Genetic Variants in Corneal Dystrophy Genes: a Maltese Cohort Study. Inhibition of TGFBI as a Treatment Modality.

Gabriella Maria Guo Sciriha

MD, MRCOphth, FEBO

Faculty of Medicine and Surgery, University of Malta

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

February 2024

L-Università ta' Malta

FACULTY/INSTITUTE/CENTRE/SCHOOL: Faculty of Medicine and Surgery, University of Malta

DECLARATION OF AUTHENTICITY FOR DOCTORAL STUDENTS

(a) **Authenticity of Thesis/Dissertation**

I hereby declare that I am the legitimate author of this Thesis/Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) **Research Code of Practice and Ethics Review Procedure**

I declare that I have abided by the University's Research Ethics Review Procedures. Research Ethics & Data Protection form code: UREC 11/2017; FRECMDS_1718_064

☒ As a Ph.D. student, as per Regulation 66 of the Doctor of Philosophy Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

☐ As a Doctor of Sacred Theology student, as per Regulation 17 (3) of the Doctor of Sacred Theology Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

☐ As a Doctor of Music student, as per Regulation 26 (2) of the Doctor of Music Regulations, I accept that my dissertation be made publicly available on the University of Malta Institutional Repository.

☒ As a Professional Doctorate student, as per Regulation 55 of the Professional Doctorate Regulations, I accept that my dissertation be made publicly available on the University of Malta Institutional Repository.

ABSTRACT

The three aims of this study were to a) establish which worldwide populations have a corneal dystrophy (CD) genetic makeup closest to that of the Maltese; b) identify mutations present in a Maltese family that exhibit granular corneal dystrophy 1 (GCD1), a subtype of *TGFBI* CDs; c) explore *TGFBI* inhibition as a treatment modality.

Genetic prevalence of CD subtypes and fixation index (F_{ST}) values for Maltese single nucleotide polymorphisms (SNPs) were calculated and compared to global cohorts. Clinical exome sequencing was performed on mouthwash samples from Maltese GCD1 phenotype individuals. A scoping literature review to identify compounds that decrease corneal TGFBIp (protein) levels was conducted to explore their potential to be used as a cost-effective approach via drug repurposing. Human corneal epithelial cells (HCECs) were cultured and shRNA mediated knockdown (KD) of *TGFBI* was effectively performed. HCECs were also exposed to lithium (Li) and mitomycin C (MMC). RNA extraction and sequencing revealed gene expression levels in control, *TGFBI* KD, Li and MMC treated HCECs. Differential expression (DE) and enrichment analysis (ORA) were performed on these samples.

F_{ST} values showed least differentiation with Puerto Rican, Mexican, and Colombian cohorts. The mutation in the GCD1 phenotype patients was identified as R555W (*TGFBI* gene). 16 compounds that can theoretically reduce the levels of mutant TGFBIp in corneal cells were identified. ORA of *TGFBI* KD DE genes showed enrichment of adhesion and signalling proteins. Surprisingly, *TGFBI* expression was found to be upregulated in the Li and MMC groups at 72hours.

Identifying the Hispanic cohorts as those with a CD genome closest to the Maltese implies that when no comparable Maltese data is available, research in these cohorts can be used to guide future treatment strategies for Maltese CD individuals. The clinical exome sequencing study is the first CD genetic study that has ever been carried out on GCD1 Maltese individuals, generating new data about the previously unknown genetic pool. This project is also the first of its kind to explore the DE of genes in KD, Li and MMC treated HCECs, leading to further understanding of *TGFBI* related molecular pathways in HCECs. Surgical treatment of *TGFBI* CDs can be associated with serious complications and recurrence is almost universal. The introduction of gene therapy as a treatment option would be a breakthrough.

ACKNOWLEDGEMENTS

Just as a “king can’t be king without the strength of his queen”, this thesis would not have been possible without the invaluable input of several individuals who assisted me along the journey of my PhD study.

I would firstly like to thank the patients and families who kindly agreed to be involved in this study, without whom, this work would not have been achievable. This PhD was generously funded by the Dean’s Grant, Faculty of Medicine and Surgery, University of Malta and this support is acknowledged with gratitude.

I am grateful to many people for supporting me during my PhD. I would like to thank my main supervisor Professor Joseph Borg. Without his unstinting support and guidance, the data presented in this thesis could not have been compiled. I would also like to thank my two other supervisors Mr. Mario Vella and Mr. Franco Mercieca for their help and expertise. I am certain that the knowledge and skills that I have learned from them will guide me throughout my clinical and academic career.

Many of my colleagues, past and present, have helped me during this PhD. I would like to thank bioinformatician Dr. Josef Borg who guided me through the process of bioinformatic analysis of the RNA seq results obtained, Mr. Mark Debono for his help in operating the R package, Dr Graziella Zahra and Mr Mark Briffa for their guidance while performing library preparation and clinical exome sequencing as well as Dr. Janet Sultana for her assistance in the pharmacological aspect of this study. I would also like to thank the researchers involved in The Malta Human Genome Project, which was led by Professor Alex Felice of the University of Malta and was funded by the Malta Council for Science and Technology (R&I 2013-041), for making available genetic data from a large Maltese cohort.

My thanks also go Professor Anthony Fenech for allowing me to use the cell culture facilities at the ARU, University of Malta and Dr. Christopher Barbara for allowing the use of pathology laboratories at Mater Dei Hospital.

Special thanks go to my family who have been vital during this PhD, their continuous encouragement, and support, have assisted me enormously.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

A2M	Alpha-2-macroglobulin
ACT	Actin, cytoplasmic 1
ACTB	Actin beta
AKT	Ak strain transforming
ALB	Serum albumin
ALDH3A1	Aldehyde dehydrogenase, dimeric NADP-preferring
ANKRD20A10P	Ankyrin repeat domain 20 family member A10
ANXA2	Annexin A2
AP-1	Activating protein-1
APCS	Serum amyloid P component
APOD	Apolipoprotein D
AQP3	Aquaporin 3
AREG	Amphiregulin
Bcl-2	B-cell lymphoma 2
BDNF	brain-derived neurotrophic factor
cAMP	Cyclic adenosine monophosphate
CCDF	Central cloudy dystrophy of François
CD	Corneal dystrophy
CD24	Cluster of differentiation 24
CHED	Congenital hereditary endothelial dystrophy
CHML	CHM Like Rab Escort Protein
CHST6	Carbohydrate sulfotransferase 6
CLU	Clusterin (Apolipoprotein J)
CMV	Cytomegalovirus

COL12A1	Collagen Type XII alpha 1 Chain
COL1A1	Collagen type I alpha 1 chain
COL1A2	Collagen type I alpha 2 chain
COL3A1	Collagen type III alpha 1 chain
COL4A1	Collagen type IV alpha 1 chain
COL4A2	Collagen type IV alpha 2 chain
COL4A3	Collagen type IV alpha 3 chain
COL5A1	Collagen type V alpha 1 chain
COL5A2	Collagen Type V Alpha 2 Chain
COL6A1	Collagen Type VI Alpha 1 Chain
COL6A2	Collagen Type VI Alpha 2 Chain
COL8A2	Collagen alpha-2(VIII) chain
CPA2	Carboxypeptidase A2
CPA4	Carboxypeptidase A4
CREB	cAMP-Response Element Binding Protein
CSCD	Congenital stromal corneal dystrophy
CSTA	Cystatin A
CTD	Comparative Toxicogenomics Database
CUL4A	Cullin 4A
DCD	Dermcidin
DCN	Decorin
DDX39B	DEAD box helicase 39b
DNA	Deoxyribonucleic acid
DSP	Desmoplakin
EBMD	Epithelial basement membrane dystrophy
EMA	European Medicines Agency

ESR2	Estrogen receptor 2
FCD	Fleck corneal dystrophy
FDA	Food and Drug Administration
FECD	Fuchs endothelial corneal dystrophy
FGF	Fibroblast growth factor
FN1	Fibronectin 1
FRCD	Franceschetti corneal dystrophy
GCD	Granular corneal dystrophy
GCD1	Granular corneal dystrophy, type 1
GCD2	Granular corneal dystrophy, type 2
GDLD	Gelatinous drop-like corneal dystrophy
Gene ontology	Gene ontology
GSK-3	Glycogen Synthase Kinase-3
HBB	Haemoglobin subunit beta
HBLV	Human B-lymphotropic virus
HCEC	human corneal epithelial cell
HSPG	heparan sulphate proteoglycan
HSPG2	heparan sulphate proteoglycan 2
ICAM4	Intercellular adhesion molecule 4
IGF	Insulin-like growth factors
IL1B	Interleukin 1B
IL6	Interleukin 6
IL6R	Interleukin 6 receptor
ITGA11	Integrin subunit alpha 11
ITGA3	Integrin subunit alpha-3
ITGA6	Integrin subunit alpha 6

ITGAV	Integrin Subunit Alpha V
ITGB5	Integrin subunit beta 5
JAK2	Janus Kinase 2
JNK	Jun N-terminal Kinase
KD	Knock down
KERA	Keratocan
KRT12	Keratin 12
KRT3	Keratin 3
LCD	Lattice corneal dystrophy
LCD1	Lattice corneal dystrophy, type 1
LECD	Lisch epithelial corneal dystrophy
Li	Lithium
LOX	Lysyl oxidase
LOXL2	Lysyl oxidase-like 2
LUM	Lumican
MAPK1	Mitogen-activated protein kinase 1
MCD	Macular corneal dystrophy
MCS	Multiple cloning site
MECD	Meesmann corneal dystrophy
MMC	Mitomycin C
MMP2	Matrix metalloproteinase 2
MMP9	Matrix metallopeptidase 9
MTPN	Myotrophin
NGS	Next-generation sequencing
NOG	Noggin
nt	Nucleotides

ORA	Over-Representation Analysis
OVOL2	Ovo like zinc finger 2
PACD	Posterior amorphous corneal dystrophy
PCR	Polymerase chain reaction
PDCD	Pre-Descemet corneal dystrophy
PDGF	Platelet-derived growth factor
PGK	Phosphoglycerokinase gene promoter
PI3-K	Phosphatidylinositol 3-kinase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PKA	Protein Kinase A
PLOD2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2
PPCD	Posterior polymorphous corneal dystrophy
PRDX1	Peroxiredoxin-1
PTGDS	Prostaglandin-H2 D isomerase
PTK2	Protein tyrosine kinase 2
PURO	Puromycin
RBCD	Reis–Bücklers corneal dystrophy
RGD	COOH-terminal Arg-Gly-Asp
RISC	RNA induced silencing complex
RN7SL385P	RNA, 7SL, Cytoplasmic 385
RNA	Ribonucleic acid
RNA5S17	RNA, 5S Ribosomal 17
RNA5SP439	RNA, 5S Ribosomal Pseudogene 439
RNAi	RNA interference
RNA-Seq	RNA sequencing

RPS27A	Ribosomal protein S27a
R-SMAD	Receptor regulated SMAD
S100A4	Protein S100-A4
S100A6	Protein S100-A6
S100A9	Protein S100-A9
S100P	S100 calcium-binding protein P
S27a	Ubiquitin-40S ribosomal protein S27a
SCD	Schnyder corneal dystrophy
scRNA-seq	Single cell RNA sequencing
SDS-PAGE	Sodium dodecyl-sulphate polyacrylamide gel electrophoresis
shRNA	Short hairpin RNAs
siRNA	Small interfering RNA
SLC4A11	Solute carrier family 4
SMAD	Suppressor of mothers against decapentaplegic
SMAD3	Suppressor of Mothers against Decapentaplegic 3
SMAD6	Suppressor of Mothers against Decapentaplegic 6
SMAD7	Suppressor of Mothers against Decapentaplegic 7
SMCD	Subepithelial mucinous corneal dystrophy
SNORD14C	Small Nucleolar RNA, C/D Box 14C
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variants
STAT2	Signal transducer and activator of transcription 2
TBCD	Thiel-Behnke corneal dystrophy
TGF-beta	Transforming growth factor-beta
TGF-betaR	TGF-beta cell-surface receptor
TGFBI	Transforming growth factor beta induced

TGFBIp	Transforming growth factor beta induced protein
TGF-β1	Transforming growth factor β-1
THBS1	Thrombospondin 1
TNC	Tenascin C
TNF	Tumor necrosis factor
U6	Type III RNA polymerase III promoter
UBIAD1	UbiA prenyltransferase domain-containing protein 1
VEGF	Vascular endothelial growth factor
VIM	Vimentin
WebGesalt	GEne SeT AnaLysis Toolkit
WGS	Whole-genome sequencing
XECD	X-linked endothelial corneal dystrophy
YAP1	Yes1 Associated Transcriptional Regulator
ZEB1	Zinc finger E-box binding homeobox 1
zsGreen	Zoanthus sp. green fluorescent protein

PUBLISHED PAPERS

Throughout my PhD research study, I published the following papers related to my PhD work. The results of these publications form part of my dissertation:

1. Guo Sciriha, G., Sultana, J., Petrucci, S., & Borg, J. (2022). Maltese Allelic Variants in Corneal Dystrophy Genes in a Worldwide Setting. *Molecular Diagnosis and Therapy, Springer Nature*, 26, 529-540.

   This publication included work, performed during my PhD study, which involves Maltese genetics in relation to corneal dystrophies and Maltese corneal dystrophy-related genetic differentiation from worldwide populations.

2. Guo Sciriha, G., Sultana, J., & Borg, J. (2022). Identifying and categorizing compounds that reduce corneal transforming growth factor beta induced protein levels: a scoping review. *Expert Review of Clinical Pharmacology, Taylor & Francis*, 15:12, 1423-1442.

   This scoping literature review allowed the identification and choice of compounds to be applied to cultured human corneal epithelial cells in phase 3 of my study.

3. Guo Sciriha, G., Sultana, J., & Borg, J. (2023). Pharmacological treatment for transforming growth factor beta induced corneal dystrophies: what is the way forward? *Expert Review of Clinical Pharmacology, Taylor & Francis*, 16:4, 275-278.

The paper brought forward licenced compounds/drugs that could possibly be applied for the use in the management of *TGFBI* corneal dystrophies. This also helped with the selection of compounds to be tested on HCECs in phase 3 of my project.

CHAPTER 1: INTRODUCTION

Population genetic variation is dependent on biological factors such as admixture among different populations, genetic drift, natural selection and mutations (Guo, Fu, Lee, *et al*., 2014). Recent advancements in molecular genetic techniques and their various applications have significantly facilitated the research of the make-up of a population and have assisted in shedding light on the role genetic mutations and variants have in the pathology and physiology of diseases (Nielsen, Akey, Jakobsson, *et al*., 2017). They assisted further in the enrichment of the principles of ancestral analyses and to clarify the geographic origin of a population (Cavalli-Sforza, 2007).

Due to its unique geographical position, Malta has been inhabited by numerous populations throughout the years. Originally, it was thought that a large percentage of Maltese men were of Phoenician origin (Zalloua, Platt, Sibai, *et al*., 2008). However, around the year 1048, when Malta was repopulated by people originating from Sicily and Calabria, and a small percentage from the Middle East, the ancestry of contemporary Maltese males changed. During this period Malta was dominated by the County of Sicily, it was dominated by the Aragonese and the Spanish until 1530 (Mercieca, 2015).

The human genetic make-up in all individuals is the same in 99.9%. The differences in the residual 0.1% hold crucial clues about our ancestral composition, which also include the risk of disease (Huang, Shu & Cai, 2015), which is one of the principal interests in this study. Single nucleotide polymorphisms (SNPs) are the most common type of sequence variations present in these 0.1% of bases and have been established as genetic markers (Chen & Sullivan, 2003). 15% of all SNPs among different

populations have been documented to be population-specific (Huang, Shu & Cai, 2015) (Spichenok, Budimlija, Mitchell, *et al*., 2011). Populations from specific geographic areas have genetic variants that are frequent among them but rare or absent in others. High throughput next-generation sequencing (NGS) is a recent technological innovation that facilitates deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) sequencing and variant/mutation identification. It uses SNPs as genetic markers, in association with bioinformatics and computational tools, in order to provide a background for the deeper comprehension of the genetic component of diseases (Shastry, 2007).

This research project is the first of its kind to look into Maltese genetics in relation to Corneal Dystrophies (CDs). CDs are a group of genetic conditions characterised by non-inflammatory, bilateral, often progressive alterations of the cornea (Weiss, Muller, Lisch, *et al*., 2008). In CDs, corneal opacification occurs due to deposition of abnormal material in the cornea that in turn leads to decompensation or scarring. Since the cornea is a clear tissue, deposition of abnormal material compromises its normal functionality with consequent decrease in vision (Klinthworth, 2009) (Bron A. J., 1990) (Glavini, 2013).

The first phase of this project addressed the genetic differentiation, with regard to CD related genes, between a large cohort of the Maltese population and various worldwide population cohorts. The objective was, to determine which worldwide population cohorts have a CD-related genetic make-up that is most related to that of the Maltese population. Identifying populations with the least genetic differentiation can provide future diagnostic and treatment strategies for Maltese individuals with CDs, especially when no similar Maltese data are available. Besides, since most of the

CDs cause substantial visual impairment (Klinthworth, 2009) (Bron A. J., 1990), these results are of significant importance in the continuous endeavour of developing improved personalised management options for individuals suffering from CDs.

To date, more than 20 subtypes of CDs have been documented and these are associated with mutations in different chromosomal loci. The Transforming Growth Factor Beta Induced gene (*TGFBI*), which is located on chromosome 5q31, is one of the genes where mutations responsible for Epithelial–Stromal Corneal Dystrophies have been identified (Klintworth G. , 2003). More than 39 mutation variants of the *TGFBI* gene have been associated with different corneal dystrophy phenotypes (Weiss, Muller, Lisch, *et al*., 2008). Granular corneal dystrophy (GCD) is one of the phenotypic subtypes described to be caused by mutations in the *TGFBI* gene (Klintworth G. , 2003). At least four variants of granular corneal dystrophy have been reported to date (Klintworth G. , 2003). These are GCD Type I (GCD1) (classic), GCD Type II (GCD2) (Avellino), GCD Type III (GCD3) (superficial), and GCD Type IV (GCD4).

The classic GCD1 phenotype is most commonly associated with the heterozygous replacement of the amino acid arginine with tryptophan at codon 555 (R555W). However, even though the diagnosis of some CD conditions on the bases of clinical features alone might seem simple, the actual underlying genotype might not be the one expected since not all affected patients exhibit a clear genotype-phenotype correlation (Hou, Hsiao, Chen, *et al*., 2012) (Kattan, Serna-Ojeda, Sharma, *et al*., 2017). In fact, population studies carried out in different countries and on different ethnicities have led to the identification of a number of genetic mutations that also cause granular deposits (Solari, Ventura, Perez, *et al*., 2007) (Ha, Bae, Yeo, *et al*., 2003) (Frising, Wildhardt, Frisch, *et al*., 2006). Moreover, individuals that have the R555W

mutation, can still exhibit phenotypic variability. In fact, it has been noted that variations in symptoms, progression and clinical manifestations of patients with GCD1 can vary not only between different families but also among members of the same family (Hou, Hsiao, Chen, *et al*., 2012) (Kattan, Serna-Ojeda, Sharma, *et al*., 2017). Thus, researchers are increasingly recommending that genetic analysis is carried out on each individual exhibiting CD changes, due to the genetic and phenotypic heterogeneity exhibited in *TGFBI* CDs (Frising, Wildhardt, Frisch, *et al*., 2006).

The second phase of this project consequently focused on determining the molecular genetics present in a cohort of Maltese patients, who phenotypically exhibited the GCD1 phenotype, by performing clinical exome sequencing. Besides uncovering the causative variant in the *TGFBI* gene present in this Maltese cohort, this part of the study analysed any other genetic mutation variants present in these individuals. The data obtained was then filtered against a list of genes that have been documented in literature to be either co-expressed, known to interact genetically or are involved in common pathways with the *TGFBI* gene or GCD1 deposits. Hypothetically, the resulting genetic mutation variants identified might play a role in the phenotype exhibited in this particular cohort of GCD1 Maltese individuals. Theoretically it is possible that genetic variants causing alterations in the ECM or modulation of *TGFBI* expression might be different in individual patients and families, thus generating variations in the onset, progression and/or severity of their condition. This would explain the intra and inter-familial phenotypic variations documented to have been observed numerous times in literature (Hou, Hsiao, Chen, *et al*., 2012) (Kattan, Serna-Ojeda, Sharma, *et al*., 2017).

In order to understand better which pathways or molecules are involved in the modulation of *TGFBI* expression, during the third phase of this project, a scoping review of related literature was performed. In literature it is reported that *TGFBI* expression is controlled by the transforming growth factor-beta (TGF-β) signalling pathway in certain cell types, such as corneal fibroblasts (Choi, Lee, Jeong, *et al*., 2010) (Figure 1). TGF-β is a cytokine that regulates cellular processes in many organs of the body, including the cornea (Tandon, Tovey, Sharma, *et al*., 2010). It is known that TGF-β in the cornea is important in maintaining corneal integrity and wound healing. It transduces its signal across the plasma membrane by binding to the TGF-β cell-surface receptors type I and II (TGF-βR1 and TGF-βR2), which are serine/threonine kinases. This leads to phosphorylation of the downstream transcription factors SMAD2 and SMAD3 (Luo, 2017). SMAD2 and SMAD3 are receptor regulated proteins (R-SMADs) that are ligand-specific and become phosphorylated, after directly interacting with activated TGF-βR1, by the TGF-βR1 kinase (ALK5) (Shen, Hu, Liberati, *et al*., 1998). SMADs are usually present in the cell cytoplasm in their basal state waiting to be exposed to activated receptors. SMAD2 and SMAD3 bind to the SMAD4 protein to form a heterotrimeric protein complex which then migrates to the cell nucleus. Here, SMAD3 interacts with the transcriptional coactivator p300 to modulate transcription of TGF-β target genes (Tandon, Tovey, Sharma, *et al*., 2010), such as *TGFBI* (Figure 1).

The Jun N-terminal Kinase (JNK) signalling cascades have been shown to perform crosstalk with signalling pathways such as the TGF-β/SMAD signalling pathway (Liu, Sheng, Peng, *et al*., 2018). In fact, it was shown that activation of JNK promoted phosphorylation of SMAD2 and SMAD3 proteins (Nie, Liu, Wang, *et al*., 2020). This subsequently caused accumulation of TGFBIp in corneal fibroblasts (Nie, Liu, Wang, *et*

al., 2020). Apart from the JNK cascade, two pathways that also affect the transcription of TGFBIp indirectly are the Phosphatidylinositol 3-kinase/Ak strain Transforming (PI3-K/AKT) and the cyclic Adenosine Monophosphate/Protein Kinase A (cAMP/PKA) cascades (Figure 1). The intracellular protein cAMP-Response Element Binding Protein (CREB) can be directly activated through these two pathways by inducing CREB serine133 phosphorylation. Both AKT and PKA also increase phosphorylated CREB indirectly through inhibition of Glycogen Synthase Kinase-3 (GSK-3), which normally inhibits phosphorylation of CREB. Phosphorylated CREB binds with p300 that has been shown to mediate transcription of proteins such as Bcl-2 and brain-derived neurotrophic factor (BDNF) (Shen, Hu, Liberati, *et al*., 1998). (BDNF activates the PI3-K/AKT signalling cascade). Thus, phosphorylated CREB competes with the SMAD2/3/4 heterotrimeric complex for p300 resulting in the blockade of SMAD-dependent transcription (Piersma, Bank & Boersema, 2015).

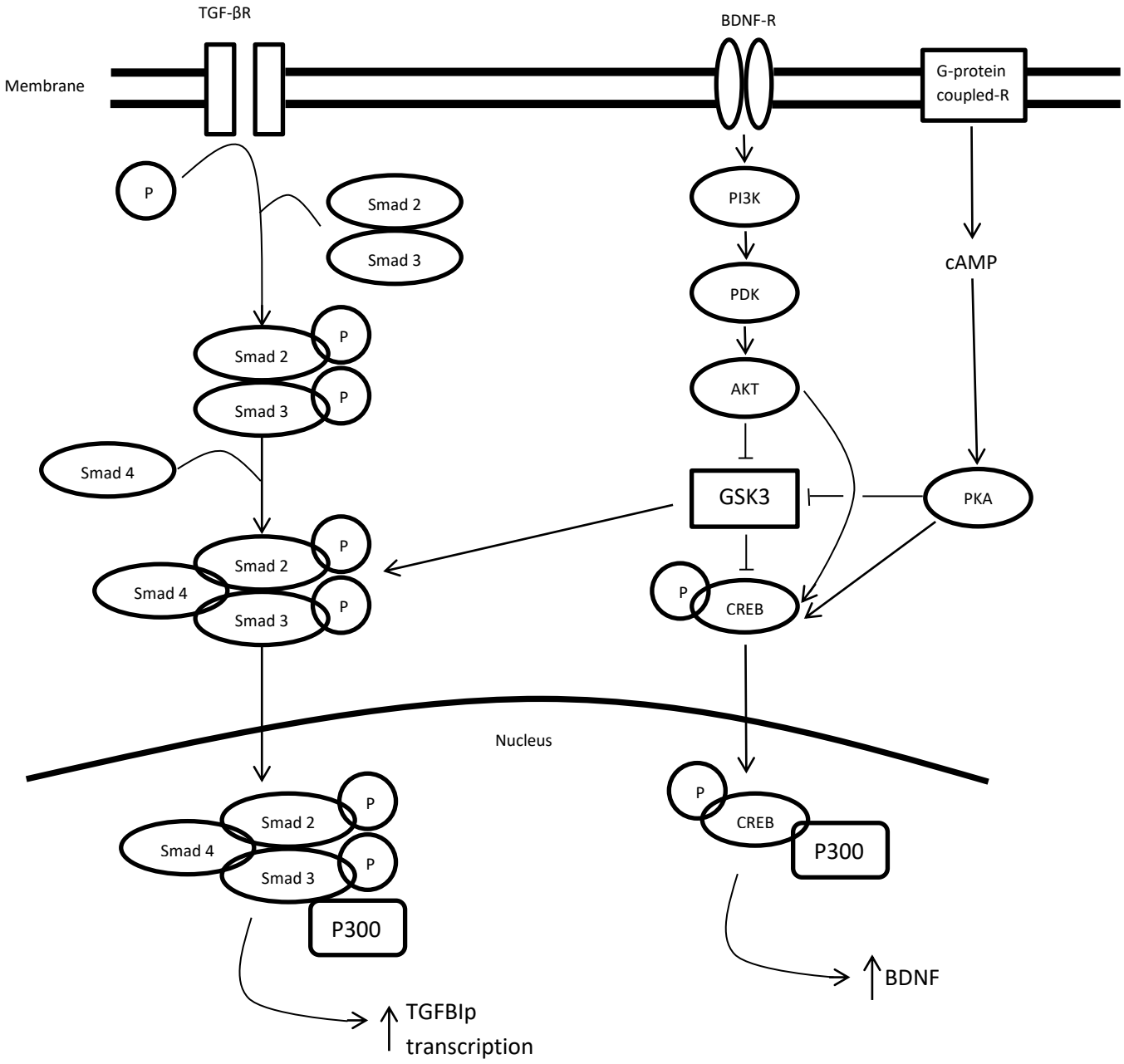


Figure 1: Modulation of TGFBIp expression by the TGF-β signalling pathway. Transcription of TGFBIp is also modulated indirectly by the PI3-K/AKT and the cAMP/PKA cascades.

The majority of cells in an organism have exactly the same genome. However, different types of cells have different combinations of genes that are expressed or suppressed, resulting in their specific structure and function. The gene expression pattern of the corneal epithelium is mirrored in the unique characteristics of this avascular and serum-free tissue. Research focussing on these unique properties will lead to the understanding of the pathophysiology involved in the clinical manifestations of diseases such as the genetically inherited *TGFBI* CDs, considering the fact that the corneal epithelium is the layer of the cornea which almost exclusively transcribes *TGFBI* (Hartz & McKusick, 2015) (Stenvag, Andreasen, Enghild, *et al*., 2013).

Several researchers have attempted to study the gene expression profile of human corneal epithelial cells. However, this was limited by the technology available at the time. The methods previously available, such as microarrays, had a limited range of detection, produced cross-hybridisation artifacts and could only detect sequences that are already known. The recent introduction of RNA sequencing (RNA-Seq) makes it possible to detect and quantify absolute RNA expression values in a tissue sample at a given point in time. This allows us to investigate which genes are turned on or switched off in different kinds of cells leading to further understanding of the various functions of different tissues. Prior to the development of RNA sequencing, quantification of gene expression and transcript identification were performed as two distinct processes. RNA-Seq utilises high-throughput next-generation sequencing in order to analyse the full transcriptome of a biological sample at a given moment in time (Kukurba & Montgomery, 2015), combining these two procedures in a single assay (Conesa, Madrigal & Tarazona, *et al*., 2016).

A disturbance in RNA transcription and processing can lead to aberrant protein expression and synthesis that in turn lead to a disease state. RNA-seq analysis results can in turn provide information about gene expression, transcript isoforms, post-transcriptional modifications, single nucleotide variants and mutations, making it possible to detect differential expression that occurs in certain diseases or treatment groups. You, Corley, Wen, *et al*. (2018) utilized RNA-seq to compare the gene expression profile of the corneal epithelium in keratoconus and myopia patients. Even more recently, Lin, Tan, Liu, *et al*. (2022) performed RNA-seq on telomerase-immortalized human corneal epithelial cell line (hTCEpi) while exploring the effect of SPARC treatment on the transcriptome of human corneal epithelial cells.

It has been documented that TGF-β-target protein expression is differentially regulated in corneal epithelial cells and fibroblasts (Guo, Hutcheon, Tran, *et al*., 2017). Nonetheless, research exploring the pathogenesis of *TGFBI* CDs, and the effect of compounds on TGFBIp production, have been largely carried out on corneal stromal fibroblasts. In view of these documented discoveries, as has been alluded to already by Poulsen *et al*., ideally when carrying out studies exploring corneal *TGFBI* pathways and its modulation, it would be best to target corneal epithelial cells and not fibroblasts in order to simulate better corneal function and produce the closest biologically comparable results (Poulsen, Runager, Nielsen, *et al*., 2018). Thus, in this third phase of the study, in order to obtain a better understanding of the role TGFBIp plays in human corneal epithelial cells (HCECs), the effect of knocking down *TGFBI* was explored. A lentiviral vector containing shRNA targeting *TGFBI* was constructed and viral transduction was then carried out on HCECs in vitro. To the best of our knowledge, to date, this is the first comparative transcriptomic study that has been carried out

that explores the differential expression of genes in *TGFBI* KD HCECs using RNA-seq generated gene expression data.

Present research is focussing on discovering more permanent and less invasive treatments to try to prevent or stop the deposition of corneal material in these dystrophies since, currently, there is no effective treatment to prevent, halt, or reverse the deposition of mutant TGFBIp. Presently, *TGFBI* CD patients with significant visual impairment are offered corneal graft surgery, the outcome of which is temporary since recurrence is high (Lyons, McCartney, Kirkness, *et al*., 1994). In fact, gene therapy is currently viewed as the most promising method for the treatment of rare hereditary genetic diseases. Having said that, certain hurdles that have been encountered during the development of this concept are still being addressed. The discovery that RNA interference (RNAi) can be used as a tool to selectively suppress gene expression in mammalian cells is a recent breakthrough that can be applied in the process of investigating gene function and possibly treat disease by silencing the causal mutant gene or allele (Leung & Whittaker, 2005).

Pharmacotherapy is also being explored as an alternative treatment option for *TGFBI* CDs. As mentioned above, TGFBIp in the cornea is mainly produced by the corneal epithelium (Escribano, Hernando, Ghosh, *et al*., 1994) (Akhtar, Meek, Ridgway, *et al*., 1999) (Nielsen, Poulsen, Lukassen, *et al*., 2020) and is subsequently secreted and then transported from the epithelium into the corneal stroma (Malkondu, Arıkoğlu, Erkoç, *et al*., 2020) (Niu, Liu, Liu, *et al*., 2012) (Poulsen, Runager, Nielsen, *et al*., 2018). However, research exploring the effect of compounds on TGFBIp production with the aim of discovering novel treatment options have also been largely carried out on corneal stromal fibroblasts. In fact, a scoping literature review, which was carried out

in this third phase of the study, to identify and catagorize compounds that reduce corneal TGFBIp levels showed that corneal fibroblasts/keratocytes had been used in 17 of the 19 relevant studies included while corneal epithelial cells were used only in one of these studies (Kim, Luo, Pflugfelder, *et al*., 2005). It has been reported that, in corneal fibroblasts, lithium (Li) suppresses SMAD3/4-dependent gene transcription leading to downregulation of TGFBIp expression by inhibiting GSK3 directly and indirectly (Nie, Peng, Li, *et al*., 2018) (Choi, Kim, Dadakhujaev, *et al*., 2011) (Liang, Wendland & Chuang, 2008). On the other hand, mitomycin C (MMC) is known to inhibit selectively the synthesis of DNA (Barnett & Brundage, 2010) and it was shown to reduce *TGFBI* messenger RNA (mRNA) levels, as well as TGFBIp levels, in GCD2 keratocytes, in a study performed by Kim *et al.* 2008. However, since it has been proven that corneal epithelial cells and fibroblasts use alternative pathways for TGF-β-target protein expression, it would be best to target corneal epithelial cells not fibroblasts when exploring potential new therapeutic options for *TGFBI* CDs. In order to address this literature gap, in the third and last phase of this research project, cultured HCECs were treated with Li and MMC, and RNA-seq and differential expression analysis were subsequently performed on these cells. The results obtained provide a comprehensive overview of the effects of these two compounds on HCECs.

Minimizing surgical interventions and evading their potential complications would definitely benefit *TGFBI* CD patients. Future research focussing on discovering more permanent and less invasive procedures to try and prevent or stop the deposition of corneal material in these dystrophies is the way forward.

CHAPTER 2: LITERATURE REVIEW

2.1 The Eye: Gross Anatomy and Function

The human eye is a sense organ that perceives light. The outer part of the globe can be described as being made up of two units: a large white spherical unit with a radius of about 12mm referred to as the sclera and a smaller spherical transparent frontal unit measuring about 11mm horizontally, called the cornea. At the junction of the cornea and sclera, there is a ring, connecting these two units, referred to as the limbus. The cornea and the lens provide the refractive power needed for the focusing of light rays onto the retina (American Academy of Ophthalmology, 2010). The inner layer of the sclera is coated by the retina, which is the visual apparatus of the eye, and its vasculature. The central region of the retina, the macula, is a highly specialized structure that allows high acuity.

Through evolution, the ocular structure has developed into a highly intricate organ that allows the perception of light and subsequently image formation. Any disruption, however minor, in the functioning of this organ will lead to visual deterioration. Currently, 285 million people are estimated to be visually impaired worldwide. Of these, 39 million are classified as blind while 246 have low vision. The leading cause of visual impairment in the world still remains cataract. Other major pathologies causing visual impairment include glaucoma, age-related macular degeneration (AMD), corneal opacities and diabetic retinopathy (WHO, 2016).

2.2 Embryology of the Human Eye with Focus on the Cornea

The eye starts to form in the fourth week of human embryonic development. The first indication is the presence of the optic sulci in the inner aspect of the neural folds.

This is followed by an evagination of each lateral wall of the forebrain (that forms from the neural folds), which goes on to form a pair of optic vesicles. The vesicles make contact with the surface ectoderm and this causes the ectoderm to become thickened forming the lens placode. Just beneath the thickening in the surface ectoderm, a neural ectodermal thickening is present. This is called the retinal disk that will eventually give rise to the retina. The lens placode subsequently invaginates and becomes completely enclosed to form the lens vesicle. When the overlying surface ectoderm becomes completely separated from the lens vesicle, the space separating them is invaded by mesenchymal cells of neural crest origin to form the corneal endothelium. A few days later, a second wave of mesenchymal cells infiltrates the space between the corneal epithelium and the endothelium in order to give rise to the corneal stroma. The surface ectoderm eventually forms the corneal epithelium and the skin of the eyelid. Thus, in a nutshell, the cornea is derived from various cells, namely, the surface ectoderm that gives rise to the epithelium and neural crest-derived mesenchyme which gives rise to the stroma, Bowman's and Decemet's layers, as well as the endothelium (Forrester, Dick, McMenamin, *et al*., 2016).

2.3 The Tear Film Structure and Function

The corneal surface is covered by a three layered structure known as the tear film (Wolff, 1954). It consists of an outer lipid layer, a middle aqueous layer and an innermost mucus layer that overlies the corneal epithelium (Knop, Knop, Millar, *et al*., 2011). The lipid layer slows the rate of tear evaporation from the corneal surface. It also acts as a hydrophobic barrier to prevent the overspill of tears and contamination by skin lipids which might cause destabilisation of the tear film (Bron, Tiffany, Gouveia, *et al*., 2004). The meibomian glands located in the tarsal plates of the eyelids are

responsible for the production phospholipids, sphingolipids, triglycerides and esters that compose this layer (McCulley & Shine, 1997). Lipid secretion is controlled by neural and hormonal stimuli, supplemented by lid action.

The aqueous layer is mainly composed of water that helps in the spreading of the tear film. It is produced mainly by the lacrimal glands.

On the other hand, the epithelial glycocalyx, also known as the mucus layer, serves as an anchor for the tear film, helping it to adhere to the surface of the eye. It is composed of a fibrillar glycoprotein produced by cells in the conjunctiva, known as goblet cells (Watanabe, Fabricant, Tisdale, *et al*., 1995) and coats the outer surface of the epithelial cell microprojections. The glycocalyx also acts as a barrier to potential pathogens by repelling bacteria which have the surface electrical charge of the same polarity. Furthermore, the mucins can bind with potentially harmful contaminants to act as a debris removal system (Gipson & Argueso, 2003) (Hodges & Dartt, 2013).

The tear film also contains substances and cells that protect the eye from infections. These include lysozyme (Fleming, 1922), lactoferrin, complement and anti-complement factor, interferon, immunoglobulins (Little, Centifanto & Kaufman, 1969) and lymphocytes. The function of lysozyme is to lyse glycosaminoglycans and plays a role in immunoglobulin A bacteriolysis, in the presence of complement factor. Lactoferrin chelates iron, thus depriving microorganisms from this substance. It is also thought to enhance the action of lysozyme (McDermott, 2013). Immunoglobulin G, which is also found in the tear film, plays a role in complement mediated bacteriolysis and also promotes phagocytosis.

Most of the refraction of light that passes through the cornea occurs at the tear-air interface. Besides providing protection to the ocular surface, the tear film produces a smooth optical surface for light to pass through the eye. An unstable tear film degrades the quality of the images formed on the retina and thus reduces visual quality.

2.4 Corneal Structure and Function

The cornea is a transparent structure which, besides acting as protective cover to the anterior part of the eye, also provides approximately two-thirds of the total optical refractive power of the eye (Land & Ferdinand, 1992). It also absorbs more than 70% of the harmful solar UV radiation that would otherwise cause oxidative damage to the ocular tissues (Kolozsvari, Hopp & Bor, 2002). It is convex on the outside and concave on the inside and measures approximately 11 to 12mm horizontally and 10 to 11mm vertically (American Academy of Ophthalmology, 2010). Its thickness varies, being 0.5 to 0.6mm centrally and increasing to 0.6 to 0.8mm peripherally (Wikipedia, 2016). The cornea is aspheric with the anterior radius of curvature being about 7.84mm and the posterior radius of curvature being 6.4mm (Vojnikovic, Gabric, Dekaris, *et al*., 2013). Its refractive index is 1.376.

The cornea is made up of water, proteins, glycosaminoglycans and salts. The proteins present in the cornea are mainly collagen type I, V and VI. The cornea contains the proteoglycans keratin sulphate and chondroitin sulphate that are made up of a core protein and covalently attached glycosaminoglycan chains. These help in regulating corneal hydration (Yanoff & Ducker, 2009).

The cornea is a very particular tissue, in that it is avascular, transparent and exhibits immunologic privilege. Its transparency is imperative for one of its main functions, that is, refraction of light that is needed to produce a clear image on the retina. Lack of corneal transparency would result in visual impairment. The clarity of the cornea depends on multiple factors (Meek & Knupp, 2015). Epithelial cells, being tightly packed, produce a layer that has a uniform refractive index and minimal light scattering. Stromal collagen fibrils are arranged in a lattice pattern in order to cause destructive interference of light and thus also reducing scattering of light. Having a water content of 78% in the stroma is also an important factor in keeping the cornea clear. This depends on having an intact epithelial layer as well as healthy endothelial cells (American Academy of Ophthalmology, 2010). The endothelial cell layers' function is to keep the cornea in a relatively dehydrated state, thus preventing corneal clouding (Schmedt, Mazzi Silva, Ziaei, *et al*., 2012).

The structure of the corneal basement membranes and stroma is such that it can resist forces inherent to the eye, such as intraocular pressure (IOP) and lamellar tension from the surrounding scleral tissue (Dupps Jr. & Wilson, 2006). It has been shown that the corneal basement membranes are the stiffest layers of the cornea, thus helping in maintaining the corneal shape (Last, Thomasy, Croasdale, *et al*., 2012). Any disruption in these membranes causes thinning and flattening of the cornea, which, as a result, affects the refractive properties of this tissue (Dupps Jr. & Wilson, 2006). Investigating ways to enhance the regenerative capacity of these membranes through targeted drug delivery or stem cell-based approaches could hold promise for improving outcomes in corneal disorders associated with basement membrane dysfunction.

2.4.1 The Layers of the Cornea

The cornea is made up of six distinct layers; the epithelium, the Bowman's layer, the stroma, Dua's layer, the Descemet's membrane and the endothelium ((Forrester, Dick, McMenamin, *et al*., 2016). Each of these is vital in the maintenance of corneal structure and function.

2.4.1.1 The Epithelium

The epithelium, which is the outer most layer of the cornea, is 50 micrometers thick and consists of five to seven layers of cells of a stratified squamous morphology. It continues at the limbus as the conjunctival epithelium. The most superficial cells are flat, while, the innermost cells of this epithelial layer are columnar. Its structure helps to maintain a smooth surface to minimize light scatter (Lu, Reinach & Kao, 2001). The superficial cells have microvilli on the outer surface that adhere to glycoproteins that form the mucin layer of the tear film. There are tight junctions between superficial epithelial cells that prevent tears and noxious agents from penetrating into the corneal stroma (American Academy of Ophthalmology, 2010). The cytoplasm of corneal epithelial cells has been found to be abundant in intermediate filaments (diameter of 8–10 nm), which are also known as tonofilaments. They are composed of keratin proteins (Sun & Vidrich, 1981) which connect neighbouring cells to one another and to the extracellular matrix, providing great tensile strength to the epithelial layer. Cytokeratins 3 and 12 are cornea-specific (Cooper, Schermer & Sun, 1985). Thus, their presence in stromal deposits in certain corneal dystrophies points towards the theory that these dystrophies are of epithelial origin (Wollensak & Witschel, 1996).

The basal columnar cells at the limbus (the border between the transparent cornea and the sclera), constitute the germinative layer of the epithelium which replaces the

cells that are constantly shed in the superficial exposed layer. There are no germinative stem cells in the basal epithelial layer of the central cornea. However, the slowly dividing limbal cells (Arpitha, Prajna, Srinivasan, *et al*., 2008) produce transiently amplifying cells, which migrate towards the centre of the cornea. These cells continue dividing and move progressively to the surface layer of the epithelium while becoming more differentiated (Davanger & Evensen, 1971) (Buschke, Friedenwald & Fleischmann, 1943) (Hanna & O'Brien, 1960).

Basal cells are attached together by desmosomes while hemidesmosomes attach these cells to the underlying basement membrane. Hemidesmosomes protect the corneal epithelium from detaching from the basement membrane ((Khodadoust, Silvestein, Kenyon, *et al*., 1968) (Gipson, Spurr-Michaud & Tisdale, 1988). Basal epithelial cells secrete the components of an underlying 50-nanometer-thick basement membrane that is composed mainly of collagens, laminins, heparan sulphate proteoglycans (HSPGs), and nidogens (Tuori, Uusitalo, Burgeson, *et al*., 1996). This basement membrane consists of two layers, the lamina lucida and the lamina densa. The lamina lucida is composed of kalinin filaments and laminin. The lamina densa contains HPSG and globular Type VII collagen. The latter is attached to Bowman's layers' anchoring plaques via anchoring fibrils that consist of helical collagen VII and glycoproteins.

Corneal epithelial basement membrane (BM) appears to have horizontal regional heterogeneity with regards to composition between the central cornea, limbus and conjunctiva. Six different collagen type IV alpha chains have been recognised. The central BM in human corneas has type IV collagen alpha3 though alpha6 chains, while

alpha1 and alpha2 chains are found only in limbal BM, in addition to alpha5 and alpha6 chains (Ljubimov, Burgeson, Butkowski, *et al*., 1995). The limbal BM also contains laminin alpha2 and beta2 chains, while the central corneal BM does not.

Studies carried out by Yurchenco *et al*. have suggested that laminins are responsible for the assembly of basement membranes (Yurchenco, 2011) (Yurchenco, Cheng & Colognato, 1992). Perlecan is the most prevalent HSPG in basement membranes and is involved in proliferation, migration and differentiation of a various cells by mediating cell signalling events (Mongiat, Taylor, Otto, *et al*., 2000). It controls the availability of growth factors, including fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), vascular endothelial growth factors (VEGF), transforming growth factor β-1 (TGF-β1), and insulin-like growth factors (IGF), to bind to receptors (Whitelock, Melrose & Iozzo, 2008) (Iozzo, 2005) (Nikolova, Strilic & Lammert, 2007). Studies carried out by Vittitow *et al*. reported that upregulation of perlecan expression occurs after corneal stromal injury, as well as after an artificial increase in intraocular pressure (Vittitow & Borras, 2004).

Injury to the cornea induces the appearance of stromal myofibroblasts and altered structure of the extracellular matrix that in turn leads to corneal opacification (Jester, Huang, Barry-Lane, *et al*., 1999) (Wilson, Liu & Mohan, 1999). The epithelial BM plays an important role in corneal wound healing. In fact, the corneal wound healing response, including keratocyte apoptosis and corneal nerve death, was found to be enhanced when the epithelial BM is removed ((Pal-Ghosh , Pajoohesh-Ganji, Tadvalkar, *et al*., 2011). Recent studies have shown that a breach in the continuity of the epithelial BM allows cytokines such as epithelium-derived TGF-β1 and PDGF to reach

the stroma. These subsequently trigger the differentiation of keratocytes to myofibroblasts. Stromal myofibroblasts are dependent on TGF-β1 for survival since there are studies proving that after the epithelial BM is repaired, resulting in a fall in levels of TGF-β1 and PDGF in the corneal stroma, apoptosis of myofibroblasts occurs (Singh, Santhiago, Barbosa, *et al*., 2011) (Kaur, Chaurasia, Agrawal, *et al*., 2009).

1.4.1.2 Bowman's Layer

Bowman's membrane is a 10 micrometers thick acellular layer, composed mainly of type I and V collagen and proteoglycans. It underlies the epithelial layer and is actually a modified anterior stromal layer. It protects the cornea by acting as a physical barrier against infections and trauma as well as preventing cytokines secreted by the epithelial layer from activating stromal keratocytes. Trauma to Bowman's layer results in it being replaced by scar tissue (Duke-Elder & Wybar, 1960).

1.4.1.3 The Corneal Stroma

The corneal stroma is about 500 micrometers thick and comprises 90% of the thickness of the cornea. It is made up of regularly arranged collagen fibres embedded in proteoglycans, with sparse interconnected keratocytes. Collagen type I and collagen type V are found interwoven with collagen type VI. It is thought that the fiber bundle size is regulated by the fact that the N-terminal of collagen-V is perpendicular to the fibril axis, protruding to the surface and sterically preventing further addition of collagen-I fibrils (Linsenmayer, Gibney, Igoe, *et al*., 1993). Collagen type III and IV have also been detected in the corneal stroma (Agarwal, Apply, Agarwal, *et al*., 2002). There is variation in the concentrations and the ratio of proteoglycans in the corneal stroma from anterior to posterior, making the posterior part of the stroma "wetter" than the anterior. There is also a variation in the length and width of the lamellae from anterior

to posterior, the anterior lamellae being short and narrow with extensive interweaving, whereas the posterior lamellae are long and wide and extend from limbus to limbus (American Academy of Ophthalmology, 2010). The major glycosaminoglycans making up the extracellular matrix are keratan and chondroitin sulphate. Keratan sulphate is found in a higher concentration in the central stroma while chondroitin sulphate is mostly present in the peripheral cornea ((Muller, Pels, Schurmans, *et al*., 2004).

The regular hexagonal lattice arrangement of these collagen fibrils running parallel to each other and to the corneal surface, with spacing between the fibrils being <200nanometers between each other, causes destructive interference of light waves thus keeping the cornea transparent (Maurice D. , 1957) ((Cox, Farrell, Hart, *et al*., 1970). This occurs since collagen type I has intrinsic light scattering properties. Also, its arrangement in an organized lattice causes light to be scattered in relative positive and negative directions causing destructive interference. The extracellular matrix surrounding the collagen lamellae helps in maintaining the distance between the fibrils (Kuwabara, 1978). Corneal transparency is also possible because the size of the collagen lattice components is smaller than the wavelength of visible light (Doutch, Quantock, Smith, *et al*., 2008).

The stroma contains semi-quiescent keratocytes that produce collagen and extracellular matrix. Intercellular communication between neighbouring keratocytes is important for their cell function (Assouline, Chew, Thompson, *et al*., 1992). They also play an important role in wound healing which is mediated via growth factor receptors. Stromal injury results in activation, migration and transformation of keratocytes into myofibroblasts that form scar tissue. Scar tissue is made up of larger diameter collagen

fibrils that are arranged in a less regular fashion resulting in corneal opacification. Subsequently, growth factors and cytokines mediate slow remodelling of the scar thus restoring, to a certain extent, corneal transparency (Imanishi, Kamiyama, Iguchi, *et al*., 2000). However, in certain pathologic conditions, deposition of material and disruption of the organisation of the collagen matrix causes corneal opacities that can even lead to blindness due to the obstruction of light through the cornea.

1.4.1.4 Dua's Layer

Recently, a novel corneal layer that lies between the stroma and Descement's membrane has been identified. It is acellular and is made up of five to eight lamellae of type-1 collagen. The spacing between the bundles of collagen is similar to that of the stroma (Dua, Faraj, Said, *et al*., 2013).

1.4.1.5 Descemet's Layer

The Descemet's layer is 10 micrometers thick and is also referred to as the posterior corneal basement membrane. It consists mainly of collagen type IV and laminin. It is secreted by the endothelial corneal cells and is made up of two layers: the anterior banded layer and the posterior nonbanded layer. The former layer is present before birth while the latter nonbanded amorphous layer is laid down slowly by the endothelium throughout life and consists of fibrillogranular material (Agarwal, Apply, Agarwal, *et al*., 2002).

1.4.1.6 The Endothelium

The endothelium is the most posterior layer of the cornea that consists of a monolayer of hexagonal cells. It is 5 micrometers thick and is considered to be amitotic, thus, the number of endothelial cells decreases with age while the size of the individual endothelial cells increases (Dua, Faraj, Said, *et al*., 1984). This mechanism is

a way of overcoming a certain degree of endothelial cell loss, however, there is a minimum threshold of endothelial cell density below which the endothelium fails to maintain its function and corneal decompensation, leading to corneal oedema, occurs (Arici, Sevki Arslan & Dikkaya, 2014). The corneal endothelium can be said to have two main functions. It contains an energy dependent Na, K-ATPase pump that controls corneal stromal hydration and electrolyte balance. Hence, its main function is to keep the cornea relatively low in water (Fischbarg, Diecke, Iserovich, *et al*., 2006). Besides this, it also functions as a physiological barrier to the aqueous. Thus, any disruption of this layer will result in swelling of the stroma and modification of the organization of the stromal matrix, causing corneal opacity.

2.4.2 Corneal Innervation

The cornea is one of the most highly innervated tissues in the body. It is supplied by the ophthalmic division of the trigeminal nerve that branches into 70–80 long ciliary nerves (Muller, 2003). The nerves enter the cornea in an unmyelinated form in the centre of the peripheral stroma (Muller, 1997). They subdivide to form three networks, namely, midstromal, subepithelial/sub-basal, and epithelial. Each nerve ending has a large receptive field and these may overlap (Cruzat, Qazi & Hamrah, 2017).

The density of the pain receptors in the cornea is 300-600 times greater than that of the skin (Belmonte & Gallar, 1996). This is why a defect in the corneal epithelial integrity causes severe pain (Karmel, 2010). Corneal nerves also help in the maintenance of the ocular epithelial integrity since they play a role in reflex blinking, tearing and the release of trophic substances that promote corneal epithelial homeostasis (Yang, Chow & Liu, 2018) (Marfut, Cox, Deek, *et al*., 2010).

2.4.3 Corneal Transparency

As mentioned above, transparency of the cornea is vital for transmittance and refraction of light to produce a clear image on the retina. Besides being avascular and being supplied by unmyelinated nerve fibers, the clarity of the cornea depends on multiple factors that keep light scattering to a minimum (Meek & Knupp, 2015). Corneal transparency is achieved due to the structural organisation of the tightly packed epithelial cells and the regular hexagonal lattice arrangement of the stromal collagen fibers. Two theories, both based on the particular arrangement of the stromal collagen lamellae, have been put forward. These are the Theory of Goldman *et al*. and the Maurice theory. Goldman *et al*. postulated that the collagen fibrils are small, and the spaces between them are short, in relation to the wavelength of visible light. Thus, these do not interfere with light transmission since they are less than one half a wavelength of light (Goldman & Benedek, 1967). On the other hand Maurice put forward the 'lattice theory' that attributed corneal transparency to the hexagonal pattern of arrangement of the collagen fibers within the lamellae (Maurice D. , 1962).

Corneal transparency also relies on maintaining a water hydration level of 78% in the stroma. In order to prevent corneal oedema, the endothelial cell layer contains a 'pump-leak' mechanism that is responsible for keeping the cornea in a relatively dehydrated state (Bonanno, 2012), (Schmedt, Mazzi Silva, Ziaei, *et al*., 2012). It has been suggested that corneal oedema leads to loss of transparency due to the alteration in the spaces between the collagen fibers that in turn, cause, scattering of light (Meek K. , 2003).

An important group of conditions that cause decreased transparency of the cornea and subsequently lead to significant visual impairment are the corneal dystrophies, which will be discussed next.

2.5 Overview of Corneal Dystrophies

2.5.1 Definition of Corneal Dystrophy

The word “dystrophy” is generally used to describe a disorder in which an organ or tissue of the body wastes away (Oxford, 2016). Corneal dystrophies (CDs) are a group of rare, genetically inherited disorders that cause a loss in corneal transparency with subsequent visual impairment and occasionally corneal blindness (Klinthworth, 2009) (Bron A. J., 1990).

CDs are characterised by non-inflammatory, bilateral, often progressive alterations of the cornea (Weiss, Moller, Lisch, *et al*., 2008). Abnormal material often accumulates in the cornea resulting in corneal opacification due to decompensation, deposition, or scarring. Since the cornea is a clear tissue, deposition of abnormal material compromises its normal functionality with consequent decrease in vision (Klinthworth, 2009) (Bron A. J., 1990) (Glavini, 2013).

According to Weiss, Moller, Aldave, *et al*. (2015), the term “corneal dystrophy” was first mentioned in literature in 1890 by Groenouw, when describing two conditions that were later identified as GCD and macular corneal dystrophy (MCD). The same term was later used by Biber, while referring to lattice corneal dystrophy (LCD) and subsequently used by other researchers including Fuchs, Uhthoff and Yoshida (Weiss, Moller, Aldave, *et al*., 2015). Since, in those days, analytical tools were limited, corneal dystrophies were classified mainly by their phenotype. Some were named even prior

to the invention of the first slit lamp biomicroscope, attributed to Allvar Gullstrand, which dates back to 1911 (Klintworth, 2018).

More than 20 subtypes of CDs have been reported in literature. Inheritance is autosomal dominant in most cases, with variations in symptoms, clinical manifestations and progression, not only between different families but occasionally also within the same family (Weiss, Moller, Lisch, *et al*., 2008) (Aldave A. J., 2011) (Vincent A. L., 2014). Despite this, classic phenotypes for the majority of CDs have been defined. Weiss *et al*. pointed out that although the above mentioned generalizations apply to most CDs, some exceptions exist (Weiss, Moller, Lisch, *et al*., 2008). For instance, Posterior Polymorphous dystrophy is typically unilateral, while MCD and Schnyder CD (SCD) can have systemic signs.

Due to the phenotypic variability seen in CDs, as has been mentioned above, molecular genetic techniques have proven to be crucial in the clinical diagnosis of individuals exhibiting these conditions. Location of the specific mutations that are responsible for the pathogenesis of most of these CDs helped to increase our understanding of these conditions. Researchers are increasingly stressing the importance of identifying the specific aetiology of corneal dystrophies, especially with regards to the genetic mutations responsible for these conditions, so as to be able to provide an accurate diagnosis and deliver effective treatment to these patients.

The first part of this research study identified CD-related allelic variants, and their frequencies, present in the Maltese ethnicity. This is the first ever CD study in which genetic variants in genes associated with CDs in a Maltese cohort of individuals were analysed.

2.5.2 Classification

In an attempt to standardise, clarify and classify case reports of patients phenotypically exhibiting CD, an International Committee for Classification of Corneal Dystrophies (IC3D) was set up. Initially, the classification of corneal dystrophies that was compiled was built only on corneal signs (Weiss, Moller, Lisch, *et al*., 2008). However, when genetic analysis became available, the shortcomings of the phenotypic method of corneal dystrophy classification were revealed. Novel mutations, which were continually being identified and documented in international literature, highlighted the importance of performing mutational analysis in the diagnosis of CDs. It was shown that corneal dystrophies show both phenotypic heterogeneity, as well as genotypic heterogeneity (Weiss, Moller, Lisch, *et al*., 2008). In order to incorporate clinical, genetic as well as histopathologic information available regarding each dystrophy, a new revised corneal dystrophy classification by the IC3D was introduced (Weiss, Moller, Aldave, *et al*., 2015). This classification grouped the CDs according to which cellular layer of the cornea is predominantly affected, namely: Epithelial and subepithelial dystrophies, Epithelial-Stromal *TGFBI* dystrophies, Stromal dystrophies, and Endothelial dystrophies (Weiss, Moller, Aldave, *et al*., 2015). This has reclassified the *TGFBI* dystrophies more accurately, recognising that these dystrophies generally affect multiple corneal layers rather than just one layer (Weiss, Moller, Aldave, *et al*., 2015). A summary of the current IC3D classification of CDs as described by the committee is shown in Table 1.

Table 1: Summary of IC3D classification of CDs outlining the category (Table 2 below this), MIM reference number, mode of inheritance and gene and locus of each CD

Corneal Dystrophy Class	Dystrophy	Category	MIM	Mode of Inheritance	Locus	Gene
Epithelial & subepithelial dystrophies	1. Epithelial basement membrane dystrophy (EBMD) (majority degenerative)	rarely C1	#121820	Unknown	5q31	*TGFBI*
	2. Epithelial recurrent erosion dystrophies (EREDs): - Franceschetti corneal dystrophy (FRCD) - Dystrophia Smolandiensis (DS) - Dystrophia Helsinglandica (DH)	C3	#122400	AD	Unknown	Unknown
	3. Subepithelial mucinous corneal dystrophy (SMCD)	C4	#612867	AD (may be X-linked)	Unknown	Unknown
	4. Meesmann corneal dystrophy (MECD)	C1	#122100	AD	12q13 17q12	*KRT3* *KRT12*
	5. Lisch epithelial corneal dystrophy (LECD)	C2	#300778	X-chr Dominant	Xp22.3	Unknown
	6. Gelatinous drop-like corneal dystrophy (GDLD)	C1	#204870	AR	1p32	*TACSTD2*
Epithelial–stromal TGFBI dystrophies	1. Reis–Bücklers corneal dystrophy (RBCD)	C1	#608470	AD	5q31	*TGFBI*
	2. Thiel-Behnke corneal dystrophy (TBCD)	C1	#602082	AD	5q31	*TGFBI*
	3. Lattice corneal dystrophy, type 1 (LCD1) Variants (III, IIIA, I/IIIA, IV) of lattice corneal dystrophy	C1	#122200	AD	5q31	*TGFBI*
	4. Granular corneal dystrophy, type 1 (GCD1)	C1	#121900	AD	5q31	*TGFBI*
	5. Granular corneal dystrophy, type 2 (GCD2)	C1	#607541	AD	5q31	*TGFBI*
Stromal dystrophies	1. Macular corneal dystrophy (MCD)	C1	#217800	AR	16q22	*CHST6*
	2. Schnyder corneal dystrophy (SCD)	C1	#21800	AD	1p36	*UBIAD1*
	3. Congenital stromal corneal dystrophy (CSCD)	C1	#610048	AD	12q21.33	*DCN*
	4. Fleck corneal dystrophy (FCD)	C1	#121850	AD	2q34	*PIKFYVE*
	5. Posterior amorphous corneal dystrophy (PACD)	C1	#612868	AD	12q21.33	*KERA/LUM/DCN/EPYC*
	6. Central cloudy dystrophy of François (CCDF)	C4	#217600	Unknown	Unknown	Unknown
	7. Pre-Descemet corneal dystrophy (PDCD)	C1/C4	None	AD/Unknown	Xp22.31/Unknown	*STS*/Unknown

Corneal Dystrophy Class	Dystrophy	Category	MIM	Mode of Inheritance	Locus	Gene
Endothelial dystrophies	1. Fuchs endothelial corneal dystrophy (FECD): - Early-onset (FECD1) - Late-onset (FECD 2-8)	 C1 C2/C3	 #136800 #610158 #613267 #613268 #613269 #613270 #613271 #615523	 AD Many unknown	 1p34.3–p32 13pter-q12.13 18q21.2-q21.3 20p13-p12 5q33.1-q35.2 10p11.2 9p24.1-p22.1 15q25	 *COL8A2* *TCF4/LOXHD1* *TCF8* *AGBL1*
	2. Posterior polymorphous corneal dystrophy (PPCD): - PPCD1 -PPCD2 and 3	 C2 C1	 #122000 #609140/1	?AD	 20p11.2-q11.2 1p34.3-p32.3 10p11.22	 Unknown *COL8A2* *ZEB1*
	3. Congenital hereditary endothelial dystrophy (CHED)	C1	#217700	AR	20p13	*SLC4A11*
	4. X-linked endothelial corneal dystrophy (XECD)	C2	None	X-chr Dominant	Xq25	Unknown

Table 2: Evidence levels, divided into categories by the IC3D Committee, supporting the existence of CDs

Category	Evidence level
1	A well-defined corneal dystrophy in which the gene has been mapped and identified and the specific mutations are known.
2	A well-defined corneal dystrophy that has been mapped to one or more specific chromosomal loci, but the gene(s) remains to be identified.
3	A well-defined corneal dystrophy in which the disorder has not yet been mapped to a chromosomal locus.
4	This category is reserved for a suspected, new, or previously documented corneal dystrophy, although the evidence for it, being a distinct entity is not yet convincing.

2.5.2.1 Epithelial and Subepithelial Dystrophies

The Epithelial and Subepithelial CDs affect the outermost layers of the cornea, namely the epithelium and the region of Bowman's membrane. These include: Epithelial basement membrane dystrophy (EBMD) Meesmann dystrophy (MECD), Gelatinous drop-like dystrophy (GDLD) and Lisch epithelial dystrophy (LECD).

2.5.2.1.1 Epithelial Basement Membrane Dystrophy

Epithelial Basement Membrane Dystrophy is also known as 'map-dot-fingerprint dystrophy' and also as 'Cogan's microcystic dystrophy', since it was first described by Cogan in 1964 (Cogan, Donaldson, Kuwabara, *et al*., 1964). The 'maps' are actually extra sheets of basement membrane that extend into the corneal epithelium while the 'fingerprints' are due to lines of thickened basement membrane. The 'dots' are formed by entrapped maturing epithelial cells in these sheets of basement membrane. Symptoms may include blurry vision, foreign body sensation, and pain if the patient develops recurrent corneal erosions (Feldman & Afshari, 2010). Even though this condition is sometimes considered to be degenerative, families with AD mutations as well as sporadic mutations in the *TGFBI* gene have been described (Boutboul, Black, Moore, *et al*., 2006).

2.5.2.1.2 Meesmann Corneal Dystrophy

Meesmann CD is characterized by a fragile corneal epithelium and bilateral multiple microcysts that extend all the way to the limbus. These cysts are usually more abundant in the interpalpebral area of the cornea (Burns, 1968)(Kuwabara & Ciccarelli, 1964). The lesions first appear in infancy with occasional slowly progressive subepithelial scarring. Symptoms include glare, photophobia, irritation, mild blurred vision, and irregular astigmatism (Fine, Yanoff, Pitts, *et al*., 1977). Genotypic

heterogeneity is seen in Meesmann corneal dystrophy since it is associated with both keratin 3 (*KRT3*) and keratin 12 (*KRT12*) genes (Weiss, Moller, Aldave, *et al*., 2015). These genes are specifically expressed in the corneal epithelium (Chen, Lin, Gee, *et al*., 2015).

2.5.2.1.3 Gelatinous Drop-Like Dystrophy

Individuals suffering from Gelatinous drop-like dystrophy develop bilateral central corneal subepithelial and stromal raised gelatinous excrescences, which consist of amyloid. Superficial and deep corneal vascularisation occurs in the areas of opacification which in turn leads to further amyloid deposition that causes severe visual loss (Weber & Babel, 1980). Clinically, patients present with blurred vision, photophobia, and foreign-body sensation due to recurrent corneal epithelial erosions (Ren, Lin & Klintworth, 2002) (Paliwal, Gupta & Tandon, 2010). Mutations in the tumor associated calcium signal transducer 2 (*TACSTD2*) gene have been recognized as being the cause for GDLD (Tsujikawa, Kurahashi & Tanaka, 1999).

2.5.2.1.4 Lisch Epithelial Dystrophy

Lisch epithelial dystrophy is characterised by intraepithelial clear microcysts arranged in a feathery, band-shaped or whorled pattern (Lisch, Steuhl, Lisch, *et al*., 1992). These opacities can cause slowly progressive impaired vision. Only a few cases have been reported to date. This condition appears to be genetic and an X-linked recessive pattern has been proposed (orphanet) with the location of the mutation being mapped to Xp22.3 (Lisch, Buttner, Oeffner, *et al*., 2000).

2.5.2.2 Epithelial–Stromal Transforming Growth Factor Beta Induced Dystrophies

Mutations in the *TGFBI* gene, which is located on chromosome 5q31, are responsible for a number of phenotypically heterogenous CDs that affect multiple

layers of the cornea (Han, Choi, Kim, *et al*., 2016) (Yang, Han, Huang, *et al*., 2010). The association between missense mutations in the *TGFBI* gene and a number of CDs was made by Munier *et al*. in 1997 (Munier, Korvatska, Djemai *et al*., 1997). To date, at least 74 distinct allelic mutations in this gene have been identified (Yang, Han, Huang, *et al*., 2010). These CDs include Granular corneal dystrophy, type 1 (GCD1) (Figure 2A) and type 2 (GCD2), Lattice corneal dystrophy (LCD) (Figure 2B) (except Type II), Reis–Bücklers corneal dystrophy (RBCD) and Thiel-Behnke corneal dystrophy (TBCD) (Weiss, Moller, Aldave, *et al*., 2015) (Klintworth, 2003) (Nielsen, Poulsen, Lukassen, *et al*., 2020). *TGFBI* corneal dystrophies are mostly inherited in an autosomal dominant fashion (Nielsen, Poulsen, Lukassen, *et al*., 2020). A very recent report claimed that there is also one *TGFBI* disease-associated variant that has been documented to be inherited in an autosomal recessive fashion (Kheir, Cortes-Gonzalez, Zenteno, *et al*., 2019).

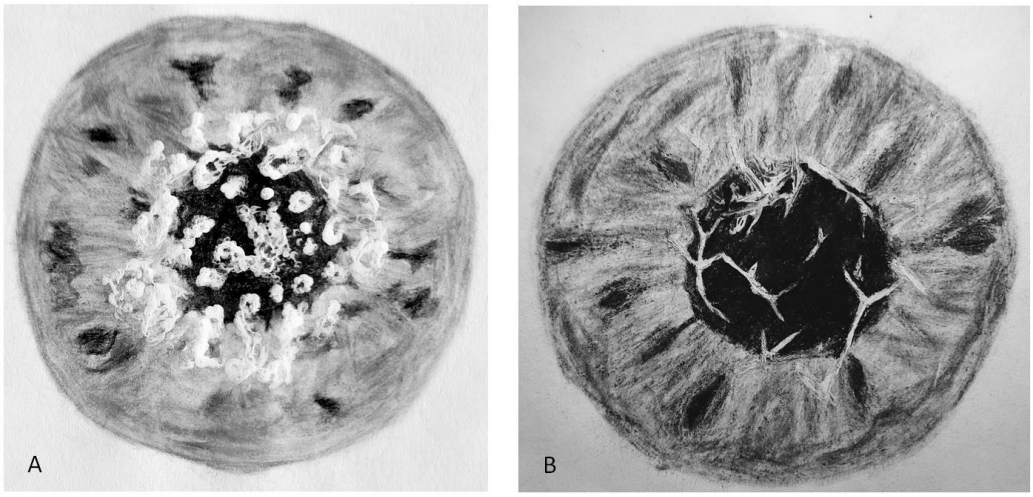


Figure 2: Diagrammatic representation of an eye exhibiting: A) granular corneal dystrophy and B) lattice corneal dystrophy.

2.5.2.2.1 Granular Corneal Dystrophy

At least four variants of GCD have been reported to date. These are GCD1 (classic), GCD2 (Avellino), GCD3 (superficial, aka Reis-Bucklers), and GCD4 (Weiss, Moller, Aldave, *et al*., 2015). It has been suggested that the different clinical phenotypes are due to differences in behaviour of various forms of mutated TGFBIp (Akhtar, Meek, Ridgway, *et al*., 1999). Interestingly, according to Okada *et al*., GCD seems to be "the first ophthalmic disease in which homozygosity for a dominant allele has been genetically identified" (Okada, Yamamoto, Watanabe, *et al*., 1998).

Since this research study focuses on *TGFBI* CDs, particularly on the GCD1 variant, this CD variant is described extensively below.

Focus on Granular Corneal Dystrophy Type 1

GCD1, the classic form of GCD, was the first CD to be mapped to chromosome 5q31 (Stone, Mathers, Rosenwasser, *et al*., 1994). The GCD1 phenotype is characterised by bilateral, discrete, small white granular deposits surrounded by intervening clear areas in the anterior and middle corneal stroma. The deposits usually appear in the first or second decade of life with opacities appearing also in the deeper stroma with increasing age. Usually, the peripheral cornea remains free of deposits and there is no amyloid deposition (Frising, Wildhardt, Frisch, *et al*., 2006). Early on in the condition, vision is usually preserved, since deposits are sparse. However, this condition is slowly progressive with the individual lesions increasing in size and number and may coalesce, causing significant deterioration in visual acuity over time. Individuals exhibiting this condition can also suffer from photophobia, glare and pain from recurrent corneal erosions.

The most prevalent mutation that causes this phenotypic variant has been identified as R555W (C to T transition at nucleotide position 1663 causing the substitution of arginine with tryptophan at position 555) in exon 12 of the *TGFBI* gene and is inherited in an autosomal dominant fashion (Klintworth, 2003). Interestingly, this mutational variant exhibits complete penetrance, as has been established by multiple researchers (Munier, Frueh, Othenin-Girard, *et al*., 2022) (Nowińska, Wylegala, Janiszewska, *et al*., 2011) (Frising, Wildhardt, Frisch, *et al*., 2006) (Kannabiran, Sridhar, Chakravarthi, *et al*., 2005). Therefore, patients carrying this mutation exhibit phenotypically the condition. Having said that, thes individuals, still exhibit phenotypic variability. Also, homozygotes have been described as having a more severe clinical course and early recurrence after surgery (Okada, Yamamoto, Watanabe, *et al*., 1998). To the best of our knowledge, GCD1 is the most common variant of GCD in most countries, except in Japan and Korea where it is quite rare (Konishi, Mashima, Yamada, *et al*., 1998) (Cho, Mok, Na, *et al*., 2012).

Histologically, the deposits are eosinophilic irregularly lobulated granules composed of hyaline. They stain bright red with Masson's trichrome and are weakly PAS positive. Transmission electron microscopy reveals discrete, rod-shaped or trapezoid bodies with discrete borders (Klintworth, 1994) (Kanai, Yamaguchi & Nakajima, 1997) (Akiya & Brown, 1970). Additional deposits appear moth-eaten with variable-shaped cavities containing fine filaments (Brownstein, Fine, Sherman, *et al*., 1974). Descemet's membrane, the corneal endothelium, the deep and peripheral stroma as well as the cornea between the deposits are unremarkable. Immunohistochemistry demonstrate the deposits reacting with antibodies to transforming growth factor beta–induced

protein (Feldman & Bernfeld, 2015), suggesting that TGFBIp forms a major part of the aggregates.

Back in 1969, in an attempt to identify the components of deposits in GCD1, Garner presented histochemical evidence for the presence of tyrosine, tryptophan, arginine and sulfur-containing amino acids (Garner, 1969). Subsequently, Akhtar *et al.* analysed the mineral content within the granular deposits of corneas exhibiting GCD by x-ray spectrometry (Akhtar, Meek, Ridgway, *et al*., 1999). They reported that the deposits contained more Ca, S, and Si than was present in collagen fibrils and corneal epithelial cells. They suggested that calcium ions might play a role in the aggregation of proteins within the deposits (Akhtar, Meek, Ridgway, et al., 1999). Other studies reported the presence of immunoglobulin (IgG and κ and λ light chains) (Moller, Bojsen-Moller, Schroder, *et al*., 1993), cytokeratin and vimentin (Wollensak & Witschel, 1996) and microfibrillar protein (Rodrigues, Streeten, Krachmer, *et al*., 1983). Courtney *et al*. identified three proteins that are present in GCD1 deposits but not present in the corneal stroma of healthy individuals (Courtney, Poulsen, Kennedy, *et al*., 2015). These are the Hb subunit beta, actin cytoplasmic 1 and desmoplakin. The proteins that have been documented by Courtney *et al*. to be present in GCD1 aggregates are listed in Table 3.

GCD1 has been reported in affected patients from Mexico (Zenteno, Correa-Gomez, Santacruz-Valdez, *et al*., 2009), New Zealand (Vincent, de Karolyi, Patel, *et al*., 2010), Taiwan (Hou, Wang, Hsiao, *et al*., 2012), Poland (Nowińska, Wylegala, Janiszewska, *et al*., 2011), India (Chakravarthi, Kannabiran, Sridhar, *et al*., 2002), Brazil (Solari, Ventura, Perez, *et al*., 2007) England (Caucasians) (Stewart, Ridgway, Dixon, *et al*., 1999) and

China(Yang, Han, Huang, *et al*., 2010), amongst others. However, as reported in literature, individuals carrying this mutation still exhibit phenotypic variability.

Table 3: The 25 most abundant proteins present in GCD1 aggregates.

Name of protein	**Relative amount of the protein found in the deposits (%)**
Transforming growth factor beta induced protein	57.1
Collagen alpha-1(I) chain	14.1
Collagen alpha-2(I) chain	9.3
Haemoglobin subunit beta	2.9
Lumican	2.2
Dermcidin	1.6
Keratocan	1.6
Collagen alpha-3(VI) chain	1.5
Aldehyde dehydrogenase, dimeric NADP-preferring	1.2
Histone H2B type 1-A	1.1
Collagen alpha-1(VI) chain	1
Collagen alpha-2(VI) chain	0.8
Decorin	0.8
Serum albumin	0.8
Alpha-enolase	0.7
Collagen alpha-1(III) chain	0.6
Collagen alpha-1(V) chain	0.5
Actin, cytoplasmic 1	0.5
Collagen alpha-2(V) chain	0.4
Clusterin (apolipoprotein J)	0.4
Biglycan	0.4
L-lactate dehydrogenase A chain	0.4
MAM domain-containing protein 2	0.2
Collagen alpha-1(XII) chain	0.1
Desmoplakin	<0.1

Table modified from Courtney *et al*. (Courtney, *et al*., 2015).

A literature search to compile a comprehensive list of the genotypic and phenotypic GCD variants documented to date was performed (Table 4). Electronic bibliographic databases were used to perform a thorough search of original peer-reviewed literature in which GCD phenotypic and genotypic variants were reported. The databases utilized were PubMed and HyDi search engine of the University of Malta. The databases and journals HyDi search engine provides access to include: American Chemical Society

Publications, BioMed Central, BMJ Journals, British National Formulary, ChemSpider, DynaMed Plus, EBSCO Information Services (includes CINAHL Complete, Cochrane Central Register of Controlled Trials, Cochrane Clinical Answers, Cochrane Database of Systematic Reviews, Cochrane Methodology Register), Journals@Ovid Full Text, Karger, MEDLINE Complete, ProQuest (includes Biological Science database, Consumer Health Database, Health & Medical Collection, Healthcare Administration Database, International Pharmaceutical Abstracts, Nursing & Allied Health Database, PTSDpubs, Public Health Database, Science Database), PubChem, PLOS, Reaxys, Scopus, ScienceDirect, SpringerLink and Taylor & Francis Online amongst others.

The search terms used in each string were: '*TGFBI*', 'cornea*' (here, '*' is used as a wildcard symbol for search term truncation), 'granular corneal dystrophy', 'GCD'. Search strings including multiple combinations of these terms were implemented in order to try and include as many relevant articles as possible. The Boolean term AND was used for the PubMed search.

The articles displayed after applying the Medical Subject Headings (MeSH) terms and database filters, which included articles in English, were vetted and duplicate articles that were featured in multiple search strings were removed. The studies were screened by title and abstract in order to apply the inclusion criteria. The inclusion criteria were that the article had to describe a classic GCD1 phenotype and the identified mutation and/or mutations that are known to cause GCD1 but with patients exhibiting a different phenotype from the classic one. The relevant full text articles were assessed to determine the applicability to this research.

A clear example of GCD phenotypic variability was reported in a study carried out in Taiwan, where even though the majority of patients with the R555W mutation showed the typical phenotypic features of classic GCD1, one affected individual was noted to have dot and line opacities (Hou, Wang, Hsiao, *et al*., 2012). Furthermore, Kattan *et al*. (Kattan, Serna-Ojeda, Sharma, *et al*., 2017) described two families, one of Mexican and the other of Italian descent, with affected individuals exhibiting two distinct phenotypes. These were the classic GCD1 phenotype and a sea fan/vortex pattern of superficial granular stromal corneal deposits originating from the inferior aspect of the cornea. The R555W mutation was present in the heterozygous state in all the affected individuals. In both the Italian and the Mexican family, single nucleotide variants (SNVs), with an allele frequency higher than that reported in the normal population, were identified in affected individuals. However, as these SNVs were present in individuals exhibiting both the classic and the vortex patterns of dystrophic deposits, these SNVs are unlikely to be related to the determination of the phenotype in this family (Kattan, Serna-Ojeda, Sharma, *et al*., 2017).

Okada *et al*. reported a R555W homozygous mutation in a Japanese family where the individuals affected were the offspring of two affected individuals who were consanguineous. The phenotype in this family was described as a continuous placoid opacification of the cornea with early recurrence after corneal graft surgery (Okada, Yamamoto, Watanabe, *et al*., 1998).

Interestingly, in a study carried out on Indian families, two thirds of the patients exhibiting the GCD1 classic phenotype with the Arg555Trp mutation were also heterozygous for Leu269Phe mutation. This was found in only 3% of their normal

control subjects. However, the authors suggested that ideally, a study involving a larger cohort of families with GCD is necessary to determine whether the Leu269Phe polymorphism is in linkage disequilibrium with the Arg555Trp mutation in this population (Chakravarthi, Kannabiran, Sridhar, *et al*., 2002).

GCD also exhibits genotypic variability. In fact, population studies carried out in different countries and on different ethnicities have led to the identification of a number of genetic variants that also cause granular deposits. Zenteno *et al.* (Zenteno, Santacruz & Ramirez, 2006) identified a mutation, H626R, located in exon 14 of the *TGFBI* gene, in two related patients who did not have the classic R555W mutation but still exhibited a classic granular type 1 phenotype. This mutation had been previously identified only in subjects with the lattice stromal corneal dystrophy variant 1/3A, which, as opposed to GCD, causes amyloid deposition in the form of lines in the corneal stroma (Zenteno, Santacruz & Ramirez, 2006). Zenteno, *et al.,* reported another novel mutation, V113I, present in exon 4 of the *TGFBI* gene (Zenteno, Ramirez-Miranda, Santacruz, *et al*., 2006). This mutation produced a phenotypic variant of granular opacities located in the anterior and midstroma, showing a 'centrifuge' distribution. There were also numerous small (< 2 mm diameter) non-coalescent opacities present in the periphery. In another study, this time carried out by Stewart *et al.* (Stewart, Ridgway, Dixon, *et al*., 1999), a mutation in exon 4 of the *TGFBI* gene, namely R124S, was found in a single Asian individual. The phenotype reported was that of numerous small white granules in the cornea with signs and symptoms of later onset than those reported in individuals with classic type 1 GCD. Ha *et al.,* (Ha, Cung, Chau, *et al*., 2003) documented a granular phenotype with deposits that are smaller and deeper than those seen in classical type 1 GCD. They located the mutation

to D123H of the *TGFBI* gene. In contrast to the complete penetrance exhibited by the mutation R555W (Munier, Frueh, Othenin-Girard, *et al*., 2022) (Nowińska, Wylegala, Janiszewska, *et al*., 2011) (Kannabiran, Sridhar, Chakravarthi, *et al*., 2005) this mutation was found to have low penetrance (Ha, Cung, Chau, *et al*., 2003).

As evidenced above, since even a slight variation in the phenotype may point towards the presence of a new genotype, it is increasingly recommended by researchers that genetic analysis is carried out on each individual exhibiting GCD (Frising, Wildhardt, Frisch, *et al*., 2006). This will ultimately lead to a more extensive and accurate diagnosis and subsequent management plan for each individual case.

Table 4: Genotypic and phenotypic GCD1 variants reported in literature.

Race/ Country of origin	**GCD variants**	**Phenotype**	**Gene locus and exon mutation location**	**Nucleotide mutation change**	**Amino acid change and codon location**	**Single nucleotide polymorphism**	**Reference**
Mexico	Type I (Classic)	Bilateral, central opacities in anterior and middle stroma with clear spaces in between. Limbal corneal region spared.	*TGFBI* (5q31) 12	c.1663C>T	R555W		(Zenteno, *et al*., 2009)
	Type I variant	Classic phenotype. Central, corneal opacities in anterior and middle stroma, with clear spaces in between. Limbal corneal region spared.	*TGFBI* (5q31) 14	1924A>G	H626R		(Zenteno, *et al*.,2006)
	Type I variant	Bilateral, symmetric, opacities showing a 'centrifuge' distribution. Intervenening stroma between the opacities was clear. The lesions were located at the anterior and midstroma. Numerous small (< 2 mm diam.) non-coalescent opacities in the periphery.	*TGFBI* (5q31) 4	c.337G>A	V113I (It is the most 5' located mutation detected so far in subjects with *TGFBI* -linked CDs)		(Zenteno, *et al*., 2006)

Mexico	variant	Bilateral diffuse small granular opacities in central cornea with almost no clear spaces between lesions. No peripheral affectation or epithelial erosions.	*TGFBI* (5q31) 12/14	1649T>C 1877A>G	L550P/H626R		(Zenteno, *et al*., 2009)
Mexico & Italy	Type I variant	Vortex pattern of superficial stromal corneal deposits originating from the inferior aspect of the cornea.	*TGFBI* (5q31) 12	c.1663C>T	R555W	<u>Mexican:</u> Exon8 c.981A>G; p.Val327Val Exon12 c.1620T>C p.Phe540Phe <u>Italian:</u> Exon6 c.651G>C p.Leu217Leu	(Kattan, *et al*., 2017)
Italy	Type II variant	Families with Type II GCD presenting with a GCD1 phenotypic spectrum in female descendants.	*TGFBI* (5q31) 4		R124H		(Mazzotta, *et al*., 2015)
New Zealand East-Indian & Caucasian	Type I	Classic phenotype	*TGFBI* (5q31) 12	c.1664C>T	R555W		(Vincent, *et al*., 2010)

Taiwan	Type I and a variant	Bread crumb like or gray–white granular opacities in the Bowman's layer and superficial corneal stroma. However, the proband of one family showed dot and line opacities.	*TGFBI* (5q31) 12	c.1664G>T	R555W		(Hou, *et al*. 2012)
	variant	Polymorphic dots with lattice lines	*TGFBI* (5q31) 12		A546D (spontaneous mutation)		(Hou, *et al*. 2012)
Poland	Type I	Granular crumb-like anterior stromal deposits in the central cornea, with age becoming located deeper in the stroma.	*TGFBI* (5q31) 12	1710C>T	R555W		(Nowińska, *et al*., 2011)
India	Type I	Granular opacities in anterior third to full stroma.	*TGFBI* (5q31) 12	1710C>T	R555W	Exon 7 c.852C>T (Leu269Phe) Found in 12 of 18 patients	(Chakravarthi, *et al*., 2002)
	Type 1 variant	Discrete gray-white deposits with clear intervening stroma.	*TGFBI* (5q31) 12	1548C>G	S516R		(Paliwal, *et al*., 2010)
Caucasian	Type I	Classic phenotype. ("crumb-like" corneal deposits)	*TGFBI* (5q31) 12	1710C>T	R555W		(Stewart, *et al*., 1999)

Asian	Type I variant	Numerous small white granules in cornea. S+S of later onset than classic type I.	*TGFBI* (5q31) 4	417C>A	R124S		(Stewart, *et al*., 1999)
Vietnam	Type I variant	Smaller and deeper deposits than in classical type I. Low penetrance		G>C	D123H (Asp123His)		(Ha, *et al*., 2003)
Brazil	Type I	Classic phenotype	*TGFBI* (5q31) 12	C>T	R555W		(Solari, *et al*., 2007)
	Type I variant	Characteristic biomicroscopic findings as in classic GCD1		No mutation in exon 4 or 12 found in 2 families			(Solari, *et al*., 2007)
China	Type I	Classic phenotype	*TGFBI* (5q31) 12		R555W		(Yang, *et al*., 2010)
	Type I variant	Possible reduced penetrance	*TGFBI* (5q31) 12		A546D (aspartic acid)		(Yu, *et al*., 2008)
China	Type II		*TGFBI* (5q31) 4		R124H+c.Δ307–308CT		(Pang & Lam, 2002)

Japan	Superficial variant	Severe form of GCD of early onset	*TGFBI* (5q31) 4	418G>T	R124L		(Mashima, *et al*., 1999)
France	Variant	Onset: Bilateral opacities at <10yrs of age Round/ snowflake opacity in superficial stroma and subepithelial layers. Initially: stroma between opacities was clear, but later it developed a ground-glass appearance	*TGFBI* (5q31) 4	418G>T + ΔACGGAG of codons 125 and 126	R124L + deletion of 6 base pairs (2 AA) corr. to residues threonine 125 and glutamic acid 126.		(Dighiero, *et al*., 2000)
Germany	Type I Variant	Patient already had bilateral PK on presentation. Surgery was performed at early age due to severe decrease in VA. Opacities in central superf stroma, at junction with recipient cornea and deep stromal deposits at site of previous corneal sutures.	*TGFBI* (5q31) 12	c.1663C>T c.1645G>A	Arginine(555) & Alanine(549) Changed to: Tryptophan & Threonine(549)		(Frising, *et al*., 2006)
Japan	Type I homozygous	Severe placoid type of GCD I. Early recurrence after corneal graft surgery.	*TGFBI* (5q31) 12	c.1663C>T homozygous	R555W homozygous		(Okada, *et. al*., 1998)

Granular Corneal Dystrophy Type 2

Similarly to GCD1, GCD2 (previously known as Avellino corneal dystrophy, GCD with amyloid, combined lattice-granular corneal dystrophy, R124H mutant *TGFBI*) is also inherited in an autosomal dominant fashion. This GCD variant is reported to be the most common type of GCD in Japan (Konishi, Mashima, Yamada, *et al*., 1998) (Mashima, Yamamoto, Inoue, *et al*., 2000) and Korea (Kim, Yoon, Cho, *et al*., 2001).

