	10.19	FLG	nonsynonymous	1q21.3	0.107753	0.28754	rs12405241	0.229	FLG:NM_002016:exon3:c.C5414T:p.A1805V
F10	10.05	FLG	nonsynonymous	1q21.3	0.209519	.	rs7522925	0.221	FLG:NM_002016:exon3:c.C6323T:p.A2108V
F10	10.05	CDH11	synonymous	16q21	0.081206	0.0625	rs35182	0.092	CDH11:NM_001797:exon3:c.C99T:p.P33P
F10	9.923	NUB1	synonymous	7q36.1	0.485389	0.384385	rs386956	0.52	NUB1:NM_001243351:exon9:c.C912T:p.Y304
F10	8.403	FLG	nonsynonymous	1q21.3	0.331462	0.514177	rs2065955	0.299	FLG:NM_002016:exon3:c.G10307C:p.G3436A
F10	8.119	ZNF469	synonymous	16q24.2	.	0.0003993	rs150435442	0.00066	ZNF469:NM_001127464:exon1:c.G2841A:p.R
F10	7.596	FOXO1	nonsynonymous	13q14.11	.	.	.	.	FOXO1:NM_002015:exon1:c.C251T:p.P84L
F10	7.482	MPDZ	synonymous	9p23	0.0049	0.0013977	rs34704118	0.00459	MPDZ:NM_001261406:exon19:c.A2580G:p.L8
F10	7.041	FLG	synonymous	1q21.3	0.180231	0.396366	rs57672167	0.269	FLG:NM_002016:exon3:c.G8673T:p.V2891V
F10	6.882	FLG	nonsynonymous	1q21.3	0.107796	0.287141	rs11588170	0.229	FLG:NM_002016:exon3:c.G1330A:p.G444R
F10	6.855	FLG	synonymous	1q21.3	0.251771	0.469649	rs9436066	0.299	FLG:NM_002016:exon3:c.G9645T:p.V3215V
F10	6.267	FLG	synonymous	1q21.3	0.108027	0.28754	rs2065956	0.229	FLG:NM_002016:exon3:c.G10017A:p.Q3339Q

F10	4.804	FLG	nonsynonymous	1q21.3	0.197846	.	rs7546186	0.219	FLG:NM_002016:exon3:c.G6134C:p.S2045T
F10	3.844	FLG	synonymous	1q21.3	0.057488	.	rs57670307	0.182	FLG:NM_002016:exon3:c.C8343T:p.D2781D
F10	3.616	COL4A3	synonymous	2q36.3	0.068087	0.105232	rs34019152	0.092	COL4A3:NM_000091:exon23:c.G1452A:p.G48
F10	1.552	NUB1	synonymous	7q36.1	0.406845	0.332867	rs2159158	0.397	NUB1:NM_001243351:exon15:c.A1905G:p.A6
F10	1.414	FLG	nonsynonymous	1q21.3	0.270567	0.464058	rs2011331	0.295	FLG:NM_002016:exon3:c.A1360G:p.T454A
F10	1.143	FLG	nonsynonymous	1q21.3	0.109565	0.290535	rs72697000	0.234	FLG:NM_002016:exon3:c.G10590T:p.R3530S
F10	1.027	FLG	nonsynonymous	1q21.3	0.281154	0.47484	rs3126072	0.297	FLG:NM_002016:exon3:c.G7633A:p.G2545R
F10	0.93	FLG	synonymous	1q21.3	.	0.453474	rs6664985	0.278	FLG:NM_002016:exon3:c.C6990T:p.H2330H
F10	0.819	FLG	nonsynonymous	1q21.3	0.198601	.	rs7512857	0.218	FLG:NM_002016:exon3:c.T6058G:p.S2020A
F10	0.424	DOCK9	synonymous	13q32.3	0.19774	0.205871	rs2296984	0.202	DOCK9:uc001vns.2:exon12:c.A1197C:p.A399A
F10	0.264	FLG	synonymous	1q21.3	0.26984	0.463658	rs2184954	0.294	FLG:NM_002016:exon3:c.T6498C:p.S2166S
