

https://dspace.org/

Differential Gene Expression in Exfoliation Syndrome and Exfoliation Glaucoma in the Conjunctiva of Black South Africans

Hulley, Michaella Robyn

2024

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/43296

Hulley, Michaella Robyn . (2024). Differential Gene Expression in Exfoliation Syndrome and Exfoliation Glaucoma in the Conjunctiva of Black South Africans [PhD thesis, University of the Witwatersrand, Johannesburg]. WireDSpace.

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

Differential Gene Expression in Exfoliation Syndrome and Exfoliation Glaucoma in the Conjunctiva of Black South Africans



Michaella Hulley

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

Johannesburg, 2024.

DECLARATION

I, Michaella Robyn Hulley (Student number: 1820405), am a student registered for the degree of Doctor of Philosophy in the academic year 2024.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" software indicating the level of plagiarism in my research document (Appendix L).

Signed: 

on 08/04/2024 at Norwood, Johannesburg.

DEDICATION

To my partner, who even through her own battle with cancer, stood by me and supported me unconditionally.

PRESENTATIONS ARISING FROM THIS STUDY

Hulley, M; Ngcungcu, T; Ramsay, M and Williams, S. (2023). Differential gene expression in exfoliation glaucoma in the conjunctiva of a South African cohort. Poster presentation at 14th International Congress of Human Genetics (ICHG2023), Cape Town.

Hulley, M; Ngcungcu, T; Ramsay, M and Williams, S. (2020). Using conjunctival cells to interrogate differential gene expression in exfoliation glaucoma in Black South Africans. Poster presentation at the Faculty of Health Sciences Research Day and Postgraduate Expo, University of the Witwatersrand, Medical School.

Hulley, M; Ngcungcu, T; Ramsay, M and Williams, S. (2019). Differential gene expression in exfoliation glaucoma in the conjunctiva of black South Africans. Poster presentation at the Biennial Congress of the Southern African Society for Human Genetics (SASHG), Century City Conference Centre.

Hulley, M; Ngcungcu, T; Ramsay, M and Williams, S. (2019). Genetics in exfoliation syndrome. (2019). Oral presentation at the Division of Ophthalmology (Wits) Research & CME Day, Glenhove Conference Centre.

Hulley, M; Ngcungcu, T; Ramsay, M and Williams, S. (2018). Differential gene expression in exfoliation syndrome and exfoliation glaucoma in the conjunctiva of black South Africans. Poster presentation at the Faculty of Health Sciences Research Day and Postgraduate Expo, University of the Witwatersrand, Medical School.

PUBLICATIONS ARISING FROM THIS STUDY

Hulley, MR; Ngcungcu, N; Ramsay, M and Williams, S. 2023. Non-invasive harvesting of conjunctival cells for whole transcriptome sequencing. *Experimental Eye Research*. 234. 109613. Available at: https://doi.org/10.1016/j.exer.2023.109613

Hulley, MR; Ngcungcu, N; Ramsay, M and Williams, S. In preparation. Differential gene expression and pathway analysis in exfoliation syndrome in a South African cohort.

ABSTRACT

Glaucoma is a heterogeneous group of clinically and genetically complex optic neuropathies. The most prevalent identifiable secondary cause of glaucoma is the ocular manifestation of exfoliation syndrome (XFS), known as exfoliation glaucoma (XFG), a severe form of glaucoma characterized by rapid progression. While XFS is systemic, XFG is defined by fibrillar extracellular matrix (ECM) deposits in the eye. XFG has a recognised genetic aetiology, specifically with contributions from *LOXL1* variants, however the molecular pathogenesis remains unclear. Conjunctival cells have not previously been used for whole transcriptome sequencing, and gene expression studies on XFG are lacking, particularly in South Africa. The aim of this study was to validate the use of conjunctival cells for whole transcriptome sequencing and interrogate the gene expression profiles of individuals with XFS and XFG. Conjunctival cells were collected from 19 cases and 15 control participants for RNA extraction, cDNA library construction and sequencing. Differential gene expression was assessed between cells from cases and controls and 81 genes were found to be differentially expressed, with 80 upregulated in cases. Interestingly, *LOXL1*, which has shown contradictory expression results in previous studies, was not identified as differentially expressed. The most significant overexpression was of the *CCN2* gene, which encodes a matricellular protein that interacts with structural ECM molecules. A potential novel pathway involving megakaryocytes and the activation of neutrophils was, however, identified. Megakaryocytes and neutrophils are both involved with interstitial fibrosis, with neutrophils also responding to chemotactic signals to induce inflammation. The inflammatory pathway has long been associated with XFS, XFG and other fibrotic disorders, such as Alzheimer's disease and rheumatoid arthritis. Genes involved in the inflammatory process were overexpressed in XFG, including *TGF-β*, *IL-1β*, *IL-33*, *EGR1* and *EGR3*. *EGR1* and *EGR3* are regulated by fibrotic signals and therefore play an important role in fibrosis, which may contribute to XFS fibrillar deposits. They are also regulated by *TGF-β*, thus supporting a complex interplay of the inflammatory process and fibrosis in XFS pathogenesis.

ACKNOWLEDGEMENTS

Aspects of the Research performed by the candidate (M Hulley) and by collaborators:

1. Recruitment process
    a. Questionnaire – Candidate, Prof Williams or her research assistant, Prescilla January
    b. Ophthalmic examination – Prof Williams or Ophthalmologist available
    c. Sample collection
        i. Conjunctival cells – Prof Williams or Ophthalmologist available
        ii. Saliva – Candidate, Prof Williams or her research assistant, Prescilla January
2. Laboratory work
    a. RNA Extraction – Candidate
    b. Library Construction – Candidate
    c. Sequencing – Centre for Proteomic and Genomic Research (CPGR)
    d. Replication using RT-qPCR – Candidate and Dr Thandiswa Ngcungcu
3. Data analysis
    a. Medical history, lifestyle habits, demographic information data analysis – Candidate
    b. Whole transcriptome sequencing
        i. QC and read count – Candidate and Dr Phelelani Mpangase
        ii. DGE analyses – Candidate, through adjustment and refinement of pipeline developed by Dr Phelelani Mpangase
        iii. Pathway analyses – Candidate and Dr Phelelani Mpangase
    c. Replication data analysis – Candidate and Dr Thandiswa Ngcungcu

I would like to thank the following people:

- My supervisors: Prof Susan Williams, Dr Thandiswa Ngcungcu and Prof Michele Ramsay, for their support, guidance, and endless patience throughout this study. Their encouragement and motivation cannot be understated, of which I will always be grateful.
- Dr Phelelani Mpangase, for his instrumental assistance with the data analysis steps, the use of his pipeline, and his never-ending guidance and support.
- Dr Nadia Carstens and Dr Jacqueline Frost, for their assistance in the early formation of this study.
- Lefentse Mapaisa and Jocelyne Gayenga, for their crucial roles in the administrative processes and assisting me with procuring all I needed for this study.
- Prescilla January, for her assistance with the recruitment process.
- The Centre for Proteomic and Genomic Research (CPGR), specifically Dr Shane Murray, for their assistance with whole transcriptome sequencing.
- The participants involved in this study, who all underwent an ophthalmic examination, provided saliva, conjunctival cells and other tissue samples, and without whom this work would have not been possible.
- The postgraduate students and staff in the Division of Human Genetics, NHLS and SBIMB for their help, support and endless encouragement.
- Funders:
    - NRF Foundation,
    - Faculty Research Committee
    - NHLS Research Trust
    - Carnegie Corporation
- Finally, I would like to thank my friends and family for their continued support and love through all the years.

Table of Contents

DECLARATION i
DEDICATION ii
PRESENTATIONS ARISING FROM THIS STUDY iii
PUBLICATIONS ARISING FROM THIS STUDY iii
ABSTRACT iv
ACKNOWLEDGEMENTS v
LIST OF FIGURES & TABLES ix
LIST OF ABBREVIATIONS xi
CHAPTER 1: INTRODUCTION 1
1.1. Glaucoma 1
1.1.1. Prevalence of Glaucoma 2
1.2. Types of Glaucoma 3
1.2.1. Primary open-angle glaucoma 3
1.2.2. Exfoliation Syndrome (XFS) 4
1.3. Exfoliation glaucoma 5
1.3.1. Prevalence of XFS 6
1.3.2. Treatment 8
1.3.3. Genetic background 8
1.3.4. Potential underlying pathways 11
1.3.5. Environmental factors 14
1.4. Gene expression and exploring target tissues 15
1.5. Study Rationale, Aims and Objectives 18
1.5.1. Study Rationale 18
1.5.2. Aims and Objectives 19
CHAPTER 2: METHODS 21
2.1. Part I: Participant demographics 21
2.1.1. Ethical clearance 21
2.1.2. Recruitment 21
2.1.3. Association between comorbidities, smoking habits, tea, and coffee consumption with XFS/XFG 22
2.2. Part II: Transcriptomics 23
2.2.1. Sample collection 23
2.2.2. RNA extraction 24
2.2.3. Library preparation 25
2.2.4. RNA Sequencing (RNA-Seq) 27

2.2.5. Data analysis: RNA-Seq 28
2.3. Part III: Candidate gene expression (Replication) 30
2.3.1. Experimental procedure and gene selection 30
2.3.2. Data analysis 31
CHAPTER 3: RESULTS 32
3.1. Part I: Participant demographics 32
3.2. Part II: Transcriptomics 35
3.2.1. Pre-transcriptomics 35
3.2.2. Differential Gene Expression Analyses 42
3.2.3. Assessing Potential Confounding Factors (Effect of sex and surgery) 71
3.2.4. Discordant Case analysis 74
3.3. PART III: Candidate gene expression (Replication) 76
3.3.1. Participants 76
3.3.2. Replication of DEGs from transcriptome study 77
3.3.3. RT-qPCR of *LOX* and *LOXL1* 80
CHAPTER 4: DISCUSSION 81
4.1. Participant recruitment outcomes 83
4.2. Novel use of an impression cytology device to harvest cells for whole transcriptome sequencing 84
4.3. Differential gene expression to detect genes that are significantly up or down regulated in XFS and/or XFG 86
4.4. Known XFS/XFG-associated pathways 91
4.5. Identification of a novel potential contributory pathway unique to our South African cohort 94
4.6. Samples from discordant eyes show similar expression patterns 96
4.7. RT-qPCR: Replication and investigation into *LOXL1* and *LOX* gene expression 96
4.8. Study Limitations/Challenges 97
4.9. Future Studies 98
CHAPTER 5: CONCLUSIONS 99
Bibliography 100
APPENDICES 110