GCD2 is characterized initially by small superficial white dots with subsequent ring or stellate-shaped granular deposits in the superficial central corneal stroma that usually develop in the first or second decades of life (Klintworth G. , 2003). Linear lattice-like opacities are usually also present in the deeper stromal layers, however, these may not appear until later. In contrast to lattice lines seen in LCD, the linear opacities in GCD2 appear whiter and rarely cross each other. Over time, stromal haze develops between the deposits. (Garner, 1969) (Akiya & Brown, 1970) ((Brownstein, Fine, Sherman, *et al*., 1974) (Iwamoto, Stuart, Srinivasan, *et al*., 1975) ((Stuart, Mund, Iwamoto, *et al*., 1975) (Kheir, Cortes-Gonzalez, Zenteno, *et al*., 2019). Clinically, when compared to GCD1, the visual acuity is less impaired. The presenting symptoms of GCD2 patients include photophobia, glare, foreign body sensation and recurrent painful epithelial erosions.

From previous studies, histology of corneal tissue from patients with GCD2 demonstrates rod-shaped hyaline crystalloid material similar to that seen in GCD1 as well as amyloid deposits similar to those seen in LCD1 (Holland, Daya, Stone, *et al*., 1992) (Akiya, Takahashi, Nakano, *et al*., 1999). It was reported that the non-amyloid accumulations present in this variant consist of a combination of the 66 kDa and 68 kDa forms of TGFBIp (Korvatska, Henry, Mashima, *et al*., 2000). Just as in GCD1, the

hyaline stains red with Masson Trichome, while, similarly to what is seen in LCD, the amyloid stains red with Congo red stain. The composition of corneal non-amyloid granular deposits from GCD2 individuals was studied by Karring *et al.* in 2012. These were reported to consist of intact TGFBIp, serum amyloid P-component, clusterin, type III collagen, keratin 3 and histone H3-like protein. They discovered that even though intact or nearly full-length TGFBIp isoforms were present in the granular deposits, there were differences in spectral count ratios for tryptic peptides from the inter-domain regions between the mutated TGFBIp in the deposits and the TGFBIp from the control tissue. They suggested that this could indicate that the aggregated TGFBIp has an altered domain arrangement compared to the control TGFBIp. They also noticed that largest difference in the spectral counts for semi-tryptic TGFBIp was for sequence L128–R172. This is situated adjacent to the mutated R124 residue. Consequently, they postulated that local structural instabilities might induce an increase in cleavages of this region in vivo that might cause local changes to structural stability. Besides, they also suggested that the TGFBIp aggregates probably cause tissue damage that in turn induce a tissue repair response with accumulation of type III collagen in the deposits (Karring, Runager, Thogersen, *et al*., 2012). Thus, even though GCD1 deposits have been identified as hyaline and considered to be similar to those found in GCD2, with reference to the above findings, there might be differences between the two. In the study mentioned above, the differences in spectral counts detected were due to changes located adjacent to the R124 mutated residue, thus, the molecular pathology in GDC1 could be different since, in GCD2, the mutation lies in position 124 while in GCD1, the mutation is at position 555 of the *TGFBI* gene. Hence, continuing studies are

needed in order to analyse further the composition of the deposits found in corneas of GCD1 patients as well as the mechanisms involved in leading to their deposition.

The phenotypical and histopathological picture of GCD2 is strongly associated with the R124H mutation in the *TGFBI* gene. However, it was noted that the phenotype of affected patients varies in severity both within and from family to family (Klintworth G. , 2003). In cases where the mutation is homozygous, the condition is more severe and these patients are also prone to early recurrence of the condition after corneal grafting procedures (Mashima, Konishi, Nakamura, *et al*., 1998) ((Fujiki, Hotta, Nakayasu, *et al*., 1998) (Okada, Yamamoto, Watanabe, *et al*., 1998)**.**

Granular Corneal Dystrophy Type 3

GCD3 is also referred to as superficial GCD, 'true" Reis-Bücklers corneal dystrophy and corneal dystrophy of Bowman's layer type I. This is because the corneal deposits are usually limited to the epithelial or subepithelial region (Haddad, Font & Fine, 1977) (Møller, 1989). The mutation causing this variant has been mapped to R124L (Okada, Yamamoto, Watanabe, *et al*., 1998) (Dighiero, Niel, Ellies, *et al*., 2001) (Yamamoto, Okada, Tsujikawa, *et al*., 2000) ((Konishi, Yamada, Nakamura, *et al*., 2000). In Sardinia, a similar clinical phenotype has been reported with the mutation being documented in the *TGFBI* gene as delta f540 (Rozzo, *et al*., 1998). Unfortunately, to the best of our knowledge, there is no documentation regarding histopathology in these patients.

Affected individuals usually present with bilateral recurrent epithelial erosions associated with pain and photophobia during the first or second decades of life. Deposits are diffuse grey-white, sand-like, in confluent or non-confluent geographic morphology with a peripheral narrow strip of normal cornea (Reis, 1917) (Bücklers,

1949). The deposits are age-dependent and progressive. This results in scarring and opacification in the axial cornea, irregular astigmatism and deteriorating visual acuity. A decrease in corneal sensitivity was also reported by a study carried out by Verdi and Filippone (Verdi & Filippone, 1958). The deep corneal stroma, endothelium and Descemet's membrane are not affected (Klintworth G. , 2003). Histochemically, it is indistinguishable from GCD1 (Haddad, Font & Fine, 1977) (Møller, 1989).

Granular Corneal Dystrophy Type 4

The fourth variant of GCD was reported in a French family by Dighiero, Drunat, D'Hermies, *et al*., (2000) and Dighiero, Niel, Ellies, *et al*., (2001). The causative genetic mutation in this family was identified to be a heterozygous R124L mutation in *TGFBI*, combined with another heterozygous mutation in the same gene that causes the deletion of 2 amino acid residues at codons 125 and 126 (d 125-d E126). The phenotype is best described to be intermediate between GCD1 and GCD3 Dighiero, Drunat, D'Hermies, *et al*., (2000).

The patients presented with pain from corneal epithelial erosions in early childhood. Deposits were round/snowflake shaped and were found in the superficial stroma and subepithelial layers. Initially the stroma between the opacities was clear, but later it developed a ground-glass appearance Dighiero, Drunat, D'Hermies, *et al*., (2000).

2.5.2.2.2 Lattice Corneal Dystrophy

Lattice corneal dystrophy is one of the more prevalent forms of the CDs in Western world. It manifests as central interdigitating branching linear opacities within the corneal stroma. These opacities have been found to consist of amyloid (Lin, Chen & Cui, 2016). LCD1 causes slowly progressive deterioration in vision and recurrent erosions are common. Also, there is often a decrease in corneal sensation with patients

occasionally presenting with recurrent corneal erosions before the corneal opacities appear (Trief, 2019). In 1997, Munier, Korvatska, Djemai, *et al.,* discovered that the mutation R124C in the *TGFBI* gene can cause LCD1. Other mutations in the *TGFBI* gene have since been associated with LCD1. These include A546D and P551Q that were reported in 2004 by Klintworth *et al*. and Aldave *et al*.

Studies have been performed in the attempt to identify the protein components of LCD deposits. The only proteins that were reported to be present in both LCD and GCD deposits are TGFBIp and clusterin (apolipoprotein J). Besides these, amyloidogenic as well as nonfibrillar amyloid associated proteins were also found to form part of the amyloid deposits in LCD (Courtney, Poulsen, Kennedy, *et al*., 2015) (Karring H. , Runager, Thogersen, *et al*., 2012) (Table 5).

Table 5: The most abundant proteins present in LCD1 deposits, modified from (Courtney, Poulsen, Kennedy, *et al*., 2015)

LCD1 Stromal deposits		**LCD1 Bowman's deposits**	
Name	**%**	**Name**	**%**
Collagen alpha-1(I) chain	14.9	Collagen alpha-1(I) chain	11.6
Transforming growth factor beta induced protein	11.2	Collagen alpha-2(I) chain	9.1
Collagen alpha-2(I) chain	11.1	Dermcidin	8.5
Serum amyloid P component	6.2	Transforming growth factor beta induced protein	7.1
Dermcidin	6.1	Protein S100-A9	6.8
Ig Kappa chain C region	5.3	Cystatin A	6.1
Protein S100-A9	4.9	Serum amyloid P component	5.2
Cystatin A	3.9	Protein S100-A6	3.6
Histone H2B type 1-A	3.7	Protein S100-A4	3.4
Haemoglobin subunit beta	2.7	Aldehyde dehydrogenase, dimeric NADP-preferring	3
Ubiquitin-40S ribosomal protein S27a	2.7	Histone H2B type 1-A	2.8
Lumican	2.6	Alpha-enolase	2.6
Clusterin	2.3	Vimentin	2.4
Apolipoprotein A-IV	2.3	Actin, cytoplasmic 1	2.3
Keratocan	2.1	Peroxiredoxin-1	2.2
Serum albumin	2.1	Fatty acid-binding protein, epidermal	2.1
Apolipoprotein D	2	Ubiquitin-40S ribosomal protein S27a	2.1
Alpha-enolase	2	Serum albumin	1.9
Collagen alpha-3(VI) chain	1.9	Lysozyme C	1.9
Apolipoprotein A-IV	1.9	Heat shock protein-beta1	1.8
Prostaglandin-H2 D isomerase	1.8	Apolipoprotein A-IV	1.6
Vimentin	1.7	Ig gamma-1 chain C region	1.5
Aldehyde dehydrogenase, dimeric NADP-preferring	1.6	Keratocan	1.5
Decorin	1.6	Annexin A2	1.4
Ig gamma-3 chain C region	1.6	Decorin	1.3

2.5.2.2.3 Reis–Bücklers Corneal Dystrophy

In another Epithelial–Stromal *TGFBI* dystrophy named Reis–Bücklers CD (RBCD), bilateral reticular opacities can be seen in the superficial central cornea. RBCD was previously classified as GCD3. The lesions are found mainly in Bowman membrane with subsequent involvement of the epithelium and superficial stroma (Kuchle, Green, Volcker, *et al.,* 1995). Eventually the cornea develops a superficial haze that leads to a reduction in vision. Epithelial erosions that cause acute episodes of pain and photophobia occur more frequently in RBCD than in the other variants of GCD (orphanet) (Hall, 1974). Analysis of the *TGFBI* gene, in individuals exhibiting RBCD, located the mutation to be R124L (Okada, Yamamoto, Watanabe, *et al*., 1998).

2.5.2.2.4 Thiel-Behnke

Thiel-Behnke corneal dystrophy can present with a phenotype similar to RBCD but it is usually less severe. The corneal opacities are usually described as sub-epithelial honeycomb-shaped opacities that cause progressive visual impairment (Thiel & Behnke, 1967) (Kuchle, Green, Volcker, *et al*., 1995). Individuals with TBCD also suffer from corneal erosions. TBCD shows genetic heterogeneity since, apart from the *TGFBI* gene mutation, TBCD appears to be also caused by mutation in chromosome 10 (Yee, Sullivan, Lai, *et al*., 1997).

2.5.2.3 Stromal Dystrophies

Corneal dystrophies categorized under the Stromal dystrophy class affect mainly the middle layer of the cornea, known as the stroma, and are not caused by dominant mutations in the *TGFBI* gene. With that said, some of these can also involve other layers of the cornea; for instance, MCD affects both the stroma and endothelium, while Schnyder CD (SCD) involves the epithelium, stroma, and endothelium. The most

common types of stromal dystrophies include Macular corneal dystrophy (MCD), Schnyder corneal dystrophy (SCD), Congenital stromal corneal dystrophy (CSCD), Posterior amorphous CD (PACD) and Fleck corneal dystrophy (FCD) (Weiss, Moller, Aldave, *et al*., 2015).

2.5.2.3.1 Macular Corneal Dystrophy

Macular corneal dystrophy is an autosomal recessive disorder which is characterized by progressive, gray, punctate opacities that merge, causing the corneal stroma to become cloudy. This eventually causes severe visual impairment. The corneal opacities seen in MCD may appear from early infancy up to the sixth decade of life. (Klinthworth, 2009). Even though the phenotypes are the same, there are mainly 2 types of MCD which are distinguished by the absence or presence of sulphated keratan sulphate in the patient serum (Akama, Nishida, Nakayama, *et al*., 2000). Patients suffer from painful recurrent erosions and there is usually a decrease in the corneal sensitivity. Most cases of MCD are caused by mutations in the carbohydrate sulfotransferase 6 (*CHST6*) gene (16q22). The *CHST6* gene encodes the protein sulfotransferase, an enzyme involved in the synthesis of keratan sulphate (Klintworth, 1980). Keratan sulphate is a glycosaminoglycan that is essential in the maintenance of corneal transparency (Pacella, Pacella, De Paolis, *et al*., 2015). Significant allelic heterogeneity has been described in the *CHST6* gene in various ethnicities around the world. In fact, to date, more than 100 mutations in the *CHST6* gene have been identified (Nowinska, Wylegala, Teper, *et al*., 2014).

2.5.2.3.2 Schnyder Corneal Dystrophy

Schnyder corneal dystrophy phenotypically presents with stromal clouding and eventual small white opacities and/or cholesterol crystals (McCarthy, Innis, Dubord, *et*

al., 1994). Symptoms include progressive decrease in vision and glare. The typical phenotype is of a ring-shaped yellow-white opacity, which is made up of fine needle-shaped crystals in Bowman layer and the adjacent anterior stroma of the central cornea. However, these crystals are usually present in only half of the individuals with this condition (Nickerson, Bosley, Weiss, *et al*., 2013). This condition is caused by various missense mutations in the ubiA prenyltransferase domain-containing protein 1 (*UBIAD1*) gene (Orr, Dube, Marcadier, *et al*., 2007).

1.5.2.3.3 Fleck Corneal Dystrophy

Another stromal CD is Fleck corneal dystrophy (FCD). Here we see many minute dot-like white opacities resembling flecks that are scattered in all layers of the corneal stroma (Francois & Neetens, 1857). The stroma between the flecks is clear and patients usually have normal vision. Some experience minor photophobia. FCD is caused by a mutation in the *PIKFYVE* gene located on chromosome 2 (Kawasaki, Yamasaki, Shinomiya, *et al*., 2012).

2.5.2.4 Endothelial Dystrophies

Fuchs endothelial corneal dystrophy (FECD), Posterior polymorphous corneal dystrophy (PPCD) and Congenital hereditary endothelial dystrophy (CHED) are classified as endothelial dystrophies in the IC3D classification (Weiss, Moller, Aldave, *et al*., 2015). They are characterized by a progressive degeneration of the innermost part of the cornea, namely, the corneal endothelium. The corneal endothelium is made up of a layer of cells whose function is to keep the cornea relatively dehydrated by pumping excess water out (Schmedt, Mazzini Silva, Ziaei, *et al*., 2012). The endothelial dystrophies are, at least in part, caused by genetic mutations. The major causal genes involved are zinc finger E-box binding homeobox 1 (*ZEB1*), Collagen alpha-2(VIII) chain

(*COL8A2*), solute carrier family 4 (*SLC4A11*) and lipoxygenase homology PLAT domains 1 (*LOXHD1*) in FCD (Iliff, Riazuddin, & Gottsch, 2012) (Gottsch, Sundin, Liu, *et al*., 2005); *COL8A2, ZEB1* and the promoter of the ovo like zinc finger 2 (*OVOL2*) gene in PPCD (Biswas, Munier, Yardley, *et al*., 2001) (Liskova, Gwilliam, Filipec, *et al*., 2012) and *SLC4A11* in CHED (Vithana, Morgan, Ramprasad, *et al*., 2006). Even though endothelial dystrophies can be identified as distinct, as one can see, they exhibit genetic heterogeneity. Biswas *et al* pointed this out in a study where they identified that gene mutations in *COL8A2* were present in both FCD and PPCD (Biswas, Munier, Yardley, *et al*., 2001). Similarly, mutations in the gene *SLC4A11* are associated with both CHED and FCD (Vithana, Morgan, Ramprasad, *et al*., 2008) (Shah, al-Rajhi, Brandt, *et al*., 2008).

2.5.2.4.1 Fuchs Endothelial Corneal Dystrophy

Fuchs endothelial corneal dystrophy is characterized by excrescences, called guttae, on a thickened Descemet membrane. Individuals suffering from this condition develop generalized corneal edema due to loss of endothelial cells which are subsequently not able to keep the cornea dehydrated any longer (Baratz, Tosakulwong, Ryu, *et al*., 2010). This causes decreased visual acuity and recurrent episodes of corneal erosions. The exact aetiology of FECD is still unknown. However, it is thought to be due to a combination of genetic and environmental factors (Hamill, Schmedt, & Jurkunas, 2013).

2.5.2.4.2 Posterior Polymorphous Corneal Dystrophy

Posterior polymorphous corneal dystrophy is a very rare type of endothelial corneal dystrophy. Phenotypically, there are vesicles bordered by a grey haze at the level of Descemet membrane. Occasionally, metaplasia and overgrowth of the corneal endothelial cells occurs with spread over the iris and nearby structures (Krafchak, Pawar, Moroi, *et al*., 2005). This causes peripheral anterior iris adhesions,

iris atrophy, pupillary ectropion, and corectopia. Vision is not usually affected. Having said that, sometimes, individuals may develop secondary glaucoma and/or corneal oedema that could lead to a severe decrease in vision (Liskova, Gwilliam, Filipec, *et al*., 2012).

2.5.2.4.3 Congenital Hereditary Endothelial Dystrophy

Congenital hereditary endothelial dystrophy is an autosomal recessive condition that is characterized by a diffuse ground-glass appearance of the corneas, which are also thickened. There is secretion of an abnormal collagenous layer at the Descemet membrane and the endothelium has an altered morphology (Vithana, Morgan, Sundaresan, *et al*., 2006). Patients suffer from nystagmus and blurred vision, and may occasionally have sensorineural deafness too (Judisch & Maumenee, 1978).

2.6 Literature Review on Transforming Growth Factor Beta Induced Protein and Gene

2.6.1 TGFBI Gene and Protein Structure

The human *TGFBI* gene codes for a 683 amino acid protein, named TGFBIp (also known as keratoepithelin, Big-h3 and RGD-CAP), which is highly conserved between species. This extracellular matrix protein was first isolated by Skonier, Neubauer, Madisen, *et al*. (1992) from human adenocarcinoma cells treated with TGF-β, while trying to identify new genes that were induced by TGF-β. Subsequently, Munier, Korvatska, Djemai, *et al*. (1997) mapped the *TGFBI* gene to chromosome 5q31 (Figure 3), in the process, identifying four corneal dystrophies caused by mutations of this gene (Munier, Korvatska, Djemai, *et al*., 1997).

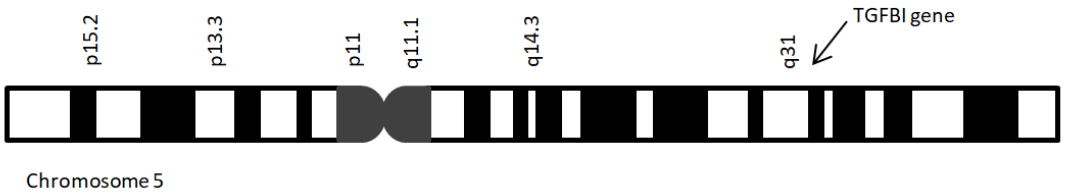


Figure 3: Genomic Location of the *TGFBI* gene.

TGFBIp consists of a secretory signal peptide sequence, an N-terminal EMI (CROPT) domain (found originally in the EMILIN-1 protein), four fasciclin-like homologous internal repeat domains (FAS1 domains) and a COOH-terminal Arg-Gly-Asp (RGD) sequence which binds to integrin. Its signal peptide is cleaved before it is secreted from the cell into the extracellular matrix (Stenvag, Andreasen, Enghild, *et al*., 2013). The EMI is a cysteine rich domain that is present in many human proteins. The *TGFBI* protein contains eleven Cys-residues. Six of these residues that are found in the EMI domain are highly conserved and they form three disulphide intra-domain bridges. There is one Cys residue in the FAS1-1 domain and two in both the FAS1-2 and FAS1-3 domains. Its function is not yet confirmed, however, it is thought that it mediates protein-protein interactions (Callebaut, Mignotte, Souchet, *et al*., 2003). The FAS1 domain contains a mix of beta-sheets and alpha-helices that are grouped together (Garcia-Castellanos, Nielsen, Runager, *et al*., 2017) and is homologous to the fasciclin-1 protein in Drosophila that functions as a cell adhesion region. Each FAS1 domain is characterised by two highly conserved sequences, H1 and H2, each consisting of about 10 amino acids (Stenvag, Andreasen, Enghild, *et al*., 2013). Structurally, TGFBIp is composed of a left moiety including the EMI, FAS1-1, and FAS1-2 domains and a right moiety consisting of the FAS1-3 and FAS1-4 domains (Figure 4). The left moiety is very compact and is held together by interdomain disulfide bridges. The domains interact

with each other through polar interactions, however, since the interfacial residues are not conserved, it is difficult to foretell how the quaternary structure of the *TGFBI* protein will be.

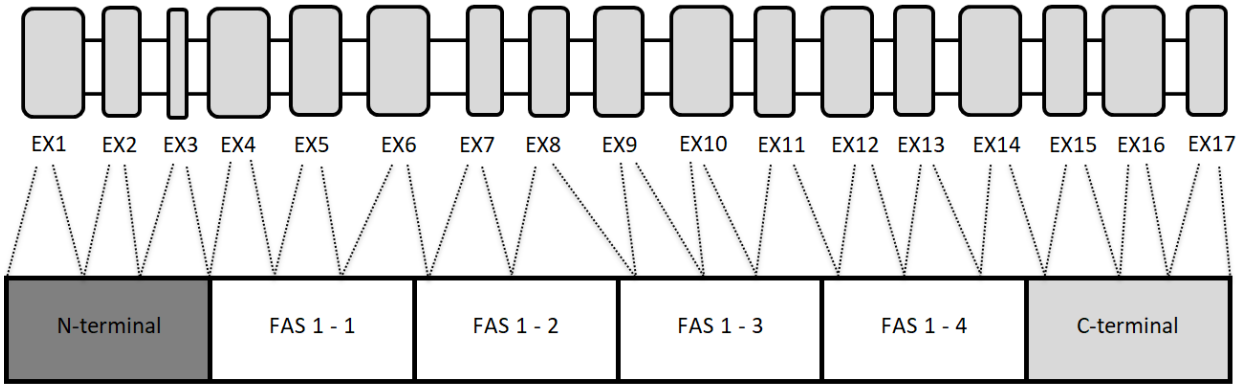


Figure 4: Diagrammatic representation of the transforming growth factor beta induced gene and protein structure.

EX: exon; FAS: Fas Cell Surface Death Receptor

Physiological function and Protein interactions of TGFBIp:

The physiological function of TGFBIp is not understood fully. Literature has revealed that TGFBIp is involved in physiologic and pathologic pathways by mediating cell proliferation, differentiation, adhesion (Park, Jung, & Kim, 2009) (Skonier, Neubauer, Madisen, *et al*., 1992) and migration (Nam, Sa, You *et al*., 2006, Park, Kim, Yim, *et al*., 2021).

Proteins that are known to be homologous to the TGFBIp, since they also contain EMI domains and multiple FAS1 domains, are periostin and stabilin. However, in the latter two proteins, the RGD motif is not present. The RGD motif, which has been identified in TGFBIp, is known to be present in many extracellular matrix proteins. It plays a role in the binding of TGFBIp to collagen, to the glycoproteins laminin and

fibronectin as well as to the proteoglycans biglycan and decorin (NCBI, *et al*., 2016) (Kim, Jeong, Nam, *et al*., 2002).

Collagens form a large part of the corneal ECM and play an important role in the maintenance of corneal clarity. TGFBIp binds via noncovalent bonds to collagen types I, II, and IV (Billings, Whitbeck, Adams, *et al*., 2002). On the other hand, collagen type VI forms both noncovalent and covalent connections with TGFBIp (Kim, Jeong, Nam, *et al*., 2002) (Reinboth, Thomas, Hanssen, *et al*., 2006) (Hanssen, Reinboth, & Gibson, 2003). More than half of the TGFBIp present in the human cornea is covalently bound to collagen type XII by a reducible bond, making it insoluble (Andersen, Karring, Moller-Pedersen, *et al*., 2004). To date, the mechanism that regulates the amount of TGFBIp that binds to collagens is still unknown (Nielsen, Poulsen, Lukassen, *et al*., 2020).

TGFBIp is documented to mediate cell adhesion and migration in vitro via interaction with integrins α1β1, α3β1, αvβ3, αvβ5, α6β4, and αmβ2 in various tissues (Kim & Kim, 2008) (Kim, Jeong, Nam, *et al*., 2002) (Nam, Sa, You, *et al*., 2006). In humans, the highly conserved residues present in the FAS1-2 and Fas1-4 domains of TGFBIp facilitate cell adhesion of corneal epithelial cells through interactions with integrin a3β1. However, this adhesion is not seen with Fas1-1 and Fas1-3 domains (Kim, Kim, Lee, *et al*., 2000). Integrin αVβ5 has also been documented to be found in the cornea. Kim, Jeong, Nam, *et al*., 2002 also reported that all four of the FAS-1 domains in the TGFBIp promote fibroblast adhesion by interacting with integrin αVβ5 that is present on the fibroblast cell surface.

2.6.2 TGFBIp Molecular Properties

In a study carried out by Andersen *et al*., the TGFBIp was purified from normal human corneas by using non-denaturing conditions and chromatography tandem mass spectroscopy. Here, no evidence of post-translational modifications was found (Andersen, Karring, Moller-Pedersen, *et al*., 2004). It was noted that the human cornea contains mostly 'mature' C-terminally processed TGFBIp (does not contain the C-terminal residues Ser_{658}-His_{683}) together with a small amount of full-length TGFBIp (Andersen, Karring, Moller-Pedersen, *et al*., 2004). Several isoforms of the full length TGFBIp have been eventually identified in the cornea in more recent studies (Karring, Thogersen, Klintworth, *et al*., 2009). Interestingly, post-translational phosphorylation and/or carboxylation of glutamate residues have been reported to be present in human corneal endothelium. This would result in changes in the isoelectric point of the protein (Karring, Thogersen, Klintworth, *et al*., 2005). It has been suggested that both non-modified and modified forms of TGFBIp might be present in a tissue or else that TGFBIp is modified in a tissue specific manner (Srivastava & Srivastava, 1999) (Couto, 2008).

To try and provide further insight about the structure and molecular properties of the TGFBIp, Basaiawmoit *et al*. used small-angle X-ray scattering (SAXS) to construct a model of human TGFBIp. The structure of the monomer consists of a "beads-on-a-string" arrangement of the four FAS domains. At high concentrations, the protein forms trimers. This was found to happen for both wild type and the R124H mutant TGFBIp. The authors also noted that there were no major structural differences between wild type and the mutant TGFBIp. This led to their hypothesis that there must

be other corneal-specific factors involved in the formation of corneal R124H deposits. Consequently it was suggested that the mutated protein accumulates due to possible alterations in the protein interactions with corneal molecules (Basaiawmoit, Oliveira, Runager, *et al*., 2011).

2.6.3 TGFBIp Processing

The TGFBIp that is synthesized is usually secreted via the endoplasmic reticulum/golgi-dependent pathway. On the other hand, internalization of this protein is mediated by the binding of the RGD motif of TGFBIp to integrin αVβ3 or αVβ5, via a caveolin-dependent endocytic pathway. It is then transported to lysosomes for degradation. Choi *et al*. also demonstrated that the ubiquitin-proteasome system regulates TGFBIp endocytic trafficking.

2.6.4 Hypotheses Regarding the Cause of Deposition of Mutated TGFBIp in Corneal Dystrophies

Multiple researchers confirmed that the deposits present in corneas of *TGFBI*-linked CDs are composed of aggregates largely containing TGFBIp (Karring H., Runager, Thogersen, Klintworth, *et al*., 2012) (Courtney, Poulsen, Kennedy, *et al*., 2015). However, even though mutations in *TGFBI* gene have been strongly linked with certain corneal dystrophy subtypes, the way the mutated TGFBIp causes corneal deposits still remains unknown.

A number of diseases such as systemic amyloidosis and Alzheimer's have been associated with inappropriate aggregation of proteins that lead to the deposition of insoluble structures. In general, there are two different aggregation pathways. One of the pathways leads to the formation of amorphous aggregates while the other causes deposition of ordered amyloid structures. Amorphous aggregates are usually formed

from large proteins in which the whole protein is involved in aggregate formation. On the other hand, small proteins or peptides that have been generated by aberrant proteolytic processing of parent proteins tend to form amyloidogenic aggregates. *TGFBI* CDs exhibit phenotypic heterogeneity, with one main reason being its dual protein aggregation tendency.

Investigations addressing the structure, molecular properties and processing of wild type and mutant TGFBIp commonly found in *TGFBI* corneal dystrophies, led to multiple hypotheses regarding possible protein aggregation mechanisms and corneal deposition.

During normal physiological processing, the TGFBIp undergoes extensive N-terminal processing by the proteolytic machinery in the cornea, leading to FAS1-4-containing products, which are subsequently cleared. Changes in the proteolytic processing and protein turnover have been implicated as the cause of deposit formation in certain *TGFBI* CDs, such as those associated with the R124C, R124H and R124L mutations (Korvatska E., Henry, Mashima, *et al*., 2000). Intriguingly, researchers showed that each of the mutations mentioned above was associated with different and distinctive protein turnover patterns. In R124C corneas it was revealed that the amount of full size TGFBIp (68-kDa) did not vary significantly from that of wild type corneas while the amount of N-terminal 44-kDa fragments was increased significantly. On the other hand, corneas with the R124H mutation contained an accumulation of 66-kDa fragments and twice the amount of full size TGFBIp while in corneas with the R124L mutation, only the full size TGFBIp was found to be accumulated. Consequently they hypothesised that various replacements of Arginine-124 in the *TGFBI* gene influence the processing

and post-translation modifications of the TGFBIp. They concluded that this possibly occurred due to misfolding and/or aggregation that led to alteration of the sites at which TGFBIp proteolysis usually occurs hence resulting in abnormal turnover of the protein that in turn would lead to the accumulation of the abnormal proteolytic fragments. This hypothesis was also pursued by Karring *et al*., where they suggested that the mutated TGFBIp can be altered in a way that causes a subsequent alteration in its proteolytic cleavage and a cause deposit formation, such as in the V624M mutation (Karring H., Runager, Thogersen, Klintworth, *et al*., 2012).

Another suggestion, this time brought about by Nielsen, Poulsen, Lukassen, *et al*., 2020, is that protein aggregation leading to formation of deposits in *TGFBI* -linked CDs could be concentration-dependent. Normally, the insoluble fraction of TGFBIp constitutes about 60% of the total TGFBIp present in the cornea (Andersen, Karring, Moller-Pedersen, *et al*., 2004). In this research they hypothesised that a shift in the ratio towards an increase in soluble TGFBIp protein will probably precipitate aggregation of these deposits (Nielsen, Poulsen, Lukassen, *et al*., 2020). It is also possible that mutations can induce abnormal protein turnover leading to higher concentrations of TGFBIp, which in turn could tip the balance towards deposits (Stenvag, Andreasen, Enghild, *et al*., 2013).

It has also been shown that TGFBIp is monomeric at low concentrations; however, at high concentrations it forms trimers, which have a tendency to form oligomers that could lead to the formation of amorphous aggregates (Basaiawmoit, Oliveira, Runager, *et al*., 2011). Furthermore, it is known that a protein's stability can be compromised when amino acid substitutions occur, especially if these involve the protein core. This

occurs due to side-chain bond formations (Fersht, 1999). Studies showed that mutations in the Fas 1-4 domain cause destabilisation that could lead to amyloid deposition or hydrophobic collapse, with subsequent amorphous aggregate formation (Runager, Basaiawmoit , Deva, *et al*., 2011) (Andreasen, Nielsen, Runager, *et al*., 2012).

It has also been suggested that the deposited mutated protein over time is damaging to cells due to the increased frequency of degenerate epithelial and stromal cells in the dystrophies. Hypothetically, cell death could add proteolytic factors, further enhancing abnormal TGFBIp polymerization (Streeten, Qi, Klintworth, *et al*., 1999).

Another proposed aggregation mechanism is that a mutation can cause a change in the charge of the protein leading to an alteration in the proteins' electrostatic interaction with other molecules. This is seen in the presence of the L550P mutation, which causes destabilisation due to a change in the charge, resulting in the formation of granular deposits (Stenvag, Andreasen, Enghild, *et al*., 2013). In contrast, the R555W, also responsible for another type of GCD, removes a charge, causing the mutated protein to become more stable (Runager, Basaiawmoit , Deva, *et al*., 2011).

Runager *et al*. carried out trials focussing on the stability of mutated TGFBIp (Runager, Basaiawmoit , Deva, *et al*., 2011). As stated previously, the majority of *TGFBI* CD mutations are missense mutations located in the FAS1 domains. The mutation hotspot Arg-124 is situated in the turn between helices α1 and α2 of FAS1 domain. On the other hand, Arg-555 is found in the helices α3-α4 of FAS1 domain 4, which is the most variable region of the FAS1 fold (Clout, Tisi, & Hohenester, 2001). Runager *et al*. (Runager, Basaiawmoit , Deva, *et al*., 2011), concluded that amino acid changes in FAS1-1 did not affect the stability of the intact TGFBIp and that they were relatively

protected against proteolysis. However, mutations in the FAS1-4, do affect the stability of TGFBIp. The *TGFBI* A546T mutation causes the protein to exhibit an increased tendency to fibrillation. On the other hand, the R555Q and R555W mutation do not cause TGFBIp to form amyloid fibrils under physiologic conditions (Kannabiran & Klintworth, 2006) (Andreasen, Nielsen, Runager, *et al*., 2012). The A546T mutated TGFBIp is considered to be partially unfolded, thus exposing hydrophobic areas in the protein which can lead to aggregation (Andreasen, Nielsen, Runager, *et al*., 2012). The stability of isolated FAS1-4 domains was found to be similar to the whole intact protein, thus, this may indicate that the stability of FAS1-4 may play a role in the development of the deposits in corneal dystrophies where the mutation lies in the FAS1-4 domain (Runager, Basaiawmoit , Deva, *et al*., 2011). On the other hand, since amino acid substitutions in FAS1-1 did not affect stability, this suggests that the molecular mechanism of corneal dystrophies caused by mutations in FAS1-1 would be different. Runager *et al*. also proposed that corneal dystrophy phenotypes might be linked to the thermodynamic stability of TGFBIp (Runager, Basaiawmoit , Deva, *et al*., 2011).

Even though R555W (GCD1) is considered to be a stabilizing mutation, it was proposed that this mutation reduced solubility of the folded monomeric state of TGFBIp (Stenvang, Schafer, Malmos, *et al*., 2018). It is thought that the replacement of the charged side-chain arginine at position 555, with a large aromatic side-chain tryptophan, will increase the protein's tendency to precipitate since it is highly exposed (Stenvang, Schafer, Malmos, *et al*., 2018). Additionally it has been shown that the mutant R555W TGFBIp shows less susceptibility to proteolysis than the wild type form. Researchers noticed that while the Arg555 residue was exposed in the wild type, the Trp555 mutation was buried in the core of the fourth FAS-1 domain, which could

be the reason for the increased proteolytic resistance observed (Underhaug, Koldso, Runager, *et al*., 2013). On the other hand, variant GCD2 mutations such as M619K and L550P result in a decrease in the protein stability rather than solubility, leading to a faster non-amyloid aggregation pathway. This contrasts with LCD mutants that basically increase the tendency of the protein to unfold.

Recent studies focusing on the pathogenesis of GCD2 proposed that GCD2 corneal fibroblasts show increased susceptibility to oxidative stress, age-dependent alteration of mitochondria, and a defective autophagy when compared to wild type (Choi, Kim, Dadakhujaev, *et al*., 2012). The cornea is constantly exposed to ultraviolet (UV) light leading to the generation of reactive oxygen species (ROS) which in turn causes oxidative injury. In fact, it has been documented that the levels of intracellular ROS and H2O2 were significantly elevated in GCD2 fibroblasts, as were the expression levels of various antioxidant enzymes. Interestingly it was reported that in spite of an increase in the mRNA level of catalase, the expression of catalase protein was decreased in these cells, and cell viability after exposure to H2O2 was also decreased, implying increased susceptibility to oxidative damage (Choi, Kim, Kim, *et al*., 2009). Furthermore, secretion of TGFBIp has been found to be delayed in corneal fibroblasts in GCD2 corneas. Studies also revealed that accumulation of mutant TGFBIp in the cytoplasmic or lysosomal compartments of fibroblasts may occur due to defective autophagy or delayed fusion between autophagosomes and lysosomes (Choi, Maeng, Kim, *et al*., 2015). Additionally, in another study, it was reported that 555 genes were found to be significantly differentially expressed in homozygous GCD2 fibroblasts when compared to wild type (Choi, Yoo, Kim, *et al*., 2010). Indeed the intricacy of inter-

related gene functions is possibly one of the factors accounting for the cellular changes recorded in the GCD2 fibroblasts.

Literature has also shown that the amount of TGFBIp in the cornea increases with age and that there are changes in the processing of the TGFBIp during postnatal maturation of the cornea. Thus, this lead to the hypothesis that age related changes in expression and processing of the *TGFBI* protein may play a role in the age of onset of *TGFBI* -associated corneal dystrophies (Karring, Runager, Valnickova, *et al*., 2009).

To summarize, various hypotheses exist regarding the cause of mutated TGFBIp deposits in *TGFBI*-linked CDs with proposals varying according to the type of mutation present. Some state that it is concentration-dependent while others postulate that certain *TGFBI* mutations affect protein-protein interactions directly while other mutation variants cause misfolding of the protein and altered TGFBIp stability and solubility (Clout & Hohenester, 2003). Additionally, other proposals focussed on the increased susceptibility of fibroblasts to oxidative stress, and the possibility of a defective autophagy leading to altered clearance of the mutant deposits (Choi, Kim, Dadakhujaev, *et al*., 2012).

However, fascinatingly, even though mutated TGFBIp is produced by cells in many organs in the body, deposits have only been located in the cornea (Runager, Enghild, & Klintworth, 2008) (Kochairi, *et al*., 2006). This supports the suggestion given by multiple researchers that there must be corneal-specific factors to trigger the deposition of mutated TGFBIp in corneal dystrophy patients (Kochairi, Letovanec, Uffer, *et al*., 2006) (Kim, Park, Choi, *et al*., 2002).

2.6.5 Tissue Location

Various studies have been carried out to try and better our understanding of TGFBIp by identifying the tissues where it is expressed. Researchers have found that *TGFBI* is expressed by cells in many organs throughout the human body (Skonier, Neubauer, Madisen, *et al*., 1992) (Ivanov, Ivanova, Salnikow, *et al*., 2008). The intensity of expression of this gene is probably regulated by the cell environment, considering the fact that lung fibroblasts have a low level of TGFBIp when compared to the high level of expression in primary dermal fibroblasts. In the human skin, TGBFIp expression has been localised to the granular layer of the epidermis (LeBaron, Bezverkov, Zimber, *et al*., 1995). Experiments involving the kidney and vascular smooth muscle also confirmed the expression of *TGFBI* in these tissues. It was also proven that TGFBIp mediated renal proximal tubular epithelial cells' adhesion and migration (Lee, Bae, Park, *et al*., 2003). In a similar experiment, this time involving vascular smooth muscle, again, it was found that TGFBIp mediated the adhesion, spreading, migration, and proliferation of rat vascular smooth muscle cells (Ha, Bae, Yeo, *et al*., 2003).

In the human eye, *TGFBI* is transcribed almost exclusively in the corneal epithelium and to a lesser extent in the stromal keratocytes, and its protein is found in abundance in corneal stroma (Hartz & McKusick, 2015) (Stenvag, Andreasen, Enghild, *et al*., 2013). TGFBIp is the second most abundant protein in the human corneal stroma and is present in all of the corneal layers (Dyrlund, Toftgaard Poulsen, Scavenius, *et al*., 2012). TGFBIp is produced in constant amounts by human corneal epithelial cells under normal conditions. However, following injury, an increase in local TGF-β occurs, which in turn causes an increase in TGFBIp synthesis by the corneal epithelium and by the

stromal keratocytes. Of note is that the expression of TGFBIp increases by 30% during the first to second decade of life (Karring, Runager, Valnickova, *et al*., 2009).

Interestingly, as mentioned above, even though TGFBIp is expressed in many tissues, the deposition of TGFBIp in individuals carrying mutations in the *TGFBI* gene appears to be specific to the cornea (Runager, Enghild, & Klintworth, 2008) (Kochairi, Letovanec, Uffer, *et al*., 2006). This leads to the postulation that for corneal deposits to occur in *TGFBI* CDs, a unique extracellular environment in the cornea is necessary (Nielsen, Poulsen, Lukassen, *et al*., 2020).

Further studies to investigate tissue specificity of pathologic TGFBIp deposition were carried out by Kochairi *et al*. (Kochairi, Letovanec, Uffer, *et al*., 2006). They confirmed that, except for the cornea, immunostaining with KE2 antisera did not reveal any deposits in any of the other 17 organs tested. This makes one question whether the breakdown of mutated TGFBIp in the cornea, produces fragments that form deposits or whether the process of clearing of TGFBIp fragments is different in affected corneas, which in turn causes aggregation and deposition of material. Also, one has to keep in mind the unique features of the cornea, namely, its avascularity and its particular extracellular matrix structural design and composition. Its avascularity limits the rate of transport of molecules locally and causes it to have a relatively slow turnover of proteins, possibly leading to the deposition of the mutated TGFBIp complexes (Korvatska, Henry, Mashima, *et al*., 2000).

Also, of note is the observation that usually there is a clear zone with no deposits, between the limbus and the deposits leading researchers to propose that the presence

of blood vessels might prevent the deposition of abnormal TGFBIp (Han, Choi, Kim, *et al*., 2016).

2.7 Management of TGFBI Corneal Dystrophies

Presently, the only treatment available for the management of *TGFBI* CD patients consists of supportive medical treatment to relieve discomfort, and surgery to treat visual loss. Ideally, the management of *TGFBI* CDs should revolve around preserving functional vision for as long as possible and minimizing surgical procedures.

Individuals suffering from *TGFBI* CDs may experience painful recurrent epithelial erosions when the dystrophic deposits are located in the outermost corneal laters. This symptom is initially managed with preservative free lubricating drops, ointments and therapeutic contact lenses. Lubricating drops that are free of benzalkonium chloride (BAC) should be used, since, it has been suggested that BAC might increase amyloidogenicity and hence deposit formation (Kato, Yagi, Kaji, *et al*., 2013). In severely symptomatic cases and when superficial corneal stromal opacification is impeding vision, superficial keratectomy or laser phototherapeutic keratectomy can be applied. Severe visual acuity deterioration due to the presence of deposits in the mid/posterior stroma is usually managed by corneal graft surgery (lamellar keratoplasty or full thickness penetrating keratoplasty). Unfortunately, each surgical technique can be associated with a number of serious complications, such as graft rejection in corneal graft surgery. Furthermore, recurrence of the dystrophy within the graft has been reported to be almost universal within a few years (Lyons, McCartney, Kirkness, *et al*., 1994). As a result, patients might have to undergo multiple procedures over a lifetime due to recurrence of the dystrophy, resulting in a poor quality of life for these individuals. Researchers are now trying to discover more permanent and non-surgical procedures to try and prevent or stop the deposition of corneal material in these dystrophies.

Minimizing surgical interventions and evading their potential complications would definitely benefit *TGFBI* CD patients. The introduction of gene therapy or drugs as a non-surgical treatment option in CDs would be a breakthrough.

2.8 Gene Expression in the Cornea

Each cell in an organism contains nearly the same genome; however, a cell expresses only a fraction of its genes. Consequently, the particular set of genes being expressed or repressed in the cell dictates its morphology and function.

Human corneal protein data sets provide a reference when analysing the molecular mechanisms involved in normal corneal homeostasis and diseases. On the other hand, RNA sequencing provides additional data on mRNA abundance, alternative splicing, nucleotide variation, and structure alteration. Even though, at the quantitative level, protein and mRNA levels are only moderately correlated, combining these two datasets validates the relevance of the results derived from the RNA-seq data as well as provides information about gene expression in posttranscriptional regulation (Wang, Liu, & Zhang, 2014).

Initial studies focused on identifying the proteins present in all the layers of the cornea. The first human cornea proteome study identified 138 unique Swiss-Prot proteins (Karring, Thogersen, Klintworth, *et al*., 2005). Subsequent studies employing more advanced technology identified 1555 unique Swiss-Prot proteins (Galiacy, Froment, Mouton-Barbosa, *et al*., 2011). Since most corneal conditions involve mainly a single layer, in order to further comprehend the molecular mechanisms involved in normal corneal homeostasis and corneal diseases, it is imperative to study the individual layers of the human cornea separately. In a study carried out by Dyrlund, Toftgaard Poulsen, Scavenius, *et al*. (2012), the proteins present in the epithelium, stroma and endothelium were identified and quantified. 2737 unique Swiss-Prot annotated proteins were identified in the human corneal epithelium and 1679 in the

stroma by tandem mass spectrometry following separation by sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) in this research. However, since not all corneal proteins are soluble in SDS, in this study by Dyrlund *et al*., proteins were also quantified based on an in-solution digest and subsequently using the Average Protocol in Mascot Distiller to calculate the intensities of the identified proteins. RNA-seq analysis of the corneal epithelium was carried out in a study by You, Corley, Wen, *et al*. (2018) where they compared epithelial gene expression between keratoconus and myopia patients. After filtering, they identified 11655 protein-coding genes in the control samples. Advancing technology led to the availability of single cell RNA sequencing (scRNA-seq) where the expression profiles of individual corneal cells can be obtained (Collin, Queen, Zerti, *et al*., 2021).

Epithelial keratins are found abundantly in the epithelial layer, with K3 and K12 being reported to be corneal specific keratins (Irvine, Corden, Swensson, *et al*., 1997). In a study carried out by Diehn, Diehn, Marmor, *et al*. (2005) utilizing DNA microarrays, numerous genes that were found to be enriched in the corneal signature encoded proteins that modulate cell-cell adhesion and stabilize epithelial sheets, such as, keratins (KRT5, KRT6B, KRT13, KRT15, KRT16, KRT17, KRT19), laminins, and desmosomal components (Diehn, Diehn, Marmor, *et al*., 2005). Other highly expressed genes encoded proteins that are vital for the maintenance of corneal structure and transparency. These included genes encoding for collagens (*COL1A1, COL5A2, COL6A3, COL12A1, COL17A1*), keratocan (*KERA*), lumican (*LUM*), aquaporin 3 (*AQP3*) and the gene for lysyl oxidase (*LOX*), an enzyme that promotes collagen cross-linking. Collagen alpha-1(I) chain was reported to be the most abundant protein in the human corneal stroma while TGFBIp was found to be the second most abundant

protein in the stroma and was identified in all of the corneal layers (Dyrlund, Toftgaard Poulsen, Scavenius, *et al*., 2012).

In the human eye, *TGFBI* is transcribed almost exclusively in the corneal epithelium and its protein is found in abundance in corneal stroma (Poulsen, Runager, Nielsen, *et al*., 2018) (Hartz & McKusick, 2015) (Stenvag, Andreasen, Enghild, *et al*., 2013). Thus, this indicates that TGFBIp is produced in the epithelium and then diffuses into the underlying stroma. About 60% of TGFBIp present in the human cornea (~65 kDa) is covalently associated with insoluble components of the extracellular matrix (Andersen, Karring, Moller-Pedersen, *et al*., 2004). TGFBIp, serum albumin, and immunoglobin κ chain have been reported to be present in the cornea in a significant number of isoforms. This indicates that post-translational additions and fragmentations take place routinely in the normal human cornea.

Interestingly, even though the cornea is avascular, a considerable number of plasma proteins were quantifiable in the human cornea including albumin, alpha-1-antitrypsin, immunoglobulin's, serrotransferrin, complement component 3 and 9 and haptoglobin. The presence of haemoglobin subunits was also detected; however, the function of these plasma proteins in the cornea still remains unclear (Dyrlund, Toftgaard Poulsen, Scavenius, *et al*., 2012).

The cornea also acts as a protective barrier against infectious organisms or physical trauma. Many of the proteins detected in the human cornea by Dyrlund, Toftgaard Poulsen, Scavenius, *et al*. (2012) were plasma proteins involved in defence responses (Table 6). These may play a role in the maintenance of corneal homeostasis and in the pathogenesis of corneal diseases. Proteins of the classical and alternative complement

activation pathways have been detected in the human cornea, as well as several complement regulatory proteins. It has been hypothesised that the latter are essential in the negative regulation of complement for prevention of autologous damage (Jha, Bora, & Bora, 2007). C-reactive protein has been identified in all layers of normal corneas (Dyrlund, Toftgaard Poulsen, Scavenius, *et al*., 2012) and was previously also documented to be present in the epithelium of corneas with lattice dystrophy (Rodrigues & Robey, 2006).

Corneal wound healing is regulated by the coagulation cascade with consequent fibrin deposition. However, the balance of fibrin deposition and degradation, which is controlled by the fibrinolytic system, is indispensable for maintaining corneal function.

Table 6: The most abundant proteins found in the human cornea (Dyrlund, Toftgaard Poulsen, Scavenius, *et al*., 2012) .

	Epithelium	**% and SD**	**Stroma**	**% and SD**	**Endothelium**	**% and SD**
1.	Keratin, type I cytoskeletal 12	8.1 ± 0.7	Collagen alpha-1(I) chain	20.0 ± 2.0	Transforming growth factor-beta-induced protein ig-h3	36.8 ± 1.7
2.	Keratin, type II cytoskeletal 5	7.3 ± 0.3	Transforming growth factor-beta-induced protein ig-h3	17.6 ± 3.3	Collagen alpha-1(VIII) chain	5.8 ± 0.2
3.	Keratin, type II cytoskeletal 6A	6.0 ± 0.5	Collagen alpha-2(I) chain	17.2 ± 3.1	Collagen alpha-2(VIII) chain	4.4 ± 0.5
4.	Keratin, type II cytoskeletal 3	4.7 ± 0.5	Decorin	5.1 ± 0.4	Collagen alpha-4(IV) chain	4.2 ± 0.7
5.	Keratin, type II cytoskeletal 2 oral	4.7 ± 0.3	Collagen alpha-3(VI) chain	4.7 ± 0.9	Collagen alpha-3(IV) chain	4.0 ± 0.3
6.	Keratin, type I cytoskeletal 15	3.8 ± 0.4	Lumican	3.5 ± 0.5	Collagen alpha-1(I) chain	3.9 ± 0.4
7.	Keratin, type II cytoskeletal 4	3.3 ± 0.1	Serum albumin	3.5 ± 0.5	Keratocan	3.5 ± 0.3
8.	Keratin, type I cytoskeletal 19	3.1 ± 0.8	Keratocan	3.4 ± 0.9	Collagen alpha-5(IV) chain	3.0 ± 0.4
9.	Aldehyde dehydrogenase, dimeric NADP-preferring	3.1 ± 0.4	Collagen alpha-1(VI) chain	2.2 ± 0.5	Clusterin	2.5 ± 0.2
10.	Alpha-enolase	3.0 ± 0.4	Ig gamma-1 chain C region	1.9 ± 0.1	Thrombospondin-4	2.3 ± 0.5

11.	Keratin, type I cytoskeletal 14	2.9 ± 0.3	Collagen alpha-2(VI) chain	1.7 ± 0.2	Collagen alpha-2(I) chain	2.3 ± 0.3
12.	Keratin, type II cytoskeletal 8	2.5 ± 0.2	Collagen alpha-1(XII) chain	1.6 ± 0.6	Ig gamma-1 chain C region	2.2 ± 0.4
13.	Histone H4	2.1 ± 0.4	Ig kappa chain C region	1.6 ± 0.3	Serum albumin	1.9 ± 0.4
14.	Histone H2B type 1-C/E/F/G/I	1.4 ± 0.1	Ig gamma-3 chain C region	1.5 ± 0.3	Ig gamma-3 chain C region	1.7 ± 0.1
15.	Transketolase	1.4 ± 0.2	Ig gamma-2 chain C region	1.3 ± 0.2	Ig gamma-4 chain C region	1.4 ± 0.1
16.	Actin, cytoplasmic 1	1.3 ± 0.1	Collagen alpha-2(V) chain	1.3 ± 0.2	Ig gamma-2 chain C region	1.3 ± 0.1
17.	Histone H2B type 1-B	1.3 ± 0.1	Ig gamma-4 chain C region	1.2 ± 0.2	Fibulin-5	1.2 ± 0.2
18.	Collagen alpha-1(I) chain	1.1 ± 0.1	Collagen alpha-1(V) chain	1.1 ± 0.2	Ig kappa chain C region	1.2 ± 0.2
19.	Annexin A2	1.1 ± 0.6	Aldehyde dehydrogenase, dimeric NADP-preferring	0.9 ± 0.2	Apolipoprotein D	0.9 ± 0.1
20.	Heat shock protein beta-1	1.1 ± 0.1	Biglycan	0.6 ± 0.1	Decorin	0.9 ± 0.1
21.	Transforming growth factor-beta-induced protein ig-h3	1.0 ± 0.3	Alpha-enolase	0.6 ± 0.1	Collagen alpha-3(VI) chain	0.9 ± 0.2
22.	Serum albumin	1.0 ± 0.1	Collagen alpha-1(III) chain	0.5 ± 0.1	Ig lambda-2 chain C regions	0.8 ± 0.0
23.	Keratin, type I cytoskeletal 13	0.9 ± 0.1	Apolipoprotein D	0.4 ± 0.1	Immunoglobulin lambda-like polypeptide 5	0.7 ± 0.0

24.	Glyceraldehyde-3-phosphate dehydrogenase	0.7 ± 0.0	Clusterin	0.4 ± 0.3	Lumican	0.7 ± 0.1
25.	L-lactate dehydrogenase A chain	0.7 ± 0.1	Alpha-1-antitrypsin	0.4 ± 0.1	Myocilin	0.6 ± 0.0
26.	Protein S100-A4	0.7 ± 0.1	MAM domain-containing protein 2	0.3 ± 0.0	Apolipoprotein E	0.6 ± 0.1
27.	Pyruvate kinase isozymes M1/M2	0.6 ± 0.0	Prolargin	0.2 ± 0.0	Serine protease HTRA1	0.6 ± 0.1
28.	Retinal dehydrogenase 1	0.6 ± 0.1	Vimentin	0.2 ± 0.0	Alpha-enolase	0.5 ± 0.1
29.	Elongation factor 1-alpha 1	0.6 ± 0.1	Ig alpha-1 chain C region	0.2 ± 0.0	Collagen alpha-1(III) chain	0.5 ± 0.0
30.	Ferritin heavy chain	0.6 ± 0.0	Alpha-1-antichymotrypsin	0.2 ± 0.0	Collagen alpha-2(VI) chain	0.4 ± 0.1

Research by Nishida, Adachi, Shimizu-Matsumoto, *et al*. (1996) exploring the expression profile of human corneal epithelium by constructing a cDNA library, showed that the most abundant transcripts were those for apolipoprotein J (clusterin), calcyclin, alpha-enolase, keratin 3 and connexin 43. They also identified active genes that were unique to the corneal epithelium. These included keratin 3, aldehyde dehydrogenase 3 (ALDH3), and three novel sequences, one of which probably being keratin 12. It is thought that the function of non-unique genes is primarily for 'housekeeping' functions, such as the genes for ribosomal proteins (Nishida, Adachi, Shimizu-Matsumoto, *et al*., 1996). On the other hand, unique genes are probably important for tissue-specific functions. These findings compare well with those reported by Dyrlund, Toftgaard Poulsen, Scavenius, *et al*. (2012). However, of note is that even though clusterin and calcyclin were reported as two of the most abundant transcripts by Nishida *et al*., these were not documented as most abundant proteins in the epithelium by Dyrlund *et al*.. Having said that, clusterin was reported to be abundant in the stroma and endothelium of the cornea (Dyrlund, Toftgaard Poulsen, Scavenius, *et al*., 2012). RNA-seq analysis of the corneal epithelium (You, Corley, Wen, *et al*., 2018) reported that the genes with the highest expression included: Keratin 3 (*KRT3*), Keratin 5 (*KRT5*), Keratin 12 (*KRT12*), aldehyde dehydrogenase 3 family member A1 (*ALDH3A1*), clusterin (*CLU*) and enolase 1 (*ENO1*).

Apolipoprotein J (clusterin) is found abundantly in the cornea. It is known to have a clear role in proteostasis and cytoprotection as well as in preventing inflammation (Aronow, Lund, Brown, *et al*., 1993). It has been described as being a potent heat shock protein-like chaperone that functions by inhibiting stress-induced amorphous protein aggregation and the fibrillar aggregation of many amyloidogenic proteins and

peptides. By forming a soluble complex with these proteins, it inhibits their precipitation (Humphreys, Carver, Easterbrook-Smith, *et al*., 1999). Research has shown that at low concentrations, apolipoprotein J incorporates into amyloid deposits. It is hypothesized that this occurs while attempting to inhibit its precipitation. However, if present in a sufficient concentration it inhibits amyloid formation and provides cytoprotection (Yerbury, Poon, Meehan, *et al*., 2007). Interestingly, apolipoprotein J has been found to be present in corneal deposits in GCD1, 2 and LCD (Karring, Runager, Thogersen, *et al*., 2012).

The high-level expression of calcyclin, which is enhanced during corneal wound healing, has been associated with the rapid growth of the corneal epithelial basal layer. This, in turn, helps renew the corneal epithelial layers (Bazan, Allan, & Bazan, 1992).

Alpha-enolase, an enzyme in the glycolytic pathway, had been reported as one of the most abundant proteins in the cornea in a proteomic study carried out by Dyrlund, Toftgaard Poulsen, Scavenius, *et al*. (2012) which tallies with the abundant expression of the alpha-enolase gene documented by Nishida, Adachi, Shimizu-Matsumoto ,*et al*. (1996) and You, Corley, Wen, *et al*. (2018). This discovery is interesting in that no other genes for the enzymes in the glycolytic pathway were correspondingly active in cornea. Thus, this might point towards alpha-enolase having another role other than energy production in the cornea.

Besides the genes coding for keratin 12 and keratin 3, a number of active genes for surface membrane proteins are also considered to be characteristic of corneal tissue. These include integrin, tyrosine kinase receptor and laminin. Similar to what was

reported in the proteomic study by Dyrlund *et al*., other active genes of note in the corneal epithelium are those of defence proteins.

Genes and biological pathways that are preferentially or uniquely active in the cornea are key players in the unique morphology and function of corneal cells. This specificity will be indeed reflected in disease states. A clear example of this is are actually *TGFBI* CDs where even though the mutation is found in all the cells of affected individuals, studies have shown that deposits only occur in the cornea (Kochairi, Letovanec, Uffer, *et al*., 2006).

2.9 Epidemiology of Corneal Dystrophies

Most of the CDs cause a decrease in visual acuity and even corneal blindness (Klinthworth, 2009) (Bron A. J., 1990). Thus, determining the frequency of these disorders and the various allele mutations responsible for them is of significant importance.

Traditionally, researchers have traditionally described how frequently a genetic disease occurs in a population by observing and diagnosing the disease clinically. This is usually done by using the standard epidemiological measures such as prevalence, incidence and lifetime risk (Musch, Niziol, Stein, *et al*., 2011). Attempts have been made to try and estimate the prevalence of CDs (Siddiqui, 2016) (Disorders) primarily by using information from corneal transplant registries, registries from hospitals and ophthalmologists (Moller H. U., 1990) (Lee, Cristol, Kim, *et al*., 2010) (Chae, Kim, Kim, *et al*., 2016), from previous published reports and case studies (Chao, DeDionisio, Ke, *et al*., 2019), and from administrative health care claims databases (Musch, Niziol, Stein, *et al*., 2011). Nonetheless, all of these data collection methods have limitations.

Relying solely on data derived from clinical diagnoses leads to an underestimation of all epidemiological measures since individuals who are asymptomatic in the early stages may be undiagnosed or incorrectly diagnosed. Bias towards the severe or symptomatic/advanced CD cases will be found in the corneal transplant registries and hospital registries. On the other hand, generating epidemiological measure from published reports and stratifying by region will result in an underestimation in countries with limited research and publishing opportunities (Chao, DeDionisio, Ke, *et al*., 2019). Similarly, the main limitation of epidemiological estimates from health care insurance claims is that only diagnosed cases are identified, leading to an

underestimation of cases (Crisafulli, Sultana, Ingrasciotta, *et al*., 2019). Despite all this, these data sources play a crucial role in providing epidemiological estimates that subsequently can be compared with each other.

Musch, Niziol, Stein, *et al*., (2011) calculated that the overall prevalence of CDs in the United States is 897 per million, using data from health care insurance claims. The most prevalent subset was endothelial CDs (specifically, Fuch's endothelial dystrophy [FECD]), which comprised 60% of all CDs (Musch, Niziol, Stein, *et al*., 2011). In the UK, FECD is also common, accounting for 22% of corneal transplants performed there (Keenan, Jones, Rushton, *et al*., 2012). Lattice corneal dystrophy (LCD) has been found to be one of the more common forms of the *TGFBI* dystrophies in Western countries. Granular corneal dystrophy type 1 (GCD1), also a *TGFBI* dystrophy, is said to be more prevalent in Europe than in other regions, while on the other hand GCD2 is more prevalent in Japan, Korea and the US (Fujiki & Nakayasu, 2006) (Klinthworth, 2009) (Lee, Cristol, Kim, *et al*., 2010); however, the actual prevalence of many of the CDs is still unknown (Weiss, Moller, Aldave, *et al*., 2015).

The recent introduction of large population genetic databases describing genetic variants from various ethnicities, has provided a wealth of information regarding allelic frequencies in populations. It helps to determine the odds of a sequence variant being pathogenic, especially when considering variants that are suspected to cause an AD disease. These databases make it possible to theoretically calculate the prevalence of a genetic disease, based on the causal allele frequency in that population. This is possible since non-synonymous single nucleotide variants (SNVs) can affect protein structure, stability and function (Dobson, 2003). These may cause diseases due to

changes in protein-protein interactions (Teng, Madej, Panchenko, *et al*., 2009), or changes in the features of the active site of the protein (Zhang, Teng, Wang, *et al*., 2010). Furthermore, large population genetic databases can also provide insight regarding current under-diagnosis, especially in such rare conditions, which could guide clinicians towards improving their diagnosis rates (Table 7).

More research is needed to form a clear picture of the worldwide frequency of CDs. Indeed, the prevalence of CDs in Malta is also unknown, making it difficult to estimate the clinical needs of persons with this disease. Thus, in order to address this research gap, the first phase of this research project consisted of an epidemiological study exploring the previously unknown CD genetic pool present in Maltese individuals, a subgroup of the human population on which CD-related literature is still lacking.

Table 7: Worldwide prevalence (orphanet) (and from literature), inheritance type and causal gene/s of the major corneal dystrophies.

Corneal Dystrophy	**Inheritance type**	**Chromosome**	**Causal gene/s**	**Worldwide Prevalence**
Epithelial and subepithelial dystrophies:				
Meesmann dystrophy	AD	12q13 17q12	*KRT3* *KRT12*	<1 per 1,000,000
Gelatinous drop-like dystrophy	AR	1p32	*TACSTD2*	Unknown 1 per 33 000 in Japan ((Kaza, Barik, Reddy, *et al*., 2017)
Lisch epithelial dystrophy	X-chr dominant	Xp22.3	Unknown	<1 per 1,000,000
Epithelial-stromal TGFBI dystrophies:				
Reis–Bücklers corneal dystrophy	AD	5q31	*TGFBI*	<1 per 1,000,000
Thiel-Behnke corneal dystrophy	AD	5q31	*TGFBI*	Unknown
Lattice corneal dystrophy	AD	5q31	*TGFBI*	Unknown One of the more common forms in Western countries (Klinthworth, 2009).
Granular corneal dystrophy type 1	AD	5q31	*TGFBI*	Unknown More prevalent in Europe than in other regions (Klinthworth, 2009).

Granular corneal dystrophy type 2	AD	5q31	*TGFBI*	Unknown More prevalent in Japan, Korea and the USA (Klinthworth, 2009). 11.5 per 10,000 in Korea (Lee, Cristol, Kim, *et al*., 2010)
Stromal dystrophies:				
Macular corneal dystrophy	AR	16q22	*CHST6*	Unknown
Schnyder corneal dystrophy	AD	1p36	*UBIAD1*	<1 per 1,000,000
Fleck corneal dystrophy	AD	2q34	*PIKFYVE*	<1 per 1,000,000
Endothelial dystrophies:				
Early-onset Fuchs endothelial corneal dystrophy Late-onset Fuchs endothelial corneal dystrophy		1p34.3-p32 Association to: 20p13-p12, 10p11.2, 18q21.2-q21.3,	*COL8A2* *ZEB1* *SLC4A11* *LOXHD1*	Unknown Extreme geographical variability. The most prevalent CD in the USA (Musch, Niziol, Stein, *et al*., 2011). Very rare in Japan ((Santo, Yamaguchi, Kanai, *et al*., 1995).
Posterior polymorphous corneal dystrophy	AD	PPCD1: 20p11.2-q11.2 PPCD2: 1p34.3-p32.3 PPCD3: 10p11.2	PPCD1: unknown. PPCD2: *COL8A2* PPCD 3: *ZEB1* Promoter of the *OVOL2*	Unknown Very rare. Higher than normal prevalence in the Czech Republic (1 per 100 000) (Liskova, Gwilliam, Filipec, *et al*., 2012)
Congenital hereditary endothelial dystrophy	AR	20p13	*SLC4A11*	Unknown

2.10 Aims of this Research Study

The three aims of this project were:

a) To establish which worldwide populations have a corneal dystrophy (CD) genetic makeup closest to that of the Maltese.

b) To perform clinical exome sequencing analysis of mouthwash samples from members of a Maltese family that phenotypically exhibit granular GCD1, a subtype of *TGFBI* CDs, in order to identify the causative mutation and any potentially associated mutation variants in this Maltese cohort, with the aim of providing a guide for future targeted treatment.

c) To explore *TGFBI* inhibition in HCECs as a treatment modality for *TGFBI* CDs and analyse the effect *TGFBI* inhibition would have in the overall homeostasis of HCECs.

2.10.1 Objectives

In order to address the first aim of this project, a cross sectional analytical study to identify and analyse variants and their allele frequency in known causal CD genes in a representative cohort of the Maltese population, was carried out. This helped to obtain insight and address the gap in the existing literature regarding the previously unknown Maltese CD genetic pool and epidemiology. The frequencies of CD-related single nucleotide variants (SNPs) that were found in this Maltese cohort were compared with those of genotypically diverse worldwide cohorts. The results established which worldwide population cohorts have a CD-related genetic make-up that is closest to that of a relatively large sample of the Maltese population. By identifying populations with the least genetic differentiation, I would be providing a guide for future diagnostic and treatment strategies for Maltese individuals with CDs, especially when no

comparable Maltese data are available. The potential pathogenicity of missense single nucleotide variations present in this Maltese cohort was also analysed by using predictive algorithmic tools.

The second aim focused around *TGFBI* CDs; specifically GCD1. The objective was to perform clinical exome sequencing on buccal mouthwash samples of Maltese individuals that phenotypically exhibited GCD1 that were members of the same family. This was performed in order to identify which GCD1 genetic variant is present in Malta and to explore whether any other mutations present in these individuals could contribute to their phenotype. In order to reach this objective, the genetic mutation variants present in the Maltese cohort exhibiting GCD1 were filtered against a list of genes that have been documented in literature to be either co-expressed, known to interact genetically or are involved in common pathways with the *TGFBI* gene or GCD1 deposits. Theoretically, the resulting genetic mutation variants identified might play a role in the phenotype exhibited in this particular cohort of GCD1 Maltese individuals. This would shed light on the possible causes of the intra and inter-familial phenotypic variations that have been reported multiple times in literature (Hou, Hsiao, Chen, *et al*., 2012) (Kattan, Serna-Ojeda, Sharma, *et al*., 2017). The proposal is that genetic variants causing alterations in the ECM or modulating *TGFBI* might be different in individual patients and families, thus generating variations in the onset, progression and/or severity of their condition. This family study is the first genetic analysis study carried out on Maltese patients who phenotypically exhibit corneal dystrophy.

Literature advocates that mutations responsible for epithelial-stromal *TGFBI* dystrophy subtypes have specific gain-of-function effects (Evans, Davidson, Carnt, *et al*., 2016). Thus, the third and final objective of this research study

was to conduct functional work to explore the inhibition of *TGFBI* as a treatment modality, as well as analyse the possible downstream effects, with the aim of preventing deposition of corneal complexes in the early stages of these conditions. By performing a scoping literature review, compounds that modulate cascades which, according to literature, could lead to a decrease in the levels of TGFBIp in corneal cells, namely, epithelial cells, endothelial cells or fibroblasts, were identified. From the results of this review, Li and mitomycin C (MMC) were shortlisted to be used in the subsequent laboratory work. Functional work that included the establishment of a cell culture of HCECs, inducing knock down (KD) of *TGFBI* in the HCECs, treatment of HCECs with Li and MMC, RNA extraction and RNA sequencing were then performed. Comprehensive transcriptome analysis using bioinformatics tools was then carried out to investigate the differential expression between normal human corneal epithelial cells (HCECs), HCECs treated with Li, HCECs treated with mitomycin C and *TGFBI* KD HCECs. The results helped elucidate which genes are associated with the *TGFBI* gene, the effect inhibition of the *TGFBI* gene would have on downstream signalling and thus on corneal structure and function, as well as imply which genes, other than *TGFBI,* might play a leading role in *TGFBI* CDs. Additionally, the differential gene expression results revealed the effect, including any side effects, Li and mitomycin C have on HCECs with the aim of exploring novel possible repurposed treatment alternatives for *TGFBI* CDs.

CHAPTER 3: METHODS and RESULTS

3.1 PHASE 1: Maltese Allelic Variants that are Present in Corneal Dystrophy Genes, in a Worldwide Setting

3.1.1 Materials and Methods

3.1.1.1 Gene Panel Selection

A list of genetic mutation variants known to cause various CDs was compiled from the Online Mendelian Inheritance in Man database (McKusick & Hamosh), the Human Gene Mutation Database (Cooper, Ball, Stenson, *et al*., 2020), and other international literature (Bron A. J., 1990) (American Academy of Ophthalmology, 2010) (Aldave A. J., 2011) (Iliff, Riazuddin, & Gottsch, 2012). The panel of causal CD genes that were analysed in this study were *KRT3* (epithelial/subepithelial dystrophies; namely Meesmann CD); *SLC4A11* and *ZEB1* (endothelial dystrophies); *TGFBI* (epithelial-stromal *TGFBI* dystrophies); and *PIKFYVE*, *CHST6* and *KERA* (stromal dystrophies) (Klinthworth, 2009) (Weiss, Moller, Aldave, *et al*., 2015).

3.1.1.2 Sample Collection and Worldwide Cohort Selection

The data of the Maltese cohort used in this study was composed of a random sample of 410 individuals from whom DNA samples were collected by the Malta Biobank at the University of Malta. The latter is a founding member of EuroBioBank (www. eurobiobank.org). Donor confidentiality was maintained according to the Declaration of Helsinki and the project was approved by the University Research Ethics Committee of the University of Malta (MD 19/2011). The collection of blood samples was carried out in 2012 and 2013. The samples consisted of randomly sampled Maltese cord blood samples. The samples were equally distributed between

males and females, and were checked for their Maltese ancestry using personal identification details and their parents' surnames going back two generations. Any reference of foreign admixture was noted, and these samples were excluded.

In order to determine the minimum sample size needed from the Maltese population for this study, a power analysis was performed. Statistical power, also referred to as the sensitivity of a test, is the likelihood of a significance test detecting a non-zero difference between two groups when this actually exists. High power in a study indicates that the hypothesis being tested is false when that is the case, whereas a low power may indicate random errors. Having sufficient statistical power is required for high accuracy in order to avoid false positives being deduced about a population using sample data.

Power is mainly affected by sample size, effect size, and significance level. In this study, the least sample size of the cohort needed to be collected from the Maltese population to ensure high power was worked out using the dedicated pwr package of the statistical computing software R, version 4.1.3 (R Team, 2022). The difference of proportion power for binomial distribution via the arcsine transformation is given by the effect size, h= 2*asin(sqrt(p_1))-2*asin(sqrt(p_2)), where h is measured in radians and p_1, p_2 are the allele frequencies of the two populations being compared (Champely, 2020). In this study, p_2 is the allele frequency of the Maltese population whereas p_1 is the allele frequency of the worldwide population cohort being compared to the Maltese.

For instance, in one of the tests: TSI vs Maltese for the mutation G>C in *CHST*. The size n1 of the TSI cohort is 107, the allele frequency p_1 for this mutation is 0.051 and

the corresponding allele frequency for the Maltese is p_2= 0.005. The calculated effect size, h, is 0.314, to 3 significant figures. The command function pwr.2p2n.test(h=.314, n1=107, power=.8, sig.level = .05) was input in the pwr package, where n1 = the worldwide population (TSI) sample size used in the 1000GP, sig. level is 0.05 corresponding to the 95% level of significance and h = effect size for the two proportions p_1 and p_2. The value, n2=311, which is the minimum sample size of the Maltese cohort required is then returned (output) (Cohen, 1988).

Genotyping data representing worldwide cohorts used in this study was obtained from the 1000 Genomes Project (1000GP) database (1000GP Consortium, 2015). The 1000GP was set up with the aim of finding common genetic variants with frequencies of at least 1% in the populations studied. It is one of the largest comprehensive, fully open-access database of whole-genome sequencing (WGS) data without access or use restrictions that was consented for public distribution and open data sharing (Byrska-Bishop, Evani, Zhao, *et al*., 2022). The final, phase III, variant call set included 2,504 unrelated samples from 26 populations across 5 continental regions. Overall, the project identified more than 88 million variants that included 84.7 million SNPs, 2.6 million short insertions/deletions (indels), and 60,000 structural variants. The 1000GP data has been applied by researchers thousands of times, such as, to aid screening for pathogenic variants in disease cohorts in exome sequencing projects, to filter out common germ line variants that are not pathogenic, as well as for population genetic studies. Newer whole genome sequencing projects are being carried out, such as the UK10K project and the 100 000 Genomes project, with even more samples being proposed to be included. However, these datasets originate solely from the British population and would not have the global genetic diversity provided by the 1000GP.

Additionally the cohorts included in these UK projects are centered on certain disease cohorts and thus may not be appropriate for certain research projects (Zheng-Bradley & Flicek, 2017). On top of that, essentially all these disease-related databases will have various access restrictions. The 1000GP was thus chosen as the worldwide cohort to compare the Maltese cohort to in this PhD study due to its distinctive combination of the unbiased health status of the participants, the global diversity it contains as well as its open data access policy. It was also especially suitable for this study since it contains human genetic variant data from many different populations across the globe making it a good resource for population genetics when investigating population ancestry and migration history.

Therefore, the genotyping data from all 26 population cohorts included in phase III of the 1000 Genomes Project (1000GP) (1000GP Consortium, 2105) was used as a representation of worldwide cohorts to compare allele frequencies between the Maltese population cohort and these worldwide genetic ancestries. The cohorts in the 1000GP were categorised into five continental groups and these pooled data were also compared with the Maltese cohort.

The allele frequencies of known CD causal mutations that were present in the population cohorts in the 1000GP database and the Maltese cohort are documented in the results section (Table 8). These were used to calculate the genetic prevalence of CD subtypes globally and in the Maltese cohort.

The SNPs in the panel of genes selected, which were present in the Maltese cohort, were analysed. Moreover, their allelic frequency was compared with those of the worldwide population cohorts from the 1000GP.

The statistical analysis performed in order to calculate and compare the allele frequencies between cohorts is explained in detail in section 3.1.1.5 below.

3.1.1.3 Sequencing

The Maltese cohort genotyping data was provided by the Malta Biobank. The DNA samples were sequenced on the HiSeq2000. An average of 413,505,519 (96.37%) reads were generated and mapped, while 15,531,435 reads were not aligned to the reference genome. They had an average coverage of 41.28X. Variants extracted from the Maltese samples were used to create the draft Maltese exome. The allele frequencies of 371,197 single nucleotide variants were considered as a draft cohort control group. The single nucleotide variants were distributed from all parts of the referenced exome. The output from the NGS sequencer constituted a set of reads (containing nucleotides). Coverage was measured by using the number of reads that cover a particular nucleotide, in a given alignment. However, this deeper coverage did not fully make up for the decreased read length as some areas from the genome are not captured. The latter was due to shortcomings such as repeats and qualities at the read ends.

Sequenced samples resulted in a Fasta file format containing all the reads. They were aligned against the human reference genome build hg19 using Burrows Wheelers Aligner (version 0.5.9) and the NARWHAL pipeline. The NARWHAL sequence analysis pipeline makes automated processing and primary analysis of datasets obtained from Illumina sequencers possible. It functions as a de-multiplexing tool that allows easy alignment of individual samples to the reference genome of choice using predefined alignment profiles (Brouwer, Van den Hout, Grosveld, *et al*., 2012).

This software tool makes use of the Qseq files from an Illumina BaseCalls folder and converts them to the FastQ format using a C tool that can be highly parallelized. Subsequently, the FastQ files are de-multiplexed using a custom C tool. The de-multiplexing consists of two steps: 1. Generation of index files in which each read is assigned to a specific sample; 2. the FastQ files are separated for each sample. Thus, multiple FastQ files for multiplexed paired-end reads can be processed efficiently. Following this, the reads in the FastQ sample-specific files are aligned using an alignment script to the reference sequence producing a SAM file (Brouwer, Van den Hout, Grosveld, *et al*., 2012).

The variants were called after genome alignment. A variant was described as a change in nucleotides between the aligned sequences and the reference genome. Different methods were involved in calling NGS data, and a number of software packages, including GATK and SAM Tools were created. Both software packages contained various programmes used for viewing or editing aligned files in BAM format and extracting variants into a variant call format. Called variants were annotated using the Annovar open source software, which was designed to work with different databases including the Single Nucleotide Polymorphism Database (dbSNP). This annotation software generated the variants' location, effect on the protein, reference SNP cluster ID and other information depending on the chosen parameters (Mizzi, 2016). Polymerase chain reaction (PCR) on specific genes was done to validate certain SNP allele frequencies as part of quality control purposes, since, back in 2016/2017 the techniques for exome and genome sequencing were still under development.

Quality assessment is also incorporated in NARWHAL. Hence, the NARWHAL software pipeline makes it possible to analyse data in an automated efficient fashion

from different applications and variable numbers of input and output files, directly from the sequencer, in contrast to other frameworks such as the Galaxy where analysing data with flexible settings is not possible (Giardine, Riemer, Hardison, *et al*., 2005). Consequently this helps in reducing manual errors and hands-on time when processing large numbers of samples (Brouwer, Van den Hout, Grosveld, *et al*., 2012).

3.1.1.4 Variant Pathogenicity Assessment

The online predictive tools PROVEAN (Choi & Chan 2015), SIFT (Ng & Henikoff, 2001), MutationAssessor (Reva, Antipin, & Sander, 2011) and Polyphen2 (Adzhubei, Schmidt, Peshkin, *et al*., 2010) were used to determine the potential pathogenicity of each single nucleotide missense mutations in the Maltese cohort by employing the pathogenicity prediction function in the Ensembl database (Ensemble). The Species (Homo_sapiens), human assembly (GRCh37), the rsID for each variant and the online predictive tools mentioned above were selected and inputed in the 'Variant Effect Predictor' in Ensembl database. These predictive computational tools apply various algorithms in order to evaluate the impact missense variants would have on the altered peptide sequence produced. The strengths of each of the above software predictive bioinformatic tools are explained in brief below while their limitations can be found in the limitations section in the discussion chapter.

PROVEAN (Protein Variation Effect Analyser) (Choi & Chan, 2015) is a software tool that can be used to predict whether any type of protein sequence variations will affect the biological function of the protein or not (Choi, Sims, Murphy, *et al*., 2012) (Choi & Chan, 2015). It compares the protein sequence containing the variation to a set of related protein sequences and an algorithm computes a semi-global pairwise sequence alignment score. Utilizing this system, a database for precomputed prediction scores

for all possible single amino acid substitutions, single amino acid insertions, and up to 10 amino acids deletions in about 91K human proteins was set up (Choi, 2012). If the score produced by PROVEAN is equal to or below the threshold (which is -2.5), the variant is considered “deleterious”. If the score is above the threshold, the protein variant is considered “neutral”.

SIFT (Sorting Intolerant from Tolerant) is a conservation based method program that predicts whether an amino acid substitution affects protein function or not (Ng & Henikoff, 2001). It is based on the physical properties of amino acids and sequence homology, in that, active sites tend to be conserved on the protein family across species. SIFT can differentiate between functionally neutral and deleterious amino acid changes in human missense single nucleotide polymorphisms occurring in protein-coding regions. It labels them as ‘tolerated’ if the score is >0.05 and ‘damaging’ if the score is <=0.05 (range from 0-1) (Ng & Henikoff, 2001).

MutationAssessor is a server that analyses and predicts the outcome of amino acid substitutions in proteins. This algorithm takes into account the evolutionary conservation of the substitution in question in protein homologs. These alterations are classified into the following categories; neutral, low, high or medium together with a rank score between 0 and 1 where variants with a higher score are more likely to be deleterious. Its predictions reflect on protein functionality and stability (Hassan, Shaalan, Dessouky, *et al*., 2019).

Polyphen2 (Polymorphism Phenotyping) (Adzhubei, Schmidt, Peshkin, *et al*., 2010) uses an algorithm to determine the likelihood of an amino acid substitution being benign or deleterious by taking into consideration a number of features, namely the sequence, the phylogenetic and the structural information of that particular

substitution. The output gives a score of between 0 and 1, 0 being “benign” and 1 “damaging”. A score of below 0.2 is labelled as “benign”, 0.2-0.85 as “possibly damaging” and above 0.85 is “probably damaging” (Adzhubei, Schmidt, Peshkin, *et al*., 2010).

In order to give a more accurate prediction of the protein variant potential pathogenicity and its possible phenotypic effects, seeing that each tool focuses on different features of a protein variant, a combination of algorithmic programs were used in this study.

3.1.1.5 Statistical Analysis

The allele frequency was used in this study to describe how common a particular allele is within the population cohorts. This was calculated as the number of derived alleles over the number of sampled alleles. The allele frequencies and genetic prevalence of known CD causal mutations that were found to be present in the population cohorts in the 1000GP database and the Maltese cohort are listed in Table 8, whereas the allele frequencies for each of the 19 variants detected in CD-related genes in the Maltese Biobank data are shown in the result section (Table 9). The allele frequencies of the populations in the 1000GP, of the 19 SNPs analysed in this study are available in the Appendix.

A comparative analysis of the data from the 1000GP and the Maltese Biobank was performed. The fixation index (FST) value was used in this paper as a measure of cohort differentiation due to genetic structure, by taking into account the differences in the allele frequencies between the cohorts. SNPs were used to calculate the FST value between the Maltese cohort and the 26 worldwide cohorts from the 1000GP to study the genetic distance in CD-related genes between the cohorts. A large genetic distance in CD-related genes among cohorts indicates little breeding among them. The FST ranges from 0 to 1, where 0 means complete sharing of genetic material and 1 means no sharing (Holsinger & Weir, 2009). Statistical analysis was performed by using the R programming language for statistical computing (R Team, 2022).

In order to identify whether CD-related genes are highly differentiated with respect to allele frequency among the 27 cohorts, an FST value was calculated for each SNP to

produce an estimator of the variance between the cohorts. The following formula was applied:

$$F_{ST} = 1 - \frac{H_S}{H_T}, \qquad H_S = \frac{\sum_{i=1}^{27} 2n_i p_i q_i}{\sum_{i=1}^{27} n_i}, \qquad \bar{p} = \frac{\sum_{i=1}^{27} 2n_i p_i}{\sum_{i=1}^{27} n_i}, \qquad H_T = 2\bar{p}(1-\bar{p}).$$

where H_s is the mean expected heterozygosities in the 27 population cohorts and H_T is the expected heterozygosities for the total population cohorts (Weir & Cockerham, 1984). FST values between the Maltese and each of the five continental cohorts were calculated using the RStudio package for Hudson's estimator, following the recommendations of Bhatia *et al*. (Bhatia, Patterson, Sankararaman, *et al*., 2013). Taking H_w to be the mean number of differences within cohorts, and H_b the mean number of diferences between cohorts:

$$F_{ST}^{Hudson} = 1 - \frac{H_w}{H_b} = \frac{(p_1 - p_2)^2 - \frac{p_1(1-p_1)}{n_1 - 1} - \frac{p_2(1-p_2)}{n_2 - 1}}{p_1(1-p_2) + p_2(1-p_1)}$$

where n_i is the sample size and p_i is the sample allele frequency in cohort i for i = 1, 2.

Then the FST values of rare variants were calculated by comparing the Maltese cohort with each of the 26 individual population cohorts in the 1000GP. As recommended by Bhatia *et al*. (Bhatia, Patterson, Sankararaman, *et al*., 2013) and Chen *et al*. (Chen, Yuan, Shriner, *et al*., 2015), the Hudson estimator was applied for pairs of population cohorts.

To obtain an estimate of the differences and similarities in the genetic composition of population cohorts, the average FST value between each pair of continental cohorts for the pooled 19 SNP sites was also calculated.

3.1.2 Results

The global genetic prevalence of Lattice CD type IIIA (LCD) in the 1000GP is 0.08% (Table 8). Among the population cohorts included in the 1000GP, only the Han Chinese in Bejing (0.97%) and the Japanese in Tokyo (0.96%) exhibited this CD subtype. The genetic prevalence of epithelial basement membrane dystrophy (EBMD) in the 1000GP is 0.16% globally (Table 8). This mutation is present in the genotype of Colombians from Medellin (1%), British in England and Scotland (1%), Iberians in Spain (0.9%), and Tuscans in Italy (0.9%). On the other hand, the ZEB1 mutation that causes Fuch's endothelial CD (FECD) is present in 0.6% of individuals globally (Table 8). The prevalence is 1% in Colombians, 1.6% in Mexicans, 3% in Utah residents, 6% in Finnish, 1% in the British in England and Scotland, 1.9% in Iberians and 1.2% in Bengalis. The mutations NC_000005.9:g.135391459C>A and NC_000005.9:g.135398363G>C in the *TGFBI* gene and the NC_000010.10:g.31810782A>C mutation in the ZEB1 gene are not present in the Maltese cohort. One established causal CD mutation has been identifed in the Maltese cohort, namely, M502V (NC_000005.9:g.135391462A>G) that is located in the *TGFBI* gene (Table 8). This mutation has been reported to cause a Bowman's layer CD/atypical Thiel-Behnke CD phenotype (Chao, DeDionisio, Ke, *et al*., 2019) (Zenteno, Correa-Gomez, Santacruz-Valdez, *et al*., 2009) and has a genetic prevalence of 0.16% among the 1000GP cohorts globally. The only cohorts that had this mutation were the Puerto Ricans (0.96%), Utah residents (1%) and Iberians (1.9%).

Table 8: The allele frequencies and genetic prevalence of gene mutations reported in literature to cause CDs identified in the 1000GP database and in the Maltese cohort.