F10	0.147	FLG	synonymous	1q21.3	0.333385	0.522564	rs3126067	0.301	FLG:NM_002016:exon3:c.T10473C:p.N3491N
F10	0.138	FLG	synonymous	1q21.3	0.262729	0.473642	rs3091276	0.284	FLG:NM_002016:exon3:c.T10194C:p.S3398S
F10	0.135	FLG	synonymous	1q21.3	0.282177	0.473243	rs3120653	0.298	FLG:NM_002016:exon3:c.T2508C:p.D836D
F10	0.113	FLG	synonymous	1q21.3	0.263611	0.463059	rs66831674	0.294	FLG:NM_002016:exon3:c.T3387C:p.S1129S
F10	0.097	XRCC1	synonymous	19q13.31	0.42032	0.321286	rs915927	0.335	XRCC1:NM_006297:exon7:c.A618G:p.P206P
F10	0.085	FLG	nonsynonymous	1q21.3	.	.	rs78974621	0.235	FLG:NM_002016:exon3:c.A6626G:p.H2209R
F10	0.046	FLG	synonymous	1q21.3	0.194449	0.327676	rs7512554	0.221	FLG:NM_002016:exon3:c.T6354C:p.H2118H
F10	0.034	FLG	nonsynonymous	1q21.3	0.143626	.	rs66954353	0.237	FLG:NM_002016:exon3:c.A6574C:p.K2192Q
F10	0.025	FLG	synonymous	1q21.3	0.255077	0.447085	rs3126066	0.278	FLG:NM_002016:exon3:c.T10491C:p.D3497D
F10	0.017	FLG	nonsynonymous	1q21.3	0.154475	0.330471	rs71625201	0.258	FLG:NM_002016:exon3:c.G7192C:p.E2398Q
F10	0.016	FLG	synonymous	1q21.3	0.203829	0.367212	rs2338554	0.241	FLG:NM_002016:exon3:c.T6603C:p.D2201D
F10	0.011	FLG	nonsynonymous	1q21.3	.	0.25639	rs7518080	0.193	FLG:NM_002016:exon3:c.G10691A:p.R3564H
F10	0.006	FLG	nonsynonymous	1q21.3	0.15426	0.343051	rs71625200	0.261	FLG:NM_002016:exon3:c.A7330G:p.K2444E
F10	0.006	FLG	nonsynonymous	1q21.3	.	0.539736	rs2184953	0.317	FLG:NM_002016:exon3:c.T6580C:p.Y2194H
F10	0.004	FLG	synonymous	1q21.3	0.156005	0.341054	rs6681433	0.261	FLG:NM_002016:exon3:c.A9966G:p.Q3322Q
F10	0.004	FLG	synonymous	1q21.3	0.172305	0.328075	rs80353812	0.22	FLG:NM_002016:exon3:c.T6072C:p.H2024H
F10	0.002	FLG	nonsynonymous	1q21.3	0.195294	0.327676	rs7512553	0.221	FLG:NM_002016:exon3:c.T6355C:p.Y2119H
F10	0.002	FLG	nonsynonymous	1q21.3	0.15713	0.343051	rs11581433	0.261	FLG:NM_002016:exon3:c.A4126G:p.R1376G

F10	0.001	FLG	nonsynonymous	1q21.3	0.269799	0.46226	rs3126074	0.294	FLG:NM_002016:exon3:c.C7521G:p.H2507Q
F10	0.001	FLG	nonsynonymous	1q21.3	0.15762	0.342452	rs55650366	0.261	FLG:NM_002016:exon3:c.T7442C:p.L2481S
F10	0.001	FLG	nonsynonymous	1q21.3	0.101799	.	rs78179835	0.208	FLG:NM_002016:exon3:c.G6891C:p.E2297D
F10	0.001	FLG	nonsynonymous	1q21.3	.	0.34385	rs74129452	0.237	FLG:NM_002016:exon3:c.A6462C:p.Q2154H