LIST OF FIGURES & TABLES

List of Figures

Figure 1.1. Anatomy of the human eye. 3
Figure 1.2. Presence of exfoliation syndrome material in ocular tissues. 5
Figure 2.1. EYEPRIM conjunctival impression device used to collect conjunctival cells 24
Figure 3.1. BioAnalyser traces showing fragment size distributions. 39
Figure 3.2. BioAnalyser traces showing fragment size distributions following additional quality control performed by CPGR.. 40
Figure 3.3. Mean quality scores (Phred scores) for each sample from FastQC. 41
Figure 3.4. Per sample percentage of mapped reads assigned with different levels of certainty according to HTSeq. 42
Figure 3.5. Workflow of the different comparisons for differential gene expression analysis. 43
Figure 3.6. Volcano plot showing significantly up-regulated or down-regulated genes identified in the case and control comparison. 44
Figure 3.7. PCA plot showing clustering of the case and control samples. Cases are in red, while control samples (ctrl) are shown in blue. 45
Figure 3.8. Heatmap for data visualisation of gene clustering according to the gene expression signals in the case control comparison. 46
Figure 3.9. Significance plots for the GO terms in the gene-set enrichment analysis of the 81 DEGs for the case and control comparison. 48
Figure 3.10. Significance plots for the pathways in the gene-set enrichment analysis of the 81 DEGs for the case and control comparison. 50
Figure 3.11. Volcano plot showing significantly up-regulated or down-regulated genes identified in the XFG and control comparison. 52
Figure 3.12. PCA plot showing clustering of the XFG-case and control samples. 53
Figure 3.13. Heatmap for data visualisation of gene clustering according to the gene expression signals in the XFG control comparison. 54
Figure 3.14. Significance plots for the GO terms in the gene-set enrichment analysis of the 85 DEGs for the XFG and control comparison. 56
Figure 3.15. Significance plots for the pathways in the gene-set enrichment analysis of the 85 DEGs for the XFG and control comparison. 58
Figure 3.16. Volcano plot showing significantly up-regulated or down-regulated genes identified in the XFS and control comparison. 61
Figure 3.17. PCA plot showing clustering of the XFS and control samples. 62
Figure 3.18. Heatmap for data visualisation of gene clustering according to the gene expression signals in the XFS control comparison. 63
Figure 3.19. Significance plots for the GO terms in the gene-set enrichment analysis of the 36 DEGs for the XFS and control comparison. 65
Figure 3.20. Significance plots for the pathways in the gene-set enrichment analysis of the 36 DEGs for the XFS and control comparison. 67
Figure 3.21. Venn diagram showing the overlap of DEGs identified in the 3 comparisons. 69
Figure 3.22. Heatmap for data visualisation of gene clustering according to the gene expression signals in the discordant-case control comparison. 75
Figure 3.23. Interaction network of all DEGs identified in the case-control comparison. 78
Figure 3.24. The calculated mean gene expression per gene compared between the control samples (n = 5) and the case samples (n = 9). 79
Figure 3.25. The calculated mean gene expression for LOXL1 and LOX compared between the control samples and the case samples. 80

List of Tables

Table 2.1. Inclusion and Exclusion criteria used for the study _____ 22
Table 2.2. Incubation steps for first strand cDNA synthesis_____ 26
Table 2.3. PCR steps for DNA enrichment of libraries _____ 27

Table 3.1. Demographics and medical history of all participants 33
Table 3.2. Lifestyle habits of all participants 34
Table 3.3. Average quality check results for extracted RNA using the Nanodrop, Qubit and Bioanalyser comparing the yield for one imprint to two imprints per eye 35
Table 3.4. Summary of case samples used in all subsequent transcriptome analyses 37
Table 3.5. Breakdown of sex and age of participants in the transcriptome study 38
Table 3.6. Library concentrations and yields per pool 38
Table 3.7. Top 10 DEGs identified in the case and control comparison. 47
Table 3.8. GO terms and genes identified based on the 81 DEGs from the case-control comparison 49
Table 3.9. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 81 DEGs from the case-control comparison 51
Table 3.10. Top 10 DEGs identified in the XFG and control comparison. 55
Table 3.11. GO terms and genes identified based on the 86 DEGs from the XFG case-control comparison 57
Table 3.12. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 86 DEGs from the XFG case-control comparison 59
Table 3.13. Top ten DEGs identified in the XFS and control comparison. 64
Table 3.14. GO terms and genes identified based on the 36 DEGs from the XFS case-control comparison 66
Table 3.15. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 36 DEGs from the XFS case-control comparison 68
Table 3.16. GO terms and genes identified based on the 17 DEGs identified in all three comparisons 70
Table 3.17. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 17 DEGs identified in all three comparisons 71
Table 3.18. GO terms and genes identified based on the 68 DEGs identified in XFG cases only 72
Table 3.19. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 68 DEGs identified in XFG cases only 73
Table 3.20. Summary of case and control samples used in the replication study, with age, sex, diagnosis, topical treatment and surgical history described. 76
Table 3.21. Summary of samples used in the replication of RNA sequencing through RT-qPCR 77

LIST OF ABBREVIATIONS

ΔCt	Change in thermal cycle
°C	Degrees Celsius
µl	Micro litre
A	Adenine
ACG	Angle closure glaucoma
bp	Base pair
cDNA	Complementary DNA
CNS	Central nervous system
Cq	Quantification/threshold cycle
DEG	Differentially expressed genes
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
ds	Double stranded
dTTPs	Deoxythymidine triphosphates
dUTP	Deoxyuridine triphosphate
ECM	Extracellular matrix
eQTL	Expression quantitative trait loci
FDR	False discovery rate
FGF	Fibroblast growth factor
FLS	Fibroblast-like synoviocytes
gDNA	Genomic DNA
GAG	Glycosaminoglycan
GO	Gene Ontology
GWAS	Genome-wide association studies
HA	Hyaluronic acid
HS	High sensitivity
Hz	Hertz (frequency per second)
IL	Interleukin
IOP	Intraocular pressure
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	Long non-coding RNA
LS	Low sample
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
MMP	Matrix metalloproteinases
mRNA	Messenger RNA
ng	Nano gram
n	Number
OAG	Open angle glaucoma
PCA	Principal component analysis
PCR	Polymerase chain reaction
POAG	Primary open angle glaucoma
PRR	Pattern-recognition receptor
QC	Quality control
RA	Retinoic acid
RIN	RNA integrity number

RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
SCI	Spinal cord injury
SD	Standard deviation
SNP	Single nucleotide polymorphism
T	Thymine
TGF	Transforming growth factor
TIMP	Tissue inhibitors of metalloproteinases
TLR	Toll-like receptor
UVR	Ultraviolet radiation
vs	Versus
WES	Whole exome sequencing
WGS	Whole genome sequencing
XFG	Exfoliation glaucoma
XFM	Exfoliation material/fibrillar extracellular matrix material
XFS	Exfoliation syndrome

CHAPTER 1: INTRODUCTION

There is a paucity of data focusing on gene expression profiling of individuals with exfoliation syndrome or exfoliation glaucoma, particularly in South Africa. This study aims to explore the use of impression cytology as a means of collecting ribonucleic acid (RNA) from these individuals and using this RNA to interrogate the transcriptome of exfoliation syndrome and glaucoma. To illustrate the utility of this device, RNA extraction, complementary deoxyribonucleic acid (cDNA) library construction, whole transcriptome sequencing and subsequent differential gene expression (DGE) analysis was completed. This was then corroborated through a small replication step using real-time quantitative polymerase chain reaction (RT-qPCR) to validate the identified differentially expressed genes (DEGs) as well as other known associated genes.

This chapter will explore the background and motivations for the study by outlining the current understanding of exfoliation syndrome and glaucoma, particularly in South Africa. First, glaucoma and its prevalence will be discussed in Section 1.1., followed by the relevant types of glaucoma (Sections 1.2). Section 1.3 will outline exfoliation syndrome and glaucoma and its prevalence, treatment, genetic background, and contributory environmental factors. Section 1.4 will focus on the steps taken in literature to understand the underlying gene expression profiles of this disease, including limitations of this type of approach. Section 1.5 will then highlight the main motivation; study aims and objectives of this study.

1.1. Glaucoma

Glaucoma refers to a heterogeneous group of clinically and genetically complex optic neuropathies. These are characterised by progressive retinal ganglion cell death due to axonal degeneration and subsequent optic nerve damage (Yasuda *et al.*, 2014). The retinal ganglion cells are central nervous system (CNS) neurons with their cell bodies in the inner retina and axons located in the optic nerve (Weinreb and Khaw, 2004). The optic nerve damage results from the loss of the retinal nerve fibres. The degeneration of these nerves results in enlargement of the cup-to-disc ratio (Waisberg and Micieli, 2021), known as *optic disc cupping*, which is a characteristic appearance of the optic disc, and visual loss. The degeneration leads to functional

changes in the visual field, as retinal ganglion cells relay visual information to the brain (Soucy *et al.*, 2023). Although glaucoma has been well studied, the underlying pathogenesis is still unclear, with limited understanding of the factors contributing to progression.

1.1.1. Prevalence of Glaucoma

Glaucoma is the leading cause of irreversible blindness globally (Pascolini and Mariotti, 2011). Globally, 57.5 million people were affected by glaucoma in 2015 (Kapetanakis *et al.*, 2016). The prevalence for glaucoma in the United States of America is 2.2% for individuals of European ancestry, while it is estimated to be 5.7% for individuals of African ancestry over the age of 40 (Wiggs and Pasquale, 2017; Kelly *et al.*, 2019). A study carried out in Germany determined the prevalence of glaucoma to be 3.22% for individuals over 50 years of age (Kreft *et al.*, 2019). A prevalence of 5.02% was identified in a Nigerian population (Kyari *et al.*, 2015).

Publications outlining statistics for the prevalence of glaucoma in South Africa are often limited to specific communities or patient groups and this information may not be transferable to other settings or population groups.One study, a population-based cross-sectional survey, determined the prevalence of glaucoma in the Zulu population of Kwa-Zulu Natal to be 4.5% (Rotchford and Johnson, 2002). The study focused on individuals of Zulu ethnic origin (n=1005), resident in South Africa. Another study, which looked at 839 black residents of Temba (considered fairly representative of the suburban black population of South Africa) in the North West Province reported the prevalence to be 5.3% (Rotchford *et al.*, 2003). The participants therefore were selected from all the main Bantu tribal groups. All participants of this study were 40 years of age or older, with 58% of all participants with glaucoma blind in at least one eye, as self-reported and confirmed by an ophthalmologist. The high prevalence in these populations suggest that glaucoma constitutes a significant public health problem in South Africa across many settings.

1.2. Types of Glaucoma

Glaucoma can be classified into two distinct categories based on the appearance of the drainage angle on gonioscopy (I.e. the angle where the iris meets the cornea: iridocorneal angle; Figure 1.1): open-angle glaucoma (OAG) and angle-closure glaucoma (ACG; Weinreb, Aung and Medeiros, 2014). Both categories are further subdivided into primary glaucoma (with no obvious identifiable cause) or secondary glaucoma (where there is a secondary ocular cause for the glaucoma).