Gene	Chr	HGVS reference sequence	CD	Genetic prevalence in 1000GP	Allele freq. of the populations included in this study in which the mutation was present
TGFBI	5q31	NC_000005.9:g.135391459C>A	LCD IIIA (Yamamoto, *et al*., 1998)	0.08%	Han Chinese (Beijing) A=0.0049 Japanese (Tokyo) A=0.0048
		NC_000005.9:g.135391462A>G	Bowman/TBCD (Chao-Shern, *et al*., 2019) (Zenteno, *et al.,* 2009) (Niel-Butschi, *et al*., 2011)	0.16%	Puerto Rican (Puerto Rico) G=0.005 Utah residents (N & W European ancestry) G=0.005 Iberian (Spain) G=0.009 Maltese G=0.0012
		NC_000005.9:g.135398363G>C	EBMD (Boutboul, *et al*., 2006)	0.16%	Colombians (Medellin) C=0.0053 British (England and Scotland) C=0.0055 Iberian (Spain) C=0.005 Toscani (Italia) C=0.0047
ZEB1	10p11.2	NC_000010.10:g.31810782A>C	FECD (Riazuddin, *et al*., 2010)	0.6%	Colombians C=0.0053 Mexican (Los Angeles) C=0.0078 Utah residents (N & W European ancestry) C=0.015 Finnish (Finland) C=0.03 Iberian (Spain) C=0.009 British (England and Scotland) C=0.0055 Bengali (Bangladesh) C=0.006

Abbreviations: AA: amino acid; CD: Corneal Dystrophy; Chr: chromosome; EBMD: Epithelial Basement Membrane Dystrophy; Freq: Frequencies; LCD IIIA: Lattice Corneal Dystrophy type IIIA; Miss: missense; Syn: synonymous; TBCD: Thiel-Behnke CD; 1000GP: 1000 Genome Project.

The Human Genome Variation Society (HGVS) nomenclature standard is used in clinical diagnostics, and is authorized by the Human Genome Organisation (HUGO). The NC number (chromosome) stands for the reference sequence accession number; g: genomic reference sequence; the following number refers to the position of the affected nucleotide; the e.g. C>A refers to substitution of the C nucleotide with A.

Table 9: SNPs allele frequencies identified in CD-related genes in the Maltese cohort.

HGVS reference sequence	Gene	Chr	Alleles	Impact	f_MLT
NC_000012.11:g.53186088G>C	*KRT3*	12	G>C	miss	0.5046
NC_000012.11:g.53189696C>G	*KRT3*	12	C>G	miss	0.1040
NC_000012.11:g.53186122G>A	*KRT3*	12	G>A	syn	0.7518
NC_000005.9:g.135391462A>G	*TGFBI*	5	A>G	miss	0.0012
NC_000005.9:g.135382989G>C	*TGFBI*	5	G>C	syn	0.3584
NC_000016.9:g.75513243G>C	*CHST6*	16	G>C	miss	0.0379
NC_000002.11:g.209190330T>C	*PIKFYVE*	2	T>C	miss	0.9538
NC_000002.11:g.209190519A>T	*PIKFYVE*	2	A>T	miss	0.9396
NC_000002.11:g.209190528C>G	*PIKFYVE*	2	C>G	miss	0.9339
NC_000002.11:g.209191082C>A	*PIKFYVE*	2	C>A	miss	0.9299
NC_000002.11:g.209184999C>T	*PIKFYVE*	2	C>T	syn	0.9058
NC_000002.11:g.209215586A>G	*PIKFYVE*	2	A>G	syn	0.9051
NC_000020.10:g.3218634G>C	*SLC4A11*	20	G>C	miss	0.5892
NC_000020.10:g.3210301G>A	*SLC4A11*	20	G>A	syn	0.0983
NC_000020.10:g.3211235C>T	*SLC4A11*	20	C>T	syn	0.0859
NC_000020.10:g.3214581C>T	*SLC4A11*	20	C>T	syn	0.1209
NC_000020.10:g.3214819T>G	*SLC4A11*	20	T>G	syn	0.3082
NC_000012.11:g.91449984C>T	*KERA*	12	C>T	syn	0.0350
NC_000012.11:g.91449990C>T	*KERA*	12	C>T	syn	0.8540

Chr: chromosome; f_MLT: allele frequency in the Maltese cohort; Miss: missense; Syn: synonymous

The FST values for the 19 SNPs among the 27 cohorts varied from 0.005 to 0.2841 (Figure 5). The variant that showed the largest degree of differentiation was NC_000012.11:g.91449984C>T (*KERA*), whereas the variant with the least degree of differentiation was NC_000005.9:g.135391462A>G (*TGFBI*). An FST value >0.15 can be

considered as significant in differentiating population cohorts (Frankham, Ballou, & Briscoe, 2022).

The FST values between the Maltese and each of the five continental cohorts are documented in Table 10. The continental cohorts that show the greatest and smallest degree of differentiation from the Maltese cohort for each SNP are depicted in Figure 6 and listed in Table 11.

When considering individual population cohorts (see the Appendix), the American cohort (specifically the Puerto Rican, Mexican in Los Angeles, and Colombian cohorts) was the most similar to the Maltese cohort for the *KRT, PIKFYVE, SLC4A11* genes. For the *KERA* gene, it was the East Asian cohort (specifically the Kinh in Vietnam, Southern Han Chinese and Chinese Dai cohorts) that was the most similar. The Maltese cohort is similar to the South and East Asian cohorts when considering the *TGFBI* gene and to the Europeans and Americans when considering the *CHST* gene. On the other hand, FST results when analysing the *TGFBI*, *PIKFYVE*, *SLC4A11* and *KERA* showed that the Maltese cohort is least similar to the Yoruba (Nigeria), Mende (Sierra Leone), Luhya (Kenya) and Gambian cohorts from the African 1000GP cohort. The Esan (Nigeria) cohort was also one of the furthest away in the *TGFBI, PIKFYVE* and *SLC4A11*. FST results for the *KRT* gene show that it is furthest away from Han Chinese, Chinese Dai and Southern Han Chinese cohorts, which are all from the East Asian cohort. The *CHST* gene results are least similar to the Maltese cohort in the Bengali and Sri Lankan cohorts (in the UK), which are from the South Asian cohort, followed by the Japanese and Kinh (Vietnam) cohorts, which are from the East Asian cohort. Overall, the results suggest that the CD-related gene variants identified in the Maltese cohort show the

greatest degree of differentiation with Africans, while the smallest degree of differentiation is seen between the Maltese and American cohorts.

The average FST value between the different genetic ancestral groups for the pooled 19 SNP sites varied from 0.0000 to 0.01896, showing that there is a certain amount of genetic differentiation among the continents considered (Table 12).

The predicted pathogenicity of each missense variant in CD-related genes in the Maltese cohort that were evaluated in silico by using a variety of established bioinformatic predictors available online is reported in Table 13. They anticipate the *KRT3* NC_000012.11:g.53186088G>C mutation to be potentially deleterious. On the other hand, the results were inconclusive for *KRT3* NC_000012.11:g.53189696C>G and CHST6 NC_000016.9:g.75513243G>C.

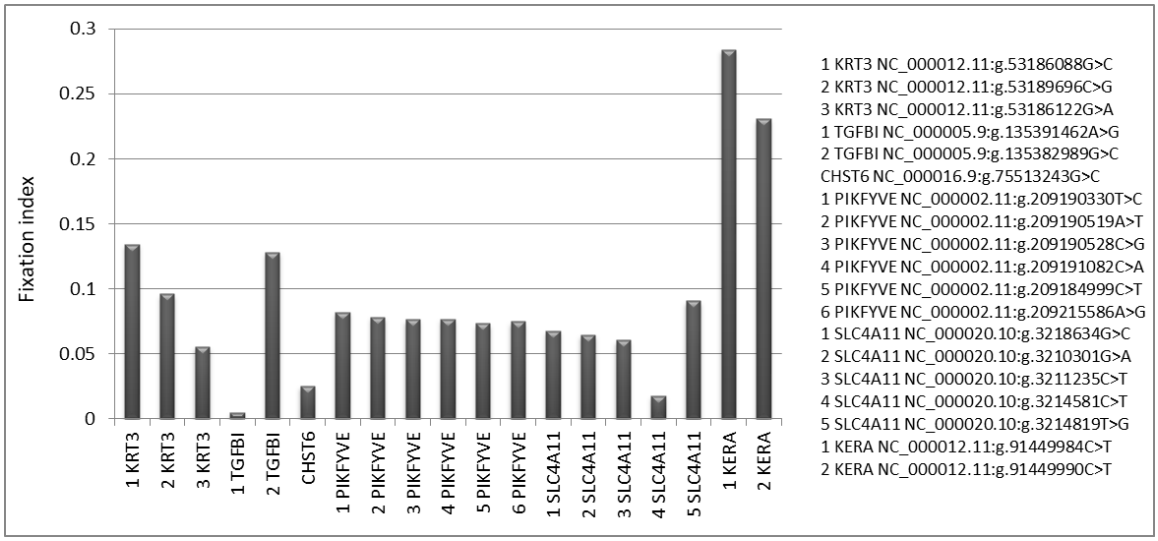


Figure 5: The fixation index values for each SNP among the 27 cohorts

Table 10: F_{ST} values obtained when calculating genetic differentiation between the Maltese cohort and each of the five continental cohorts from the 1000GP

HGVS reference sequence	Gene	F_{ST} of Maltese cohort vs Continental cohorts				
		African	American	East Asian	Europe	South Asian
NC_000012.11:g.53186088G>C	*KRT3*	0.0306	0.0026	0.2669	0.0538	0.108
NC_000012.11:g.53189696C>G	*KRT3*	0.1057	0	0.0964	0.0129	0.0116
NC_000012.11:g.53186122G>A	*KRT3*	0.0052	0.001	0.089	0.0108	0.0453
NC_000005.9:g.135391462A>G	*TGFBI*	0	0	0	0	0
NC_000005.9:g.135382989G>C	*TGFBI*	0.2981	0.0139	0	0.0247	0.0023
NC_000016.9:g.75513243G>C	*CHST6*	0.0276	0	0.0355	0.0006	0.0226
NC_000002.11:g.209190330T>C	*PIKFYVE*	0.1129	0.0051	0	0.0314	0.0094
NC_000002.11:g.209190519A>T	*PIKFYVE*	0.0908	0.0144	0	0.0453	0.0019
NC_000002.11:g.209190528C>G	*PIKFYVE*	0.0808	0.0186	0	0.051	0
NC_000002.11:g.209191082C>A	*PIKFYVE*	0.0774	0.0217	0	0.055	0
NC_000002.11:g.209184999C>T	*PIKFYVE*	0.0498	0.0419	0.0057	0.079	0
NC_000002.11:g.209215586A>G	*PIKFYVE*	0.0502	0.0425	0.006	0.0797	0
NC_000020.10:g.3218634G>C	*SLC4A11*	0.12	0	0.0141	0	0.0263
NC_000020.10:g.3210301G>A	*SLC4A11*	0.0574	0.0008	0.0883	0.0089	0.0543
NC_000020.10:g.3211235C>T	*SLC4A11*	0.0444	0	0.0808	0.0101	0.0504
NC_000020.10:g.3214581C>T	*SLC4A11*	0.0315	0	0	0.0106	0.002
NC_000020.10:g.3214819T>G	*SLC4A11*	0.2314	0.0264	0.0988	0.0144	0
NC_000012.11:g.91449984C>T	*KERA*	0.5211	0.065	0.0167	0.046	0.0048
NC_000012.11:g.91449990C>T	*KERA*	0.5451	0.0604	0.0089	0.0093	0.0616

Abbreviations: F_{ST} fixation index.

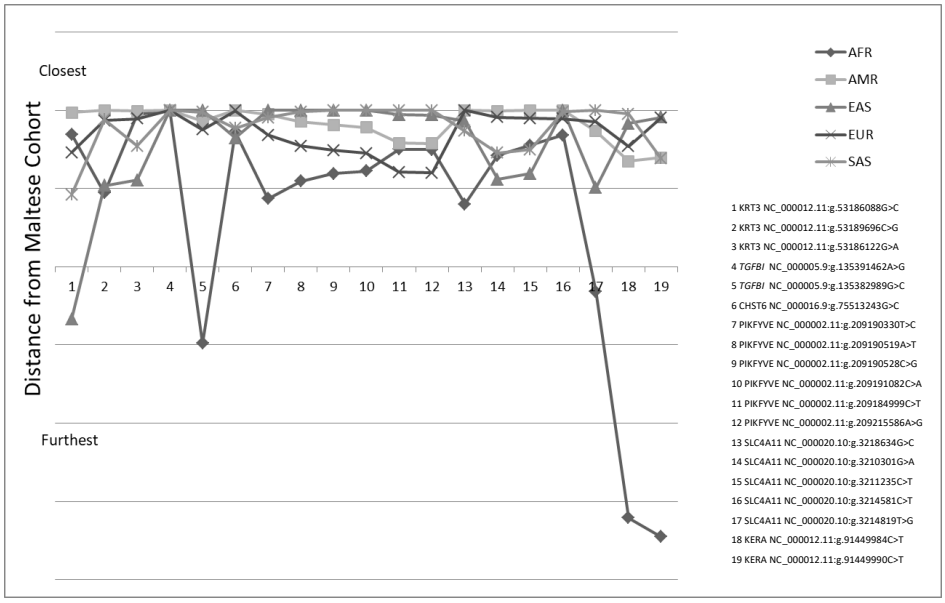


Figure 6: The continental cohorts that show the greatest and smallest degree of differentiation for each variant found in the Maltese cohort.
AFR African, *AMR* American, *EAS* East Asian, *EUR* Europe, *SAS* South Asian

Table 11: The continental cohorts that exhibit the greatest and smallest degree of differentiation for each SNP variant from the Maltese cohort.

HGVS reference sequence	**Gene**	**Greatest degree of differentiation from the Maltese cohort**	**Smallest degree of differentiation from the Maltese cohort**
NC_000012.11:g.53186088G>C	*KRT3*	East Asians	**Americans**
NC_000012.11:g.53189696C>G	*KRT3*	**Africans**	**Americans**
NC_000012.11:g.53186122G>A	*KRT3*	East Asians	**Americans**
NC_000005.9:g.135391462A>G	*TGFBI*	NA	**All cohorts**
NC_000005.9:g.135382989G>C	*TGFBI*	**Africans**	East Asians
NC_000016.9:g.75513243G>C	*CHST6*	East Asians	**Americans**
NC_000002.11:g.209190330T>C	*PIKFYVE*	**Africans**	East Asians
NC_000002.11:g.209190519A>T	*PIKFYVE*	**Africans**	East Asians
NC_000002.11:g.209190528C>G	*PIKFYVE*	**Africans**	East Asians, South Asians
NC_000002.11:g.209191082C>A	*PIKFYVE*	**Africans**	East Asians, South Asians
NC_000002.11:g.209184999C>T	*PIKFYVE*	Europe	South Asians
NC_000002.11:g.209215586A>G	*PIKFYVE*	Europe	South Asians
NC_000020.10:g.3218634G>C	*SLC4A11*	**Africans**	**Americans**, Europe
NC_000020.10:g.3210301G>A	*SLC4A11*	East Asians	**Americans**
NC_000020.10:g.3211235C>T	*SLC4A11*	East Asians	**Americans**
NC_000020.10:g.3214581C>T	*SLC4A11*	**Africans**	**Americans**, East Asians
NC_000020.10:g.3214819T>G	*SLC4A11*	**Africans**	East Asians
NC_000012.11:g.91449984C>T	*KERA*	**Africans**	South Asians
NC_000012.11:g.91449990C>T	*KERA*	**Africans**	East Asians

The continental cohort that shows the greatest degree of differentiation from the Maltese cohort for each SNP is the African cohort while the continental cohort that shows the smallest degree of differentiation from the Maltese cohort is the American cohort. These are shown in bold font.

Table 12: The average FST between the different continental cohorts describing the genetic differentiation calculated for the pooled 19 SNP sites.

	Cohort					
	Maltese	**African**	**American**	**E Asian**	**Europe**	**S Asian**
Maltese	NA	NA	NA	NA	NA	NA
African	0.00006	NA	NA	NA	NA	NA
American	0	0.00355	NA	NA	NA	NA
East Asian	0	0	0	NA	NA	NA
Europe	0.00432	0.01401	0.00008	0.00626	NA	NA
South Asian	0.00194	0	0.00662	0.00098	0.01896	NA

NA: not applicable.

Table 13: Predicted pathogenicity of missense variants.

Gene mutation	**Variant pathogenicity (Score)**				**Literature**
	PROVEAN	**SIFT**	**PolyPhen2**	**Mutation Assessor**	
KRT3 NC_000012.11:g.53186088G>C	Deleterious (-5.03)	Damaging (0)	Probably Damaging (0.944)	High (3.975)	No documented phenotype
KRT3 NC_000012.11:g.53189696C>G	Neutral (-1.90)	Tolerated (0.2)	Probably Damaging (0.89)	Low (1.625)	No documented phenotype
TGFBI NC_000005.9:g.135391462A>G	Neutral (-0.32)	Tolerated (0.2)	Benign (0)	Low (1.555)	Pathologic: Bowman's layer CD /atypical Thiel-Behnke (Chao-Shern *et al*., 2019) (Zenteno *et al*., 2009) (Niel-Butschi, *et al.*, 2011)
CHST6 NC_000016.9:g.75513243G>C	Neutral (-1.53)	Damaging (0.03)	Benign (0.003)	Neutral (0.345)	Heterozygotes are considered as a polymorphism (Abbruzzese *et al*., 2004)
PIKFYVE NC_000002.11:g.209190330T>C	Neutral (1.49)	Tolerated (0.98)	Benign (0)	Neutral (-0,92)	Fleck CD (Ensemble)
PIKFYVE NC_000002.11:g.209190519A>T	Neutral (-1.41)	Tolerated (0.27)	Benign (0.014)	Neutral (0.205)	Fleck CD (Ensemble)
PIKFYVE NC_000002.11:g.209190528C>G	Neutral (-0.10)	Tolerated (0.9)	Benign (0.01)	Neutral (0.14)	Fleck CD (Ensemble)
PIKFYVE NC_000002.11:g.209191082C>A	Neutral (0.17)	Tolerated (1)	Benign (0)	Neutral (-0.805)	Fleck CD (Ensemble)
SLC4A11 NC_000020.10:g.3218634G>C	Neutral (0.25)	Tolerated (0.67)	Benign (0)	n/a	No documented phenotype

3.2 PHASE 2: Comprehensive Clinical Exome Sequencing Analysis of Maltese GCD1 Patients Uncovers the Causative Variant in the TGFBI Gene and Highlights Potential Alterations in the Corneal Extracellular Matrix and Integrin Interactions. Focus on Future Targeted Treatment.

3.2.1 Materials and Methods

Maltese individuals over 18 years of age that phenotypically exhibited GCD1 were recruited prospectively during corneal clinics performed over a period of 2 years from beginning of 2018 to end of 2019, at the Ophthalmic Outpatients Department at Mater Dei Hospital. Genetic analysis of the recruited seven consenting individuals, who formed part of the same family, was performed in order to determine which GCD1 genetic variant is present in Malta. A complete ophthalmological examination consisting of visual acuity testing and slit-lamp biomicroscopy, as well as a detailed family history of all the participants was performed. During the interviews, a family tree demonstrating their relation was drawn. Templates of pro forma used during the interviews are included in the Appendix.

3.2.1.1 Ethical Considerations

Ethical approval to perform this study was obtained from the Faculty Research Ethics Committee (FREC), as well as from the University of Malta Research Ethics Committee (UREC). The ethics reference numbers are 11/2017 and FRECMDS_1718_064. The study had approval and permission from the CEO of Mater Dei Hospital of Malta, the Head of Department of Ophthalmology as well as from the Ophthalmology consultants under whose care the participating subjects were.

3.2.1.2 Patient Recruitment and Sample Collection

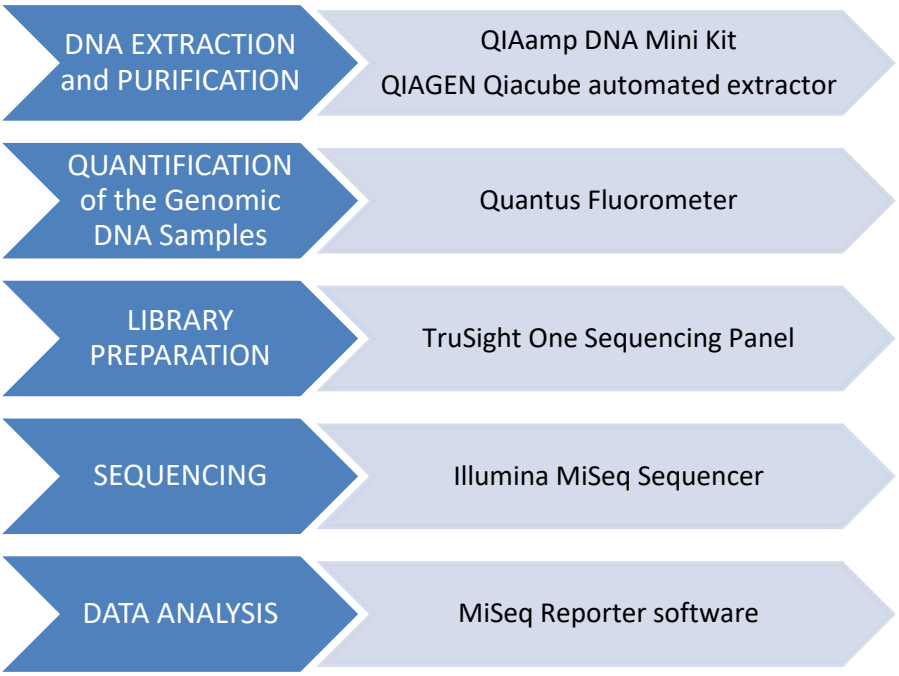
Patients who were clinically diagnosed with Granular Corneal Dystrophy type I were introduced to this study and given a "Patient Information Sheet" to explain further the procedures involved. The patients who gave their consent after reading and understanding the patient information sheet were given consent forms to give their approval to take part in this study. Templates of the patient information sheet and consent forms are included in the Appendix.

Genomic DNA from all subjects was extracted from mouthwash samples. The subject was asked to provide a buccal mouthwash sample by rinsing their mouth for about 30 seconds with 50mls of a standard mouthwash solution and collecting it in a provided sterile container. Two samples, one early in the morning and the other in the evening, were collected from each individual in order to maximise the collection of DNA material. The subjects providing the samples were asked not to consume food or drink in the 30 minutes prior to the collection of the mouthwash sample.

Buccal mouthwash samples were chosen as the method of obtaining DNA samples in this study since it is non-invasive, safe and cost-effective. It has been shown that DNA degradation that might occur did not significantly prevent successful amplification of PCR fragments (Zayats, Young, Mackey, *et al*., 2009).

3.2.1.3 Laboratory Work

Clinical exome sequencing workflow:



3.2.1.3.1 Extraction of Genomic DNA

Genomic DNA was extracted from the mouthwash samples using the QIAamp DNA Mini Kit for the QIAGEN Qiacube automated extractor (QIAGEN, Hilden, Germany). This is a form of automated DNA extraction that makes use of a spin column principle. It helps to reduce significantly the time needed for this process when compared to manual DNA extraction (Qiagen, 2016).

The protocol for extraction of genomic DNA from mouthwash was performed according to the kit manufacturer instructions:

Since Buffer AW1 and AW2 were supplied as a concentrate, they were prepared by adding ethanol (96-100%) according to the instructions on the bottles and stored at room temperature (15–25°C). Four millilitres of phosphate buffered saline (PBS) were added to each mouthwash sample and centrifuged at 1800 x g for 5 min. The supernatant was then decanted and the pellets were resuspended in 180 µl PBS. The

Buffer labelled AL was mixed thoroughly and Proteinase K and 200 µl Buffer AL were added to each sample. These were further mixed by vortexing for 15 s. (The activity of the proteinase K solution is 600 mAU/ml solution (or 40 mAU/mg protein)). The tubes were then incubated at 56°C for 10 min after which 200 µl of ethanol (96–100%) were added to each sample, and mixed again by vortexing. The lysate buffering conditions stabilised the nucleic acids and allowed for optimal binding of the DNA to the QIAamp membrane before the sample is loaded onto the QIAamp Mini spin column. QIAamp Spin Columns were placed in a 2 ml collection tubes and the mixtures were added to the QIAamp Spin Columns without moistening the rim and this was centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp Spin Columns were placed in clean 2 ml collection tubes and the tubes containing the filtrate were discarded. DNA is adsorbed onto the QIAamp silica membrane during this centrifugation. Salt and pH conditions in the lysate ensure that protein and other contaminants, which can inhibit PCR and other downstream enzymatic reactions, are not retained on the QIAamp membrane. The QIAamp Spin Columns were opened and 500 µl of Buffer AW1 was added to each. These were centrifuged at 6000 x g (8000 rpm) for 1 min and the QIAamp Spin Columns were then placed in clean 2 ml collection tubes and the tubes containing the filtrate were discarded. The QIAamp Spin Columns were opened and 500 µl of Buffer AW2 was added to each and subsequently centrifuged at full speed for 3 min. The use of these 2 different wash buffers, Buffer AW1 and Buffer AW2, eliminated any impurities which would otherwise have inhibited the analysis to be done on the DNA samples. The QIAamp Spin Columns were placed in clean 1.5ml microcentrifuge tubes and the tubes containing the filtrate were discarded. The QIAamp Spin Columns were opened and the DNA was eluted with 100 µl of Buffer AE. Finally these were incubated

at room temperature for 1 min then centrifuged at 6000 x g (8000 rpm) for 1 min. The spin column was then discarded.

3.2.1.3.2 Automated DNA Purification on the QIAcube:

The preset protocol for purification of genomic DNA was selected and the samples and reagents were loaded onto the QIAcube worktable. The QIAcube was then closed and the protocol was started. Automated sample lysis and purification using QIAGEN spin columns was performed.

3.2.1.3.3 Quantification of the Genomic DNA Samples

The concentrations of the genomic DNA (gDNA) samples were established using a Quantus Fluorometer (Promega). The protocol for calibration of the fluorometer and the quantification of the genomic samples collected from the GCD1 patients was performed as described in the paragraph below:

The 1X TE Buffer was prepared by diluting the 20X TE Buffer 20-fold with nuclease-free water and mixed. The Working Solution was prepared by diluting the QuantiFluor® dsDNA Dye 1:400 in 1X TE buffer. (10µl of QuantiFluor® dsDNA Dye was added to 3,990µl of 1X TE buffer and mixed.)

The Blank Sample was prepared by adding 200µl of QuantiFluor® dsDNA Dye working solution to an empty 0.5ml PCR tube whilst at the same time making sure to keep the tube away from direct light. The 200ng Standard Sample was prepared by adding 2µl of the provided DNA Standard (100ng/µl) to 200µl of QuantiFluor® dsDNA Dye working solution in an empty 0.5ml PCR tube. It was vortexed well and protected from light. The GCD1 patient samples were prepared by adding 2µl of each of the GCD1 patient samples to 200µl of QuantiFluor® dsDNA Dye working solution in separate 0.5ml PCR tubes. These samples were also kept away from light.

The prepared samples were incubated at room temperature for 5 minutes while being protected from light. The dsDNA protocol was selected on the Quantus™ Fluorometer. The Quantus™ Fluorometer was calibrated and a blank sample was placed in the reader. Following this, the standard was also placed in the reader and proceeded with the next steps. The volume of the GCD1 patient sample (2µl) was inserted and the 'Units' were set to ng/µL. The first sample was loaded and its concentration was automatically displayed when the lid was closed. Five out of the seven participant samples collected, were of sufficient good quality to be processed and analysed by next generation sequencing to yield accurate results. The range of concentrations of the samples was 1.6 ng/µL to 2512 ng/µL. The concentrations of the samples that were sequenced ranged from 225 ng/µL to 2512 ng/µL.

The advantage of using fluorometry over UV spectrophotometry is that the DNA estimation by the former is more accurate and precise. Spectrophotometry leads to an overestimation of DNA content since they do not distinguish between dsDNA, RNA, proteins or free nucleotides, while in fluorometry the use of different dyes are used for dsDNA, ssDNA or RNA thus increasing its specificity (Paul, Raiput, Joshi, *et al*., 2021).

3.2.1.3.4 Preparation of the Libraries and Exome Sequencing

Clinical exome sequencing looks at the exome, i.e. the protein-coding regions, of the genome that are already known to be associated clinically with disease. This method helps to facilitate establishing causality when a variant is present in these regions than if a variant is found elsewhere.

Clinical exome sequencing was used in this setting since it is advantageous in providing evidence for potential phenotypic variations associated with recognized human diseases as well as to identify new candidate genes associated with disease. By

providing an in-depth base-by-base view of the exome, single nucleotide variants and insertions or deletions can be detected in exons which are ultimately the protein-coding regions most likely to be associated with disease (Saarela & Kettunen, 2017). It is also relatively cheaper and time efficient when compared to whole exome sequencing. Thus, in this study, clinical exome sequencing was applied rather than whole genome sequencing for these reasons. Having said that, whole genome sequencing can be subject for future study.

The exome sequencing panel used to analyse the specimens was the TruSight One Sequencing Panel (ilumina). This clinical exome panel covers 12Mb of genomic content including >4,800 genes associated with specific clinical phenotypes (illumina, 2018).

The protocol on the TruSight One Sequencing Panel Reference Guide was used to prepare the libraries (Figure 7) (illumina).

Initially, the gDNA samples were normalized and diluted in Tris-HCl to 10ng/µl, and then to 5ng/µl in a 2-step dilution. Tagmentation of these samples was performed by means of Nextera transposomes that cut the DNA into fragments. The ends of the DNA fragments were tagged with primer-binding adapter regions. Sample purification beads (SPB) were used to purify the tagmented DNA from the transposomes in order to avoid downstream inhibition. A PCR program was then used to amplify the tagmented DNA whilst adding unique indices. Targeted regions of interest of the DNA were bound with capture biotin probes, and subsequently separated from the rest of the material by means Streptavidin Magnetic Beads (SMB). The samples were amplified and cleaned again. The libraries were checked by fluorometry in order to quantify them and subsequently diluted accordingly to avoid over or under-clustering upon sequencing (illumina, 2018).

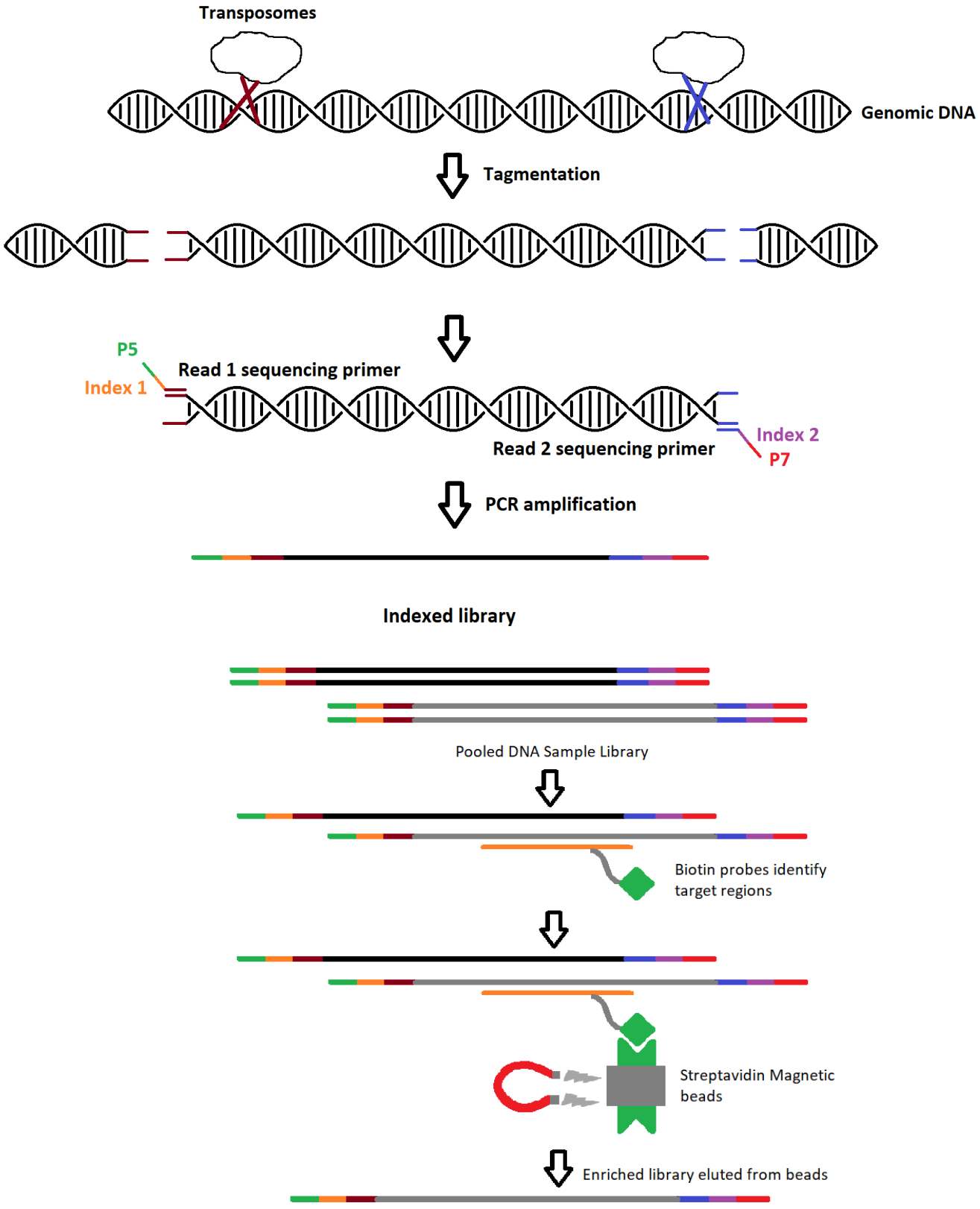


Figure 7: Preparation of libraries using the TruSight One Sequencing Panel

Next-generation sequencing (NGS) is a high-throughput methodology that enables rapid sequencing of the base pairs in DNA or RNA samples. The illumina MiSeq sequencer (illumina, 2018) was used for sequencing in this study. It leverages sequencing by synthesis (SBS) technology i.e. tracking the addition of fluorescent labelled nucleotides as the DNA chain is copied (illumina, 2018). Millions or billions of DNA strands can be sequenced in parallel, yielding substantially more throughput (Nature, NGS, 2018). Illumina MiSeq sequencer enables the testing of customised gene panels and has been proven to be a dependable tool to identify known and new mutations in patients with ophthalmic genetic disorders (Van Cauwenbergh, Van Schil, Bauwens, *et al*., 2022). The concept of SBS technology, applied by this sequencer, is a process where individual base-pairs are added to single stranded DNA, producing a complementary DNA strand and thus determining the DNA sequence. Upon the addition of an individual base-pair to a DNA strand template, a fluorescence is emitted. The base-pair which was added to the DNA strand template will be indicated by the colour of the fluorescence. The combination of these colours will create a full genomic DNA sequence (Dadlani, 2023). This method has been proven to be highly precise as well as allowing the amount of genomic data potentially generated to be increased significantly.

3.2.1.3.5 Analysis of Genetic Data

The MiSeq Reporter software, which is pre-installed on MiSeq sequencers, was utilized to convert the base cells produced by primary analysis to raw Variant Call Format files. This secondary analysis procedure included demultiplexing the data (separation of data from pooled samples based on short index sequences from different libraries), alignment of the sequences against the human hg19 reference

genome as well as variant calling, where SNVs and other structural variants were recorded (illumina, MiSeq® Reporter 2018). The sequencer also provided information regarding the quality of the sequencing by producing technical data reports.

Nucleotide sequences were compared with the published sequence of the human genome to detect the sequence changes.

In order to refine the variants detected, filtering techniques were implemented to the annotated excel file that was obtained by using wANNOVAR (Yang & Wang, 2015). In order for a variant to be accepted, a coverage filter which was set at 10X or higher was applied for each specific variant position. The genotype quality filter, which reflects the Phred-scaled confidence that there was a correct genotype assignment, was also applied.

Exome analysis led to the identification of the DNA mutation variant present in the *TGFBI* gene of this cohort of Maltese subjects that is linked to GCD1. The genetic variants identified in the genome of these Maltese GCD patients were filtered using the PyCharm software so as to eventually evaluate any potential role they might play in the pathogenesis of the Maltese GCD1 individuals.

3.2.1.4 Bioinformatic Analysis

In order to identify the genetic variants present in the Maltese cohort exhibiting GCD1 that might be relevant to the pathogenesis and phenotype of these individuals, the variants obtained were filtered using the PyCharm software. A program was written using Python language to exclude intron variants and synonymous variants present in the GCD patient samples. This filtering was performed since introns are noncoding, thus, intronic mutations do not have an effect on protein coding sequences since they are spliced out during protein synthesis. Having said that, pathogenic deep

intronic variants have occasionally been reported to cause activation of atypical splice sites, or changes in splicing regulatory elements (Vaz-Drago, Custódio, & Carmo-Fonseca, 2017) (Zhang, Yan, Zhou, *et al*., 2023). On the other hand, a synonymous mutation is a nucleotide change in the DNA sequence of a gene that does not affect the amino acid sequence of the encoded protein. As a result, these variants are unlikely to be pathogenic. Infrequently, synonymous mutations have been reported to affect RNA splicing, translational efficiency, and mRNA stability (Dhindsa, Wang, Vitsios, *et al*., 2022).

Following this filtering, the variants that are present in all the GCD1 patient samples were selected. Only the variants common to all the samples were kept for further analysis. This was done so as to make sure that the variants found in the genes that are being shortlisted are key genes that might be playing a role in the pathogenesis of the Maltese GCD1 cohort of patients. Using PyCharm software, a dataframe containing all variant specific columns and sample specific columns for each GCD1 sample was created. A deepcopy of the filtered data was made and this data was used to create a dataframe containing variant information that included the chromosome number, coordinate and the specific variant. A temporary variant list was created to check if these were present in all the samples by using the 'for loop' function. By making use of the 'if-else' statement, when a variant was found to be present in all the GCD1 samples, the said variant data was appended to the final dataframe, while, variants that were not present in all the samples were rejected. The final dataframe contained variants that were present in all the GCD1 patient samples. The code written in order to perform this filtering together with the list of TGFBI associated genes derived from databases and literature, as well as the filtered list of variants present in all the GCD1

patient samples can be accessed via the repository: https://github.com/gabysci/corneal_dystrophy.

A list of genes coding for the top proteins found to be present in GCD1 deposits (Table 3) (Courtney, Poulsen, Kennedy, *et al*., 2015), genes that have been documented in literature and databases; namely Genemania, Pathway Commons, Ndexbio and KEGG pathways**;** to be co-expressed or modified by the *TGFBI* gene or known to bind to the *TGFBI* protein (Table 14 and Table 15) was compiled, and these were cross-checked with the mutation variants present in all 5 of the Maltese GCD1 patient samples. This was performed by using the PyCharm software by calling the 'isin' function. This was carried out in order to identify genes containing variants that could be of significance in the mentioned patients.

Table 14: Genes that have been documented in literature to be related to *TGFBI* gene or protein.

Gene symbol	**Name**	**Reference**
AKT1	Ak strain transforming	(Wen, *et al*., 2011)
CUL4A	Cullin 4A	(Hung, *et al*., 2015)
IL1B	Interleukin 1B	(Wilson S. , 2021)
IL6	Interleukin 6	(Zhang, *et al*., 2005)
JAK2	Janus kinase 2	(Zhang, *et al*., 2005)
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	(Yu, *et al*., 2015)
SMAD2	Suppressor of Mothers against Decapentaplegic 2	(Derynck & Zhang, 2003)
SMAD3	Suppressor of Mothers against Decapentaplegic 3	(Derynck & Zhang, 2003)
SMAD4	Suppressor of Mothers against Decapentaplegic 4	(Derynck & Zhang, 2003)
SMAD7	Suppressor of Mothers against Decapentaplegic 7	(Derynck & Zhang, 2003)
STAT3	Signal transducer and activator of transcription 3	(Zhang, *et al*., 2005)

Table 15: List of genes that have been documented in literature to be co-expressed/bind/pathway-related with the *TGFBI* gene.

Gene	**Genemania**				**Pathway commons**				**Ndexbio**		**KEGG pathways**
	Co-expressed	Genetic interaction	pathway		Co-expressed	Modification	Binding		Binding		
APCS	y	y	y								
A2M							y				
COL1A1	y		y						y		y
COL1A2	y	y	y				y		y		
COL2A1	y	y	y				y				
COL4A1							y		y		
COL4A2							y		y		
COL4A3							y		y		
COL4A4							y		y		
COL6A3											y
COL6A5	y	y									
CPA2	y	y									
CPA4	y	y									

EMILIN3	y	y									
ESR1					y						
ESR2					y						
FLOT1						y					
FLOT2						y					
FN1	y	y	y				y		y		y
HSPG2	y		y								
ITGA3	y	y	y								
ITGA5											y
ITGA11											y
ITGAM	y	y	y								
ITGAV	y		y			y			y		
LOXL2	y										
ITGB5						y					
MAPK1						y					
MAPK3						y					

MMP2											y
NOG					y						
PLOD2	y										
POSTN	y	y									
RAP2B	y										
RELA						y					
SELE					y						
STAB1	y	y									
STAB2	y	y									
S100P							y				
THBS1											y
TNC											y
TNF						y					
VCAM1					y						
ZC2HC1A		y									

3.2.1.5 Enrichment Analysis

To identify the gene ontology (GO) molecular and biological functions, as well as KEGG and Reactome pathways enriched for the gene mutation variants present in the Maltese GCD1 patients, the WEB-based GEne SeT AnaLysis Toolkit (WebGesalt) was used to perform an Over-Representation Analysis (ORA) in function and pathways. WebGesalt is a web-based integrated data mining system for exploring gene sets in various biological contexts. It offers a unique online resource for information retrieval, organization, visualization and statistical analysis of gene sets. This analysis helps pinpoint the significance these mutations can potentially have in the final phenotype of these patients, together with the potential response to future treatment options.

The parameters used to run ORA included the following: Organism of Interest: Homo sapiens; Method of interest: Over-Representation Analysis; Functional Database: Geneontology- Biological process, Molecular function or Pathway- KEGG, Reactome for every set of DE genes; Gene ID type inserted was the Gene symbol; Gene lists uploaded were the up or downregulated DE expressed genes for comparison group; and Reference set chosen: genome. The advanced parameters included Minimum number of genes for a category: 5; Maximum number of genes for a category: 2000; Multiple test adjustment: BH; Significance level: top 25; Number of catagories expected from set cover: 10; and Number of categories visualized in the report: 40. The significance threshold applied was <=0.05 for each group of enriched genes.

3.2.2 Results

3.2.2.1 Clinical Findings

A total of seven patients that form part of the same family and that exhibited the GCD1 phenotype by clinical evaluation using slit lamp biomicroscopy, took part in the study. The disease was bilateral in all patients. These individuals all showed a very similar phenotype with patterns of central whitish-grey, crumb-like corneal opacities at the level of the anterior stroma separated from each other by clear spaces. No stromal lattice lines were seen in any of these patients. Elderly patients had opacities that were more numerous and larger in size than those seen in the younger probands. With age, the opacities were seen to coalesce and extend into the deeper stroma, causing significant deterioration in visual acuity. However, the peripheral cornea, in general, remained free of deposits. Unfortunately, the age of onset of each individual could not be determined since most of the candidates were examined ophthalmologically only when they became symptomatic. Similarly, no remark can be made regarding the rate of progression and severity of GCD1 in these individuals since this would need long term follow up examinations.

The recruited patients formed part of the same family, making it possible to design a genogram that tallied with previous international studies confirming autosomal dominant inheritance (Figure 8) (Klinthworth, 2009) (Nowińska, Wylegala, Janiszewska, *et al*., 2011) (Le, Nguyen, Hoang, *et al*., 2004) (Kannabiran, Sridhar, Chakravarthi, *et al*., 2005).

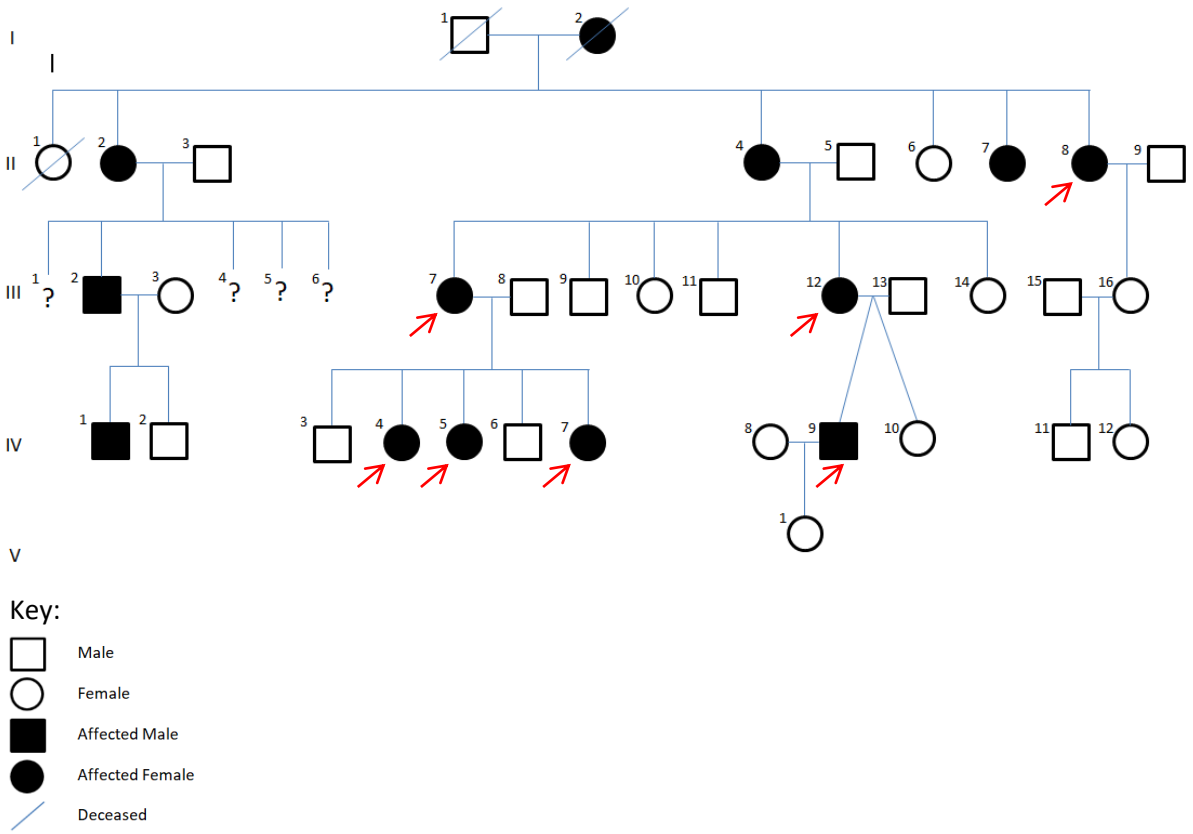


Figure 8: Genogram depicting the family relationships of the Maltese individuals exhibiting GCD1 in this study. The red arrows mark the individuals that participated in the study.

3.2.2.2 Case Reports

Case II-8 is an 80 year old lady who was diagnosed with GCD1 at the age of 24. The first symptoms she experienced were photophobia and foreign body sensation. She underwent penetrating keratoplasty (PK) in both eyes, twice in her right eye and once in her left eye. Recurrence of the condition in her right eye was noted about 6 years after the second PK, and 4 years after the PK in her left eye.

Case III-7 is a 58 year old lady who besides suffering from GCD1 is also a high myope. She underwent left eye PK in 2009. The first traces of recurrence of granular deposits were recorded in 2012. Unfortunately, she also developed myopic degeneration and an associated Fuch's haemorrhage in her right eye which caused her VA to drop significantly. She has five children, two boys and three girls. All the girls exhibit GCD1 while the males do not. The corneal opacities present currently in these three females, cases IV-4, 5 and 7, are at the level of the anterior stroma separated from each other by clear spaces. This confirms what has been previously documented in literature i.e. the age related progressive increase in deposits.

Case III-12 is 54 year old lady who was diagnosed with GCD1 in her childhood years. She was initially diagnosed as having high myopia as a child and was later diagnosed with GCD1 when she presented with a foreign body sensation and photophobia. She underwent left eye PK two years ago and to date has no signs of recurrence. Her son, case IV-9 was diagnosed with granular corneal dystrophy when he was 10years old while the twin sister does not phenotypically exhibit the condition.

3.2.2.3 Molecular Analysis

Five out of the seven samples collected could be analysed by exome sequencing to yield reliable results. All the sequenced samples revealed a heterozygous nucleotide

missense mutation in chromosome 5q31, namely a nucleotide change of cytosine being replaced with thymine within exon 12. This causes the amino acid arginine to be replaced with tryptophan at codon 555 (R555W) (Figure 9). The described mutation is characteristic of granular corneal dystrophy type I.

Due to the phenotype-genotype variability seen in GCD1 individuals that has been documented multiple times in literature, researchers have suggested that all individuals exhibiting *TGFBI* CD phenotypes should have genetic screening in order to identify the exact variant in question. By identifying the mutation variant present in this Maltese cohort one would be able to understand better the local Maltese GCD1 disease course and phenotype.

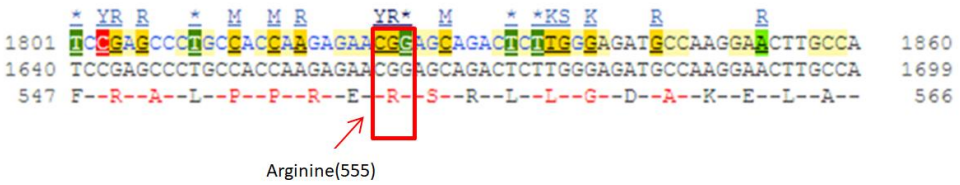


Figure 9: Location of Arginine 555 in exon 12 of the *TGFBI* gene.

3.2.2.4 Bioinformatic Analysis and Enrichment Analysis

In this part of the research study, the objective was to identify mutations, other than the known causal *TGFBI* gene mutations, that could potentially play a role in the pathogenesis of the Maltese GCD1 cohort.

The list comparison between genes coding for proteins present in GCD1 deposits (Courtney, Poulsen, Kennedy, *et al*., 2015) and *TGFBI*-related genes, and the mutation variants present in all 5 of the Maltese GCD1 patients returned the following key genes: alpha-2-macroglobulin (*A2M*), collagen type XII alpha chain (*COL12A1*), collagen type I alpha 1 chain (*COL1A1*), collagen type I alpha 2 chain (*COL1A2*), collagen type III

alpha 1 chain (*COL3A1*), collagen type IV alpha 1 chain (*COL4A1*), collagen type IV alpha 2 chain (*COL4A2*), collagen type IV alpha 3 chain (*COL4A3*), collagen type V alpha 1 chain (*COL5A1*), collagen type VI alpha 5 chain (*COL6A5*), desmoplakin (*DSP*), fibronectin 1 (*FN1*), heparan sulphate proteoglycan 2 (*HSPG2*), interleukin 6 receptor (*IL6R*), integrin subunit alpha 11 (*ITGA11*), matrix metallopeptidase 9/ type IV collagenase (*MMP9*), transforming growth factor beta induced (*TGFBI*), tenascin C (*TNC*) (Table 16)

Table 16: Key genes related to *TGFBI* that were found to contain mutation variants in the Maltese GCD1 patient cohort.

Gene	**Variant**	**Chr**	**dbSNP ID**	**Transcript**	**Consequence**	**Protein Position**	**AA**	**Codons**	**ClinVar Condition**	**Clinincal significance**
A2M	T>C/C	12	rs226405	NM_000014.4	missense_variant	639	N/D	Aat/Gat	n/a	Benign
COL12A1	C>T/T	6	rs970547	NM_004370.5	missense_variant	3058	G/S	Ggc/Agc	n/a	Probably damaging
COL1A1	T>C/C	17	rs1800215	NM_000088.3	missense_variant	1075	T/A	Acc/Gcc	Osteogenesis imperfecta, Ehlers-Danlos syndrome	Benign
COL1A2	C>G/G	7	rs42524	NM_000089.3	missense_variant	549	P/A	Cct/Gct	Osteogenesis imperfecta, Ehlers-Danlos syndrome, intracranial aneurysms, risk allele for nARMD and PCV	Benign
COL3A1	T>G/G	2	rs1516446	NM_000090.3	missense_variant	1353	H/Q	caT/caG	n/a	Unknown
COL4A1	T>G/G	13	rs3742207	NM_001845.4	missense_variant	1334	Q/H	caA/caC	n/a	Benign
COL4A1	T>G/G	13	rs536174	NM_001845.4	missense_variant	555	T/P	Aca/Cca	n/a	Unknown
COL4A2	T>A/A	13	rs439831	NR_046583.1	downstream_gene					

					_variant					
COL4A2	T>C/C	13	rs409858	NR_046583.1	downstream_gene_variant					
COL4A3	T>C/C	2	rs10178458	NM_000091.4	missense_variant	141	L/P	cTg/cCg		Benign
COL4A3	A>G/G	2	rs6436669	NM_000091.4	missense_variant	162	E/G	gAa/gGa		Benign
COL5A1	G>G/A	9	rs3827848	NM_000093.3	splice_region_variant, synonymous_variant	1374	T	acG/acA	Stomach carcinoma	
COL6A5	A>G/G	3	rs9883988	NM_153264.5	missense_variant	2188	Q/R	cAa/cGa		Benign
COL6A5	G>A/A	3	rs819085	NM_153264.5	missense_variant	2205	G/D	gGt/gAt		
DSP	G>A/A	6	rs1016835	NM_004415.2	splice_region_variant, synonymous_variant	877	R	agG/agA		
FN1	C>T/T	2	rs1250209	NM_212482.1	missense_variant	2261	V/I	Gtc/Atc		Benign
FN1	T>G/G	2	rs2577301	NM_212482.1	missense_variant	817	T/P	Acg/Ccg		Benign
HSPG2	T>C/C	1	rs989994	NM_005529.5	missense_variant	765	N/S	aAt/aGt		Benign

HSPG2	T>C/C	1	rs1874792	NM_005529.5	missense_variant	638	M/V	Atg/Gtg		Benign
IL6R	A>C/C	1	rs2228145	NM_000565.3	missense_variant	358	D/A	gAt/gCt		Benign
ITGA11	C>T/T	15	rs898591	NM_001004439.1	splice_region_variant, intron_variant					
ITGA11	G>G/A	15	rs4777035	NM_001004439.1	missense_variant	972	P/L	cCc/cTc		Benign
ITGA11	G>G/A	15		NM_001004439.1	missense_variant	104	R/C	Cgc/Tgc		probably_damaging
MMP9	G>C/C	20	rs2250889	NM_004994.2	missense_variant	574	R/P	cGg/cCg		Benign
MMP9	C>C/T	20	rs20544	NM_004994.2	3_prime_UTR_variant	0				
TNC	T>C/C	9	rs1757095	NM_002160.3	missense_variant	539	Q/R	cAg/cGg		Benign

ORA analysis revealed that, with regards to the GO annotation of biological and molecular function of the proteins coded by key mutated genes in the GCD1 patients, the most significantly affected processes would include extracellular structure and matrix organization, the collagen-activated tyrosine kinase receptor signalling pathway, integrin binding and cell adhesion (Figure 10) (Table 17 and Table 18). The over-representation of extracellular structure and matrix organization is due to the significant number of mutations that were found to be present in the genes coding for corneal extracellular matrix proteins such as collagens, fibronectin, the proteoglycan heparan sulphate and tenascin. Collagens and fibronectin are also related to integrin binding and cell adhesion. Additionally, this was reflected in the enriched KEGG and Reactome pathways which were related to ECM-receptor interaction, ECM protein degradation, integrin cell surface interactions and the PI3K-Akt signalling pathway (Table 19 and Table 20).

Also, interestingly, missense variants in genes known to be causal for CD subtypes were also found to be present in the Maltese GCD1 patients. The variants found in *ZNF469*, have been reported to be associated with Brittle cornea syndrome (Micheal, Siddiqui, Zafar, *et al*., 2019) while the variant found in the *TCF4* gene was associated with Fuch's Endothelial CD (Wieben, Aleff, Eckloff, *et al*., 2014). On the other hand, the variant in the *PIKFYVE* gene has been reported in literature to be associated with Fleck CD (Table 21) (Gee, Frausto, Chung, *et al*., 2015).

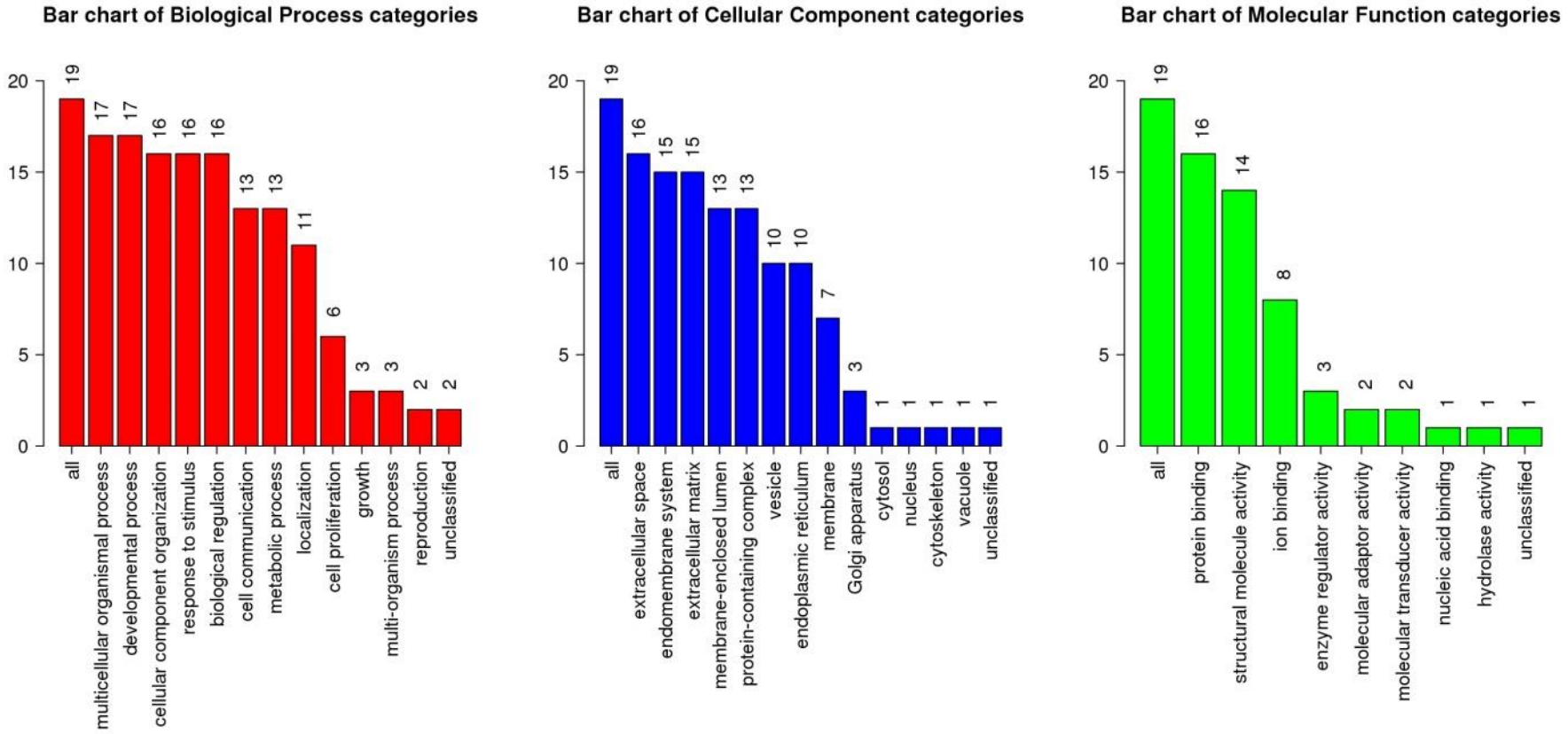


Figure 10: Bar Chart Depicting GO for Biological Process, Cellular Component and Molecular Function category is represented by a red, blue and green bar, repectively. The height of the bar represents the number of IDs in the user list and also in the category

Table 17: GO annotation of biological function of the proteins coded by key mutated genes in the GCD1 patients

GeneSet	GO Biological Process description	Overlap	Enrichment Ratio	pValue	FDR	Genes
GO:0043062	extracellular structure organization	15	34.71667	<2.2e-16	<2.2e-16	*A2M;COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;FN1;HSPG2;ITGA11;MMP9;TGFBI ;TNC*
GO:0030198	extracellular matrix organization	15	40.01921	<2.2e-16	<2.2e-16	*A2M;COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;FN1;HSPG2;ITGA11;MMP9;TGFBI ;TNC*
GO:0038065	collagen-activated signalling pathway	5	356.0684	1.03E-12	3.11E-09	*COL1A1;COL4A1;COL4A2;COL4A3;ITGA11*
GO:0038063	collagen-activated tyrosine kinase receptor signalling pathway	4	370.3111	1.99E-10	4.53E-07	*COL1A1;COL4A1;COL4A2;COL4A3*
GO:0001568	blood vessel development	10	14.15562	2.67E-10	4.85E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;FN1;HSPG2;TGFBI*
GO:0001944	vasculature development	10	13.57445	4.02E-10	5.93E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;FN1;HSPG2;TGFBI*
GO:0072358	cardiovascular system development	10	13.39765	4.56E-10	5.93E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;FN1;HSPG2;TGFBI*
GO:0035987	endodermal cell differentiation	5	102.8642	9.53E-10	9.29E-07	*COL12A1;COL4A2;COL5A1;FN1;MMP9*

GO:0072359	circulatory system development	11	9.925493	9.76E-10	9.29E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;DSP;FN1;HSPG2;TGFBI*
GO:0007155	cell adhesion	12	8.114926	1.05E-09	9.29E-07	*COL12A1;COL1A1;COL3A1;COL4A3;COL5A1;COL6A5;DSP;FN1;IL6R;ITGA11;TGFBI ;TNC*

Table 18: GO annotation of molecular function of the proteins coded by key mutated genes in the GCD1 patients

GeneSet	**GO Molecular Function description**	**Overlap**	**Enrichment Ratio**	**pValue**	**FDR**	**User ID**
GO:0005201	extracellular matrix structural constituent	13	76.18987	<2.2e-16	<2.2e-16	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;FN1;HSPG2;TGFBI ;TNC*
GO:0030020	extracellular matrix structural constituent conferring tensile strength	9	225.2432	<2.2e-16	<2.2e-16	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5*
GO:0005198	structural molecule activity	14	16.38938	6.66E-16	4.17E-13	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;DSP;FN1;HSPG2;TGFBI ;TNC*
GO:0048407	platelet-derived growth factor binding	5	420.9091	3.68E-13	1.73E-10	*COL1A1;COL1A2;COL3A1;COL4A1;COL5A1*
GO:0019838	growth factor binding	7	46.97101	6.75E-11	2.53E-08	*A2M;COL1A1;COL1A2;COL3A1;COL4A1;COL5A1;IL6R*
GO:0002020	protease binding	5	37.33871	1.67E-07	5.03E-05	*A2M;COL1A1;COL1A2;COL3A1;FN1*
GO:0005178	integrin binding	5	36.45669	1.88E-07	5.03E-05	*COL3A1;COL4A3;COL5A1;FN1;TGFBI*
GO:0005518	collagen binding	4	55.28358	6.99E-07	1.64E-04	*FN1;ITGA11;MMP9;TGFBI*
GO:0043394	proteoglycan binding	3	79.37143	6.77E-06	0.001401	*COL5A1;FN1;TNC*
GO:0050839	cell adhesion molecule binding	6	11.62343	7.46E-06	0.001401	*COL3A1;COL4A3;COL5A1;DSP;FN1;TGFBI*

Table 19: KEGG pathways associated with the proteins coded by key mutated genes in the GCD1 patients

GeneSet	KEGG pathway description	Overlap	Enrichment Ratio	pValue	FDR	User ID
hsa04512	ECM-receptor interaction	10	49.26152	7.77E-16	2.53E-13	*COL1A1;COL1A2;COL4A1;COL4A2;COL4A3;COL6A5;FN1;HSPG2;ITGA11;TNC*
hsa04974	Protein digestion and absorption	9	40.39444	2.01E-13	3.28E-11	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5*
hsa04510	Focal adhesion	9	18.26884	2.83E-10	3.07E-08	*COL1A1;COL1A2;COL4A1;COL4A2;COL4A3;COL6A5;FN1;ITGA11;TNC*
hsa05146	Amoebiasis	7	29.45428	1.59E-09	1.11E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;FN1*
hsa04933	AGE-RAGE signalling pathway in diabetic complications	7	28.56173	1.97E-09	1.11E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;FN1*
hsa04151	PI3K-Akt signalling pathway	10	11.41086	2.04E-09	1.11E-07	*COL1A1;COL1A2;COL4A1;COL4A2;COL4A3;COL6A5;FN1;IL6R;ITGA11;TNC*
hsa04926	Relaxin signalling pathway	7	21.75085	1.34E-08	6.25E-07	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;MMP9*
hsa05165	Human papillomavirus infection	9	10.72419	3.14E-08	1.28E-06	*COL1A1;COL1A2;COL4A1;COL4A2;COL4A3;COL6A5;FN1;ITGA11;TNC*
hsa05222	Small cell lung cancer	4	17.5628	6.41E-05	0.002323	*COL4A1;COL4A2;COL4A3;FN1*
hsa05200	Pathways in cancer	6	4.625318	0.001195	0.038949	*COL4A1;COL4A2;COL4A3;FN1;IL6R;MMP9*

Table 20: Reactome pathways associated with the proteins coded by key mutated genes in the GCD1 patients.

GeneSet	Reactome pathway description	Overlap	Enrichment Ratio	pValue	FDR	User ID
R-HSA-1474244	Extracellular matrix organization	15	29.21650055	<2.2e-16	<2.2e-16	*A2M;COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;FN1;HSPG2;ITGA11;MMP9;TNC*
R-HSA-1474228	Degradation of the extracellular matrix	13	54.440079337	<2.2e-16	<2.2e-16	*A2M;COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;FN1;HSPG2;MMP9*
R-HSA-1474290	Collagen formation	10	65.14197531	<2.2e-16	<2.2e-16	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;MMP9*
R-HSA-216083	Integrin cell surface interactions	12	82.768627455	<2.2e-16	<2.2e-16	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;FN1;HSPG2;ITGA11;TNC*
R-HSA-3000178	ECM proteoglycans	11	84.855994155	<2.2e-16	<2.2e-16	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;FN1;HSPG2;TNC*
R-HSA-1442490	Collagen degradation	10	91.605902778	<2.2e-16	<2.2e-16	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;MMP9*
R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	10	96.111111111	<2.2e-16	<2.2e-16	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5;MMP9*

R-HSA-3000171	Non-integrin membrane-ECM interactions	10	99.36911488	<2.2e-16	<2.2e-16	*COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;FN1;HSPG2;TNC*
R-HSA-8948216	Collagen chain trimerization	9	119.9204545	<2.2e-16	<2.2e-16	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5*
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	9	78.75373134	4.44E-16	7.67E-14	*COL12A1;COL1A1;COL1A2;COL3A1;COL4A1;COL4A2;COL4A3;COL5A1;COL6A5*

Table 21: Variants in CD related genes present in GCD patients.

Gene	Variant	Chr	dbSNP ID	Exonic	Transcript	Consequence	Protein Position	Amino Acids	Codons	ClinVar Condition	Clinincal significance
PIKFYVE	G>A/A	2	rs10932258	yes	NM_015040.3	missense_variant	696	S/N	aGt/aAt	Fleck CD: heterozygous c.4166_4169delAAGT mutation causing p.Glu1389AspfsX16 flame-shift mutation	Benign
TGFBI	C>C/T	5	rs121909208	yes	NM_000358.2	missense_variant	555	R/W	Cgg/Tgg	GCD1, GCD2, LCD3A,EBMD, LCD, RBCD, TBCD	Pathogenic
ZNF469	T>C/C	16	rs11648572	yes	NM_001127464.1	missense_variant	357	S/P	Tcc/Ccc	Brittle cornea syndrome	Benign
ZNF469	A>C/C	16	rs11640794	yes	NM_001127464.1	missense_variant	366	R/S	agA/agC	Brittle cornea syndrome	Benign
ZNF469	G>C/C	16	rs7199961	yes	NM_001127464.1	missense_variant	510	G/A	gGc/gCc	Brittle cornea syndrome	Benign
ZNF469	A>G/G	16	rs7197071	yes	NM_001127464.1	missense_variant	1162	K/E	Aag/Gag	Brittle cornea syndrome	Benign
ZNF469	C>T/T	16	rs4782300	yes	NM_001127464.1	missense_variant	1420	P/L	cCg/cTg	Brittle cornea syndrome	Benign
ZNF469	G>G/C	16	rs12598474	yes	NM_001127464.1	missense_variant	2358	G/R	Ggg/Cgg	Brittle cornea	Benign

										syndrome	
ZNF469	T>T/A	16	rs3812956	yes	NM_001127464.1	missense_variant	2670	L/Q	cTg/cAg	Brittle cornea syndrome	Benign
ZNF469	A>G/G	16	rs1983014	yes	NM_001127464.1	missense_variant	2848	H/R	cAt/cGt	Brittle cornea syndrome	Benign
ZNF469	G>G/C	16	rs1105066	yes	NM_001127464.1	missense_variant	3630	E/Q	Gag/Cag	Brittle cornea syndrome	Benign
ZNF469	A>G/G	16	rs904783	yes	NM_001127464.1	missense_variant	3636	T/A	Acc/Gcc	Brittle cornea syndrome	Benign
ADAMTS18	C>C/A	16	rs12935394	yes	NM_199355.2	missense_variant	946	A/S	Gcc/Tcc		Benign
TCF4	C>G/G	18	rs611326	yes	NM_001243226.1	missense_variant	10	A/P	Gca/Cca	Not reported in clinvar (Fuch's endothelial dystr.)	

3.3 PHASE 3: Inhibition of *TGFBI* as a Treatment Modality: Functional Work

The adult human corneal epithelium is the primary source of TGFBIp production (Malkondu, Arıkoğlu, Erkoç Kaya, *et al*., 2020) (Niu, Liu, Liu, *et al*., 2012). In this PhD study, in order to explore the inhibition of *TGFBI* as a treatment modality, the first step involved establishing a cell culture of human corneal epithelial cells.

3.3.1 Culture of Human Corneal Epithelial Cells

3.3.1.1. Protocol for Media Preparation for HCECs

Human Corneal Epithelial cells (HCEC) and EpiGRO Human Ocular Epithelia Complete Media Kit were purchased from Merck Millipore (Sigma-Aldrich). This medium was chosen for the culture of HCECs in this study since it is a serum-free medium optimized for HCECs. It does not contain antimicrobials or phenol red, i.e. compounds that could possibly interfere with experimental results or cause cell stress. It is also especially formulated to promote cell growth in an incubator with an atmosphere of 5% CO_2. 0.25% Trypsin-EDTA digestion solution was also needed in order to detach cells from tissue culture dishes.

The growth factors were defrosted at room temperature for 15 minutes and then in a 37˚C water bath for 5 minutes. The closed containers containing the growth factors and T-25 flask were sprayed with 70% EtOH and placed in the biosafety cabinet that was previously cleaned with 70% alcohol and treated with UV irradiation. The growth factors which included L-Glutamine, EpiFactor O, Epinephrine, EpiFactor P, rh Insulin, Apo-Transferrin and Hydrocortisone hemisuccinate (Table 22) were placed individually into the 485mL of Basal Medium.

Table 22: Growth factor supplements.

Growth factor Supplement	**Volume**	**Final concentration after addition to the basal medium**
L-Glutamine	15mL	6mM
EpiFactor O	1mL	Proprietary
Epinephrine	0.5mL	1.0 micrM
EpiFactor P	2mL	0.4%
rh Insulin	0.5mL	5microg/mL
Apo-Transferrin	0.5mL	5microg/mL
Hydrocortisone hemisuccinate	0.5mL	100mg/mL

3.3.1.2 Protocol for Thawing of HCEC

The vial containing the HCEC was wiped with 70% EtOH. The cap of the vial was twisted open to release pressure and retightened. The HCECs were thawed in a 37˚C water bath. The vial was wiped again with 70% EtOH and the suspension was pipetted into a T-25 flask. The vial was rinsed with 1ml of pre-warmed medium and this was added to the cells in the T-25 flask.

3.3.1.3 Passaging of Cells

When cells become confluent, growth slows down and can stop. Thus, cells are subcultured to keep them healthy and prevent them from going into apoptosis. The cell cultures were examined and fed with supplemented medium every other day until they reached 80% confluency. Trypsin-EDTA digestion solution was added to detach cells from tissue culture dishes and to dissociate cells from one another. The plate was gently rocked to make sure the solution covered the cell layer completely and left there for 3 minutes. The suspension containing the dissociated cells was transferred to a conical tube and centrifuged at 200 × g for 5 to 10 min. The pellet was collected and re-suspended in 10 ml of complete culture media. This was transferred to a 10 cm

culture plate and incubated at 37°C with 5% CO2 for 3 days. When the cells reached 80% confluency, the suspension was subcultured for 3 more generations.

3.3.1.4 Protocol for Cell Culture of HCEC

4mls of the pre-warmed (37˚C) complete medium were pipetted into the T-25 flask containing the HCECs and they were incubated at 37˚C with 5% CO_2. The culture was observed under the microscope for any microbial contamination and to assess growth every other day. When cells reached about 50 to 60% confluency, the spent media was removed and replaced with fresh supplemented medium and incubated again at 37˚C with 5% CO_2.

An aseptic working technique including sterilization and disinfection of surfaces, equipment and media containers used was followed during all steps of the cell culture to avoid any microbial contamination. Daily microscopic observation of the cultures was carried out to look for any cloudiness in the medium and to look for any biological contaminants.

Maintaining a temperature of 37˚C with 5% CO_2 was possible by using a CO_2 incubator. This equipment is critical in the maintenance of an optimal environment for cell growth. Besides providing an optimal constant temperature and humidity, CO_2 is needed to act as a buffer to help keep the pH within the ideal physiological range for the cells to grow (Tayebi-khorami, Chegeni, Tahmasebi Birgani, *et al*., 2022).

3.3.1.2 Results

HCECs can be passaged over 3 generations (Figure 11).

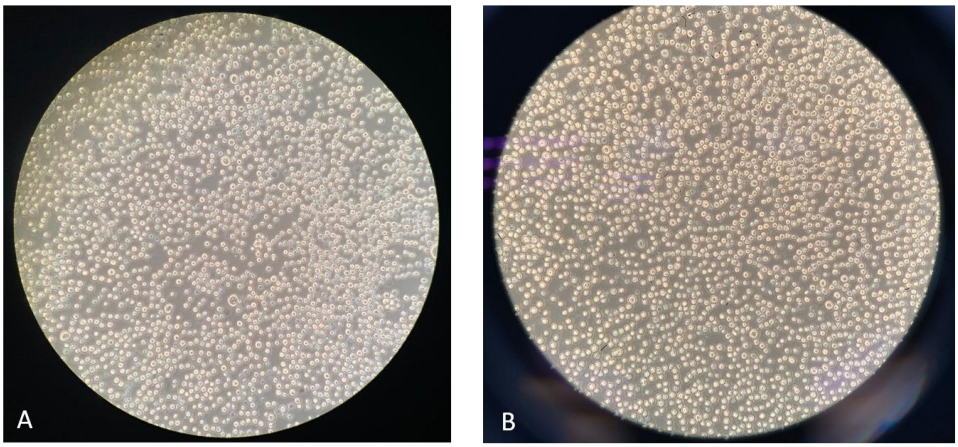


Figure 11: 80% confluent human corneal epithelial cells from passage 1 (A) and 3 (B) under a microscope

HCECs were used in this project to address two issues in which there is a research void to date:

The first was to explore a potential method of inhibiting TGFBIp production in the human cornea since in patients with *TGFBI* dystrophies, mutated TGFBIp forms deposits that subsequently lead to visual impairment. In this research study, the effectiveness of shRNA in the inhibition of TGFBIp production was explored. Additionally, the effect the inhibition of *TGFBI* would have on downstream signalling and subsequently on the overall corneal structure and function were investigated.

The second was to investigate the effects of two compounds/drugs on HCECs to explore their possible use as a treatment modality in *TGFBI* CDs. These compounds were shortlisted after a literature review on compounds that were reported to reduce the production of TGFBIp in the cornea was performed. These compounds were selected to be tested in HCECs in this study since in corneal fibroblasts it has been documented that they lead to the inhibition of mutant TGFBIp production. Yet, to the

best of our knowledge, the effect of Li and MMC on corneal epithelial cells, which would be of greater interest since TGFBIp is mainly produced by HCECs, has never been investigated.

3.3.2 Induction of *TGFBI* Knock Down in Human Corneal Epithelial Cells

One of the aims of this project was to seek a potential method of inhibiting TGFBIp production in the human cornea. In this research study, the effectiveness of RNAi, specifically shRNA, in the inhibition of TGFBIp production was explored.

Mutations in the *TGFBI* gene can potentially have various downstream effects on other genes involved in various cellular processes especially considering the high level of expression of *TGFBI* in the human cornea and the multiple physiological roles TGFBIp has been documented to have. Once knockdown of the shRNA-targeted *TGFBI* gene was established, RNA sequencing and subsequent bioinformatic analysis could be performed to investigate the effects it would have in the context of specific pathways within the cell, and ultimately within the context of *TGFBI* CD. Furthermore, shRNA-induced KD of the mutant *TGFBI* gene in corneal cells of patients with *TGFBI* CDs, can eventually be looked into as a prospective treatment approach, since theoretically, it can prevent deposit recurrence after corneal transplantation or even inhibit primary corneal deposit formation.

3.3.2.1 Materials and Methods for Induction of TGFBI KD in HCECS:

The induction of *TGFBI* knockdown in HCECs using a lentiviral vector containing shRNA sequences targeting the *TGFBI* gene was planned and subsequently conducted through a collaborative project with LabOmics, a specialized laboratory equipped for viral vector manipulation and transduction. Due to the stringent biohazard

containment requirements and the need for specialized facilities for viral vector work, which are not available within our laboratories in Malta, this step was outsourced to LabOmics laboratories.

This collaboration ensured the successful steps of the experimental protocol while adhering to the highest standards of safety and precision. All samples and data generated from this collaborative endeavor formed the next downstream processes which I resumed working on in my capacity.

3.3.2.1.1 Pre-project Testing on Primary Human Corneal Epithelial Cells (HCEC):

3.3.2.1.1.1 Method for Determination of the Minimum Antibiotic Concentration

The materials used for this part of the project were: Growth medium: DMEM Complete Medium (90% DMEM high glucose medium supplemented with 10% fetal bovine serum); Selection medium: growth medium supplemented with Puromycin at the appropriate concentration for cell selection; and Puromycin.

The cells were plated in a 96-well plate and cultured overnight at 37°C with 5% CO2. The spent media was replaced with fresh complete Growth Medium supplemented with 1, 2, 3, 4, 5 µg/mL Puromycin, respectively. The control did not have Puromycin added to the fresh Growth Medium used to replace the spent media. Cell growth in each well was monitored using a microscope for 72 hours. After 48 hours the medium was replaced with fresh Selection Medium containing the above range of Puromycin concentrations to be tested. The percentage of surviving cells was noted daily and the minimum antibiotic concentration required to kill 100% of untreated control cells in 72 hours from the start of antibiotic selection was determined.

3.3.2.1.1.2 Results of Pre-project Testing and Minimum Antibiotic Concentration Determination

The minimum antibiotic concentration was found to be 2 µg/mL Puromycin (Figure 12).

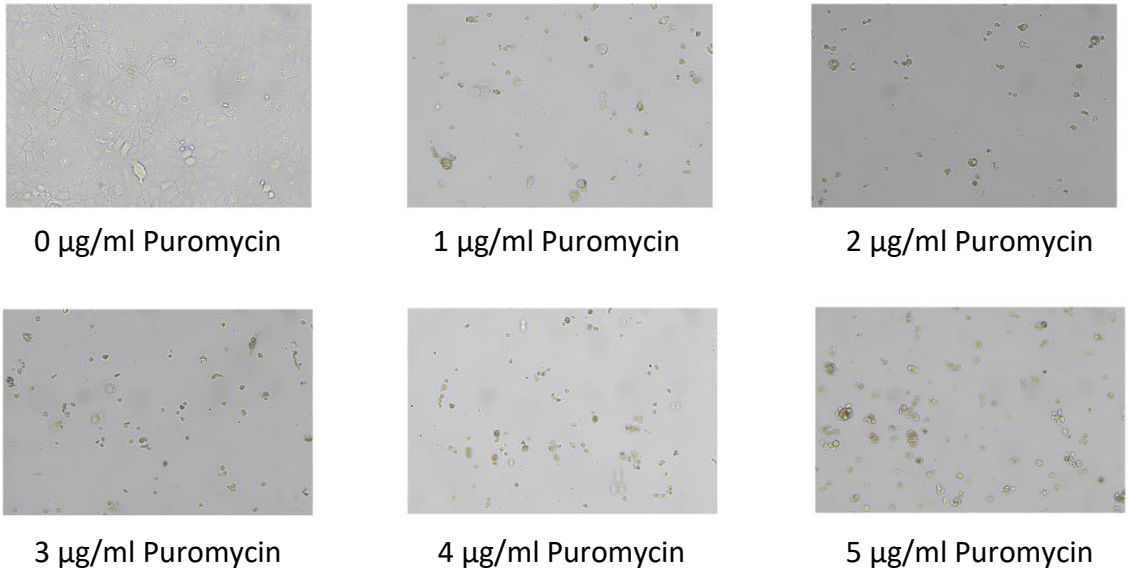


Figure 12: Minimum antibiotic concentration determination. Human corneal epithelial cells treated with a range of Puromycin concentrations observed under a 100X lens with a microscope

3.3.2.1.2 h-TGFBI shRNA Lentiviral Vector Construction and Virus Packaging

RNAi is a process by which double-strand RNAs inhibit gene expression through the use of nucleotide sequences that are complementary to the targeted messenger RNA (Mello & Conte, 2004) (Dykxhoorn & Lieberman, 2005). It can be carried out by using small interfering RNAs (siRNAs) or vector-based expression of short hairpin RNAs (shRNAs) that are synthesised within the cell itself after transfection by plasmid vectors or by transduction with virally produced vectors (Figure 13). ShRNAs consist of two complementary 19–22 bp RNA sequences that are linked by a short loop sequence.

The shRNA sequence is transcribed in the nucleus and then exported to the cytosol. An endogenous enzyme, Dicer, processes the shRNA by removing the loop to release small interfering RNA (siRNA) duplexes. The duplex is then unwound and incorporated into the RNA induced silencing complex (RISC). This process favours one of the two strands due to the difference in thermodynamic stability at the ends of the duplex. The selected strand guides the RISC to bind to the target mRNA in order to degrade it (Taxman, Livingstone, Zhang, *et al*., 2006). ShRNA can be used to produce stable knockdown cell lines and has less chance of off target effects than synthetically produced siRNA. This method of transduction helps eliminate the need for multiple rounds of transfection and increases the reproducibility of results (Moore, Guthrie, Huang, *et al*., 2010), however, creating a stable shRNA cell line is time-consuming.

Lentiviruses form part of the group of retroviruses. They are unique in that they are able to infect both dividing and non-dividing cells since it can pass through the nuclear membrane. It also expresses reverse transcriptase and integrase which makes it possible to convert the viral RNA to double stranded DNA and insert this viral DNA into the host chromosomes (Anderson & Hope, 2005). Thus, they are ideal for difficult to transfect cells as well as in situations where high efficiency is needed.

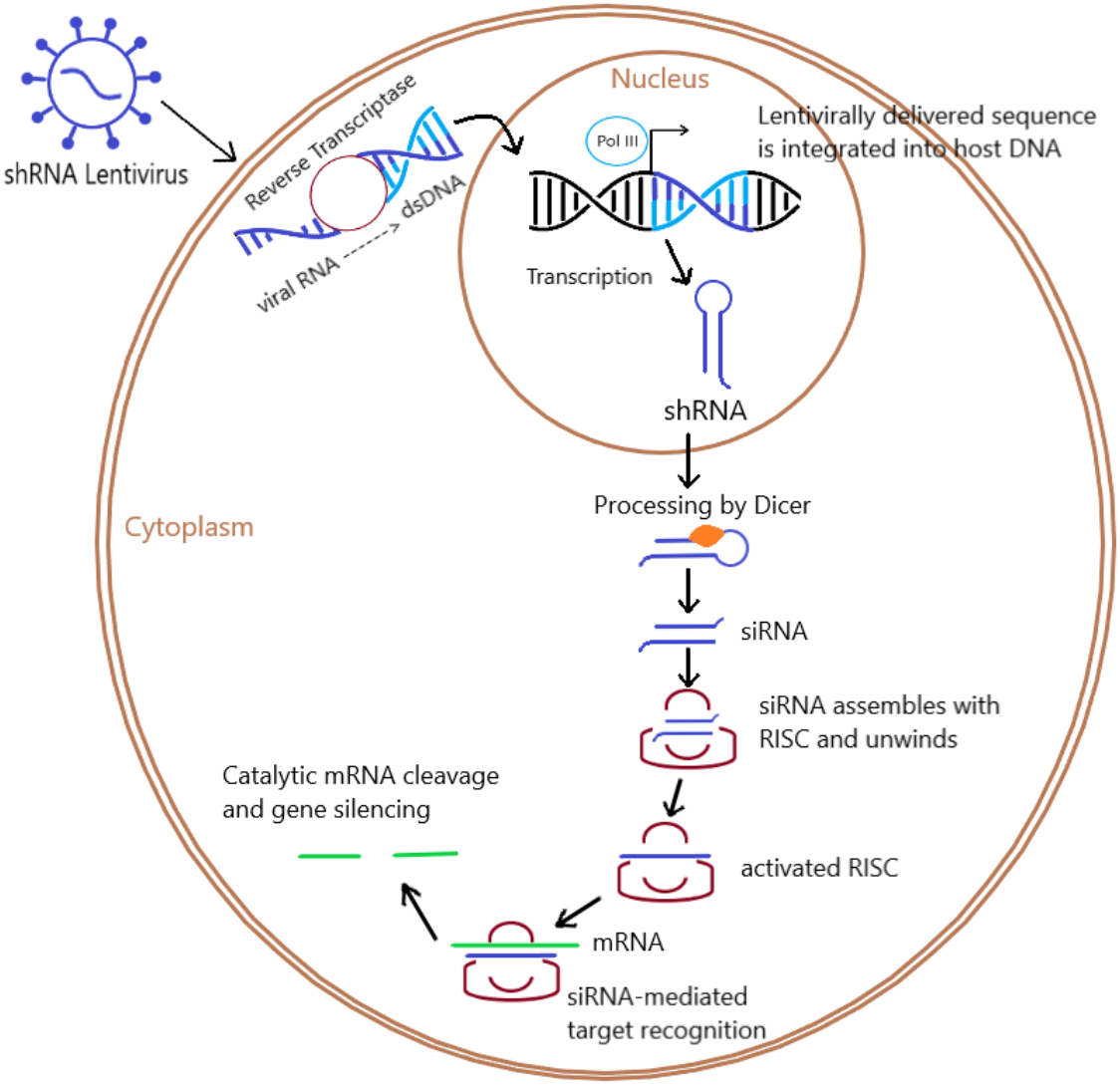


Figure 13: Diagram illustrating targeted gene silencing by shRNA lentiviral transduction

3.3.2.1.2.1 Construction of shRNA Lentiviral Vector

Materials for construction of shRNA lentiviral vector:

Reagent Names	**Instrument Names**
Vector	PCR Thermocycler
DH5α Competent Cells	Centrifuge
DNA Polymerase	Constant Temperature Incubator
PCR Master Mix	Constant Temperature Water Bath
Plasmid Purification Kit	Micropipette
DNA Gel Extraction Kit	UV Imager
DNA Ladder	Gel Electrophoresis Instrument
Restriction Endonuclease	

Materials for production of lentivirus:

Reagent Names	**Instrument Names**
Fetal Bovine Serum	Cell Incubator
E. coli strain DH5α	Biological Safety Cabinets
Trypsin	High Speed Refrigerated Centrifuge
Plasmid DNA Extraction Kit	Inverted Microscope
DMEM	Ultra-High Speed Refrigerated Centrifuge
Lipofactamine	
PBS	

ShRNA lentiviral vector construction workflow followed is depicted below (Figure 14):

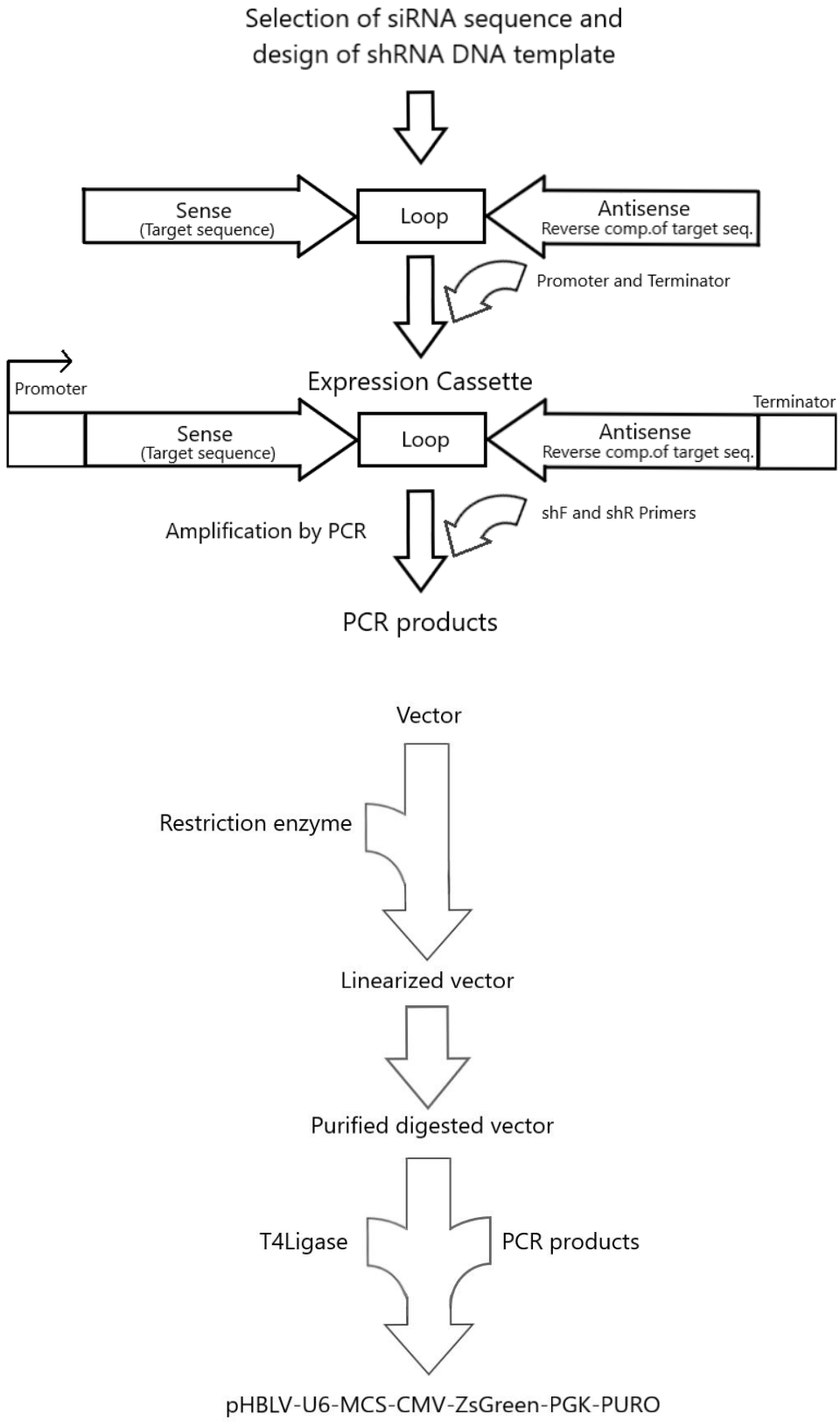


Figure 14: ShRNA lentiviral vector construction workflow.