F10	0.001	FLG	nonsynonymous	1q21.3	0.346917	0.538139	rs3126079	0.317	FLG:NM_002016:exon3:c.C5883A:p.H1961Q
F10	0.001	FLG	nonsynonymous	1q21.3	0.310165	0.534145	rs58001094	0.306	FLG:NM_002016:exon3:c.C3500G:p.A1167G
F10	0.001	FLG	nonsynonymous	1q21.3	0.158312	0.344649	rs11584340	0.261	FLG:NM_002016:exon3:c.C1432T:p.P478S
F10	0.001	COL4A3	nonsynonymous	2q36.3	0.073534	0.111422	rs11677877	0.094	COL4A3:NM_000091:exon22:c.A1352G:p.H45
F11	28.4	COL4A1	nonsynonymous	13q34	0.000231	0.0001996	rs141395813	0.00010	COL4A1:NM_001845:exon45:c.G3997A:p.D13
F11	24.2	FLG	nonsynonymous	1q21.3	0.017684	0.0249601	rs36006086	0.023	FLG:NM_002016:exon3:c.C5671T:p.R1891W
F11	23.7	ZNF469	nonsynonymous	16q24.2	.	0.0027955	rs75288466	0.00411	ZNF469:NM_001127464:exon2:c.G10277A:p.R Q
F11	23	COL4A3	nonsynonymous	2q36.3	0.16577	0.115815	rs55703767	0.167	COL4A3:NM_000091:exon17:c.G976T:p.D326Y
F11	22.6	FLG	nonsynonymous	1q21.3	0.139243	0.327676	rs74129461	0.256	FLG:NM_002016:exon3:c.G2263A:p.E755K
F11	22.5	IL1A	nonsynonymous	2q13	0.261648	0.217452	rs17561	0.272	IL1A:NM_000575:exon5:c.G340T:p.A114S
F11	22.3	FLG	nonsynonymous	1q21.3	0.004306	0.0047923	rs12750571	0.00495	FLG:NM_002016:exon3:c.C4309T:p.R1437C
F11	18.89	COL8A2	nonsynonymous	1p34.3	0.000616	0.0005990	rs142307403	0.00075	COL8A2:NM_001294347:exon4:c.G1504A:p.G
F11	17.89	FLG	nonsynonymous	1q21.3	0.01753	0.0249601	rs35621145	0.023	FLG:NM_002016:exon3:c.C9808T:p.R3270C
F11	15.37	FLG	nonsynonymous	1q21.3	0.155774	0.340655	rs41267154	0.26	FLG:NM_002016:exon3:c.G995T:p.G332V
F11	14.38	COL5A1	synonymous	9q34.3	0.031601	0.0415335	rs78511105	0.041	COL5A1:NM_000093:exon36:c.C2892T:p.G964
F11	13.89	COL4A2	synonymous	13q34	0.262849	0.238618	rs76425569	0.231	COL4A2:NM_001846:exon19:c.G1095A:p.P365
F11	13.84	COL27A1	synonymous	9q32	0.080194	0.0796725	rs13290696	0.093	COL27A1:NM_032888:exon8:c.G2148T:p.P716
F11	12.89	COL4A2	synonymous	13q34	0.284764	0.246805	rs74941798	0.295	COL4A2:NM_001846:exon19:c.C1179T:p.I393I
F11	12.76	FLG	nonsynonymous	1q21.3	0.017684	0.0249601	rs12750081	0.023	FLG:NM_002016:exon3:c.C4568T:p.T1523I
F11	12.72	COL4A2	nonsynonymous	13q34	0.103497	0.103834	rs3803230	0.106	COL4A2:NM_001846:exon27:c.G2048C:p.G68
F11	12.41	COL5A1	synonymous	9q34.3	0.425188	0.408147	rs2228560	0.375	COL5A1:NM_000093:exon58:c.G4482C:p.P149
F11	12.11	FLG	nonsynonymous	1q21.3	0.017607	0.0247604	rs71626704	0.023	FLG:NM_002016:exon3:c.C6045A:p.D2015E