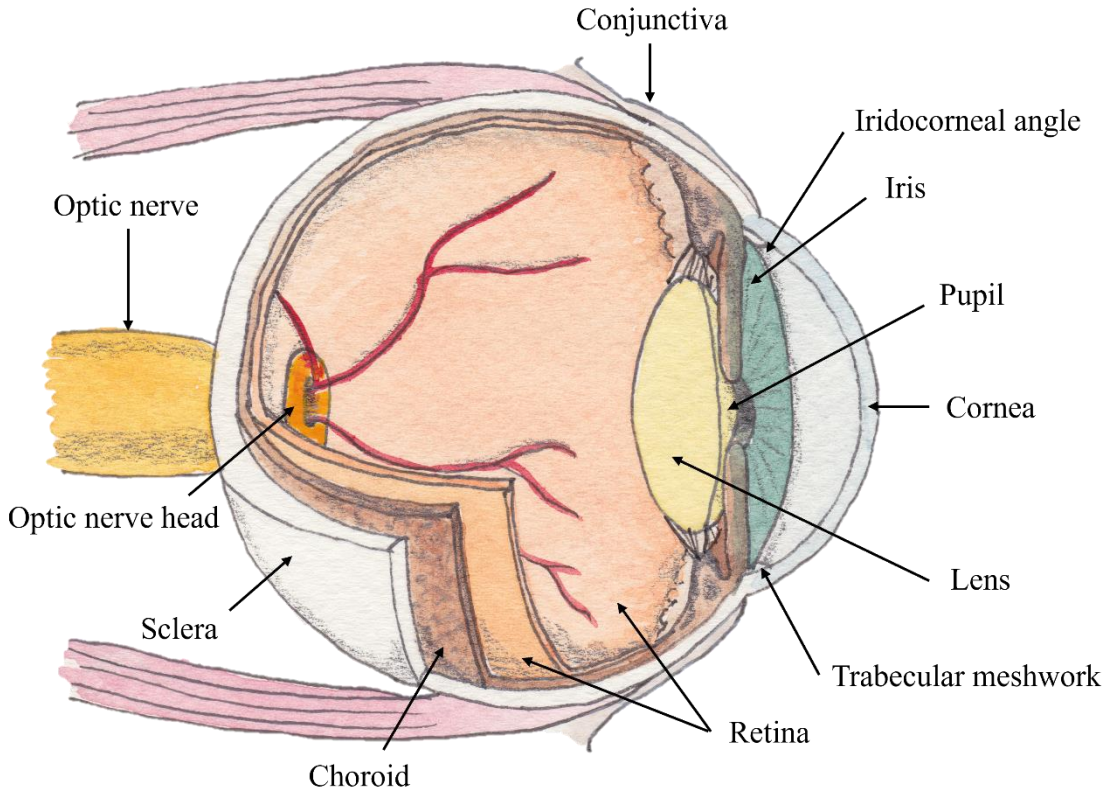


Figure 1.1. Anatomy of the human eye. Cross-section view. Image drawn by Prof Susan Williams.

1.2.1. Primary open-angle glaucoma

Primary OAG (POAG) is the most common type of glaucoma worldwide, comprising about 80% of glaucoma cases in the United States (Weinreb, Aung and Medeiros, 2014). POAG is the most prevalent form of glaucoma in South Africa, ranging from 2.7% to 2.9% of the population (Rotchford and Johnson, 2002; Rotchford *et al.*, 2003). Approximately 87% of participants with POAG in the study by Rotchford *et al*. (2003) had not previously been diagnosed.

OAG, specifically POAG, is the most studied and understood form of glaucoma, however there is still much about the pathogenesis that is not known. About 10% of POAG is of monogenic inheritance, with the remainder being complex, and multifactorial with both genetic and environmental components contributing to the aetiology. There is also a frequent occurrence of families with multiple individuals affected with glaucoma. A positive family history is associated with increased risk of POAG (Chiam *et al.*, 2018) and among POAG cases, positive family historyalso correlated with younger age and female sex (O'Brien *et al.*, 2018). Having access to a patient's family history has also been shown to improve glaucoma-related diagnostic classifications both in population studies and in a clinical setting (Green *et al.*, 2007). One older study (Hampton *et al.*, 1975) identified that correct diagnosis was achieved in 82.5% of cases in a general medical outpatient setting, but was dependent on adequate medical and family history. The ocular manifestation of exfoliation syndrome (XFS) is the most common cause of secondary OAG in the South African populations (Rotchford and Johnson, 2002; Rotchford *et al.*, 2003).

1.2.2. Exfoliation Syndrome (XFS)

XFS was first described by John G. Lindberg in his doctoral thesis in 1917. Lindberg noticed greyish fringes at the pupillary border, which also formed a membrane on the anterior lens capsule (Tarkkanen and Kivelä, 2002) (Figure 1.1A and B). These observations were made in patients, both cataractous and non-cataractous, older than 55 years. Lindberg noted this phenomenon in 50% of chronic glaucoma cases. XFS is an age-related disease, characterised by the accumulation of fibrillar extracellular matrix material (XFM) throughout the body, found in both ocular (Figure 1.1) and extraocular tissues (Zenkel and Schlötzer-Schrehardt, 2014). The extracellular matrix (ECM) forms a complex three-dimensional network in the cellular microenvironment, and is composed of the basement membrane and the interstitial XFM (Marchand *et al.*, 2019). The ECM is highly dynamic and plays a crucial role in cell behaviour and characteristics, such as proliferation, adhesion, migration, polarity, differentiation, and apoptosis (Yue, 2014). Basement membranes of the ECM are formed through interactions among laminins, type IV collagens, nidogens, and proteoglycans, and form cell adhesion platforms and act as solid-phase agonists, specifically for tissue and organ morphogenesis. Dysregulation of the basement membrane can lead to diseases of muscle, nerve, brain, eye, skin, vasculature, and kidney (Yurchenco, 2011).

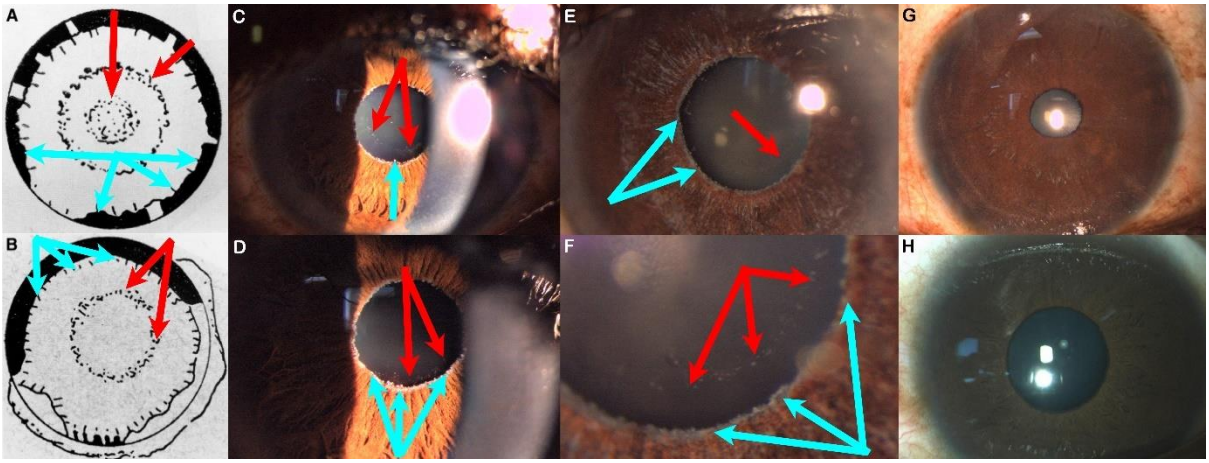


Figure 1.2. Presence of exfoliation syndrome material in ocular tissues. A and B) Drawings of the presentation of XFS in the eye by Lindberg (Tarkkanen and Kivelä, 2002). C and D) Eyes of African patients with XFS. E and F) Eyes of patients with XFS, also exhibiting iris atrophy. G and H) Eyes from African individuals without XFS, therefore no XFM present. Blue arrows indicate XFM on the pupil margin. Red arrows indicate XFM on the lens surface.

Histological studies have been carried out to determine the nature of the fibrillar deposits associated with XFS. The exfoliation material that constitutes these deposits is composed of various basement membrane and elastic fiber constituents. These include, but are not limited to, various ECM constituents: fibrillin-1, elastin, laminin, fibronectin, latent transforming growth factor-β (TGF-β) proteins, clusterin as well as the cross-linking enzyme lysyl oxidase like 1 (LOXL1) (Zenkel and Schlötzer-Schrehardt, 2014; Marchand *et al.*, 2019). Through the deposition of this material throughout the body, XFS, therefore, increases the risk for a variety of non-ocular diseases, including coronary artery disease, hypertension, cerebrovascular disease, abdominal aortic aneurysm, peripheral vascular disease, renal artery stenosis, erectile dysfunction, Alzheimer's-like dementia, sensorineural hearing loss and pelvic organ prolapse (Aboobakar *et al.*, 2017). However, the most recognised disease manifestation of XFS is exfoliation glaucoma (XFG).

1.3. Exfoliation glaucoma

XFG occurs when these extracellular matrix deposits affect the fluid dynamics within the eye. The exfoliation material can be found in deposits within all anterior segment structures of the eye, as well as on the surface of these structures. The anterior segment structures of the eye

include the lens capsule, trabecular meshwork, iris, ciliary body and zonules (Ritch and Schlötzer-Schrehardt, 2001). XFS is diagnosed through the detection of white deposits of exfoliation material on the pupillary border and/or anterior lens surface. The accumulation of exfoliation material deposits and pigment in the trabecular meshwork can lead to a blockage of drainage of aqueous humour from the eye and tissue damage, leading to elevated intraocular pressure (IOP) and progression to XFG.

Compared to other forms of glaucoma, specifically that of primary open-angle glaucoma, XFG is associated with a greater mean IOP, more severe visual field loss at diagnosis, as well as a poorer treatment response (Ritch, 2008). This severity in glaucoma and associated damage to the optic nerve are correlated with the amount of exfoliation material within the trabecular meshwork. This was identified in studies in human donor eyes (Gottanka *et al.*, 1997). Additionally, a study assessing factors for progression in the Early Manifest Glaucoma Trial (Leske *et al.*, 2003), showed that while 53% of 255 patients with open-angle glaucoma progressed over the course of two years, 83% of those with exfoliation glaucoma showed progression.

XFS also increases the risk for other ocular conditions. These include angle-closure glaucoma, climatic droplet keratopathy, cataract formation, zonular weakness, and retinal vein occlusion (Davis and Schuman, 2013; Ritch, 2014). These damaging associations with XFS highlight the need for earlier interventions and thus early diagnosis of XFS. Early identification of XFS and subsequent management may reduce or even prevent XFG and glaucoma-related visual loss.