Where HBLV = Human B-lymphotropic virus; U6 = a type III RNA polymerase III promoter; MCS = multiple cloning site; CMV = cytomegalovirus; zsGreen = green fluorescent protein; PGK = phosphoglycerokinase gene promoter and PURO = puromycin.

The vector used in this study is pHBLV-U6-MCS-CMV-ZsGreen-PGK-PURO. In order to construct the h-*TGFBI* shRNA Lentiviral Vector, first, the primers for the target fragment were selected. This was followed by the designing and synthesis of the shRNA oligonucleotides. The vector was digested with appropriate restriction enzymes a linearized vector was obtained by using a gel extraction kit. The target fragment was inserted into the linearized vector using T4 ligase. Competent DH5α cells were transformed and incubated for 12-16 hrs. Single colonies for sequencing validation were then selected and the plasmid was extracted from validated clones.

3.3.2.1.2.1.1 Vector Information

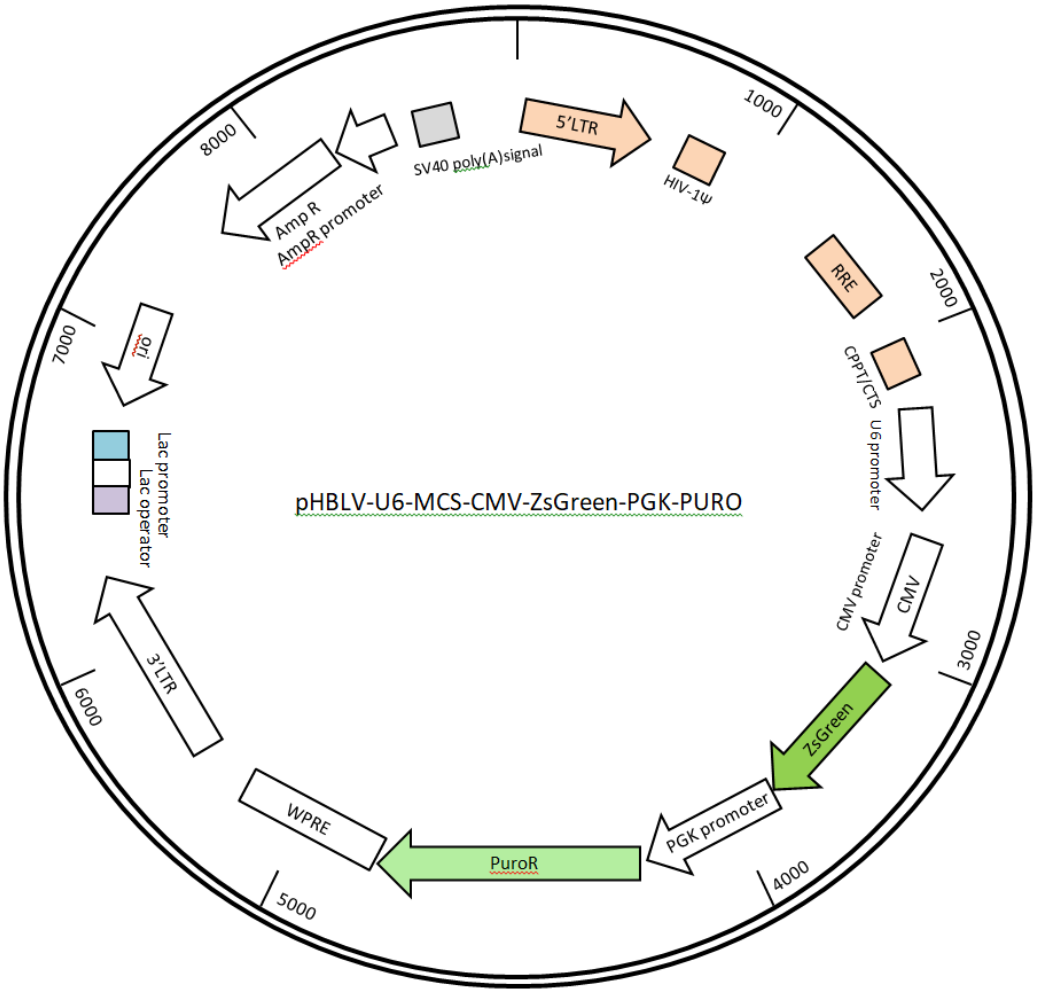


Figure 15: Vector map of pHBLV-U6-MCS-CMV-ZsGreen-PGK-PURO
Where HBLV = Human B-lymphotropic virus; U6 = a type III RNA polymerase III promoter; MCS = multiple cloning site; CMV = cytomegalovirus; zsGreen = green fluorescent protein; PGK = phosphoglycerokinase gene promoter and PURO = puromycin.

3.3.2.1.2.1.2 Selection of siRNA Sequences and Designing shRNA Oligonucleotides

The shRNA sequence generally consists of a stem of 19–29 nucleotides, which is the target sequence (sense sequence), a loop of at least 4 nucleotides (nt), and a 19–29 nucleotide antisense sequence, to form a sense-loop-antisense sequence. An overhang sequence is also required, complementary to the overhang of the pRNAi vectors, to restore the promoter and terminator sequences in the vector (Ge, Ilves, Dallas, *et al*., 2010).

When designing the shRNA oligonucleotides to be used in this study, general guidelines that have been documented in literature, which result in highly effective gene silencing in mammalian cells, were followed. These include: the preference for A/T at the 3' end of the sense strand; the preference for G/C at the 5' end of the sense strand; the presence of an A/T-rich stretch at positions 3–6; and the strong preference for A/T at positions 3, 6, 13, and 18 (Li, Lin, Khvorova, *et al*., 2007). When the U6 promoter is used, such as in this study, the target sequence should start with guanosine (G). Intron regions and stretches of 4 or more repeated bases were avoided. Four consecutive T in the target shRNA sequence would cause early transcription termination so this was also avoided. Of note is that for shRNAs that are to be expressed from a vector under control of a pol III promoter, the target sequence should not end with more than one T, since, in combination with the two T's of the loop this would constitute a termination sequence for RNA pol III. Literature suggests that it is usually best to choose sequences with a GC content between 30-60%. However, having said that, there are multiple examples of active siRNAs that contain a high GC content (Reynolds, Leake, Boese, *et al*., 2004). Since studies have also shown that the 9-nt loop that is based on a naturally occurring miRNA sequence (TTCAAGAGA)

exhibits a robust target knockdown, this was used in the construction of the shRNA in this project (Li, Lin, Khvorova, *et al*., 2007) (Brummelkamp, Bernards , & Agami, 2002). The shRNA oligonucleotides were designed to consist of two complementary oligonucleotides, a top and bottom strand for each shRNA target site. The above guidelines were followed as much as possible during the selection and design of the shRNA oligonucleotide sequences so as to design potent shRNAs.

Furthermore, target sequences were tested for homology using the BLAST function in NCBI databse (NCBI, 2016). Sequences with significant homology to other genes were eliminated. The sequences chosen were those that had the fewest amounts of BLAST matches that had a perfect match of 17nt or more to an off target gene and those that target different regions within the gene itself. Nevertheless, since the effectiveness of the shRNA sequence cannot be predicted, and can only be determined by gene suppression analysis, three target sequences were selected in this study. Besides, a control shRNA sequence was also designed to ensure the validity of the study (Jackson & Linsley, 2004).

3.3.2.1.2.1.3 Control Viral Vector siRNA and shRNA Sequence

The control viral vector sequence selected is documented in Table 23.

Table 23: The control viral vector siRNA and shRNA sequence used.

Name	**Sense/Antisense**	**Sequence**
Control siRNA		TTCTCCGAACGTGTCACGTAA
Control shRNA	Sense	GATCCGTTCTCCGAACGTGTCACGTAATTCAAGAGATTACGTGACACGTTCGGAGAATTTTTTC
	Antisense	AATTGAAAAAATTCTCCGAACGTGTCACGTAATCTCTTGAATTACGTGACACGTTCGGAGAACG

3.3.2.1.2.1.4 Target Viral Vector siRNA and shRNA Sequence

Using the ensembl database, the MANE select and Ensembl Canonical transcript of the *TGFBI* gene was selected (ENST00000442011.7). The exons link was selected in order to see the sequences of the different exons and their lengths. This helped in determining the possible sequences to be considered in the design of the shRNA, which were subsequently inputted in the BLAST function of the ncbi database in order to select sequences with the least possible off-target effects (An example of the blast results is shown in the Appendix). The target sequences selected together with the short hairpin sequences for the three shRNAs used in this study are documented in Table 24.

Table 24: The target viral vector sequences and the three shRNA sequences used in this study.

Name	**Sense/Antisense**	**Sequence**
siRNA1		CACCACTATCCTAATGGGATTGTAA
shRNA1	Sense	5'-GATCCGCACCACTATCCTAATGGGATTGTAATTCAAGAGATTAC AATCCCATTAGGATAGTGGTGTTTTTTG-3'
	Antisense	5'-AATTCAAAAAACACCACTATCCTAATGGGATTGTAATCTCTTGA ATTACAATCCCATTAGGATAGTGGTGCG-3'
siRNA2		TCCACTACATTGATGAGCTACTCAT
shRNA2	Sense	5'-GATCCGTCCACTACATTGATGAGCTACTCATTTCAAGAGAATGA GTAGCTCATCAATGTAGTGGATTTTTTG-3'
	Antisense	5'-AATTCAAAAAATCCACTACATTGATGAGCTACTCATTCTCTTGA AATGAGTAGCTCATCAATGTAGTGGACG-3'
siRNA3		CACATTGGTGATGAAATCCTGGTTA
shRNA3	Sense	5'-GATCCGCACATTGGTGATGAAATCCTGGTTATTCAAGAGATAA CAGGATTTCATCACCAATGTGTTTTTTG-3'
	Antisense	5'-AATTCAAAAAACACATTGGTGATGAAATCCTGGTTATCTCTTGA ATAACCAGGATTTCATCACCAATGTGCG-3'

The sequences in red are the target sequence and its complimentary sequence. The middle sequence between these is the short loop that links them together.

3.3.2.1.2.1.5 Construction of shRNA Cassettes

shRNA expression cassettes consist of PCR products that include promoter and terminator sequences flanking a hairpin siRNA template. In this project, these were developed so as to be incorporated into the lentiviral vector in order to permit stable integration into and expression from the host genome. The synthesized oligonucleotides were diluted to 100 µM. The constructs were amplified using sh.forward and sh.reverse primers by PCR. The PCR components (Table 25) and program (Table 26) are shown as below:

Table 25: PCR Components

Reagent	**Volume (µl)**
10 *oligo Buffer	2
Primer F	1
Primer R	1
H2O	16
Total	20

Table 26: PCR Program

Temperature	**Time (min)**
95°C	10min
75°C	10min
55°C	10min
35°C	10min
15°C	10min

3.3.2.1.2.1.6 Cloning of shRNA Cassettes into Lentiviral Vector

In this step, a linearized vector was prepared by digesting with corresponding restriction enzyme 1 & 2 (Table 27) at 37 °C for 15 min, and purified using Gel Extraction Kit. PCR products were then ligated to the digested vector (Table 28) at 22 °C for 1-2 hrs. The completed ligation reaction generated pHBLV-U6-MCS-CMV-ZsGreen-PGK-PURO.

Table 27: Linearization of Lentiviral Vector

Reagent	**Volume (µl)**
Vector DNA(1ug/uL)	1
10*buffer	4
DdH2O	32
Restriction enzyme 1	1.5
Restriction enzyme 2	1.5
Total	40

Table 28: Ligation Program

Reagent	**Volume (µl)**
PCR products	4
Digested vector	X(⩾50ng)
T4 ligase buffer	2
T4 ligase	1
H2O	13-X
Total	20

3.3.2.1.2.1.7 Transformation of Competent Cells

Competent DH5α cells were used for cloning since they are known to increase cloned DNA stability since it contains the endA mutation to prevent cleavage and degradation of DNA. The competent DH5α cells were transformed as follows: The cells that were stored in a -80°C freezer were thawed on ice. 50 µl of cells were added to an eppendorf containing 5 µl of the vector. This was left on ice for 20-30 min. This cell suspension was placed in a 42°C water bath for 60 sec to heat-shock the bacteria and then the tubes were placed immediately in the ice bucket for 2-3 min. 900 µl of LB broth (without antibiotic) was then added to the cell suspension. The cell culture was mixed gently and then the tubes were incubated at 37°C and 230 rpm for 45-60 minutes. 50 -150 µl cell suspension was added from each transformation onto LB agar plates containing ampicillin. The plates were then incubated at 37°C overnight (Figure 16).

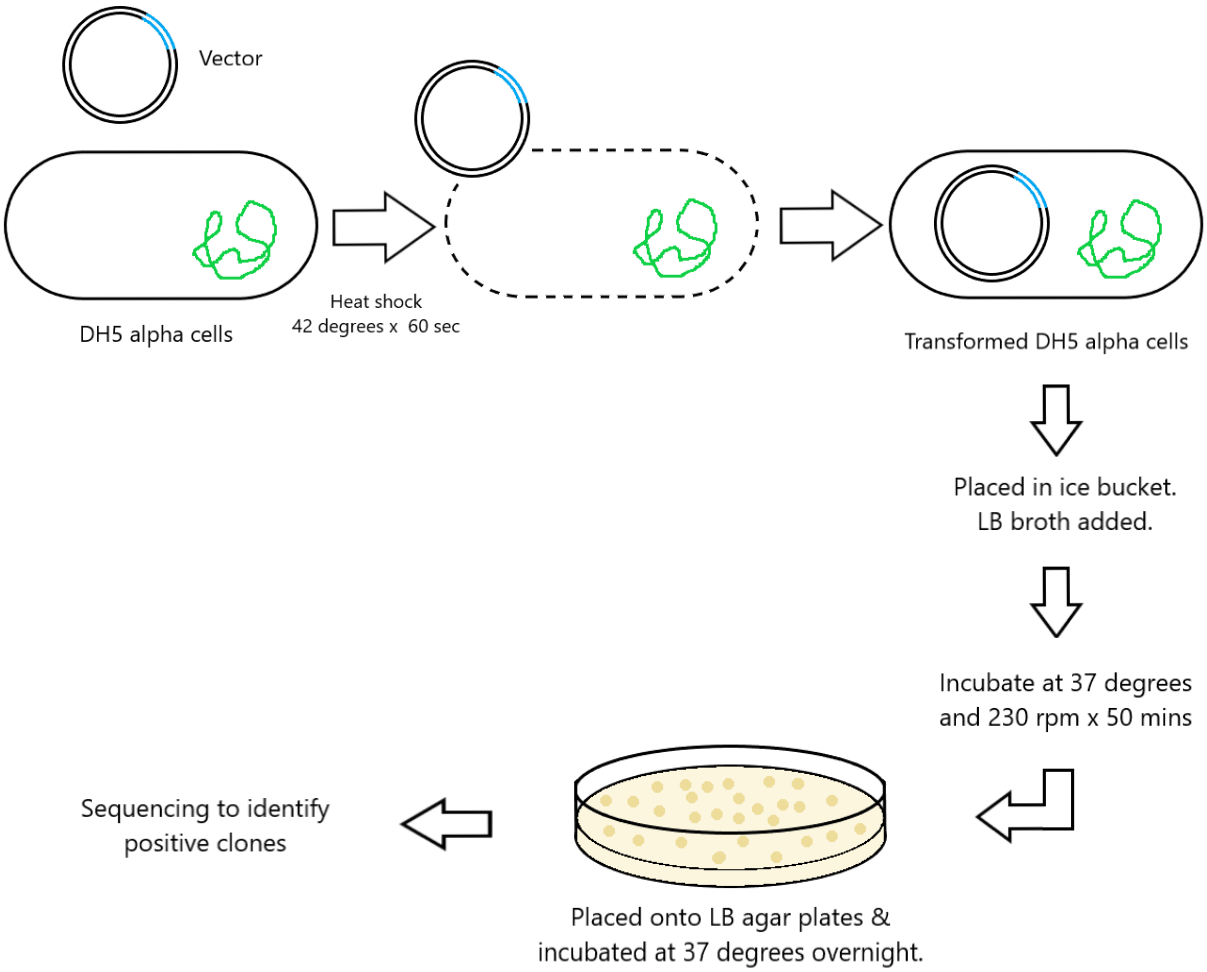


Figure 16: Transformation of Competent Cells

3.3.2.1.2.1.8 Sequencing

A single colony from each plate was picked and positive clones were identified by sequencing.

Comparison of sequencing results are shown as below：

Sequencing results of h-TGFBI shRNA1：

```
----------gatccgcaccactatcctaatgggattgtaattcaagagattacaatcccattaggatagtggtgttttttg---------
:ggaaaggacgaggatccgcaccactatcctaatgggattgtaattcaagagattacaatcccattaggatagtggtgttttttgaattctagt
```

Sequencing results of h-TGFBI shRNA2:

```
----------gatccgtccactacattgatgagctactcatttcaagagaatgagtagctcatcaatgtagtggatttttttg---------
ggaaaggacgaggatccgtccactacattgatgagctactcatttcaagagaatgagtagctcatcaatgtagtggatttttttgaattctagt
```

Sequencing results of h-TGFBI shRNA3:

```
----------gatccgcacattggtgatgaaatcctggttattcaagagataaccaggatttcatcaccaatgtgttttttg---------
ggaaggacgaggatccgcacattggtgatgaaatcctggttattcaagagataaccaggatttcatcaccaatgtgttttttgaattctagt
```

3.3.2.1.2.1.9 Plasmid Extraction

After positive clones were identified, the plasmid vector was extracted using the plasmid miniprep kit (Figure 17).

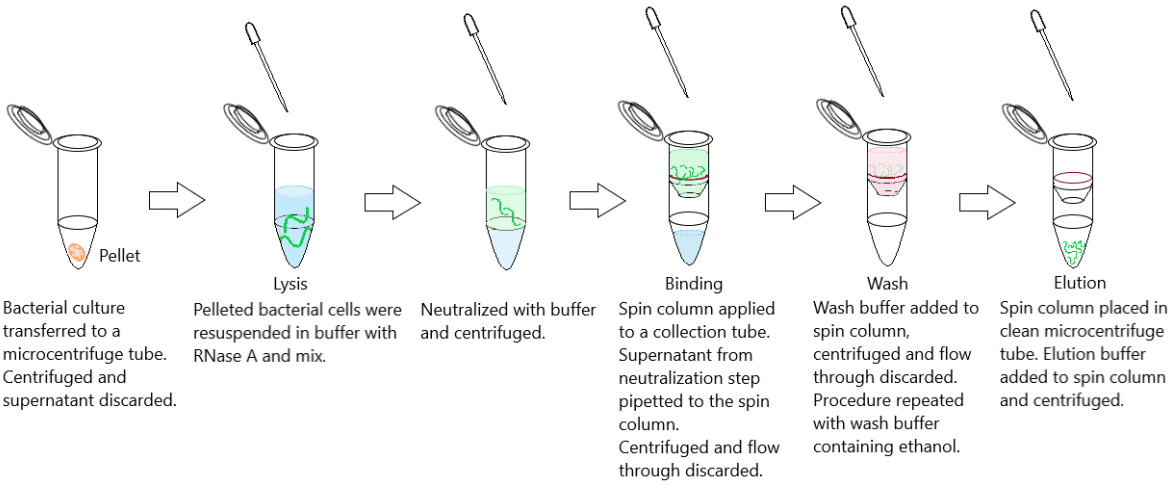


Figure 17: Plasmid Extraction

3.3.2.1.2.2 Production of Lentivirus

3.3.2.1.2.2.2.1 Lentivirus Packaging and Concentration/ Purification:

In order to produce high-titer lentiviral particles, producer cells (in this case HEK293 cells) need to be co-transfected with the lentiviral vector carrying the shRNA, a packaging vector (psPAX2) and an envelope vector (pMD2G). HEK293 cells are immortalized robust human embryonic kidney cells that have been transfected with sheared human adenovirus type 5DNA. The incorporation of adenoviral genes makes it possible for these cells to produce recombinant proteins efficiently making it ideal to be used as a host for gene expression studies and gene therapy. The HEK293T is a daughter cell line derived from HEK293 original cell line that has been transfected with a plasmid vector carrying the SV40 origin of replication that can greatly increase the amount of recombinant protein. This cell line is mostly used for protein expression and production of recombinant retroviruses.

The protocol followed for production of lentivirus in this project included the following steps:

a) Transfection of HEK293T cells with the above packaging system (Figure 18).

b) Collection of lentiviral supernatants at 48 hr and 72 hr after the start of transfection and concentration of lentiviruses by centrifugation.

c) Quality control.

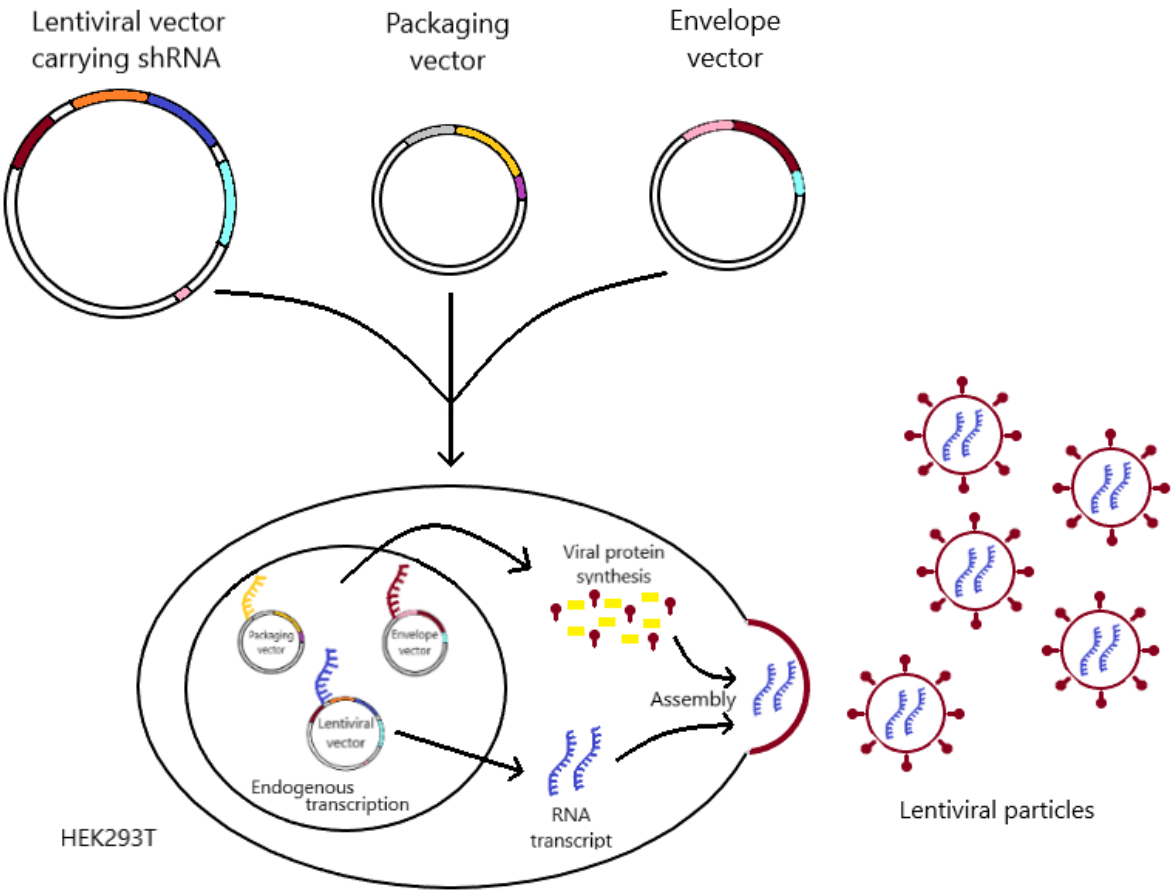


Figure 18: Transfection of HEK293T cells

a) Transfection: 24 hours before transfection, HEK293T cells were seeded into 100 mm tissue culture plates and incubated at 37°C, 5% CO2 overnight. When 70% confluency was reached, the cells were transfected and incubated at 37°C, 5% CO2. After 16 hr of transfection, transfection medium was replaced with 10ml of fresh

complete growth medium containing 10% FBS. At 48 and 72 hours after the start of transfection, transfection medium were collected into 50 ml centrifugation tubes, followed by centrifugation (4°C, 2000 x g, 10 min) to remove cellular debris.

b) Lentiviral supernatants were then collected and centrifuged again (4°C, 82700 x g, 120 min). Lentiviral particles were resuspended in complete medium, aliquoted and stored at -80°C for future use.

c) Quality control included: 1) sterility testing, and 2) mycoplasma detection.

Quality control was completed as shown below in Table 29.

Table 29: Quality control

Testing	**Specification**	**Results**
	Method: 10 µl of virus were added to Hela cells seeded in a 96-well plate and incubated for 24 hr. The cell culture was examined under the microscope.	
Sterility testing	QC standard: It was made sure that the medium culture was clear and transparent and that no large particles in the intercellular space were present. These were also examined to make sure there was no bacterial or fungal contamination.	Sterile
Mycoplasma detection	Method: A test tube containing 10µl of virus was placed into a water bath at 96°C for 15 min. The supernatant was used as a template in PCR testing (the target gene was 16S rRNA of mycoplasma). The results were examined by electrophoresis. QC standard: No band at 500bp.	Negative

HeLa cells were used for quality control due to their well-characterized nature and widespread use in biomedical research. They have robust growth characteristics and are easily transfectable (Ma, 2017). Thus, they are ideal for assessing the efficacy and consistency of the lentiviral vector-mediated knockdown approach before then applying it to the HCECs used in this study.

3.3.2.1.2.2.2.2 Viral Titer Calculation

24 hr before transfection, HEK293 cells were diluted to 1 x 105 cells/ml and seeded 96-well plate at 100 µl per well, 6 wells per sample. These were then incubated at 37°C, 5% CO2 overnight.

On Day 2, six 1.5 ml EP tubes were prepared. 10 µl of viral solution was added to the 1st EP tube, followed by a 3-fold serial dilution for a total of 6 dilutions.

On Day 3, for wells that need to be screened with puromycin, 100 µl of culture medium was aspirated, replaced with 100 µl of 10% FBS complete medium containing 1.5 µg/ml puromycin.

On Day 5, 6 hr before final observation, 80 µl of culture containing virus was aspirated and replaced with 80 µl of fresh complete medium containing 10% FBS, followed by incubation at 37°C, 5% CO2. After 6 h, the wells were observed under a fluorescence microscope. Viral titer was calculated using wells with fluorescence or cell variability of 10-50%.

The following equation was used to calculate the viral titer:

Titer (TU/ml) = cell number × percentage of positive clones × MOI (1) × dilution factor × 10^3 TU/ml

Where MOI = Multiplicity of infection (number of viral particles present relative to host cells)

3.3.2.1.2.3 Development of Knockdown Cell Pools

The protocol followed for the development of knockdown cell pools in this project included the following steps:

a) Transduction of target cells

b) Antibiotic selection of positive knockdown

c) Isolation of RNA from stable knockdown pool/control cells and confirmation by PCR

The reagents required were DMEM Complete Medium, Fetal Bovine Serum (FBS), 0.25% Trypsin-EDTA, Phenol Red, Puromycin, and Polybrene while the equipments used were CO_2 Incubator, Biological Safety Cabinets, Fluorescence Microscope, and High Speed Refrigerated Centrifuge.

a) Transduction of target cells: Human corneal epithelial cells were seeded into 6-well plate ($2x10^5$ cells/well) with DMEM complete medium and incubated overnight. 20 µL of HBLV-h-*TGFBI* shRNA1-ZsGreen-PURO viruses were added to each well. Subsequently, polybrene (5 µg/ml) was also added to each well and the suspension was mixed gently to increase interaction between the cells and the virus. The same procedure was performed for HBLV-h-*TGFBI* shRNA2-ZsGreen-PURO, HBLV-h-*TGFBI* shRNA3-ZsGreen-PURO, HBLV-ZsGreen-PURO negative control, and HBLV-ZsGreen-PURO scrambled control. The cells were incubated with the virus for 24 hr at 37°C in a 5% CO_2 incubator. The next day, the cell culture medium was replaced with fresh DMEM complete medium supplemented with 10% FBS DMEM complete medium and incubated again for 24 hr at 37°C in a 5% CO_2 incubator. Cell growth was observed under a fluorescence microscope.

b) Antibiotic Selection of Positive Knockdown: Puromycin (6 µg/ml) was added to treated cells and controls. Over a 20 day period the cells were passaged and media was replaced every 3 days. A fluorescence microscope was used to identify surviving cells.

3.3.2.1.2.4 Calculation of TGFBI Knockdown Efficiency Using qPCR

TGFBI -shRNA and non-silencing shRNA-transfected cells were harvested after 48-hour incubation. The RNA was subsequently purified using RNA purification kit. The purified RNA was reverse-transcribed into cDNA, transferred to a separate tube and used as template for qPCR. Amplification: the qPCR was performed on a Real-Time PCR Detection System using gene-specific primers for either *TGFBI* or GAPDH.

The percentage of siRNA-induced knockdown of *TGFBI* gene in the human corneal epithelial cells was calculated using the comparative CT method (ΔΔCT) by applying the following formula:

ΔCT (sample) = CT target gene – CT reference gene

ΔCT (calibrator) = CT target gene – CT reference gene

ΔΔCT = ΔCT (sample) – ΔCT (calibrator)

Normalized target gene expression in sample = 2-ΔΔCT

Knockdown efficiency (%) = (1 - normalized target gene expression in sample) x 100%

Where: sample = the *TGFBI* -specific–siRNA transfection; calibrator = the non-silencing-siRNA transfection.

3.3.2.2 Results

3.3.2.2.1 ShRNA Target Sequences BLAST Results

ShRNA1 sense and complementary antisense strand BLAST results returned the *TGFBI* gene as the sequence producing the most significant alignment both with an E value of 0.58 and with the sense strand having a perfect match of 18nt and the

antisense strand having a perfect match of 17nt. The homologous sequences next in line had only 14 identities that matched and E values of 9.

ShRNA2 sense and complementary antisense strand BLAST results also returned the *TGFBI* gene as the top hit with both having an E value of 4e-05 and a perfect match of 25nt. The homologous sequences next in line had only 15 identities that matched and E values of 0.037.

ShRNA3 sense and complementary antisense strand BLAST results again returned the *TGFBI* gene as the top hit with both having an E value of 4e-05 and a perfect match of 25nt. The homologous sequences next in line both had 17 identities that matched and E values of 0.15. Even though 17nt matched in the runner up sequence, this shRNA design was still implemented in the project since the E value was far apart from that given for *TGFBI.* Besides, having a third shRNA helps increase confidence of the results.

On the other hand, the top hit from the BLAST results of the control shRNA gave an E value of 81 and only 14nt giving a perfect match.

3.3.2.2.2 Transduction Check

The transduction check was performed by observing under a fluorescence microscope the HCECs transduced with the scrambled shRNA, empty vector, HBLV-h-*TGFBI* shRNA1-ZsGreen-PURO, HBLV- h-*TGFBI* shRNA2-ZsGreen-PURO and HBLV- h-*TGFBI* shRNA3- ZsGreen-PURO (Figure 19).

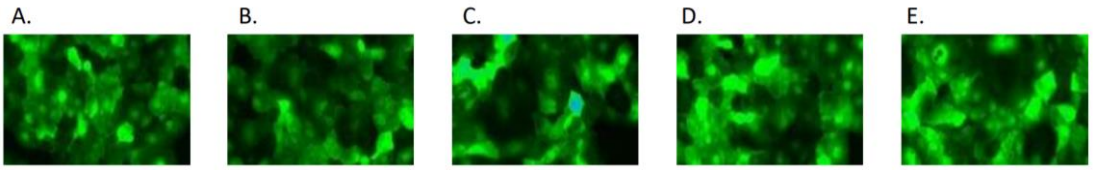


Figure 19: Green flourescence protein expression in human corneal epithelial cells at day 3 post transduction. The above are fluorescence images of hCECs transduced with (A) scrambled control, (B) empty vector control, (C) HBLV-h-*TGFBI* shRNA1-ZsGreen-PURO, (D) HBLV-h-*TGFBI* shRNA2-ZsGreen-PURO and (E) HBLV-h-*TGFBI* shRNA3-ZsGreen-PURO.

3.3.2.2.3 Positive Knock Down Check

Lentivirus knockdown efficiency was checked under a fluorescence microscope (Figure 20).

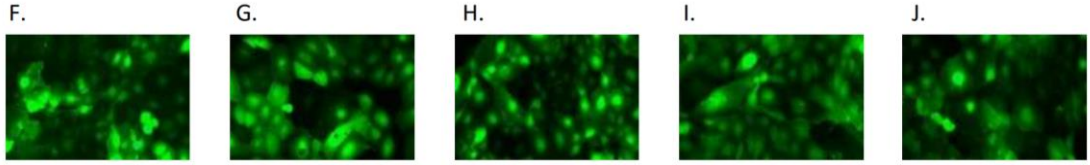


Figure 20: Lentivirus knockdown efficiency recorded under a fluorescence microscope: (F) scrambled control, (G) empty vector control, (H) HBLV-h-*TGFBI* shRNA1-ZsGreen-PURO, (I) HBLV-h-*TGFBI* shRNA2-ZsGreen-PURO and (J) HBLV-h-*TGFBI* shRNA3-ZsGreen-PURO.

3.3.3 A Scoping Review to Identifying and Categorize Compounds that Reduce Corneal Transforming Growth Factor Beta Induced Protein Levels.

In this review the crosstalk between cascades that regulate *TGFBI* expression were explored. Compounds that modulate these pathways leading to a decrease in the expression of *TGFBI* mRNA directly and/or indirectly, or to a reduction in the levels of TGFBIp in human corneal cells were identified and explored. This study helped categorize the identified compounds and describe their mode of action as well as focus on their potential to be used as a cost-effective approach via drug repurposing.

Through a scoping review I was able to provide an overview of the published literature in this area of research since this approach is ideal to employ in developing research fields (Hanneke, Asada, Lieberman, *et al*., 2017), while identifying areas where future research is needed. This methodology allowed me to consider a wide range of study designs (Arksey & O'Malley, 2005) in order to collect most of the literature available for this condition while maintaining a rigorous and transparent outlook. The scoping review is reported in accordance with the Preferred Reporting

Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR).

3.3.3.1 Eligibility Criteria

Database-specific filters were applied in order to retrieve articles written in the English language. The search was conducted on PubMed from inception to 26th November 2021 and on HyDi from inception to 28th January 2022. No time restrictions were applied. The articles taken into consideration in this study included all types of study designs relating to human corneal cells. Articles were included in this study if reference was made to compounds that modulate directly or indirectly *TGFBI* expression resulting in decreased secretion or production of TGFBIp or promote TGFBIp degradation in human corneal cells. Articles that did not analyse levels of *TGFBI* specifically but reported compounds that affected the SMAD, JNK, PI3-K/AKT and the cAMP/PKA cascades, theoretically decreasing the expression of *TGFBI*, were also included in this review. The exclusion criteria included articles that focussed only on the genotypic/phenotypic aspect and surgical management options of *TGFBI* CDs. Studies were also excluded if they were not carried out on human corneal cells. These criteria were applied in order to optimize the search strategy.

3.3.3.2 Information Source and Search Strategy

Electronic bibliographic databases were used to perform a thorough search of original peer-reviewed literature in which compounds that decreased the expression of *TGFBI* mRNA and/or decreased the secretion or production of TGFBIp were documented. The databases utilized were PubMed, HyDi search engine of the University of Malta and The Comparative Toxicogenomics Database (CTD). The databases and journals HyDi search engine provides access to include: American

Chemical Society Publications, BioMed Central, BMJ Journals, British National Formulary, ChemSpider, DynaMed Plus, EBSCO Information Services (includes CINAHL Complete, Cochrane Central Register of Controlled Trials, Cochrane Clinical Answers, Cochrane Database of Systematic Reviews, Cochrane Methodology Register), Journals@Ovid Full Text, Karger, MEDLINE Complete, ProQuest (includes Biological Science database, Consumer Health Database, Health & Medical Collection, Healthcare Administration Database, International Pharmaceutical Abstracts, Nursing & Allied Health Database, PTSDpubs, Public Health Database, Science Database), PubChem, PLOS, Reaxys, Scopus, ScienceDirect, SpringerLink and Taylor & Francis Online amongst others.

The search terms used in each string in this review are: '*TGFBI*', 'cornea*'(here, '*' is used as a wildcard symbol for search term truncation), 'treatment', 'therapy', 'GSK-3', 'SMAD', 'compounds', 'drugs', 'GCD1', 'R555W' (arginine to tryptophan change at position 555 in the *TGFBI* gene), 'medication' and 'decreased expression of *TGFBI*'. Search strings including multiple combinations of these terms were implemented in order to try and include as many relevant articles as possible. The Boolean term AND was used for the PubMed search.

In addition to the above-mentioned databases, the Comparative Toxicogenomics Database (CTD) was used to look specifically at chemicals documented to interact with the TGF-β/SMAD pathway. The CTD is a curated database that provides information about how exposure to various compounds affects humans. It contains data about chemical–gene/protein interactions, chemical–disease and gene–disease relationships. The Gene query applied was '*TGFBI*' and 'SMAD2/3' independently in two different

searches. The Chemical-gene interaction chosen was ‘decreases’. The organism chosen was ‘homo sapiens TAXON:9606’ and related disease was ‘Corneal dystrophies, hereditary (curated)’.

Also, Clinicaltrials.gov, a web-based resource that includes clinical studies, was used to identify whether any clinical research projects studying GCD were or have been carried out.

3.3.3.3 Selection Process

The articles displayed after applying the MeSH terms and database filters were vetted and duplicate articles that were featured in multiple search strings were removed. The studies were screened by title and abstract in order to apply the inclusion and exclusion criteria. The relevant full text articles were assessed to determine the applicability to this research (Figure 21). From the relevant studies, results were extracted to identify the mode of action of the said compounds.

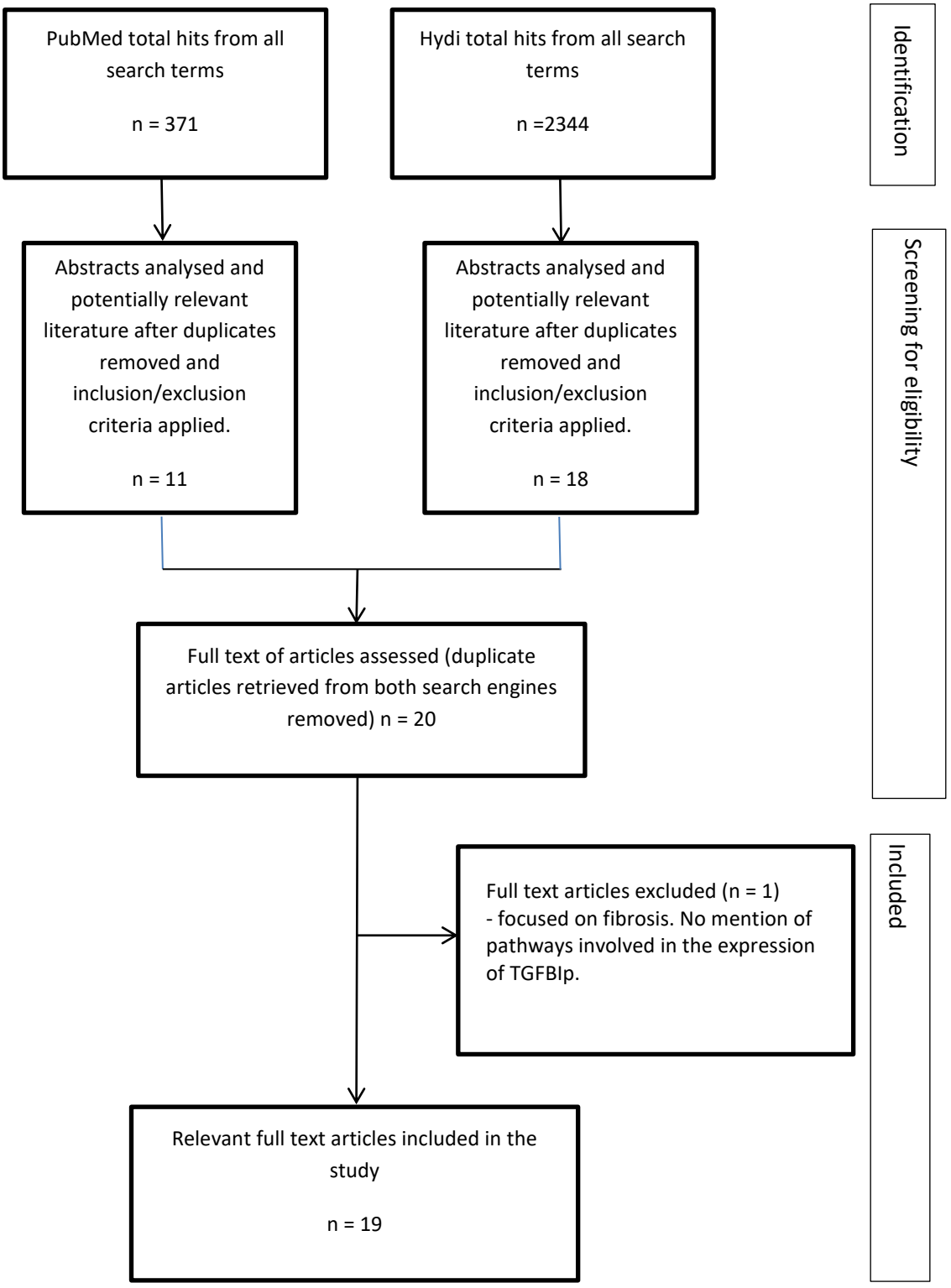


Figure 21: A quantitative overview of the selection process followed (PRISMA flow diagram)

3.3.3.4 Results

Eligible studies were described in detail, including the methods used and outcomes; the evidence identified for each drug was summarised. The drug class for each compound was obtained from PubChem while information on prior use in humans was obtained from the European Medicines Agency (EMA) and Food and Drug Administration (FDA) websites or online repositories of medicine summary of product characteristics. Information on prior use in humans was obtained as a partial indicator of feasibility of repurposing.

The search terms applied in the PubMed and HyDi databases, as well as the number of hits and the relevant hits obtained, are documented in the Appendix. The total number of articles retrieved from all the search terms applied to PubMed was 387, while those from Hydi were 2344. One search query, namely 'inhibition of SMAD AND cornea', resulted in a large number (1010) of retrieved articles in the Hydi search engine. This was due to the retrieval of studies performed on cells other than corneal cells as well articles where SMAD was mentioned but no reference to compound/s that inhibited this pathway was made. The number of relevant full text articles that were included in this study after duplicates were removed added up to 19 (Figure 21). The type of study design used in all of the articles that were chosen to be included in this study turned out to be all analytical experimental laboratory trials (Table 30). Investigators used different types of corneal cells in their studies. Corneal fibroblasts/keratocytes were used in the majority of the studies (in 17 of the studies) while corneal epithelial cells were used in the study by Kim *et al*. (2005) and corneal endothelial cells were used by Okumura *et al*. (Okumura, Kay, Nakahara, *et al*., 2013).

In the study by Lim *et al*. (Lim, Tan, Liu, *et al*., 2016), a rabbit model of corneal fibrosis was also employed besides corneal fibroblasts in cell culture (Table 30).

Table 30: Overview of the papers analysed in the scoping review.

Authors	**Article type**	**Method**	**Outcome**
Ni *et al*., 2018.	Experimental research	Primary corneal fibroblasts were isolated and transfected with wild type or mutant type *TGFBI* over-expressed vectors, and then treated with LiCl and/or autophagy inhibitor 3-methyladenine (3-MA).	Treatment with LiCl inhibited the expression of *TGFBI* and enhanced autophagy.
Choi S. *et al*., 2011.	Experimental research	Levels of TGFBIp and its mRNA in corneal fibroblasts treated with various Li (LiCl) concentrations were analysed.	LiCl treatment reduced the expression levels of normal and mutant TGFBIp in a dose- and a time-dependent manner
Kim, *et al*., 2015.	Experimental research	Analysed the effects of tranilast on phosphorylated SMAD2 (pSMAD2) in GCD2 corneal fibroblasts and determined the changes in the expression of TGFBIp and *TGFBI* mRNA.	Application of tranilast in WT and HO corneal fibroblasts inhibited the expression of TGFBIp by blocking TGF-β signalling.
Chen *et al*., 2020.	Experimental research	Examined the cyclooxygenase-2 (COX-2) expression of 2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b]pyridine-3-carbonitrile (N1) in the human corneal fibroblasts (HCFs) under the treatment TGF-β1.	N1 exerts anti-fibrotic and anti-inflammatory effects through suppression of COX-2, SMAD2, STAT3, iNOS and NF-κB expressions, which suggests they are potential therapeutic targets in the treatment of corneal fibrosis.
Sharma *et al*., 2015.	Experimental research	Human corneal fibroblast cultures were exposed to vorinostat (2.5 µM) to test its effect on TGIF mRNA and protein levels. Myofibroblast formation was induced with TGF-β1 and changes in fibrosis parameters were recorded.	Human corneal fibroblasts demonstrate the expression of TGIF1 and TGIF2 transcription factors. These transcriptional repressors mediate the antifibrotic effect of vorinostat in the cornea.

Lim *et al*., 2016.	Experimental research	Investigated the mechanism of action of a HDACi, ITF2357 in TGF-β-stimulated in vitro primary human cornea stromal fibroblasts (pHCSFs) and in vivo in a photorefractive keratectomy-treated rabbit model of corneal fibrosis.	ITF2357 decreased collagen I, collagen IV, fibronectin, integrin αVβ3 expression; it reduced myofibroblast formation, suppressed phosphorylation of SMAD proteins in TGF-β pathway and inhibited key responsive protein, P4HA1 involved in pro-collagen synthesis.
Kim *et al*., 2005.	Experimental research	Primary human corneal epithelial cells in culture were treated with different concentrations of TGF-β1, with or without TGF-β1–neutralizing mAb, or doxycycline.	Doxycycline inhibits TGF-β1–induced MMP-9 production and activity, perhaps through the SMAD and MAPK signalling pathways.
Nuwormegbe, Park, & Kim, 2021.	Experimental research	Human primary corneal fibroblasts were cultivated and treated with TGF-β1 to induce fibrosis, with or without pre-treatments with different concentrations of lobeglitazone.	Lobeglitazone attenuated TGF-β1-induced ECM synthesis and myofibroblast differentiation of corneal fibroblasts. This was due to the inhibition of the TGF-β1-induced SMAD signalling. It also blocked TGF-β1-induced ROS generation.
Nelson *et al*., 2012.	Experimental research	Human corneal fibroblasts were incubated in TGF-β (to stimulate pro-fibrotic responses from corneal fibroblasts) under halofuginone treatment.	Halofuginone reduced the expression of α-SMA, fibronectin, and type I collagen. It also led to reduced protein expression of SMAD3.
Sarenac *et al*., 2016.	Experimental research	Primary human corneal keratocytes were studied under TGF-β2 stimulation. Keratocytes were treated with IGF-1, and suberoylanilidehydroxamic acid (SAHA) or halofuginone±IGF-1.	IGF-1, SAHA and halofuginone inhibited the TGF-β/SMAD pathway of fibrosis.

Chen *et al*., 2019.	Experimental research	Human corneal fibroblasts (HCFs) were treated with GlcN. The expression of Krüppel-like factor 4 (KLF4) and its downstream signalling effects were determined in the presence and absence of TGF-β1.	GlcN suppresses TGF-β1-induced fibroblast-to-myofibroblast differentiation through the upregulation of KLF4 expression. This effect appeared to be mediated through suppression of SMAD2 phosphorylation and ERK-dependent signalling.
Park *et al*., 2021.	Experimental research	Human keratocytes were induced by exposing them to TGF-β1. Resting (control) and stimulated keratocytes were exposed to various concentrations of sodium nitrite.	TGF-β1-stimulated keratocytes that were exposed to sodium nitrite exhibited decreased SMAD3 phosphorylation.
Nie *et al*., 2020.	Experimental research	JNK signalling activation and inhibition in primary corneal fibroblasts were obtained by treatments with anisomycin and SP600125, respectively.	JNK signalling was activated in GCD2 corneal tissues, and it mediated the TGF-β-induced TGFBIp accumulation and apoptosis of corneal fibroblasts during GCD2 pathogenesis.
Okumura *et al*., 2013.	Experimental research	Human corneal endothelial cells were exposed to SB431542.	SB431542 blocked the transformation of the endothelial cells to fibroblastic phenotypes thus reducing the production of fibrillar ECM proteins.
Choi *et al*., 2016.	Experimental research	GCD2 and WT corneal fibroblasts as well as primary cultured WT corneal fibroblasts that expressed mutant TGFBIp following adenovirus-mediated gene transfer were analysed to determine whether ER stress is associated with GCD2 pathogenesis.	GCD2 cells were found to be more susceptible to ER stress-induced cell death than were wild-type corneal fibroblasts. Treatment with 4-PBA protected against the GCD2 cell death induced by ER stress.

Choi *et al*., 2013.	Experimental research	Primary corneal fibroblasts were treated with 3-MA, bafilomycin A1, and rapamycin.	Melatonin activates autophagy in both wild-type (WT) and GCD2-homozygous (HO) corneal fibroblast cell lines via the mammalian target of rapamycin (mTOR)-dependent pathway.
Choi *et al*., 2017.	Experimental research	WT and GCD2 corneal fibroblasts were divided into control and ER stress (tunicamycin treatment) groups. To investigate the role of melatonin receptor signalling in mediating the protective effects of melatonin, the ER stress group was also treated with luzindole (a nonselective melatonin receptor antagonist), besides melatonin.	Melatonin significantly inhibited GCD2 corneal cell death caused by the ER stress inducer, tunicamycin.
Wang *et al*., 2021.	Experimental research	WT and Thiel-Behnke CD corneal fibroblasts were exposed to ATP-competitive MTOR inhibitor torin 1.	Torin 1 caused the degradation of accumulated mutant-TGFBIp (in Thiel-Behnke CD) via the ameliorative lysosomal function and autophagic flux by increasing transcription factor EB activity.
Kim *et al*., 2008.	Experimental research	Analysed the effect of mitomycin C (MMC) on cell viability, apoptosis, and TGFBIp expression in cultured normal corneal fibroblasts and heterozygote or homozygote granular corneal dystrophy type II (GCD II) corneal fibroblasts.	MMC reduced the production of TGFBIp in all three types of corneal fibroblasts. It induced apoptosis especially in GCD II homozygote cells.

This literature review led to the identification of 16 compounds that can theoretically reduce the levels of mutant TGFBIp in the cornea (Table 31). By focussing on the mode of action of the compounds identified, they were categorized into the following five groups: compounds targeting TGF-β/SMAD and Akt/GSK-3 signalling cascades, compounds targeting JNK signalling cascade, compounds inhibiting DNA synthesis and function, compounds inducing increased elimination of TGFBIp and compounds binding to mutant TGFBIp. From the articles analysed, Li chloride, tranilast and MMC were proven to specifically reduce the expression of *TGFBI*. SB431542 selectively inhibits the TGF-β receptor directly while 9 of these 16 compounds (Table 31) acted further downstream, targeting the SMAD proteins. These included 2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b] pyriine-3-carbonitrile, vorinostat, givinostat (ITF2357), doxycycline, lobeglitazone, halofuginone, glucosamine, nitric oxide and SP600125. Thus it can be postulated that these compounds also lead to a decrease in the expression of *TGFBI*. On the other hand, 3 of these 16 compounds, namely 4-phenylbutyric acid (4-PBA), melatonin and torin-1 were reported to decrease the levels of TGFBIp by increasing the degradation of TGFBIp (Table 31).

In the search query for '*TGFBI*' applied to CTD, there were 281 hits reporting an interacting chemical with this pathway. Of these, 44 were reported to decrease the expression of *TGFBI* mRNA and/or decrease secretion or production of TGFBIp. After going through the literature that was quoted for each of these 44 compounds, only 1 was found to be relevant to this review. This was due to results containing articles that reported reactions occurring in organs/cells other than corneal cells. The only compound that was tested on corneal cells reported in this database was MMC (Kim, Choi, Lee, *et al*., 2008).

A search applied to clinicaltrials.gov, a Web-based resource that includes clinical studies, with the term 'corneal dystrophy' yielded 98 hits. Studies investigating the effect of compounds on CD subtypes targeted Fuchs Endothelial Corneal Dystrophy and Congenital Hereditary Endothelial Dystrophy, which are not *TGFBI* CDs. There were no studies reported that have been carried out on the *TGFBI* CD subtypes.

This review guided the selection process of the compounds that were chosen to be used in the subsequent functional work performed on human corneal epithelial cells. The compounds selected were Li and MMC. Li was chosen since studies have shown that in certain types of cells it downregulates TGFBIp expression via multiple pathways while MMC has been documented to reduce *TGFBI* mRNA levels as well as TGFBIp levels via directly inhibiting DNA synthesis and function.

Table 31: Class and mode of action of compounds that were documented to decrease the levels of TGFBIp in the cornea.

Pharmaceutical Compound	**Drug class**	**Mode of action**	**Summary of evidence**	**Prior use in human populations**
Li chloride	Mood Stabilizer	Reduces SMAD3 phosphorylation. Negatively regulates GSK-3beta by directly inactivating it and indirectly via PI3-K/AKT and PKA cascade. Inhibits the expression of *TGFBI* and enhances autophagy.	Two studies performed in vitro on corneal fibroblasts both showed that Li inhibited the expression of *TGFBI*.	In Europe: Licensed by EMA as a mood stabiliser and has been granted orphan drug status for Huntington's. In USA: for manic disorder.
Tranilast	Anti-Inflammatory Agent, Non-Steroidal Anti-Allergic Agent Histamine H1 Antagonist Platelet Aggregation Inhibitor Calcium Channel Blocker	Reduces levels of pSMAD2 and pSMAD3. Inhibited the expression of TGFBIp by blocking TGF-β signalling.		In Europe: Given orphan drug status by EMA since 2010 for the prevention of scarring post glaucoma filtration surgery. South Korea and Japan: Licensed for allergic conditions.

2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b] pyriine-3-carbonitrile	Anti-fibrotic agent Anti-inflammatory agent	Reduces SMAD2 phosphorylation. Upregulates mRNA of SMAD2 and downregulates mRNA of SMAD3.		No
Vorinostat	Histone deacetylase (HDAC) inhibitor	Transcriptional repressors of TGF-β1 signalling via the SMAD pathway.		In USA: Licenced by FDA for use in the treatment of cutaneous manifestations of cutaneous T-cell lymphoma since 2006. In Europe: Granted orphan drug status by EMA for malignant mesothelioma in 2010 (withdrawn from the register of orphan drugs in 2013 on request of the sponsor).
Givinostat (ITF2357)	Histone deacetylase (HDAC) inhibitor	Suppresses phosphorylation of SMAD proteins in TGF-β pathway.		In Europe: Granted orphan drug status by EMA for Duchenne's muscular dystrophy in 2012.

Doxycycline	Tetracycline antibacterial and anti-malarial	Inhibits activation of SMAD2 signalling pathways. Inhibits JNK signalling activation.		In USA: Licenced by FDA as an antibacterial. In Europe: Granted orphan drug status by EMA for the treatment of systemic amyloidosis caused by beta-2 microglobulin in 2012; for the treatment of familial amyloid polyneuropathy in 2012; for the treatment of retinitis pigmentosa in 2012 (withdrawn from the register of orphan drugs at the request of the sponsor in 2016).
Lobeglitazone	Thiozolidinedione anti-diabetic agent Peroxisome Proliferator-Activated Receptors (PPAR) gamma activator	Inhibits the TGF-β 1-induced SMAD signalling.		No

Halofuginone	Coccidiostat agent Protein Synthesis Inhibitor Angiogenesis Inhibitor Antineoplastic Agent	Reduces protein expression of SMAD3 leading to disruption of TGF-β signalling.	Two studies performed in vitro on corneal fibroblasts both showed that halofuginone inhibited the SMAD pathway.	In Europe: Granted orphan drug status by EMA for the treatment of systemic sclerosis in 2001 (withdrawn from the register of orphan drugs in 2019 at the request of the sponsor); for the treatment of Duchenne muscular dystrophy in 2012 (withdrawn from the register of orphan drugs in 2019 at the request of the sponsor).
Glucosamine	Glycosaminoglycan	Reduces SMAD2 phosphorylation.		Used as a dietary supplement
Nitric oxide	Bronchodilator Agent Neurotransmitter Agent Endothelium-Dependent Relaxing Factor Free Radical Scavenger Gasotransmitter	Reduces SMAD3 phosphorylation.		In Europe: Licensed as a medical gas and granted orphan drug status by EMA for the treatment of cystic fibrosis in 2015 (withdrawn from the register of orphan drugs in 2019 at the request of the sponsor). Also licenced by the FDA as a medical gas.

SB431542	TGF-β type I receptor kinase activity inhibitor	SB431542 is a selective inhibitor of the TGF-βR.		No
SP600125	Selective ATP-competitive inhibitor of c-Jun N-terminal kinase JNK	Reduces levels of pSMAD2 and pSMAD3. Inhibits JNK signalling activation.		No
4-phenylbutyric acid (4-PBA)	Antineoplastic Agent Ammonium Ion Binding Agent Histone deacetylase inhibitor	Reduces level of TGFBIp by increasing its degradation via Endoplasmic-reticulum-associated protein degradation (ERAD) pathway.		No Related compounds include Phenylbutyrate, sodium phenylbutylrate and glyceryl tri-(4-phenylbutyrate). Phenylbutyrate: Granted orphan designation by EMA in 2020 for the treatment of amyotrophic lateral sclerosis. Sodium phenylbutylrate: Licenced by EMA for urea cycle disorders. Glyceryl tri-(4-phenylbutyrate): Granted orphan drug designation by EMA in 2010 for the treatment of argininosuccinic aciduria and for the treatment of hyperargininaemia.

Melatonin	Hormone Central Nervous System Depressant Antioxidants	Reduces level of TGFBIp by increasing its degradation by inducing ERAD pathway activation.	The studies performed in vitro on corneal fibroblasts showed that melatonin activates autophagy in both wild-type (WT) and GCD2-homozygous (HO) corneal fibroblast cell lines and that it reduces ER stress.	In Europe: Licensed by EMA for insomnia. Also licensed by EMA for insomnia in children with developmental disorders. In USA: Available as a dietary supplement.
Torin-1	mTOR Inhibitor	Reduces level of TGFBIp by improving lysosomal function thus increasing degradation of mutant TGFBIp.		No
Mitomycin C	Cross-Linking Reagent Alkylating Agent Antineoplastic Nucleic Acid Synthesis Inhibitor	MMC reduces *TGFBI* mRNA levels as well as TGFBIp levels.		Licenced in Europe and in USA for specific types of cancer.

Licencing data was obtained from EMA and FDA databases (EMA, European Medicines Agency) (FDA, 2016). EMA: European Medicines agency; FDA: U.S. Food and Drug Administration; HDAC: Histone deacetylase; JNK: c-Jun N-terminal kinase; PI3-K/AKT: Phosphatidylinositol-3-Kinase and Protein Kinase B; PPAR: Peroxisome Proliferator-Activated Receptors; *TGFBI*: Transforming Growth Factor Beta Induced.

3.3.4 Treatment of HCECs with Li and Mitomycin C in Vitro

Following the literature review described in the previous section, it was decided that Li and MMC were the compounds the cultured HCECs would be exposed to in this study. Researchers had proposed in previous literature that Li might be a possible compound that could be used in the future in the treatment of *TGFBI* CDs. Besides, it was the only compound licensed to be used in Europe that was previously tested on fibroblasts in GCD2, a subtype of *TGFBI* CDs (Choi, Kim, Dadakhujaev, *et al*., 2011) (Nie, Peng, Li, *et al*., 2018). MMC was the other compound chosen to be tested since, apart from also being licensed in Europe, is routinely used in ophthalmic surgery especially during glaucoma filtration procedures and in the prevention of scarring during refractive laser surgery (Ichhpujani, Singla, Kalra, *et al*., 2022).

3.3.4.1 Materials and Methods for HCEC Treatment with Li

3.3.4.1.1 Protocol for Preparation of LiCl Solutions

0.4239g of LiCl (>99%) was dissolved in 10mls of phosphate-buffered saline (PBS), at room temperature, to obtain a concentration of 1M. It was decided that the concentration of the solution of LiCl the HCEC'S would be exposed to were 5mM and 10mM. Using the equation c1v1=c2v2, it was calculated that 60µL and 30µL of this 1M solution were needed to obtain 6mls of 10mM and 5mM of LiCl solution respectively.

3.3.4.1.2 Protocol for HCEC Treatment with Li

The HCEC's were incubated with the 5mM and 10mM solutions of LiCl at 37˚C with 5% CO_2 for 72 hours in order to examine the effect Li has on the expression of *TGFBI* and *TGFBI*-related genes.

3.3.4.2 Materials and Methods for HCEC Treatment with MMC

3.3.4.2.1 Protocol for Preparation of MMC Solutions

The 20mg MMC vial was reconstituted with 10mls of normal saline. Using the same dilution equation above, 1.2mls and 0.6mls of this solution were made up to 6mls in order to obtain a concentration of 0.4mg/ml and 0.2mg/ml.

3.3.4.2.2 Protocol for HCEC Treatment with MMC

Similarly, the HCEC'S were then incubated in these MMC solutions at 37˚C with 5% CO_2 for 72 hours in order to examine the effect Li has on the expression of *TGFBI* and *TGFBI*-related genes.

3.3.5 RNA Extraction and RNA Sequencing

RNA extraction followed by RNA sequencing, were performed in this study on HCECs, *TGFBI* KD HCECs and Li and MMC exposed HCECs. These results would provide snapshot of the transcriptome of each of these cultured groups of HCECs. It provides information regarding which genes are turned on as well as their level of transcription under these different conditions. These results would subsequently be analysed further by using bioinformatic tools to obtain differential expression of genes between the control HCECs and each of the experimental groups.

RNA extraction, in preparation for RNA-seq, was performed first by employing the spin column method. Consequently, RNA-seq was carried out to detect the presence and quantity of RNA in the samples, thus providing us with a snapshot of the cells' pool of RNAs that could be subsequently used for differential expression analysis between the samples.

3.3.5.1 Materials and Methods for RNA Extraction

The ReliaPrep™ RNA Tissue Miniprep System (Promega) was used to perform RNA extraction (Figure 22). RNase-free 95% ethanol and 100% isopropyl alcohol were also needed in order to carry out this procedure.

The steps involved in isolating intact RNA are 1. cell lysis, 2. denaturation of proteins that may bind to nucleic acids or degrade RNA, 3. inactivation of endogenous ribonuclease (RNase) and 4. separation of RNA from other macromolecules.

The ReliaPrep™ RNA Tissue Miniprep System was used for RNA extraction (Promega, ReliaPrep™ RNA Tissue Miniprep System Technical Manual, 2021). This kit makes use of the chaotropic and protective properties of guanidine thiocyanate (GTC) and 1-Thioglycerol to inactivate the endogenous ribonuclease in the cell samples. The GTC causes release of the RNA into solution which subsequently binds to the ReliaPrep™ Minicolumns by centrifugation. In order to digest any contaminating DNA, RNase-free DNase I was applied directly to the membrane. This was followed by washing steps so as to purify the bound RNA from contaminating salts. Finally, DEPC water was used to elute the total RNA from the membrane.

During this procedure, special care was taken to prevent RNAase contamination by using a sterile technique, utilizing sterile disposable plasticware and wearing gloves throughout the procedure.

The four solutions needed for this protocol were prepared accordingly:

1. DNase I: 275µl of Nuclease-Free Water was added to the vial and it was mixed gently by swirling.

2. LBA + TG Buffer: 3000µl of 1-Thioglycerol were added to 150ml of BL Buffer

3. RNA Wash Solution: 350ml of 95% ethanol was added to the bottle containing 206ml of concentrated RNA Wash Solution (RWA)

4. Column Wash Solution: 36ml of 95% ethanol was added to the bottle containing 24ml of concentrated Column Wash Solution (CWE)

250µl of LBA + TG Buffer were added to the cells. The cell pellet was dispersed and the suspension mixed by pipetting. 85µl of 100% isopropanol was added and mixed by vortexing for 5 seconds.

Using clean gloves, two Collection Tubes were placed in a microcentrifuge tube rack and one ReliaPrep™ Minicolumn was placed in one of the Collection Tubes.

The lysate was transferred to a ReliaPrep™ Minicolumn and centrifuged at 12,000 × g for 30 seconds at 20–25°C. The ReliaPrep™ Minicolumn was removed and the liquid present in the Collection Tube was discarded. The ReliaPrep™ Minicolumn was then placed back into the Collection Tube. 500µl of RNA Wash Solution was added to the ReliaPrep™ Minicolumn and centrifuged at 12,000 × g for 30 seconds. The Collection Tube was emptied again and placed in the microcentrifuge rack.

The DNase I incubation mix was prepared in a sterile tube by combining 24µl of Yellow Core Buffer 3µl, 0.09M MnCl2 and 3µl of DNase I enzyme per sample. This was mixed by gentle pipetting and kept on ice. 30µl of this DNase I incubation mix was applied directly to the membrane inside the minicolumn. The solution was placed in direct contact with and thoroughly covering the membrane.

This was incubated for 15 minutes at room temperature (+20°C to +25°C).

200µl of Column Wash Solution was added to the ReliaPrep™ Minicolumn and this was centrifuged at 12,000 × g for 15 seconds. 500µl of RNA Wash Solution was added

and the solution was centrifuged at 12,000 × g for 30 seconds. The wash solutions were emptied and the Collection Tube discarded.

The ReliaPrep™ Minicolumn was placed into a new Collection Tube and 300µl of RNA Wash Solution was added. This was centrifuged at high speed for 2 minutes. The ReliaPrep™ Minicolumn was then transferred from the Collection Tube to the 1.5ml Elution Tube, and 15µl of Nuclease-Free Water was added to the membrane. It was made sure that the surface of the membrane was covered with the water. The ReliaPrep™ Minicolumn was placed in the centrifuge with the lids of the Elution Tubes facing out. This was centrifuged at 14,000 × g for 1 minute. The minicolumn was removed and discarded. The Elution Tube containing the purified RNA was capped and stored at –70°C.

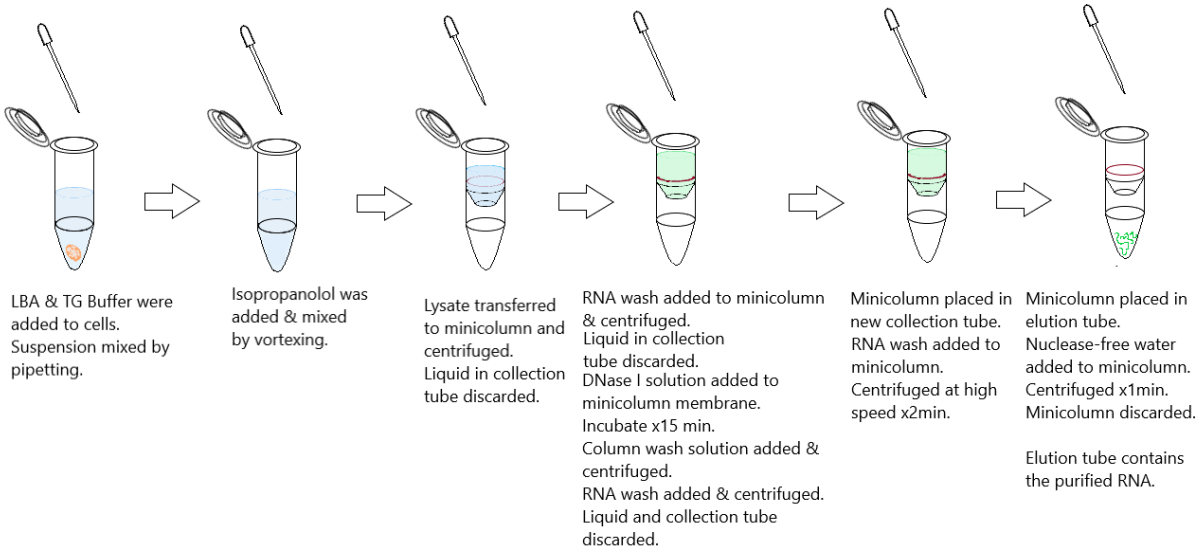


Figure 22: RNA Extraction

3.3.5.1.1 Determination of RNA Yield and Quality by Spectrophotometric Analysis:

Performing spectrophotometric analysis of the RNA samples is imperative prior to carrying out RNA sequencing in order to normalize the samples and quantify RNA. The yield, purity and quality of all the RNA samples were determined by using the NanoDrop spectrophotometer (Thermo Scientific). This instrument allowed

measurement of the concentration of the RNA in just a 1 to 2-µL drop, making analysis very efficient.

Besides the samples, the materials required included 1 blank control, 70% ethanol and deionized water. The samples were thawed and vortexed to homogenize them. The samples were kept on ice and gloves were worn to prevent sample degradation. The spectrophotometer was switched on 15 minutes prior to starting the analysis to allow it to warm up. The RNA mode option was selected from the NanoDrop software. This setting uses a 260nm wavelength to quantify RNA. The pedestal was cleaned with a lint-free wipe. Instrument calibration was performed by pipetting 1 to 2-µL of a blanking solution containing deionized water onto the Nanodrop sample holder and 'Measure blank' selected. The pedestal was cleaned with a lint-free wipe and 1 to 2-µL of one of the samples was pipetted onto the pedestal. The 'RNA' sample type was selected and the 'measure' button was pressed for the instrument to analyse the sample and determine its RNA concentration and quality. The instrument was then cleaned with 70% ethanol in order to remove any residual RNA and wiped with a lint-free wipe. It was allowed to dry before applying the next sample.

RNA quality is evaluated by the Nanodrop spectrophotometer by using the ratios A260/A280 and A260/A230. The ratio A260/A280 (absorbance at wavelengths of 260nm and 280nm) indicates whether there is protein contamination, and thus the purity of the sample. A ratio value of a sample that is between 1.8 and 2.0 is considered to be pure. Lower ratio values indicate contamination and RNA degradation. Furthermore, the ratio A260/A230 can also be calculated. This gives an indication whether contaminants such as salts are present.

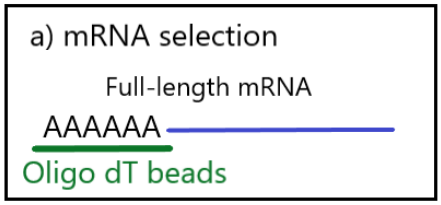
3.3.5.2 Materials and Protocol for RNA Sequencing

RNA-Seq involves ribosomal RNA depletion, complementary DNA (cDNA) library preparation and sequencing (Figure 23). It is based on the principle of converting RNA to cDNA by using a reverse transcriptase with the aim of defining the quantity and sequence of RNA transcripts.

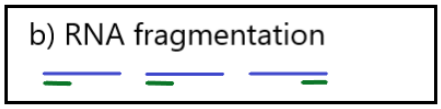
The isolated RNA, which can be total or fractionated, serves as a template for reverse transcription to generate a library of double stranded cDNA fragments. This is done since DNA is more stable than RNA and also can be easily amplified and modified. Sequencing adaptors are then attached to one or both ends of each cDNA fragment. Adaptors allow the sequencing machine to recognize the fragments and make it possible to sequence different samples at the same time since different adapters can be used for different samples. The fragments with attached adapters are amplified.

Each amplified molecule is then sequenced using high-throughput sequencing technology to obtain short sequences. The short sequences can be either from one end (single-end sequencing) or both ends (pair-end sequencing). The generated reads are usually between 30 to 400bp, depending on the DNA sequencing technology used.

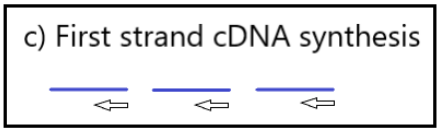
Following sequencing, the reads can either be aligned to a known annotated reference genome and reference transcripts, or else, when there is no reference genome available, they can be assembled de novo. This in turn produces a transcription map comprising of the transcriptional structure and an estimation of level of gene and transcript expression.



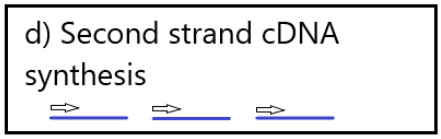
Poly-A containing mRNA molecules are purified using poly-T oligo attached magnetic beads.



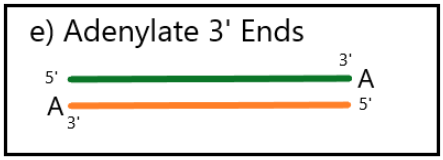
mRNA is fragmented into small pieces using divalent cations.



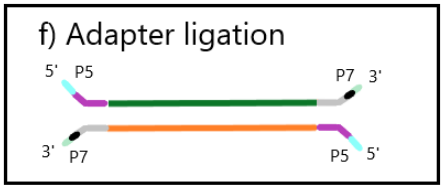
Reverse transcriptase and primers are used to synthesize first strand cDNA from the cleaved RNA fragments.



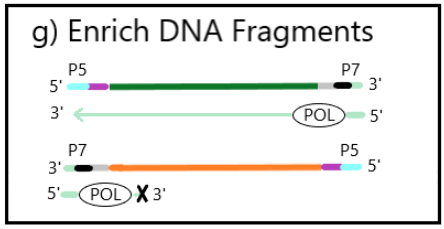
DNA Polymerase I and RNase H are used to synthesize second strand cDNA. The use of dUTP in second strand synthesis quenches the second strand during amplification.



To prevent the fragments from ligating to each other during the adapter ligation reaction, a single 'A' nucleotide is added to the 3' ends of the fragments.



On the 3' end of the adapter, a single 'T' nucleotide provides a complementary overhang for ligating the adapter to the fragment. Adapter ligation is needed for subsequent hybridization of the ds cDNA onto the flow cell.



Since the polymerase used does not incorporate dUTP, the second strand is quenched during amplification. PCR is performed, products purified and cDNA library created.

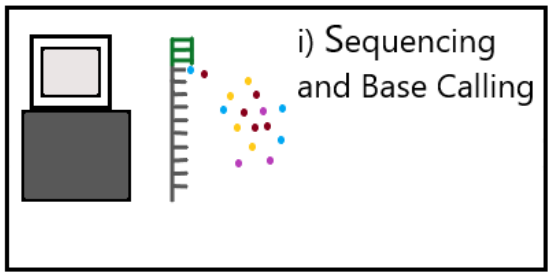


Figure 23: Protocol for RNA Sequencing

In this project, RNA sequencing was performed on HCECs as per cultured protocol. The samples included the five controls, three *TGFBI* KD samples and four samples with differing concentrations of LiCl and Mitomycin C. RNA sequencing libraries were prepared using TruSeq Stranded mRNA library kit (Illumina Cambridge Ltd., UK) according to manufacturer's instructions at Dante Labs, Italy. Sequencing was performed on the Illumina HiSeq2500 sequencer. Samples were sequenced to a depth of 15x10^6 x 50bp single end reads. RNA sequencing reads were mapped to GRCh38 using STAR.

3.3.6 Comprehensive Transcriptome Analysis to Investigate the Differential Expression of Genes Between Normal HCECs, TGFBI KD HCECs, Li and MMC Treated HCECs: Implications for TGFBI CD Treatment.

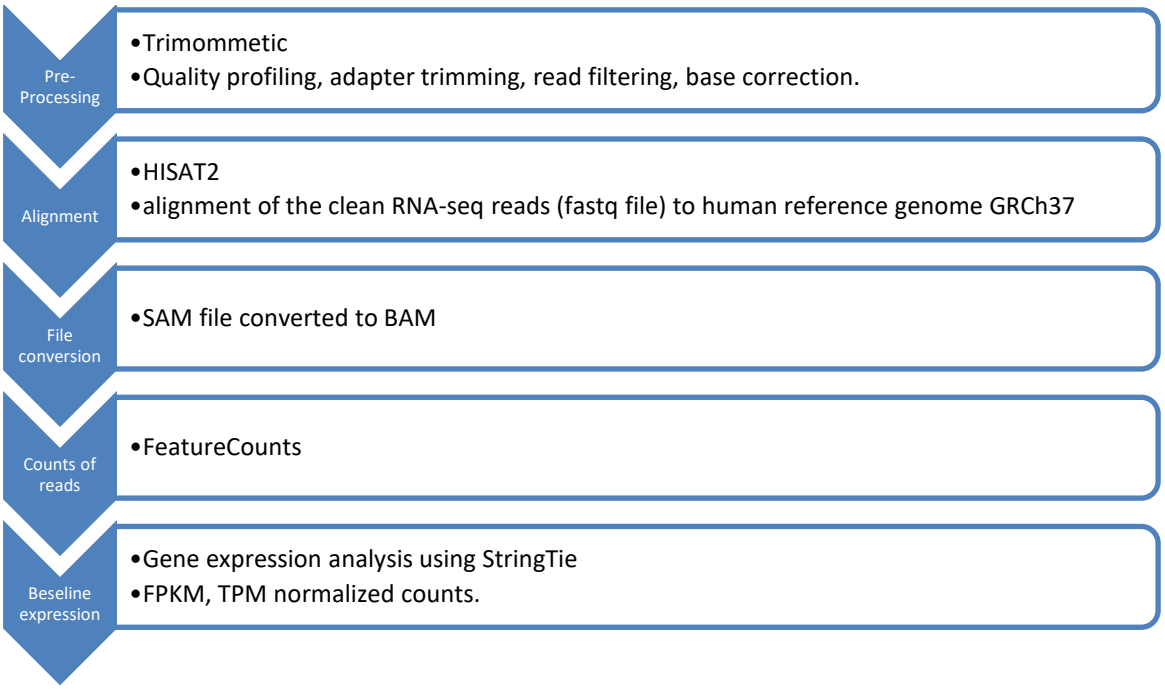
Comprehensive transcriptome analysis using bioinformatics tools was carried out with the aim of investigating the differential expression between normal human corneal epithelial cells (HCECs), HCECs treated with Li, HCECs treated with mitomycin C and *TGFBI* knock down HCECs. The results would help decipher which genes are associated with the *TGFBI* gene, the effect inhibition of the *TGFBI* gene would have on downstream signalling and thus on corneal structure and function. Additionally, the differential gene expression results would show the effects Li and MMC would have on HCECs to investigate whether they could be potentially used as future treatment alternatives for *TGFBI* CDs.

3.3.6.1 Materials and Methods for Bioinformatic Analysis

RNA-seq analysis was performed on a short read dataset that was obtained using Illumina next generation sequencing technology. The RNA-seq workflow analysis steps included the pre-processing of the raw Illumina sequencing data (raw data quality

control, adapter trimming, read-filtering), alignment of short reads, feature counting and differential expression analysis.

RNA-seq workflow analysis steps:



3.3.6.1.1 Data Preprocessing and Alignment

Pre-processing of the raw files was performed using trimmometic, which provides functions including quality profiling, adapter trimming, read filtering and base correction. The alignment of the clean RNA-seq reads (FastQ file) to the human reference genome GRCh37 was performed using the HISAT2 software, a graph-based genome alignment tool. Each SAM file was converted to BAM format, sorted and then indexed, in order to obtain the analysis-ready reads.

3.3.6.1.2 Quantification

In order to generate counts of reads mapped to known genes, the alignment files were provided as input to featureCounts for the quantification step. This counts raw mapped reads for genomic features such as genes and exons. The count data denotes the number of sequence reads from a particular gene. A high number of counts gives a

high number of reads associated with that particular gene, thus implying a higher level of expression of that gene in the sample.

The counts were summarized and reported using the gene_id feature in the annotation file and were saved to count files, which serve as input for downstream RNA-seq differential expression analysis. Summary of counting results can be found in the “RAW_counts_matrix.csv” file.

3.3.6.1.3 Normalised Counts

To enable comparison of gene/transcript expression across all samples outside of the context of differential expression analysis, FPKM (fragments per kilobase of exon per million reads mapped) and TPM (transcripts per million) values were calculated with StringTie v2.2.1 and provided in the FPKM-TPM directory.

3.3.6.1.4 Correlation and Differential Expression Analysis

To carry out the analysis on the various data sets collected, a number of software packages were used. PyCharm, being an integrated development environment, was used for programming in Python language. Python, being a general-purpose programming language, was used for data filtering, for correlation analysis as well as for differential gene expression analysis. In order to calculate Pearson's correlation coefficient between two data samples and to build the volcano plots, the packages SciPy, pandas, NumPy and Plotly were used. Differential gene expression analysis was carried out with a Python implementation of the DESeq2 package. The Microsoft Excel spreadsheet software was used to convert files to csv and xlsx format to convert the data into a more readable form.

The RNA-seq raw count data were used as an input for the pyDESeq package, which is a python implementation of DESeq2. DESeq2 is a statistical package, within the R

platform, that analyses RNA-seq data to obtain differential expression of genes. The code written in order to compare the KD, Li exposed and MMC exposed with their controls can be found in the repository https://github.com/gabysci/corneal_dystrophy.

The DESeq2 package performs normalization internally using the median of ratios method. It then compares gene expression between two customer-defined conditions or groups of samples and finds differentially expressed genes. Groups have typically small sample size and over-dispersion and an expression curve model based on a negative binomial distribution and local regression is used to estimate the relationship between the mean and variance of each gene. Furthermore, it allows scaling factors to be easily included in the statistical test.

The main output result, of the differential expression analysis carried out, are lists of differentially expressed genes (DE genes). The lists with the most significantly over/under expressed genes for each of the following comparisons: *TGFBI* KD HCECs vs Control; Li treated HCECs vs Control; MMC treated HCECs vs Control, can be found in the repository https://github.com/gabysci/corneal_dystrophy.

3.3.6.1.5 Enrichment Analysis

The web-based GEne SeT AnaLysis Toolkit (WebGesalt), a functional enrichment analysis web tool, was employed to perform an Over-Representation Analysis (ORA) in function and pathways. Bioinformatics were used to perform functional enrichment analysis by delineating critical pathways, via KEGG and Reactome databases, and to identify molecular and biological functional GO annotations of the differentially expressed genes. The parameters used to run ORA included the following: Organism of Interest: Homo sapiens; Method of interest: Over-Representation Analysis; Functional Database: Geneontology- Biological process, Molecular function or Pathway- KEGG,

Reactome for every set of DE genes; Gene ID type inserted was the Gene symbol; Gene lists uploaded were the up or downregulated DE expressed genes for comparison group; and Reference set chosen: genome. The advanced parameters included Minimum number of genes for a category: 5; Maximum number of genes for a category: 2000; Multiple test adjustment: BH; Significance level: top 25; Number of catagories expected from set cover: 10; and Number of catagories visualized in the report: 40. The significance threshold applied was <=0.05 for each group of enriched genes.

The genes that showed the largest log_2FCs were looked into. Additionally this study also honed in on and identified the GO annotations and related pathways of 'genes of interest' known to be related to *TGFBI* , which were shortlisted from literature and databases as well as proteins that have been shown to be present in *TGFBI* CD deposits. The 'genes of interest' panel used in this phase of this study is the same as that used in phase 2, with an added list of proteins that were documented to form part of the deposits found in corneas of LCD individuals (Table 5) (Courtney, Poulsen, Kennedy, *et al*., 2015).

This comprehensive transcriptome and enrichment analysis helps to give insight into the significance these DE genes have with regards to their association with *TGFBI*, together with the potential response to future treatment options.

3.3.6.2 Results

3.3.6.2.1 Correlation Analysis

Correlation analysis to determine the strength of association between all the RNA sequenced samples (which consist of 2 negative controls, 2 scrambled controls, drug negative control, KD1, KD2, KD3, high and low Li and high and low MMC HCECs) was carried out using Python language in PyCharm software. Subsequently, a heatmap showing Pearson's correlation coefficients between dataset samples was constructed (Figure 24), with positive numbers indicating positive correlations and describing the strength of the correlation between the two samples. Pearson's correlation reflects the linear relationship between two variables by taking into account differences in their mean and standard deviation. Thus, at a glance, this heatmap helps us understand better the inter- and intra-group variability by displaying the distance between samples, which is represented as correlation. The individual correlation plots between all samples together with the code needed to generate them can be found in the repository https://github.com/gabysci/corneal_dystrophy.