F11	10.4	MPDZ	synonymous	9p23	0.318029	0.393171	rs2274856	0.286	MPDZ:NM_001261406:exon20:c.G2754A:p.S9
F11	9.923	NUB1	synonymous	7q36.1	0.485389	0.384385	rs386956	0.52	NUB1:NM_001243351:exon9:c.C912T:p.Y304Y

F11	9.85	COL4A2	splicing	13q34	0.192017	0.190895	rs3803228	0.175	NM_001846:exon32:c.2759-5T>C
F11	8.831	VSX2	synonymous	14q24.3	0.033133	0.0195687	rs62006815	0.031	VSX2:NM_182894:exon5:c.G831A:p.L277L
F11	7.041	FLG	synonymous	1q21.3	0.180231	0.396366	rs57672167	0.269	FLG:NM_002016:exon3:c.G8673T:p.V2891V
F11	6.916	IL1RN	synonymous	2q13	0.333154	0.35004	rs315952	0.308	IL1RN:NM_173842:exon4:c.T390C:p.S130S
F11	6.855	FLG	synonymous	1q21.3	0.251771	0.469649	rs9436066	0.299	FLG:NM_002016:exon3:c.G9645T:p.V3215V
F11	1.678	FLG	nonsynonymous	1q21.3	0.017684	0.0249601	rs62623409	0.023	FLG:NM_002016:exon3:c.C5617A:p.Q1873K
F11	1.552	NUB1	synonymous	7q36.1	0.406845	0.332867	rs2159158	0.397	NUB1:NM_001243351:exon15:c.A1905G:p.A6
F11	1.414	FLG	nonsynonymous	1q21.3	0.270567	0.464058	rs2011331	0.295	FLG:NM_002016:exon3:c.A1360G:p.T454A
F11	1.027	FLG	nonsynonymous	1q21.3	0.281154	0.47484	rs3126072	0.297	FLG:NM_002016:exon3:c.G7633A:p.G2545R
F11	0.384	FLG	nonsynonymous	1q21.3	0.089428	.	rs80221306	0.208	FLG:NM_002016:exon3:c.A8807G:p.D2936G
F11	0.264	FLG	synonymous	1q21.3	0.26984	0.463658	rs2184954	0.294	FLG:NM_002016:exon3:c.T6498C:p.S2166S
F11	0.034	FLG	nonsynonymous	1q21.3	0.143626	.	rs66954353	0.237	FLG:NM_002016:exon3:c.A6574C:p.K2192Q
F11	0.012	FLG	synonymous	1q21.3	0.017607	0.0249601	rs12732920	0.023	FLG:NM_002016:exon3:c.T4410C:p.H1470H
F11	0.006	FLG	nonsynonymous	1q21.3	.	0.539736	rs2184953	0.317	FLG:NM_002016:exon3:c.T6580C:p.Y2194H
F11	0.004	FLG	synonymous	1q21.3	0.156005	0.341054	rs6681433	0.261	FLG:NM_002016:exon3:c.A9966G:p.Q3322Q
F11	0.002	FLG	nonsynonymous	1q21.3	0.15713	0.343051	rs11581433	0.261	FLG:NM_002016:exon3:c.A4126G:p.R1376G
F11	0.002	FLG	nonsynonymous	1q21.3	0.017684	0.0249601	rs7512779	0.023	FLG:NM_002016:exon3:c.C2181A:p.H727Q
F11	0.001	FLG	nonsynonymous	1q21.3	0.269799	0.46226	rs3126074	0.294	FLG:NM_002016:exon3:c.C7521G:p.H2507Q
F11	0.001	FLG	nonsynonymous	1q21.3	0.15762	0.342452	rs55650366	0.261	FLG:NM_002016:exon3:c.T7442C:p.L2481S
F11	0.001	FLG	nonsynonymous	1q21.3	0.346917	0.538139	rs3126079	0.317	FLG:NM_002016:exon3:c.C5883A:p.H1961Q