1.3.1. Prevalence of XFS and XFG

The prevalence of XFS varies widely across populations, reported both with and without XFG. This may be a true observed difference, but could be the result of the focus that each study reported. While some studies focus on the prevalence in the general population, others have looked at the prevalence of XFS/XFG in glaucoma cases. XFG is most prevalent in Scandinavian countries (Traboulsi and Sarfarazi, 2008), as well as countries such as Ireland, Greece and Saudi Arabia, where it comprises over 50% of OAG cases. XFG has also been

found at a high prevalence in European Americans, especially when compared to African Americans (Ritch, 2014).

Few studies have focused on the prevalence of XFS in African countries, with the prevalence between populations varying widely, but generally found to be lower than those reported in European studies. A study was carried out looking at the prevalence of glaucoma in a rural East African population (Buhrmann *et al.*, 2000). Six villages in the Kongwa district in Tanzania were randomly selected and all residents over the age of 40 years were recruited to complete a comprehensive ophthalmic examination. The final cohort number was 3247, with an average age of 53.3 years. Exfoliation syndrome was not detected in any individual. A study on 2142 Congolese patients found 37 individuals with exfoliation material, at a frequency of 1.73% (Kaimbo, 2012).

Another study focused on the prevalence of XFS in a Nigerian population (Olawoye *et al.*, 2012). A total of 448 individuals were recruited and completed an ophthalmic examination. Patients with exfoliation material on the anterior lens surface or pupillary margin in either of their eyes were considered to have XFS. Most of the recruited individuals (94.2%) were of the Yoruba tribe from southwestern Nigeria, with 5.8% from southeastern Nigeria. They identified 12 patients with XFS (2.7%), all from the Yoruba tribe, with a mean age of 65.6 years. They also identified exfoliation material predominantly in males (66.7%). A follow up study on 620 patients living in Northern Nigeria identified nine patients with XFS, a frequency of 1.5% (Idakwo *et al.*, 2018). When looking at a cohort of individuals with glaucoma, POAG was found to be the underlying diagnosis in 94.5% of all glaucoma diagnoses in a study on 362 individuals in a West African population (Budenz *et al.*, 2013), thus supporting a low prevalence of XFS-related glaucoma.

The prevalence of XFS is higher in black South African populations, specifically when compared to South Africans of European ancestry. A study looking at the prevalence of XFS in an urban South African clinic population of European (n=431) and African (n=234) ancestry was carried out (Luntz, 1972). They detected exfoliation material in 1.4% of the European ancestry population and in 20% of the African ancestry population. More recent studies have

replicated the finding of increased prevalence of XFS and XFG in Black South Africans. The presence of exfoliation material was found in 21.6% of 51 glaucoma cases in a South African Zulu population (Rotchford and Johnson, 2002). The South African-based study in Temba, which included most ethnicities, found XFS to be responsible for 16% of all glaucoma cases (Rotchford *et al.*, 2003). The increased prevalence of the more progressive form of glaucoma, that with XFS, within the Black South African population needs to be addressed as this enhances the burden of the disease.

1.3.2. Treatment

Although XFS-related glaucoma is associated with faster progression and greater severity, the treatment of patients with XFG is similar to the treatment of patients with POAG (Ritch and Schlötzer-Schrehardt, 2001). Treatment has not been tailored specifically to XFG. The management typically involves topical ocular hypotensive therapy as with POAG, however poorer outcomes and responses are more common in patients with XFG compared to POAG (Konstas *et al.*, 2004). This is most likely due to the higher baseline IOP in patients. Increased observation and more aggressive treatment are often required.

1.3.3. Genetic background

Family history is a risk factor in exfoliation syndrome. A study focusing on exfoliation syndrome in six Icelandic families, with at least one affected member over 70 years of age, retrospectively observed at least one other affected individual in the second generation of each family. An affected parent was also identified in all cases either by physical examination or through ophthalmic records; with no cases where an affected individual had unaffected living parents (Allingham *et al.*, 2001).

The link to family history suggests strong genetic underpinnings to exfoliation glaucoma. This, along with the variability in prevalence across different ethnic populations, has motivated the study of inherited risk factors within South African populations. Studies have identified several genes and genetic variants that may play a role in the development of XFS, highlighting the hereditary aspects of the disorder. A large portion of the identified genes, which will be

described below, are implicated in a functional capacity within the production or maintenance of the extracellular matrix.

XFS is a complex trait with multifactorial aetiology involving both genetic and environmental factors. Some XFS cases lead to the development of XFG, while others do not. The genetic susceptibility for XFS, and the subsequent development of XFG, is poorly understood but several studies have identified associated genetic variants. The two variants of most interest have been identified through genome-wide association studies (GWAS). The first GWAS on XFS was reported in 2007 on a Scandinavian population (Thorleifsson *et al.*, 2007). One of the strongest candidates involved in XFS and XFG, lysyl oxidase-like 1 (*LOXL1*) was identified. *LOXL1* is a member of the lysyl oxidase gene family, a group of enzymes involved in the synthesis and maintenance of elastic fibers. The involvement of *LOXL1* in XFS pathogenesis has biological relevance due to the nature of XFS as a disorder of elastic tissues (Schlötzer-Schrehardt, 2009).

Variants at the *LOXL1* locus are associated with increased XFS susceptibility. This gene is located on chromosome 15q24.1, with three significant XFG-associated single nucleotide polymorphisms (SNPs) identified; namely rs1048661, rs3825942 and rs2165241 (Thorleifsson *et al.*, 2007). The estimated effect of rs2165241, located in intron 1 of *LOXL1*, was weak for POAG, but this signal was significantly stronger for XFG. The other two SNPs, rs1048661 and rs3825942, are both non-synonymous SNPs and showed strong association with XFG. Since then, various studies have replicated the significant association of *LOXL1* variants with XFS in other populations.

Variants in *LOXL1* for both XFG and POAG were also analysed in a black population from South Africa (Williams *et al.*, 2010). The results strongly confirmed the association between *LOXL1* variants, rs1048661 and rs3825942, and XFG. However, importantly, the A allele of rs3825942 was found to be the risk allele. This contrasts to previous studies reporting the G allele as the risk allele, suggesting other causal variants within or close to *LOXL1* may contribute to XFG genetic risk, rather than rs3825942 (Williams *et al.*, 2010). Due to this discrepancy in the risk allele, other studies have also commented on the implied complexity of the genetic architecture underlying XFS and the need for further study (Aung *et al.*, 2017).

Larger population studies analysing the functional role for the *LOXL1* variants have also included some South African cases. While variants found to be associated with South African XFS cases disrupted promoter activity, functional analysis showed that this did not lead to increased promoter activity (Hauser *et al.*, 2015). Deep sequencing identified a critical region of XFS association in a 70-kb region containing the *LOXL1* exon 1/intron 1 junction, immediately upstream of the *LOXL1* antisense RNA (*LOXL1-AS1*) (Hauser *et al.*, 2015). This finding was replicated in the US Caucasian, German and Japanese populations. *LOXL1-AS1* is a long non-coding RNA (lncRNA) encoded on the opposite strand of *LOXL1*. The research suggested the strongly associated XFS risk alleles were of functional significance, but rather through modulation of the activity of the *LOXL1-AS1* promoter. Both oxidative and cyclic mechanical stress were found to alter *LOXL1-AS1* expression significantly in different human ocular tissues (Hauser *et al.*, 2015), supporting a functional role for the risk variants through the cellular stress response and modulation of *LOXL1-AS1* expression.

The second most reported locus associated with XFS susceptibility, is *CACNA1A*. *CACNA1A* encodes the α1A subunit of the type P/Q voltage-dependent calcium channel. Calcium channels are responsible for the transport of calcium ions across cell membranes and regulate the cell's ability to generate and transmit electrical signals. The presence of high calcium concentrations is in direct association with aggregating XFS fibrils, thus providing supporting evidence for the potential contributory role *CACNA1A* may play (Schlötzer-Schrehardt, Körtje and Erb, 2001).

The *CACNA1A* locus was reported following a GWAS study in a Japanese XFS dataset and has subsequently been replicated (Aung *et al.*, 2015). One SNP marker (rs4926244) identified was mapped within the *CACNA1A* gene. The associated SNP, rs4926244, lies within an intronic region near the 3' end of *CACNA1A* and the allelic association is consistent across all major populations studied. The risk allele frequency ranges between 10%-40% within these populations. However, this risk allele increases XFS risk by 1.16-fold, compared to a 10-fold risk increase for the *LOXL*-1 variants.

Aung *et al*. carried out another study in 2017, a GWAS focusing on XFS using 13,838 cases and 110,275 controls from 24 countries, including South Africa (Aung *et al.*, 2017). Association signals at five additional loci, *POMP* (rs7329409), *TMEM136* (rs11827818), *AGPAT1* (rs3130283), *RBMS3* (rs12490863) and *SEMA6A* (rs10072088) were identified. A rare, protective allele at *LOXL1* (rs201011613) was observed in the Japanese population (Aung *et al.*, 2017). The presence of XFS was also significantly associated with functionally deficient *CYP39A1* sequence variants (Li *et al.*, 2021), observed in a cohort that included South African samples. Individuals with XFS were significantly more likely to carry the deleterious *CYP39A1* sequence variants than controls. They also demonstrated that CYP39A1 transcript expression was 47% lower in ciliary body tissues from individuals with XFS.

The identification of novel variants or altered risk alleles for XFS and XFG within a black South African population can provide additional and novel genetic information on these populations. The gap in knowledge within the African context for glaucoma needs to be addressed in the hope of both validating known glaucoma loci and identifying candidate XFG susceptibility genes within specific African populations, which may elucidate the underlying pathogenetic pathways.

1.3.4. Potential underlying pathways

Several cellular pathways have been implicated in XFS pathogenesis. Some of these include transforming growth factor (TGF)-β1 signalling, matrix metalloproteinases (MMP) and tissue inhibitors of metalloproteinases (TIMP) function, homocysteine metabolism, oxidative stress, autophagy, inflammatory cytokines, and hypoxia.

One of the most studied pathways focuses on TGF-β1 signalling, with research into its role within XFS dating back to the early 2000s. The accumulation of ECM material throughout the body suggests that the pathways involved with the development of XFS may contribute toward aberrant synthesis, maintenance, or degradation of the ECM. The transforming growth factor family of proteins, which includes TGF-β1, modulate ECM formation and have been implicated in the development of XFS (Koliakos *et al.*, 2001). Levels of active and latent forms of TGF-β1 are elevated in XFS when compared to matched controls. A positive correlation has also been observed between levels of TGF-β1 and IOP, suggestive of TGF-β1's role in

modulating the ECM in exfoliation glaucoma (Takai, Tanito and Ohira, 2012). The mechanism by which TGF-β1 modulates the ECM is through ECM accumulation via upregulation of various ECM constituents' synthesis. This includes increased expression of *LOXL1*, type I and type III collagens, elastin, fibrillin-1 and fibulin-2 (Zenkel *et al.*, 2011). However, the mechanism of increased activity and upregulation of TGF-β1, thereby affecting ECM constituents, is unknown.