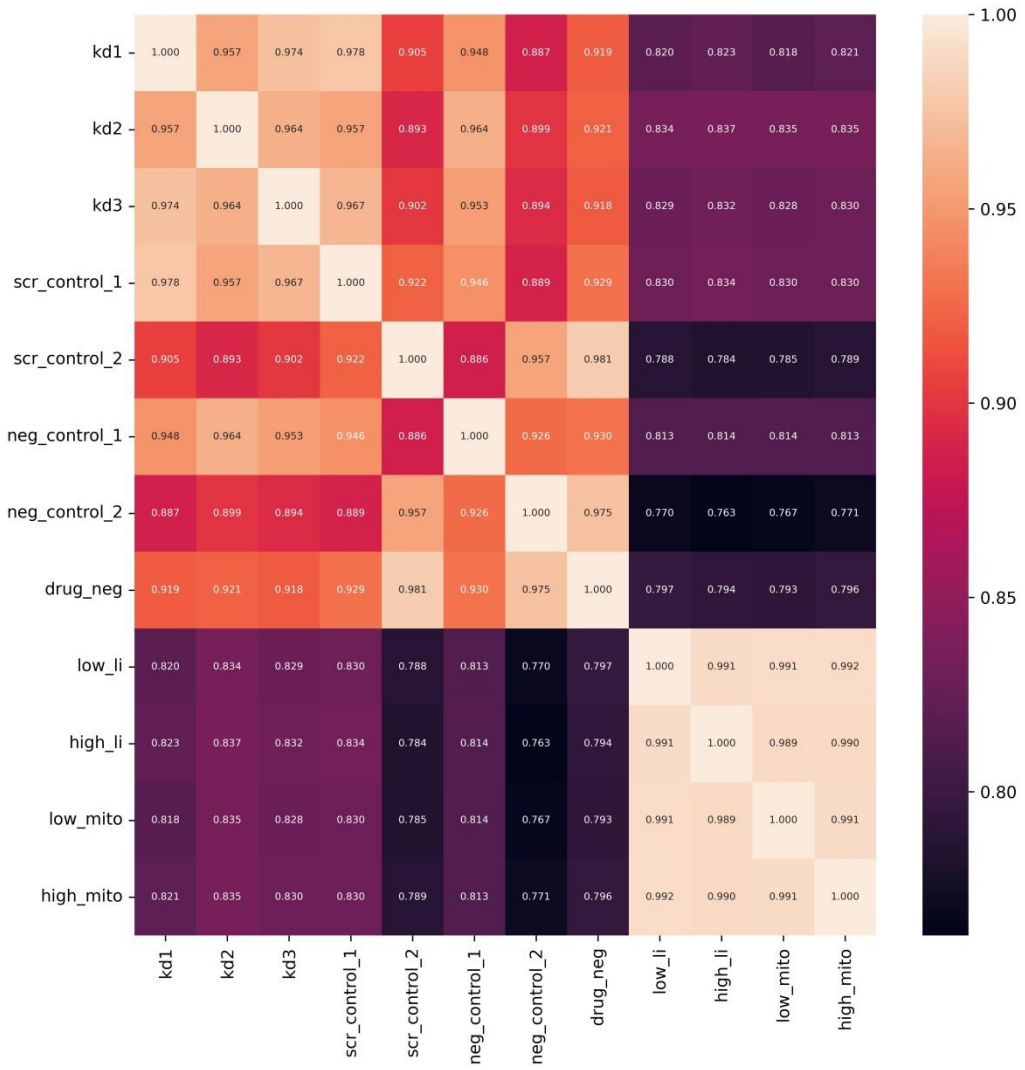


Figure 24: Heatmap showing Pearson correlation between all samples sequenced.

The scalebar on the right represents the range of the correlation coefficients displayed, with -1 meaning a total negative correlation, 0 being no correlation and +1 meaning a total positive correlation. KD1, 2, 3 = *TGFBI* KD samples brought about by shRNA1, 2, 3 respectively; scr_control = scrambled control; neg_control = empty vector negative control; drug_neg = negative control consisting of HCECs not exposed to anything; low_li and high_li = HCECs exposed to low and high concentrations of Li; low_mito and high_mito = HCECs exposed to low and high concentrations of MMC.

3.3.6.2.2 Differential Expression Analysis

The lists of genes obtained from differential expression analysis with DEseq, to identify differences in transcriptome between *TGFBI* KD HCECs and control group

HCECs; Li treated HCECs and control group HCECs and MMC treated HCECs and control group HCECs, can be found in the repository: https://github.com/gabysci/corneal_dystrophy. The total number of DE genes found in *TGFBI* KD HCECs was 2296, with a $\log_2$FC cutoff of 1.5 decreasing the number to 742. Li treated HCECs exhibited a total of 5775 DE genes, with a $\log_2$FC cutoff of 1.5 decreasing the number to 3246 while MMC treated HCECs had a total of 5572 DE genes, with a $\log_2$FC cutoff of 1.5 decreasing the number to 3195. Volcano plots displaying the DE genes from differential expression analysis of the RNAseq data obtained in the *TGFBI* KD HCEC group vs control group, Li treated HCECs vs control group and MMC treacted HCECs vs control group are shown in Figure 25, Figure 26 and Figure 27 respectively. The number of up and downregulated DE expressed genes for each of these comparisons can be found in Table 32.

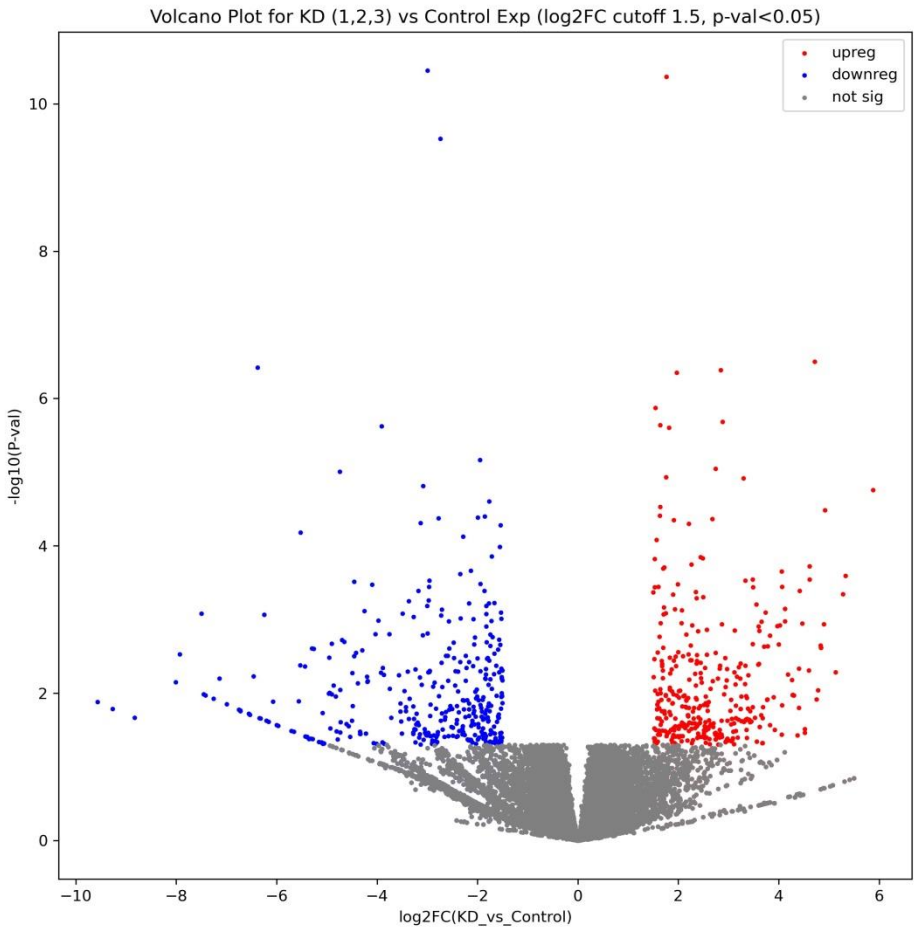


Figure 25: Volcano plot representation of DE genes from differential expression analysis of the RNAseq data obtained in the *TGFBI* KD HCEC group vs Control group

The red and blue points represent the genes with significantly increased or decreased expression respectively. The grey dots represent differentially expressed genes with a log_2FC of less than 1.5. The x-axis shows log_2fold-changes in expression and the y-axis the $–log_{10}$ (P-value) of gene expression alterations.

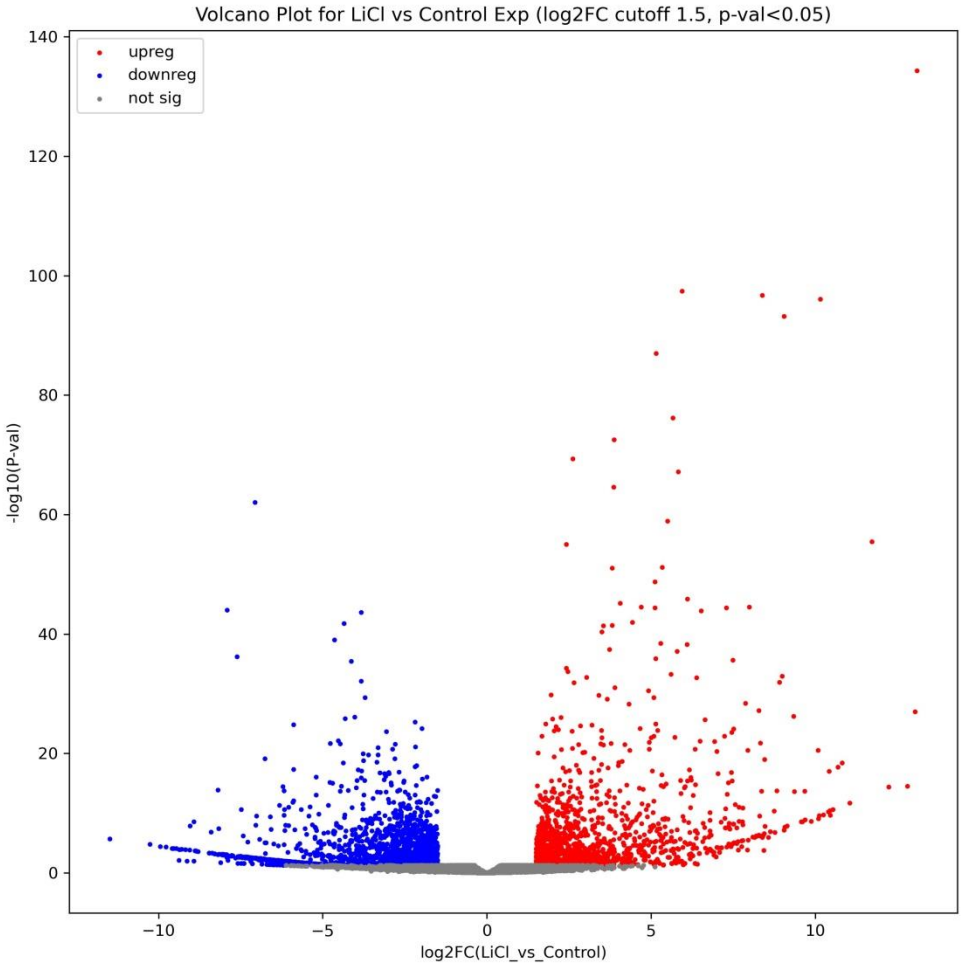


Figure 26: Volcano plot representation of DE genes from differential expression analysis of the RNAseq data obtained in the Li treated HCEC group vs Control group.

The red and blue points represent the genes with significantly increased or decreased expression respectively. The grey dots represent differentially expressed genes with a $\log_2$FC of less than 1.5. The x-axis shows $\log_2$fold-changes in expression and the y-axis the $-\log_{10}$(P-value) of gene expression alterations.

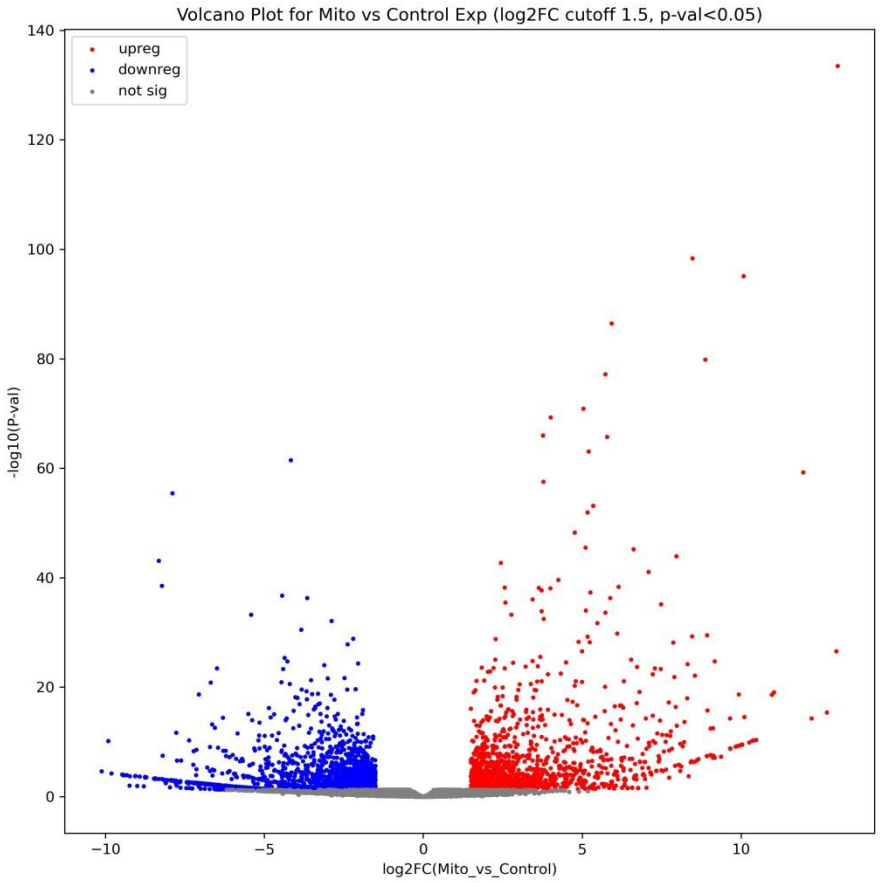


Figure 27: Volcano plot representation of DE genes from differential expression analysis of the RNAseq data obtained in the MMC treated HCEC group vs Control group.

The red and blue points represent the genes with significantly increased or decreased expression respectively. The grey dots represent differentially expressed genes with a $\log_2$FC of less than 1.5. The x-axis shows $\log_2$fold-changes in expression and the y-axis the $-\log_{10}$(P-value) of gene expression alterations.

Table 32: Number of total up and downregulated DE expressed genes (P value <=0.05) and number of DE genes with cut off of 1.5 log_2FC for *TGFBI* KD HCECs vs control HCECs; Li treated HCECs vs control group HCECs and MMC treated HCECs vs control group HCECs.

	TGFBI KD HCECs	**Li treated HCECs**	**MMC treated HCECs**
Total Downregulated DE genes	1180	2874	2820
Total Downregulated DE genes with log_2FC >= -1.5	365	1722	1710
Total Upregulated DE genes	1116	2901	2752
Total Upregulated DE genes with log_2FC >= 1.5	377	1524	1485

A total of 1180 downregulated DE genes in KD HCECs with p value of <=0.05 were identified by DE analysis. When a log_2FC cut-off of >-1.5 was implemented, the number of significantly downregulated genes went down to 365. When eliminating pseudogenes/unclassified GO annotated genes, this number was further reduced to about 152 genes. The downregulated genes that showed the largest log_2FCs were looked into (Table 33). Furthermore, this study also honed in on and analysed the 'genes of interest' known to be related to *TGFBI* (Table 34), which were shortlisted from literature and databases, as well as proteins that have been shown to be present in *TGFBI* CD deposits.

Table 33: Downregulated genes in *TGFBI* KD HCECs that showed the largest log_2FCs.

Genes with largest log_2FC Downregulated in KD	**Name**	**Log_2FoldChange**	**P value**
CD24	Cluster of differentiation 24	-9.56923	0.01317
MTPN	Myotrophin	-9.26364	0.016398
DDX39B	DEAD box helicase 39b	-8.83188	0.021397
AMMECR1L	AMME Chromosomal Region Gene 1-Like	-8.01021	0.007045
CHML	CHM Like Rab Escort Protein	-7.92606	0.002959
AREG	Amphiregulin	-7.49931	0.000831

Table 34: Genes of interest that were found to be downregulated in *TGFBI* KD HCECs.

Genes of interest Downregulated in KD	**Name**	**Log_2FoldChange**	**P value**
CPA4	Carboxypeptidase A4	-2.49358	0.037592
IL6	Interleukin 6	-1.53083	0.046589
ITGA11	Integrin subunit alpha-11	-1.34462	0.042883
S100A9	S100 calcium-binding protein A9	-0.93364	0.04691
IL1B	Interleukin 1B	-0.90163	0.01318
JAK2	Janus Kinase 2	-0.68498	0.020046
STAT2	Signal transducer and activator of transcription 2	-0.59388	0.000137

A total of 1116 upregulated DE genes in KD HCECs with p value of <=0.05 were identified by DE analysis. When a log_2FC cut-off of >-1.5 was implemented, the number of significantly upregulated genes went down to 377. Eliminating pseudogenes/unclassified GO annotated genes reduced the number further to 140 genes. The upregulated genes that showed the largest log_2FCs were looked into (Table 35). Furthermore, similarly to what was done with the upregulated genes, the downregulated 'genes of interest' known to be related to *TGFBI* (Table 36), which

were shortlisted from literature and databases, as well as proteins that have been shown to be present in *TGFBI* CD deposits, were analysed.

Table 35: Upregulated genes in *TGFBI* KD HCECs that showed the largest log_2FCs.

Genes with largest log_2FC Upregulated in KD	**Name**	**Log_2FoldChange**	**P value**
ICAM4	Intercellular adhesion molecule 4	5.881498	1.75E-05
RNA5SP439	RNA, 5S Ribosomal Pseudogene 439	5.330029	0.000255
RNA5S17	RNA, 5S Ribosomal 17	5.27924	0.000454
SNORD14C	Small Nucleolar RNA, C/D Box 14C	5.133837	0.005183
RN7SL385P	RNA, 7SL, Cytoplasmic 385	4.918127	3.28E-05
ANKRD20A10P	Ankyrin repeat domain 20	4.902211	0.001153

Table 36: Genes of interest that were found to be upregulated in *TGFBI* KD HCECs.

Genes of interest Upregulated in KD	**Name**	**Log_2FoldChange**	**P value**
MMP2	Matrix metalloproteinase 2	1.253716544	0.022905
SMAD6	Suppressor of Mothers against Decapentaplegic 6	1.23267	0.001155
COL6A2	Collagen Type VI Alpha 2 Chain	0.999639247	0.023426
COL6A1	Collagen Type VI Alpha 1 Chain	0.916271641	0.025146
AKT1	Ak strain transforming	0.583362324	0.002419
ITGA6	Integrin subunit alpha 6	0.353225934	0.02616

In this study, transcriptome analysis revealed 2874 downregulated DE genes in Li treated HCECs with p value of <=0.05, while, in MMC treated HCECs, 2820 of the DE genes were found to be downregulated. The number of downregulated DE genes with p value of <=0.05 and this time with a log_2FC cut-off of >-1.5 was 1722 for Li and 1710 for MMC. The downregulated 'genes of interest' in Li and MMC treated HCECs known to be related to *TGFBI* (Table 37 and Table 38), which were shortlisted from literature

and databases, as well as proteins that have been shown to be present in *TGFBI* CD deposits, were analysed.

Table 37: Genes of interest that were found to be downregulated in Li treated HCECs.

Genes of interest Downregulated in Li	**Name**	**Log_2FoldChange**	**P value**
S100P	S100 calcium-binding protein P	-9.37133	0.000127
COL6A2	Collagen Type VI Alpha 2 Chain	-5.83665	0.015221
APOD	Apolipoprotein D	-5.6968	0.021012
ITGA11	Integrin subunit alpha-11	-4.15566	0.006596
IL1B	Interleukin 1B	-3.77589	1.03E-20
S100A9	S100 calcium-binding protein A9	-2.97142	9.05E-10
COL6A1	Collagen Type VI Alpha 1 Chain	-2.92575	0.000034
ESR1	Estrogen Receptor 1	-1.98436	0.023359
COL5A1	Collagen Type V Alpha 1 Chain	-1.69864	4.56E-08
FLOT1	Flotillin 1	-1.55721	0.000173
JAK2	Janus Kinase 2	-1.37901	0.001603
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	-1.29478	0.015619
PTK2	Protein tyrosine kinase 2	-0.85677	0.003061
RPS27A	Ribosomal protein S27a	-0.84421	0.036271
FN1	Fibronectin-1	-0.74048	0.000593
SMAD3	Suppressor of Mothers against Decapentaplegic 3	-0.52011	0.050043

Table 38: Genes of interest that were found to be downregulated in MMC treated HCECs.

Genes of interest Downregulated in MMC	**Name**	**Log_2FoldChange**	**P value**
S100P	S100 calcium-binding protein P	-9.23832	0.000168
COL6A2	Collagen Type VI Alpha 2 Chain	-5.70476	0.018225
APOD	Apolipoprotein D	-5.56411	0.024756
IL1B	Interleukin 1B	-3.83393	2.52E-20
ESR1	Estrogen Receptor 1	-3.1056	0.003177
S100A9	S100 calcium-binding protein A9	-2.9592	4.63E-09
COL6A1	Collagen Type VI Alpha 1 Chain	-2.461	0.000154
FLOT1	Flotillin 1	-1.82905	2.62E-05
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	-1.41162	0.022784
RPS27A	Ribosomal protein S27a	-0.85708	0.030828
PTK2	Protein tyrosine kinase 2	-0.82236	0.007425
FN1	Fibronectin-1	-0.71306	0.001124
SMAD3	Suppressor of Mothers against Decapentaplegic 3	-0.55262	0.044621

In this study, transcriptome analysis revealed 2901 upregulated DE genes in Li treated HCECs with p value of <=0.05, while, in MMC treated HCECs, 2752 of the DE genes were found to be upregulated. The number of upregulated DE genes with p value of <=0.05 and this time with a log_2FC cut-off of >1.5 was 1524 for Li and 1485 for MMC. The upregulated 'genes of interest' in Li and MMC treated HCECs known to be related to *TGFBI* (Table 39 and Table 40), which were shortlisted from literature and databases, as well as proteins that have been shown to be present in *TGFBI* CD deposits, were analysed.

Table 39: Genes of interest that were found to be upregulated in Li treated HCECs.

Genes of interest Upregulated in Li	**Name**	**Log_2Fold Change**	**P value**
CPA4	Carboxypeptidase A4	7.351172	8.01E-16
ALDH3A1	Aldehyde dehydrogenase 3A1	6.038055	3.59E-15
COL1A2	Collagen Type I Alpha 2 Chain	6.023159	7.51E-06
VIM	Vimentin	5.946396	3.78E-98
S100A4	S100A4 S100 calcium binding pr A4	4.943884	2E-21
THBS1	Thrombospondin 1	4.105671	1.01E-10
COL12A1	Collagen Type XII Alpha 1 Chain	3.813417	9.08E-52
MMP2	Matrix metalloproteinase 2	3.386495	6.24E-10
NOG	Noggin	3.105164	0.001845
CPA2	Carboxypeptidase A2	2.836112	0.025952
MAMDC2	MAM Domain Containing 2	2.793728	6.34E-07
TGFBI	Transforming growth factor beta induced	2.538253	9.19E-08
COL4A1	Collagen Type IV Alpha 1 Chain	2.512776	7.4E-18
FABP5	Fatty acid-binding protein 5	2.506098	6E-07
COL1A1	Collagen Type I Alpha 1 Chain	2.4748	5.67E-11
PLOD2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	2.230431	0.000514
SMAD7	Suppressor of Mothers against Decapentaplegic 7	2.12018	1.88E-06
COL4A2	Collagen Type IV Alpha 2 Chain	2.101886	3.01E-25
LDHA	Lactate dehydrogenase A	2.041573	1.71E-24
FAT4	FAT Atypical Cadherin 4	1.988777	0.001314
ITGA6	Integrin subunit alpha 6	1.94668	1.6E-30
ACTB	Actin beta	1.884444	1.23E-09
COL5A2	Collagen Type V Alpha 2 Chain	1.817187	5.57E-07
CLU	Clusterin (Apolipoprotein J)	1.807268	0.00026
DSP	Desmoplakin	1.668326	1.24E-23
ENO1	α-Enolase	1.596114	3.52E-12
CSTA	Cystatin A	1.579822	1.58E-07
ESR2	Estrogen receptor 2	1.336916	0.008383
ITGA3	Integrin subunit alpha-3	1.313993	4.94E-07
ITGB5	Integrin subunit beta 5	1.289922	8.85E-07
LOXL2	Lysyl oxidase-like 2	1.152428	0.02485
YAP1	Yes1 Ass. Transcriptional Regulator	1.102519	3.88E-09
ITGAV	Integrin Subunit Alpha V	1.038009	4.46E-06
ANXA2	Annexin A2	0.767988	2.92E-06
PRDX1	Peroxiredoxin 1	0.722669	0.002207
CUL4A	Cullin 4A	0.713854	0.005849
S100A6	S100 calcium binding protein A6	0.538309	0.001414
MAPK1	Mitogen-activated protein kinase 1	0.529024	0.015095

Table 40: Genes of interest that were found to be upregulated in MMC treated HCECs.

Genes of interest Upregulated in MMC	**Name**	**Log2FoldChange**	**P value**
COL1A2	Collagen Type I Alpha 2 Chain	6.494055	1.11E-07
VIM	Vimentin	5.924655	3.33E-87
S100A4	S100A4 S100 calcium binding protein A4	4.537373	1.99E-18
TNF	Tumor necrosis factor	4.145475	0.029959
THBS1	Thrombospondin 1	4.130674	8.11E-11
COL12A1	Collagen Type XII Alpha 1 Chain	4.011143	4.99E-70
MMP2	Matrix metalloproteinase 2	3.327616	2.76E-09
FABP5	Fatty acid-binding protein 5	2.812885	6.03E-09
NOG	Noggin	2.716764	0.015092
COL1A1	Collagen Type I Alpha 1 Chain	2.569094	1.26E-11
COL4A4	Collagen Type IV Alpha 4 Chain	2.543208	6.54E-05
COL4A1	Collagen Type IV Alpha 1 Chain	2.49293	1.05E-20
SMAD7	Suppressor of Mothers against Decapentaplegic 7	2.467148	2.78E-08
TGFBI	Transforming growth factor beta induced	2.458737	2.19E-07
PLOD2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	2.378331	0.000309
COL4A2	Collagen Type IV Alpha 2 Chain	2.097939	1.28E-23
LDHA	Lactate dehydrogenase A	2.036425	1.44E-23
CLU	Clusterin (Apolipoprotein J)	1.962546	4.45E-05
CSTA	Cystatin A	1.932268	1.79E-09
COL5A2	Collagen Type V Alpha 2 Chain	1.884631	4.49E-08
ACTB	Actin beta	1.851494	1.71E-09
ITGA6	Integrin subunit alpha 6	1.842386	2.46E-24
DSP	Desmoplakin	1.677993	6.74E-22
ESR2	Estrogen receptor 2	1.591085	0.001341
ENO1	α-Enolase	1.537326	5.59E-11
ITGA3	Integrin subunit alpha-3	1.235537	1.99E-06
ITGB5	Integrin subunit beta 5	1.17089	1.36E-05
LOXL2	Lysyl oxidase-like 2	1.142773	0.022863
ITGAV	Integrin Subunit Alpha V	1.032147	8.7E-06
ANXA2	Annexin A2	0.687423	8.27E-05
ITGA5	Integrin subunit alpha 5	0.663986	0.030142
CUL4A	Cullin 4A	0.545363	0.032551
S100A6	S100 calcium binding protein A6	0.396087	0.024294

The lists of DE genes in Li treated and MMC treated HCECs obtained by bioinformatic analysis were further filtered to identify the ones that were significantly differentially expressed (with cut off of log_2FC 1.5) in one treatment group but not the other. The numbers of down and upregulated DE genes that were found to be present only in Li and not in MMC were 347 and 500, respectively. Whereas, the numbers of down and upregulated DE genes that were found to be present only in MMC and not in Li were 336 and 461, respectively. This would help in identifying which genes and pathways are specific to each treatment group.

3.3.6.2.3 Enrichment Analysis

ORA analysis was performed using the WebGestalt functional enrichment tool for GO annotation of biological and molecular function as well as KEGG and Reactome pathways for upregulated and downregulated DE genes identified in *TGFBI* KD HCECs vs. control, Li vs control and MMC vs control. A summary of the results can be found in the figures (Figures: 28, 29, 30, 31, 32, 33) and tables below (Tables: 41, 42, 43, 44, 45, 46,), where the first 25 descriptions (P<=0.05) for each set are listed. The full set of results including the GO annotation or pathway, gene overlap, p value and gene IDs for each of the described processes can be found in the repository: https://github.com/gabysci/corneal_dystrophy.

1. **KD vs Control: Downregulated genes**

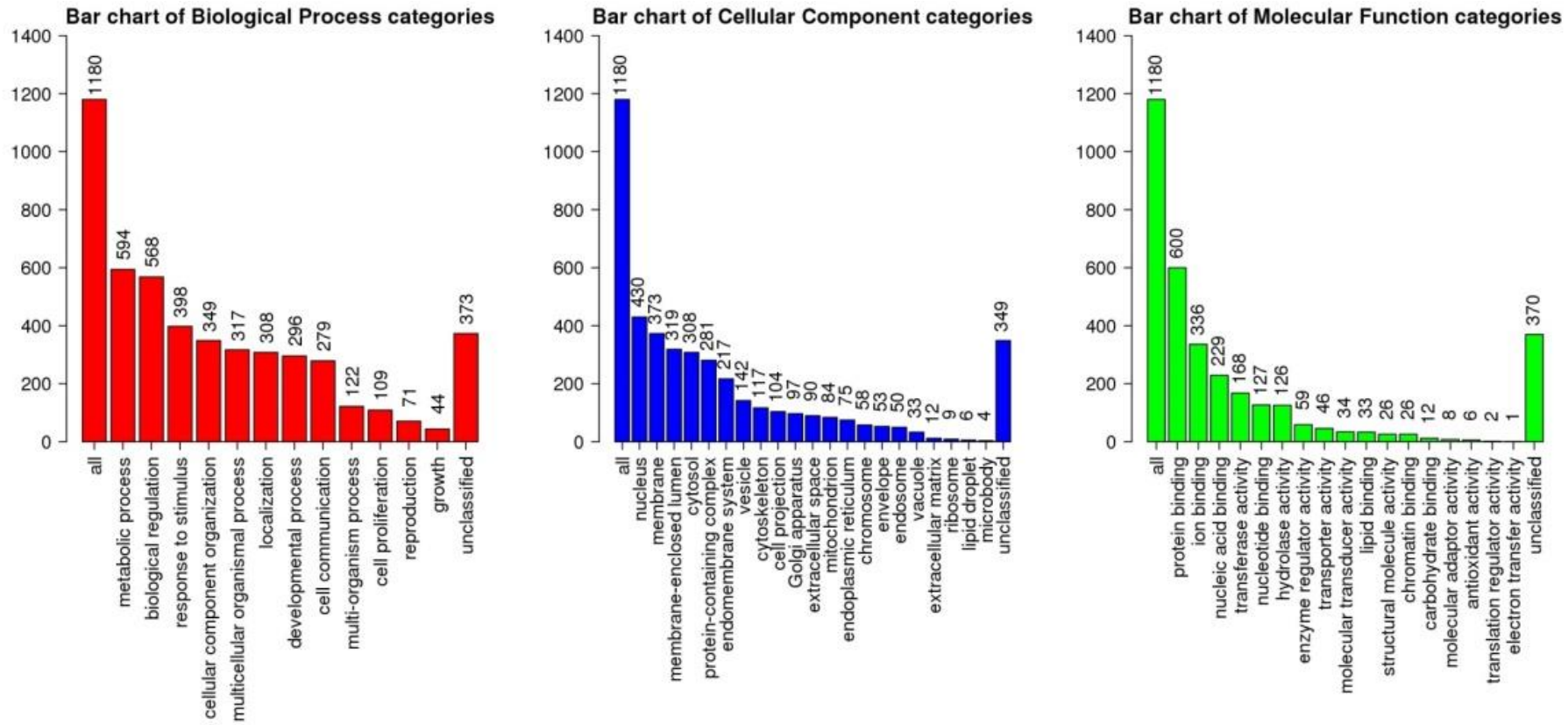


Figure 28: Bar charts displaying enrichment results of GO annotations of downregulated genes identified from differential expression analysis of the RNAseq data in KD vs control groups.

Table 41: GO annotations and pathway associations of downregulated genes identified from differential expression analysis of the RNAseq data in KD vs control groups.

GO Biological Process	GO Molecular Function	KEGG Pathway	Reactome Pathway
cellular response to stress	transferase activity, transferring phosphorus-containing groups	Fanconi anaemia pathway	Regulation of TP53 Activity
regulation of signal transduction by p53 class mediator	histone demethylase activity	Autophagy	Regulation of TP53 Activity through Phosphorylation
regulation of response to DNA damage stimulus	modification-dependent protein binding	Endocytosis	Gene expression (Transcription)
cellular response to DNA damage stimulus	kinase activity	p53 signalling pathway	RNA polymerase II transcribes snRNA genes
cell cycle	phosphotransferase activity, alcohol group as acceptor	RNA transport	RNA Polymerase II Transcription
histone demethylation	ATP binding	Basal transcription factors	PI Metabolism
protein demethylation	purine ribonucleoside triphosphate binding	NOD-like receptor signalling pathway	Late Phase of HIV Life Cycle
protein dealkylation	adenyl nucleotide binding	Selenocompound metabolism	Transcriptional Regulation by TP53
histone lysine demethylation	adenyl ribonucleotide binding	FoxO signalling pathway	Phospholipid metabolism

non-motile cilium assembly	purine nucleotide binding	Mitophagy	HIV Life Cycle
peptidyl-serine phosphorylation	ribonucleotide binding	IL-17 signalling pathway	Synthesis of PIPs at the late endosome membrane
regulation of cell cycle	K63-linked polyubiquitin modification-dependent protein binding	NF-kappa B signalling pathway	Metabolism of RNA
DNA repair	purine ribonucleotide binding	Cell cycle	HDR through Single Strand Annealing (SSA)
regulation of DNA damage response, signal transduction by p53 class mediator	protein serine/threonine kinase activity	Glycerophospholipid metabolism	Presynaptic phase of homologous DNA pairing and strand exchange
signal transduction by p53 class mediator	demethylase activity	Phosphatidylinositol signalling system	Fanconi Anaemia Pathway
snRNA transcription by RNA polymerase II	drug binding	Necroptosis	SUMOylation of DNA damage response and repair proteins
snRNA transcription	carbon-sulfur lyase activity	Platinum drug resistance	Homologous DNA Pairing and Strand Exchange
regulation of cellular component biogenesis	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	Inositol phosphate metabolism	Synthesis of PIPs at the plasma membrane
histone H3-K4 demethylation	histone demethylase activity (H3-K4 specific)	Amino sugar and nucleotide sugar metabolism	Infectious disease

organelle assembly	phosphatidylinositol bisphosphate phosphatase activity	mTOR signalling pathway	Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer
cellular response to external stimulus	protein kinase activity	TNF signalling pathway	HDR through Homologous Recombination (HRR)
cellular response to starvation	polyubiquitin modification-dependent protein binding	Phototransduction	Generic Transcription Pathway
Golgi organization	alpha-mannosidase activity	Tuberculosis	Negative regulators of DDX58/IFIH1 signalling
cellular response to nutrient levels	phosphatidylinositol phosphate phosphatase activity	Legionellosis	Interleukin-20 family signalling
peptidyl-serine modification	mannosidase activity	Apoptosis	Regulation of TP53 Degradation

2. KD vs Control: Upregulated genes

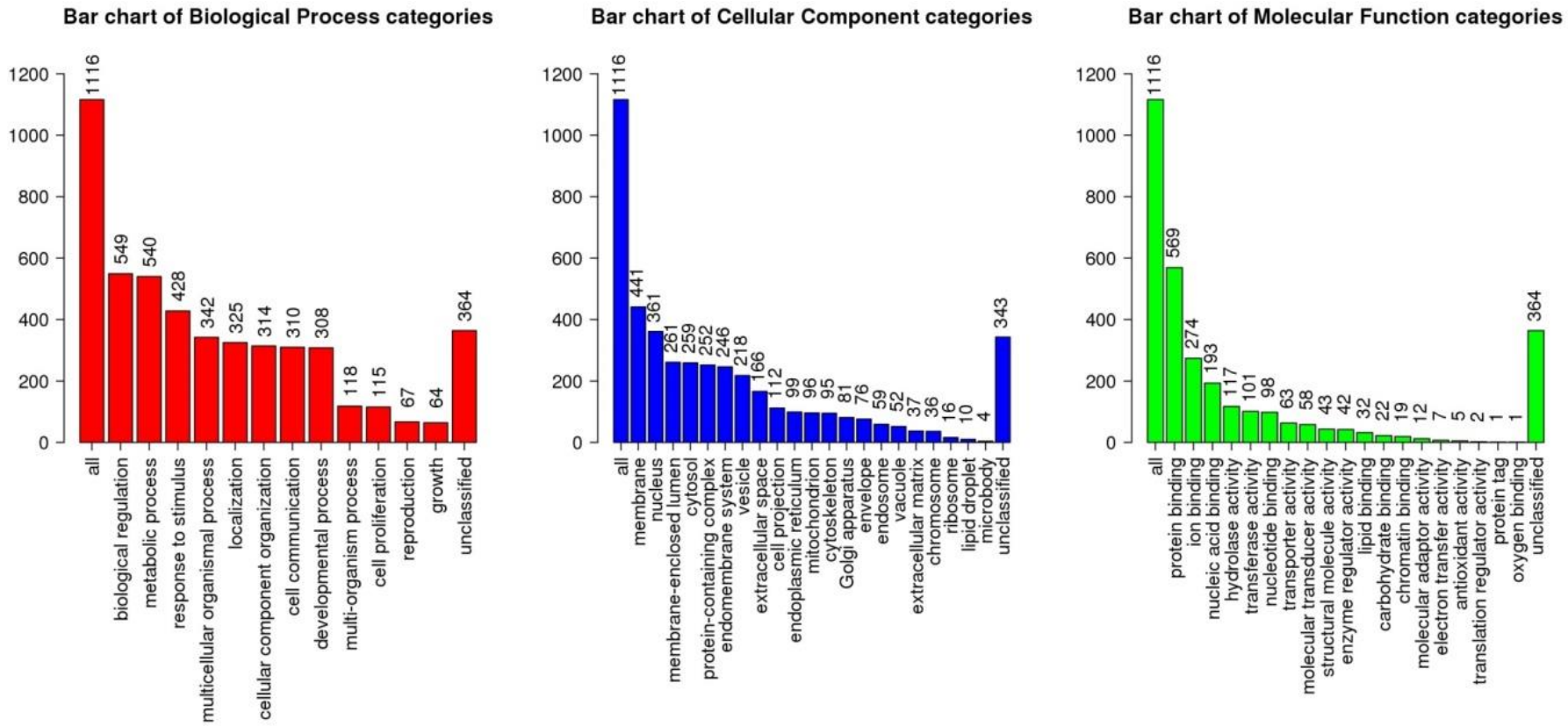


Figure 29: Bar charts displaying enrichment results of GO annotations of upregulated genes identified from differential expression analysis of the RNAseq data in KD vs control groups.

Table 42: GO annotations and pathway associations of upregulated genes identified from differential expression analysis of the RNAseq data in KD vs control groups.

GO Biological Process	**GO Molecular Function**	**KEGG Pathway**	**Reactome Pathway**
regulation of apoptotic signalling pathway	nucleobase-containing compound transmembrane transporter activity	Alzheimer disease	Intrinsic Pathway for Apoptosis
positive regulation of apoptotic signalling pathway	carbohydrate derivative transmembrane transporter activity	Thermogenesis	Apoptosis
mitochondrial membrane organization	2-acylglycerol-3-phosphate O-acyltransferase activity	Oxidative phosphorylation	Programmed Cell Death
apoptotic signalling pathway	identical protein binding	Huntington disease	Metabolism of nitric oxide
regulation of protein localization to membrane	tau protein binding	Parkinson disease	eNOS activation and regulation
exocytosis	DNA polymerase binding	Vasopressin-regulated water reabsorption	Tetrahydrobiopterin (BH4) synthesis, recycling, salvage and regulation
developmental growth	transcription regulatory region DNA binding	Non-alcoholic fatty liver disease (NAFLD)	eNOS activation
regulation of binding	regulatory region nucleic acid binding	Human papillomavirus infection	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling

			proteins.
activation of cysteine-type endopeptidase activity	MHC protein complex binding	Folate biosynthesis	Activation of BH3-only proteins
positive regulation of interleukin-8 production	DNA-binding transcription activator activity, RNA polymerase II-specific	Gap junction	Activation of BAD and translocation to mitochondria
positive regulation of cell communication	1-acylglycerol-3-phosphate O-acyltransferase activity	Axon guidance	Neutrophil degranulation
cell activation	lysophosphatidic acid acyltransferase activity	Tuberculosis	Translocation of SLC2A4 (GLUT4) to the plasma membrane
positive regulation of signalling	lysophospholipid acyltransferase activity	Glucagon signalling pathway	The citric acid (TCA) cycle and respiratory electron transport
secretion	transferase activity, transferring acyl groups	Focal adhesion	Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters
animal organ morphogenesis	transferase activity, transferring acyl groups other than amino-acyl groups	Glycosphingolipid biosynthesis	Metabolism of cofactors
myeloid leukocyte mediated immunity	proton transmembrane transporter activity	Cortisol synthesis and secretion	EPH-ephrin mediated repulsion of cells

ossification	carbohydrate binding	Metabolic pathways	Cristae formation
apoptotic mitochondrial changes	transcription coactivator activity	Fluid shear stress and atherosclerosis	Transport of nucleosides and free purine and pyrimidine bases across the plasma membrane
cellular response to external stimulus	RNA polymerase II regulatory region sequence-specific DNA binding	Renin secretion	Deregulated CDK5 triggers multiple neurodegenerative pathways in Alzheimer's disease models
positive regulation of cell death	myosin binding	Central carbon metabolism in cancer	Neurodegenerative Diseases
regulated exocytosis	RNA polymerase II regulatory region DNA binding	Acute myeloid leukaemia	Disease
leukocyte degranulation	nucleoside transmembrane transporter activity	ECM-receptor interaction	Other interleukin signalling
neutrophil mediated immunity	transcription regulatory region sequence-specific DNA binding	Hepatitis B	Nef Mediated CD4 Down-regulation
carbohydrate derivative transport	sequence-specific double-stranded DNA binding	Pathways in cancer	Detoxification of Reactive Oxygen Species
vesicle-mediated transport	MHC class II protein complex binding	Glutathione metabolism	COPI-mediated anterograde transport

3. Li vs Control: downregulated genes

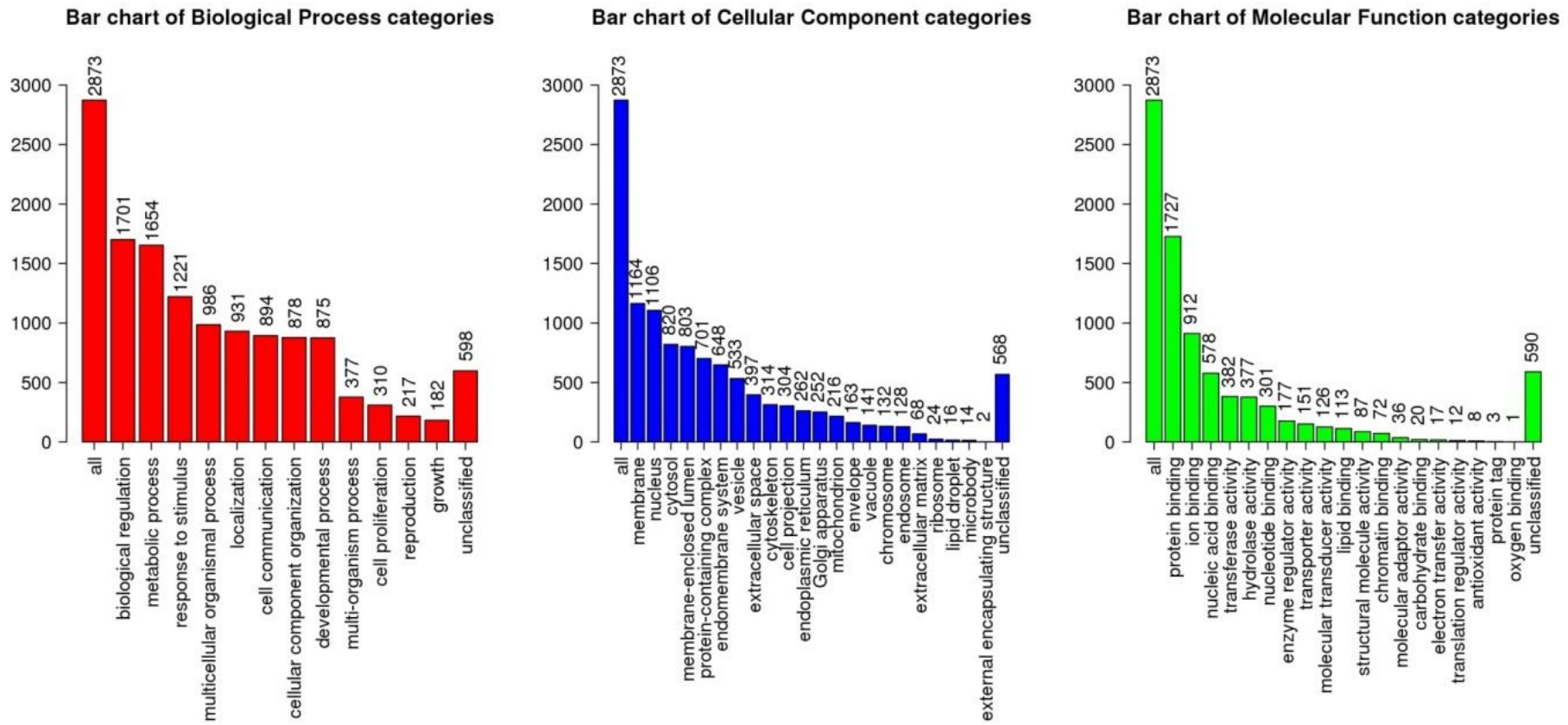


Figure 30: Bar charts displaying enrichment results of GO annotations of downregulated genes identified from differential expression analysis of the RNAseq data in Li vs control groups.

Table 43: GO annotations and pathway associations of downregulated genes identified from differential expression analysis of the RNAseq data in Li vs control groups

GO Biological Process	**GO Molecular Function**	**KEGG Pathway**	**Reactome Pathway**
regulation of intracellular signal transduction	DNA-binding transcription factor activity	NOD-like receptor signalling pathway	Cytokine Signalling in Immune system
positive regulation of molecular function	DNA-binding transcription factor activity, RNA polymerase II-specific	Autophagy	Death Receptor Signalling
positive regulation of RNA biosynthetic process	enzyme activator activity	Mitophagy	Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signalling pathways
positive regulation of nucleic acid-templated transcription	RNA polymerase II regulatory region sequence-specific DNA binding	Hepatitis C	Immune System
response to cytokine	RNA polymerase II regulatory region DNA binding	TNF signalling pathway	NOD1/2 Signalling Pathway
regulation of response to stress	sequence-specific DNA binding	NF-kappa B signalling pathway	Interferon alpha/beta signalling
positive regulation of RNA metabolic process	transcription regulatory region sequence-specific DNA binding	Human papillomavirus infection	Interferon Signalling
positive regulation of transcription, DNA-templated	sequence-specific double-stranded DNA binding	FoxO signalling pathway	Cell death signalling via NRAGE, NRIF and NADE
cellular response to cytokine	transcription regulatory region DNA	Inositol phosphate metabolism	Signalling by Interleukins

stimulus	binding		
positive regulation of gene expression	regulatory region nucleic acid binding	Measles	p75 NTR receptor-mediated signalling
positive regulation of cellular biosynthetic process	kinase activity	Fluid shear stress and atherosclerosis	Rho GTPase cycle
cytokine-mediated signalling pathway	identical protein binding	AGE-RAGE signalling pathway in diabetic complications	Macroautophagy
positive regulation of catalytic activity	phosphotransferase activity, alcohol group as acceptor	Necroptosis	Interleukin-4 and Interleukin-13 signalling
positive regulation of biosynthetic process	RNA polymerase II proximal promoter sequence-specific DNA binding	Endocytosis	Endosomal/Vacuolar pathway
positive regulation of nucleobase-containing compound metabolic process	double-stranded DNA binding	Prolactin signalling pathway	NRAGE signals death through JNK
regulation of small GTPase mediated signal transduction	protein domain specific binding	C-type lectin receptor signalling pathway	Interleukin-20 family signalling
positive regulation of macromolecule biosynthetic process	kinase binding	Tight junction	Phospholipid metabolism

negative regulation of response to stimulus	GTPase activator activity	Small cell lung cancer	Negative regulators of DDX58/IFIH1 signalling
regulation of anatomical structure morphogenesis	Rho guanyl-nucleotide exchange factor activity	Osteoclast differentiation	TNF signalling
positive regulation of transcription by RNA polymerase II	enzyme regulator activity	Influenza A	NrCAM interactions
positive regulation of cell differentiation	proximal promoter sequence-specific DNA binding	Cellular senescence	Regulation of TNFR1 signalling
positive regulation of signal transduction	SH3 domain binding	Human immunodeficiency virus 1 infection	Interleukin-1 family signalling
positive regulation of signalling	phosphatase activity	Chagas disease (American trypanosomiasis)	TNFR1-induced NFkappaB signalling pathway
positive regulation of cell communication	protein kinase binding	Pathways in cancer	DDX58/IFIH1-mediated induction of interferon-alpha/beta
nephron tubule development	protein serine/threonine kinase activity	Choline metabolism in cancer	Signalling by cytosolic FGFR1 fusion mutants

4. Li vs Control: upregulated genes

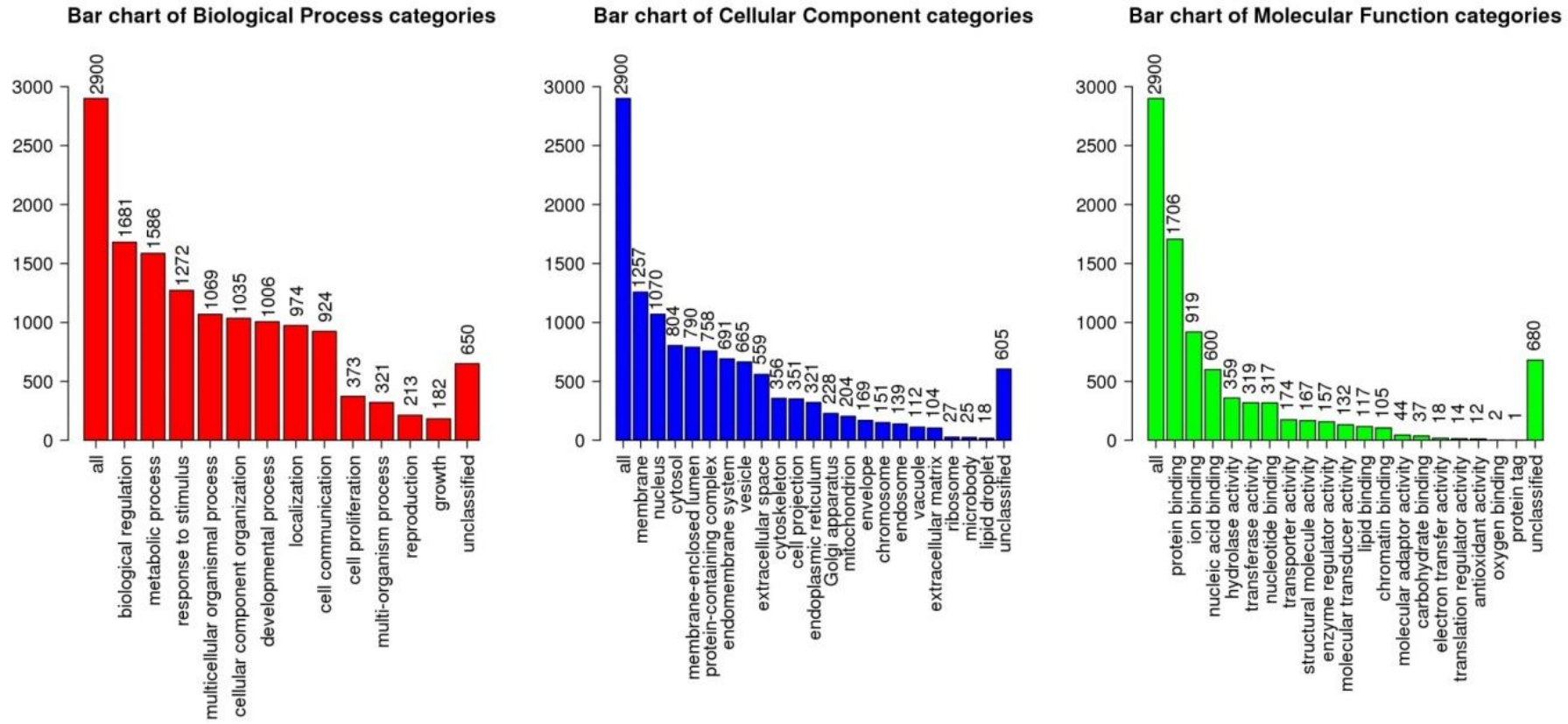


Figure 31: Bar charts displaying enrichment results of GO annotations of upregulated genes identified from differential expression analysis of the RNAseq data in Li vs control.

Table 44: GO annotations and pathway associations of upregulated genes identified from differential expression analysis of the RNAseq data in Li vs control groups.

GO Biological Process	**GO Molecular Function**	**KEGG Pathway**	**Reactome Pathway**
movement of cell or subcellular component	cell adhesion molecule binding	Regulation of actin cytoskeleton	Collagen formation
tissue development	cadherin binding	Focal adhesion	EPH-Ephrin signalling
locomotion	protein-containing complex binding	Axon guidance	Cholesterol biosynthesis
response to endogenous stimulus	structural molecule activity	Pathogenic Escherichia coli infection	Axon guidance
neurogenesis	cytoskeletal protein binding	Bacterial invasion of epithelial cells	Extracellular matrix organization
cell motility	kinase binding	Central carbon metabolism in cancer	Type I hemidesmosome assembly
localization of cell	protein domain specific binding	Fluid shear stress and atherosclerosis	Signalling by VEGF
cell migration	structural constituent of cytoskeleton	Endocytosis	Syndecan interactions
cellular response to endogenous stimulus	protein kinase binding	Adherens junction	Assembly of collagen fibrils and other multimeric structures
neuron projection development	RNA binding	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Signalling by Receptor Tyrosine Kinases
cell morphogenesis involved in	actin binding	Salmonella infection	RHO GTPases activate IQGAPs

differentiation			
plasma membrane bounded cell projection organization	identical protein binding	ECM-receptor interaction	Recycling pathway of L1
cell morphogenesis	signalling receptor binding	Biosynthesis of unsaturated fatty acids	VEGFA-VEGFR2 Pathway
cell junction organization	oxidoreductase activity	Rap1 signalling pathway	Collagen biosynthesis and modifying enzymes
neuron development	chromatin binding	Pathways in cancer	Non-integrin membrane-ECM interactions
generation of neurons	disordered domain specific binding	AGE-RAGE signalling pathway in diabetic complications	Cell-extracellular matrix interactions
cell projection organization	SMAD binding	Cellular senescence	Cell junction organization
cytoskeleton organization	calcium ion binding	Relaxin signalling pathway	Developmental Biology
neuron differentiation	integrin binding	Estrogen signalling pathway	Gap junction trafficking
cellular component morphogenesis	transition metal ion binding	Proteoglycans in cancer	EPHB-mediated forward signalling
regulation of cellular component movement	cell-cell adhesion mediator activity	Terpenoid backbone biosynthesis	L1CAM interactions
cell projection morphogenesis	cell adhesion mediator activity	Shigellosis	EPH-ephrin mediated repulsion of

			cells
cell part morphogenesis	scaffold protein binding	Endocrine and other factor-regulated calcium reabsorption	Gap junction trafficking and regulation
cell proliferation	laminin binding	Glutathione metabolism	Anchoring fibril formation
plasma membrane bounded cell projection morphogenesis	intermediate filament binding	Steroid biosynthesis	RHO GTPases Activate WASPs and WAVEs

5. MMC vs Control: downregulated genes

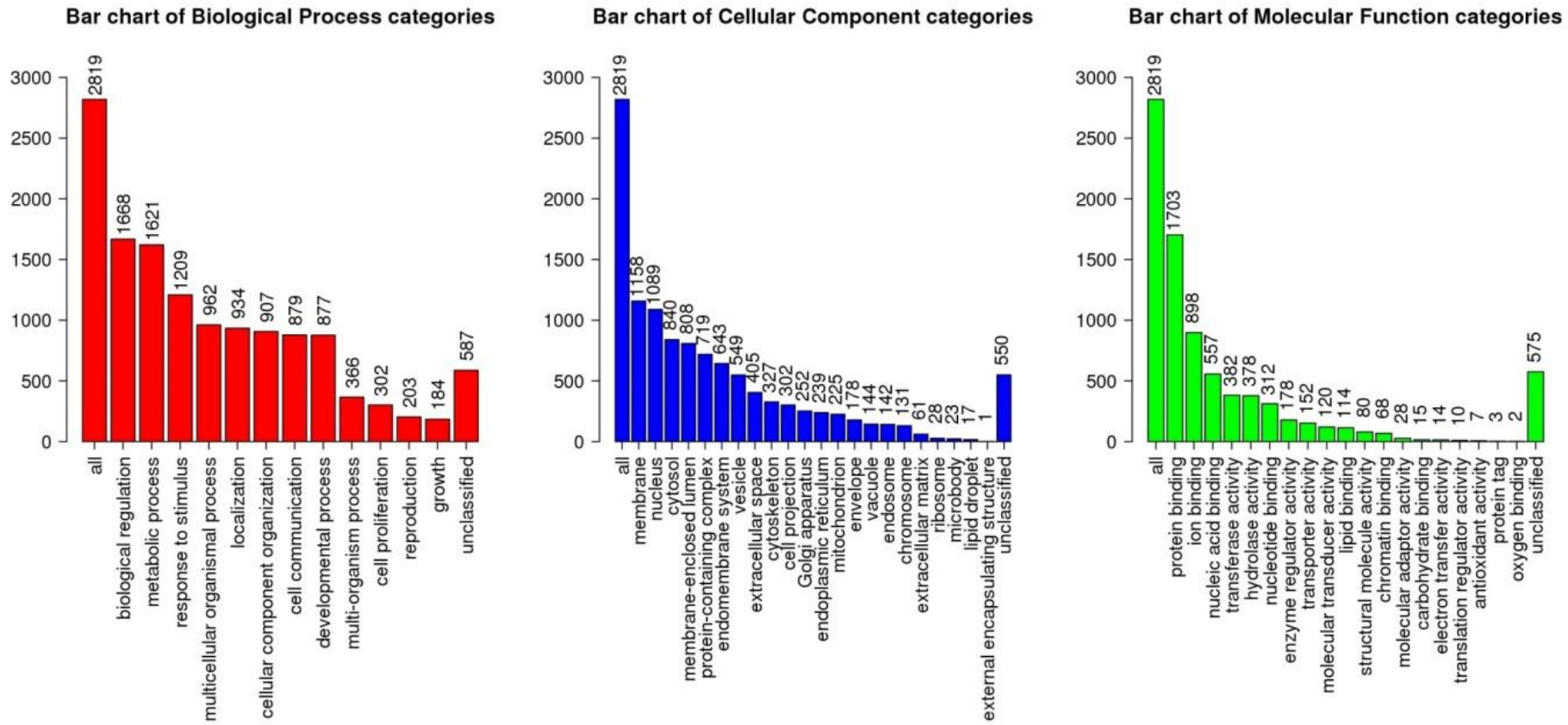


Figure 32: Bar charts displaying enrichment results of GO annotations of downregulated genes identified from differential expression analysis of the RNAseq data in MMC vs control groups

Table 45: GO annotations and pathway associations of downregulated genes identified from differential expression analysis of the RNAseq data in MMC vs control groups.

GO Biological Process	**GO Molecular Function**	**KEGG Pathway**	**Reactome Pathway**
positive regulation of molecular function	enzyme activator activity	NOD-like receptor signalling pathway	Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signalling pathways
regulation of intracellular signal transduction	kinase activity	Autophagy	Cytokine Signalling in Immune system
response to cytokine	phosphotransferase activity, alcohol group as acceptor	TNF signalling pathway	Immune System
cellular response to cytokine stimulus	DNA-binding transcription factor activity	Mitophagy	Signalling by Interleukins
positive regulation of catalytic activity	kinase binding	NF-kappa B signalling pathway	NOD1/2 Signalling Pathway
regulation of response to stress	protein serine/threonine kinase activity	FoxO signalling pathway	Interferon alpha/beta signalling
cytokine-mediated signalling pathway	transferase activity, transferring phosphorus-containing groups	Measles	Death Receptor Signalling
regulation of organelle organization	protein kinase activity	Apoptosis	Macroautophagy
cellular response to stress	protein kinase binding	Fluid shear stress and atherosclerosis	Interleukin-4 and Interleukin-13

			signalling
positive regulation of gene expression	DNA-binding transcription factor activity, RNA polymerase II-specific	Human immunodeficiency virus 1 infection	Innate Immune System
positive regulation of protein metabolic process	enzyme regulator activity	Apoptosis	Ovarian tumor domain proteases
positive regulation of RNA metabolic process	identical protein binding	Neurotrophin signalling pathway	Interferon Signalling
positive regulation of cellular biosynthetic process	SH3 domain binding	Hepatitis C	Interleukin-1 family signalling
positive regulation of cellular protein metabolic process	nuclear hormone receptor binding	Necroptosis	Interleukin-15 signalling
positive regulation of RNA biosynthetic process	protein domain specific binding	Pathways in cancer	Endosomal/Vacuolar pathway
growth	RNA polymerase II regulatory region DNA binding	Pertussis	Interleukin-20 family signalling
positive regulation of nucleic acid-templated transcription	RNA polymerase II regulatory region sequence-specific DNA binding	Small cell lung cancer	Negative regulators of DDX58/IFIH1 signalling
positive regulation of biosynthetic process	molecular function regulator	Legionellosis	Cell death signalling via NRAGE, NRIF and NADE

positive regulation of transcription, DNA-templated	sequence-specific DNA binding	Cellular senescence	DDX58/IFIH1-mediated induction of interferon-alpha/beta
developmental growth	transcription regulatory region DNA binding	Endocytosis	Regulation of TNFR1 signalling
protein phosphorylation	ubiquitin-protein transferase activity	Epstein-Barr virus infection	p75 NTR receptor-mediated signalling
positive regulation of nucleobase-containing compound metabolic process	regulatory region nucleic acid binding	Inositol phosphate metabolism	TNF signalling
macroautophagy	small GTPase binding	Tight junction	Rho GTPase cycle
positive regulation of proteolysis	adenyl ribonucleotide binding	Prolactin signalling pathway	RNA polymerase II transcribes snRNA genes
positive regulation of DNA-binding transcription factor activity	GTPase binding	AGE-RAGE signalling pathway in diabetic complications	Class I MHC mediated antigen processing & presentation

6. MMC vs Control: upregulated genes

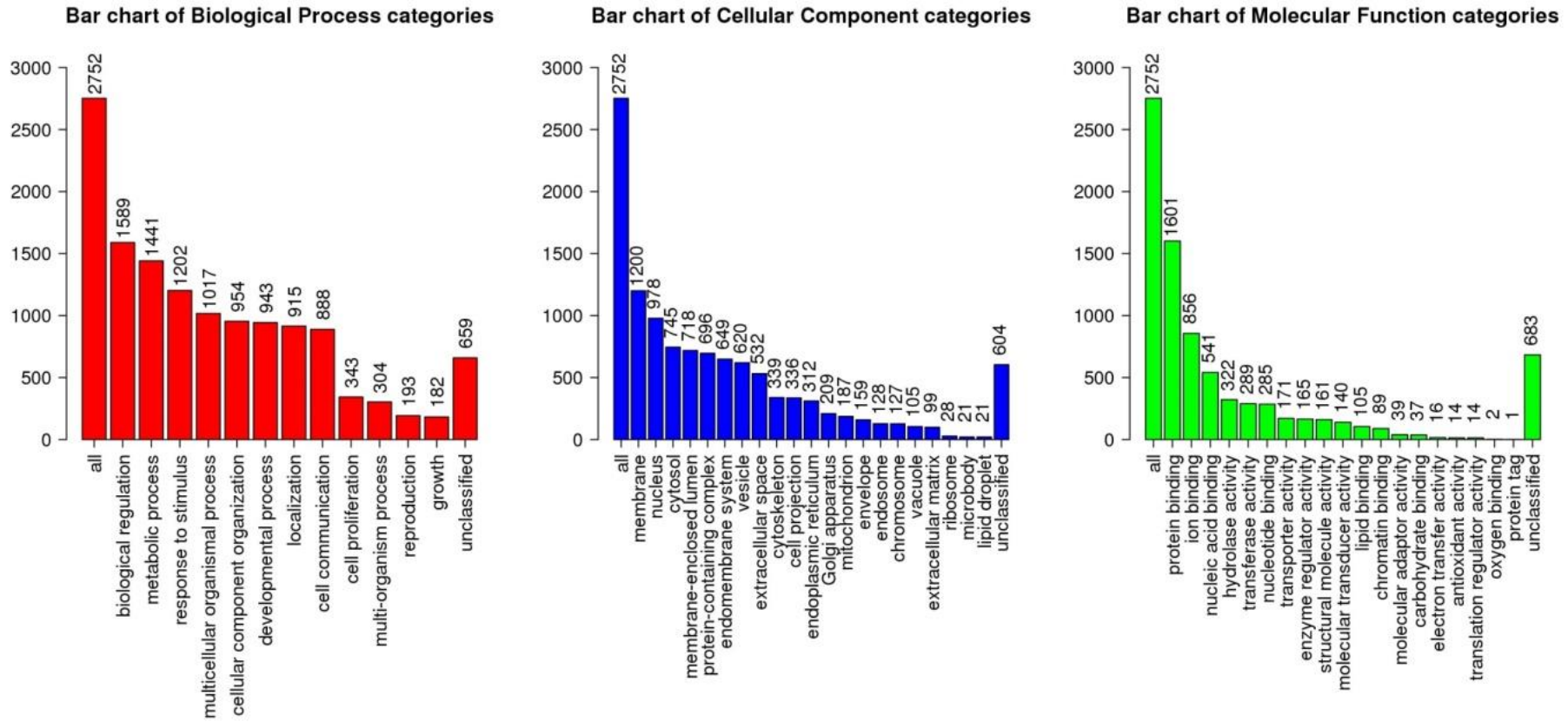


Figure 33: Bar charts displaying enrichment results of GO annotations of upregulated genes identified from differential expression analysis of the RNAseq data in MMC vs control groups.

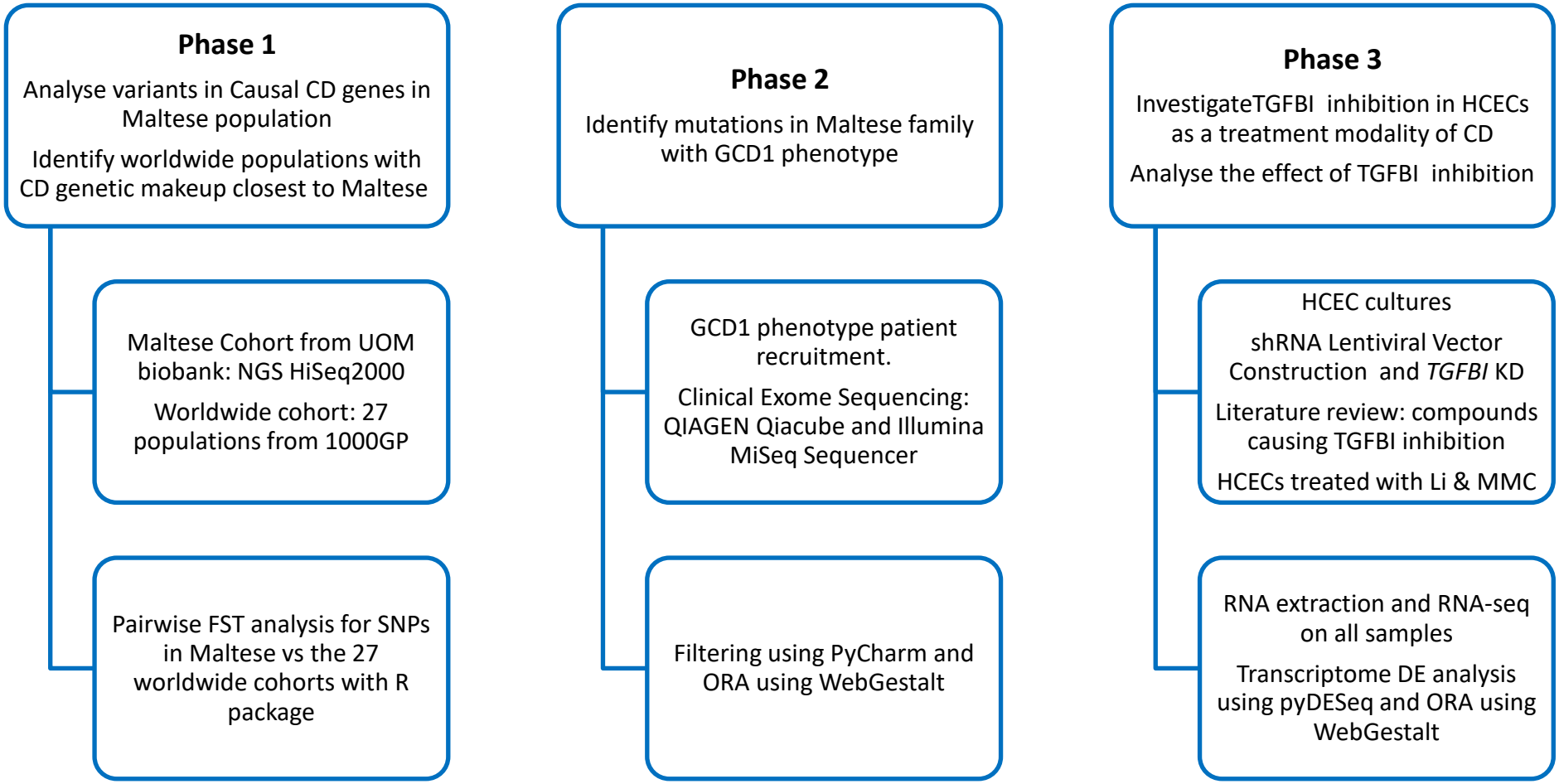
Table 46: GO annotations and pathway associations of upregulated genes identified from differential expression analysis of the RNAseq data in MMC vs control groups.

GO Biological Process	**GO Molecular Function**	**KEGG Pathway**	**Reactome Pathway**
movement of cell or subcellular component	cell adhesion molecule binding	Focal adhesion	Collagen formation
tissue development	structural molecule activity	Regulation of actin cytoskeleton	Extracellular matrix organization
locomotion	protein-containing complex binding	Fluid shear stress and atherosclerosis	Assembly of collagen fibrils and other multimeric structures
cell motility	cytoskeletal protein binding	Axon guidance	Type I hemidesmosome assembly
localization of cell	cadherin binding	Bacterial invasion of epithelial cells	Axon guidance
cell migration	actin binding	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Signalling by Receptor Tyrosine Kinases
cell junction organization	identical protein binding	Pathogenic Escherichia coli infection	Signalling by VEGF
epithelium development	protein domain specific binding	Glutathione metabolism	Non-integrin membrane-ECM interactions
cytoskeleton organization	structural constituent of cytoskeleton	Rap1 signalling pathway	RHO GTPases activate IQGAPs
neurogenesis	kinase binding	ECM-receptor interaction	EPH-Ephrin signalling

regulation of locomotion	RNA binding	Pathways in cancer	Syndecan interactions
cell morphogenesis involved in differentiation	protein kinase binding	Proteoglycans in cancer	Apoptotic cleavage of cellular proteins
cell junction assembly	calcium ion binding	Estrogen signalling pathway	Anchoring fibril formation
cell morphogenesis	integrin binding	Hippo signalling pathway	Cell-extracellular matrix interactions
response to endogenous stimulus	scaffold protein binding	p53 signalling pathway	Collagen biosynthesis and modifying enzymes
regulation of cell migration	signalling receptor binding	Cellular senescence	Formation of the cornified envelope
regulation of cell motility	laminin binding	Endocytosis	Cholesterol biosynthesis
neuron projection development	growth factor binding	Biosynthesis of unsaturated fatty acids	Cell junction organization
regulation of cellular component movement	enzyme regulator activity	AGE-RAGE signalling pathway in diabetic complications	Apoptotic execution phase
vasculature development	disordered domain specific binding	Central carbon metabolism in cancer	Gap junction trafficking
regulation of multicellular organismal development	GTPase binding	Human papillomavirus infection	Laminin interactions
cardiovascular system development	extracellular matrix binding	Glioma	VEGFA-VEGFR2 Pathway
regulation of cell differentiation	fibronectin binding	Adherens junction	Response to metal ions

generation of neurons	cell adhesion mediator activity	Hypertrophic cardiomyopathy (HCM)	Developmental Biology
cellular component morphogenesis	cell-cell adhesion mediator activity	Oxytocin signalling pathway	Gap junction trafficking and regulation

3.4 SUMMARY FLOWCHART: Methodology



CHAPTER 4: DISCUSSION

4.1 PHASE 1: Maltese Allelic Variants that are Present in Corneal Dystrophy Genes, in a Worldwide Setting

In the first phase of this project the first ever epidemiological genetic analysis study on a large cohort of the Maltese population regarding CDs was carried out. It analyses the genetic differentiation, with regard to CD-related genes, between the Maltese and various continental population cohorts. Since most of the CDs cause substantial visual impairment (Klinthworth, 2009) (Bron, 1990), these results are of significant importance since they can lead to positive measures in clinical practice and influence public health management strategies.

Diverse ethnic populations and individuals from various geographical regions have been documented to have different prevalences of CD subtypes. In the United States the overall prevalence of CDs was quoted to be 897 per million, with the largest percentage being FECD (Musch, Niziol, Stein, *et al*., 2011). It has also been reported that FECD is also very common in the UK (Keenan, Jones, Rushton, *et al*., 2012). Additionally, literature indicates LCD to be one of the more common forms of the *TGFBI* dystrophies in West while the prevalence of GCD1 is documented to be highest in Europe (Trief & Fay, 2019). In contrast to this, GCD2 has been reported to be more prevalent in Japan, Korea and the US (Fujiki, Nakayasu, & Kanai, 2001) (Klinthworth, 2009) (Lee, Cristol, Kim, *et al*., 2010). Thus, being primarily a genetic disorder, these population-specific distinctive prevalences give insight into the historical genetic composition of these populations.

It is well known that 99.9% of the genetic make-up of all humans is the same. The differences in the remaining 0.1% include SNPs that enclose important clues about

ancestral composition and risk of disease (Huang, Shu, & Cai, 2015). SNPs are defined as single base changes in a DNA sequence that occur in more than 1% in a random set of individuals in a large population (Keats, 2013) and have been established as genetic markers (Chen & Sullivan, 2003). The average density of SNPs distributed throughout the human genome is quoted as one every 1.9 kilobases (The International SNP Map Working group, 2001). Studies have shown that about 15% of all SNPs are population-specific (Huang, Shu, & Cai, 2015) (Spichenok, Budimlija, Mitchell, *et al*., 2011). The recent introduction of high throughput next-generation sequencing (NGS) uses SNPs as genetic markers to provide the background for understanding the genetic component of diseases and to facilitate the production of personalised medicine (Shastry, 2007).

Due to the relative rarity of most CDs, a large population-based sample would be required to produce an accurate measurement of the frequency of the various CDs. Implementing this method would be extremely expensive and labour intensive. The recent introduction of large population genetic databases describing genetic variants from various ancestries has provided a wealth of information regarding allelic frequencies in populations. They provide an indication of the likelihood of a sequence variant being pathogenic. Moreover, in autosomal dominant conditions with complete penetrance, an inheritance pattern found in many of the CDs, the calculation of the allele frequency of a causal mutation of a population would be a direct measurement of the prevalence of the disease in that population. Furthermore, large population genetic databases can also provide a guidance with regard to current underdiagnosis, especially in such rare conditions, which could guide clinicians towards improving their diagnostic rates.

In this study, the genetic frequency of the CD subtypes of the population cohorts analysed in the 1000GP showed that FECD was the most prevalent at 0.6%, followed by EBMD at 0.16% and LCD IIIA at 0.08%. These CD subtypes were not present in the Maltese cohort used in this study.

While keeping in mind that corneal dystrophies are rare conditions, when analysing a relatively large sample of the Maltese population from the Malta biobank, one Maltese individual was identified to have a mutation located in the *TGFBI* gene that causes methionine to be replaced by valine at codon position 502. This has been reported in literature to cause a Bowman's layer CD /atypical Thiel-Behnke CD phenotype. The phenotype of this mutation has been documented to cause small round, refractive deposits located mainly in the anterior corneal stroma bilaterally in certain individuals while severe opacifications of the cornea were observed in others (Zenteno, Correa-Gomez, Santacruz-Valdez, *et al*., 2009). This mutation was present in only 3 of the 26 population cohorts of the 1000GP, namely the Puerto Ricans, Utah residents with Northern and Western European ancestry, and Iberian population cohorts in Spain, and has a global prevalence of 0.16%. Spaniards settled in Puerto Rico in 1493 (Malavet, 2000), in the same era that Malta was occupied by the Spanish. This might explain the presence of this mutation in the Maltese cohort. Interestingly, none of the usually more prevalent subtypes of CDs, such as FECD and LCD, were present in the Maltese cohort. Three of the nine missense SNPs present in CD-related genes in the Maltese cohort were described as possibly damaging by predictive algorithms. The *KRT3* NC_000012.11:g.53186088G>C mutation was predicted to be potentially deleterious by all four predictive algorithmic tools used in this study. Even though this mutation has not been reported in literature to specifically cause a particular CD, since

the *KRT3* gene is strongly linked to the cornea, there is a reason to believe that it might increase the risk of CD.

Genetic population studies are fundamental tools in the understanding of population health (Belsky, Moffitt, & Caspi, 2013). The results of the first phase of this project identify the prevalence of known CD associated mutations, any potential mutations that might contribute to the development of these conditions, as well as the CD associated ancestral origin of a the Maltese population. This is of great value in delineating the distribution and origin of CDs. Additionally, these data prove useful in assessing the overall burden of this disease in the Maltese population. Therefore, the results obtained in this first phase of this project may serve as a valuable evidence-based measure that can be applied by healthcare administrators in the decision-making processes of healthcare policies and when assessing the need for services and treatment facilities.

From the FST results measuring the degree of population cohort differentiation of the CD-related gene SNP variants among all the population cohorts considered, which varied from 0.005 to 0.2841 (Figure 5), it can be deduced that CD related genes vary substantially among the studied population cohorts.

The Hudson's estimator used in this study generates estimates that are the average of FST according to the definition given by Weir and Hill (Weir & Hill, 2022). It has the advantageous properties of not being sensitive to the ratio of sample sizes and of not overestimating FST (Bhatia, Patterson, Sankararaman, *et al*., 2013). The FST values for the variants in the *KRT3, TGFBI, CHST6* and *SLC4A11* genes in the American group, when compared with the Maltese, show values that are indistinguishable from panmixia (Table 10). On the other hand, the variants present in the *PIKFYVE* gene show

FST values below 0.01 when the Maltese are compared with East Asians and South Asians (Table 10). The largest degree of differentiation from the Maltese cohort is seen in the African cohort, where 11 SNPs, out of a total of 19, have the largest FST values out of all the cohorts (Figure 6). The pattern of diversity of CD-related gene variants between the Maltese and other continental populations helps to shed light on the hereditary CD-related genetic differences relative to the Maltese. The discrepancy in the FST results when comparing the pooled SNP FSTs with the FSTs of the individual SNPs occurs since the data contain outliers and thus the differences even out. Besides, pooled datasets lead to considerable loss of information. These results show that it is more reasonable to focus on the individual SNPs in causal CD genes when analysing the data and deriving results on genetic differentiation. In general, it can be deduced that the CD-related gene variants identified in the Maltese cohort showed the greatest degree of differentiation with Africans, while the least degree of differentiation was seen between the Maltese and American cohorts of the 1000GP; namely, the Puerto Rican, Mexican, and Colombian cohorts. Here, the similarity between the Maltese and Hispanic cohorts emerged again. Mexicans have also been shown to have a significant European ancestral contribution (Phillips, 2015). The same can be said for Colombians but to a lesser degree (Sirugo, Williams, & Tishkoff, 2019). Thus, this leads to the rational postulation that, when considering CD-related genes, in the absence of Maltese data, any medical research on future CD genetic targeted treatment in which individuals from these populations are involved, can be used as a guide for the Maltese population.

Recent literature has drawn attention to the fact that genome-wide association studies on human disease have failed to include the level of diversity that is present

globally (Sirugo, Williams, & Tishkoff, 2019). Including diverse genetic ancestral populations in genetic research is imperative in order to fully understand the genetic component of disease in humans. Focusing on only a subgroup of the human population will lead to the underrepresentation of understudied populations, leading to the omission of important determinants of health. Large-scale genetic studies are being translated into clinical practice and public health policies, since analysis of genomes can be used to predict disease risk and prognosis of individuals, as well as for targeted drug development. Thus, the underrepresentation of certain ancestral populations in these studies may not only be misleading but also possibly unsafe. Additionally, the inclusion of genetically diverse ancestral populations in genetic research increases the global genetic bank that researchers rely on for medical progress. With that said, one should keep in mind that environmental factors such as trauma and smoking (Zhang, Igo Jr, Fondran, *et al*., 2013) may play a role in the earlier onset of this disease (Chao-Shern, Me, DeDionisio, *et al*., 2017). Besides, although still largely unexplored, other highly possible environmental factors, such as exposure to UV light, smoking and high levels of free radicals might induce a more severe phenotype or earlier onset of the condition.

This study is the first epidemiological study to describe the previously unknown CD genetic pool present in Maltese individuals, a subgroup of the human population on which CD-related literature was still lacking. Therefore, by increasing our knowledge of the epidemiology of CDs, early detection and subsequent management of individuals, particulary those with a known family history of this pathology, can be facilitated.

4.2 PHASE 2: Comprehensive Clinical Exome Sequencing Analysis of Maltese GCD1 Patients Uncovers the Causative Variant in the TGFBI Gene and Highlights Potential Alterations in the Corneal Extracellular Matrix and Integrin Interactions. Focus on Future Targeted Treatment.

The second phase of this project reports the first genetic study that has ever been carried out on Maltese individuals who phenotypically exhibit corneal dystrophy, thus generating new material about the previously unknown genetic pool present.

Identifying the mutation variant present in these Maltese patients was important in more ways than one, since, even though the diagnosis of some CD conditions on the bases of clinical features alone might seem simple, the actual underlying genotype might not be the one expected, especially in view of the fact that not all affected patients exhibit a clear genotype-phenotype correlation (Frising, Wildhardt, Frisch, *et al*., 2006) (Zenteno, Correa-Gomez, Santacruz-Valdez, *et al*., 2009). Additionally, determining the mutation variant present in this Maltese family would allow for accurate diagnosis, paving the way towards understanding the causal CD genetic mutation variants for the phenotype present within our population while at the same time making it easier for clinicians to provide proper management to subjects at risk..

The individuals in this study phenotypically exhibited corneal deposits that were identical to those described classically in previous literatures in patients with GCD1 (Klinthworth, 2009). The mutation variant identified to be responsible for the phenotype in this Maltese cohort, a heterozygous p.R555W change, is the mutation that is most recognized to cause the classic GCD1 phenotype. However, having said that, phenotypic variation can also manifest itself as a variable age of onset of the deposits, to a difference in the rate of progression, severity and prognosis of the condition as well as a difference in the time of onset of recurrence after graft surgery.

These factors are of utmost importance especially when it comes to councel patients about prognosis and make decisions regarding their personalized treatment plan.

The phenotype-genotype correlation in CDs is noteworthy since abnormalities in different genes can cause a similar phenotype, and different defects in a single gene can cause different phenotypes (Nowińska, Wylegala, Janiszewska, *et al*., 2011) (Patel, Chang, Harocopos, *et al*., 2010). In fact, the phenotype of GCD1, the initial manifestation of GCD2, as well as some of the superficial variants, can be confused, especially during the early stages, and when the diagnosis is based on the clinical features and histopathological findings alone (Mashima, Nakamura, Noda, *et al*., 1999) (Solari, Ventura, Perez, *et al*., 2007). It has been reported in certain studies that, in contrast to GCD1, patients with GCD2 may exhibit reduced penetrance (Cao, Ge, Cui, *et al*., 2009) (Kim, Kim, & Song, 2008). Thus, accurate identification of the mutation responsible for the phenotypic manifestation in patients suffering from granular corneal dystrophies is essential, especially when it comes to counselling patients regarding inheritance.

Hou *et al*., clearly reported an affected Taiwanese individual carrying the R555W mutation who was noted to have dot and line corneal opacities instead of the classic GCD1 phenotype (Hou, Wang, Hsiao, *et al*., 2012). Additionally, in another study, this time describing two families, one of Mexican and the other of Italian descent, affected individuals carrying the R555W heterozygous mutation, exhibited two distinct phenotypes, namely, the classic GCD1 phenotype and a sea fan/vortex pattern initiating from the inferior aspect of the cornea(Kattan, Serna-Ojeda, Sharma, *et al*., 2017).

Furthermore, although the majority of corneal dystrophies are caused by heterozygous mutations in the *TGFBI* gene, homozygous mutations in both GCD1 and Type II have also been reported in literature (Okada, Yamamoto, Watanabe, *et al*., 1998) (Cho, Mok, Na, *et al*., 2012). These patients usually exhibit a more severe phenotype than that seen in heterozygous patients. In a literature report by Okada, Yamamoto, Watanabe, *et al*. (1998), a R555W homozygous mutation was documented to be present in a Japanese family where the individuals affected were the offspring of two affected individuals who were consanguineous. The phenotype in this family was described as a continuous placoid opacification of the cornea with early recurrence after corneal graft surgery (Okada, Yamamoto, Watanabe, *et al*., 1998). Thus this condition is also referred to as a semidominant condition (Yang, Han, Huang, *et al*., 2010). Establishing the zygosity of the individual suffering from GCD is important to be able to establish the prognosis of the patient as well as to be able to provide guidance regarding gene transmission.

Even though there are many reports documenting the phenotypic-genotypic variability in individuals exhibiting a GCD1 phenotype, since *TGFBI* mutations have been determined to be the causal mutation in most cases of these CDs, most research studies focused only on identifying the mutation variants present in the *TGFBI* gene itself (Hou, Hsiao, Chen, *et al*., 2012) (Paliwal, Sharma, Tandon, *et al*., 2010) (Kattan, Serna-Ojeda, Sharma, *et al*., 2017) (Stewart, Ridgway, Dixon, *et al*., 1999) (Zenteno, Correa-Gomez, Santacruz-Valdez, *et al*., 2009) (Zenteno, Ramirez-Miranda, Santacruz-Valdes, *et al*., 2006). This approach has led to lack of literature about other gene mutations that might be present in individuals exhibiting CDs and has not contributed to the clarification of the causes of phenotypic variability seen in *TGFBI* CDs. Therefore,

it is quite possible that mutations altering the age of onset of the deposits, the rate of progression and severity of the condition have not yet been identified (Zhu, Wang, Wang, *et al*., 2023).