F11	0.001	FLG	nonsynonymous	1q21.3	.	0.0339457	rs12756586	0.025	FLG:NM_002016:exon3:c.C2938G:p.H980D
F11	0.001	FLG	nonsynonymous	1q21.3	0.158312	0.344649	rs11584340	0.261	FLG:NM_002016:exon3:c.C1432T:p.P478S
F11		VSX1	splicing	20p11.21	.	0.230831	rs11431377	0.221	NM_014588:exon6:c.809-5->T
F12	17.52	MUTYH	nonsynonymous	1p34.1	0.050054	0.0203674	rs3219484	0.049	MUTYH:NM_001048171:exon2:c.G64A:p.V22I
F12	10.76	FLG	nonsynonymous	1q21.3	0.107873	0.286941	rs77422831	0.229	FLG:NM_002016:exon3:c.G11213A:p.R3738H
F12	1.143	FLG	nonsynonymous	1q21.3	0.109565	0.290535	rs72697000	0.234	FLG:NM_002016:exon3:c.G10590T:p.R3530S
F12	0.025	FLG	synonymous	1q21.3	0.255077	0.447085	rs3126066	0.278	FLG:NM_002016:exon3:c.T10491C:p.D3497D
F12	0.147	FLG	synonymous	1q21.3	0.333385	0.522564	rs3126067	0.301	FLG:NM_002016:exon3:c.T10473C:p.N3491N
F12	6.267	FLG	synonymous	1q21.3	0.108027	0.28754	rs2065956	0.229	FLG:NM_002016:exon3:c.G10017A:p.Q3339Q

F12	0.004	FLG	synonymous	1q21.3	0.156005	0.341054	rs6681433	0.261	FLG:NM_002016:exon3:c.A9966G:p.Q3322Q
F12	6.855	FLG	synonymous	1q21.3	0.251771	0.469649	rs9436066	0.299	FLG:NM_002016:exon3:c.G9645T:p.V3215V
F12	0.384	FLG	nonsynonymous	1q21.3	0.089428	.	rs80221306	0.208	FLG:NM_002016:exon3:c.A8807G:p.D2936G
F12	7.041	FLG	synonymous	1q21.3	0.180231	0.396366	rs57672167	0.269	FLG:NM_002016:exon3:c.G8673T:p.V2891V
F12	0.097	FLG	synonymous	1q21.3	.	.	rs201652079	0.00017	FLG:NM_002016:exon3:c.T7842C:p.A2614A
F12	0.001	FLG	nonsynonymous	1q21.3	.	0.0027955	rs138143041	0.00016	FLG:NM_002016:exon3:c.T7687C:p.W2563R
F12	1.027	FLG	nonsynonymous	1q21.3	0.281154	0.47484	rs3126072	0.297	FLG:NM_002016:exon3:c.G7633A:p.G2545R
F12	0.001	FLG	nonsynonymous	1q21.3	0.269799	0.46226	rs3126074	0.294	FLG:NM_002016:exon3:c.C7521G:p.H2507Q
F12	0.006	FLG	nonsynonymous	1q21.3	0.15426	0.343051	rs71625200	0.261	FLG:NM_002016:exon3:c.A7330G:p.K2444E
F12	0.006	FLG	nonsynonymous	1q21.3	.	0.539736	rs2184953	0.317	FLG:NM_002016:exon3:c.T6580C:p.Y2194H
F12	0.264	FLG	synonymous	1q21.3	0.26984	0.463658	rs2184954	0.294	FLG:NM_002016:exon3:c.T6498C:p.S2166S
F12	0.001	FLG	nonsynonymous	1q21.3	0.346917	0.538139	rs3126079	0.317	FLG:NM_002016:exon3:c.C5883A:p.H1961Q
F12	11.35	FLG	nonsynonymous	1q21.3	0.10795	0.28754	rs12407748	0.229	FLG:NM_002016:exon3:c.G5672A:p.R1891Q
F12	10.19	FLG	nonsynonymous	1q21.3	0.107753	0.28754	rs12405241	0.229	FLG:NM_002016:exon3:c.C5414T:p.A1805V