While TGF-β2 has also been investigated in relation to glaucoma, its elevation has consistently been found to be correlated with POAG rather than XFG or XFS. However, in human donor eyes, exposure of TGF-β2 increases fibronectin concentration in the anterior segments of the eye, as well as IOP and extracellular matrix production (Gottanka *et al.*, 2004). As mentioned previously, fibronectin is one of the components of the exfoliation material deposits. As a profibrotic cytokine, when activated, TGF-β2 upregulates several ECM proteins, not only fibronectin, but also elastin, and several forms of collagens. It is also known to induce various growth factors, such as the connective tissue growth factor (*CCN2*/*CTGF*) and fibroblast growth factors (*FGFs*) (Dillinger *et al.*, 2023). *CCN2/CTGF* is a member of the CCN family of matricellular proteins, which exhibit ECM-like structural features and growth factor-like activities which include, but are not limited to, modulation of cell motility, adhesion, proliferation, survival, ECM protein expression and reprogramming of gene expression. *CCN2/CTGF* has been linked to multiple ocular diseases, including diabetic retinopathy, proliferative vitreoretinopathy, neovascular age-related macular degeneration, glaucoma and exfoliation glaucoma (Chaqour and Karrasch, 2020). Levels of *CCN2/CTGF* have also been found to be elevated in individuals with XFG compared to those with XFS, POAG and controls (Ho *et al.*, 2005; Browne *et al.*, 2011). Exposure to CCN2/CTGF of the trabecular meshwork cells resulted in the up-regulation of fibrillin-1, one of the ECM constituents (Browne *et al.*, 2011).

Another pathway, affecting ECM in a more direct manner, involves matrix metalloproteinases (MMPs) and their endogenous inhibitors (TIMPs) (Arpino, Brock and Gill, 2015). MMPs and TIMPs are both involved in critically maintaining ECM homeostasis. While MMPs promote ECM degradation and turnover, TIMPs work to inhibit this degradation. If there is an imbalance in the ratio of MMPs to TIMPs, which is usually 1:1, an accumulation of abnormal

ECM material can result within both ocular and systemic tissues (Weinreb *et al.*, 2020). This imbalance has been observed in individuals with XFS, as well as POAG (Ashworth Briggs *et al.*, 2015), suggesting it is not XFS-specific. The MMPs and TIMPs most associated with XFS and inducing abnormal deposition of extracellular matrix material are *MMP-2*, *MMP-9* and *TIMP-1* (Chen *et al*. 2021; Starikova *et al*. 2021).

Another important pathway with associations to XFS and ECM modulation is the inflammatory cytokine response pathway. Inflammatory cytokines, which include chemokines, regulate acute inflammation and immune response, as well as regulate the deposition and remodeling of the ECM (Nian *et al.*, 2004). Additionally, inflammatory cytokines produce pro-fibrotic effects. Increased expression of pro-inflammatory cytokines, IL-6 and IL-8, has been observed in anterior segment tissues, however these samples were from early stage XFS, and this increase was not observed in late-stage tissue samples (Zenkel *et al.*, 2010). This increased expression of the pro-inflammatory cytokines leads to an association with XFS (Zenkel *et al.*, 2011; Aung *et al.*, 2017) . Tenon's capsule fibroblast cells were treated with IL-6, which led to an increase in expression of elastin, fibrillin-1, fibulin-2 and TGF-β1, all constituents of the ECM or involved in its modulation. Cytokines have also been linked to activation of MMPs and TIMPs (Ribeiro Vitorino *et al.*, 2023) and may therefore may direct ECM modulation. It is unclear whether the pro-inflammatory response is a direct contributor to XFS development or is a response to XFS.

One of the more recently identified pathways is the retinoic acid receptor-mediated signalling pathway. Due to the strength of previous associations of *LOXL1* with XFS and XFG but conflicting results, *LOXL1* genetic analyses extended to include the flanking intergenic regions. One such study used nine case-control datasets from multiple populations, including Japan, Italy, South Africa, Greece, India, Pakistan, Mexico, Russia, America and Germany (Berner *et al.*, 2019). They detected a non-coding SNP, rs7173049A>G, located downstream of *LOXL1*. The minor allele, G, was found to be enriched in controls across all studied populations, suggesting a significant functionally protective effect. Functional assay results supported a regulatory role of the genomic region surrounding rs7173049. The potentially protective SNP was associated with increased expression of *STRA6*, stimulated by retinoic acid receptor 6, and *ISLR2*, immunoglobulin superfamily containing leucine-rich repeat protein 2, both involved in

retinoic acid (RA) signalling. *STRA6* and *ISLR2* were both significantly down-regulated in ocular tissues of individuals with XFS, along with other genes involved in STRA6 receptor-driven RA signalling. Down-regulation of RA signalling results in significant upregulation of *LOXL1* and other genes involved in fibrosis associated with XFS. The protective effect observed here adds to the complexity of XFS.

1.3.5. Environmental factors

Another aspect adding to the complexity of studying genes and pathways associated with XFS and XFG is an additional element that may contribute to XFS risk - environmental and lifestyle factors. The development of XFS most likely stems from an interaction between genetic and environmental influences rather than having a single influence. One of the most studied potential environmental contributors is exposure to ultraviolet radiation (UVR). This has been identified through epidemiological studies which suggest that geography and in turn increased UVR exposure for prolonged periods may contribute toward XFS risk. A retrospective study looked at 626 901 eye care recipients from 47 states from the USA, dating from 2001 to 2007 (Stein *et al.,* 2011). This study determined that having a northern-tier residence (above 42°N) led to an increase in the risk of XFS, while southern-tier residence (below 37°N) was associated with a reduced risk.

In addition to UVR exposure, dietary habits and intake have also been studied in connection to XFS risk. The association between caffeine and caffeinated beverage consumption in relation to the risk of XFS and XFG was examined in a cohort from the United States (Pasquale *et al.*, 2012). This cohort consisted of 78 977 women from the Nurses Health Study and 41 202 men from the Health Professionals Follow-up Study, who were at least 40 years of age, did not have glaucoma and underwent eye examinations from 1980 or 1986 until 2008. Validated questionnaires were used to determine the consumption statistics of caffeine-containing beverages throughout this period. They compared participants with a total caffeine consumption of <125 mg/day to those who consumed >500 mg/day. Those in the higher caffeine consumption group showed a non-significant trend toward increased risk of XFS and XFG. However, compared to those who abstained from coffee, those with increased coffee consumption were found to be at a 1.66 increased risk of developing XFS and/or XFG (RR = 1.66; 95% CI, 1.09 – 2.54; p = 0.02). These associations were stronger with women and when

family history of glaucoma was present. They did not observe correlations with increased risk with other caffeinated products though, such as soda, tea, or chocolate.

In addition to caffeine consumption, folate intake, specifically through Vitamin B, has also been implicated in increased XFS/XFG risk (Kang *et al.*, 2014). A national prospective cohort study with 20 years of follow-up data from the same cohort as Pasquale *et al*. (2012) was used; the Nurses Health Study (78 980 individuals) and the Health Professionals Follow-up Study (41 221 individuals). B vitamin intake, which included folate, vitamin B6, and vitamin B12, was recorded through repeated administration of validated questionnaires. While vitamin B6 and vitamin B12 did not correlate with increased XFS/XFG risk, a suggested trend of reduced risk was observed with higher intake of folate. This association was observed only for supplemental folate intake and not for dietary folate. The authors suggest that this association could potentially link homocysteine in XFS development, due to increased homocysteine in folate deficiency. Additionally, folate is involved in vascular health via their contributory effects in reducing oxidative stress. The effect of folate in exfoliation syndrome may be modulated by UVR, specifically in individuals of African descent, as darkly pigmented skin is protective of UVR exposure and potential damage, including folate degradation (Wolf and Kenney, 2019).

The summary of the discussed studies highlights the complexity of XFS and XFG development, with potential contributory factors extending from genetic to environmental. This stresses the need to record and study XFS and XFG from a multidisciplinary approach, attempting to analyse as many of these factors as possible to obtain impactful data that can move research and understanding of XFS/XFG forward.

1.4. Gene expression and exploring target tissues

One way to identify or expand current knowledge of the pathways and genes involved in XFS is to identify changes in gene expression in the affected tissues compared to unaffected controls. To accomplish this, the methodology for studying these changes and identifying DEGs should allow for quantitative analysis to determine which genes are being differentially expressed between the XFS (with and without XFG) and control groups. The detection of DGE

can aid in elucidating the cellular processes involved, as functional pathways usually require the interaction of multiple genes (Nickells and Pelzel, 2015).

Most investigations into gene expression in an ocular context have focused on POAG, specifically on changes within the optic nerve, optic nerve head, retina (Yasuda *et al.*, 2014), the trabecular meshwork (Liu *et al*. 2013) and the endothelial cells of Schlemm's canal (Porter, Epstein and Liton, 2012). However, there are several studies which have conducted gene expression profiling of XFS and XFG. A whole transcriptome study was carried out in Australia (Mullany *et al.*, 2022). The authors of this study investigated novel pathways in XFS through RNA sequencing of the lens capsular epithelium. They completed DGE analysis between their cohorts of eight individuals with XFG, 17 with XFS, and 27 controls. This study also included 11 individuals with non-XFS related glaucoma. When combined, their comparison involved 25 individuals with XFS/XFG and 38 without XFS. They identified novel associations between upregulated genes in the mitochondrial and ribosomal pathways in XFS compared to the non-XFS samples. Additionally, they observed XFS-specific down-regulation of genes involved in pathways associated with cell-cell adhesion and plasma membrane function. They also observed the upregulation of TGF β-2, however no significant difference in any matrix metallopeptides or their inhibitors was identified.

Another study investigated XFS through genome-wide RNA sequencing, however this study utilised human trabecular meshwork cells treated with TGF-β1 (Roodnat *et al.*, 2022). Their goal was to elucidate the role of TGF-β1 in the changes observed in the trabecular meshwork in XFS. The trabecular meshwork cells were harvested from donor eyes (n = 4). These cells were stimulated with recombinant human TGF-β1. DGE analysis was carried out by comparing non-stimulated cells to the stimulated cells. The remodeling of the basement proteins involved in ECM organisation was identified as the most enriched pathway, suggesting that TGF-β1 strongly affects the remodeling of the ECM through the basement proteins.

One of the biggest challenges for analysing gene expression profiles of any disease is the focus tissue or cell type. The regulation of genes and their expression is highly tissue/cell dependent and can thus influence results. The ocular cells obtained for the above-mentioned studies were either obtained from mice, human post-mortem tissues or through surgical procedures on patients. The phenotypic analysis of the patient is completed retrospectively through clinical

records when working with post-mortem tissue, which can be limiting and unreliable. Additionally, the interval between donor death and subsequent RNA extraction varies between studies and may also impact the results (Tian *et al.*, 2015).