Inter-familial and inter-individual phenotypic variability seen in individuals suffering from GCD1 CDs include a varied age of onset, a difference in the rate of advancement, severity and prognosis of the condition as well as a difference in the time of onset of recurrence after graft surgery. As evidenced above, since even a slight variation in the phenotype may point towards the presence of a new genotype, it is increasingly recommended by researchers that genetic analysis is carried out on each individual exhibiting GCD (Frising, Wildhardt, Frisch, *et al*., 2006). Apart from possible environmental factors, which have not yet been identified, it can be hypothesized that variants present in genes related to *TGFBI,* or even, variants in genes coding for proteins known to be deposited together with the *TGFBI* protein in corneas of GCD1 patients account for this variability.

Clinical or whole genome sequencing of all individuals exhibiting *TGFBI* CDs will lead to a more extensive and accurate diagnosis and will ultimately unravel the cause of the variability in phenotype and genotype that has been reported universally.

In order to select the genetic variants present in the Maltese cohort exhibiting GCD1 that might be relevant to the pathogenesis and phenotype of these individuals, the variants were filtered to exclude intron variants and synonymous variants and select only the variants that were present in all the GCD patient samples. Excluding synonymous and intronic variants and selecting only the variants common to all the GCD1 patient samples ensures that, the remaining mutation variants present in key

genes that were shortlisted are possibly contributing to the overall phenotype, onset, severity and progression of the Maltese GCD1 individuals. The shortlisted variants were thus all missense mutations, downstream gene variants, splice region variants or 3 prime UTR variants.

A list of genes coding for the top proteins found to be present in GCD1 deposits (Courtney, Poulsen, Kennedy, *et al*., 2015), genes that have been documented in literature and databases to be co-expressed or modified by the *TGFBI* gene or known to bind to the *TGFBI* protein was cross-checked with the mutation variants present in all 5 of the Maltese GCD1 patient samples. The genes that were returned were further analysed by applying ORA in the WebGestalt tool.

Interestingly, the variants found to be present in the genes *COL12A1* and *ITGA11* were classified as 'probably damaging' by predictive computational tools. However, even though the other mutation variants were classified as 'benign', further research investigating their possible effect in combination with the *TGFBI* R555W mutation would prove interesting.

Corneal transparency relies on the structural organisation of the ordered hexagonal lattice network of the stromal collagen fibers. This corneal collagen arrangement consists mainly of lamellae of collagens type I and type V, intertwined with collagen type VI. Collagens type III and IV have also been detected in the corneal stroma (Agarwal, Apple, Agarwal, *et al*., 2002). Collagen type III has been implicated in corneal scarring while on the other hand, mutations in the gene coding for collagen type IV have been implicated in anterior segment ocular dysgenesis which can be associated with corneal opacification. It has been reported that collagen type XII, a non-fibrillar

collagen, interacts with collagen fibrils and together with the proteoglycans keratocan, lumican, decorin, biglycan and fibromodulin, provides structural support for collagen fibril organization and spacing (Massoudi, Malecaze, & Galiacy, 2016). Additionally, it has been proposed that one of the important physiological roles of TGFBIp is to bind covalently with collagens type VI and XII (Hanssen, Reinboth, & Gibson, 2003), thereby playing a role in collagen matrix organization (Poulsen, Runager, Nielsen, *et al*., 2018). Thus, significant missense mutations in the genes coding for these collagens may theoretically alter the course of *TGFBI* CDs. Intriguingly, the exome sequencing carried out in the Maltese GCD1 cohort revealed noteworthy mutation variants in all of these central collagen types, hence laying the foundations for future research exploring the potential pathogenicity of these mutations.

The gene *ITGA11* encodes an alpha integrin that binds to a beta chain to form part of an integral membrane protein. It appears to be a receptor for collagen and plays a central role in collagen deposition during corneal development. It has been reported that TGF-β 1 is a potent activator of *ITGA11* (Grella, Kole, Holmes, *et al*., 2016). KEGG, Reactome pathways, and GO annotations link *ITGA11* to the collagen-activated signalling pathway, PI3K-Akt signalling pathway and extracellular matrix organization category.

Fibronectin is an ECM glycoprotein that is known to interact with TGFBIp (Billing, Whitbeck, Adams, *et al*., 2002). It is also known to bind to cell surfaces and to several compounds such as collagen, besides to TGFBIp. It is worth noting that, *TNC* is a paralog of *FN1* and missense mutations were found to be present in both of these genes in the Maltese GCD1 individuals. There are also reports that tenascin binds to

fibronectin and together they promote the development of corneal neovascularisation (Baleriola-Lucas, Fukuda, Willcox, *et al*., 1997). According to GO annotations, TNC is involved in cell adhesion as well as proteoglycan binding and is thus involved in extracellular matrix structure organization. Among the related pathways of *TNC* that are listed in KEGG and Reactome are the PI3K-Akt signalling pathway and Integrin cell surface interactions.

Glycosaminoglycans are a major constituent of the corneal stroma ECM. *HSPG2*, another gene that was found to contain missense mutations in the Maltese GCD1 patients, encodes the perlecan protein that is made up of a core protein and glycosaminoglycans. In fact, the main glycosaminoglycan synthesized by epithelial cells was found to be heparan sulphate, (Michelacci, 2003) which is required in the maintenance of corneal homeostasis (Coulson-thomas, Chang, Yeh, *et al*., 2015). GO annotations and related pathways of *HSPG2* include extracellular matrix organization with structural molecule activity as well as involvement in integrin cell surface interactions.

The gene *IL6R* encodes part of the receptor for interleukin 6, a cytokine that modulates cell growth and differentiation. It also plays a role in the immune response. PI3K-Akt signalling pathway, cell adhesion and growth factor binding are listed among the GO annotations. The GO annotation for the DSP gene included cell adhesion molecule binding with no pathway involvement listed in KEGG and Reactome databases. Mutations in the DSP gene have been reported to cause congenital erythrokeratodermia with associated corneal opacities (Boyden, Kam, Hernandez-Martin, *et al*., 2016).

Missense mutations were also found to be present in the *A2M* and *MMP9* gene in the Maltese GCD1 patients. The protein encoded by the *A2M* gene acts as a protease inhibitor, inhibiting a number of proteases such as trypsin, thrombin and collagenase, hence regulating the modulation of extracellular matrix organization. It also functions as a cytokine transporter, thus inhibiting inflammatory cascades while also modulating the clearance and degradation of A-beta, the main component of beta-amyloid deposits. On the other hand, *MMP9* encodes an enzyme that degrades collagens type IV, V and fibronectin. Therefore, these two proteins both play an important role in local proteolysis of the extracellular matrix and extracellular matrix organization as indicated by the Reactome pathway classification and GO annotations.

Further research into the effect all these mutations may have, especially in the presence of a mutated TGFBIp, would help understand further the pathogenesis of this subset of *TGFBI* CDs. Ideally, in order to help pinpoint better contributing mutation variants, future studies revolving around multiple families carrying *TGFBI* mutations should be carried out. Genetic analysis and regular follow up of these individuals longterm, starting from their childhood years and recording onset, progression and disease severity, would determine definitively the role mutation variants, other than the *TGFBI* gene mutation, play, in the phenotypic heterogeneity exhibited by these individuals.

The results obtained from this phase of the study set the foundation towards understanding better the local Maltese GCD1 disease course and phenotype. This study can serve as a model for future international studies to be carried out. Documenting potential mutation variants present, together with the phenotype

exhibited by families suffering from GCD1 from all over the world will help hone in and confirm any contributing gene mutations, while helping to understand better the pathophysiology of these conditions.

Furthermore, knowledge of a patient's genetic profile would also be crucial in developing personalised treatment that might even target genes other than *TGFBI* , such as genes related to collagen organisation and extracellular matrix modulation,which might affect severity or onset, should *TGFBI* inhibition in vivo prove to be too detrimental. Additionally, this data can help anticipate or comprehend the effects and side-effects of future proposed drug treatments, when the molecular targets of which would involve any of the above mentioned genes and/or pathways.

4.3 PHASE 3: Inhibition of *TGFBI* as a Treatment Modality: Functional Work

4.3.1 *TGFBI* Knock Down in Human Corneal Epithelial Cells

4.3.1.1 The Elusive Physiological Role of TGFBIp

Different types of cells exist because a cell usually expresses only a fraction of its genes. Each cell type expresses a different set of genes making it unique. When, how and which genes are expressed, in order for the appropriate proteins to be synthesised, contributes to the proper functioning of the cell. This regulation of gene expression occurs from information encoded in their DNA (Shen C.-H. , 2023).

TGFBIp has been identified as an ECM protein that is expressed ubiquitously in multiple tissues (Skonier, Neubauer, Madisen, *et al*., 1992) (Ivanov, Ivanova, Salnikow, *et al*., 2008). Recently, extensive research has focussed on TGFBIp structure, regulation of its expression as well as on its functional and pathological roles. However, most of the mechanisms are still not satisfactorily understood.

The function and structural integrity of the cornea relies on the specific and orderly interaction of the constituting macromolecular components in the ECM. Literature has shown that *TGFBI* mediates the binding of ECM proteins such as collagen, laminin and fibronectin to integrins which in turn has been associated with the activation of cell proliferation, adhesion, migration and differentiation (Kim, Jeong, Nam, *et al*., 2002). The tendency of TGFBIp to associate with these ECM molecules could be important for the biological functions of ECM matrix and TGFBIp itself. However, the physiological importance of such interactions has not yet been illustrated and the TGFBIp binding

sites involved in the interaction with other ECM proteins have yet to be mapped (Runager, Klintworth, Karring, *et al*., 2013).

In literature it was reported that *TGFBI* expression in corneal epithelial cells can be induced by TGF-β 1 (Wang, Munier, Araki-Saski, *et al*., 2002). The binding of TGF-β to its receptor can theoretically lead to the activation of either the SMAD-signalling pathway or else non-SMAD-signalling pathways that include: PI3K/AKT, MAPK pathways (ERK, JNK, and p38 MAPK) as well as NF-κB and Rho/Rac1 pathways, amongst others. Besides the direct regulation by the TGF-β/SMAD signalling pathway, the expression of the *TGFBI* gene is also modulated indirectly by the PI3-K/AKT, the cAMP/PKA and the JNK signalling cascades. Additionally, it has also been documented that in various cell types, *TGFBI* expression is also altered by retinoid, IL-4, IL-1, IL-6, TNF-α and microRNA (miR-21) (Nam, Sa, You, *et al*., 2006) (Dokmanovic, Chang, Fang, *et al*., 2002). However there are still large gaps in literature with regards to the molecular understanding of how *TGFBI* is induced in different cell types and its regulatory factors have yet to be improved. In fact, even though TGFBIp is one of the most abundant proteins within the human cornea, its physiological role remains elusive. Transcriptome analysis of *TGFBI* KD HCECs carried out in this PhD study, helps define the role TGFBIp plays in homeostasis of HCECs by delineating the key genes and pathways regulating various cellular functions which are associated with this ECM protein.

4.3.1.2 RNAi

The maintenance of the cell's physiological balance is crucial in human health, and when this balance is tipped, a variety of disorders can arise. Mutations or alterations in expression of the *TGFBI* gene can potentially have various downstream effects on

other genes involved in various cellular processes. RNA interference (RNAi), using shRNAs, is a reproducible biological process enabling sequence-specific silencing of gene expression. Once knockdown of the shRNA-targeted gene is confirmed, a variety of assays can be performed. This enables researchers to monitor and analyse the effects gene-specific knock downs have in the context of a specific pathways within the cell, and ultimately within the context of human disease. It provides the opportunity of unravelling the uncertainties of individual gene function in cell physiology and disease states as well as in identifying previously unknown interacting genes.

The specific genes affected by shRNAs will depend on the cell type and type of gene alteration. Since TGFBIp in the cornea is mainly produced by the corneal epithelium (Escribano, Hernando, Ghosh, *et al*., 1994) (Akhtar, Meek, Ridgway, *et al*., 1999) (Nielsen, Poulsen, Lukassen, *et al*., 2020) the human corneal epithelium should be the target of further research. Recently, an increasing number of studies have looked into the use of RNAi in HCECs in the inhibition of pathologic processes that occur in the cornea. The large majority used siRNAs in their studies while a smaller number chose the more stable shRNA approach (Cao, Wang, Deng, *et al*., 2023) (Yellore, Rayner, & Aldave, 2011). To the best of my knowledge, only two studies have reported the use of RNAi to target specifically the *TGFBI* gene in human corneal epithelial cells, and this was done by using siRNA (Courtney, Atkinson, Moore, *et al*., 2014) (Yellore, Rayner, & Aldave, 2011). RNAi was applied by Courtney *et al*. to specifically target a LCD mutant allele (Courtney, Atkinson, Moore, *et al*., 2014). While this allele-specific strategy could be applicable to other *TGFBI* mutations, this method would be only effective in heterozygous patients since it still relied on the production of wild-type TGFBIp. Thus, in homozygous patients one would still have to consider complete *TGFBI* knockdown.

Recent advancements in molecular genetic techniques have led to an ever increasing number of genetic mutations that have been attributed to various phenotypes of *TGFBI*-associated corneal dystrophies. Additionally, research testing on all the known variants would be tremendously expensive and labour intensive. Therefore, preferably, treatment plans that target all the subtypes in the *TGFBI* CD group would be desirable by aiming to suppress the disease-causing protein. Yellore *et al*. investigated the efficacy of siRNA in decreasing TGFBIp expression in TGF-β 1-induced and noninduced HCECs, and concluded that TGFBIp production in HCECs can be inhibited with RNA interference (Yellore, Rayner, & Aldave, 2011). These studies provided insight into the potential of partial or complete knockdown of *TGFBI* as a future therapeutic strategy to treat *TGFBI* -linked CDs. Nevertheless, the effect on corneal matrix structure and potential consequences on corneal homeostasis when reducing or stopping TGFBIp production in the cornea had not been explored (Poulsen, Runager, Nielsen, *et al*., 2018). Subsequently, in order to investigate the possible side-effects of total KD of *TGFBI* on corneal integrity, Poulsen *et al*. analysed the histological structure of corneas in *TGFBI*-deficient mice and reported that even after eliminating TGFBIp totally from the mouse cornea, the mouse cornea still exhibited structural integrity with only minor structural changes seen. Nevertheless the effect on human corneal epithelial cells has not yet been fully investigated.

4.3.1.3 Targeting HCECs

Human corneal epithelial cells are the main source of mutant TGFBIp production in individuals suffering from *TGFBI* CD diseases. Therefore, *TGFBI* KD targeting specifically corneal epithelial cells of these individuals should subsequently inhibit the deposition of visually significant corneal deposits. Since corneal epithelial cells are located

superficially and are easily accessible, therapeutic delivery of such therapy would be less complicated than having to target all layers of the cornea (Sarkar, Panikker, D'Souza, *et al*., 2023). The results obtained from the research carried out by Poulsen, Runager, Nielsen, *et al*. (2018), Courtney, Atkinson, Moore, *et al*. (2014) and Yellore, Rayner & Aldave (2011) encourage further research on novel therapeutic options exploring the effect of KD of *TGFBI*, this time specifically targeting human corneal epithelial cells, since one has to keep in mind that mice express much lower corneal TGFBIp levels when compared to humans. Besides, even though Poulsen, Runager, Nielsen, *et al*. (2018) ventured along the path of investigating the corneal proteomic profiles in *TGFBI*-null mice, none of these projects looked into the downstream effects *TGFBI* KD could subsequently have on human corneal epithelial cell pathways and the expression of *TGFBI* related genes, as opposed to mice corneas. Thus, the potential side-effects of knocking down one of the most abundant proteins present in the human cornea still remain unclear and definitely worthwhile being explored.

This PhD study is the first study in which RNAseq and subsequent differential gene expression analysis were carried out in shRNA-induced *TGFBI* KD in human corneal epithelial cells. Considering the high level of expression of *TGFBI* in the human cornea and the multiple physiological roles TGFBIp has been documented to have, the results obtained from the transcriptomic and pathway analysis in this project shed light on any alterations in homeostasis *TGFBI* KD has in HCECs.

4.3.2 Comprehensive Transcriptome Analysis to Investigate the Differential Expression of Genes between Normal HCECs and TGFBI KD HCECs: Implications for TGFBI CD Treatment.

Transcriptome analysis of *TGFBI* KD human corneal epithelial cells carried out in this study is the first step towards identifying the key genes and pathways associated with this ECM protein that has been documented to be involved in cell growth, differentiation, wound healing as well as cell adhesion (Reinboth, Thomas, Hanssen, *et al*., 2006). The results obtained from this PhD study additionally help to elucidate the role *TGFBI* KD may have as a potential therapeutic target for *TGFBI* CD disease treatment, since, decreasing *TGFBI* expression would possibly inhibit mutant TGFBIp deposition (Yellore, Rayner, & Aldave, 2011).

In this study, we used bioinformatic analysis to identify DEGs between *TGFBI* KD HCECs and the normal control HCECs. Pin pointing the up and down regulated genes can give us a guide regarding *TGFBI* associated genes and genes encoding for proteins that are associated with pathways *TGFBI* is involved in. It also delineates the effect silencing of *TGFBI* would have on corneal epithelial cells if implemented as a proposed treatment for *TGFBI* CDs.

Through the use of WebGestalt (WEB-based Gene SeT AnaLysis Toolkit), the the enriched GO biological and molecular function, as well as the KEGG and Reactome pathways associated with the significant DE genes in the *TGFBI* KD HCECs, were identified and analysed. The genes that showed the largest $\log_2$FCs were looked into. Additionally this study also honed in on and identified the GO annotations and related pathways of 'genes of interest' known to be related to *TGFBI* , which were shortlisted

from literature and databases as well as proteins that have been shown to be present in *TGFBI* CD deposits (Courtney, Poulsen, Kennedy, *et al*., 2015).

4.3.2.1 Downregulated Differentially Expressed Genes in TGFBI KD HCECs

The downregulated genes that were found to have the largest log_2FC were *CD24, MTPN, DDX39B, AMMECR1L, CHML* and *AREG*, while, the seven 'genes of interest' that were found to be significantly downregulated in the KD HCECs were *S100A9, IL1B, IL6, JAK2, STAT2, CPA4* and *ITGA11*.

The *CD24* gene encodes a sialoglycoprotein that is expressed on mature granulocytes and B cells and plays a role in signalling associated with growth and differentiation in these cells. *CD24* has been implicated in angiogenesis via Hsp90-mediated STAT3/VEGF signalling pathway (Wang, Zhang, Zhao, *et al*., 2016). It has also been reported that TGF-β stimulates *CD24* expression (Nakamura, Terai, Tanabe, *et al*., 2017).

The gene *DDX39B* encodes a member of the DEAD box family of RNA-dependent ATPases that mediate ATP hydrolysis during pre-mRNA splicing. It also plays a role in exporting spliced and unspliced mRNA from the nucleus to the cytoplasm, including export of FUT3. In turn, FUT3 mediates the fucosylation of TGFβR-I, which activates the TGFβ signalling pathway (He, Li, Lai, *et al*., 2021). This could be a reflection of the mechanism brought about by the shRNA itself.

The protein encoded by the *AREG* gene, amphiregulin, is a transmembrane glycoprotein related to epidermal growth factor (EGF) and transforming growth factor alpha (TGF-alpha). It is known to interact with the EGF/TGF-alpha receptor to promote cell proliferation of normal epithelial cells. Generally, TGF-alpha is known to promote

cell proliferation while the versatile TGFβ may stimulate or inhibit proliferation depending on the type of cells and growth factors involved (Partridge, Green, Langdon, *et al*., 1989). For instance, in human lung epithelial cells, *AREG* activates the EGFR/JNK/AP-1 pathway that in turn mediates the TGF-β-induced epithelial mesenchymal transition (Cheng, Kao, Chen, *et al*., 2022).

The gene S100A9 codes for a calcium- and zinc-binding protein that modulates inflammatory and immune responses. It has also been implicated in the pro-inflammatory IL-17 signalling pathway identified during KEGG pathway analysis. The canonical IL-17 signalling pathway activates downstream pathways that include NF-kappaB and MAPKs to induce the expression of antimicrobial peptides, cytokines and chemokines. A relation between TGF-beta 1 and *S100A9* has also been documented in inflammation and pancreatic cancer cells in previous literature (Ahmed, Bradshaw, Gera, *et al*., 2017). Interestingly, Courtney, Poulsen, Kennedy, *et al*. (2015) reported the accumulation of S100A9 in LCD1 TGFBIp deposits, while usually this protein is not found to be present within the healthy cornea (Courtney, Poulsen, Kennedy, *et al*., 2015). Thus, downregulation of S100A9 seen in *TGFBI* knocked down corneal epithelial cells, could additionally lead to decreased deposition of complexes if implemented as treatment in LCD1 individuals, providing further benefit.

IL6 is a cytokine that has been documented to have both pro- and anti-inflammatory properties (Scheller, Chalaris, Schmidt-Arras, *et al*., 2011). GO annotations related to the IL6 gene included ‘cellular response to stress’ and ‘peptidyl-serine phosphorylation’ while ‘NOD-like receptor signalling pathway’,’ FoxO signalling pathway’, ‘IL-17 signalling pathway ‘and ‘TNF signalling pathway‘ were identified as

related KEGG pathways. Previous studies carried out on HK-2 cells (human renal proximal tubular epithelial cells immortalized by transduction with human papilloma virus 16 E6/E7 genes) have indicated that IL-6 increased trafficking of TGF-β1 receptors that in turn resulted in amplified TGF-β1 SMAD signalling (Zhang, Topley, Ito, *et al*., 2005). In this study *TGFBI* KD showed downregulation of *IL6* expression; this might imply that these two proteins are positively correlated when it comes to corneal healing and inflammatory response.

IL6 is also known to activate *JAK2* (Janus Kinase 2 gene) via its receptor that in turn activates transcription factor STAT3. *JAK2* is a non-receptor tyrosine kinase that modulates cellular growth and proliferation. In this study both *IL6* and *JAK2* were found to be significantly downregulated. *JAK2* is implicated in 'cellular response to stress', 'phosphotransferase activity', 'protein kinase activity', 'ATP binding', 'purine ribonucleoside triphosphate binding' and 'adenyl ribonucleotide binding' with regards to GO annotations and to 'necroptosis' in KEGG pathways.

Another downregulated gene found in *TGFBI* KD HCECs in this study was *IL1B*. *IL1B* also forms part of the group of pro-inflammatory cytokines, with research reports stating that overexpression of *IL1B* by the corneal epithelium causes persistent severe inflammation of the ocular surface (Liu, Yuan, Zhang, *et al*., 2012). Interestingly, TGF-β and *IL1* have been labelled as 'master regulators' of the corneal wound healing reaction to injury with reports of opposing but occasionally also supporting roles (Wilson, 2021). GO annotations associated to this gene included 'cellular response to stress', 'cell cycle' and 'cellular response to external stimulus'. Similarly to *IL6, IL1B* was listed in KEGG pathways to be implicated in 'NOD-like receptor signalling pathway', 'IL-

17 signalling pathway ‘ and the ‘TNF signalling pathway‘ with additional pathways including ‘NF-kappa B signalling pathway’ and ‘Necroptosis’.

The gene *ITGA11* encodes integrin alpha-11 protein. It has been reported in one paper that TGF-β1 induces alpha11 via SMAD2 and other elements in primary fibroblasts (Lu, Carracedo, Ranta, *et al*., 2010). In another study, this time investigating keratoconic corneas, also reported increased expression of alpha11 integrin chain when cultures were treated with TGF beta and they also noted its enhanced expression in scarred keratoconus corneas implying a role in collagen deposition when corneal scarring occurs (Bystrom, Carracedo, Behndig, *et al*., 2009). In the *TGFBI* KD HCECs in this study, *ITGA11* was found to be downregulated, thus inferring possibly decreased expression of pathways leading to scarring when *TGFBI* is silenced.

On the other hand, the main role of STAT2 is to transduce signals downstream of the receptors for IFN-I and -III. It acts as both a positive and negative regulator of IFN-I signalling. Additionally, type I interferons can be either anti-inflammatory and tissue protective or proinflammatory and stimulate autoimmunity (Kalliolias & Ivashkiv, 2010).

In summary, Webgestalt analysis indicated that, the main GO Annotations associated with all the significantly ($p<=0.05$) downregulated DE genes in KD HCECs were related to the cell cycle (specifically signal transduction by p53 class mediator), cellular response to stress and external stimuli and transferring of phosphorus-containing groups. Related pathways included the NOD-like receptor signalling pathway, IL-20 family signalling pathway, IL-17 signalling pathway, TNF signalling pathway, NF-kappa B signalling pathway and the mTOR signalling pathway, together

with regulation of tp53 activity and transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer.

When only the downregulated genes in KD HCECs with p value of <=0.05 and log_2FC of >-1.5 were taken into consideration, the main associated GO annotations included 'regulation of tyrosine phosphorylation of STAT protein'and 'positive regulation of JAK-STAT cascade', ' homophilic cell adhesion via plasma membrane adhesion molecules', 'calcium ion binding' and 'coupled ATPase activity'. KEGG and Reactome pathway descriptions together included mainly 'RNA transport' and 'Metabolism of RNA', 'FoxO signalling pathway, ' PI3K-Akt signalling pathway', ' JAK-STAT signalling pathway', 'Cytokine-cytokine receptor interaction', 'Collagen formation' and 'Keratan sulphate biosynthesis'.

When the function of the downregulated DE genes with a log_2FC of >-1.5 were analysed in more detail (Table 47), most were associated with cell adhesion and signalling (such as *CD24, NEDD9, SH3GL2, ST3GAL6, ST6GALNAC5, MPPE1, IFT80, IZUMO1, CLDN16 COL8A2, S100A8, CTNND1, MEGF11, IGSF1, SGK3, GNGT1, GNB3, RGS9*); immune response regulation and inflammation (including *CD24, IL24, IL20, IL23A, IL6, IRAK3, S100A8*); transcriptional regulation (such as *GTF2IRD2B, DDX39B,DQX1, PNRC2, SCX*) and neurotransmitter/neuronal system (*DNAJC6, SH3GL2, PCDHGC3, PCDHGB5, PCDHGB6, PCDHGA9, PCDHGA10, PCDHGB7, IGSF1, GABRP, GNGT1, GNB3*). Other genes of note included those related to glycolipid biosynthesis (*ST3GAL6, ST6GALNAC5, B3GALT2, MPPE1*). The genes encoding for collagen structural proteins *COL8A2* and *COL9A2*, as well as *LOX*, the gene encoding for an enzyme that catalyzes the crosslinking of collagen and elastin fibers in the extracellular matrix, and

P4HA3, encoding for an enzyme involved in collagen synthesis, were also found to be downregulated. A subset of genes involved in RNA processing, were also enriched (*SNORD3B-2, SNORD3B-1, SNORD3C, RPP14, HNRNPH2, SRSF10, SNRPN* and *TRMT13*). However these might be due to the effect of shRNA transduction.

Thus the general trend of downregulated genes observed in *TGFBI* KD HCECs points towards a decrease in inflammation via downregulation of *IL1B, IL6* and *JAK2*; decreased angiogenesis via downregulation of *CD4, IL6* and *JAK2*; decreased TGF-β signalling via downregulation of *DDX39B* and *IL6*, and decreased fibrinectin production and corneal scarring via downregulation of *AREG* and *ITGA11*. These general outcomes, which would be theoretically seen by the mentioned downregulated genes, were subsequently analysed in combination with the upregulated genes in *TGFBI* KD HCECs.

Table 47: Categorization of downregulated DE genes with a log_2FC of >-1.5 according to physiological function based on current literature.

Cell adhesion and signalling	**Immune response**	**Growth factor** **Cell proliferation**	**Inflamma-tion**	**Healing and tissue repair**	**Neurotransmitter /neuronal system involvement**	**Transcriptional regulation**
CD24, NEDD9 SH3GL2 ST3GAL6 ST6GALNAC5 MPPE1 PCDHGC3, PCDHGB5, PCDHGB6, PCDHGA9, PCDHGA10, PCDHGB7 IFT80, IZUMO1 CLDN16 COL8A2 S100A8 CTNND1 MEGF11 IGSF1, SGK3 GNGT1, GNB3 RGS9, ANXA8	*CD24 IL24, IL20 IL23A, IL6, IRAK3 S100A8*	*AREG SH3GL2 CSNK1E(Wnt) BOP1*	*AREG IL24, IL20 IL6, IRAK3 S100A8*	*AREG IL20 RTEL1*	*DNAJC6, SH3GL2 PCDHGC3, PCDHGB5, PCDHGB6, PCDHGA9, PCDHGA10, PCDHGB7, IGSF1 GABRP, GNGT1, GNB3*	*GTF2IRD2B DDX39B DQX1, PNRC2 SCX, TRPS1*

4.3.2.2 Upregulated Differentially Expressed Genes in TGFBI KD HCECs

The upregulated genes that were found to have the largest log_2FC were *ICAM4, RNA5SP439, RNA5S17, SNORD14C, RN7SL385P* and *ANKRD20A10P*. Most of these are pseudogenes, which are non-protein coding genes. However, lately, it has been discovered that they can be processed into short interfering RNAs that can regulate the expression of coding genes at transcriptional and post-transcriptional level (Ranganathan & Sivasankar, 2014). Additionally, the six 'genes of interest' that were found to be significantly upregulated in the KD HCECs were *SMAD6, AKT1, ITGA6, MMP2, COL6A1* and *COL6A2*.

The upregulated *ICAM4* gene encodes the intercellular adhesion molecule 4, which has been shown to serve as a ligand for alpha-4/beta-1 and alpha-V integrins. *TGFBI* is known to bind to alpha-V integrins with high affinity.

SMAD6 is an inhibitory SMAD and is known to negatively regulate TGF-β type1 signalling. It has been shown to preferentially inhibit BMP-induced SMAD signalling (Miyazawa & Miyazono, 2017). However, it has also been shown to impede the phosphorylation of SMAD2 and the subsequent heteromerization with SMAD4 (Inamura, Takase, Nishihara, *et al*., 1997). DEseq analysis of DE genes in *TGFBI* KD HCECs showed an upregulation of *SMAD6* thus exhibiting a negative correlation with *TGFBI* expression.

AKT1 encodes AKT Serine/Threonine kinase1. It is a regulatory protein kinase that is critical in the PI3K/AKT signalling pathway. Thus, it plays a role in the modulation cell proliferation, metabolism and angiogenesis in both normal and malignant cells and is a known oncogene. GO annotations for *AKT1* included 'apoptotic signalling pathway', 'developmental growth', 'positive regulation of cell communication', 'cell activation' and 'positive regulation of signalling' with pathway associations of 'Glucagon signalling pathway', 'Focal adhesion', 'Pathways in cancer', 'Intrinsic Pathway for Apoptosis', 'Programmed Cell Death', 'eNOS activation and regulation', 'Translocation of SLC2A4 (GLUT4) to the plasma membrane', 'Metabolism of cofactors' and 'Metabolism of nitric oxide'. PI3K/AKT signalling pathway molecules are known to mediate crosstalk with the TGFb pathway that in turn regulates TGFBIp production (Yu, Ramasamy, Murphy, *et al*., 2015). Similarly to *AKT1*, *ITGA6* has also been implicated in the 'regulation of apoptotic signalling pathway', 'focal adhesion', 'ECM-receptor interaction' and 'Pathways in

cancer' according to the ORA carried out, thus functioning as a cell surface adhesion molecule binding ECM molecules of the laminin family (Stelzer, Rosen, Plaschkes, *et al*., 2016).

In this study, *COL6A1* and *COL6A2*, encoding for collagen VI proteins, were both found to be upregulated in *TGFBI* KD HCECs. *COL6*, similarly to *ITGA6*, is also associated mainly with 'Focal adhesion', 'ECM-receptor interaction' as well as with 'animal organ morphogenesis'. Collagens have been reported to interact with several ECM proteins and cell surface molecules such as integrin proteins. Besides its important structural role, collagen VI affects various cellular functions including adhesion, migration and cytoprotective functions, which include cell cycle progression by counteracting apoptosis and oxidative damage, as well as regulating autophagy (Cescon, Gattazzo, Chen, *et al*., 2015). It has been reported that overexpression of *COL6* leads to remodeling of the ECM and occurs during the healing response of corneal wounds (Esquenazi, Esquenazi, Grunstein, *et al*., 2009). *COL6* upregulation was also reported by Poulsen *et al*. in *TGFBI*-null mice, where they concluded that the covalent binding of TGFBIp to COL6 must be one of the important physiological roles of TGFBIp, assisting subsequent regulation of matrix organization and collagen constituent composition. However, as mentioned previously, this deregulation didn't produce significant structural changes in corneas of *TGFBI*-null mice (Poulsen, Runager, Nielsen, *et al*., 2018).

The gene *MMP2* that encodes the enzyme matrix metallopeptidase 2 was also found to be upregulated. It is known to play an important role in extracellular-remodeling during corneal wound healing (Li, Lokeshwar, Solomon, *et al*., 2001). Its increased expression has been positively associated with increased corneal fibrosis and

neovascularisation in corneal disease (Wolf, Clay, Oldenburg, *et al*., 2019). It has also been documented to cleave type IV collagen that is a major structural component of basement membranes. KEGG and Reactome pathways enriched for *MMP2* were 'Pathways in cancer' and 'EPH-ephrin mediated repulsion of cells' while GO annotations were related to 'ossification' and 'animal organ morphogenesis'.

Thus, the enriched pathways and GO annotations associated with all the significantly (p<=0.05) upregulated DE genes in KD HCECs were mainly associated with cell cycle regulation (including the apoptotic signalling pathway), cell communication and signalling, ECM-receptor interaction, focal adhesion and identical protein binding.

When only the upregulated genes in KD HCECs with p value of <=0.05 and $\log_2$FC of >1.5 were taken into consideration, the main GO annotations associated with these genes were 'defense response', 'cell adhesion', 'extracellular matrix structural constituent', 'proteoglycan binding' and 'immune system' and 'signalling receptor activity binding and regulation'. The major associated pathways included 'JAK-STAT signalling pathway', 'Cell adhesion molecules', 'Cytokine-cytokine receptor interaction' and 'O-linked glycosylation of mucins'.

When the function of the upregulated DE genes with a $\log_2$FC of >1.5 were looked into in more detail (Table 48), most were associated with cell adhesion and signalling (*MUC1, MAPK8IP2, LRP1, IFNAR2, ADCY8, VSTM2L, ITGB6, L1CAM, LY6D, CADM2, SELP, WNT10A, PDE1A, LRRC4B*); immune response regulation (*CFD, IL17D, IRF4, PRG2, CLEC4A, IL13, IFNL1, KLRC3, LY6D, LCK, CLEC4A, IL34, FCGR1A, CFI, CFH, SELP, IFNAR2*); transcriptional regulation (*IRF4, CITED1, SOX8, EGR3, CNTN1, L1CAM, CADM2, CNTN5*) and neurotransmitter/neuronal system (*PDXP, NPY4R, NPFFR1, NTSR1, GABRQ, KCNC3,*

CPLX1, LRRC4B, KCNJ6, HTR3E, ADCY8). Other genes of note included genes encoding components of ribosomal subunits that are involved in ribosome assembly and protein synthesis (*RPL3L, RNA5S9, RNA5S13, RNA5S17*) and a few that are associated with inflammatory processes (*CFD, IL17D, MMP12, PRG2, IL13, SELP*). Genes that might also be of interest included *FBLN5*, a gene that encodes for fibulin-5, an extracellular matrix protein that is important for tissue structure and function; *TP73*, a gene that encodes for a member of the p53 family of proteins involved in cell cycle control, apoptosis and tumor suppression and *MAPK8IP2*, a gene encoding an adaptor protein that regulates the activity of the JNK pathway, which is involved in cellular responses to stress and apoptosis.

'Transcription' and 'RNA polymerase II regulatory region sequence-specific DNA binding' annotations were also noted. However similarly to what was noted in the downregulated genes, these might be related to the method employed to induce the *TGFBI* KD in the HCECs, namely shRNA.

Therefore, the general trend of upregulated genes observed when *TGFBI* was KD in HCECs, theoretically points towards the inhibition of TGF-β/SMAD and BMP signalling due to the upregulation of *SMAD6*; the promotion of remodelling of the ECM together with fibrosis and neovascularisation due to upregulation of *MMP2*, *AKT1*, *COL6A1* and *COL6A2*; increased cell proliferation modulated by overexpression of *AKT1* and *ITGA6*; as well as increased integrin signalling due to upregulation of *ICAM* and *ITGA6*.

Table 48: Categorization of upregulated DE genes with a log_2FC of >-1.5 according to physiological function based on current literature.

Cell adhesion and signalling	**Immune response**	**Growth factor** **Cell proliferation**	**Inflammation**	**Healing and tissue repair**	**Neurotransmitter /neuronal system involvement**	**Transcriptional regulation**
MUC1 MAPK8IP2(JNK) LRP1 IFNAR2 ADCY8 VSTM2L ITGB6 L1CAM LY6D CADM2 SELP WNT10A PDE1A LRRC4B ADCY8	*CFD IL17D IRF4 PRG2 CLEC4A IL13 IFNL1 KLRC3 KLRC1EGR3 LY6D LCK CLEC4A IL34 FCGR1A CFI CFH SELP IFNAR2*	*PDGFRB FGF3*	*CFD IL17D MMP12 PRG2 IL13 SELP*	*FGF3*	*PDXP NPY4R NPFFR1 NTSR1 GABRQ KCNC3 CPLX1 LRRC4B KCNJ6 HTR3E ADCY8*	*IRF4 CITED1 SOX8 EGR3 CNTN1 L1CAM CADM2 CNTN5*

Also, one must look at the global effects of the resulting upregulated and downregulated genes since most of these DE genes, having opposite effects, possibly balance out each other. On one hand one finds upregulation of *MMP2*, *AKT1*, *COL6A1* and *COL6A2* genes that play a role in fibrosis and angiogenesis while on the other hand there is concurrent downregulation of *CD4, IL6, IL1B, JAK2, AREG* and *ITGA11* which, being under expressed, would lead to a decrease in inflammation, angiogenesis and corneal fibronectin production and scarring.

TGFBI is also known to act as a tumour suppressor gene, and, together with the upregulation of *AKT1* (a known oncogene) seen in this study, the possibility of oncogenic consequences must be addressed. The proposal for the use of RNAi to

silence *TGFBI* as a new treatment modality for *TGFBI* CDs would aim to be tissue specific where only the human corneal epithelial cells would be exposed to the *TGFBI* KD therapy. This would reduce significantly the risk of oncogenic consequences given that the cornea is rarely the primary site for neoplasms (Rao & Shields, 2019). Furthermore, interestingly, in a study where proteomic profiling was carried out in *TGFB* -null mouse corneas, the authors reported only minor changes in corneal matrix composition. These results should reasonably be analysed with caution when extrapolating them to humans. This is because, as correctly pointed out by the authors themselves, the corneal *TGFBI* expression levels in humans is about 10-fold more than that found in mice (Poulsen, Runager, Nielsen, *et al*., 2018).

4.3.2.3 TGFBI KD in HCECs and it's Potential Implications for Understanding CDs

TGFBI CDs are a complex group of conditions. The pathways leading to complex formation and deposition of mutated TGFBIp in CDs and the reason why the pathological manifestation of *TGFBI* mutations is confined only to the cornea has yet to be uncovered.

KD experiments provide a controlled and precise way to study the role of *TGFBI* in the disease process, offering valuable insights that may not be readily apparent when comparing normal and diseased cells directly. In this research study, knockdown of the *TGFBI* gene in human corneal epithelial cells brought about a number of differentially expressed genes; hence in response to this knockdown, some genes were overexpressed and others underexpressed. Therefore, from the results of the transcriptome and overexpression analysis carried out in this study, one can observe all the genes and pathways that are directly influenced by the normal *TGFBI* gene.

The corneal epithelial cells of individuals with *TGFBI* CDs, carry a mutated *TGFBI* gene that leads to the production of mutated TGFBI protein. The synthesis of a mutated TGFBI protein can lead to alterations in the folding and stability of the protein, in the interaction with ECM proteins and cell surface molecules as well as in functional levels or the overall role of the protein. This would subsequently cause a number of genes to be differentially expressed; hence, once again, overexpression and underexpression of certain genes would occur. Thus in this scenario, transcriptome analysis of various subtypes of *TGFBI* CD epithelial cells would delineate the differential expression of certain genes that are brought about by these *TGFBI* gene mutations. Apart from genes or pathways that would be directly affected by the impaired production of normal TGFBIp, alterations and deregulation of other pathways or genes might occur as a result of changes induced directly by the mutated protein itself. These might include pathways that influence the degradation and elimination of the abnormal protein, thus resulting in accumulation and precipitation of the abnormal deposits with subsequent visual impairment.

The transcriptome analysis of *TGFBI* KD HCECs performed in this study sheds new light on the key genes that are associated with TGFBIp and provides us with a skeleton of TGFBIp-associated pathways. Additionally, one has to keep in mind that even though a number of expression changes were identified at the molecular level, the actual side effects these changes might have at a functional level in the cornea might be minimal and would still need to be confirmed in future studies. Therefore by using the results of this KD study as a baseline, future studies can compare differentially expressed genes in various *TGFBI* CDs in order to identify changes that would be brought about directly by the abnormal protein itself. This allows us to gain insight

into how *TGFBI* dysfunction impacts gene expression and cellular function, thus leading to a better comprehension of the genetic basis and pathogenesis of the disease.

Furthermore, this study confirmed the efficacy of shRNA-induced knockdown in HCECs. By applying the same principle and performing KD of the mutant *TGFBI* gene in corneal cells of individuals with *TGFBI* CDs, this method can eventually be looked into as a new treatment approach, since theorectically, it can prevent deposit recurrence after corneal transplantation or even inhibit primary corneal deposit formation, hence paving the way for future studies to implement this treatment approach in vivo.

Even though gene therapy is currently considered as the most promising method for the treatment of rare hereditary genetic diseases, possible compounds that were documented in literature to decrease TGFBIp production, were also looked into in this research project. In doing so, additional treatment options that can play a role in the future management of *TGFBI* CDs were taken into consideration and compared to *TGFBI* KD.

4.3.3 A Scoping Review to Identifying and Categorize Compounds that Reduce Corneal Transforming Growth Factor Beta Induced Protein Levels.

To the best of our knowledge, this is the first published review of compounds that can potentially decrease corneal levels of TGFBIp. Such research is essential to promote therapeutic drug identification, including potential repurposing, as has been done in other areas, such as Alzheimer's disease (Ballard, Aarsland, Cummings, *et al*., 2020).

Corneal dystrophies cause substantial visual impairment in affected individuals (Klinthworth, 2009) (Bron, 1990). Most of the articles recalled from the search terms related to *TGFBI* CDs that were applied to the electronic bibliographic databases were found to focus mainly on the clinical, nosology and genetic aspects of these conditions. While these are essential in order to classify and identify novel causative mutations responsible for various CD phenotypes, in order to provide more effective treatment options, it is also crucial to identify less invasive non-surgical treatments to try and prevent or stop the deposition of corneal material in the early stages of these dystrophies.

Medications to treat eye conditions can be administered both topically and systemically. By far, the commonest and least invasive route for the treatment of anterior segment conditions is topical application (Kitamoto, Taketani, Fuji, *et al*., 2020). However, low ocular bioavailability can decrease their efficacy due to the presence of multiple physiological barriers that are present in the cornea. To mention a few: the limited surface area of the cornea, reflex blinking that removes most of the volume applied, the tight junctions between the hydrophobic epithelial cells and the hydrophilic stroma of the cornea, all impede the passive diffusion of drug molecules

(Gaudana, Jwala, Boddu, *et al*., 2009). Furthermore, the unwanted absorption of the drug systemically can result in adverse effects. Recent advances in nanotechnology have contributed to the development of new drug delivery systems to help counteract these issues (Jumelle, Gholizadeh, Annabi, *et al*., 2020). Novel ophthalmic formulations such as those containing tranilast nanoparticles (Nagai, Ono, Hashino, *et al*., 2014), and mesoporous silica nanoparticles for nitric oxide delivery are already being tested (Hu, Sun, Zhang, *et al*., 2018). Other nitric oxide donor delivery systems such as polyamino acid-based *S*-nitrosothiols with high cationic charge density are also being tested to try to address the problem of corneal penetration (Jia, Han, Ha, *et al*., 2020).

My research identified and categorized compounds, tested on human corneal cells, according to their mode of action that were reported in literature to theoretically reduce the levels of mutant TGFBIp in the cornea.

4.3.3.1 Mechanism of Action of Compounds

4.3.3.1.1 Compounds Targeting the TGF-β/SMAD and Akt/GSK-3 Signalling Cascades:

Li, the first natural GSK-3 inhibitor to be discovered, is currently used in the medical field in the treatment of bipolar disorder. Studies carried out by Nie, Peng, Li, *et al*. (2018) and Choi, Kim, Dadakhujaev, *et al*. (2011) on GCD2 fibroblasts revealed that Li suppresses SMAD3/4-dependent gene transcription leading to the downregulation of TGFBIp expression by inhibiting GSK3 directly and also indirectly through interactions with cAMP/PKA and PI3-K/AKT signalling cascades (Figure 34) (Liang, Wendland & Chuang, 2008). The inhibition of GSK3 allows the activation of CREB via increased phosphorylation of CREB and subsequent pCREB-p300 complex formation. The sequestration of p300 by Li-induced increase in pCREB leads to decreased SMAD3-

p300 interaction that in turn results in the blockade of SMAD-dependent transcription (Liang, Wendland, & Chuang, 2008).

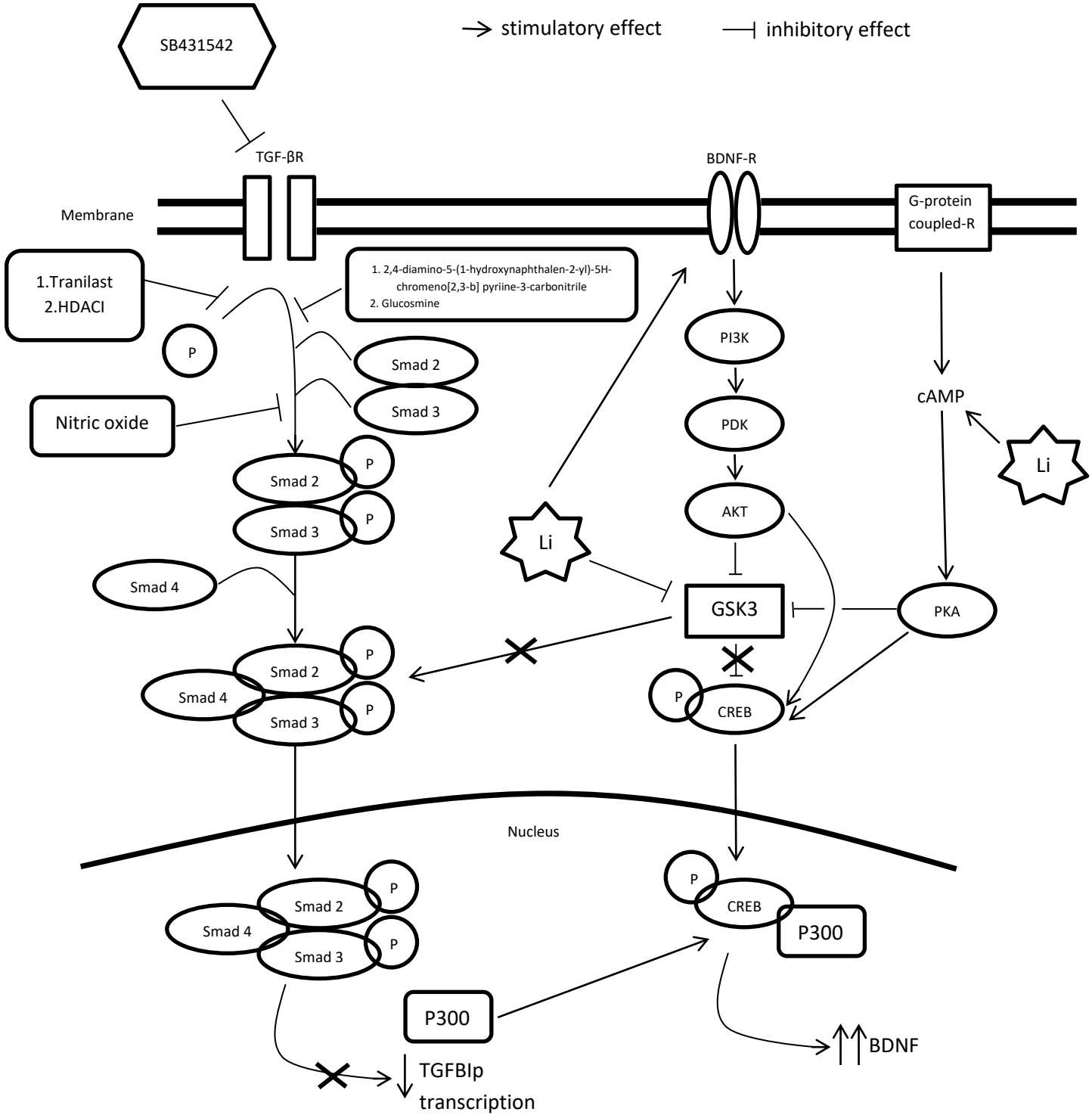


Figure 34: Suppression of TGF-β/Smad induced gene transcription by compounds that alter this pathway. Figure drawn from results obtained from papers of Nie, Peng, Li, et al. (2018), Choi, Kim, Dadakhujaev, et al. (2011), Liang, Wendland & Chuang (2008), Park, Kim, Yim, et al. (2021). Downregulation of TGFBIp expression is modulated by Li, in a direct and indirect way: Li causes an increase in pCREB levels & subsequent sequestration of p300. This causes a decrease in Smad dependent transcription by alterating the smad-p300 complex. TGF-βR is selectively inhibited by SB431542. Phoshorylation of SMAD proteins is suppressed by Tranilast and HDACI while SMAD 3 phosphorylation is reduced by nitric oxide. Smad2 phosphorylation is reduced by 2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b] pyriine-3-carbonitrile & glucosmine. HDACI: Histone deacetylase inhibitor.

To inhibit GSK3 directly, Li competes with magnesium ions for GSK3 binding. On the other hand, it inhibits GSK3 indirectly via two pathways: cAMP-dependent activation of PKA that potentiates activation of CREB and inhibits GSK3; PI3-K/AKT cascade that also causes phosphorylation of CREB and inhibition of GSK3 (Figure 34) (Liang, Wendland, & Chuang, 2008). LiCl also acts on the TGF-β pathway directly by reducing SMAD3 phosphorylation.

To the best of our knowledge, Li has not been tested in vivo on animal models of *TGFBI* CDs. One of the disadvantages of Li is that it has a narrow therapeutic index and thus requires close monitoring of serum levels when it is administered systemically so as to limit adverse effects of treatment. However, exploring topical and local routes of ocular administration is an attractive alternative that might help address the latter issue.

Besides Li, compounds that were documented to result in reduced levels of TGBFIp by acting on SMAD proteins included: tranilast (Kim, Lee, Hong, *et al*., 2015), 2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b] pyriine-3-carbonitrile (Chen, Huang, Tai, *et al*., 2020), histone deacetylase (HDAC) inhibitors: vorinostat (Sharma, Sinha, Siddiqui, *et al*., 2015) and givinostat (ITF2357) (Lim, Tan, Liu, *et al*., 2016), doxycycline (Kim, Luo, Pflugfelder, *et al*., 2005), lobeglitazone (Nuwormegbe, Park, & Kim, 2021), halofuginone (Nelson, Huang, Ewel, *et al*., 2012) (Sarenac, Trapecar, Gradisnik, *et al*., 2016), glucosamine (Chen, Huang, Tai, *et al*., 2019), and nitric oxide (Park, Kim, Yim, *et al*., 2021) (Figure 34).

Tranilast has been marketed as an anti-allergic medication available as an ophthalmic solution and in oral form to be used in Japan, South Korea and China.

When administered systemically, its adverse effects include liver damage, anaemia and kidney failure (Holmes Jr, Savage, LaBlanche, *et al*., 2002). As a result, in 2016 the FDA proposed that it is excluded from the list of substances that can be prescribed in the US (FDA, FDA Proposed Rules, 2016). It is considered an orphan drug in the prevention of scarring following glaucoma surgery, in Europe (EMA, European Medicines Agency). In 2014, Nagai, Ono, Hashino, *et al*. prepared ophthalmic formulations containing tranilast nanoparticles to evaluate corneal toxicity and permeability. They reported increased corneal penetration and better tolerability by human corneal epithelial cells with the nanoparticle formulation (Nagai, Ono, Hashino, *et al*., 2014). Therefore, further studies are needed in order to establish whether any adverse effects are seen with the administration of topical ocular formulations. Measures to reduce systemic absorption of drugs from eye drops such as applying just one drop to the eye and digital nasolacrimal occlusion or eyelid closure for few minutes immediately after application of the drops have been shown to reduce plasma concentrations significantly thus reducing dose-related adverse effects (Flach, 2008).

The efficacy and safety of vorinostat, a histone deacetylase (HDAC) inhibitor, was investigated in the treatment of laser-induced corneal haze in rabbits in vivo (Tandon, Tovey, Sharma, *et al*., 2012). The drug was applied to rabbit corneas once for 5 minutes immediately after refractive laser surgery. It was reported that vorinostat reduced TGF-β 1-induced fibrosis in human and rabbit corneas in vitro, without affecting the viability or proliferation of the cells. In vivo application did not cause redness or inflammation to rabbit eyes. Thus, the authors concluded that vorinostat is non-cytotoxic and safe for the eye (Tandon, Tovey, Sharma, *et al*., 2012). Systemically, vorinostat is used in the treatment of cutaneous T cell lymphoma when the disease

does not improve, progresses or recurs during or after treatment with other medicines. It is the first HDAC inhibitor to be approved for the treatment of cancer and is considered to be a well-tolerated drug (Kumar Bubna, 2015).

Givinostat is a second generation histone deacetylase (HDAC) inhibitor and is being tested on various diseases such as myeloproliferative neoplasms (Passamonti, Maffioli, & Caramazza, 2012), juvenile idiopathic arthritis (Vojinovic, Gabric, Dekaris, *et al*., 2011), inflammatory bowel diseases (Felice, Lewis, Armuzzi, *et al*., 2015) and cardiac fibrosis due to its varied applicability and safety in humans (Milan, Pace, Maiullari, *et al*., 2018). It has also been granted orphan drug status for the treatment of Duchenne muscular dystrophy (EMA, European Medicines Agency). Like vorinostat, givinostat was studied in vivo in a photorefractive keratectomy-treated rabbit model of corneal fibrosis (Lim, Tan, Liu, *et al*., 2016).

Doxycycline is a licenced tetracycline antibacterial and anti-malarial medication. It is also widely used topically or systemically in the treatment of several ocular conditions (not FDA approved) such as ocular/acne rosacea (Snodgrass & Motaparthi, 2021), meibomian gland dysfunction, recurrent corneal erosion (Wang, Tsang, & Coroneo, 2008) and corneal wounds. Animal studies to investigate the effect of doxycycline in alkaline/chemically burnt corneas and dry eye have been performed with favourable outcomes (Yi & Zou, 2019) (Zhang, Yang, Zhu, *et al*., 2014) (Bian, Pelegrino, Henriksson, *et al*., 2016).

Lobeglitazone is currently being investigated as an antidiabetic drug in ongoing clinical trials (Dahlén, Dashi, Maslov, *et al*., 2022) as well as for thyroid cancer treatment (Jin, Han, Ha, *et al*., 2021) amongst others. To the best of our knowledge no

in vivo animal studies have been carried out to examine the effect of lobeglitazone on corneal cells and the study by Nuwormegbe *et al*. (Nuwormegbe, Park, & Kim, 2021) is the only one that was found to be performed on corneal cells.

Halofuginone is mainly used in veterinary medicine. It was given orphan drug status by EMA for the treatment of systemic sclerosis and Duchenne muscular dystrophy. However they were both withdrawn at the request of the sponsor. No in vivo animal studies of the administration of halofuginone on the cornea have been performed to date.

Glucosamine is widely used as a dietary supplement. The only article that was found where glucosamine was tested supposedly in vivo on a model of post-traumatic keratitis was written in Russian with no access to the full report. In the abstract it was reported that glucosamine hydrochloride stimulated repair of dystrophic post-traumatic processes in the connective-tissue structures (Zupanets, Bezdetko, Dedukh, *et al*., 2002).

Nitric oxide is known to cause relaxation of smooth muscle cells leading to dilation of blood vessels and a decrease in blood pressure. Due to its mode of action, nitric oxide is used in the respiratory and cardiovascular fields (EMA, European Medicines Agency). Nitric oxide is lately being evaluated for its potential benefits in the treatment of glaucoma (Aliancy, Stamer, & Wirostko, 2017). To overcome the issue of corneal penetration, various nitric oxide donor delivery systems are being tested such as mesoporous silica nanoparticles (Hu, Sun, Zhang, *et al*., 2018) and polyamino acid-based S-nitrosothiols with high cationic charge density (Jia, Han, Ha, *et al*., 2020).

Okumura *et al*. (Okumura, Kay, Nakahara, *et al*., 2013) studied the effect of SB431542, a selective inhibitor of the TGF-βR, on human corneal endothelial cells. They reported that this compound blocked the transformation of the endothelial cells to fibroblastic phenotypes. This led to the reduction in the production of fibrillar ECM proteins. However, the aim of the study by Okumura, *et al*. (Okumura, Kay, Nakahara, *et al*., 2013) did not include investigating the effect of SB431542 on downstream pathways. This study has been included in this review since direct inhibition of the TGF-βR will theoretically result in inhibition of the SMAD cascade. To the best of our knowledge, no in vivo animal studies have been carried out to study the effect of SB431542.

Lycium barbarum polysaccharides (LBP) has been reported to attenuate TGF-β 1-induced fibrosis in primary human corneal keratocytes. However, the underlying mode of action was still inconclusive (Kwok, Wong, Shih, *et al*., 2020). Thus, this paper was excluded in the final count of this scoping review since the decrease in fibrosis could be the result of modulation of other downstream pathways involved specifically in fibrosis, other than the TGF-β/SMAD, JNK, PI3-K/AKT and the cAMP/PKA cascades.

4.3.3.1.2 Compounds Targeting the JNK Signalling Cascade:

The compound SP600125 (selective ATP-competitive inhibitor of c-Jun N-terminal kinase JNK) (Nie, Liu, Wang, *et al*., 2020) and doxycycline (Kim, Luo, Pflugfelder, *et al*., 2005) act by inhibiting JNK signalling activation that in turn suppresses phosphorylation of SMAD2/3 and TGFBIp accumulation. SP600125 has also been used in mouse models of corneal keratitis to investigate the host inflammatory response (Zhang, Meng, Liu, *et al*., 2018).

4.3.3.1.3 Compounds Inhibiting DNA Synthesis and Function:

In a study by Kim *et al.* (Kim, Choi, Lee, *et al*., 2008), it was shown that exposure of normal and GCD2 keratocytes to MMC reduces *TGFBI* mRNA levels as well as TGFBIp levels. MMC is an alkylating agent with antineoplastic properties that selectively inhibits the synthesis of DNA (Barnett & Brundage, 2010). Systemically it is administered in the treatment of gastrointestinal cancers and topically for bladder and intraperitoneal tumours (EMC, Electronic Medicines Compendium). It is widely used for the prevention of scarring post glaucoma (Ichhpujani, Singla, Kalra, *et al*., 2022) and pterygium surgery and is recently being evaluated for its in vivo effect on corneal haze post-photorefractive keratectomy (Ouerdane, Zaazouee, Mohamed, *et al*., 2021) (Mohan, Gogri, Murthy, *et al*., 2021).

4.3.3.1.4 Compounds Inducing Increased Elimination of TGFBIp:

On the other hand compounds that were documented to decrease the level of TGFBIp by increasing its degradation included: 4-PBA (Choi, Lee, Jeong, *et al*., 2016), melatonin (Choi, Kim, Oh, *et al*., 2013) (Choi, Lee, Akuzum, *et al*., 2017) and Torin1 (Wang, Zhao, Zheng, *et al*., 2021). The former two worked by increasing Endoplasmic-reticulum-associated protein degradation (ERAD) activity, a pathway that leads to the elimination of unfolded proteins. On the other hand, Torin 1 was shown to improve lysosomal function leading to increased autophagy and degradation of mutant TGFBIp in Thiel-Behnke CD.

4-PBA is a histone deacetylase inhibitor that is currently used to treat urea cycle disorders (Kolb, Ayaub, Zhou, *et al*., 2015). The mode of action and effectivity of 4-PBA has also been investigated in a study conducted on a mouse model where ocular eye

drops of PBA were shown to lower significantly elevated intraocular pressure (Maddineni, Kasetti, Kodati, *et al*., 2021).

Melatonin is a hormone that is produced by many tissues in the body, including the eye. Systemically is mainly used in the management of insomnia (EMA, European Medicines Agency). Recently it has been investigated as a potential drug in the management of dry eye (Wang, Zhao, Zheng, *et al*., 2021) and the control of intraocular pressure (Lundmark, Pandi-Perumal, Srinivasan, *et al*., 2007). Melatonin receptor agonists dissolved in formulations already in use for other ocular topical treatments have also been tested on in vivo animal models with favourable results (Andrés-Guerrero, Alarma-Estrany, Molina-Martínez, *et al*., 2009). Navarro-Gil *et al*. (Navarro-Gil, Huete-Toral, Domínguez-Godínez, *et al*., 2022) also investigated the use of contact lenses loaded with melatonin analogs as an alternative drug delivery system.

4.3.3.1.5 Compounds that Bind to Mutant TGFBIp:

In a study carried out by Venkatraman *et al*. (Venkatraman, Duong-Thi, Pervushin, *et al*., 2020) small molecule compounds that bind to mutant TGFBIp were identified from Maybridge RO3 fragment library of 2500 compounds. The authors postulated that binding of a chemical modulator to the mutant TGFBIp would limit its processing into smaller fragments and thus limit aggregation and subsequent deposition. The lead compounds reported to be effective in delaying or preventing the generation of amyloidogenic peptides in the R555W mutant were MO07617, RJF00203 and, BTB05094 while those reported to delay/prevent the generation of amyloidogenic peptides in the H572R mutant were the compounds RJF00203 and BTB05094.

4.3.3.2 Animal Models

The generation of physiologically relevant animal models is imperative in the understanding of the mechanisms leading to the phenotypic expression of *TGFBI* CDs. Besides, animal models can be used to test the relevance of potential drugs that would have been previously tested in culture. Only few attempts have been made to develop physiologically relevant CD animal models. However, quite recently, Kitamoto *et al*. (2020) managed to generate mutant mice using the CRISPR/Cas9 approach with the *TGFBI* R124C mutation, which actually exhibited the LCD phenotype.

4.3.3.3 Drug Repurposing and Choice of Compounds

CDs are a rare group of conditions and, to date, there is still a lack of research that investigates the prevalence, pathogenesis and possible treatment of these conditions. More research focussing on non-surgical procedures to try and prevent or stop the deposition of corneal material in these dystrophies is needed. Drug repurposing and repositioning offers the advantage that the compound would have already been amply studied and its safety profile established. Thus, this method can be applied in order to reduce considerably the cost and time required for these compounds to be progressed to the stage of clinical trials. To summarize, Li, doxycycline, nitric oxide and MMC are currently licenced for use in the treatment of various medical conditions by both the EMA and FDA. Vorinostat is also in the list of licenced drugs by the FDA. On the other hand, glucosamine and melatonin are widely available, as dietary supplements, both in the USA and in Europe. Currently, tranilast and givinostat are considered as orphan drugs by the EMA while vorinostat and halofuginone had also been granted orphan drug status in the past, however, this was withdrawn on request of the sponsor. A summary of pharmaceutical compounds that have been shown in literature to reduce

the expression of *TGFBI* or decrease the amount of TGFBIp in the cornea, therefore showing potential to progress to the stage of clinical trials is shown in Table 49.

In this study, from all the previously mentioned compounds, Li and MMC were chosen to be the compounds the cultured HCECs were exposed to in this phase of the study. Li was chosen since it was the only compound licensed to be used in Europe that was tested on fibroblasts in GCD2, a subtype of *TGFBI* CDs. Furthermore, researchers strongly suggested that it might be possible to implement it as a treatment in the management of *TGFBI* CDs (Choi, Kim, Dadakhujaev, *et al*., 2011) (Nie, Peng, Li, *et al*., 2018). On the other hand, MMC, apart from also being licensed in Europe, is routinely used in ophthalmic surgery especially during glaucoma filtration procedures and in the prevention of scarring during refractive laser surgery (Ichhpujani, Singla, Kalra, *et al*., 2022).

However, there are a number of studies that suggested a potential toxic effect of MMC on corneal cells. Certain studies exploring the effects of MMC during PRK reported poorly differentiated epithelial cells and delay in wound healing which some authors debated this outcome by attributing post-operative steroid use as the cause (Arranz-Marquez, Katsanos, Kozobolis, *et al*., 2019). Furthermore, in another study carried out by (Mohammadi, Ashrafi, Norouzi, *et al*., 2014) the use of MMC for only a short period of time failed to elicit significant epithelial toxicity. Having said that, one must be careful of its use with regards to dosage and exposure time (Lalitha, Meena, & Pallavi, 2021). Therefore, the effect of MMC on HCECs was worth further investigation.

Table 49: Pharmaceutical compounds studied that were shown to reduce the expression of TGFBI or decrease the amount of TGFBIp in the cornea.

Pharmaceutical Compound	**Authors**	**Mode of action**	**Model/phase**	**Outcome**
Li chloride	Ni, *et al*., 2018. Choi, *et al*., 2011.	Reduces SMAD3 phosphorylation. Negatively regulates GSK-3beta.	In vitro	Inhibits the expression of *TGFBI* and enhances autophagy in fibroblasts.
Tranilast	Kim, *et al*., 2015.	Reduces levels of pSMAD2 and pSMAD3 in WT and GCD2 fibroblasts.	In vitro	Inhibited the expression of *TGFBI* by blocking TGF-β signalling.
2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b] pyriine-3-carbonitrile	Chen, *et al*., 2020.	Reduces SMAD2 phosphorylation. Upregulates mRNA of SMAD2 and downregulates mRNA of SMAD3 in human corneal fibroblasts.	In vitro	Anti-fibrotic and anti-inflammatory.
Vorinostat	Sharma, *et al*., 2015.	Transcriptional repressors of TGF-β1 signalling via the SMAD pathway.	In vitro	Antifibrotic effect.
Givinostat (ITF2357)	Lim, *et al*., 2016.	Suppresses phosphorylation of SMAD proteins in TGF-β pathway.	In vitro and in vivo animal model	Decreased collagen I, collagen IV, fibronectin, integrin αVβ3 expression and it reduced myofibroblast formation.

Doxycycline	Kim, *et al*., 2005.	Inhibits activation of SMAD2 signalling pathways. Inhibits JNK signalling activation.	In vitro	Inhibits TGF-β1–induced MMP-9 production and activity.
Lobeglitazone	Nuwormegbe, Park, & Kim, 2021.	Inhibits the TGF-β 1-induced SMAD signalling.	In vitro	Attenuates TGF-β1-induced ECM synthesis and myofibroblast differentiation of corneal fibroblasts.
Halofuginone	Nelson, *et al*., 2012. Sarenac, *et al*., 2016.	Reduces protein expression of SMAD3 leading to disruption of TGF-β signalling.	In vitro	Halofuginone reduced the expression of α-SMA, fibronectin, and type I collagen.
Glucosamine	Chen, *et al*., 2019.	Reduces SMAD2 phosphorylation.	In vitro	Suppresses TGF-β1-induced fibroblast-to-myofibroblast differentiation.
Nitric oxide	Park, *et al*., 2021.	Reduces SMAD3 phosphorylation.	In vitro	Reduces SMAD3 phosphorylation.
Melatonin	Choi, *et al*., 2013.	Reduces level of TGFBIp by increasing its degradation by inducing ERAD pathway activation.	In vitro	Activates autophagy in both wild-type (WT) and GCD2-homozygous (HO) corneal fibroblast cell lines and also reduces ER stress.
Mitomycin C	Kim, *et al*., 2008.	MMC reduces *TGFBI* mRNA levels as well as TGFBIp levels.	In vitro	Reduced the production of TGFBIp in normal, heterozygote and homozygote (GCD2) corneal fibroblasts. It induced apoptosis in GCD II homozygote cells.

4.3.4 Comprehensive Transcriptome Analysis to Investigate the Differential Expression of Genes between Normal HCECs and HCECs Exposed to Li and MMC: Implications for TGFBI CD Treatment.

Studies have suggested that since TGFBIp is robustly expressed in the adult corneal epithelium, it is secreted and then transported from the epithelium into the corneal stroma (Malkondu, Arıkoğlu, Erkoç, *et al*., 2020) (Niu, Liu, Liu, *et al*., 2012) (Poulsen, Runager, Nielsen, *et al*., 2018). The distribution of the normal pattern of TGFBIp and that of the dystrophic aggregates is similar, with the largest concentration found in Bowman region. Expression of the *TGFBI* gene has been found to be very low in adult corneal stromal cells (Streeten, Qi, Klintworth, *et al*., 1999), however, in response to injury or surgery, it has been reported that TGFBIp production is increased in both corneal epithelial cells and stromal fibroblasts (Aldave, Sonmez, Forstot, *et al*., 2007). *TGFBI* expression seems to be modulated by various pathways and different molecules depending in the cell type. Thus, more research is needed to investigate further the modulation of its expression in corneal epithelial cells.

The large majority of studies carried out on corneal cells exploring the effect of compounds on TGFBIp production have been performed on corneal stromal fibroblasts. Studies reported that *TGFBI* expression in corneal fibroblasts is controlled directly by the TGF-β signalling pathway (Choi, Yoo, Kim, *et al*., 2010) (Nie, Liu, Wang, *et al*., 2020) and indirectly via the JNK signalling cascades (Liu, Sheng, Peng, *et al*., 2018) the PI3-K/AKT and the cAMP/PKA cascades (Shen, Hu, Liberati, *et al*., 1998) (Piersma, Bank, & Boersema, 2015). In fibroblasts it was reported that Li downregulates TGFBIp expression by inhibiting GSK3 directly and indirectly (Nie, Peng, Li, *et al*., 2018) (Choi,

Kim, Dadakhujaev, *et al*., 2011) while MMC reduces TGFBIp levels by inhibiting DNA synthesis (Kim, Choi, Lee, *et al*., 2008).

In order to determine whether there are differences in pathway activation between corneal epithelial cells and fibroblasts, Guo *et al*. studied the regulation of TGF-β target genes, which included thrombospondin-1, in these two cell types. Surprisingly and interestingly they reported that not only did the pathways responsible for the expression of TGF-β-target genes vary for different proteins, but also that different pathways were used to modulate the expression of the same protein they studied, thrombospondin-1, in epithelial cells and fibroblasts. Consequently, they came to the conclusion that not all TGF-β-target proteins are induced through the SMAD pathway, and that corneal epithelial cells and fibroblasts use alternative pathways for TGF-β-target protein expression. They also hypothesised that an internal check mechanism might occur where the continued expression of TGF-β-target proteins could be blocked since there have been studies showing that some of the SMAD proteins are downregulated by TGF-β stimulation. Therefore, the expression of certain genes could be limited in one corneal cell type but not in the other (Guo, Hutcheon, Tran, *et al*., 2017). Furthermore, it is well known that TGF-β1 inhibits the proliferation in certain cell types, such as epithelial cells, while acting as a proliferative inducer of others, such as corneal stromal fibroblasts (Ten Dijke, Goumans, Itoh, *et al*., 2022).

Differences in gene expression and metabolic pathway activation between corneal epithelial cells and fibroblasts were also documented in research carried out by Yam *et al*. (Yam, Fuest, Zhou, *et al*., 2019). Here they found that changes in the epithelial proteome centred around cell metabolism and mitochondrial involvement, whereas

the stromal proteome showed more changes in cellular assembly and tissue organization.

In view of the above proven facts, in this study it was decided to explore the effect of Li and MMC on human corneal epithelial cells, rather than on fibroblasts so as to mirror better the actual biological environment in the human cornea.

4.3.4.1 Downregulated Differentially Expressed Genes in Li and MMC Exposed HCECs

The downregulated genes that were found to have the largest log_2FC were *UCA1, POTEF, MAGEA4, LAMA4, USP9Y* and *RPS4Y1* in Li treated cells. Interestingly, *POTEF, UCA1, MAGEA4, USP9Y* and *RPS4Y1* were also on top of the list for MMC. In MMC the other top downregulated gene was found to be *IRX4*. Most of these top downregulated genes such as *UCA1, POTEF* and *MAGEA4* are all known to play a regulatory role in the cell cycle. *UCA1* produces a long non-coding RNA that plays a regulatory role in cell proliferation while *MAGEA4* promotes growth by precluding cell cycle arrest and by inhibiting apoptosis mediated by the p53 transcriptional targets. On the other hand *POTEF* encodes proteins that form part of a family of proapoptotic proteins. In a study carried out on cultured retinal ganglion cells it was documented that downregulation of *LAMA4* might reduce oxidative stress-induced apoptosis by inhibiting the activation of the MAPK signalling pathway (Wang, Ren, Zhai, *et al*., 2019). The fact that the top downregulated genes in both Li and MMC groups were so similar, both with very large log_2FCs, might indicate that these differentially expressed genes were downregulated in relation to the cell's response to an exogenous factor and might not be directly related to the effect of the specific drug in question.

On further analysis, the 'genes of interest' that were found to be significantly downregulated when HCECs were exposed to Li or MMC were *COL6A1, COL6A2, COL5A1, S100A9, RPS27a, APOD, ESR1, FLOT1, FN1, S100P, IL1B, PIK3CA, JAK2, PTK2, SMAD3* and *ITGA11*.

Four of these 'genes of interest', namely *S100A9, IL1B, JAK2* and *ITGA11* were also noted to be downregulated in the *TGFBI* KD HCECs. On the other hand *COL6A1, COL6A2* were found to be upregulated in *TGFBI* KD HCECs while Li or MMC exposure caused their downregulation. *TGFBI* has a role as a linker protein in facilitating cell-collagen interactions and is known to bind covalently to collagen VI (Reinboth, Thomas, Hanssen, *et al*., 2006).

TGF-β can induce ECM synthesis through the SMAD signalling pathway. SMAD heterogenous oligomer binding site (SBE) has been identified in promoters of many genes such as *FN*, that encodes fibronectin, and *COL6*, that encodes collagen VI (Zhang, Topley, Ito, *et al*., 2005) with TGF-β_1 stimulation of various cell types being associated with deregulation of various genes encoding for collagens and glycoproteins such as *COL5, COL6, FN1* and *TGFBI,* amongst others (Zhang, Topley, Ito, *et al*., 2005). *COL6, COL5* and *FN1* were all found to be downregulated in Li and MMC exposed HCECs. Besides being stimulated by a common cytokine, *TGFBI* is known to interect with ECM proteins such as collagen and fibronectin in the process of mediating integrin binding.

Collagen VI, in addition to its structural role, affects various cellular functions including adhesion, migration and cytoprotective functions which include cell cycle progression by counteracting apoptosis and oxidative damage as well as regulating autophagy (Cescon, Gattazzo, Chen, *et al*., 2015). In line with this, *COL6A1* KD has been

reported to suppress cell proliferation in human vascular smooth muscle cells (Chen, Wu, Yan, *et al*., 2019). *COL5A1* encodes an alpha chain of collagen V, and similarly to collagen VI it functions as an extracellular matrix structural constituent. Collagen V is known to bind to receptor tyrosine kinase that plays a role in the development and growth of organs. It has also been reported to interact with integrins. Of interest is that collagen V was found to suppress nuclear translocation of the transcription factor YAP. While collagen V was found to be downregulated by Li exposure in this study, *YAP* was in turn found to be upregulated. In pancreatic beta cells, collagen V has been associated with inhibition of proliferation via integrin β1 and E-cadherin/β-catenin signalling pathways (Zhu, Chen, Liu, *et al*., 2021). *RPS27a*, another gene found to be downregulated in Li and MMC treated cells, is a ribosomal protein that is also involved in cell cycle progression.

Fibronectin has been found to be present on the cell surface and in the extracellular matrix. *TGFBI* has been reported to mediate integrin binding to ECM proteins including fibronectin (Kim, Kim, Lee, *et al*., 2000) that subsequently play a role in cell adhesion, migration, differentiation and fibrosis (Dou, Liu, Lv, *et al*., 2023) (Kim, Kim, Lee, *et al*., 2000). In line with this, downregulation of *FN1* has been associated with suppressing proliferation and inducing apoptosis in colorectal carcinogenic and melanoma cells (Cai, Liu, Zhang, *et al*., 2018)**.**

This study revealed that GO annotation descriptions obtained by ORA of the downregulated genes as a result of exposure of HCECs to Li or MMC generally cause an effect on regulation of intracellular signal transduction, regulation of molecular function as well as regulation of gene expression and transcription. ‘Response to

cytokine' and 'cytokine mediated signalling pathway' , 'regulation of intracellular signal transduction', 'DNA-binding transcription factor activity' and 'regulation of cell differentiation' were amongst the most common annotations that cropped up. These annotations were mostly associated with the genes *S100A9, ESR1, FLOT1, FN1, IL1B*, *PIK3CA, JAK2* and *PTK2*.

The KEGG and Reactome pathways identified to be enriched in Li treated HCECs were the NOD-like receptor signalling pathway, TNF signalling pathway, NF-kappa B signalling pathway, Interferon Signalling, JAK-STAT signalling pathway, FoxO signalling pathway and Cytokine Signalling in Immune system as well as Autophagy and Endocytosis.

On further in-depth analysis of these downregulated genes, *S100A9* has been documented to regulate inflammatory and immune responses and activate downstream pathways that include NF-kappaB and MAPKs (Stelzer, Rosen, Plaschkes, *et al*., 2016). *JAK2* and *IL1B* are also known to be major modulators in signalling pathways including JAK-STAT and NF-kappa B signalling, the JNK and p38 MAPK pathways respectively (Weber, Wasiliew, & Kracht, 2010). IL1B is a potent pro-inflammatory cytokine and as already noted above, TGF-β and IL1 can have both opposing and supporting roles (Wilson S. , 2021). The *PTK2* gene encodes a cytoplasmic protein tyrosine kinase that is located in the focal adhesions present among cells attaching to extracellular matrix components. It has been reported to be involved in the integrin pathway, in the activation of PI3K/Akt signalling cascade as well as the MAP kinase signalling cascade (Ladewig, Michelini, Jhaveri, *et al*., 2022). Similarly, PIK3CA is also involved in the PI3K/Akt/mTOR signalling pathway. The

PI3K/Akt pathway is activated downstream of a many of growth factor receptors, which induce proliferation and stimulate growth and survival of cells (Fruman, Chiu, Hopkins, *et al*., 2018).

The gene *ITGA11* encodes for integrin subunit alpha 11 and is a receptor for collagen. Since it has been reported to exhibit enhanced expression in scarred keratoconus corneas, downregulation by Li theoretically might lead to decreased scar formation in the cornea (Bystrom, Carracedo, Behndig, *et al*., 2009). To the best of my knowledge, integrins are the only *TGFBI* cell surface receptors identified to date. The interaction of *TGFBI* with cell surface integrin receptors is intricate, and is probably cell-type specific (Thapa, Lee, & Kim, 2007).

In summary, the results obtained from molecular and enrichment analysis of downregulated genes in HCECs exposed to Li or MMC gives us a mixed picture. On one side, we have a number of genes which, when downregulated, decrease cell proliferation such as *UCA1, MAGEA4, S100A9, COL6, COL5* and *FN1* while, on the other hand, downregulation of genes that usually induce apoptosis such as *POTEF* and *LAMA* will theoretically allow increased cell proliferation. Other downstream effects point towards a downregulation of inflammation and scarring due to the downregulation of *JAK2, IL1B* and *ITGA11*.

4.3.4.2 Upregulated Differentially Expressed Genes in Li and MMC Exposed HCECs

The upregulated genes that were found to have the largest $\log_2$FC were *KRT14*, *MEG3, CA12, CCDC8, XIST* and *RTL1* in Li treated cells. Similarly to what was noted in the downregulated gene sets, *KRT14, MEG3, CA12, CCDC8, XIST* were also on top of

the list for MMC. The similarity in the top upregulated genes in the two groups also points towards the possibility of these being a response to exogenous factors.

However, having said that, some of these genes exhibit interesting functions in relation to this study. The gene that showed the largest positive log_2FC change, *KRT14,* encodes a member of the keratin family, which function as intermediate filaments. Keratin 14 has been reported to be present all over the cornea and predominantly in the basal epithelial layers (Ligocki, Fury, Gutierrez, *et al*., 2021). Interestingly, increased corneal K14 expression has been reported in Meesmann epithelial CD which is usually associated with mutations within the *KRT3* or *KRT12* genes (Allen, Courtney, Aktinson, *et al*., 2016). In the skin it has been noted that following *TGFBI* treatment, wound healing was significantly enhanced with significant increases in the expression of *KRT14* thus exhibiting a positive correlation to each other (Li, Zhang, Huang, *et al*., 2022).

MEG3, another highly expressed gene, is a long noncoding RNA that has been found to modulate the activity of TGF-β genes by binding to distal regulatory elements (Mondal, Subhash, Vaid, *et al*., 2015).

On further analysis, the 'genes of interest' that were found to be significantly upregulated when HCECs were exposed to Li, and similarly MMC in this study , were *TGFBI , COL1A1, COL1A2, COL5A2, COL12A1, COL4A1, COL4A2, ENO1, ACTB, CLU, LDHA, DSP, CSTA, VIM, S100A6, S100A4, FABP5, ANXA2, ESR2, ITGA3, ITGA6, ITGB5, ITGAV, LOXL2, MMP2, NOG, PLOD2, THBS1, SMAD7, CUL4A, ALDH3A1, CPA4, MAMDC2, YAP1* and *FAT4*.

Li has been reported to modulate a number of cellular signalling pathways including adenylyl cyclase/cAMP, PKC, GSK3, BDNF and Bcl-2 systems. It has been reported to inhibit PKC, GSK3 while activating CREB (a downstream target of cAMP) which in turn facilitates the production of BDNF and Bcl-2 in neural cells (Malhi, Tanious, Das, *et al*., 2013) (Liang, Wendland, & Chuang, 2008). The mode of action of Li, which after 12 hours of exposure has been reported to lead to the downregulation of TGFBIp expression in corneal fibroblasts, has been attributed to the inhibition of GSK3 through interactions with cAMP/PKA and PI3-K/AKT signalling cascades leading to suppression of SMAD3/4-dependent gene transcription (Choi, Kim, Dadakhujaev, *et al*., 2011).

Contrastingly, in this study, exposure of HCECs to Li for 72hours resulted in a significant upregulation of *TGFBI* instead of a downregulation. While Li was reported to inhibit GSK3 in fibroblasts, here we find that *GSK3a* and *GSK3b*, the two isoforms of GSK-3 with similar substrate specificity and functions, were found to be mildly upregulated, however with a p value of 0.4. Similarly, *SMAD2* and *SMAD4* were also noted to be mildly upregulated with a p value of 0.2 and 0.7 respectively. On the other hand, *SMAD3* and *SMAD7*, which encodes an inhibitory SMAD, were found to be significantly downregulated and upregulated respectively.

Additionally, in the Li and MMC treated HCECs, a number of genes involved in the PI3K/Akt signalling pathway were found to be significantly differentially expressed. Upregulation of *PIK3CD* and *PIK3R1* was noted, while, on the other hand, *PIK3CA*, *PIK3CB* and *PIK3C2B* were found to be downregulated. *AKT1* was noted to be mildly upregulated but not at a 0.05 level of significance. On an interesting note, researchers

have also postulated that *TGFBI* has a suppressive effect on PI3K/Akt/mTOR signalling pathway via decreased binding of PI3K to EGF receptor (Wen, Hong, Li, *et al*., 2011).

The DE analysis of Li treated HCECs in this study also revealed the upregulation of *YAP1* and *TEAD4*, a primary mediator and transcription factor respectively, of the Hippo pathway. YAP1 phosphorylation has been reported to be triggered by TGFβ signalling. Phosphorylated YAP1 forms a complex in the nucleus that leads to the induction of expression of early target genes. Subsequently, when the TGFβ signal decreases, YAP1 associates with TEAD4 in the nucleus to form the YAP1/TEAD4/SMAD3/p300/AP1 complex with subsequent induction of late target gene expression (Kim, Han, Park, *et al*., 2023). *YAP* has been reported to promote *AP-1* expression in tubular epithelial cells (Liu, Xu, Li, *et al*., 2023) and *AP-1* has in turn been found to promote the transcription of *SMAD7* (Brodin, Ahgren, Dijke, *et al*., 2000). SMAD7 has also been implicated in the mediation of crosstalk of TGF-β signalling pathway with other pathways besides reports of it being transcriptionally induced by TGF-β/BMPs signalling and inhibiting their signalling by negative feedback. Various research reports suggest that the role of SMAD7 is context dependent since although it is a well documented antagonist of TGF-β, it has been shown to induce TGF-β-mediated apoptosis in certain cell lines while inhibiting this process in others.

Similarly, opposite roles of YAP1 and TGFβ have been documented which have been attributed to the fact that their functions are probably determined by the cell line and context (Brodin, Ahgren, Dijke, *et al*., 2000).

In the cornea, the ordered network of collagen consists mainly of lamellae of collagens type I and type V, intertwined with collagen type VI. Furthermore, it has

been proposed that one of the significant physiological roles of TGFBIp is to bind covalently to collagen type XII (Hanssen, Reinboth, & Gibson, 2003). In this study, upregulated DE genes of interest included *COL1A1, COL1A2, COL5A2* as well as *COL12A1*, the deregulation of which might cause an effect on corneal structure and ECM organization.

Besides being stimulated by a common cytokine TGF-β, *TGFBI* contains an RGD-containing protein that is known to bind to collagen type I and IV (Patricelli, Lehmann, Oxford, *et al*., 2023) (Silbiger, Lei, Ziyadeh, *et al*., 1998). *TGFBI* is known to interect with ECM proteins such as collagen in the process of mediating integrin binding. It has been reported that silencing *COL1A1* downregulated the expression levels of *TGFBI* in bladder epithelial carcinoma cells showing that *COL1A1* and *TGFBI* interacted closely with each other (Zhu, Chen, Wang, *et al*., 2019). PLOD2 gene, another upregulated gene in Li and MMC exposed HCECs, encodes a membrane-bound enzyme that is critical for the stability of the intermolecular collagen cross-links. Its related pathways include collagen chain trimerization and extracellular matrix organization. On the other hand, *MMP2*, also found to be upregulated, encodes a metalloproteinase that is involved in remodeling and degrading of extracellular matrix proteins.

The binding of *TGFBI* to integrins has been related to the activation of cell proliferation, adhesion, migration and differentiation (Kim, Kim, Lee, *et al*., 2000). Integrin genes which were found to be upregulated here included *ITGA3* and *ITGA6*. These genes encode alpha chains of the integrin family of proteins that function as cell surface adhesion and signalling molecules. *ITGA3* also functions as a receptor for fibronectin, laminin and collagen. *ITGAV* and *ITGB5* also form part of the integrin

cell-surface receptor family and function as a receptor for fibronectin. Similarly, *THBS1*, which was also upregulated, encodes a multifunctional adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions by binding to fibrinogen, fibronectin, laminin, type V collagen and integrins alpha-V/beta-1, amongst others.

LOXL2, another gene found to be upregulated in the Li and MMC treated HCECs, forms part of the lysyl oxidase (*LOX*) gene family, which encode enzymes that covalently cross-link collagens and elastin. This cross-linking reaction provides additional mechanical strength to the ECM and makes it more resistant to degradation. In trabecular meshwork cells, *LOX* has been found to be regulated by the TGFβ/SMAD as well as by the JNK/AP-1 signalling pathway (Sethi, Mao, Wordinger, *et al*., 2011).