F12	23.8	FLG	nonsynonymous	1q21.3	0.10772	0.28774	rs12405278	0.229	FLG:NM_002016:exon3:c.C5095T:p.R1699C
F12	17.66	FLG	nonsynonymous	1q21.3	0.118714	0.296725	rs12407807	0.232	FLG:NM_002016:exon3:c.G5051A:p.R1684H
F12	20.7	FLG	nonsynonymous	1q21.3	0.10772	0.287141	rs11204978	0.229	FLG:NM_002016:exon3:c.C4445A:p.S1482Y
F12	15.37	FLG	nonsynonymous	1q21.3	0.107676	0.28774	rs11586631	0.229	FLG:NM_002016:exon3:c.G4079A:p.R1360H
F12	0.135	FLG	synonymous	1q21.3	0.282177	0.473243	rs3120653	0.298	FLG:NM_002016:exon3:c.T2508C:p.D836D
F12	22.6	FLG	nonsynonymous	1q21.3	0.139243	0.327676	rs74129461	0.256	FLG:NM_002016:exon3:c.G2263A:p.E755K
F12	0.001	FLG	nonsynonymous	1q21.3	0.158312	0.344649	rs11584340	0.261	FLG:NM_002016:exon3:c.C1432T:p.P478S
F12	1.414	FLG	nonsynonymous	1q21.3	0.270567	0.464058	rs2011331	0.295	FLG:NM_002016:exon3:c.A1360G:p.T454A
F12	6.882	FLG	nonsynonymous	1q21.3	0.107796	0.287141	rs11588170	0.229	FLG:NM_002016:exon3:c.G1330A:p.G444R
F12	15.37	FLG	nonsynonymous	1q21.3	0.155774	0.340655	rs41267154	0.26	FLG:NM_002016:exon3:c.G995T:p.G332V
F12	23	PARP1	nonsynonymous	1q42.12	0.000077	0.0001996	rs200470832	0.00004	PARP1:NM_001618:exon6:c.G727A:p.D243N
F12	22.5	IL1A	nonsynonymous	2q13	0.261648	0.217452	rs17561	0.272	IL1A:NM_000575:exon5:c.G340T:p.A114S
F12	6.916	IL1RN	synonymous	2q13	0.333154	0.35004	rs315952	0.308	IL1RN:NM_173842:exon4:c.T390C:p.S130S
F12	20.3	COL4A3	nonsynonymous	2q36.3	0.063256	0.100639	rs34505188	0.091	COL4A3:NM_000091:exon21:c.G1223A:p.R40
F12	0.001	COL4A3	nonsynonymous	2q36.3	0.073534	0.111422	rs11677877	0.094	COL4A3:NM_000091:exon22:c.A1352G:p.H45

F12	3.616	COL4A3	synonymous	2q36.3	0.068087	0.105232	rs34019152	0.092	COL4A3:NM_000091:exon23:c.G1452A:p.G48
F12	12.19	LOX	nonsynonymous	5q23.2	0.175339	0.156749	rs1800449	0.17	LOX:NM_002317:exon1:c.G473A:p.R158Q
F12	8.103	SPARC	synonymous	5q33.1	0.190912	0.197684	rs2304052	0.11	SPARC:NM_003118:exon3:c.A66G:p.E22E
F12	22.4	HGF	nonsynonymous	7q21.11	0.050284	0.0197684	rs5745687	0.045	HGF:NM_000601:exon8:c.G910A:p.E304K
F12	1.296	HGF	synonymous	7q21.11	0.104061	0.103435	rs5745635	0.042	HGF:NM_000601:exon3:c.A333G:p.E111E
F12	9.923	NUB1	synonymous	7q36.1	0.485389	0.384385	rs386956	0.52	NUB1:NM_001243351:exon9:c.C912T:p.Y304Y 1:NM_016118:exon9:c.C912T:p.Y304Y
F12	15.01	RXRA	synonymous	9q34.2	0.012917	0.0075878	rs1805348	0.012	RXRA:NM_001291921:exon9:c.G1080A:p.A36