Additionally, one of the primary purposes of glaucoma research is to identify biomarkers or clinical indicators of early-stage disease. This requires a suitable sample from the patient in question and therefore a non-invasive, easily acquired tissue is needed for potential future genetic/transcriptome testing. One alternative to the widely used tissues in glaucoma and XFG studies is conjunctival cells, which can be harmlessly harvested from the surface of the eyes of living individuals. Conjunctiva regulate the ocular surface immune response and facilitate smooth movements of the eye through mucus-producing goblet cells (Pflugfelder and de Paiva, 2017). The degree to which these cells could be used to detect changes in gene expression between patients with XFS-related glaucoma, XFG, and a control group is unclear. However, through conjunctival biopsies, exfoliation material has been observed in the conjunctiva prior to its appearance in the anterior chamber of the eye (Ritch *et al.*, 2010). Conjunctival biopsies can be used to identify evidence of early deposition of exfoliation material in patients with XFS, representing the earliest stage at which XFS has been detected (Oliveira *et al.*, 2006). Conjunctival cell-thickness has also been shown to change based on type of glaucoma (Van Ginderdeuren *et al.*, 2016).

Various techniques, including microarray or RT-qPCR (Tong *et al.*, 2009; Shibata *et al.*, 2020), and whole transcriptome sequencing studies (Xiang *et al.*, 2019) have previously been used for gene expression profiling of conjunctival tissue. However, these studies utilised surgically excised tissues and focused on other ocular diseases. An alternative non-invasive route to surgical excision is impression cytology. Impression cytology can be performed using the EYEPRIM™ conjunctival cell impression device (OPIA Technologies, Paris, France), a validated and reliable tool for conjunctival impression cytology. Previously, conjunctival cells collected via this method have been used to extract sufficient high-quality RNA (López-Miguel *et al.*, 2017). This technique has been used for gene expression profiling through quantitative PCR and droplet digital PCR (Ganesalingam *et al.*, 2019) or in microarray studies (Latta *et al.*, 2021), however, it has not been completed for whole transcriptome sequencing (Thia and Tong, 2019). Therefore, one aim of this study was to demonstrate that transcriptome analysis using

conjunctival cells obtained by the EYEPRIM device would lead to the identification of DEGs in XFS and XFG study groups that might provide novel insights into early glaucoma pathogenesis and the associated genes and pathways involved.

1.5. Study Rationale, Aims and Objectives

1.5.1. Study Rationale

Understanding the pathological processes underlying glaucoma is essential for early detection (diagnosis) and to develop novel or improved therapeutic strategies that may delay the onset of XFG. Such knowledge could lead to the identification of biomarkers for earlier detection of the disease, as well as subsequent prevention and management (Nickells and Pelzel, 2015). Therefore, it is imperative to address the knowledge gap regarding the contributory or causative genetic or molecular pathways. Due to the prevalence of glaucoma worldwide, its complex multifactorial nature, including genetic susceptibility, as well as its more frequent and more severe effect on populations of African descent, this makes it especially important to study in the South African context.

The main purpose of this study was to analyse the gene expression profile in participants with XFS and XFG to identify changes in the mRNA expression of genes across these disease states, as well as compared to a control group. The analysis of these gene expression profiles was completed using RNA-Sequencing (RNA-Seq), making use of next-generation sequencing technology. This served to provide detailed and practical knowledge regarding the mechanisms of disease, elucidating the association of certain pathogenetic pathways.

Studies focusing on pathway or functional analyses are lacking in the black South African patients with XFS/XFG. While studies have observed genetic associations in South African populations, data regarding this population is sparse, in particular expression and functional studies. The underlying changes in expression or gene regulation and pathways involved in the development of exfoliation syndrome and glaucoma are still unclear. The validation of proposed contributory pathways or identification of new associations can shed light on the reason that black South African populations have a higher prevalence of exfoliation glaucoma

compared to other populations of African ancestry (Luntz, 1972; Rotchford and Johnson, 2002; Rotchford *et al.*, 2003; Kaimbo, 2012; Olawoye *et al.*, 2012; Idakwo *et al.*, 2018). This could advance research toward the translation of information into the clinical management of exfoliation glaucoma. New discoveries regarding the pathophysiology of XFS may lead to more XFS-specific management of patients and improved treatment options.

1.5.2. Aims and Objectives

The core aim of this quantitative study was to investigate the underlying genetic pathways involved in exfoliation syndrome, and subsequent development of XFG. To achieve this, three approaches were used to examine potential contributors to the development and progression of exfoliation syndrome, namely environmental or lifestyle factors, transcriptome-wide expression profiling (transcriptomics), and finally candidate gene expression and replication. These three approaches contributed three parts to this study; the specific aims and objectives of each part are outlined below.

Part I: Participant demographics

The main goal of this section was to examine potential contributing environmental factors, such as caffeine intake, careers, smoking, and geographical location, to the development of XFS.

The two main objectives within this section were:

- Collect key information regarding environmental factors for participants with XFS or XFG and for a control group.
- Compare these factors between the two groups to identify any associations

Part II: Transcriptomics

The second aim of this study was to identify DEGs in conjunctival cells of black South Africans and identify pathways in which these genes are enriched. I hypothesised that genes involved within the development and maintenance of the fibrillar extracellular matrix would be identified as differentially expressed between the XFS and control group. RNA-Seq was used

to perform genome-wide expression profiling in conjunctival cells from living individuals of African descent with XFS and XFG.

The three main objectives within this section were:

1. To determine the feasibility of performing RNA-sequencing using conjunctival cell impression.
2. To identify DEGs associated with exfoliation syndrome in the conjunctiva of black South Africans compared to unaffected controls.
3. To perform enrichment and pathway analysis using the identified DEGs.

Part III: Candidate gene expression (Replication)

The main goal of this section was to use RT-qPCR to replicate the top findings from the transcriptome analysis (Part II) as well as to interrogate a known candidate gene, *LOXL1*.

The two main objectives within this section were:

1. To examine and replicate the expression of the top ranking DEGs in a separate cohort of affected individuals and controls.
2. To compare the expression of *LOXL1* and *LOX* between these affected individuals and controls.

CHAPTER 2: METHODS

This chapter is structured such that the recruitment procedure for participants is included at the beginning as it relates to all the sections, and the research methodology is presented in three sections that align with the three main study objectives. This includes Participant demographics (Part i), Transcriptomics (Part ii) and Candidate Gene Expression and Replication (Part iii).

2.1. Part I: Participant demographics

2.1.1. Ethical clearance

This research study is part of two broader parent studies: "Differentially Expressed Genes in Exfoliation Syndrome, Exfoliation Glaucoma and Primary Open-Angle Glaucoma in Black South Africans" (M160701) (Appendix A) and "Gene Identification in Exfoliation Syndrome and Exfoliative Glaucoma" (M160706) (Appendix B). The participant recruitment was approved under the ethical clearances obtained for these studies. However, additional ethical clearance was obtained for this PhD study through the Human Research Ethics Committee (HREC) at the University of the Witwatersrand, Johannesburg (M170883) (Appendix C), to cover additional research not outlined in the objectives of the original studies.

2.1.2. Recruitment

Participants, both cases and controls, were recruited from the eye clinics at Charlotte Maxeke Johannesburg Academic Hospital and St John Eye Hospital in Johannesburg, Gauteng, South Africa. Written, informed consent was obtained after a thorough explanation of the research study. Each participant was then invited to participate, and upon acceptance, a 30-minute interview session was conducted. A translator was used when the participants were not fluent in English. An information sheet was used to gather demographic information, medical history, as well as lifestyle habits, such as smoking and caffeine consumption (Appendix D).

Each participant then underwent an ophthalmological examination by ophthalmologists at Charlotte Maxeke Johannesburg Academic Hospital or St John Eye Hospital. The ophthalmological examination was performed to determine whether participants met inclusion or exclusion criteria (Table 2.1).

Table 2.1. Inclusion and Exclusion criteria used for the study

	Inclusion criteria	**Exclusion criteria**
Cases	Black Southern African ancestry	Other ancestry
	>18 years of age	<18 years of age
	Presence of XFS material	Other forms of glaucoma
Controls	Black Southern African ancestry	Other ancestry
	>18 years of age	<18 years of age
	No presence of XFS material	All forms of glaucoma

2.1.3. Association between comorbidities, smoking habits, tea, and coffee consumption with XFS/XFG

This part of the study utilised the information sheet mentioned in Section 2.2. (Recruitment), which was used to gather demographic information, medical history, as well as lifestyle habits, such as smoking and caffeine consumption. The values for all information from the 49 cases and 39 controls were converted into contingency tables. Chi-squared tests or Fisher's exact tests (when sample sizes were small, or zero values were included) were used to determine any significant differences between the case and control cohorts in R (R Core Team 2021).

2.2. Part II: Transcriptomics

2.2.1. Sample collection

Once consent and the necessary information was obtained, each participant provided conjunctival cells for RNA extraction purposes. Conjunctival cells were collected by an ophthalmologist using the EYEPRIM conjunctival impression device, following manufacturer's instructions (Figure 2.1; IFU-Eyeprim-2020, instructions for use available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://eyeprim.com/shop/wp-content/uploads/2020/02/IFU-Eyeprim-2020.pdf). The EYEPRIM device contains a membrane at one end, which is exposed upon depression of a trigger mechanism. Following the ophthalmological examination, the EYEPRIM device membrane was applied to the ocular surface fifteen minutes after installation of a topical anaesthetic eye drop and for a three second period, after which the membrane was ejected directly into a container containing RNAlater™ Stabilization Solution (Invitrogen, Carlsbad, California, United States) to limit degradation of the RNA. The samples were incubated in the stabilisation buffer at room temperature for 24 hours to allow for RNALater permeation, following which the samples were stored at -80°C until RNA isolation. Initial results (when one sample was used) yielded insufficient RNA for the purpose of whole transcriptome sequencing, therefore two samples were taken from each eye and combined to make the final sample. One sample was taken from the superior bulbar conjunctiva and one from the inferior bulbar conjunctiva to maximise surface area coverage.

Additionally, in some cases, the presentation of XFS/XFG was discordant, whereby one eye of an individual was affected either by XFG or XFS, while the other eye presented with the opposite or had no visible XFM at all. In these discordant cases, samples were taken from both eyes and collected and analysed separately.