Other DE genes involved in signalling pathways included the *CLU* gene that functions as extracellular chaperone that prevents aggregation of non native proteins, inhibits formation of amyloid fibrils and modulates NF-kappa-B transcriptional activity, as well as *ESR2* that is known to play a role in modulation of PI3K/AKT signalling. Additionally, *MAPK1* gene encodes a serine/threonine kinase, which is an essential constituent of the MAP kinase signal transduction pathway that is involved in multiple biological functions including cell growth, adhesion, survival and differentiation through the regulation of transcription, translation and cytoskeletal rearrangements.

Other upregulated genes were *S100A4* and *S100A4*. *S100A4* is a paralog of *S100A6. S100A6* encodes a calcium-binding protein that modulates various cellular processes including motility, angiogenesis, cell differentiation, apoptosis, and autophagy. It also plays a role in the pro-apoptotic function of TP53 by binding to its C-terminal transactivation domain within the nucleus and reducing its protein levels.

In summary, the results obtained from the biological and molecular function as well as pathway enrichment analysis of upregulated genes in HCECs exposed to Li and MMC, centered on extracellular matrix structural constituents especially collagens and related genes, cell adhesion molecule binding such as integrins as well as cell signalling pathways.

4.3.4.3 DE Genes Identified in Li Exposed HCECs but not in MMC Exposed Cells

Downregulated

Even though a large number of DE expressed genes in Li and MMC treated HCECs were identified to be common to both, a significant number of DE genes specific to each group were also identified.

A large number of transcription factors involved in the regulation of DNA binding and differentiation, especially related to neuronal cells, were enriched in ORA analysis of the downregulated genes in the Li treated HCECs. These included multiple genes encoding zinc finger proteins as well as other transcriptional regulators such as *HES7, LHX6, ONECUT2, PITX1, WT1* and *PBX1*.

A large number of upregulated genes were also linked to DNA binding and transcriptional regulation, namely, *GBX2, NR2C2, E2F2, HIRA, MAFB, PROX1, NFE4, HIF3A, ESRRG, MLLT1, HES2, ZSCAN23, EMX2, ZSCAN4, HEY2, POU4F2* and *TRIM29*, as well as a number of Zinc finger proteins. CREB5, the function of which is to activate transcription was also found to be upregulated. It is also known to be involved in the PI3K-Akt signalling pathway.

Genes involved in neuronal migration, axon guidance regulating the development and maintenance of the nervous system, namely *CNTN2, LHX6, SEMA6B* and *SEMA4G*, were also found to be downregulated in Li treated HCECs.

A number of downregulated DE genes crucial in intracellular signalling pathways were also shortlisted. *WNT4, WNT3* and *TCF7* are involved in the Wnt signalling pathway that plays a central role in the regulation of cell homeostasis and differentiation. The *PIK3CB* gene encodes the catalytic subunit of the PI3K enzyme while *PIK3R3* encodes a regulatory subunit of the PI3K complex, both of which are involved in the PI3K/AKT cell signalling pathway that regulates cell growth, proliferation, and survival. Genes encoding proteins involved in the NF-κB signalling and Notch signalling were also found to be differentially expressed.

Another group of downregulated genes found to be enriched in ORA analysis were those related to cell cycle regulation: *PPP2R3C, SPDYA, INCA1, CCNO, CCNG1, PP2R3C*; as well as immune and inflammatory responses: *IRF1, LGALS9, LITAF, CHAD*.

Upregulated

A significant number of upregulated DE genes were noted to play important roles in various metabolic pathways such as cholesterol biosynthesis, glycolysis / gluconeogenesis fatty acid metabolism, and nucleotide metabolism. These included *MVK, HMGCS1, PMVK, ACOT1, ACOT6, ACOT4, ENPP1, NUDT7, B4GAT1, ATP5MC3, ITPKB* and *SMPD3* amongst others.

A good number of upregulated genes were also found to be involved in transmembrane transporter activity. These comprised genes encoding for members of

the solute carrier (SLC) group of transport proteins as well as others, such as *ABCC4, CACNB1, ATP13A2, KCNJ9, PKD2L2, GRIN3B, SCN2B, FXYD4* and *GLRA2*.

Interestingly, *CHST14* and *VCAN* genes were also found to be upregulated. These are involved in the biosynthesis of dermatan sulphate proteoglycans, a group of proteins known to be present in the corneal stroma which have been associated with the regulation of interfibrillar spacing and lamellar adhesion properties of corneal collagens (Maccarana *et al*., 2009)

Since these DE genes were not found to be significantly deregulated in MMC exposed cells, it can be hypothesised that these changes were induced specifically by Li.

4.3.4.4 DE Genes Identified in MMC Exposed HCECs but not in Li Exposed Cells

Downregulated

A significant number of downregulated DE genes specific to MMC treated HCECs were noted to have a role in metabolic pathways, namely, *GCNT4, ENO3, SLC27A5, MAN1C1, B3GALT2, CYP3A5, B3GALNT1, NDUFC2, CYP24A1, ACOT2, NADK2* and *ASNS*, amongst others.

A significant number of downregulated transcriptional regulator genes, such as *TP53, AGO4, RAD1, TSC1, CCNK, BID, RMI2, MRE11* and *COX14*, resulted to be enriched during ORA analysis. Other downregulated groups of genes included those involved in the Hippo signalling pathway (*ITGB2, CRB2, FZD7, WNT2B, BMPR1B, PPP2R2C*) and those involved in proton transmembrane transporter activity (*SLC15A2, SLC9A2, SLC25A27, ATP6V1C2, SLC9A5, SURF1*). *ITGB2*, a gene that encodes an integrin beta chain which combines with multiple different alpha chains to form different integrin heterodimers, was also noted to be downregulated in MMC treated HCECs.

Upregulated

The most predominant groups of upregulated genes identified by ORA analysis included cell surface receptor signalling pathways involved in cell-cell signalling (*SOX7, CTNNBIP1, RBMS3, CHRNA5, NKD1, DLG4, EDA, FUZ, TLE6, LRRK2, GRIN2A, ROR2, APCDD1*) and genes involved in the negative regulation of Wnt signalling pathway (*CTNNBIP1, RBMS3, NKD1, FUZ, TLE6, GLI1, DKK2, ROR2, APCDD1, SOX7, EDA, LRRK2, DKK2*). KEGG and Reactome pathway analysis also identified groups of upregulated genes involved in the TNF signalling pathway (*JAG1, TNF, MMP9, CCL2*), cAMP signalling pathway (*CALML6, CALML4, PPP1R1B, GLI1, GRIN2A, CALML5, PPP1R1B, CYP4V2, TAS2R14, TAS2R20*), as well as the Hippo signalling pathway (TEAD4, WTIP).

These DE genes that were not found to be deregulated in Li exposed cells are probably due to the effect of MMC on the HCECs.

4.3.4.5 Unexpected Effect of Li and MMC on *TGFBI* in Human Corneal Epithelial cells

The obvious question that emerges from the data obtained in this study is: What are the mechanisms that led to the indisputable significant upregulation of *TGFBI* in all the Li and MMC treatment groups when previous studies focusing on the effect of these compounds on TGFBIp in fibroblasts were documented to have caused its downregulation? The first postulation is that there might be an alternative pathway being involved in corneal epithelial cells exposed to Li and MMC that is not majorly involved in corneal fibroblasts. Seeing that TGFBIp is mainly produced by epithelial cells and not by fibroblasts in the cornea, when testing compounds with the goal of studying their effect on the modulation of *TGFBI* expression, logically, corneal epithelial cells and not fibroblasts should be targeted. This approach should

theoretically produce more accurate and more comparable results to what would happen in vivo. As mentioned above, it has been proven that different cell lines employ different pathways to modulate the expression of TGF-β target genes. This mechanism of pathway activation has been also tested specifically on corneal epithelial cells and fibroblasts and it was concluded that not all TGF-β-target proteins are induced through the SMAD pathway, and that corneal epithelial cells and fibroblasts use alternative pathways for TGF-β-target protein expression. They also stated that when there is a continued expression of the TGF-β-target protein, some of the SMAD proteins were found to be downregulated implying that there is an internal check mechanism in place (Guo, Hutcheon, Tran, *et al*., 2017).

Another possible theory is that in this study, upregulation of *TGFBI* expression could be due to the difference in the duration of exposure of HCECs to Li and MMC when compared to the previous studies carried out on fibroblasts. In previous studies, the effect of fibroblast exposure to these compounds was studied for only 12 hours (Choi, Sims, Murphy, *et al*., 2012) while in this study HCECs were exposed for 72hours. This timepoint was chosen since during cell culture it was noted that 72hours was a timeframe long enough for cells to divide 1–2 times and to observe drug effects without cells reaching confluence (Berrouet, Dorilas, Rejniak, *et al*., 2020). Besides, when a treatment is given, the long term effect of a drug on the cells has to be taken into account, considering the fact that, since CDs are due to a genetic mutation, this would not be a one-time treatment.

From the above quoted literature, one can see that the stimulation of a given pathway does not always have a predefined outcome. Elucidating the mechanisms accountable for this observation has proved to be far from simple due to different

pathways sharing components and an increasing number of interactions being identified between components of different pathways. Thus, the resulting differential expression of genes identified at a given point in time has to be looked at in the context of the complex spacial and temporal mechanisms of these pathways, their crosstalk and interacting feedback loops.

4.3.5 Socioeconomic Burden and Quality of Life:

Blurred or cloudy vision is the most common symptom of *TGFBI* CDs, in addition to watery or dry eyes, light sensitivity, glare, pain, and irritation (Klintworth, 2009). Corneal deposits in GCD1 individuals usually appear within the first decade of life, with early symptoms including glare and photophobia and followed by a decrease in visual acuity as opacity progresses with age (Cho, Mok, Na, *et al*., 2012). GCD2 symptoms also appear within the first decade with painful sensation accompanying corneal erosions (Akhtar, Meek, Ridgway, *et al*., 1999), while here, visual acuity usually remains good until later in life. On the other hand, LCD1 symptoms usually develop towards the end of the first decade, while some symptoms appear during middle age (Aldave, Kawasaki, Khemichian, *et al*., 2011).

Being one of the most prominent senses humans possess, vision is crucial for the well being and all daily aspects of life. As explained above, in *TGFBI* CDs, visual impairment occurs early on in life, leaving a severe impact on the quality of life of these individuals and their families. Visual disability impairs social interactions, independence, life satisfaction and working abilities. This inevitably leads to socio-economic burdens that are ever increasing with an aging population. In fact, in a study carried out by Singh *et al*. in the Medicare population in the United States, the economic burden relating to FECD alone has increased by twenty percent over a period of six years (Singh, Parmar, Kahale, *et al*., 2023).

These facts encourage the development and research of new ophthalmic interventions, in the pursuit of alleviating the financial and social negative impacts for people affected by these conditions all around the world.

4.3.6 Implementing these Research Findings into Practice

The results from the first phase of this study delineate the CD associated genetic variants present in a large cohort of the Maltese population. These genetic findings can be implemented in future studies by serving as a guide during the development process of potential personalized medicine approaches and targeted treatments based on individual genetic profiles.

The data showing that the Maltese cohort exhibited the least degree of CD-related genetic differentiation with the Hispanic cohorts leads to the rational suggestion that, in the absence of Maltese data, any medical research on future CD genetic targeted treatment carried out in these populations, can be used as a guide for the Maltese population.

Furthermore the data obtained contributes to new CD-related genetic knowledge about the Maltese population, bridging the gap and increasing the global genetic bank that scientists rely on for medical progress. Additionally, these results can be implemented into clinical practice and public health policies, since genomic analysis can be used to study traits or predict inherited disease risk and prognosis of individuals, as well as for targeted drug development.

The results obtained from the second phase of the study set the foundation towards understanding better the local Maltese GCD1 disease course and phenotype. The significant variants found to be present in the GCD1 Maltese individuals, especially those classified as ‘probably damaging’ by functional effect software tools, may theoretically alter the phenotype and severity of this condition in these individuals. Applying this research methodology to international GCD1 families in future research

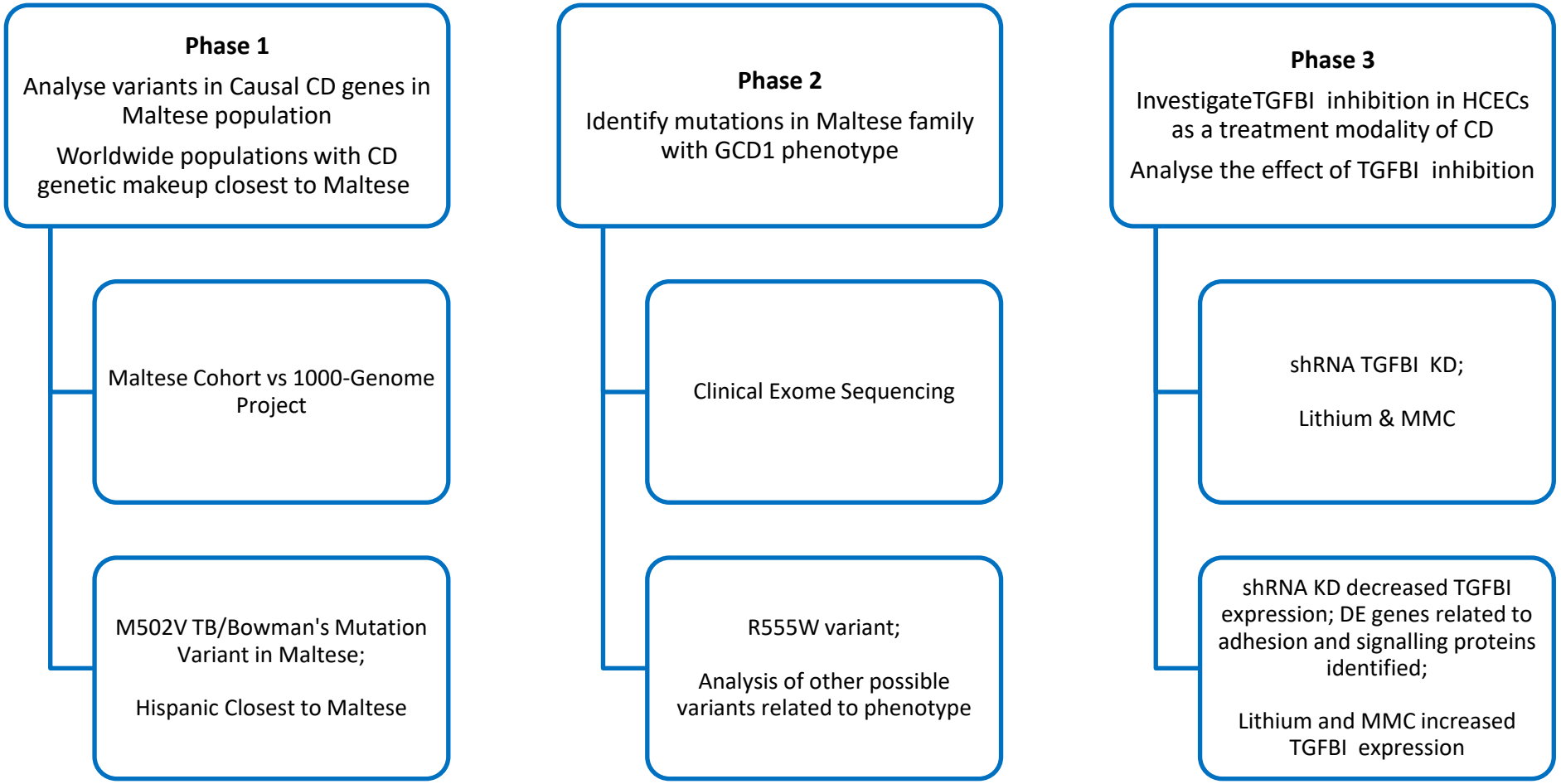
would determine definitively the role mutation variants, other than the *TGFBI* gene mutation, play in the phenotypic heterogeneity exhibited by these individuals, further substantiating these results. Moreover, this data opens the door to future personalised treatment strategies that might even target genes other than *TGFBI* ; genes that affect onset or severity, if *TGFBI* inhibition in vivo would prove to be too damaging.

Gene therapy is at present considered as the ideal and most promising method for the treatment of rare hereditary genetic diseases. This is especially true for CDs, where no effective treatments beyond corneal transplant are currently available. Despite appearing to be one of the most easily accessible organs in the body, delivery of medication to the cornea is met with a unique set of challenges. This is due to a combination of factors including physical barriers such as the tear film and the lipophilic nature of the epithelium as well as drug contact time and tear dilution. Implementing RNA interference (RNAi) in personalised treatment for genetic conditions by silencing the causal mutant gene or allele has shown great potential. Additionally it can be applied in the process of investigating gene function. Researchers have recently embarked on developing an effective, safe and non-invasive means of delivering RNAi to the cornea. These include hybrid silicon-lipid nanoparticles, modified cell-penetrating peptides and pH-sensitive vehicles amongst others (Schiroli, Gomara, Maurizi, *et al*., 2019) (Cao, Wang, Deng, *et al*., 2023) (Baran-Rachwalska, Torabi-Pour, Sutera, *et al*., 2020).

The successful KD of *TGFBI* in this study together with the analysis results of the subsequent RNA sequencing leads us to conclude that shRNA mediated gene KD in HCECs is effective and shows strong potential for the use as personalized treatment in

TGFBI CDs. The relatively small number of significant off-target effects, especially when compared to the number of differentially expressed genes documented in Li and MMC, further substantiates this proposal. These results pave the way towards the next step in research, in vivo studies.

4.3.7 SUMMARY FLOWCHART: Genetic Variants in Maltese Corneal Dystrophy Genes. Inhibition of TGFBI as a Treatment Modality.



4.4 Limitations

The results of the first phase of this project indicate that genetic ancestry plays a critical role in the distribution of allele frequencies in CD-related genes. Nevertheless, there are some limitations to the study. CDs are considered rare. Thus, the number of individuals exhibiting causal mutations was expected to be low. However, this was overcome by analysing a relatively large sample from the Maltese population so as to provide sufficient statistical validity. With regards to sample size of the 1000GP cohorts, even though the large majority of worldwide cohorts from the 1000GP had an adequate sample size, a small number (five) consisted of a sample size that was less than 96. The minimum sample size found to be required for the Maltese cohort to have a confidence level of 95%, when compared to the vast majority of the worldwide cohorts, was well below 410, which is the Maltese sample size utilized in this study. For the five cohorts from the 1000GP that had a sample size of less than 96, by performing power analysis (using the equation described in the methods section 3.1.1.2), it was determined that when compared to each other, the Maltese sample size of 410 will return a confidence level of 90%. Another limitation is that when data from individual cohorts is pooled into continental cohorts, the distinct variation between the individual ancestral cohorts is lost. Thus, in this study, the focus was on the individual ancestral cohorts when analysing the data. However, the results give a good indication regarding which ancestries have a genetic make-up that is closest to the Maltese population with regard to CD genes.

With regards to utilizing computational predictions in order to predict the pathogenicity of missense variants, some limitations are unfortunately inevitable. On analysing the performance of mutation pathogenicity prediction tools on missense

variants, researchers quoted Mutation Assessor to have the highest sensitivity, although SIFT and Polyphen-2 exhibited sensitivity of >0.90. The latter two were however found to have a lower specificity (Montenegro, Lerario, Nishi, *et al*., 2021). One of the main limitations of the applied computational tools in this study is that they are sequence-based. Thus, no structural information is taken into consideration. Furthermore, while these are particularly useful for loss-of function mutation interpretation, they provide less pronounced change when predicting gain-of-function mutations (Flanagan, Patch, & Ellard, 2010). Computational predictions can identify only a small percentage of pathogenic variants with high confidence. Thus, ideally, in order to determine definitively whether these missense mutations are actually pathogenic or not, functional assays would have to be performed. However, performing experimental characterization of all variants would be impractical considering cost, time and the large number of variants present. Therefore computational tools are indispensable in the prediction of variants on phenotype (Liu, Yeung, Chiu, *et al*., 2022).

The number of patients I came across that satisfied the inclusion criteria for recruitment for the second phase of the study were limited since CDs are a rare group of disorders. Additionally, since individuals were recruited from the cornea clinic at Mater Dei Hospital, asymptomatic GCD1 individuals and unaffected relatives were not encountered. Unaffected relatives would have made a robust control group in giving strength to the causal mutation variant that was identified in the GCD1 patients. However, the T*GFBI* R555W mutation identified in the individuals in the study has been confirmed to be causal for GCD1 in literature. In view of multiple literature reports of intra and inter-familial phenotypic variability, it is logical to question whether mutation

variants other than that in the *TGFBI* gene might be contributing to the phenotype GCD1 individuals. Thus, when analysing mutation variants other than the causal *TGFBI* gene that might have a role in the phenotype exhibited by this family of affected GCD1 individuals, filtering variants of the patients against unaffected relatives or any other group of individuals as controls could eliminate potentially relevant mutations. Even though the mutation variants that were listed as potential variants contributing to the phenotype of this family cannot be confirmed at this point, future studies combining genetic mutation testing results with longterm follow up to monitor progression and severity would help strengthen results.

With regards to the functional work, I personally performed the HCEC cell cultures, passaging and the exposure of the cultures to Li and MMC as well as the RNA extraction of the samples. During Clinical Exome Sequencing, I was assisted by Dr. Graziella Zahra and Mr. Mark Briffa who aided in operating the laboratory equipment and machines. Regrettably, since biohazard containment, viral labs and RNA sequencers are not currently available in Malta, KD of TGFBI via a shRNA lentiviral vector and RNA sequencing had to be carried out by LabOmics and Dante laboratories respectively.

A few limitations of shRNA technology must be mentioned. Unfortunately, the complete evasion of off-target effects caused by shRNA gene silencing is probably impossible; however, in this study these were minimized as much as possible. Off-target effects can arise due to nucleotide sequence homology with non-target transcripts, anomalous processing of endogenous miRNAs and disruption in cell homeostasis owing to the large presence of exogenous vectors and RNAs (Fellmann &

Lowe, 2014). Off target-effects due to sequence based homology were minimized by a) identifying shRNAs that are likely to target non-target transcripts by employing the BLAST function in NCBI and b) experimentally by using multiple shRNAs (Jackson & Linsley, 2010). Aberrant processing of endogenous miRNAs can also lead to off-target effects due to saturation of miRNA processing by Dicer complexes (Gu, Jin, Zhang, *et al*., 2012), experimental and endogenous RNAi competing for incorporation in RISC as well as unintended guide sequences being produced after processing of shRNA precursors (Kawahara, Zinshteyn, Sethupathy, *et al*., 2007).

In order to have a complete outlook on the effect Li and MMC have on HCECs, it would have been ideal if samples were taken, RNA extracted and sequencing performed at different timepoints. Although in this study we were more interested in the effect Li and MMC would have on HCECs in the long term, this would have given a more comprehensive outlook on the effect of these compounds on HCECs at various stages of exposure. Additionally, increasing the number of samples would also have increased the statistical power.

RNA-seq can be used in various analysis strategies that include transcript analysis of samples of organisms with known, well annotated genomes or even analysis of samples of organisms without known sequenced genomes. However even though it is very sensitive and can measure gene expression levels accurately, results can be affected by the challenging process of library preparation, sequencing depth, and bioinformatics analysis pipelines.

CHAPTER 5: CONCLUSION and FUTURE DIRECTIONS

In the first study of its kind, this epidemiological research focuses on Maltese genetics in the CD area of ophthalmology. One established mutation, which has been reported to cause a Bowman's layer CD/atypical Thiel-Behnke CD phenotype, was identified in this cohort, namely M502V, which is located in the *TGFBI* gene. Of the nine missense SNPs present in CD-related genes in the Maltese cohort, the *KRT3* NC_000012.11:g.53186088G>C mutation was predicted to be potentially deleterious by all four predictive algorithmic tools used in this study, namely PROVEAN, SIFT, MutationAssessor and Polyphen2. From the statistical analysis carried out, it can be deduced that the genetic composition of the Maltese cohort from the Malta biobank is most similar to that of the Puerto Rican, Mexican, and Colombian cohorts, when considering CD-related genes. Genetic studies such as the present shed light on the responsible mutations, novel or known, responsible for causing CDs in diverse populations, in this case the Maltese, bridging the gap and increasing the global genetic bank that scientists rely on for medical progress. Additionally, the data obtained can be employed in clinical practice and public health policies, since genomic analysis can be used to study traits or predict inherited disease risk and prognosis. Furthermore these results provide the opportunity of applying future CD related medical research carried out in these populations as a guide for the Maltese population, thus guiding clinicians towards developing the best management strategies to help improve the quality of life of individuals susceptible to these conditions.

To date, multiple GCD mutation variants with varying phenotypes have been detected in various ethnic population and family studies. Nevertheless, to date, the possible cause or mechanism that accounts for the phenotypic variability seen in GCD1 individuals carrying the R555W mutation has not yet been elucidated. This reinforces the fact that molecular analysis of cohorts of CD patients from different ethnic backgrounds is crucial in order to establish the actual mutation variants, their prevalence and the potential cause of the phenotypic-genotypic variability seen in these conditions (Paliwal, Sharma, Tandon, *et al*., 2010) (Long, Gu, Han, *et al*., 2011) (Cho, Mok, Na, *et al*., 2012) (Mazzotta, Traversi, Baiocchi, *et al*., 2015) (Zhao, Zhu, Shentu, *et al*., 2013) (Weiss, Moller, Lisch, *et al*., 2008). Thus, identifying mutation variants in key *TGFBI* associated genes in the GCD1 Maltese individuals is the first stepping stone towards understanding further the phenotype exhibited locally. Additionally, this data can guide future medical research focussing on personalized targeted treatment strategies for Maltese individuals with GCD1, with the aim of maximizing effectiveness while at the same time minimizing potential side-effects. Moreover, these results introduce the possibility of future personalised treatment strategies targeting genes which affect onset or severity of this condition, if *TGFBI* inhibition in vivo would prove to have undesirable side-effects.

In view of the unmet clinical needs and treatment options available for the management of *TGFBI* CDs, the third and last phase of this project revolved around exploring *TGFBI* inhibition in HCECs as a possible treatment modality. A scoping review was performed in which several compounds that decrease corneal TGFBIp levels were identified and the various cascades that modulate the production, secretion and degradation of TGFBIp were explored. However, the large majority of these studies

were performed on corneal stromal fibroblasts, while, it is known that TGFBIp is produced constitutively by corneal epithelial cells and that they are the primary source of production of mutant TGFBIp in persons with a *TGFBI* CD. Interestingly it has been documented that there are differences in pathway activation between corneal epithelial cells and fibroblasts and in fact, it was proven that not all TGF-β-target proteins are induced through the SMAD pathway, and that corneal epithelial cells and fibroblasts use alternative pathways for TGF-β-target protein expression (Guo, Hutcheon, Tran, *et al*., 2017). For *these* reasons, in this project, HCECs were used and not fibroblasts in order to mimic more closely what happens in vivo and produce the most biologically equivalent results. The hypothesis-driven method of using targeted repurposing of drugs led to testing the effect Li and MMC exposure would have on HCECs by subsequently analysing RNAseq results. In parallel to this, seeing that gene therapy is at present considered as the ideal method for the treatment of rare hereditary genetic diseases, the promising therapeutic approach of RNAi was also explored. *TGFBI* KD in HCECs was successfully carried out using shRNA lentiviral vectors and the DE genes identified from the RNAseq results were then analysed. These results revealed that the inhibition of *TGFBI* by shRNA caused significantly less deregulation of genes than that seen in the drug exposed group. ORA performed on DE genes from RNAseq results of *TGFBI* KD HCECs identified enriched associations with the JAK-STAT cascade, adhesion molecules, extracellular matrix structural constituents, RNA transport and Metabolism, FoxO signalling pathway and PI3K-Akt signalling pathway.

The transcriptome analysis of *TGFBI* KD HCECs performed in this research study sheds new light on the key genes that are associated with TGFBIp and provides us with a skeleton of TGFBIp-associated pathways. Besides, even though a number of

expression changes were identified at the molecular level, the actual side effects these changes might have at a functional level in the cornea might be minimal. Thus this leads us to conclude that shRNA mediated gene KD in HCECs is effective and shows strong potential for subsequent future in vivo studies, with the aim of finally implementing this method as a novel treatment in *TGFBI* CDs. This study provides insight into the function of TGFBIp in HCECs, helping us understand better its role in corneal epithelial cell homeostasis, while, at the same time providing a new perspective on future clinical treatment of *TGFBI* CDs.

REFERENCES

Abbruzzese, C., Khun, U., Molina, F., Rama, P., & De Luca, M. *(*2004). Novel mutations in the CHST6 gene causing macular corneal dystrophy. *Clin Genet, 65*, 120-125.

Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nature, 7*(4), 248-249.

Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V., Gerasimova, A., Bork, P., et al. (2010). *PolyPhen2*. Retrieved October 2020, from PolyPhen2: http://genetics.bwh.harvard.edu/pph2

Afshari, N. A., Bouchard, C. S., Colby, K. A., de Freitas, D., Rootman, D. S., Tu, E. Y., et al. (2014). Afshari NA, Bouchard CS, Colby KA, de Freitas D, Rootman DS,. In *Basic and Clinical Science Course, External Disease and Cornea: Corneal Dystrophies and ectasias.* (2014-2015 ed., pp. 253-287). San Francisco: American Academy of Ophthalmology.

Agarwal, S., Apple, D., Agarwal, A., Buratto, L., Alio, J., Pandey, S., et al. (2002). *Textbook of Ophthalmology.* New Delhi: Jaypee Brothers Medical Publishers Ltd.

Ahmed, S., Bradshaw, A., Gera, S., Dewan, M., & Xu, R. (2017). The TGF-β/Smad4 Signaling Pathway in Pancreatic Carcinogenesis and Its Clinical Significance. *6*(1), 5.

Akama, T., Nishida, K., Nakayama, J., Watanabe, H., Ozaki, K., Nakamura, T., et al. (2000). Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene. *Nat Genet, 26*, 237-241.

Akhtar, S., Meek, K., Ridgway, A., Bonshek, R., & Bron, A. (1999). Deposits and Proteoglycan Changes in Primary and Recurrent Granular Dystrophy of the Cornea. *Clinical Sciences, 117*(3), 310-321.

Akiya, S., & Brown, S. (1970). Granular dystrophy of the cornea. Characteristic electron microscopic lesion. *Arch Ophthalmol , 84*, 179-192.

Akiya, S., Takahashi, H., Nakano, N., Hirose, N., & Tokuda, Y. (1999). Granular-lattice (Avellino) corneal dystrophy. *Ophthalmologica, 213*, 58-62.

Aldave, A. J. (2011). The genetics of the corneal dystrophies. *Dev. Ophthalmol., 48*, 51-66.

Aldave, A. J., Kawasaki, S., Khemichian, A. J., Kinoshita, S., Murthy, S. J., Weiss, J. S., et al. (2011). Corneal Dystrophies. In F. Bandello, W. Lisch, & B. Seitz (Eds.), *Developments in Ophthalmology* (p. Vol. 48).

Aldave, A., Gutmark, J., Yellore, V., Affeldt, J., Meallet, M., Udar, N., et al. (2004). Lattice corneal dystrophy associated with the ala546asp and pro551gln missense changes in the TGFBI gene. *Am. J. Ophthal., 138*, 772-781.

Aldave, A., Sonmez, B., Forstot, S., Rayner, S., Yellore, V., & Glasgow, B. (2007). A clinical and histopathologic examination of accelerated TGFBIp deposition after LASIK in combined granular-lattice corneal dystrophy. *Am J Ophthalmol, 143*(3), 416–419.

Aliancy, J., Stamer, W., & Wirostko, B. (2017). A Review of Nitric Oxide for the Treatment of Glaucomatous Disease. *Ophthalmology and Therapy, 6*(2), 221-232.

Allen, E., Courtney, D., Atkinson, S., Moore, J., Mairs, L., Poulsen, E., et al. (2016). Keratin 12 missense mutation induces the unfolded protein response and apoptosis in Meesmann epithelial corneal dystrophy. *Human Molecular Genetics, 25*(6), 1176-1191.

American Academy of Ophthalmology. (2010). *External Disease and Cornea, Basic and Clinical Science Course.* American Academy of Ophthalmology.

Andersen, R., Karring, H., Moller-Pedersen, T., Valnickova, Z., Thogersen, I., Hedegaard, C., et al. (2004). Purification and Structural Characterisation of Transforming Growth Factor Beta Induced Protein (TGFBIp) from Porcine and Human Corneas. *Biochemistry, 43*(51), 16374-16384.

Anderson, J., & Hope, T. (2005). Intracellular trafficking of retroviral vectors: obstacles and advances. *Gene Therapy, 12*, 1667-1678.

Andreasen, M., Nielsen, S., Runager, K., Christiansen, G., Nielsen, N., & Enghild, J. (2012). Polymorphic fibrillation of the destabilized 4th fasciclin-1 domain mutant A546T of the transformig growth factor beta induced protein (TGFBIp) occurs through multiple pathways with different oligomeric intermediates. *J Biol Chem, 287*(41), 34730-34742.

Andrés-Guerrero, V., Alarma-Estrany, P., Molina-Martínez, I., Peral, A., Herrero-Vanrell, R., & Pintor, J. (2009). Ophthalmic formulations of the intraocular hypotensive melatonin agent 5-MCA-NAT. *Experimental Eye Research, 88*(3), 504-511.

Arici, C., Sevki Arslan, O., & Dikkaya, F. (2014). Corneal endothelial cell density and morphologyin healthy Turkish eyes. *Journal of Ophthalmology, 2014*.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice, 8*(1), 19-32.

Aronow, B., Lund, S., Brown, T., Harmony, J., & Witte, D. (1993). Apolipoprotein J expression at fluid-tissue interface: Potential role in barrier cytoprotection. *Proc Natl Acad Sci USA, 90*, 725-729.

Arpitha, P., Prajna, N. V., Srinivasan, M., & Muthukkaruppan, V. (2008). A subset of human limbal epithelial cells with greater nucleus-to-cytoplasm ratio expressing high levels of p63 possesses slow-cycling property. *Cornea*(27), 1164-1170.

Arranz-Marquez, E., Katsanos, A., Kozobolis, V., Konstas, A., & Teus, M. (2019). A critical overview of the biological effects of mitomycin C application on the cornea following refractive surgery. *Advances in Therapy, 36*(4), 786-797.

Assouline, M., Chew, S., Thompson, H., & Beuerman, R. (1992). Effect of growth factors on collagen lattice contraction by human keratocytes. *Invest Ophthalmol Vis Sci, 33*, 1742-1755.

Bae, J., Lee, S., Kim, J., Choi, J., Park, R., Park, J., et al. (2002). Beta ig–h3 supports keratinocyte adhesion, migration, and proliferation through alpha3beta1 integrin. *Bioachem Biophys Res Commun, 294*(5), 940-948.

Baleriola-Lucas, C., Fukuda, M., Willcox, M., Sweeney, D., & Holden, B. (1997). Fibronectin concentration in tears of contact lens wearers. *Exp Eye Research, 64*, 37-43.

Ballard, C., Aarsland, D., Cummings, J., O'Brien, J., Mills, R., Molinuevo, J., et al. (2020). Drug repositioning and repurposing for Alzheimer disease. *Nat Rev Neurol, 16*(12), 661-673.

Baran-Rachwalska, P., Torabi-Pour, N., Sutera, F., Ahmed, M., Thomas, K., Nesbit, M., et al. (2020). Topical siRNA delivery to the cornea and anterior eye by hybrid silicon-lipid nanoparticles. *Journal of Controlled Release, 326*, 192-202.

Baratz, K., Tosakulwong, N., Ryu, E., Brown, W., Branham, K., Chen, W., et al. (2010). E2-2 protein and Fuchs's corneal dystrophy. *New Eng. J. Med., 363*, 1016-1024.

Barnett, J., & Brundage, K. (2010). Immunotoxicology of Pesticides and Chemotherapies. In *Comprehensive Toxicology.* Science Direct.

Basaiawmoit, R., Oliveira, C., Runager, K., Sørensen, C., Behrens, M., Jonsson, B., et al. (2011). SAXS Models of TGFBIp Reveal a Trimeric Structure and Show that the

Overall Shape is not Affected by the Arg124His Mutation. *Journal of Molecular Biology*(408), 503-513.

Bazan, H., Allan, G., & Bazan, N. (1992). Enhanced expression of the growth-regulated calcyclin gene during corneal wound healing. *Exp Eye Res, 55*(1), 173-177.

Belmonte, C., & Gallar, J. (1996). *Neurobiology of Nociceptors. Chapter6: Corneal Nociceptors.* Oxford University Press.

Belsky, D., Moffitt, T., & Caspi, A. (2013). Genetics in Population Health Science: Strategies and Opportunities. *Am J Public health, 103*(1), S73-S83.

Berrouet, C., Dorilas, N., Rejniak, K., & Tuncer, N. (2020). Comparison of Drug Inhibitory Effects (IC50) in Monolayer and Spheroid Cultures. *Bull Math Biol, 82*(6), 68.

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Res, 23*(9), 1514-1521.

Bian, F., Pelegrino, F., Henriksson, J., Pflugfelder, S., Volpe, E., Li, D., et al. (2016). Differential Effects of Dexamethasone and Doxycycline on Inflammation and MMP Production in Murine Alkali-Burned Corneas Associated with Dry Eye. *Ocul Surf, 14*(2), 242-254.

Billings, P., Whitbeck, J., Adams, C., Abrams, W., Cohen, A., Engelsberg, B., et al. (2002). The transforming growth factor-beta-inducible matrix protein (beta)ig-h3 interacts with fibronectin. *J. Biol. Chem., 277*, 28003–28009.

Biswas, S., Munier, F., Yardley, J., Hart-Holden, N., Perveen, R., Cousin, P., et al. (2001). Missense mutations in COL8A2, the gene encoding the alpha-2 chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Hum. Molec. Genet., 10*, 2415-2423.

Bonanno, J. (2012). Molecular mechanisms underlying the corneal endothelial pump. *Experimental Eye Research, 95*(1), 2-7.

Boutboul, S., Black, G., Moore, J., Sinton, J., Menasche, M., Munier, F., et al. (Eds.). (2006). A subset of patients with epithelial basement membrane corneal dystrophy have mutations in TGFBI/BIGH3. *Hum Mutat, 27*, 553-557.

Boyden, L., Kam, C., Hernandez-Martin, A., Zhou, J., Craiglow, B., Sidbury, R., et al. (2016). Dominant de novo DSP mutations cause erythrokeratodermia-cardiomyopathy syndrome. *Human Molecular Genetics, 25*(2), 348-357.

Brodin, G., Ahgren, A., Dijke, P., Heldin, C., & Heuchel, R. (2000). Efficient TGF-beta induction of the Smad7 gene requires cooperation between AP-1, Sp1, and Smad proteins on the mouse Smad7 promoter. *J Biol Chem, 275*, 29023-29030.

Bron, A. J. (1990). The corneal dystrophies. *Current Opinion in Ophthalmology*, 333-346.

Bron, A. J., Tiffany, J. M., Gouveia, S. M., Yokoi, N., & Voon, L. W. (2004). Functional aspects of the tear film lipid layer. *Experimental Eye Research*(78), 347-360.

Brouwer, R., Van den Hout, M., Grosveld, F., & Van IJcken, W. (2012). NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics, 28*(2), 284-285.

Brownstein, S., Fine, B., Sherman, M., & Zimmerman, L. (1974). Granular dystrophy of the cornea. Light and electron microscopic confirmation of recurrence in a graft. *Am J Ophthalmol, 77*, 701-710.

Brummelkamp, T., Bernards , R., & Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science, 296*, 550-553.

Bücklers, M. (1949). Über eine weitere familiare Hornhautdystrophie. *Klin Monatsbl Augenheilkd, 114*, 386.

Burns, R. (1968). Meesman's (sic) corneal dystrophy. *Trans. Am. Ophthal. Soc, 66*, 530-635.

Buschke, W., Friedenwald, J. S., & Fleischmann, W. (1943). Studies on the mitotic activity of the corneal epithelium; methods; the effect of colchicine, ether, cocaine and ephedrin. *Bull Johns Hopkins Hospital*(73), 143-167.

Byrska-Bishop, M., Evani, U., Zhao, X., Basile, A., Abel, H., Regier, A., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *CellPress, 185*, 3426–3440.

Bystrom, B., Carracedo, S., Behndig, A., Gullberg, D., & Pedrosa-Domellof, F. (2009). α11 Integrin in the Human Cornea: Importance in Development and Disease. *IOVS, 50*, 5044-5053.

Cai, X., Liu, C., Zhang, T., Zhu, Y., Dong, X., & Xue, P. (2018). Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *J Cell Biochem, 119*(6), 4717-4728.

Callebaut, I., Mignotte, V., Souchet, M., & Mornon, J. (2003). EMI domains are widespread and reveal the probable orthologs of the Caenorhabditis elegans CED-1 protein. *Biochem Biophys Res Commun*(300), 619-23.

Cao, W., Ge, H., Cui, X., Zhang, L., Bai, J., Fu, S., et al. (2009). Reduced penetrance in familial Avellino corneal dystrophy associated with TGFBI mutations. *Mol Vis, 15*, 70-75.

Cao, X., Wang, C., Deng, Z., Zhong, Y., & Chen, H. (2023). Efficient ocular delivery of siRNA via pH-sensitive vehicles for corneal neovascularization inhibition. *Int J Pharm X, 5*, 100183.

Capelli, C., Redhead, N., Romano, V., Cali, F., Lefranc, G., Delague, V., et al. (2005). Population structure in the Mediterranean basin: A Y chromosome perspective. *Ann Hum Genet, 69*, 1-20.

Cavalli-Sforza, L. (2007). Human evolution and its relevance for genetic epidemiology. *Annal Review of Genomics and Human Genetics, 8*, 1-15.

Cescon, M., Gattazzo, F., Chen, P., & Bonaldo, P. (2015). Collagen VI at a glance. *J Cell Sci, 128*, 3525-3531.

Chae, H., Kim, M., Kim, Y., Kim, J., Kwan, A., Choi, H., et al. (2016). Mutational spectrum of Korean patients with corneal dystrophy. *Clinical Genetics, 89*, 678-689.

Chakravarthi, K., Kannabiran, C., Sridhar, M. S., & Vemuganti, G. K. (2002). TGFBI Gene Mutations Causing Lattice and Granular Corneal Dystrophies in Indian Patients. *IOVS, 43*(4), 949-954.

Champely, S. (2020). *pwr: Basic Functions for Power Analysis.* Retrieved from https://CRAN.R-project.org/package=pwr

Chao-Shern, C., DeDionisio, L., Jang, J.-H., Chan, C., Thompson, V., Christie, K., et al. (2019). Evaluation of TGFBI corneal dystrophy and molecular diagnostic testing. *Eye, 33*, 874-881.

Chao-Shern, C., Me, R., DeDionisio, L., Ke, B., Nesbit, M., Marshall, J., et al. (2017). Post-LASIK exacerbation of granular corneal dystrophy type 2 in members of a chinese family. *Eye, 32*, 39-43.

Chen, G., Yuan, A., Shriner, D., Tekola-Ayele, F., Zhou, J., Bentley, A., et al. (2015). An improved Fst estimator. *PLoS ONE, 10*(8), e0135368.

Chen, J., Lin, B., Gee, K., Gee, J., Chung, D., Frausto, R., et al. (2015). Identification of presumed pathogenic KRT3 and KRT12 gene mutations associated with Meesmann corneal dystrophy. *Molecular Vision, 21*, 1378-1386.

Chen, K., Wu, Y., Zhu, M., Deng, Q., Nie, X., Li, M., et al. (2013). Lithium chloride promotes host resistance against Pseudomonas aeruginosa keratitis. *Mol Vis, 19*(19), 1502-1514.

Chen, X., & Sullivan, P. (2003). Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *The Pharmacogenomics journal, 3*, 77-96.

Chen, Y., Huang, S., Tai, M., Chen, J., & Liang, C. (2019). Glucosamine impedes transforming growth factor β1-mediated corneal fibroblast differentiation by targeting Krüppel-like factor 4. *Journal of Biomedical Science, 26*, 72.

Chen, Y., Huang, S., Tai, M., Chen, J., Lee, A., Huang, R., et al. (2020). The anti-fibrotic and anti-inflammatory effects of 2,4-diamino-5-(1-hydroxynaphthalen-2-yl)-5H-chromeno[2,3-b] pyriine-3-carbonitrile in corneal fibroblasts. *Pharmacol Rep, 72*(1), 115-125.

Chen, Z., Wu, Q., Yan, C., & Du, J. (n.d.). COL6A1 knockdown suppresses cell proliferation and migration in human aortic vascular smooth muscle cells. *Experimental and Therapeutic Medicine, 18*(3), 1977-1984.

Cheng, W., Kao, S., Chen, C., Yuliani, F., Lin, L., Lin, C., et al. (2022). Amphiregulin induces CCN2 and fibronectin expression by TGF-β through EGFR-dependent pathway in lung epithelial cells. *Respiratory Research, 23*, 381.

Cho, K., Mok, J. W., Na, K. S., Rho, C. R., Byun, Y., Hwang, H., et al. (2012). TGFBI gene mutations in a Korean population with corneal dystrophy. *Molecular vision, 18*, 2012-2021.

Choi, S., Kim, B., Dadakhujaev, S., Jester, J., Ryu, H., & Kim, T. (2011). Inhibition of TGFBIp Expression by Lithium: Implications for TGFBI-linked Corneal Dystrophy Therapy. *Invest Ophthalmol Vis Sci, 52*, 3293-3300.

Choi, S., Kim, K., Oh, J., Jin, J., Lee, G., & Kim, E. (2013). Melatonin induces autophagy via an mTOR-dependent pathway and enhances clearance of mutant-TGFBIp. *Journal of Pineal Research, 54*(4), 361-372.

Choi, S., Kim, T., Kim, K., Kim, B., Ahn, S., Cho, H., et al. (2009). Decreased catalase expression and increased susceptibility to oxidative stress in primary cultured corneal fibroblasts from patients with granular corneal dystrophy type II. *Am J Pathol, 175*, 248-261.

Choi, S., Lee, E., Akuzum, B., Jeong, J., Maeng, Y., Kim, T., et al. (2017). Melatonin reduces endoplasmic reticulum stress and corneal dystrophy‐associated TGFBIp through activation of endoplasmic reticulum‐associated protein degradation. *Journal of Pineal Research, 63*(3), 12426.

Choi, S., Lee, E., Jeong, J., Akuzum, B., Meang, Y., Kim, T., et al. (2016). 4-Phenylbutyric acid reduces mutant-TGFBIp levels and ER stress through activation of ERAD

pathway in corneal fibroblasts of granular corneal dystrophy type 2. *Biochem Biophys Res Commum, 477*(4), 841-846.

Choi, S., Yoo, Y., Kim, B., Kim, T., Cho, H., Ahn, S., et al. (2010). Involvement of TGF-β Receptor– and Integrin-Mediated Signaling Pathways in the Pathogenesis of Granular Corneal Dystrophy II. *Investigative Ophthalmology and Visual Science, 51*, 1832-1847.

Choi, S.-i., Maeng, Y.-S., Kim, T.-i., Yangsin, L., Kim, Y.-S., & Kim, E. K. (2015). Lysosomal Trafficking of TGFBIp via Caveolae-Mediated Endocytosis. *PLoS One*(10).

Choi, S.-l., Kim, B.-Y., Dadakhujaev, S., Oh, J.-Y., Kim, T.-I., Kim, J. Y., et al. (2012). Impaired autophagy and delayed autophagic clearance of transforming growth factor beta-induced protein (TGFBI) in granular corneal dystrophy type 2. *Autophagy*(8), 1782-1797.

Choi, Y. (2012). A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 414-417.

Choi, Y., & Chan, A. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics, 31*(16), 2745-2747.

Choi, Y., Sims, G., Murphy, S., Miller, J., & Chan, A. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE, 7*(10), e46688.

Clout, N. J., & Hohenester, E. (2003). A model of FAS1 domain 4 of the corneal protein beta(ig)-h3 gives a clearer view on corneal dystrophies. *Molecular vision*(11), 440-448.

Clout, N. J., Tisi, D., & Hohenester, E. (2001). Novel fold revealed by the structure of a FAS1 domain pair from the insect cell adhesion molecule fasciclin I. *Structure*(11), 197-203.

Cogan, D., Donaldson, D., Kuwabara, T., & Marshall, D. (1964). Microcystic dystrophy of the corneal epithelium. *Trans Am Ophthal Soc, 62*, 213-225.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York.

Collin, J., Queen, R., Zerti, D., Bojic, S., Dorgau, B., Moyse, N., et al. (2021). A single cell atlas of human cornea that defines its development, limbal progenitor cells and their interactions with the immune cells. *The Ocular Surface, 21*, 279-298.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology, 17*(13), doi.org/10.1186/s13059-016-0881-8.

Consortium, T. 1. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74.

Cooper, D., Ball, E., Stenson, P., Phillips, A., Evans, K., Heywood, S., et al. (n.d.). *Human Gene Mutation Database*. Retrieved October 2020, from http://www.hgmd.cf.ac.uk/ac/index.php

Cooper, D., Schermer, A., & Sun, T. (1985). Classification of human epithelia and their neoplasms using monoclonal antibodies to keratins: strategies, applications, and limitations. *Lab Invest, 52*, 243-56.

Coulson-Thomas, V., Chang, S., Yeh, L., Coulson-Thomas, Y., Yamaguchi, Y., Esko, J., et al. (2015). Loss of Corneal Epithelial Heparan Sulfate Leads to Corneal Degeneration and Impaired Wound Healing. *IOVS, 56*, 3004-3014.

Courtney, D., Atkinson, S., Moore, J., Maurizi, E., Serafini, C., Pellegrini, G., et al. (2014). Development of Allele-Specific Gene-Silencing siRNAs for TGFBI Arg124Cys in Lattice Corneal Dystrophy Type I. *IOVS, 55*, 977-985.

Courtney, D., Poulsen, E., Kennedy, S., Moore, J., Atkinson, S., Maurizi, E., et al. (2015). Protein Composition of TGFBI-R124C- and TGFBI-R555W- Associated Aggregates Suggests Multiple Mechanisms Leading to Lattice and Granular Corneal Dystrophy. *Investigative Ophthalmology & Visual Science, 56*(8), 4653-4661.

Couto, D. L. (2008). Periostin, a member of a novel family of vitamin K-dependent proteins, is expressed by mesenchymal stromal cells. *J Biol Chem, 283*(26), 179911-18001.

Cox, J. L., Farrell, R. A., Hart, R. W., & Langham, M. E. (1970). The transparency of the mammalian cornea. *The Journal of Physiology, 3*(210), 601-616.

Crisafulli, S., Sultana, J., Ingrasciotta, Y., Addis, A., Cananzi, P., Cavagna, L., et al. (2019). Role of healthcare databases and registries for surveillance of orphan drugs in the real-world setting: the Italian case study. *Expert Opin Drug Saf, 18*(6), 497-509.

Cruzat, A., Qazi, Y., & Hamrah, P. (2017). In vivo confocal microscopy of corneal nerves in health and disease. *Ocul Surf, 15*(1), 15-47.

Dadlani, M. (2023). Illumina Sequencing: Process and Advantages Explained. *Sequencing*.

Dahlén, A., Dashi, G., Maslov, I., Attwood, M., Jonsson, J., Trukhan, V., et al. (2022). Trends in antidiabetic drug discovery: FDA approved drugs, new drugs in clinical trials and global sales. *Front Pharmacol, 12*, 807548.

Davanger, M., & Evensen, A. (1971). Role of the pericorneal papillary structure in renewal of corneal epithelium. *Nature*(229), 560-561.

Derynck, R., & Zhang, Y. (2003). Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature, 425*, 577-584.

Dhindsa, R., Wang, Q., Vitsios, D., Burren, O., Hu, F., DiCarlo, J., et al. (2022). A minimal role for synonymous variation in human disease. *AJHG, 109*(12), 2105-2109.

Diehn, J., Diehn, M., Marmor, M., & Brown, P. (2005). Differential gene expression in anatomical compartments of the human eye. *Genome Biology, 6*(9), R74.

Dighiero, P., Drunat, S., D'Hermies, F., Renard, G., Delpech, M., & Valleix, S. (2000). A novel variant of granularcorneal dystrophy caused by association of 2 mutations in the TGFBI gene R124L and deltaT125-deltaE126. *Archives of Ophthalmology, 118*, 814-818.

Dighiero, P., Niel, F., Ellies, P., D'Hermies, F., Savoldelli, M., Renard, G., et al. (2001). Histologic phenotype-genotype correlation of corneal dystrophies associated with eight distinct mutations in the TGFBI gene. *Ophthalmology, 108*, 818-823.

Disorders, N. O. (n.d.). *Rare Disease Database*. Retrieved November 2020, from https://rarediseases.org/rare-diseases/corneal-dystrophies/

Dobson, C. M. (2003). Protein folding and misfolding. *Nature, 426*(6968), 884–890.

Dokmanovic, M., Chang, B., Fang, J., & Roninson, I. (2002). Retinoid-induced growth arrest of breas tcarcinoma cells involves co-activation of multiple growth-inhibitory genes. *Cancer Biol Ther, 1*, 24-27.

Dou, F., Liu, Q., Lv, S., Xu, Q., Wang, X., Liu, S., et al. (2023). FN1 and TGFBI are key biomarkers of macrophage immune injury in diabetic kidney disease. *Medicine (Baltimore), 120*(45), e35794.

Doutch, J., Quantock, A., Smith, V., & Mee, K. (2008). Light Transmission in the Human Cornea as a Function of Position across the Ocular Surface: Theoretical and Experimental Aspects. *Biophysical Journal, 95*, 5092–5099.

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*(9), 1778-1785.

Duke-Elder, S., & Wybar, K. C. (1960). The anatomy of the visual system. In *Duke-Elders System of Ophthalmology* (pp. 92-131). London: Kimpton.

Dupps Jr., W., & Wilson, S. (2006). Biomechanics and Wound Healing in the Cornea. *Exp Eye Res., 83*(4), 709-720.

Dykxhoorn, D., & Lieberman, J. (2005). The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu Rev Med, 56*, 401-423.

Dyrlund, T., Toftgaard Poulsen, E., Scavenius, C., Lund Nikolajsen, C., Thogersen, I., Vorum, H., et al. (2012). Human Cornea Proteome: Identification and Quantitation of the Proteins of the Three Main Layers Including Epithelium, Stroma, and Endothelium. *Journal of Proteome Research, 11*(8), 4231-4239.

Elavazhagan, M., Lakshminarayanan, R., Zhou, L., Ting, L., Tong, L., & Beuerman, R. (2012). Expression, purification and characterisation of fourth FAS1 domain of TGFbetaIp-associated corneal dystrophic mutants. *Protein Expr Purif, 84*, 108-115.

EMA. (n.d.). *European Medicines Agency*. Retrieved 2022, from Science Medicines Health: https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu-3-10-756

EMC. (n.d.). *Electronic Medicines Compendium*. Retrieved 2022, from https://www.medicines.org.uk/emc/product/11327/smpc

Ensemble. (n.d.). Retrieved from http://www.ensembl.org/Homo_sapiens/Variation/Citations?db=core;r=2:208325106-208326106;v=rs2363468;vdb=variation;vf=183290727

Escribano, J., Hernando, N., Ghosh, S., Crabb, J., & Coca-Prados, M. (1994). cDNA from human ocular ciliary epithelium homologous to beta ig-h3 is preferentially expressed as an extracellular protein in the corneal epithelium. *J Cell Physiol, 160*(3), 511-521.

Esquenazi, S., Esquenazi, I., Grunstein, L., He, J., & Bazan, H. (2009). Immunohistological Evaluation of the Healing Response at the Flap Interface in Patients With LASIK Ectasia Requiring Penetrating Keratoplasty. *J Refract Surg, 25*(8), 739-7446.

Evans, C., Davidson, A., Carnt, N., Rojas Lopez, K., Veli, N., Thaung, C., et al. (2016). Genotype-Phenotype Correlation for TGFBI Corneal Dystrophies Identifies p.(G623D) as a Novel Cause of Epithelial Basement Membrane Dystrophy. *IOVS, 57*(13), 5407-5414.

FDA. (2016). FDA Proposed Rules. *Federal Register, 81*(242), 91071-91082.

Feldman, B., & Afshari, N. (2010). Corneal Dystrophies. In D. Dartt (Ed.), *Encyclopedia of the eye.* Science Direct.

Felice, C., Lewis, A., Armuzzi, A., Lindsay, J., & Silver, A. (2015). Review article: selective histone deacetylase isoforms as potential therapeutic targets in inflammatory bowel diseases. *Aliment Pharmacol, 41*, 26-38.

Fellmann, C., & Lowe, S. (2014). Stable RNA interference rules for silencing. *Nat Cell Biol, 16*(1), 10-18.

Fersht, A. (1999). *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding.* New York: Freeman.

Fine, B., Yanoff, M., Pitts, E., & Slaughter, F. (1977). Meesmann's epithelial dystrophy of the cornea. *Am. J. Ophthal., 83*, 633-642.

Fischbarg, J., Diecke, F. P., Iserovich, P., & Rubashkin, A. (2006). The Role of the Tight Junction in Paracellular Fluid Transport across Corneal Endothelium. Electro-osmosis is a Driving Force. *J Membr Biol, 210*(2), 117-130.

Flach, A. (2008). The importance of eyelid closure and nasolacrimal occlusion following the ocular instillation of topical glaucoma medications, and the need for the universal inclusion of one of these techniques in all patient treatments and clinical studies. *Trans Am Ophthalmol Soc, 106*, 138-145.

Flanagan, S., Patch, A., & Ellard, S. (2010). Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. *Genetic Testing and Molecular Biomarkers, 14*(4), 533-537.

Fleming, A. A. (1922). One remarkable bacteriolytic element found in tissues and secretions. *Proceedings of the Royal Society London (Biol)*(93).

Folberg, R., Alfonso, E., Croxatto, O., Driezen, N. G., Panjwani, N., Laibson, P., et al. (1988). Clinically Atypical Granular Corneal Dystrophy with Pathologic Features of Lattice-like Amyloid Deposits. *Ophthalmology, 95*(1), 46-51.

Forrester, J., Dick, A., McMenamin, P., Roberts, F., & Pearlman, E. (2016). *The Eye, Basic Sciences in Practice.* Elsevier.

Francois, J., & Neetens, A. (1857). Nouvelle dystrophie heredofamiliale du parenchyme corneen (heredo-dystrophie mouchetee). *Bull. Soc. Belge Ophtal., 114*, 641-646.

Frankham, R., Ballou, J., & Briscoe, D. (2022). *Introduction to conservation genetics.* Cambridge: Cambridge University Press.

Frising, M., Wildhardt, G., Frisch, L., & Pitz, S. (2006). Recurrent Granular Dystrophy of the Cornea: An Unusual Case. *Cornea, 25*(5), 614-617.

Fruman, D., Chiu, H., Hopkins, B., Bagrodia, S., Cantley, L., & Abraham, R. (2018). The PI3K pathway in human disease. *Cell, 170*(4), 605-635.

Fujiki, K., Hotta, Y., Nakayasu, K., & Kanai, A. (1998). Homozygotic patient with betaig-h3 gene mutation in granular dystrophy. *Cornea, 17*, 288-292.

Fujiki, K., Nakayasu, K., & Kanai, A. (2001). Corneal dystrophies in Japan. *J. Hum. Genet., 46*, 431-435.

Galiacy, S., Froment, C., Mouton-Barbosa, E., Erraud, A., Chaoui, K., Desjardins, L., et al. (2011). Deeper in the human cornea proteome using nanoLC-Orbitrap MS/MS: An improvement for future studies on cornea homeostasis and pathophysiology. *J. Proteomics, 74*(1), 81-92.

Garcia-Castellanos, R., Nielsen, N., Runager, K., Thogersen, I., Lukassen, M., Poulsen, E., et al. (2017). Structural and Functional Implications of Human Transforming Growth Factor β-Induced Protein, TGFBIp, in Corneal Dystrophies. *Structure, 25*(11), 1740-1750.

Garner, A. (1969). Histochemistry of corneal granular dystrophy. *Br J Ophthalmol, 53*, 799-807.

Gaudana, R., Jwala, J., Boddu, S., & Mitra, A. (2009). Recent Perspectives in Ocular Drug Delivery. *Pharm Res, 26*, 1197-1216.

Ge, Q., Ilves, H., Dallas, A., Kumar, P., Shorenstein, J., Kazakov, S., et al. (2010). Minimal-length short hairpin RNAs: The relationship of structure and RNAi activity. *RNA, 16*(1), 106-117.

Gee, J., Frausto, R., Chung, D., Tangmonkongvoragul, C., Le, D., Wang, C., et al. (2015). Identification of novel PIKFYVE gene mutations associated with Fleck corneal dystrophy. *Mol Vis, 21*, 1093-1100.

Giardine, B., Riemer, C., Hardison, R., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Research, 15*, 1451-1455.

Gipson, I. K., Spurr-Michaud, S. J., & Tisdale, A. S. (1988). Hemidesmosomes and anchoring fibril collagen appear synchronously during development and wound healing. *Dev Biol., 2*(126), 253-262.

Gipson, I., & Argueso, P. (2003). Role of mucins in the function of the corneal and conjunctival epithelia. *Int Rev Cytol, 231*, 1-49.

Glavini, A. (2013). *Role of keratoepithelin in inherited corneal dystrophies.* Ann Arbor, US: ProQuest.

Goldman, J., & Benedek, G. (1967). The relatioinship between morphology and transparency in the nonswelling corneal stroma of the shark. *Invest Ophthalmol, 6*(6), 574-600.

Gottsch, J., Sundin, O., Liu, S., Jun, A., Broman, K., Stark, W., et al. (2005). Inheritance of a novel COL8A2 mutation defines a distinct early-onset subtype of Fuchs corneal dystrophy. *Invest. Ophthal. Vis. Sci., 46*, 1934-1939.

Grella, A., Kole, D., Holmes, W., & Dominko, T. (2016). FGF2 Overrides TGFβ1-Driven Integrin ITGA11 Expression in Human Dermal Fibroblasts. *J Cell Biochem, 117*(4), 1000-1008.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P., et al. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of Dicer processing in vivo. *Cell, 151*(4), 900-911.

Guo, G., Fu, Y., Lee, H., Cai, T., & Harris, K. (2014). Recognizing a Small Amount of Superficial Genetic Differences Across African, European and Asian Americans Helps Understand Social Construction of Race. *Demography, 51*, 2337-2342.

Guo, X., Hutcheon, A., Tran, J., & Zieske, J. (2017). TGF-β-target genes are differentially regulated in corneal epithelial cells and fibroblasts. *New Front Ophthalmol, 3*, DOI: 10.15761.

Ha, N., Cung, l. X., Chau, H., Thanh, T., Fujiki, K., Murakami, A., et al. (2003). A novel mutation of the TGFBI gene found in a Vietnamese family with atypical granular corneal dystrophy. *Jpn J Ophthalmol, 47*(3), 246-248.

Ha, S.-W., Bae, J.-S., Yeo, H.-J., Lee, S.-H., Choi, J.-Y., Sohn, Y.-K., et al. (2003). TGF-β-induced protein βig-h3 is upregulated by high glucose in vascular smooth muscle cells. *Journal of Cellular Biochemistry, 88*(4), 774-784.

Haddad, R., Font, R., & Fine, B. (1977). Unusual superficial variant of granular dystrophy of the cornea. *American Journal of Ophthalmology, 83*, 213-218.

Hall, P. (1974). Reis-Bucklers dystrophy. *Arch. Ophthal., 91*, 170-173.

Hamill, C., Schmedt, T., & Jurkunas, U. (2013). Fuchs endothelial cornea dystrophy: a review of the genetics behind disease development. *Semin Ophthalmol., 28*, 281–286.

Han, K., Choi, S.-i., Kim, T.-i., Maeng, Y.-S., Doyle Stulting, R., Ji, Y., et al. (2016). Pathogenesis and treatments of TGFBI corneal dystrophies. *Progress in Retinal and Eye Research*(50), 67-88.

Hanna, C., & O'Brien, J. E. (1960). Cell production and migration in the epithelial layer of the cornea. *Arch Ophthalmol.*(64), 536-539.

Hanneke, R., Asada, Y., Lieberman, L., Neubauer, L., & Fagan, M. (2017). *The Scoping Review Method: Mapping the Literature in "Structural Change" Public Health Interventions.* Department of Public Health Scholarship and Creative Works.

Hanssen, E., Reinboth, B., & Gibson. (2003). Covalent and non-covalent interactions of betaig-h3 with collagen VI. Beta ig-h3 is covalently attached to the amino-terminal region of collagen VI in tissue microfibrils. *J. Biol. Chem., 278*, 24334–24341.

Hartz, P. A., & McKusick, V. A. (2015). *OMIM*. Retrieved 2016, from http://www.omim.org/entry/601692

Hassan, M. S., Shaalan, A. A., Dessouky, M., Abdelnaiem, A., & ElHefnawi, M. (2019). Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics, 111*(4), 869-882.

He, C., Li, A., Lai, Q., Ding, J., Yan, Q., Liu, S., et al. (2021). The DDX39B/FUT3/TGFβR-I axis promotes tumor metastasis and EMT in colorectal cancer. *Cell Death Dis, 12*(1), 74.

Hodges, R., & Dartt, D. (2013). Tear Film Mucins: Front Line Defenders of the Ocular Surface. *Exp Eye Res, 117*, 62-78.

Holland, E., Daya, S., Stone, E., Folberg, R., Dobler, A., Cameron, J., et al. (1992). Avellino corneal dystrophy: clinical manifestations and natural history. *Ophthalmology, 99*, 1564-1568.

HolmesJr, D., Savage, M., LaBlanche, J., Grip, L., Serruys, P., Fitzgerald, P., et al. (2002). Results of Prevention of REStenosis with Tranilast and its Outcomes (PRESTO) Trial. *Circulation, 106*, 1243-1250.

Holsinger, K., & Weir, B. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews Genetics, 10*(9), 639-650.

Hou, Y.-C., Wang, I.-J., Hsiao, C.-H., Chen, W.-L., & Hu, F.-R. (2012). Phenotype–genotype correlations in patients with TGFBI-linked corneal dystrophies in Taiwan. *Molecular Vision, 18*, 362-371.

Hu, C., Sun, J., Zhang, Y., Chen, J., Lei, Y., Sun, X., et al. (2018). Local Delivery and Sustained-Release of Nitric Oxide Donor Loaded in Mesoporous Silica Particles for Efficient Treatment of Primary Open-Angle Glaucoma. *Adv Healthc Mater, 7*(23), e1801047.

Huang, T., Shu, Y., & Cai, Y.-D. (2015). Genetic differences among ethnic groups. *BMC Genomics*, 1093.

Humphreys, D., Carver, J., Easterbrook-Smith, S., & Wilson, M. (1999). Clusterin has chaperone-like activity similar to that of small heat shock proteins. *The Journal of biological chemistry, 274*(11), 6875-6881.

Hung, M., Chen, I., You, L., Jablons, D., Li, Y., Mao, J., et al. (2015). Knockdown of Cul4A increases chemosensitivity to gemcitabine through upregulation of TGFBI in lung cancer cells. *Oncol Rep., 34*(6), 3187-3195.

Ichhpujani, P., Singla, E., Kalra, G., Bhartiya, S., & Kumar, S. (2022). Surgical trends in glaucoma management: The current Indian scenario. *Int Ophthalmol, 42*(6), 1661-1668.

Iliff, B., Riazuddin, S., & Gottsch, J. (2012). The genetics of Fuchs′ corneal dystrophy. *Expert Rev Ophthalmol., 7*(4), 363-375.

illumina. (2018). *MiSeq® Reporter*. Retrieved August 24, 2018, from https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/miseqreporter/miseq-reporter-software-guide-15042295-05.pdf

illumina. (n.d.). *TruSight One Sequencing Panel Series*. Retrieved August 21, 2018, from https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/trusight_one/trusight-one-sequencing-panel-reference-guide-15046431-03.pdf

Imanishi, J., Kamiyama, K., Iguchi, I., Kita, M., Sotozono, C., & Kinoshita, S. (2000). Growth factors: importance in wound healing and maintenance of transparency of the cornea. *Progress in Retinal and Eye Research, 19*(1), 113-129.

Inamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J., Kawabata, M., et al. (1997). Smad6 inhibits signalling by the TGF-beta superfamily. *Nature, 389*(6651), 622-626.

Iozzo, R. V. (2005). Basement membrane proteoglycans: from cellar to ceiling. *Nat Rev Mol Cell Biol*(6), 646-656.

Irvine, A., Corden, L., Swensson, O., Swensson, B., Moore, J., Frazer, D., et al. (1997). Mutations in cornea-specific keratin K3 or K12 genes cause Meesmann's corneal dystrophy. *Nat. Genet, 16*(2), 184–187.

Ivanov, S., Ivanova, A., Salnikow, K., Timofeeva, O., Subramaniam, M., & Lerman, M. (2008). Two novel VHL targets, TGFBI (BIGH3) and its transactivator KLF10, are up-regulated in renal clear cell carcinoma and other tumors. *Biochem Biophys Res Commun.*(370), 536-540.

Iwamoto, T., Stuart, J., Srinivasan, B., Mund, M., Farris, R., Donn, A., et al. (1975). Ultrastructural variation in granular dystrophy of the cornea. *Albrecht von Graefes Arch Klin Exp Ophthalmol, 194*, 1-9.

Jackson, A., & Linsley, P. (2004). Noise amidst the silence: off-target effects of siRNAs? *Trends Genet, 20*, 521-524.

Jackson, A., & Linsley, P. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature reviews Drug discovery, 9*(1), 57-67.

Jester, J. V., Huang, J., Barry-Lane, P. A., Kao, W. W., Petroll, W. M., & Cavanagh, H. D. (1999). Transforming growth factor(beta)-mediated corneal myofibroblast differentiation requires actin and fibronectin assembly. *Invest Ophthalmol Vis Sci.*(40), 1959–1967.

Jha, P., Bora, P., & Bora, N. (2007). The role of complement system in ocular diseases including uveitis and macular degeneration. *Mol. Immunol, 44*(16), 3901–3908.

Jia, F., Li, L., Fang, Y., Song, M., Man, J., Jin, Q., et al. (2020). Macromolecular Platform with Super-Cation Enhanced Trans-Cornea Infiltration for Noninvasive Nitric Oxide Delivery in Ocular Therapy. *ACS Nano*, 33289535.

Jin, J., Han, J., Ha, J., Baek, H., & Lim, D. (2021). Lobeglitazone, a peroxisome proliferator-activated receptor-gamma agonist, inhibits papillary thyroid cancer cell migration and invasion by suppressing p38 MAPK signalling pathway. *Endocrinol Metab (Seoul), 36*(5), 1095-1110.

Judisch, G., & Maumenee, I. (1978). Clinical differentiation of recessive congenital hereditary endothelial dystrophy and dominant hereditary endothelial dystrophy. *Am. J. Ophthal., 85*, 606-612.

Jumelle, C., Gholizadeh, S., Annabi, N., & Dana, R. (2020). Advances and limitations of drug delivery systems formulated as eye drops. *J Control Release, 321*, 1-22.

Kabza, M., Karolak, J., Rydzanicz, M., Szcześniak, M., Nowak-Malczewska, D., Ginter-Matuszewska, B., et al. (2017). Collagen synthesis disruption and

downregulation of core elements of TGF-β, Hippo, and Wnt pathways in keratoconus corneas. *Eur. J. Hum. Genet., 25*, 582-590.

Kalliolias, G., & Ivashkiv, L. (2010). Overview of the biology of type I interferons. *Arthritis Research Therapy, 12*(1), doi:10.1186/ar2881.

Kanai, A., Yamaguchi, T., & Nakajima, A. (1977). The histochemical and analytical electron microscopy studies of the corneal granular dystrophy. *Nippon Ganka Gakkai Zasshi, 81*, 145-154.

Kannabiran, C., & Klintworth, G. K. (2006). TGFBI gene mutations in corneal dystrophies. *Human Mutation, 27*(7), 615-625.

Kannabiran, C., Sridhar, M., Chakravarthi, S., Vemuganti, G., & Lakshmipathi, M. (2005). Genotype-phenotype correlation in 2 Indian families with severe granular corneal dystrophy. *Arch Ophthalmol , 123*, 1127-33.

Karmel, M. (2010). Addressing the pain of corneal neuropathy. *EyeNet. American Association of Ophthalmology*.

Karring, H., Runager, K., Thogersen, I., Klintworth, G., Hojrup, P., & Enghild, J. (2012). Composition and proteolytic processing of corneal deposits associated with mutations in the TGFBI gene. *Exp Eye Res, 96*(1), 163-170.

Karring, H., Runager, K., Valnickova, Z., Thogersen, I., Moller-Pedersen, T., Klintworth, G., et al. (2009). Differential expression and processing of TGFBIp in the normal human cornea during postnatal development and aging. *Experimental Eye Research*.

Karring, H., Thogersen, I., Klintworth, G., Moller-Pedersen, T., & Enghild, J. (2005). A Dataset of Human Cornea Proteins Identified by Peptide Mass Fingerprinting and Tandem Mass Spectrometry. *Molecular Cell Proteomics, 4*, 1406-1408.

Kato, Y., Yagi, H., Kaji, Y., Oshika, T., & Goto, Y. (2013). Benzalkonium chloride accelerates the formation of the amyloid fibrils of corneal dystrophy-associated peptides. *J Biol Chem, 288*, 25109-25118.

Kattan, J., Serna-Ojeda, J. C., Sharma, A., Kim, E., Ramirez-Miranda, A., Cruz-Aguilar, M., et al. (2017). Vortex pattern of corneal deposits in granular corneal dystrophy associated with the p. (ArgR555WTrp) mutation in TGFBI. *Cornea, 36*(2), 210-216.

Kaur, H., Chaurasia, S. S., Agrawal, V., Suto, C., & Wilson, S. E. (2009). Corneal myofibroblast viability: opposing effects of IL-1 and TGF beta1. *Exp Eye Res.*(89), 152-158.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A., & Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science, 315*(5815), 1137-1140.

Kawasaki, S., Yamasaki, K., Nakagawa, H., Shinomiya, K., Nakatsukasa, M., Nakai, Y., et al. (2012). A novel mutation (p.Glu2389AspfsX16) of the phosphoinositide kinase, FYVE finger containing gene found in a Japanese patient with fleck corneal dystrophy. *Molec. Vis., 18*, 2954-2960.

Kaza, H., Barik, M., Reddy, M., Mittal, R., & Das, S. (2017). Gelatinous drop-like corneal dystrophy: a review. *Br J Ophthalmol, 101*, 10-15.

Keats B. S., & Rimoin D. P. (2013). Population genetics. *Emery and Rimoin's principles and practice of medical genetics* (Vol. 13, pp. 1-12).

Keenan, T., Jones, M., Rushton, S., & Carley, F. (2012). Trends in the indications for corneal graft surgery in the United Kingdom: 1999 Through 2009. *Archives of Ophthalmology, 130*, 621-628.

Kheir, V., Cortes-Gonzalez, V., Zenteno, J., & Schorderet, D. (2019). Mutation update: TGFBI pathogenic and likely pathogenic variants in corneal dystrophies. *Human Mutation, 40*(6).

Khodadoust, A. A., Silvestein, A. M., Kenyon, A. M., & Dowling, J. R. (1968). Adhesion of regenerating corneal epithelium. *American Journal of Ophthalmology, 3*(65), 339-348.

Kim, E., Meng, H., & Jun, A. (2013). Lithium treatment increases endothelial cell survival and autophagy in a mouse model of Fuchs endothelial corneal dystrophy. *Br J Ophthalmol, 97*(8), 1068-1073.

Kim, H., & Kim, I. (2008). Transforming growth factor-beta-induced gene product, as a novel ligand of integrin alphaMbeta2, promotes monocytes adhesion, migration and chemotaxis. *Int. J. Biochem. Cell Biol., 40*, 991-1004.

Kim, H., Luo, L., Pflugfelder, S., & Li, D. (2005). Doxycycline inhibits TGF-beta1-induced MMP-9 via Smad and MAPK pathways in human corneal epithelial cells. *Inves Ophthalmol Vis Sci, 46*(3), 840-848.

Kim, H., Yoon, S., Cho, B., Kim, E., & Joo, C. (2001). BIGH3 gene mutations and rapid detection in Korean patients with corneal dystrophy. *Cornea, 20*, 844-849.

Kim, J. E., Jeong, H. W., Nam, J. O., Lee, B. H., Choi, J. Y., Park, R. W., et al. (2002). Identification of motifs in the fasciclin domains of the transforming growth factor-beta-induced matrix protein betaig-h3 that interact with the alphavbeta5 integrin. *Journal of Biological Chemistry*(277), 46159–46165.

Kim, J. E., Kim, S. J., Lee, B. H., Park, R. W., Kim, K. S., & Kim, I. S. (2000). Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, beta ig-h3. *Journal of Biological Chemistry*(275), 30907-30915.

Kim, J. E., Park, R. W., Choi, J. Y., Bae, Y. C., Kim, K. S., Joo, C. K., et al. (2002). Molecular properties of wild-type and mutant beta-IG-H3 proteins. *Invest. Ophthal. Vis. Sci.*(43), 656-661.

Kim, J., Kim, E., Han, E., Park, R., Park, I., Jun, S., et al. (2000). A TGF-beta-inducible cell adhesion molecule, betaig-h3, is downregulated in melorheostosis and involved in osteogenesis. *J Cell Biochem, 77*, 169-178.

Kim, J., Kim, H., & Song, J. (2008). Phenotypic non-penetrance in granular corneal dystrophy type II. *Graefes Arch Clin Exp Ophthalmol, 246*, 1629-1631.

Kim, J., Kim, S., Lee, B., Park, R., Kim, K., & Kim, I. (2000). Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, betaig-h3. *J Biol Chem, 275*, 30907-30915.

Kim, M., Han, S., Park, T., Song, S., Lee, J., Lee, Y., et al. (2023). The TGF β →TAK1→ LATS→YAP1 Pathway Regulates the Spatiotemporal Dynamics of YAP1. *Mol Cells, 46*(10), 592-610.

Kim, T., Choi, S., Lee, H., Cho, Y., & Kim, E. (2008). Mitomycin C induces apoptosis in cultured corneal fibroblasts derived from type II granular corneal dystrophy corneas. *Mol Vis, 14*, 1222-1228.

Kim, T., Lee, H., Hong, H., Kim, K., Choi, S., Maeng, Y., et al. (2015). Inhibitory Effect of Tranilast on Transforming Growth Factor-Beta-Induced Protein in Granular Corneal Dystrophy Type 2 Corneal Fibroblasts. *Cornea, 34*(8), 950-958.

Kim, Y., Kwon, H., & Kim, D. (2012). Matrix metalloproteinase 9 (MMP-9)-dependent processing of βig-h3 protein regulates cell migration, invasion, and adhesion. *J Biol Chem, 287*, 38957-38969.

Kitamoto, K., Taketani, Y., Fujii, W., Inamochi, A., Toyono, T., Miyai, T., et al. (2020). Generation of mouse model of TGFBI-R124C corneal dystrophy using CRISPR/Cas9-mediated homology-directed repair. *Scientific Reports, 10*, doi.org/10.1038/s41598-020-58876-w.

Klinthworth, G. K. (2009). Corneal Dystrophies. *Orphanet Journal of Rare Diseases, 4*(7).

Klintworth, G. (1980). Research into the pathogenesis of macular corneal dystrophy. *Trans Ophthalmol Soc U K, 100*, 186-194.

Klintworth, G. (1994). Proteins in Ocular Disease. In A. Garner, & G. K. Klintworth (Eds.), *Pathobiology of Ocular Disease: A Dynamic Approach* (pp. 973-1032). New York.

Klintworth, G. (2003). THE MOLECULAR GENETICS OF THE CORNEAL DYSTROPHIES - CURRENT STATUS. *Frontiers in Bioscience*, 687-713.

Klintworth, G. (2009). Corneal dystrophies. *Orphanet J Rare Dis, 4*(7), PMC2695576.

Klintworth, G., Bao, W., & Afshari, N. (2004). Two mutations in the TGFBI (BIGH3) gene associated with lattice corneal dystrophy in an extensively studied family. *Invest. Ophthal. Vis. Sci., 45*, 1382-1388.

Knop, E., Knop, N., Millar, T., Obata, H., & Sullivan, D. (2011). The International Workshop on Meibomian Gland Dysfunction: Report of the Subcommittee on Anatomy, Physiology, and Pathophysiology of the Meibomian Gland. *Investigative Ophthalmology and Visual Science, 52*(4).

Kochairi, E. I., Letovanec, I., Uffer, S., Munier, F. L., Chaubert, P., & Schorderet, D. F. (2006). Systemic investigation of keratoepithelin deposits in TGFBI/BIGH3-related corneal dystrophy. *Molecular Vision*(10), 461-466.

Kolb, P., Ayaub, E., Zhou, W., Yum, V., Dickhout, J., & Ask, K. (2015). The therapeutic effects of 4-phenylbutyric acid in maintaining proteostasis. *The International Journal of Biochemistry and Cell Biology, 61*, 45-52.

Kolozsvari, L., Hopp, B., & Bor, Z. (2002). UV absorbance of the human cornea in the 240- to 400-nm range. *Invest Ophthalmol Vis Sci, 43*, 2165-8.

Konishi, M., Mashima, Y., Yamada, M., Kudoh, J., & Shimizu, N. (1998). The classic form of granular corneal dystrophy associated with R555W mutation in the BIGH3 gene is rare in Japanese patients. *American Journal of Ophthalmology, 126*, 450-452.

Konishi, M., Yamada, M., Nakamura, Y., & Mashima, Y. (2000). Immunohistology of kerato-epithelin in corneal stromal dystrophies associated with R124 mutations of the BIGH3 gene. *Curr Eye Res, 21*, 891-896.

Korvatska, E., Henry, H., Mashima, Y., Yamada, M., Bachmann, C., Munier, F., et al. (2000). Amyloid and non-amyloid forms of 5q31-linked corneal dystrophy resulting from kerato-epithelin mutations at Arg-124 are associated with abnormal turnover of the protein. *Journal of Biological Chemistry, 275*, 11465-11469.

Krafchak, C., Pawar, H., Moroi, S., Sugar, A., Lichter, P., Mackey, D., et al. (2005). Mutations in TCF8 cause posterior polymorphous corneal dystrophy and

ectopic expression of COL4A3 by corneal endothelial cells. *Am. J. Hum. Genet., 77*, 694-708.

Kreidberg, J. (2000). Functions of alpha3beta1 integrin. *Current Opinion in Cell Biology, 12*(5), 548-553.

Kuchle, M., Green, W., Volcker, H., & Barraquer, J. (1995). Reevaluation of corneal dystrophies of Bowman's layer and the anterior stroma (Reis-Bucklers and Thiel-Behnke types): a light and electron microscopic study of eight corneas and a review of the literature. *Cornea, 14*, 333-354.

Kukurba, K., & Montgomery, S. (2015). RNA Sequencing and Analysis. *Cold Spring Harb Protoc., 2015*(11), 951-969.

Kumar Bubna, A. (2015). Vorinostat—An Overview. *Indian Journal of Dermatology, 60*(4), 419.

Kuwabara, T. (1978). Current concepts in anatomy and histology of the cornea. *Contact Intraocular Lens Med J*(4), 101.

Kuwabara, T., & Ciccarelli, E. (1964). Meesmann's corneal dystrophy: a pathological study. *Arch. Ophthal., 1964*, 672-682.

Kwok, S., Wong, F., Shih, K., Chan, Y., Bu, Y., Chan, T., et al. (2020). Lycium barbarum Polysaccharide Suppresses Expression of Fibrotic Proteins in Primary Human Corneal Fibroblasts. *Journal of clinical medicine, 9*(11), 3572.

Ladewig, E., Michelini, F., Jhaveri, K., Castel, P., Carmona, J., Fairchild, L., et al. (2022). The Oncogenic PI3K-Induced Transcriptomic Landscape Reveals Key Functions in Splicing and Gene Expression Regulation. *Cancer Research, 82*(12), 2269-2279.

Lalitha, K., Meena, G., & Pallavi, J. (2021). Corneal epithelial toxicity following intraoperative use of mitomycin C during trabeculectomy - A case report. *ACTA Scientific Ophthalmology, 4*(6), 53-56.

Land, M., & Ferdinand, R. (1992). The evolution of eyes. *Annu Rev Neurosci, 15*, 1-29.

Last, J., Thomasy, S., Croasdale, C., Russel, P., & Murphy, C. (2012). Compliance profile of the human cornea as measured by atomic force microscopy. *Micron, 43*(12), 1293-1298.

Le, X. C., Nguyen, T. H., Hoang, M. C., Ton, K. T., Keiko, F., Akira, M., et al. (2004). Mutation Analysis of the TGFBI Gene in Vietnamese with Granular and Avellino Corneal Dystrophy. *Jpn J Ophthalmol , 48*, 12–16.

LeBaron, R. G., Bezverkov, K. I., Zimber, M. P., Pavelec, R., Skonier, J., & Purchio, A. F. (1995). Beta IG-H3, a novel secretory protein inducible by transforming growth factor-beta, is present in normal skin and promotes the adhesion and spreading of dermal fibroblasts in vitro. *Journal of Investigative Dermatology*(104), 844-849.

Lee, J. H., Cristol, S. M., Kim, W. C., Chung, E. S., Tchah, H., Kim, M. S., et al. (2010). Prevalence of Granular Corneal Dystrophy Type 2 (Avellino Corneal Dystrophy) in the Korean Population. *Ophthalmic Epidemiology, 17*(3), 160-165.

Lee, S.-H., Bae, J.-S., Park, S.-H., & Kim, I.-S. (2003). Expression of TGF-β-induced matrix protein βig-h3 is up-regulated in the diabetic rat kidney and human proximal tubular epithelial cells treated with high glucose. *Kidney international, 3*(64), 1012-1021.

Leng, N., Dawson, J., Thomson, J., Rissman, A., Smits, B., Haag, J., et al. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics, 29*(8), 1035-1043.

Leung, R. & Whittaker, P. (2005). RNA interference: from gene silencing to gene-specific therapeutics. *Pharmacol Ther*, *107*(2), 222-39.

Li, D., Lokeshwar, B., Solomon, A., Monroy, D., Ji, Z., & Pflugfelder, S. (2001). Regulation of MMP-9 Production by Human Corneal Epithelial Cells. *Exp Eye Research, 73*(4), 449-459.

Li, H., Zhang, H., Huang, G., Bing, Z., Xu, D., Liu, J., et al. (2022). Loss of RPS27a expression regulates the cell cycle, apoptosis, and proliferation via the RPL11-MDM2-p53 pathway in lung adenocarcinoma cells. *Journal of Experimental and Clinical Cancer Research, 41*, 33.

Li, J., Ma, J., Zhang, Q., Gong, H., Gao, D., Wang, Y., et al. (2022). Spatially resolved proteomic map shows that extracellular matrix regulates epidermal growth. *Nature Communications, 13*, 4012.

Li, L., Lin, X., Khvorova, A., Fesik, S., & Shen, Y. (2007). Defining the optimal parameters for hairpin-based knockdown constructs. *RNA, 13*(10), 1765-1774.

Liang, M., & Chuang, D. (2006). Differential roles of glycogen synthase kinase-3 isoforms in the regulation of transcriptional activation. *J Biol Chem, 281*, 30479-30484.

Liang, M., Wendland, J., & Chuang, D. (2008). Lithium inhibits Smad3/4 transactivation via increased CREB activity induced by enhanced PKA and AKT signalling. *Mol Cell Neurosci, 37*(3), 440-453.

Ligocki, A., Fury, W., Gutierrez, C., Alder, C., Yang, T., Ni, M., et al. (2021). Molecular characteristics and spatial distribution of adult human corneal cell subtypes. *Scientific Reports, 11*, 16323.

Lim, R., Tan, A., Liu, Y., Barathi, V., Mohan, R., Mehta, J., et al. (2016). ITF2357 transactivates Id3 and regulate TGFβ/BMP7 signaling pathways to attenuate corneal fibrosis. *Sci Rep, 6*, 20814.