F12	24.9	COL5A1	nonsynonymous	9q34.3	0.031601	0.0235623	rs61735045	0.035	COL5A1:NM_000093:exon13:c.G1588A:p.G53
F12	12.41	COL5A1	synonymous	9q34.3	0.425188	0.408147	rs2228560	0.375	COL5A1:NM_000093:exon58:c.G4482C:p.P149
F12	11.9	STK24	synonymous	13q32.2	0.000231	0.0003993	rs55640122	0.00034	STK24:NM_001286649:exon6:c.G852C:p.S284
F12	14.3	DOCK9	synonymous	13q32.3	0.13827	0.167532	rs12428661	0.181	DOCK9:NM_001130049:exon5:c.C477T:p.V159
F12	9.85	COL4A2	splicing	13q34	0.192017	0.190895	rs3803228	0.175	NM_001846:exon32:c.2759-5T>C
F12	6.676	COL4A2	splicing	13q34	0.035794	0.117212	rs2296849	0.049	NM_001846:exon38:c.3455-5C>G
F12	8.443	COL4A2	splicing	13q34	0.053195	0.128395	rs2274544	0.091	NM_001846:exon39:c.3634+4C>T
F12		POLG	nonframeshift deletion	15q26.1	.	.	.	0.003212	POLG:NM_001126131:exon2:c.153_158del:p. del
F12	19.86	CDH11	synonymous	16q21	0.322159	0.223642	rs28216	0.33	CDH11:NM_001797:exon7:c.G945A:p.S315S
F12	28.2	CDH11	nonsynonymous	16q21	0.260111	0.18151	rs35195	0.273	CDH11:NM_001797:exon6:c.C764T:p.T255M
F12	2.599	BANP	splicing	16q24.2	.	.	.	.	.
F12	10.23	ZNF469	synonymous	16q24.2	.	.	.	.	ZNF469:NM_001127464:exon1:c.G2808T:p.A9
F12	8.17	ZNF469	synonymous	16q24.2	.	.	.	0.00003	ZNF469:NM_001127464:exon2:c.G4836A:p.A1
F12	9.066	ZNF469	nonsynonymous	16q24.2	.	.	.	.	ZNF469:NM_001127464:exon2:c.C8342T:p.P2
F12		ZNF469	nonframeshift deletion	16q24.2	.	0.00179712	.	0.003504	ZNF469:NM_001127464:exon2:c.9010_9024d 004_3008del
F12	10.27	SLC4A11	synonymous	20p13	0.093803	0.0485224	rs41281860	0.091	SLC4A11:NM_032034:exon13:c.C1659T:p.N55
F12	10.54	SLC4A11	synonymous	20p13	0.090804	0.0451278	rs6084312	0.085	SLC4A11:NM_032034:exon11:c.G1389A:p.T46
F13	28.2	CDH11	nonsynonymous	16q21	0.260111	0.18151	rs35195	0.273	CDH11:NM_001797:exon6:c.C764T:p.T255M
F13	24.9	COL5A1	nonsynonymous	9q34.3	0.031601	0.0235623	rs61735045	0.035	COL5A1:NM_000093:exon13:c.G1588A:p.G53
F13	19.86	CDH11	synonymous	16q21	0.322159	0.223642	rs28216	0.33	CDH11:NM_001797:exon7:c.G945A:p.S315S

F13	13.89	COL4A2	synonymous	13q34	0.262849	0.238618	rs76425569	0.231	COL4A2:NM_001846:exon19:c.G1095A:p.P365
F13	12.89	COL4A2	synonymous	13q34	0.284764	0.246805	rs74941798	0.295	COL4A2:NM_001846:exon19:c.C1179T:p.I393
F13	12.87	SLC4A11	nonsynonymous	20p13	.	0.0640974	rs79057061	0.031	SLC4A11:NM_001174090:exon1:c.A148G:p.R5