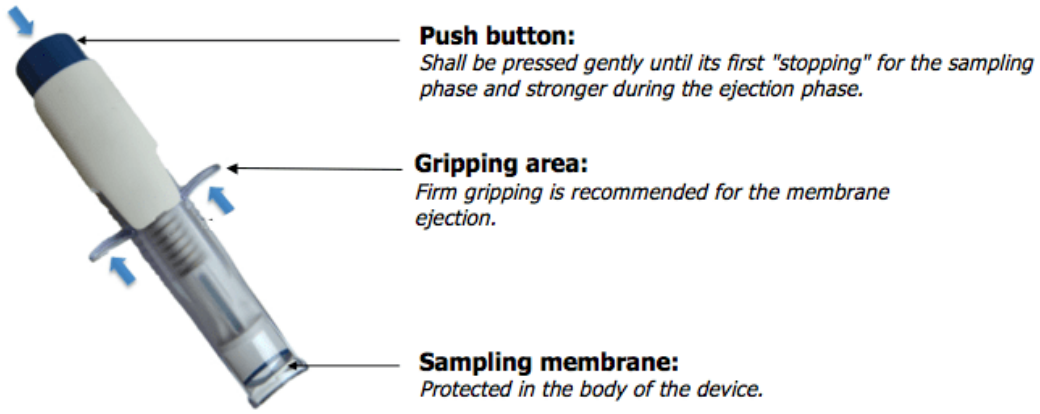


Figure 2.1. EYEPRIM conjunctival impression device used to collect conjunctival cells (*https://eyeprim.com/shop/#1613249909171-72823ed5-47d2*)

2.2.2. RNA extraction

Total RNA was extracted from the conjunctival cells using the InviTrap® Spin Cell RNA Mini Kit (Invitek Molecular, Berlin, Germany). This kit was selected to maximise the RNA yield considering the relatively small number of epithelial conjunctival cells in each sample. The InviTrap® Spin Cell RNA Mini Kit (Invitek Molecular, Berlin, Germany) is listed as an ideal tool for isolation of high-quality total RNA from samples of human cells up to 1 x 10^{-7} cells/ml. Using a spin-filter format, the kit combines an efficient lysis of the starting material with rapid inactivation of RNases, total RNA stabilization and near complete separation from the deoxyribonucleic acid (DNA). The genomic DNA (gDNA) is fixed at the surface of the mineral carrier particles provided within the Lysis Solution R; this occurs during cell lysis. The DNA is then selectively bound by optimized buffer conditions and the mineral carrier particles attached to the DNA are removed via centrifugation on the spin filter surface. The remaining filtrate contains the RNA which is then bound onto the membrane of the RNA-RTA Spin Filter and the RNA-Binding Spin Filters. Contaminants are removed by repeated washing steps and the purified total RNA is eluted in RNase free water.

The manufacturer's instructions were followed for the InviTrap® kit, with two modifications to optimize the RNA yield and adjust to the source material. First, the EYEPRIM™ membrane was removed from the RNALater using RNase-free forceps and placed in the lysis buffer provided with the InviTrap® kit. Second, this buffer containing the membrane was subjected to additional disruption and homogenization using the TissueLyser LT (Qiagen, Hilden,

Germany). This step involved spinning the samples at 30 Hz for 30 seconds at room temperature.

The quality and integrity of the RNA was assessed using the NanoDrop™ One Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, United States), the Agilent RNA 6000 Nano Kit on the Bioanalyser 2100 System (Agilent Technologies, Santa Clara, California, United States) and with the Qubit™ RNA HS Assay Kit using the Qubit™ 3.0 Fluorometer (Invitrogen, Carlsbad, California, United States).

2.2.3. Library preparation

Following RNA extraction, RNA samples with an RNA integrity (RIN) value above 7 and NanoDrop™ 260/280 ratio above 1.8 were used for downstream steps. Complementary DNA (cDNA) libraries were then constructed using these samples with the TruSeq Stranded mRNA kit (Illumina, San Diego, California, United States), according to manufacturer's instructions for the Low Sample (LS) Protocol. The LS Protocol was selected as it can accommodate up to 48 samples being processed at one time, sufficient for the purposes of this research. Stranded RNA-seq was chosen as it retains strand of origin information for each transcript provides a more accurate estimate of transcript expression compared with non-stranded RNA-seq (Zhao *et al.*, 2015). The libraries were constructed in two batches based on when recruitment took place and when the conjunctival cells were sampled.

The workflow for this kit comprised various steps. The first step involved purifying the poly-A containing mRNA molecules using poly-T oligo attached magnetic beads to remove highly abundant rRNAs for efficient gene detection. The purified mRNA was then fragmented into smaller pieces using divalent cations under high temperatures. The cleaved RNA fragments were converted and multiplied into first strand cDNA using reverse transcriptase (Superscript II, Thermo Fisher Scientific, Massachusetts, United States), and random primers (Table 2.2).

Table 2.2. Incubation steps for first strand cDNA synthesis

Temperature	Time
25°C	10 minutes
42°C	15 minutes
70°C	15 minutes
4°C	Hold

Second strand cDNA synthesis was then completed to remove the RNA template and synthesize a replacement, incorporating deoxyuridine triphosphates (dUTPs) in place of deoxythymidine triphosphates (dTTPs) to generate double stranded (ds) cDNA. The incorporation of the dUTP terminates the second strand during amplification. AMPure XP beads (Beckman Coulter Life Sciences, Indianapolis, United States) were then used to separate the ds cDNA from the second strand reaction mix, resulting in blunt-ended cDNA.

To prevent the blunt ends from ligating to one another during the adaptor ligation reaction, a single 'A' nucleotide was added to the 3' ends of the blunt fragments. The corresponding single 'T' nucleotide on the 3' end of the adaptor provided a complementary overhang for ligating the adapter to the fragment. Multiple indexing adaptors (TruSeq RNA Single Indexes Set A, added to distinguish the different samples from one another) were then ligated to the ends of the ds cDNA to prepare for hybridization to the flow cell in the sequencing process.

The next step involved enriching the cDNA fragments. This process entailed utilising specific primers for PCRs to selectively enrich the DNA fragments with bound adaptor molecules on both ends of the fragments and thus amplifying the amount of DNA in the libraries. Table 2.3 outlines the PCR steps required for DNA enrichment.

Table 2.3. PCR steps for DNA enrichment of libraries

Temperature	Time	Cycles
98°C	30 seconds	1
98°C	10 seconds	15
60°C	30 seconds	
72°C	30 seconds	
72°C	5 minutes	1
4°C	Hold	1

AMPure XP Beads were then used to clean up the PCR amplicons/fragments and the resulting libraries were validated using quality control analysis. Quality, integrity, and purity of the libraries was confirmed using the Agilent DNA 1000 Kit (Agilent Technologies, Santa Clara, California, United States) to confirm the library fragments were of the correct size and the Qubit™ dsDNA High Sensitivity (HS) Assay Kit (Invitrogen, Carlsbad, California, United States) to confirm sufficient yield and concentration for sequencing purposes.

2.2.4. RNA Sequencing (RNA-Seq)

Forty of the TruSeq Stranded mRNA libraries were sequenced at the Centre for Proteomic and Genomic Research (CPGR) (Cape Town, South Africa) using the Illumina NextSeq 500/550 Sequencer (Illumina, San Diego, California, United States). These samples were separated into four pools; pool 1 to pool 4. Pools 1 and 2 each contained 12 samples from the first library construction, while pools 3 and 4 contained eight samples each from the second library construction.

Additional quality checks using the Qubit™ and Bioanalyzer were performed by CPGR on the submitted libraries to confirm they met quality requirements. The libraries were quantified using the Qubit™ DNA HS Assay (Invitrogen, Carlsbad, California, United States). Libraries were then pooled by adding an equal amount of each sample to the batch. The average fragment

size of the four pools containing the dsDNA libraries was then determined on a HS DNA Bioanalyzer chip (Agilent Technologies, Santa Clara, California, United States). The libraries were denatured at 95°C for 5 minutes as an additional quality control step and analysed on an RNA 6000 Nano Bioanalyzer chip to confirm the library fragment sizes (Agilent Technologies, Santa Clara, California, United States).

The pooled libraries were quantified using the KAPA Library Quantification Kit for Illumina Platforms (Roche, Basel, Switzerland). A library dilution series of 1:100 000, 1:1 000 000, 1:10 000 000 and 1:100 000 000 was quantified in triplicate. The calculated concentration was then adjusted by size. The pooled libraries were diluted to 4nM, denatured with 0.2N NaOH, diluted to 1.8pM and then combined with denatured PhiX positive control at a spike-in concentration of 1% v/v for the sequencing runs, according to standard Illumina instructions. Sequencing was done at CPGR and completed per pool, with each set of libraries loaded on the Illumina NextSeq 500 or 550 instrument and sequenced using the NextSeq High Output Reagent Kits v2 (150 cycle). The sequencer was programmed to run a paired-end, single indexed 2 x 76 cycle sequencing run.

2.2.5. Data analysis: RNA-Seq

FASTQ files were generated from the sequencing runs using Bcl2Fastq software on an Illumina DRAGEN server. To determine the success of the sequencing runs and viability of using the impression cytology device to collect conjunctiva cells as a source material for RNA-Sequencing, quality control measures were completed.

Illumina's Sequence Analysis Viewer was used to assess the sequencing run through the Quality Metrics. These include the cluster density (K/mm²) and the average Q-scores. The optimal cluster density falls between 170 K/mm² and 220 K/mm². The Q-scores indicate run performance and assess the quality of each called base, where a Q30 score of 30 is equivalent to a 0.1% probability that an incorrect base was called by the sequencer.

Additional quality confirmation checks, read alignment and read count was completed using Nf-Rnaseqcount, a tailored Nextflow pipeline (Mpangase *et al.*, 2021). All references and

scripts for this pipeline and all analyses in R are in a public Github repository (https://github.com/MichaellaHulley/transcriptomic_work). Nf-Rnaseqcount is a Nextflow pipeline for obtaining raw read counts for RNA-seq data using a given reference genome and annotation. This pipeline uses FastQC which provides a simple way to perform some quality control checks on raw sequence data coming from high throughput sequencing pipelines (Andrews, 2010). These quality control checks include per base or sequence quality scores, N content, sequence duplication levels, overrepresented sequences. A per base quality score of 30 is equivalent to at least 99.9% base call accuracy, while N content provides the percentage of base calls at each position for which an N was called when the correct base call is unclear. Trimmomatic was used to trim poor quality reads and adaptor sequences (Bolger, Lohse and Usadel, 2014).

The use of the STAR software package in the pipeline enabled highly accurate and fast alignment of the RNA-seq reads to a reference genome (GRCh38 release 94) (Dobin *et al.*, 2013). HTSeq Count (Putri *et al.*, 2022) and featureCounts (Liao, Smyth and Shi, 2014) were used to count reads. This was completed by taking a file with aligned sequencing reads, as well as a list of genomic features and counting how many reads mapped to each feature.