Lin, J., Shen, X., Pfeifer, C., Shiau, F., Santeford, A., Ruzycki, P., et al. (2022). Dry eye disease in mice activates adaptive corneal epithelial regeneration distinct from constitutive renewal in homeostasis. *Proc Natl Acad Sci U S A, 120*(2), e2204134120.

Lin, Z., Chen, J., & Cui, H. (2016). Characteristics of corneal dystrophies: a review from clinical, histological and genetic perspectives. *Int. J. Ophthal., 9*, 904-913.

Linsenmayer, T. F., Gibney, E., Igoe, F., Gordon, M. K., Fitch, J. M., Fessler, L. I., et al. (1993). Type V collagen: molecular structure and fibrillar organization of the chicken alpha 1(V) NH2-terminal domain, a putative regulator of corneal fibrillogenesis. *Journal of Cellular Biology, 121*(5), 1181-1189.

Lisch, W., Buttner, A., Oeffner, F., Boddeker, I., Engel, H., Lisch, C., et al. (2000). Lisch corneal dystrophy is genetically distinct from Meesmann corneal dystrophy and maps to Xp22.3. *Am. J. Ophthal., 130*, 461-468.

Lisch, W., Steuhl, K., Lisch, C., Weidle, E., Emmig, C., Cohen, K., et al. (1992). A new, band-shaped and whorled microcystic dystrophy of the corneal epithelium. *Am. J. Ophthal., 114*, 35-44.

Liskova, P., Gwilliam, R., Filipec, M., Jirsova, K., MerjavaS, Deloukas, P., et al. (2012). High prevalence of posterior polymorphous corneal dystrophy in the Czech Republic: linkage disequilibrium mapping and dating an ancestral mutation. *PLoS, 7*, e45495.

Little, J. M., Centifanto, Y. M., & Kaufman, H. E. (1969). Immunoglobulins in human tears. *American Journal of Ophthalmology*(68), 898-905.

Liu, H., Yuan, Y., Zhang, J., Jone, A., & Kao, W. (2012). A Dry Eye Model By Overexpression Of Interleukin-1β In Corneal Epithelium Of Krt12rtta/Tet-O-IL1b Mice. *IOVS, 53*(14), 2334.

Liu, Y., Huang, H., Sun, G., Alwadani, S., Semba, R., Lutty, G., et al. (2017). Gene Expression Profile of Extracellular Matrix and Adhesion Molecules in the Human Normal Corneal Stroma. *Curr Eye Res., 42*(4), 520-527.

Liu, Y., Xu, C., Li, J., Zhang, Y., Wang, X., Wang, Y., et al. (2023). YAP promotes AP-1 expression in tubular epithelial cells in the kidney. *Molecular Basis of Kidney Injury and Repair, 324*(6), 581-589.

Liu, Y., Yeung, W., Chiu, P., & Cao, D. (2022). Computational approaches for predicting variant impact: An overview from resources, principles to applications. *Frontiers in Genetics, 13*, 10.3389/fgene.2022.981005.

Liu, Z., Sheng, J., Peng, G., Yang, J., Chen, W., & Li, K. (2018). TGF-β1 regulation of P-JNK and L-type calcium channel Cav1.2 in cortical neurons. *J Mol Neurosci, 64*, 374-384.

Ljubimov, A. V., Burgeson, R. E., Butkowski, R. J., Michael, A. F., Sun, T. T., & Kenny, M. C. (1995). Human corneal basement membrane heterogeneity: topographical differences in the expression of type IV collagen and laminin isoforms. *Lab Invest., 4*(72), 461-473.

Long, Y., Gu, Y.-s., Han, W., Li, X.-y., Yu, P., & Qi, M. (2011). Genotype-phenotype correlations in Chinese patients with TGFBI gene-linked corneal dystrophy. *Journal of Zhejiang University SCIENCE B, 12*(4), 287-292.

Louro, B., Marques, J., Manchado, M., Power, D., & Campinho, M. (n.d.). Sole head transcrip-tomics reveals a coordinated developmental program during metamorphosis. *Genomics, 112*(1), 592-602.

Lu, L., Reinach, P., & Kao, W. (2001). Corneal Epithelial wound healing. *Exp Biol Med (Maywood), 226(7)*, 653-64.

Lu, N., Carracedo, S., Ranta, J., Heuchel, R., Soininen, R., & Gullberg, D. (2010). The human alpha11 integrin promoter drives fibroblast-restricted expression in vivo and is regulated by TGF-beta1 in a Smad- and Sp1-dependent manner. *Matrix Biol, 29*(3), 166-167.

Lundmark, P., Pandi-Perumal, S., Srinivasan, V., Cardinali, D., & Rosenstein, R. (2007). Melatonin in the eye: implications for glaucoma. *Experimental Eye Research, 84*(6), 1021-1030.

Luo, K. (2017). Signaling Cross Talk between TGF-β/Smad and Other Signaling Pathways. *Cold Spring Harb Perspect Biol, 9*(1), a022137.

Lyons, C., McCartney, A., Kirkness, C., Ficker, L., Steele, A., & Rice, N. (1994). Granular corneal dystrophy. Visual results and pattern of recurrence after lamellar or penetrating keratoplasty. *Ophthalmology, 101*(11), 1812-1817.

Ma, H. (2017). HeLa cella and immortality. *Cancer Biology*, 7(3), 71-78.

Ma, X., & Leng, N. (2023). *EBSeq: An R package for gene and isoform differential expression analysis of RNA-seq data.* Retrieved from Bioconductor: https://bioconductor.org/packages/release/bioc/html/EBSeq.html

Maddineni, P., Kasetti, R., Kodati, B., Yacoub, S., & Zode, G. (2021). Sodium 4-Phenylbutyrate Reduces Ocular Hypertension by Degrading Extracellular Matrix Deposition via Activation of MMP9. *International Journal of Molecular Sciences, 22*(18), 10095.

Malavet, P. (2000). Puerto rico: cultural nation, American Colony. *Michigan J Race Law, 6*, 1-106.

Malhi, G., Tanious, M., Das, P., Coulston, C., & Berk, M. (2013). Potential mechanisms of action of lithium in bipolar disorder. *CNS drugs, 27*(2), 135-153.

Malkondu, F., Arıkoğlu, H., Erkoç Kaya, D., Bozkurt, B., & Özkan, F. (2020). Investigation of TGFBI (transforming growth factor beta-induced) Gene Mutations in Families with Granular Corneal Dystrophy Type 1 in the Konya Region. *Turk J Ophthalmol, 50*(2), 64-70.

Marfut, C., Cox, J., Deek, S., & Dvorscak, L. (2010). Anatomy of the human corneal innervation . *Experimental Eye Research*, 478-492.

Mashima, Y., Konishi, M., Nakamura, Y., Imamura, Y., Yamada, M., Ogata, T., et al. (1998). Severe form of juvenile corneal stromal dystrophy with homozygous R124H mutation in the keratoepithelin gene in five Japanese patients. *British Journal of Ophthalmology, 82*, 1280-1284 .

Mashima, Y., Nakamura, Y., Noda, K., Konishi, M., Yamada, M., Kudoh, J., et al. (1999). A Novel Mutation at Codon 124 (R124L) in the BIGH3 Gene is Associated with a Superficial Variant of Granular Corneal Dystrophy. *Arch Ophthalmol, 117*(1), 90-93.

Mashima, Y., Yamamoto, S., Inoue, Y., Yamada, M., Konishi, M., Watanabe, H., et al. (2000). Association of autosomal dominantly inherited corneal dystrophies with BIGH3 gene mutations in Japan. *American Journal of Ophthalmology, 130*, 516-517.

Massoudi, D., Malecaze, F., & Galiacy, S. (2016). Collagens and proteoglycans of the cornea: importance in transparency and visual disorders. *Cell and Tissue Research, 363*(2), 337-349.

Maurice, D. (1957). The Structure and Transparency of the Cornea. *The Journal of Physiology, 136*(2), 263-286.

Maurice, D. (1962). Clinical physiology of the cornea. *Int Ophthalmol Clin, 2*, 561-72.

Mazzotta, C., Traversi, C., Baiocchi, S., Barabino, S., Mularoni, A., & Ferreras, A. (2015). Phenotypic Spectrum of Granular Corneal Dystrophy Type II in Two Italian Families Presenting an Unusual Granular Corneal Dystrophy Type I Clinical Appearance. *Hindawi Journal, Case Reports in Ophthalmological Medicine, 2015*.

McCarthy, M., Innis, S., Dubord, P., & White, V. (1994). Panstromal Schnyder corneal dystrophy: a clinical pathologic report with quantitative analysis of corneal lipid composition. *Ophthalmology, 101*, 895-901.

McCulley, J. P., & Shine, W. (1997). A compositional based model for the tear film lipid layer. *Transactions of the American Opthalmological Society*(95), 79-88.

McDermott, A. (2013). Antimicrobial Compounds in Tears. *Exp Eye Res*(117), 53-61.

McKusick, V., & Hamosh, A. (n.d.). *Online Mendelian Inheritance in Man*. Retrieved October 2020, from https://www.omim.org

Meek, K. (2003). Transparency, swelling and scarring in the corneal stroma. *Eye, 17*(8), 927-36.

Meek, K., & Knupp, C. (2015). Corneal structure and transparency. *Progress in Retinal and Eye Research, 49*, 1-16.

Mello, C., & Conte, D. (2004). Revealing the world of RNA interference. *Nature, 431*(7006), 338-342.

Mercieca, S. (2015). Anatomies of Spanish settlers in malta between 1580 and 1648: their family stories. *Symposia Melitensia, 11*, 145-170.

Micheal, S., Siddiqui, S., Zafar, S., Gabriela Niewold, I., Khan, M., & Bergen, A. (2019). Identification of a Novel ZNF469 Mutation in a Pakistani Family With Brittle Cornea Syndrome. *Cornea, 38*(6), 718-722.

Michelacci, Y. (2003). Collagens and proteoglycans of the corneal extracellular matrix. *Brazilian Journal of Medical and Biological Research, 36*, 1037-1046.

Milan, M., Pace, V., Maiullari, F., Chirivì, M., Denisa Baci, D., Maiullari, S., et al. (2018). Givinostat reduces adverse cardiac remodeling through regulating fibroblasts activation. *Cell Death and Disease, 9*(2), 108.

Miyazawa, K., & Miyazono, K. (2017). Regulation of TGF-β Family Signaling by Inhibitory Smads. *Cold Spring Harb Perspect Biol., 9*(3), a022095.

Mizzi, C. (2016). How does Bioinformatics Analysis of Next Generation Sequences on Critical Families and Populations serve to Identify Globin Gene Control

Mechanism and Maltese Genetic Variations? Faculty of Medicine and Surgery, University of Malta. PhD thesis.

Mohammadi, S., Ashrafi, E., Norouzi N, Abdolahinia, T., Mir-Abou Talebi, M., & Jabbarvand, M. (2014). Effects of mitomycin C on tear film, corneal biomechanics, and surface irregularity in mild to moderate myopic surface ablasion: preliminary results. *Journal of Cataract and Refractive Surgery, 40*(6), 937-942.

Mohan, S., Gogri, P., Murthy, S., Chaurasia, S., Mohamed, A., & Dongre, P. (2021). Evaluation of the Effect of Mitomycin-C on Corneal Endothelium after Photorefractive Keratectomy for Myopia Correction. *Middle East Afr J Ophthalmol, 28*(2), 111-115.

Møller, H. (1989). Granular corneal dystrophy Groenouw type I (GrI) and Reis-Bücklers' corneal dystrophy (R-B). One entity? *Acta Ophthalmol (Copenh) , 67*, 678-684.

Moller, H. U. (1990). Granular corneal dystrophy Groenouw type I 115 Danish patients An epidemiological and genetic population study. *Acta Ophthalmologica, 68*, 287-303.

Moller, H., Bojsen-Moller, M., Schroder, M., Nelson, M., & Gegge, T. (1993). Immunoglobulins in granular corneal dystrophy Groenouw type I. *Acta Ophthalmol*, 71548-71551.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nat Commun, 24*(6), 7743.

Mongiat, M., Taylor, K., Otto, J., Aho, S., Uitto, J., Whitelock, J. M., et al. (2000). The protein core of the proteoglycan perlecan binds specifically to fibroblast growth factor-7. *Journal of Biological Chemistry, 10*(275), 7095-7100.

Montenegro, L., Lerario, A., Nishi, M., Jorge, A., & Mendonca, B. (2021). Performance of mutation pathogenicity prediction tools on missense variants associated with 46,XY differences of sex development. *Clinics, 76*, e2052.

Moore, C., Guthrie, E., Huang, M., & Taxman, D. (2010). Short Hairpin RNA (shRNA): Design, Delivery, and Assessment of Gene Knockdown. *Methods Mol Biol, 629*, 141-158.

Muller, L. (1997). Architecture of human corneal nerves. *Invest Ophthalmol Vis Sci, 38*(5), 985-94.

Muller, L. (2003). Corneal nerves: structure, contents and function. *Exp Eye Res, 76*(5), 521-42.

Muller, L., Pels, E., Schurmans, L., & Vrensen, G. (2004). A new three-dimensional model of the organisation of proteoglycans and collagen fibrils in the human corneal stroma. *Exp Eye Res, 78*(3), 493-501.

Munier, F. L., Korvatska, E., Djemai, A., Le Paslier, D., Zografos, L., Pescia, G., et al. (1997). Kerato-epithelin mutations in four 5q31-linked corneal dystrophies. *Nature Genetics*(15), 247-251.

Munier, F., Frueh, B., Othenin-Girard, P., Uffer, S., Cousin, P., Wang, M., et al. (2002). BIGH3 Mutation Spectrum in Corneal Dystrophies. *Investigative Ophthalmology & Visual Science, 43*(4), 949-954.

Murphy, C., Alvarado, J., Juster, R., & Maglio, M. (1984). Prenatal and Postnatal cellularity of the human corneal endothelium. A quantitative histologic study. *Invest Ophthalmol Vis Sci, 25*(3), 312-22.

Musch, D. C., Niziol, L. M., Stein, J. D., Kamyar, R. M., & Sugar, A. (2011). Prevalence of Corneal Dystrophies in the United States: Estimates from Claims Data. *Investigative Ophthalmology and Visual Science, 52*(9), 6959-6963.

Nagai, N., Ono, H., Hashino, M., Ito, Y., Okamoto, N., & Shimomura, Y. (2014). Improved corneal toxicity and permeability of tranilast by the preparation of ophthalmic formulations containing its nanoparticles. *J Oleo Sci, 63*(2), 177-186.

Nakamura, K., Terai, Y., Tanabe, A., Ono, Y., Hayashi, M., Maeda, K., et al. (2017). CD24 expression is a marker for predicting clinical outcome and regulates the epithelial-mesenchymal transition in ovarian cancer via both the Akt and ERK pathways. *Oncology reports, 37*(6), 1791-2431.

Nam, E., Sa, K., You, D., Cho, J., Seo, J., Han, S., et al. (2006). Up-regulated transforming growth factor beta-inducible gene h3 in rheumatoid arthritis mediates adhesion and migration of synoviocytes through αvβ3 integrin: Regulation by cytokines. *Arthritis Rheum, 54*, 2734-2744.

nature. (2018). *nature.com*. Retrieved August 24, 2018, from https://www.nature.com/subjects/next-generation-sequencing

Navarro-Gil, F., Huete-Toral, F., Domínguez-Godínez, C., Carracedo, G., & Crooke, A. (2022). Contact Lenses Loaded with Melatonin Analogs: A Promising Therapeutic Tool against Dry Eye Disease. *J Clin Med, 11*(12), 3483.

NCBI, Sayers, E., Bolton, E., Brister, J., Canese, K., Chan, J., et al. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 7*(50), D20-D26.

Nelson, E., Huang, C., Ewel, J., Chang, A., & Yuan, C. (2012). Halofuginone down-regulates Smad3 expression and inhibits the TGFbeta-induced expression of fibrotic markers in human corneal fibroblasts. *Molecular Vision, 18*, 479-487.

Ng, P., & Henikoff. (2001). (Bioinformatics Institute, Singapore) Retrieved October 2020, from Sorting Intolerant From Tolerant: https://sift.bii.a-star.edu.sg/

Nickerson, M., Bosley, A., Weiss, J., Kostiha, B., Hirota, Y., Brandt, W., et al. (2103). The UBIAD1 prenyltransferase links menaquinone-4 synthesis to cholesterol metabolic enzymes. *Hum. Mutat., 34*, 317-329.

Nie, D., Liu, X., Wang, Y., He, W., Li, M., Peng, Y., et al. (2020). Involvement of the JNK signaling in granular corneal dystrophy by modulating TGF-β-induced TGFBI expression and corneal fibroblast apoptosis. *In Vitro Cell Dev Biol Anim, 56*(3), 234-242.

Nie, D., Peng, Y., Li, M., Liu, X., Zhu, M., & Ye, L. (2018). Lithium chloride (LiCl) induced autophagy and downregulated expression of transforming growth factor β-induced protein (TGFBI) in granular corneal dystrophy. *Experimental Eye Research, 173*, 44-50.

Niel-Butschi, F., Kantelip, B., Iwaszkiewicz, J., Zoete, V., Boimard, M., Delpech, M., et al. (2011). Genotype-phenotype correlations of TGFBI p.Leu509Pro, p.Leu509Arg, p.Val613Gly, and the allelic association of p.Met502Val-p.Arg555Gln mutations. *Mol Vis, 17*, 1192-1202.

Nielsen, N., Poulsen, E., Lukassen, M., Shern, C., Mogensen, E., Weberskov, C., et al. (2020). Biochemical mechanisms of aggregation in TGFBI-linked corneal dystrophies. *Prog Retin Eye Res, 77*, 100843.

Nielsen, R., Akey, J., Jakobsson, M., Pritchard, J., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature, 541*, 302-310.

Nikolova, G., Strilic, B., & Lammert, E. (2007). The vascular niche and its basement membrane. *Trends Cell Biol.*(17), 19-25.

Nishida, K., Adachi, W., Shimizu-Matsumoto, A., Kinoshita, S., Mizuno, K., Matsubara, K., et al. (1996). A gene expression profile of human corneal epithelium and the isolation of human keratin 12 cDNA. *Invest Ophthalmol Vis Sci, 37*(9), 1800-1809.

Niu, J., Liu, J., Liu, L., Lu, Y., Chen, J., Xu, J., et al. (2012). Construction of eukaryotic plasmid expressing human TGFBI and its influence on human corneal epithelial cells. *Int J Ophthalmol, 5*(1), 38-44.

Nowińska, A. K., Wylegala, E., Janiszewska, D. A., & Dobrowolski, D. (2011). Genotype-phenotype correlation of TGFBI corneal dystrophies in Polish patients. *Molecular Vision, 17*, 2333-2342.

Nowinska, A., Wylegala, E., Teper, S., Wróblewska-Czajka, E., Aragona, P., Roszkowska, A., et al. (2014). Phenotype and genotype analysis in patients with macular corneal dystrophy. *British Journal of Ophthalmology, 98*, 1514-1521.

Nuwormegbe, S., Park, N., & Kim, S. (2021). Lobeglitazone attenuates fibrosis in corneal fibroblasts by interrupting TGF-beta-mediated Smad signaling. *Greafe's archive for clinical and experimental ophthalmology, 260*(1), 149-162.

O'Grady, T., Baddoo, M., & Flemington, E. (2017). Analysis of EBV Transcription Using High-Throughput RNA Sequencing. *Epstein Barr Virus*, 105-121.

Okada, M., Yamamoto, S., Watanabe, H., Inoue, Y., Tsujikawa, M., Maeda, N., et al. (1998). Granular corneal dystrophy with homozygous mutations in the kerato-epithelin gene. *American Journal of Ophthalmology, 162*(2), 169-176.

Okumura, N., Kay, E., Nakahara, M., Hamuro, J., Kinoshita, S., & Koizumi, N. (2013). Inhibition of TGF-b signaling enables human corneal endothelial cell expansion in vitro for use in regenerative medicine. *PLOS one, 8*(2), e58000.

Orphanet. (n.d.). Retrieved September 2020, from https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=en&Expert=98954.

Orr, A., Dube, M., Marcadier, J., Jiang, H., Federico, A., George, S., et al. (2007). Mutations in the UBIAD1 gene, encoding a potential prenyltransferase, are causal for Schnyder crystalline corneal dystrophy. *PloS One, 2*, e685.

Ouerdane, Y., Zaazouee, M., Mohamed, M., Hasan, M., Hamdy, M., Ghoneim, A., et al. (2021). Mitomycin C application after photorefractive keratectomy in high, moderate, or low myopia: Systematic review and meta-analysis. *Indian J Ophthalmol, 69*(12), 3421-3431.

Oxford. (2016). *Oxford Dictionaries*, Web. Retrieved February 22, 2016, from http://oxforddictionaries.com/definition/english/dystrophy

Pacella, E., Pacella, F., De Paolis, G., Romana Parisella, F., Turchetti, P., Anello, G., et al. (2015). Glycosaminoglycans in the Human Cornea: Age-Related Changes. *Ophthalmology and Eye Diseases, 7*, 1-5.

Paliwal, P., Gupta, J., & Tandon, R. (2010). Identification and characterization of a novel TACSTD2 mutation in gelatinous drop-like corneal dystrophy. *Molecular Vision, 16*, 729-739.

Paliwal, P., Sharma, A., Tandon, R., Sharma, N., Titiyal, J. S., Sen, S., et al. (2010). TGFBI mutation screening and genotype-phenotype correlation in north Indian patients with corneal dystrophies. *Molecular Vision*(16), 1429-1438.

Pal-Ghosh , S., Pajoohesh-Ganji, A., Tadvalkar, G., & Stepp, M. A. (2011). Removal of the basement membrane enhances corneal wound healing. *Exp Eye Res*(93), 927–936.

Pang, C., & Lam, D. (2002). Differential occurrence of mutations causative of eye diseases in the Chinese population. *Human Mutations, 19*, 189-208.

Pappa, A., Estey, T., Manzer, R., Brown, D., & Vasiliou, V. (2003). Human aldehyde dehydrogenase 3A1 (ALDH3A1): biochemical characterization and immunohistochemical localization in the cornea. *Biochem J, 376*(2), 615-623.

Park, J.-H., Kim, M., Yim, B., & Park, C. (2021). Nitric oxide attenuated transforming growth factor-beta induced myofibroblast differentiation of human keratocytes. *Nature: Scientific reports, 11*(1), DOI:10.1038/s41598-021-87791-x.

Park, S., Jung, M., & Kim, I. (2009). Stabilin-2 mediates homophilic cell-cell interactions via its FAS1 domains. *FEBS Lett., 583*, 1375-1380.

Partridge, M., Green, M., Langdon, J., & Feldmann, M. (1989). Production of TGF-alpha and TGF-beta by cultured keratinocytes, skin and oral squamous cell carcinomas--potential autocrine regulation of normal and malignant epithelial cell proliferation. *British Journal of Cancer, 60*(4), 542-548.

Passamonti, F., Maffioli, M., & Caramazza, D. (2012). New generation small-molecule inhibitors in myeloproliferative neoplasms. *Curr Opin Hematol, 19*, 117-123.

Patel, D., Chang, S., Harocopos, G., Vora, S., Thang, D., & Huang, A. (2010). Granular and lattice deposits in corneal dystrophy caused by R124C mutation of TGFBIp. *Cornea, 29*, 1215-22.

Patricelli, C., Lehmann, P., Oxford, J., & Pu, X. (2023). Doxorubicin-induced modulation of TGF-β signaling cascade in mouse fibroblasts: insights into cardiotoxicity mechanisms. *Scientific Reports, 13*, 18944.

Paul, P., Raiput, S., Joshi, P., Naithani, M., Chowdhury, N., Rao, S., et al. (2021). Comparison of Fluorometric and UV Spectrophotometric Findings for DNA Isolated From Formalin-Fixed Paraffin-Embedded Blocks, Fine Needle Aspiration Cytology Smears, and Blood. *Cureus, 13*(11), e19583.

Phillips, C. (2015). Forensic genetic analysis of bio-geographical ancestry. *Forensic Sci Int Genet, 18*, 49-65.

Piersma, B., Bank, R., & Boersema, M. (2015). Signalling in fibrosis: TGF-β, Wnt, and YAP/TAZ converge. *Frontiers in Medicine, 2*(10), 3389.

Poulsen, E., Runager, K., Nielsen, N., Lukassen, M., Thomsen, K., Snider, P., et al. (2018). Proteomic profiling of TGFBI-null mouse corneas reveals only minor changes in matrix composition supportive of TGFBI knockdown as therapy against TGFBI-linked corneal dystrophies. *FEBS, 285*(1), 101-114.

Promega. (2021). *ReliaPrep™ RNA Tissue Miniprep System Technical Manual.* Retrieved from https://worldwide.promega.com/-/media/files/resources/protocols/technical-manuals/101/reliaprep-rna-tissue-miniprep-system-protocol.pdf?rev=33df947efabb4114a72b7a358908fa60&sc_lang=en.

Promega. (n.d.). *Quantus™ Fluorometer*. Retrieved October 25, 2108, from https://www.promega.com/-/media/files/resources/protocols/technical-manuals/101/quantus-fluorometer-operating-manual.pdf

Qiagen. (2016). *QIAamp® DNA Mini and Blood Mini Handbook.*

QIAGEN. (n.d.). *QIAGEN*. Retrieved October 25, 2018, from https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/qiaamp-dna-mini-kit/#resources

Qin, Z., Xia, W., Fisher, G., Voorhees, J., & Quan, T. (2018). YAP/TAZ regulates TGF-β/Smad3 signaling by induction of Smad7 via AP-1 in human skin dermal fibroblasts. *Cell Communucation and Signaling, 16*, 18.

Ranganathan, K., & Sivasankar, V. (2014). MicroRNAs - Biology and clinical applications. *JnOral Maxillofac Pathol, 18*(2), 229-234.

Rao, R., & Shields, C. (2019). Overview and Classification of Conjunctival and Corneal Tumors. In H. S. Chaugule S (Ed.), *Surgical Ophthalmic Oncology.*

Rare Disease Database. (n.d.). Retrieved November 6, 2020, from National Organisation for Rare Disorders: https://rarediseases.org/rare-diseases/corneal-dystrophies/

Reinboth, B., Thomas, J., Hanssen, E., & Gibson, M. (2006). Beta ig-h3 interacts directly with biglycan and decorin, promotes collagen VI aggregation, and participates in ternary complexing with these macromolecules. *J. Biol. Chem., 281*, 7816-7824.

Reis, W. (1917). Fämiliare, fleckige Hornhautetartung. *Dtsch Med Wochenschr, 43*, 575.

Ren, Z., Lin, P., & Klintworth, G. (2002). Allelic and locus heterogeneity in autosomal recessive gelatinous drop-like corneal dystrophy. *Human Genetics, 110*, 568-577.

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the Functional Impact of Protein Mutations: Application to Cancer Genomics. *Nucleic Acids Research*.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W., & Khvorova, A. (2004). Rational siRNA design for RNA interference. *Nat Biotechnol, 22*(3), 326-330.

Riazuddin, S., Zaghloul, N., Al-Saif, A., Davey, L., Meadows, D., Eghrari, A., et al. (2010). Missense mutations in TCF8 cause late-onset Fuchs corneal dystrophy and interact with FCD4 on chromosome 9p. *Am J Hum Genet, 86*(1), 45-53.

Rodrigues, M., & Robey, P. (1982). C-reactive protein in human lattice corneal dystrophy. *Curr. Eye Res, 2*(10), 721–4.

Rodrigues, M., Streeten, B., Krachmer, J., & al, e. (1983). Microfibrillar protein and phospholipid in granular corneal dystrophy. *Invest Ophthalmol Vis Sci, 24*, 274.

Rozzo, C., Fossarello, M., Galleri, G., Sole, G., Serru, A., Orzalesi, N., et al. (1998). A common beta ig-h3 gene mutation (delta f540) in a large cohort of Sardinian Reis Bucklers corneal dystrophy patients. *Human Mut , 12*, 215-216.

Runager, K., Basaiawmoit , R. V., Deva, T., Andreasen, M., Valnickova, Z., Sorensen, C. S., et al. (2011). Human phenotypically distinct TGFBI corneal dystrophies are linked to the stability of the fourth FAS1 domain of TGFBIp. *Journal of Biological Chemistry*(286), 4951-4958.

Runager, K., Enghild, J., & Klintworth, G. (2008). Focus on molecules: Transforming growth factor beta induced protein (TGFBIp). *Experimental Eye Research, 87*(4), 298-299.

Runager, K., Klintworth, G., Karring, H., & Enghild, J. (2013). The Insoluble TGFBIp-fraction of the Cornea is Covalently Linked via a Disulfide Bond to Type XII Collagen. *Biochemistry, 52*(16), 2821-2827.

Saarela, J., & Kettunen, K. (2017). Who would benefit from exome sequencing? *Duodecim., 133*(5), 481-488.

Santo, R., Yamaguchi, T., Kanai, A., Okisaka, S., & Nakajima, A. (1995). Clinical and histopathologic features of corneal dystrophies in Japan. *Ophthalmology, 102*, 557-567.

Sarenac, T., Trapecar, M., Gradisnik, L., Rupnik, M., & Pahor, D. (2016). Single-cell analysis reveals IGF-1 potentiation of inhibition of the TGF-[beta]/Smad

pathway of fibrosis in human keratocytes in vitro. *Scientific Reports (Nature), 6*, 34373.

Sarkar, S., Panikker, P., D'Souza, S., Shetty, R., Mohan, R., & Ghosh, A. (2023). Corneal Regeneration Using Gene Therapy Approaches. *Cells, 12*(9), 1280.

Scheller, J., Chalaris, A., Schmidt-Arras, D., & Rose-John, S. (2011). The pro- and anti-inflammatory properties of the cytokine interleukin-6. *BBA Molecular cell research, 1813*(5), 878-888.

Schiroli, D., Gomara, M., Maurizi, E., Atkinson, S., Mairs, L., Christie, K., et al. (2019). Effective in vivo topical delivery of siRNA and gene silencing in intact corneal epithelium using a modified cell-penetrating peptide. *Molecular therapy, 17*, 891-906.

Schmedt, T., Mazzini Silva, M., Ziaei, A., & Jurkunas, U. (2012). Molecular bases of corneal endothelial dystrophies. *Experimental Eye Research, 95*(1), 24-34.

Sethi, A., Mao, W., Wordinger, R., & Clark, A. (2011). Transforming Growth Factor–β Induces Extracellular Matrix Protein Cross-Linking Lysyl Oxidase (LOX) Genes in Human Trabecular Meshwork Cells. *Invest Ophthalmol Vis Sci, 52*(8), 5240-5250.

Shah, S., al-Rajhi, A., Brandt, J., Mannis, M., Roos, B., Sheffield, V., et al. (2008). Mutation in the SLC4A11 gene associated with autosomal recessive congenital hereditary endothelial dystrophy in a large Saudi family. *Ophthalmic Genet, 29*, 41-45.

Sharma, A., Sinha, N., Siddiqui, S., & Mohan, R. (2015). Role of 5'TG3'-interacting factors (TGIFs) in Vorinostat (HDAC inhibitor)-mediated Corneal Fibrosis Inhibition. *Mol Vis, 21*, 974-984.

Shastry, B. (2007). SNPs in disease gene mapping, medicinal drug development and evolution. *Journal of Human Genetics, 52*(11), 871-880.

Shen, C. (2023). Gene Expression. *Diagnostic Molecular Biology.*

Shen, X., Hu, P., Liberati, N., Datto, M., Frederick, J., & Wang, F. (1998). TGF-β–induced Phosphorylation of Smad3 Regulates Its Interaction with Coactivator p300/CREB-binding Protein. *Molecular Biology of the Cell, 9*(12), 3309-3319.

Shi, Y., & Massague, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell, 113*, 685-700.

Siddiqui, S. (2016). Genetic Analysis of Corneal Dystrophies. University of Leeds.

Silbiger, S., Lei, J., Ziyadeh, F., & Neugarten, J. (1998). Estradiol reverses TGF-β1-stimulated type IV collagen gene transcription in murine mesangial cells. *Renal Physiology, 274*(6), 1113-1118.

Singh, R., Parmar, U., Kahale, F., Jeng, B., & Jhanji, V. (2023). Prevalence and Economic Burden of Fuchs Endothelial Corneal Dystrophy in the Medicare Population in the United States. *Cornea*, 37906001.

Singh, V., Santhiago, M. R., Barbosa, F. L., Agrawal, V., Singh, N., Ambati, B. K., et al. (2011). Effect of TGFbeta and PDGF-B blockade on corneal myofibroblast development in mice. *Exp Eye Res.*(93), 810-817.

Sirugo, G., Williams, S., & Tishkoff, S. (2019). The Missing Diversity in Human Genetic Studies. *Cell, 177*(1), 26-31.

Skonier, J., Neubauer, M., Madisen, L., Bennett, K., Plowman, G. D., & Purchio, A. F. (1992). cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. *DNA Cell Biology*(11), 511-522.

Snodgrass, A., & Motaparthi, K. (2021). Systemic Antibacterial Agents. In S. Wolverton, *Comprehensive Dermatologic Drug Therapy (Fourth Edition).* Science Direct.

Solari, H., Ventura, M., Perez, A., Sallum, J., Burnier Jr, M., & Belfort Jr, R. (2007). TGFBI gene mutations in Brazilian patiens with corneal dystrophy. *Eye , 21*, 587–590.

Son, J., Jeong, H., Park, D., No, S., Lee, E., Lee, J., et al. (2017). miR-10a and miR-204 as a potential prognostic indicator in low-grade gliomas. *Cancer Informatics, 16*, 1176935117702878.

Spichenok, O., Budimlija, Z., Mitchell, A., Jenny, A., Kovacevic, L., Marjanovic, D., et al. (2011). Prediction of eye and skin colour in diverse populations using seven SNPs. *Forensic Sci Int Genet, 5*(5), 472-478.

Srivastava, O. P., & Srivastava, K. (1999). cAMP-dependent Phosphorylation of Betaig-h3 Protein in Human Corneal endothelial cells. *Curr Eye Res, 19*(4), 348-357.

Stachon, T., Nastaranpour, M., Seitz, B., Meese, E., Latta, L., Taneri, S., et al. (2022). Altered Regulation of mRNA and miRNA Expression in Epithelial and Stromal Tissue of Keratoconus Corneas. *IOVS, 63*(8), doi.org/10.1167/iovs.63.8.7.

Stelzer, G., Rosen, R., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics, 54*(1), 10.1002.

Stenvag, M., Andreasen, M., Enghild, J., & Otzen, D. (2013). *Bio-nanoimaging: Protein Misfolding and Aggregation.* (V. Uversky, & Y. Lyubchenko, Eds.) Academic Press.

Stenvang, M., Schafer, N., Malmos, K., Perez, A., Niembro, O., Sormanni, P., et al. (2018). Corneal Dystrophy Mutations Drive Pathogenesis by Targeting TGFBIp Stability and Solubility in a Latent Amyloid-forming Domain. *J Mol Biol, 430*, 1116-1140.

Stewart, H., Ridgway, A., Dixon, M., Bonshek, R., Parveen, R., & Black, G. (1999). Heterogeneity in granular corneal dystrophy: identification of three causative mutations in the TGFBI (BIGH3) gene-lessons for corneal amyloidogenesis. *Human Mutation, 14*(2), 126-132.

Stix, B., Leber, M., Bingemer, P., Gross, C., Ruschoff, J., Fandrich, M., et al. (2005). Hereditary Lattice Corneal Dystrophy Is Associated with Corneal Amyloid Deposits Enclosing C-Terminal Fragments of Keratoepithelin. *Investigative Ophthalmology & Visual Science, 46*, 1133-1139.

Stone, E., Mathers, W., Rosenwasser, G., Holland, E., Folberg, R., Krachmer, J., et al. (1994). Three autosoma ldominant corneal dystrophies map to chromosome 5q. *Nat Genet, 6*, 47-51.

Streeten, B., Qi, Y., Klintworth, G., Eagle, R., Strauss, J., & Bennett, K. (1999). Immunolocalization of βig-h3 Protein in 5q31-Linked Corneal Dystrophies and Normal Corneas. *Arch Ophthalmol, 117*(1), 67-75.

Stuart, J., Mund, M., Iwamoto, T., Troutman, R., White, H., & DeVoe, A. (1975). Recurrent granular corneal dystrophy. *Am J Ophthalmol, 79*, 18-24.

Sun, T.-T., & Vidrich, A. (1981). Keratin filaments of corneal epithelial cells. *Vision Research, 21*(1), 55-63.

Tai, T., Damani, M., Rayner, V., Glasgow, B., & Hofbauer, J. (2009). Keratoconus associated with corneal stromal amyloid deposition containing TGFBIp. *Cornea, 28*, 589-593.

Tandon, A., Tovey, J., Sharma, A., Gupta, R., & Mohan, R. (2010). Role of TGFbeta in corneal function, biology and pathology. *Current Molecular Medicine, 10*(6), 565-578.

Tandon, A., Tovey, J., Waggoner, M., Sharma, A., Cowden, J., Gibson, D., et al. (2012). Vorinostat: a potent agent to prevent and treat laser-induced corneal haze. *J Refract Surg, 28*(4), 285-290.

Taxman, D., Livingstone, L., Zhang, J., Conti, B., Iocca, H., Williams, K., et al. (2006). Criteria for effective design, construction, and gene knockdown by shRNA vectors. *BMC Biotechnology, 6*(7), https://doi.org/10.1186/1472-6750-6-7.

Tayebi-khorami, M., Chegeni, N., Tahmasebi Birgani, M., Danyaei, A., Fardid, R., & Zafari, J. (2022). Construction a CO2 Incubator for Cell Culture with Capability of Transmitting Microwave Radiation. *Journal of Medical Signals and Sensors, 12*(2), 127-132.

Team, R. C. (2022). R: A language and environment for statistical computing, v4.1.3. R Foundation for Statistical Computing, Vienna, Austria.

Ten Dijke, P., Goumans, M., Itoh, F., & Itoh, S. (2022). Regulation of cell proliferation by Smad proteins. *J Cell Physiol, 191*(1), 1-16.

Teng, S., Madej, T., Panchenko, A., & Alexov, E. (2009). "Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophysical Journal, 96*(6), 2178–2188.

Thapa, N., Lee, B., & Kim, I. (2007). TGFBIp/βig-h3 protein: A versatile matrix molecule induced by TGF-β. *International journal of biochemistry and cell biology, 39*(12), 2183-2194.

The International SNP Map Working group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*, 928-933.

Thiel, H.-J., & Behnke, H. (1967). Eine bisher unbekannte subepitheliale hereditaere Hornhautdystrophie. *Klin. Monatsbl. Augenheilkd., 150*, 862-874.

Trief, D., & Fay, J. (2019). Lattice Corneal Dystrophy. *Medscape Ophthalmology*, 1193793.

Tsujikawa, M., Kurahashi, H., & Tanaka, T. (1999). Identification of the gene responsible for gelatinous drop-like corneal dystrophy. *Nat Genet, 21*, 420-423.

Tuori, A., Uusitalo, H., Burgeson, R. E., Terttunen, J., & Virtanen, I. (1996). The immunohistochemical composition of the human corneal basement membrane. *Cornea, 3*(15), 286-294.

Underhaug, J., Koldso, H., Runager, K., Nielsen, J., Sorensen, C., Kristensen, T., et al. (2013). Mutation in transforming growth factor beta induced protein associated with granular corneal dystrophy type 1 reduces the proteolytic susceptibility through local structural stabilization. *Biochem Biophys Acta, 1834*, 2812-2822.

Van Cauwenbergh, C., Van Schil, K., Bauwens, M., & De Baere, E. (2022). Genetic Testing Techniques. In *Clinical Ophtahlmic Genetics and Genomics* (pp. 7-12). Academic Press.

Vaz-Drago, R., Custódio, N., & Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. *Hum Genet, 136*, 1093-1111.

Venkatraman, A., Duong-Thi, M., Pervushin, K., Ohlson, S., & Mehta, J. (2020). Pharmaceutical modulation of the proteolytic profile of Transforming Growth Factor Beta induced protein (TGFBIp) offers a new avenue for treatment of TGFBI-corneal dystrophy. *Journal of Advanced Research, 24*, 529-543.

Verdi, G., & Filippone, A. (1958). Osservazione di un caso di degenerazione eredofamiliare della cornea, tipo Reis-Bücklers. *Boll Oculist, 37*, 410-430.

Vincent, A. L. (2014). Corneal dystrophies and genetics in the International Committee for Classification of Corneal Dystrophies era: a review. *Clin. Exp. Ophthalmol., 42*(1), 4-12.

Vincent, A. L., de Karolyi, B., Patel, D. V., Wheeldon, C. E., & McGhee, C. N. (2010). TGFBI mutational analysis in a New Zealand population of inherited corneal dystrophy patients. *British Journal of Ophthalmology , 94*, 836-842.

Vithana, E., Morgan, P., Ramprasad, V., Tan, D., Yong, V., Venkataraman, D., et al. (2008). SLC4A11 mutations in Fuchs endothelial corneal dystrophy. *Human Molecular Genetics, 17*, 656-666.

Vithana, E., Morgan, P., Sundaresan, P., Ebenezer, N., Tan, D., Mohamed, M., et al. (2006). Mutations in sodium-borate cotransporter SLC4A11 cause recessive congenital hereditary endothelial dystrophy (CHED2). *Nature Genet., 38*, 755-757.

Vittitow, J., & Borras, T. (2004). Genes expressed in the human trabecular meshwork during pressure-induced homeostatic response. *Journal of Cell Physiology*(201), 126-137.

Vojinovic, J., Damjanov, N., D'Urzo, C., Furlan, A., Susic, G., Pasic, S., et al. (2011). Safety and efficacy of an oral histone deacetylase inhibitor in systemic-onset juvenile idiopathic arthritis. *Arthritis Rheum, 63*, 1452–1458.

Vojnikovic, B., Gabric, N., Dekaris, I., & Juric, B. (2013). Curvature Analyses of the Corneal Front and Back Surface. *Coll. Antropol., 37*, 93-96.

Wang, C., Ren, Y., Zhai, J., Zhou, X., & Wu, J. (2019). Down-regulated LAMA4 inhibits oxidative stress-induced apoptosis of retinal ganglion cells through the MAPK signaling pathway in rats with glaucoma. *Cell Cycle, 18*(9), 932-948.

Wang, L., Tsang, H., & Coroneo, M. (2008). Treatment of recurrent corneal erosion syndrome using the combination of oral doxycycline and topical corticosteroid. *Clin Exp Ophthalmol, 36*(1), 8-12.

Wang, L., Zhao, C., Zheng, T., Zhang, Y., Liu, H., Wang, X., et al. (2021). Torin 1 alleviates impairment of TFEB-mediated lysosomal biogenesis and autophagy in TGFBI (p.G623_H626del)-linked Thiel-Behnke corneal dystrophy. *Autophagy*, 1-18.

Wang, M., Munier, F., Araki-Saski, K., & Schorderet, D. (2002). TGFBI gene transcript is transforming growth factor-beta1-responsive and cell density-dependent in a human corneal epithelial cell line. *Ophthalmic genetics, 23*(4), 237-245.

Wang, X., Liu, Q., & Zhang, B. (2014). Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics*, 23-24.

Wang, X., Zhang, Y., Zhao, Y., Liang, Y., Xiang, C., Zhou, H., et al. (2106). CD24 promoted cancer cell angiogenesis via Hsp90-mediated STAT3/VEGF signaling pathway in colorectal cancer. *Oncotarget, 7*, 55663–55676.

Watanabe, H., Fabricant, M., Tisdale, A. S., Spurr-Michaud, S. J., Lindberg, K., & Gipson, I. K. (1995). Human Corneal and Conjunctival Epithelia Produce a Mucin-like Glycoprotein for the Apical Surface. *Investigative Ophthalmology and Visual Science, 36*, 337-344.

Weber, A., Wasiliew, P., & Kracht, M. (2010). Interleukin-1 (IL-1) Pathway. *Science Signalling, 3*(105).

Weber, F., & Babel, J. (1980). Gelatinous drop-like dystrophy. A form of primary corneal amyloidosis. *Arch Ophthalmol, 98*, 144-148.

Weir, B., & Cockerham, C. (1984). Estimating F-Statistics for the analysis of population-structure. *Evolution, 38*, 1358-1370.

Weir, B., & Hill, W. (2022). Estimating F-statistics. *Annu Rev Genet, 36*, 721-750.

Weiss, J. S., Moller, H. U., Aldave, A. J., Seitz, B., Bredrup, C., Kivela, T., et al. (2015). IC3D Classification of Corneal Dystrophies—Edition 2. *Cornea, 34*(2), 117-159.

Weiss, J. S., Moller, H. U., Lisch, W., Kinoshita, S., Aldave, A. J., Belin, M. W., et al. (2008). The IC3D Classification of the Corneal Dystrophies. *Cornea, 27*(2), S1-S42.

Weiss, J., Moller, H. U., Aldave, A., Seitz, B., Bredrup, C., Kivela, T., et al. (2015). IC3D Classification of Corneal Dystrophies. *Cornea, 34*(2), 117-159.

Weiss, J., Moller, H. U., Aldave, A., Seitz, B., Bredrup, C., Kivela, T., et al. (2015). IC3D Classification of Corneal Dystrophies—Edition 2. *Cornea, 34*(2), 117-159.

Wen, G., Hong, M., Li, B., Liao, W., Cheng, S., Hu, B., et al. (2011). Transforming growth factor-β-induced protein (TGFBI) suppresses mesothelioma progression through the Akt/mTOR pathway. *Int J Oncol., 39*(4), 1001-1009.

Whitelock, J. M., Melrose, J., & Iozzo, R. V. (2008). Diverse cell signaling events modulated by perlecan. *Biochemistry*(47), 11174–11183.

WHO. (2016). *World Health Organisation*. Retrieved April 2016, from http://www.who.int/mediacentre/factsheets/fs282/en/

Wieben, E., Aleff, R., Eckloff, B., Atkinson, E., Baheti, S., Middha, S., et al. (2014). Comprehensive Assessment of Genetic Variants Within TCF4 in Fuchs' Endothelial Corneal Dystrophy. *IOVS, 55*(9).

Wikipedia. (2016). *Cornea*. Retrieved April 2016, from https://en.wikipedia.org/wiki/Cornea

Wilson, S. (2021). Interleukin-1 and Transforming Growth Factor Beta: Commonly Opposing, but Sometimes Supporting, Master Regulators of the Corneal Wound Healing Response to Injury. *IOVS, 62*(4), 8.

Wilson , S. E., Liu, J. J., & Mohan, R. R. (1999). Stromal-epithelial interactions in the cornea. *Prog Retin Eye Res.*(18), 293-309.

Wolf, M., Clay, S., Oldenburg, C., Rose-Nussbaumer, J., Hwang, D., & Chan, M. (2019). Overexpression of MMPs in Corneas Requiring Penetrating and Deep Anterior Lamellar Keratoplasty. *Invest Ophthalmol Vis Sci, 60*(5), 1734–1747.

Wolff, E. (1954). *The Anatomy of the Eye and Orbit* (4th ed.). London: H. K. Lewis and Co.

Wollensak, G., & Witschel, H. (1996). Vimentin and cytokeratin pattern in granular corneal dystrophy. *Graefes Arch Clin Exp Ophthalmol, 234*, 110-4.

Wu, P., Quan, H., Kang, J., He, J., Luo, S., Xie, C., et al. (2017). Downregulation of Calcium-Binding Protein S100A9 Inhibits Hypopharyngeal Cancer Cell Proliferation and Invasion Ability Through Inactivation of NF-κB Signaling. *Oncol Research, 25*(9), 1479-1488.

Yam, G., Fuest, M., Zhou, L., Liu, Y., Deng, L., Chan, A., et al. (2019). Differential epithelial and stromal protein profiles in cone and non-cone regions of keratoconus corneas. *Sci Rep, 9*, 2965.

Yamamoto, S., Okada, M., Tsujikawa, M., Morimura, H., Maeda, N., Watanabe, H., et al. (2000). The spectrum of big-h3 gene mutations in Japanese patients with corneal dystrophy. *Cornea, 19*, S21-S23.

Yamamoto, S., Okada, M., Tsujikawa, M., Shimomura, Y., Nishida, K., Inoue, Y., et al. (1998). A Kerato-Epithelin (βig-h3) mutation in lattice corneal dystrophy type IIIA. *Am J Hum Genet, 62*(3), 719-722.

Yan, L., Ma, J., Wang, Y., Zan, J., Wang, Z., Zhu, Y., et al. (2018). miR-21-5p induces cell proliferation by targeting TGFBI in non-small cell lung cancer cells. *Exp Ther Med, 16*(6), 4655-4663.

Yang, A., Chow, J., & Liu, J. (2018). Corneal innervation and sensation: the eye and beyond. *Yale Journal of Biology and Medicine, 91*(1), 13-21.

Yang, H. & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and Wannovar. *Nature Protocols*, 10, 1556-1566.

Yang, J., Han, X., Huang, D., Yu, L., Zhu, Y., Tong, Y., et al. (2010). Analysis of TGFBI gene mutations in Chinese patients with corneal dystrophies and review of the literature. *Molecular Vision , 16*, 1186-1193.

Yanoff, M., & Ducker, J. (2009). *Ophthalmology.* (J. Wiggs, D. Miller, D. Azar, M. Goldstein, E. Rosen, J. Duker, et al., Eds.) Mosby Elsevier.

Yee, R., Sullivan, L., Lai, H., Stock, E., Lu, Y., Khan, M., et al. (1997). Linkage Mapping of Thiel–Behnke Corneal Dystrophy (CDB2) to Chromosome 10q23–q24. *Genomics, 46*(1), 152-154.

Yellore, V. S., Rayner, S. A., & Aldave, A. J. (2011). TGFB1-Induced Extracellular Expression of TGFBIp and Inhibition of TGFBIp Expression by RNA Interference in a Human Corneal Epithelial Cell Line. *Investigative Ophthalmology and Visual Science, 52*, 757-763.

Yerbury, J., Poon, S., Meehan, S., Thompson, B., Kumita, J., Dobson, C., et al. (2007). The extracellular chaperone clusterin influences amyloid formation and toxicity by interacting with prefibrillar structures. *.FASEB journal, 21*(10), 2312-2322.

Yi, Q., & Zou, W. (2019). The Wound Healing Effect of Doxycycline after Corneal Alkali Burn in Rats. *Journal of Ophthalmology (Hindawi)*, 5168652.

Yokobori, T., & Nishiyama, M. (2017). TGF-beta Signalling an Gastrointestinal Cancers: Progress in Basic and Clinical Research. *J. Clin. Med., 6*(1), 11.

You, J., Corley, S., Wen, L., Hodge, C., Hollhumer, R., Madigan, M., et al. (2018). RNA-Seq analysis and comparison of corneal epithelium in keratoconus and myopia patients. *Scientific Reports, 8*, 389.

Yu, J., Ramasamy, T., Murphy, N., Holt, M., Czapiewski, R., Wei, S., et al. (2015). PI3K/mTORC2 regulates TGF-β/Activin signalling by modulating Smad2/3 activity via linker phosphorylation. *Nature Communications, 6*, 7212.

Yu, P., Yangshun, G., Fan, J., Rongrong, H., Lili, C., & Xiaoyi, Y. (2008). p.Ala546 > Asp and p.Arg555 > Trp mutations of TGFBI gene and their clinical manifestations in two large Chinese families with granular corneal dystrophy type I. . *Genetic Testing, 12*(3), 421+.

Yuan, C., Zins, E., Clark, A., & Huang, A. (2007). Suppression of keratoepithelin and myocilin by small interfering RNAs (siRNA) in vitro. *Mol Vis, 13*, 2083-2095.

Yurchenco, P. D. (2011). Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb Perspect Biol., 2*(3).

Yurchenco, P. D., Cheng, Y. S., & Colognato, H. (1992). Laminin forms an independent network in basement membranes. *Journal of Cell Biology*(117), 1119-1133.

Zalloua, P., Platt, D., Sibai, M., Khalife, J., Makhoul, N., Haber, M., et al. (2008). Identifying genetic traces of historical expansions: phoenician footprints in the Mediterranean. *Am J Hum Genet, 83*(5), 633-642.

Zayats, T., Young, T., Mackey, D., Malecaze, F., Calvas, P., & Guggenheim, J. (2009). Quality of DNA Extracted from Mouthwashes. *PLos One, 4*(7), e6165.

Zenteno, J. C., Correa-Gomez, V., Santacruz-Valdez, C., Suarez-Sanchez, R., & Villanueva-Mendoza, C. (2009). Clinical and genetic features of TGFBI-linked corneal dystrophies in Mexican population: Description of novel mutations and novel genotype–phenotype correlations. *Experimental Eye Research, 89*, 172–177.

Zenteno, J. C., Ramirez-Miranda, A., Santacruz-Valdes, C., & Suarez-Sanchez, R. (2006). Expanding the mutational spectrum in TGFBI-linked corneal dystrophies: Identification of a novel and unusual mutation (Val113Ile) in a family with granular dystrophy. *Molecular Vision 2006, 12*, 331-335.

Zenteno, J., Santacruz, V., & Ramirez, M. (2006). Autosomal Dominant GranularCorneal Dystrophy Caused by a TGFBI gene mutation in a Mexican Family. *Arch Soc Esp Oftalmol, 81*, 369-374.

Zhang, C., Yan, Y., Zhou, B., Wang, Y., Tian, X., Hao, S., et al. (2023). Identification of deep intronic variants of PAH in phenylketonuria using full-length gene sequencing. *Orphanet Journal of Rare Diseases, 18*, 128.

Zhang, H., Liu, F., Liu, Y., Peng, Y., Liao, Q., & Zhang, K. (2005). Effect of TGF-β1 Stimulation on the Smad Signal Transduction Pathway of Human Peritoneal Mesothelial Cells. *INt J Biomed Sci, 1*(1), 8-15.

Zhang, R., & Xia, T. (2017). Long non-coding RNA XIST regulates PDCD4 expression by interacting with miR-21-5p and inhibits osteosarcoma cell growth and metastasis. *Int J Oncol, 51*(5), 1460-1470.

Zhang, S., Meng, P., Liu, G., Liu, K., & Che, C. (2018). ATF4 Involvement in TLR4 and LOX-1-Induced Host Inflammatory Response to Aspergillus fumigatus Keratitis. *J Ophthalmol*, 5830202.

Zhang, X., Igo Jr, R., Fondran, J., Mootha, V., Oliva, M., Hammersmith, K., et al. (2013). Association of smoking and other risk factors with Fuchs' endothelial corneal dystrophy severity and corneal thickness. *54*(8), 5829-5835.

Zhang, X., Topley, N., Ito, T., & Phillips, A. (2005). Interleukin-6 regulation of transforming growth factor (TGF)-beta receptor compartmentalization and turnover enhances TGF-beta1 signaling. *J Biol Chem, 280*(13), 12239-12245.

Zhang, Y., Wen, G., Shao, G., Wang, C., Lin, C., Fang, H., et al. (2009). TGFBI deficiency predisposes mice to spontaneous tumor development. *Cancer Res, 69*(1), 37-44.

Zhang, Z., Teng, S., Wang, L., Schwartz, C., & Alexov, E. (2010). Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Human Mutation, 31*(9), 1043–1049.

Zhang, Z., Yang, W., Zhu, Z., Hu, Q., Chen, Y., He, H., et al. (2014). Therapeutic effects of topical doxycycline in a benzalkonium chloride-induced mouse dry eye model. *Invest Ophthalmol Vis Sci, 55*(5), 2963-2974.

Zhao, S. J., Zhu, Y., Shentu, X. C., & Miao, Q. (2013). Chinese family with atypical granular corneal dystrophy type I caused by the typical R555W mutation in TGFBI. *Intenational Journal of Ophthalmology, 6*(4), 458-462.

Zheng-Bradley, X., & Flicek, P. (2017). Applications of the 1000 Genomes Project resources. *Briefings in Functional Genomics, 16*(3), 163-170.

Zhu, D., Wang, J., Wang, Y., Jiang, Y., Li, S., Xiao, X., et al. (2023). Variant Landscape of 15 Genes Involved in Corneal Dystrophies: Report of 30 Families and Comprehensive Analysis of the Literature. *International Journal of Molecular Sciences, 24*(5), 5012.

Zhu, H., Chen, H., Wang, J., Zhou, L., & Liu, S. (2019). Collagen stiffness promoted non-muscle-invasive bladder cancer progression to muscle-invasive bladder cancer. *Onco Targets Ther, 12*, 3441-3457.

Zhu, Y., Chen, S., Liu, W., Zhang, L., Xu, F., Hayashi, T., et al. (2021). Collagens I and V differently regulate the proliferation and adhesion of rat islet INS-1 cells through the integrin β1/E-cadherin/β-catenin pathway. *Connect Tissue Res, 62*(6), 658-670.

Zupanets, I., Bezdetko, N., Dedukh, N., & Otrishko, I. (2002). Experimental study of the effect of glucosamine hydrochloride on metabolic and repair processes in connective tissue structures. *Eks Klin Farmakol, 65*(6), 67-69.

APPENDIX

Participant Information Statement

Granular corneal dystrophy in the Maltese Islands

Dear Sir/Madam

Granular corneal dystrophy is an inherited disorder of the eye. It results in permanent, gradual decrease in vision that cannot be corrected by any means of visual aids. Symptoms may begin in a person's early twenties and progresses with time. In order to improve the management of this condition, the gene responsible would need to be identified. Please take a moment to read the information and procedure below.

Procedure

The procedure is simple and non-invasive, and will take only a few seconds. You will be asked to provide buccal cells by means of a mouthwash. Once the buccal sample has been examined, you will be notified with the results, should you desire. Any personal information obtained from you, will be strictly confidential, and will be destroyed upon completion of this study. Your name will not be mentioned in any part of the final results but rather your assigned number in order to ensure anonymity.

I would like to inform you that your participation is entirely voluntary and that you are free to withdraw at any time without giving any reasons.

Finally you are free to ask any questions regarding the above. It would be my pleasure to provide you with satisfactory answers to your questions at all times.

Dr. Gabriella Guo Sciriha MD, MRCOphth, FEBO
Ophthalmologist
Specialist Registrar, Mater Dei Hospital
Email [redacted]
Tel: [redacted]

Dr. Joseph Borg B.Sc. (Hons.), M.Sc., Ph.D.
Project Supervisor
Senior Lecturer, Faculty of Health Sciences
Email: [redacted]
Tel: [redacted]

Informazzjoni lill-Parteċipanti

Granular Corneal Dystrophy fil-Familji Maltin

Għażiż Sinjur / Sinjura

Granular Corneal Dystrophy hija kundizzjoni li tintiret li taffettwa l-għajn. Din tikkawża tnaqqis progressiv fil-vista li ma tistax tigi ikkoreġuta b'nuċċali. Is-sintomi jistgħu jibdew minn eta' żgħira. Sabiex intejbu t-trattament ta' din il-kundizzjoni, il-ġene li qed tikkawża' il-Granular Corneal Dystrophy trid tiġi identifikata. Sabiex ngħinek tifhem aktar dwar dan il-proġett, nitolbok tieħu ftit ħin biex taqra l-informazzjoni u l-proċeduri hawn taħt infissra.

Il- Proċedura

Il-proċedura meħtieġa biex tkun tista' tiġi identifikata din il-ġene hija semplici u tieħu biss ftit sekondi. Ħa nistaqsuk biex tipprovdi celluli mill-ħalq permezz ta' 'mouthwash'. Jekk tixtieq, nistgħu ninfurmawk bir-rizultat ta' dan il-buccal swab wara li din tiġi eżaminata u maħduma. Kull informazjoni personali miksuba ser tiġi ttrattata bl-akbar riżervatezza u titqies bħala waħda kunfidenzjali. Din l-informazjoni tiġi wżata biss għall-iskop ta' dan il-proġett u tiġi meqruda hekk kif jintemm dan l-eżerċizzju. Ismek m'hu se jiġi mxandar f'ebda' parti u bl-ebda' mod tul dan il-proġett. Inti tkun magħruf biss permezz ta' numru li jingħatalek u dan biex aktar inħarsu l-anonimità tiegħek.

Nixtieq nifurmak illi l-parteċipazjoni tiegħek hija għal kollox fuq bażi volontarja u li tista' tirtira minn dan meta trid bla ma tintalablek ebda' spjegazzjoni.

Fl-aħħarnett, jekk tixtieq tagħmel xi mistoqsijiet, ħossok liberu li tikkuntattjani. F'dan il-każ, ikun dmir tiegħi inwieġeb għall-mistoqsijiet tiegħek f'kull ħin.

Dr. Gabriella Guo Sciriha MD, MRCOphth, FEBO Ophthalmologist
Specialist Registrar, Mater Dei Hospital
Email: [redacted]
Tel: [redacted]

Dr. Joseph Borg B.Sc. (Hons.), M.Sc., Ph.D.
Project Supervisor
Senior Lecturer, Faculty of Health Sciences
Email [redacted]
Tel: [redacted]

Participant Consent Form

Date: ____ / ____/ _____

I hereby voluntarily consent to provide buccal cells by means of a mouthwash which will be solely used to identify the causative gene/s in granular corneal dystrophy.

I understand that my participation is entirely voluntary. I acknowledge that I have the right to question any part of the procedure and can withdraw at any time without this being held against me.

I have received and read the "Participant Information Statement" sheet, which describes the procedure involved, and if I have any questions I may contact Dr. Gabriella Guo Sciriha on her email address: [redacted].

I understand that the information obtained from this study is strictly confidential and that my personal details will not be revealed to anybody at any time.

Therefore I give my consent to participate.

_____ _____

Signature of Participant Name in Block Letters

_____ _____

Dr. Gabriella Guo Sciriha MD, MRCOphth, FEBO Ophthalmologist
Specialist Registrar, Mater Dei Hospital
Email: [redacted]
Tel: [redacted]

Dr. Joseph Borg B.Sc. (Hons.), M.Sc., Ph.D.
Project Supervisor
Senior Lecturer, Faculty of Health Sciences
Email: [redacted]
Tel: [redacted]

Formola ta' Kunsens mill-Parteċipanti

Data:____/ ____/ _____

Jiena qed naċċetta li volontarjament ngħati ċelluli mill-ħalq permezz ta' 'mouthwash' li ħa tintuża' biss biex tiġi identifikata il-gene/s li tikkawża' il-Granular Corneal Dystrophy.

Jiena nifhem illi l-parteċipazzjoni tiegħi f'dan l-istudju hija fuq bażi għal kollox volontarja. Nifhem illi għandi dritt nitlob aktar tagħrif dwar kwalunkwe parti minn din il-proċedura u li nista' ntemm il-parteċipazzjoni tiegħi kif u meta rrid bla ma dan jittieħed kontra tiegħi.

Jiena rċevejt u qrajt id-Dikjarazzjoni ta' Informazzjoni għall-Parteċipanti illi tiddeskrivi il-proċedura nvoluta. Nifhem dak li huwa mitlub minn għandi u li jekk hemm xi mistoqsijiet li nixtieq nagħmel dwar dan nista' nikkuntattja lil Gabriella Guo Sciriha fuq l-indirizz e.mail: [redacted].

Nifhem illi l-informazzjoni miksuba f'dan l-istudju hija kunfidenzjali u li d-dettalji personali tiegħi ma jiġu mxandra lil ħadd f'ebda ħin.

Għalhekk, jien nagħti l-kunsens tiegħi biex nipparteċipa f'dan il-proġett.

_____ _____

Firma tal-Parteċipant Isem il-Parteċipant b'ittri kbar

_____ _____

Dr. Gabriella Guo Sciriha MD, MRCOphth, FEBO Ophthalmologist
Specialist Registrar, Mater Dei Hospital
Email: [redacted]
Tel: [redacted]

Dr. Joseph Borg B.Sc. (Hons.), M.Sc., Ph.D.
Project Supervisor
Senior Lecturer, Faculty of Health Sciences
Email: [redacted]
Tel: [redacted]

Calculation for determining the minimum sample size needed for the Maltese cohort in Phase1 of this PhD study.

The following is a sample run using the pwr package of the statistical computing R program for the mutation G>C in *CHST* for TSI vs MLT.

```
> pwr.2p2n.test(h=.314,n1=107,power=.8, sig.level = .05)

     difference of proportion power calculation for binomial distribut
ion (arcsine transformation)

              h = 0.314
             n1 = 107
             n2 = 310.9425
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

Allele frequencies of the populations in the 1000GP of the 19 SNPs analysed in this study.

HGVS ref. seq.	NC_000012.11:g.53186088G>C	NC_000012.11:g.53189696C>G	NC_000012.11:g.53186122G>A	NC_000005.9:g.135391462A>G	NC_000005.9:g.135382989G>C	NC_000006.9:g.75513243G>C	NC_000002.11:g.209190330T>C	NC_000002.11:g.209190519A>T	NC_000002.11:g.209190528C>G	NC_000002.11:g.209191082C>A	NC_000002.11:g.209184999C>T	NC_000002.11:g.209215586A>G	NC_000020.10:g.3218634G>C	NC_000020.10:g.3210301G>A	NC_000020.10:g.321 1235C>T	NC_000020.10:g.321 4581C>T	NC_000020.10:g.321 4819T>G	NC_000012.11:g.91449984C>T	NC_000012.11:g.91449990C>T
Gene	KRT3	KRT3	KRT3	TGFBI	TGFBI	CHST6	PIKFYVE	PIKFYVE	PIKFYVE	PIKFYVE	PIKFYVE	PIKFYVE	SLC4A11	SLC4A11	SLC4A11	SLC4A11	SLC4A11	KERA	KERA
ACB	0.318	0.276	0.766	0	0.062	0.005	0.76	0.76	0.76	0.76	0.76	0.76	0.828	0.042	0.042	0.214	0.714	0.609	0.266
ASW	0.369	0.262	0.746	0	0.107	0.008	0.861	0.861	0.861	0.861	0.861	0.861	0.779	0.025	0.025	0.123	0.615	0.516	0.295
ESN	0.348	0.379	0.854	0	0.005	0	0.712	0.712	0.712	0.712	0.712	0.702	0.843	0.01	0.01	0.217	0.657	0.535	0.253
GWD	0.456	0.221	0.832	0	0.022	0.009	0.81	0.81	0.819	0.81	0.81	0.81	0.863	0.022	0.022	0.261	0.686	0.593	0.274
LWK	0.338	0.333	0.778	0	0.025	0	0.823	0.823	0.823	0.823	0.823	0.823	0.778	0.005	0.005	0.207	0.611	0.631	0.242
MSL	0.453	0.247	0.812	0	0	0	0.8	0.8	0.806	0.8	0.8	0.8	0.847	0.018	0.018	0.194	0.724	0.571	0.182
YRI	0.356	0.329	0.801	0	0	0	0.782	0.782	0.782	0.782	0.782	0.782	0.796	0	0.005	0.264	0.676	0.606	0.208
CLM	0.527	0.16	0.824	0	0.441	0.059	0.973	0.973	0.973	0.973	0.973	0.973	0.585	0.128	0.133	0.112	0.463	0.101	0.739
MXL	0.406	0.102	0.805	0	0.453	0.016	0.984	0.984	0.984	0.984	0.984	0.984	0.641	0.062	0.055	0.125	0.469	0.148	0.648
PEL	0.294	0.059	0.729	0	0.5	0.012	0.988	0.988	0.988	0.988	0.988	0.988	0.582	0.024	0.018	0.076	0.418	0.147	0.659
PUR	0.548	0.135	0.793	0.005	0.409	0.048	0.966	0.966	0.966	0.966	0.966	0.966	0.538	0.077	0.082	0.178	0.37	0.168	0.745
CDX	0.113	0	0.946	0	0.349	0	0.925	0.925	0.925	0.925	0.925	0.925	0.731	0	0	0.118	0.57	0.081	0.823
CHB	0.117	0	0.903	0	0.354	0	0.937	0.932	0.932	0.932	0.932	0.932	0.733	0	0	0.121	0.597	0.121	0.757
CHS	0.086	0	0.919	0	0.362	0	0.929	0.929	0.929	0.929	0.929	0.929	0.71	0	0.005	0.129	0.576	0.052	0.79
JPT	0.207	0	0.885	0	0.404	0	0.971	0.971	0.971	0.971	0.971	0.971	0.587	0.014	0	0.159	0.409	0.082	0.803
KHV	0.162	0.01	0.924	0	0.374	0	0.939	0.939	0.939	0.939	0.939	0.939	0.626	0	0	0.111	0.53	0.066	0.818
CEU	0.697	0.182	0.823	0.005	0.47	0.051	1	1	1	1	1	1	0.551	0.101	0.091	0.192	0.359	0.111	0.773
FIN	0.631	0.101	0.808	0	0.515	0.04	1	1	1	1	1	1	0.591	0.111	0.111	0.086	0.5	0.086	0.818
GBR	0.681	0.198	0.841	0	0.451	0.06	1	1	1	1	1	1	0.577	0.143	0.126	0.231	0.346	0.11	0.824
IBS	0.678	0.178	0.832	0.009	0.463	0.065	0.986	0.986	0.986	0.986	0.986	0.986	0.617	0.229	0.21	0.187	0.421	0.136	0.785
TSI	0.673	0.159	0.79	0	0.467	0.051	0.991	0.991	0.991	0.991	0.991	0.991	0.561	0.145	0.131	0.196	0.35	0.145	0.785
BEB	0.256	0.058	0.622	0	0.36	0	0.913	0.913	0.913	0.913	0.913	0.913	0.547	0.017	0.017	0.163	0.355	0.052	0.721
GIH	0.267	0.049	0.573	0	0.374	0.005	0.961	0.961	0.961	0.961	0.961	0.961	0.417	0.024	0.019	0.17	0.228	0.058	0.704
ITU	0.25	0.074	0.549	0	0.412	0.005	0.951	0.951	0.951	0.951	0.951	0.951	0.422	0.01	0.01	0.147	0.25	0.059	0.74
PJL	0.349	0.062	0.646	0	0.406	0.016	0.896	0.896	0.896	0.896	0.896	0.896	0.552	0.026	0.01	0.151	0.354	0.078	0.667
STU	0.23	0.049	0.647	0	0.466	0	0.858	0.858	0.858	0.858	0.858	0.858	0.426	0.015	0.015	0.132	0.275	0.054	0.686

ACB: African Caribbean in Barbados; ASW: African ancestry in Southwest US; ESN: Esan in Nigeria; GWD: Gambian in Western Division, The Gambia; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria; CLM: Colombian in Medellin, Colombia; MXL: Mexican ancestry in Los Angeles, California; PEL: Peruvian in Lima, Peru; PUR: Puerto Rican in Puerto Rico; CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Bejing, China; CHS: Southern Han Chinese, China; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; CEU: Utah residents with Northern and Western European ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian populations in Spain; TSI: Toscani in Italy; BEB: Bengali in Bangladesh; GIH: Gujarati Indian in Houston, TX; ITU: Indian Telugu in the UK; PJL: Punjabi in Lahore, Pakistan; STU: Sri Lankan Tamil in the UK.

F_{ST} of Maltese cohort vs individual population cohorts.

	KRT			TGFBI			CHST			PIKFYVE			SLC4A11			KERA
MLT PUR	0		MLT BEB	0		MLT FIN	0		MLT CLM	0.002891		MLT CLM	0.002779		MLT KHV	0.000333
MLT MXL	0.000376		MLT CDX	0		MLT PUR	0		MLT MXL	0.00411		MLT MXL	0.004188		MLT CHS	0.002121
MLT CLM	0.003417		MLT CHB	0		MLT CEU	0		MLT CDX	0.005505		MLT PUR	0.009979		MLT CDX	0.00384
MLT FIN	0.007688		MLT CHS	0		MLT TSI	0		MLT PUR	0.005539		MLT CDX	0.010177		MLT FIN	0.006503
MLT GWD	0.018671		MLT KHV	0		MLT CLM	0		MLT CHB	0.009235		MLT PEL	0.010717		MLT JPT	0.007603
MLT TSI	0.018866		MLT GIH	0		MLT GBR	0		MLT CHS	0.009287		MLT CHB	0.015681		MLT GBR	0.014345
MLT MSL	0.020747		MLT PUR	0		MLT IBS	0.000734		MLT JPT	0.010362		MLT CHS	0.015744		MLT ITU	0.016026
MLT PJL	0.022623		MLT PJL	0		MLT MXL	0.002668		MLT PEL	0.012441		MLT ASW	0.015848		MLT BEB	0.019617
MLT PEL	0.027939		MLT JPT	0		MLT PJL	0.004265		MLT BEB	0.015687		MLT JPT	0.016287		MLT CEU	0.024232
MLT IBS	0.028252		MLT ITU	0		MLT PEL	0.008992		MLT KHV	0.016005		MLT ACB	0.024736		MLT GIH	0.027836
MLT ASW	0.029089		MLT CLM	0.003114		MLT GWD	0.014419		MLT GIH	0.016056		MLT BEB	0.025618		MLT CLM	0.029181
MLT CEU	0.031833		MLT MXL	0.003955		MLT ASW	0.01485		MLT PJL	0.017022		MLT KHV	0.025777		MLT IBS	0.0327
MLT GBR	0.033645		MLT GBR	0.004791		MLT ACB	0.022128		MLT ITU	0.017417		MLT GIH	0.025802		MLT CHB	0.03273
MLT ACB	0.047423		MLT CEU	0.00634		MLT ITU	0.022201		MLT GBR	0.020169		MLT PJL	0.026285		MLT STU	0.033672
MLT BEB	0.052249		MLT IBS	0.006917		MLT GIH	0.022212		MLT CEU	0.020944		MLT ITU	0.026482		MLT TSI	0.036735
MLT YRI	0.057498		MLT STU	0.008099		MLT ESN	0.035548		MLT IBS	0.021232		MLT GBR	0.027858		MLT PJL	0.047031
MLT STU	0.060373		MLT TSI	0.008432		MLT LWK	0.035548		MLT STU	0.021823		MLT CEU	0.028246		MLT PUR	0.057386
MLT LWK	0.06038		MLT PEL	0.015891		MLT MSL	0.035548		MLT TSI	0.02199		MLT IBS	0.02839		MLT PEL	0.077167
MLT GIH	0.062208		MLT FIN	0.020611		MLT YRI	0.035548		MLT FIN	0.028079		MLT STU	0.028685		MLT MXL	0.079667
MLT ITU	0.068169		MLT ASW	0.077906		MLT CDX	0.035548		MLT ASW	0.036534		MLT TSI	0.028769		MLT ASW	0.45981
MLT ESN	0.083591		MLT ACB	0.114656		MLT CHB	0.035548		MLT ACB	0.054709		MLT FIN	0.031814		MLT ESN	0.49821
MLT JPT	0.108949		MLT LWK	0.150337		MLT CHS	0.035548		MLT LWK	0.073946		MLT LWK	0.035751		MLT GWD	0.516808
MLT KHV	0.134919		MLT GWD	0.153522		MLT JPT	0.035548		MLT GWD	0.075458		MLT GWD	0.036426		MLT ACB	0.529708
MLT CHB	0.156495		MLT ESN	0.172076		MLT KHV	0.035548		MLT ESN	0.083838		MLT ESN	0.039719		MLT LWK	0.556376
MLT CDX	0.178745		MLT MSL	0.177795		MLT BEB	0.035548		MLT MSL	0.087679		MLT MSL	0.042622		MLT MSL	0.561076
MLT CHS	0.17986		MLT YRI	0.177795		MLT STU	0.035548		MLT YRI	0.089264		MLT YRI	0.044999		MLT YRI	0.564349

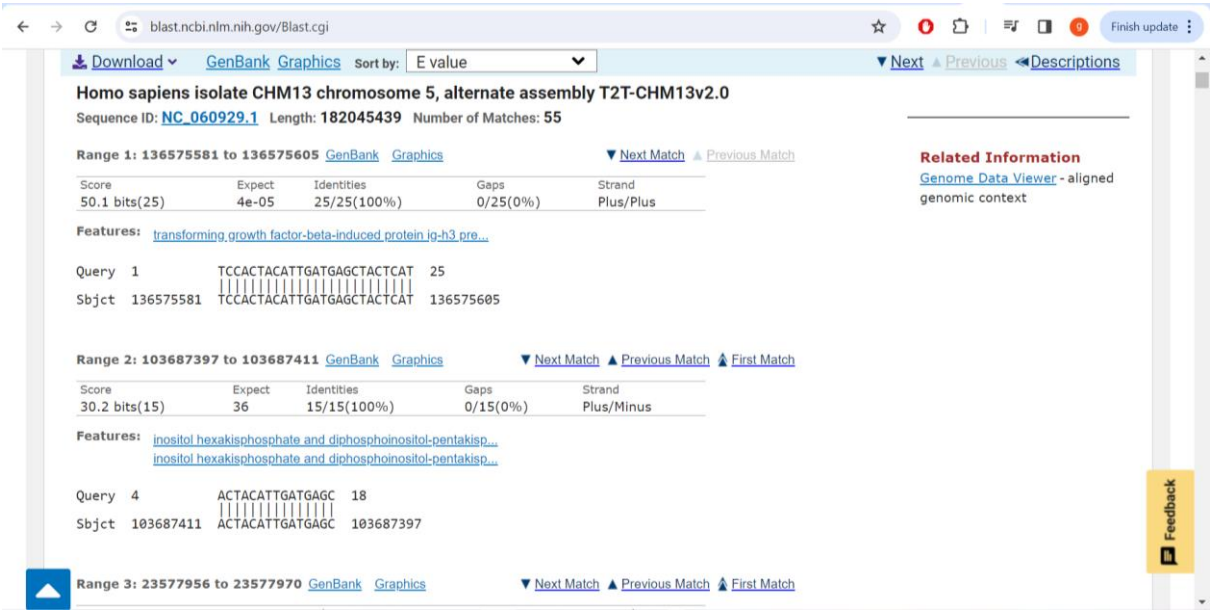
ACB: African Caribbean in Barbados; ASW: African ancestry in Southwest US; ESN: Esan in Nigeria; GWD: Gambian in Western Division, The Gambia; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria; CLM: Colombian in Medellin, Colombia; MXL: Mexican ancestry in Los Angeles, California; PEL: Peruvian in Lima, Peru; PUR: Puerto Rican in Puerto Rico; CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Bejing, China; CHS: Southern Han Chinese, China; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; CEU: Utah residents with Northern and Western European ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian populations in Spain; TSI: Toscani in Italy; BEB: Bengali in Bangladesh; GIH: Gujarati Indian in Houston, TX; ITU: Indian Telugu in the UK; PJL: Punjabi in Lahore, Pakistan; STU: Sri Lankan Tamil in the UK; MLT: Maltese.

Ensembl exon sequence and NCBI BLAST result for siRNA2 sequence.

Ensembl exon sequence of siRNA2.

```
1073 TAGCCACCAACGGGGTGATCCACTACATTGATGAGCTACTCATCCCAGACTCAGCCAAGA   1132
 358 L--A--T--N--G--V--I--H--Y--I--D--E--L--L--I--P--D--S--A--K--    377
```

NCBI BLAST result for siRNA2 sequence.



A scoping review of compounds that reduce corneal TGFBI protein levels.

Search strategy for PubMed

Search term	**Number of hits**	**Number of relevant articles**	**Studies**
TGFBI AND Cornea* AND treatment AND therapy	59	4	1. (Nie D, et al., 2018) 2. (Choi S, et al., 2011) 3. (Kim T, et al., 2015) 4. (Choi S, et al., 2016)
TGFBI AND Cornea* AND treatment	79	6	**1. (Nie D, et al., 2018)** **2. (Choi S, et al., 2011)** **3. (Kim T, et al., 2015)** **4. (Choi S, et al., 2016)** 5. (Nie D, et al., 2020) 6. (Wang B, et al., 2021)
TGFBI AND Cornea* AND therapy	60	4	**1. (Nie D, et al., 2018)** **2. (Choi S, et al., 2011)** **3. (Kim T, et al., 2015)** **4. (Choi S, et al., 2016)**
TGFBI AND Cornea* AND GSK-3	3	2	**1. (Nie D, et al., 2018)** **2. (Choi S, et al., 2011)**
TGFBI AND Cornea* AND therapy AND SMAD	1	0	Nil
TGFBI AND Cornea* AND treatment AND SMAD	2	0	Nil
TGFBI AND Cornea* AND SMAD	3	0	Nil
TGFBI AND Cornea* AND compounds	7	0	Nil
TGFBI AND Cornea* AND drugs	2	0	Nil
GCD1 AND Cornea* AND treatment	0	0	Nil
GCD1 AND Cornea* AND therapy	0	0	Nil
R555W AND Cornea* AND therapy	8	0	Nil
R555W AND Cornea* AND treatment	9	0	Nil
TGFBI AND Cornea* AND medication	109	2	**1. (Wang B, et al., 2021)** **2. (Kim T, et al., 2015)**
decreased expression of TGFBI AND cornea*	11	4	**1. (Nie D, et al., 2018)** **2. (Choi S, et al., 2011)** **3. (Kim T, et al., 2015)** 4. (Kim T, et al., 2008)
inhibition of SMAD AND cornea	18	5	1. (Chen Y, et al., 2020) 2. (Sharma A, et al., 2015) 3. (Lim R, et al., 2016) 4. (Kim H, et al., 2005)

Note: Articles highlighted in bold are duplicates that were found across the search strings. These have been included in the duplicate count.

Search strategy for Hydi

Search term	Number of hits	Number of relevant articles	Studies
TGFBI AND cornea AND treatment AND therapy	170	8	1. (Choi S, et al., 2011) 2. (Nie D, et al., 2018) 3. (Wang B, et al., 2021) 4. (Choi S, et al., 2016) 5. (Kim T, et al., 2008) 6. (Choi S, et al., 2017) 7. (Choi S, et al. 2013) 8. (Chen K, et al., 2019)
TGFBI AND cornea AND treatment	379	7	1. **(Choi S, et al., 2011)** 2. (Venkatraman A, et al., 2020) 3. **(Wang B, et al., 2021)** 4. **(Choi S, et al., 2016)** 5. **(Kim T, et al., 2008)** 6. **(Choi S, et al., 2017)** 7. **(Choi S, et al. 2013)**
TGFBI AND cornea AND therapy	205	6	1. **(Choi S, et al., 2011)** 2. **(Wang B, et al., 2021)** 3. **(Choi S, et al., 2016)** 4. **(Kim T, et al., 2008)** 5. **(Choi S, et al., 2017)** 6. **(Choi S, et al. 2013)**
TGFBI AND Cornea AND GSK-3	3	1	1. **(Choi S, et al., 2011)**
TGFBI AND cornea AND therapy AND SMAD	12	0	nil
TGFBI AND Cornea AND treatment AND SMAD	24	0	nil
TGFBI AND Cornea AND SMAD	32	0	nil
TGFBI AND Cornea AND compounds	64	3	1. **(Venkatraman A, et al., 2020)** 2. **(Wang B, et al., 2021)** 3. **(Choi S, et al. 2013)**
TGFBI AND Cornea AND drugs	123	6	1**. (Choi S, et al., 2011)** 2. **(Wang B, et al., 2021)** 3. **(Choi S, et al., 2016)** 4. **(Kim T, et al., 2008)** 5. **(Choi S, et al., 2017)** 6. **(Choi S, et al. 2013)**
GCD1 AND Cornea AND treatment	35	0	nil
GCD1 AND Cornea AND therapy	18	0	nil
R555W AND Cornea AND therapy	26	1	1. **(Kim T, et al., 2008)**
R555W AND Cornea AND treatment	58	2	1. **(Venkatraman A, et al., 2020)** 2. **(Kim T, et al., 2008)**
TGFBI AND Cornea AND	4	0	nil

medication			
decreased expression of tgfbi AND cornea	181	6	1. (Nie D, et al., 2020) 2**. (Choi S, et al., 2011)** 3. **(Wang B, et al., 2021)** 4. **(Kim T, et al., 2008)** 5. **(Choi S, et al., 2017)** 6. **(Choi S, et al. 2013)**
inhibition of SMAD AND cornea	1010	9	1. **(Nie D, et al., 2020)** 2. (Sarenac T, et.al., 2016) 3. (Nuwormegbe S, et al., 2021) 4. (Sharma A, et al., 2015) 5. (Lim R, et al., 2016) 6. (Nelson E, et al., 2012) 7. (Park JH, et al., 2021) 8. (Chen Y, et al., 2020) 9. (Okumura M, et al., 2013)

Note: Articles highlighted in bold are duplicates that were found across the search strings. These have been included in the duplicate count.

Institutional Permission from Mater Dei Hospital CEO

Proceed as per applicable protocol.

Ivan Falzon
Chief Executive Officer | TeaMDH

T

M

E

Mater Dei Hospital, Triq tal-Qroqq, Msida, Malta MSD 2090 | Tel +356 2545 0000 | www.materdeihospital.org.

On Thursday, October 1, 2015 7:04 AM, Data Protection at MDH <datapro.mdh@gov.mt> wrote:

Dear Dr Sciriha

Good Morning

On the basis of the documentation you submitted, from the MDH data protection point of view you have been cleared to proceed with your study provided that you obtain approval from MDH CEO and the University Ethics Committee.

Please contact Ms. Nadine Buhagiar on or Ms. Graziella Aquilina on to present a copy of your approvals and fill in the appropriate Data Protection Form.

Remember that in no way should you retain any personal details you obtain from your research and this should be destroyed at the end of your study.

All medical records are to be viewed at the Medical Records Department MDH.

You are requested to submit a copy of your findings to this office at the end of your study.

Regards

Sharon Young

Data Protection Officer

Mater Dei Hospital

L-Università ta' Malta

Faculty of Medicine & Surgery

University of Malta
Msida MSD 2080, Malta

Tel: +356 2340 1879/1891/1167
umms@um.edu.mt

www.um.edu.mt/ms

Ref No: **FRECMDS_1718_064**

Friday 28th September 2018

Dr Gabriella Guo Sciriha

Dear Dr Guo Sciriha,

Please refer to your application submitted to the Research Ethics Committee in connection with your research entitled:

The mutation variant, Novel to Established, Present in a Cohort of Maltese Patients that Phenotypically Exhibit Granular Corneal Dystrophy. (Inhibition of Mutated Keratoepithelin Deposition as a Treatment Modality)

The University Research Ethics Committee granted ethical approval for the above mentioned protocol.

Yours sincerely,

Professor Pierre Mallia
Chairman
Research Ethics Committee

L-UNIVERSITÀ TA' MALTA

Msida – Malta
Skola Medika
Sptar Mater Dei

UNIVERSITY OF MALTA

Msida – Malta
Medical School
Mater Dei Hospital

Ref No: **11/2017**

Wednesday 24th May 2017

Dr Gabriella Guo Sciriha

Dear Dr Gabriella Gio Sciriha,

Please refer to your application submitted to the Research Ethics Committee in connection with your research entitled:

The Mutation Variant, Novel or Established, Present in a Cohort of Maltese Patients that Phenotypically Exhibit Granular Corneal Dystrophy.

The University Research Ethics Committee granted ethical approval for the above mentioned protocol.

Yours sincerely,

Dr. Mario Vassallo
Chairman
Research Ethics Committee

L-Università ta' Malta

Doctoral School
University of Malta
Msida MSD 2080, Malta

Tel: +356 2340 3608/3254
doctoralschool@um.edu.mt

www.um.edu.mt

13 May 2022

Dr Gabriella Maria Guo Sciriha

Student Code:

Dear Dr Guo Sciriha

Change in Title

Reference is made to correspondence dated 11 January 2022.

I am pleased to inform you that Senate has approved the change in title of your Doctor of Philosophy Thesis to "Genetic Variants in Corneal Dystrophy Genes: A Maltese Cohort Study. Inhibition of TGFBI as a Treatment Modality".

Yours sincerely

Nikolai Papagiorcopulo
Assistant Registrar

c.c Dean, Faculty of Medicine and Surgery
Professor Joseph Borg, Principal Supervisor
Mr Mario Vella, Co-Supervisor
Mr Franco Mercieca, Co-Supervisor
Director of Finance
Officer i/c Faculty of Medicine and Surgery
SIMS Office