F13	12.41	COL5A1	synonymous	9q34.3	0.425188	0.408147	rs2228560	0.375	COL5A1:NM_000093:exon58:c.G4482C:p.P149
F13	12.19	LOX	nonsynonymous	5q23.2	0.175339	0.156749	rs1800449	0.17	LOX:NM_002317:exon1:c.G473A:p.R158Q
F13	11.27	ZNF469	nonsynonymous	16q24.2	0.056338	0.0894569	rs141218390	0.075	ZNF469:NM_001127464:exon2:c.C7424A:p.A2
F13	10.27	SLC4A11	synonymous	20p13	0.093803	0.0485224	rs41281860	0.091	SLC4A11:NM_032034:exon13:c.C1659T:p.N55
F13	1.652	ZNF469	synonymous	16q24.2	.	0.0942492	rs12919507	0.079	ZNF469:NM_001127464:exon2:c.G6468T:p.T2
F13	0.097	XRCC1	synonymous	19q13.31	0.42032	0.321286	rs915927	0.335	XRCC1:NM_006297:exon7:c.A618G:p.P206P
F13	0.039	COL4A1	synonymous	13q34	0.252499	0.20647	rs995224	0.258	COL4A1:NM_001845:exon21:c.T1257C:p.P419
KF	22.6	FLG	nonsynonymous	1q21.3	0.077095	.	rs75448155	0.191	FLG:NM_002016:exon3:c.G10903A:p.D3635N
KF	14.99	COL5A1	synonymous	9q34.3	.	.	.	.	COL5A1:NM_000093:exon54:c.C4191T:p.P139
KF	14.66	FLG	nonsynonymous	1q21.3	0.099721	.	rs71625202	0.206	FLG:NM_002016:exon3:c.G7097C:p.S2366T
KF	9.808	FLG	nonsynonymous	1q21.3	0.000077	.	rs376237195	0.00001	FLG:NM_002016:exon3:c.G5212A:p.G1738R
KF	8.403	FLG	nonsynonymous	1q21.3	0.331462	0.514177	rs2065955	0.299	FLG:NM_002016:exon3:c.G10307C:p.G3436A
KF	8.201	HGF	nonsynonymous	7q21.11	.	0.0021964	.	0.00143	HGF:NM_000601:exon16:c.G1765A:p.V589I
KF	0.93	FLG	synonymous	1q21.3	.	0.453474	rs6664985	0.278	FLG:NM_002016:exon3:c.C6990T:p.H2330H
KF	0.819	FLG	nonsynonymous	1q21.3	0.198601	.	rs7512857	0.218	FLG:NM_002016:exon3:c.T6058G:p.S2020A
KF	0.017	FLG	nonsynonymous	1q21.3	0.154475	0.330471	rs71625201	0.258	FLG:NM_002016:exon3:c.G7192C:p.E2398Q
KF	0.006	FLG	nonsynonymous	1q21.3	0.15426	0.343051	rs71625200	0.261	FLG:NM_002016:exon3:c.A7330G:p.K2444E
KF	0.004	FLG	synonymous	1q21.3	0.172305	0.328075	rs80353812	0.22	FLG:NM_002016:exon3:c.T6072C:p.H2024H
KF	0.001	FLG	nonsynonymous	1q21.3	0.15762	0.342452	rs55650366	0.261	FLG:NM_002016:exon3:c.T7442C:p.L2481S
KF	0.001	FLG	nonsynonymous	1q21.3	0.101799	.	rs78179835	0.208	FLG:NM_002016:exon3:c.G6891C:p.E2297D
KF	0.001	FLG	nonsynonymous	1q21.3	0.346917	0.538139	rs3126079	0.317	FLG:NM_002016:exon3:c.C5883A:p.H1961Q
KF	0.001	FLG	nonsynonymous	1q21.3	0.068046	0.0369409	rs34806697	0.067	FLG:NM_002016:exon3:c.A5741G:p.Q1914R
KF	0.001	FLG	nonsynonymous	1q21.3	0.310165	0.534145	rs58001094	0.306	FLG:NM_002016:exon3:c.C3500G:p.A1167G