Once the read counts were obtained, DGE was computed using DESeq2 to identify DEGs (Love, Huber and Anders, 2014). The pipeline includes a step to adjust for sex and age in the analyses. Multiple comparisons were performed using DESeq2 to identify DEGs that may contribute to pathophysiology of XFS. These include: all cases (both XFS and XFG individuals) to controls, XFG individuals to controls, XFS individuals to controls. To account for any confounding factors, all males were also compared to all female participants. A comparison including only the discordant samples was also conducted. The initial Log2 Fold cutoff used for DESeq2 was 1.3, while the False Discovery Rate (FDR) was 0.05. These thresholds were chosen based on a publication by Mullany *et al* (2022), who also analysed DGE in exfoliation syndrome. However, the Log2 Fold cutoff was reduced to ±0.5 (~1.5-fold change) in the final analyses as very few genes met the criterion of 1.3. Analysis was conducted in R (R Core Team 2021) and data visualisation was carried out using the package ggplot2 (Wickham, 2016) in R. All tools and packages used in R are listed in Appendix M, along with the version numbers. Heatmaps were generated using pheatmap_1.0.12 in R, while principal

component analyses were also performed (functions.R at https://github.com/MichaellaHulley/transcriptomic_work).

Gene-set enrichment was carried out using the enrichR (version 1.0) package (Chen *et al*. 2013; Kuleshov *et al*. 2016; Xie *et al*. 2021) in R. The databases for GO terms (Molecular Function 2018, Biological Process 2018 and Cellular Component 2018) and pathways (WikiPathways 2016, Kyoto Encyclopedia of Genes and Genomes (KEGG) 2016, Reactome 2016) were selected from the databases offered by the enrichR package.

2.3. Part III: Candidate gene expression (Replication)

2.3.1. Experimental procedure and gene selection

Replication for the DEGs identified through RNA-sequencing (Part I) was performed using Taqman Gene Expression Assays and the Taqman Fast Advanced Master Mix (Thermo Fisher Scientific, Massachusetts, United States) on the QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific, Massachusetts, United States). Eight of the DEGs from the case-control comparison were selected for validation purposes, based on the Log2 fold value, FDR, literature review of each gene and through interrogation of links between DEGs using STRING v12 (Szklarczyk *et al.*, 2021). These include *PTGS2*, *TGFB2*, *OLFM4*, *CCN2*, *EGR1*, *CDH4*, *RPTN* and *CCL3*.

LOXL1 and *LOX* were also selected, making a total of nine genes, to assess gene expression within the cases compared to control samples, as *LOXL1* is the most studied XFS-associated gene to date and *LOX* is its isoform. Therefore, 10 genes were used for replication, with three additional house-keeping genes, namely eukaryotic translation initiation factor 4A2 (*EIF4A2)*, ribosomal protein L13a (*RPL13A)* and tata-binding protein (*TBP)*. These genes were selected as house-keeping genes due to their high stability observed within ocular surface epithelial tissues, specifically the conjunctiva (Van Acker *et al.*, 2019).

A non-template control was used for each assay, while a no-reverse-transcriptase-control and a calibrator sample was also included for the house-keeping assays. The calibrator sample was

included to account for batch effects across the different runs. All assays were run in triplicate to account for any technical error. The biological conclusions from the RNA-Seq experiments can be further validated by using biological replicates, rather than technical replicates (Fang and Cui, 2011). The replication experiments were performed on a different sample group, which included 9 cases and 5 control samples that were not part of the original DGE study. Different samples were used to act as biological replicates for replication. Validation using qRT-PCR on the samples used for the RNA-Seq analysis would only validate the technology.

2.3.2. Data analysis

Following each run on the QuantStudio, the data was exported and analysed in Microsoft Excel (2016). The threshold cycle, reported as Ct by QS5 software, will be referred to as Cq (quantification cycle) as per MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines (Bustin *et al.*, 2009). This value refers to the number of cycles required for the sample to reach the critical point of quantification and/or detection. The average Cq values for each sample were used in the gene expression analysis based on the △△Ct method (Winer *et al.*, 1999). The Cq values for each sample were extracted and the mean Cq from the associated technical replicates were calculated for each sample and then each biological group (cases and control). The change in Cq (ΔCq) was then determined by subtracting the mean Cq of all the samples from the control group from the mean Cq of the technical replicates for each sample. The relative quantity was then calculated by raising one + the PCR efficiency to the ΔCq. The relative quantity of each gene assay was divided by the geometric mean of the relative quantities of the three housekeeping genes and was normalised through log transformation (ΔΔCq). The average ΔΔCq of the genes for each sample in each biological group was then calculated, along with the standard deviation and confidence intervals. Independent Samples t-tests were used to perform statistical analysis of the log transformed normalised expression per sample between cases and control samples for each gene assay.

CHAPTER 3: RESULTS

To interrogate the hypotheses posed in the introduction, the results are presented in three parts: (i) Participant demographics (ii) Transcriptome and DGE on a subset of cases and controls (iii) Replication of selected DEGs. Participant demographics includes a summary of all participants recruited for the study, along with association studies investigating the link with co-morbidities (hypertension and diabetes), as well as lifestyle habits such as smoking, and tea and coffee consumption. Part II is divided into pre-transcriptomics (sample pre-processing and library construction) and DGE and pathway analyses. Finally, Part III describes the replication of selected DEGs, and expression of the most well-known XFS associated gene, *LOXL1* and its isoform, *LOX*, in cases and controls.

3.1. Part I: Participant demographics

The demographic, medical and lifestyle history data collected through the questionnaire administered during the recruitment process was interrogated to further understand the external factors affecting XFS/XFG.

Recruitment, performed from March 2017 to February 2019, at the eye clinics at Charlotte Maxeke Johannesburg Academic Hospital and St John Eye Hospital in Johannesburg, Gauteng, South Africa, was concluded at the time when 88 participants had been recruited for the study. The 88 participants comprised 39 unaffected controls and 49 affected patients with XFG and/or XFS (cases). In some individuals, the disease presented in a discordant manner, either with one affected eye and one unaffected eye, or one eye presenting with XFS and the second with XFG. Samples were therefore collected from each eye of these individuals. The remaining cases, where both eyes were affected in the same manner, were considered concordant cases, and samples were collected from only one of the eyes.

A structured questionnaire was used to collect demographic information, medical features and lifestyle habits. First, demographic and medical information were used to investigate associations with age, sex and comorbidities (Table 3.1). No significant difference in sex or

age was observed between the case and control groups. Medical history included diabetes and hypertension status of all individuals and no significant differences were observed between the two groups (Table 3.1).

Table 3.1. Demographics and medical history of all participants

		Controls (n=39)	**Cases (XFS and /or XFG) (n=49)**	**P-values**
Mean age ± SD		68.10 ± 8.19	70.16 ± 8.88	0.27
Sex (%)	Male	18 (46.2)	29 (59.2)	
	Female	21 (53.8)	20 (40.8)	0.32
Diabetes (%)	Yes	12 (30.8)	9 (18.4)	
	No	27 (69.2)	40 (81.6)	0.27
Hypertension (%)	Yes	27 (69.2)	25 (51.0)	
	No	12 (30.8)	24 (49.0)	0.13

Then, the lifestyle habits were used to assess the impact of smoking, and tea and coffee consumption on XFS/XFG (Table 3.2). No significant differences were observed when comparing the status of smoking, or consuming tea or coffee between the two groups (Table 3.2). However, a significant difference was observed for amount of coffee consumed (0.03), with higher consumption observed in the control group.

Table 3.2. Lifestyle habits of all participants

		Controls (n=39)	Cases (XFS and/or XFG) (n=49)	P-values
Smoking (%)	Yes	7 (17.9)	17 (34.8)	
	No	32 (82.1)	32 (65.2)	0.13
Pack years	0	32 (82.1)	32 (65.3)	
	1 - ≤4	5 (12.8)	8 (16.3)	
	>4	2 (5.1)	9 (18.4)	0.13
Tea consumption (%)	Yes	35 (89.7)	45 (91.8)	
	No	4 (10.3)	4 (8.2)	1.00
Cups/day	0	4 (10.3)	4 (8.2)	
	0>- ≤2	20 (51.3)	20 (40.8)	
	>2	15 (38.4)	25 (51.0)	0.50
Coffee consumption (%)	Yes	21 (53.8)	23 (46.9)	
	No	18 (46.2)	26 (53.1)	0.67
Cups/day	0	18 (46.2)	26 (53.1)	
	1	12 (30.7)	21 (42.8)	
	>1	9 (23.1)	2 (4.1)	0.03*

* denotes significance of $p < 0.05$

Participants for the subsequent transcriptomic analyses were chosen based on the quality of RNA extraction, as described in the next section. In the next part of the study, the demographics and treatment of only the 35 individuals included in the transcriptome experiments are included and therefore the treatment status was not interrogated further in Part I.

3.2. Part II: Transcriptomics

3.2.1. Pre-transcriptomics

This section reports the results from the RNA extraction and library preparation on a sub-selection of participants. These results explore the feasibility of utilising the EYEPRIM™ impression cytology device to collect conjunctival cells for the purpose of whole transcriptome sequencing.

3.2.1.1. RNA yield and quality

RNA was extracted for all recruited participants. The RNA concentration was too low for the Bioanalyzer System to generate RIN values with only one imprint sample per eye (Table 3.3). With two imprint samples per eye and the additional disruption step using the TissueLyser LT, the concentrations increased, and quality of the RNA improved (Table 3.3). The Bioanalyzer System results showed a mean RIN value of 9.3 for the samples indicating that the RNA was of good quality for the downstream applications.

Table 3.3. Average quality check results for extracted RNA using the Nanodrop, Qubit and Bioanalyser comparing the yield for one imprint to two imprints per eye

Parameter	**Nanodrop Readings**				**Qubit Readings**		**RNA Integrity Number (RIN)**
	RNA Conc. (ng/µL)	**280/260 ratio**	**280/230 ratio**	**RNA yield (ng)**	**Conc. (ng/µL)**	**RNA yield (ng)**	
One imprint per eye (n = 2)	19.40	1.77	0.43	776	8.32	332	NA*
Two imprints per eye (n = 10)	52.90	2.15	1.25	2116	44.53	1780	9.30

RNA quantification showed varying results from using the Nanodrop and the Qubit; *NA = the concentration of the sample was too low to generate a RIN value; Conc: concentration

3.2.1.2. Demographics of Participants included in the transcriptomic analyses

From the total 88 participants, the RNA from 20 cases and 15 control individuals were selected for library construction, based on RNA yield and quality. Five of the cases selected presented with discordant expressions of XFG and/or XFS. Two of these individuals presented with XFG in one eye and XFS in the second, one presented with XFS in one eye and no presence of XFS

in the second, while the remaining two presented with XFG in one eye and no presence of XFS in the second. Impressions from both eyes were taken in these five participants and used for additional DGE analyses, while impressions from only one eye were taken from the concordant cases and from the controls. Therefore, the total number of samples selected for library construction and subsequent RNA sequencing was 40, comprising 25 samples (eyes) from 20 case participants (five discordant cases - two samples per case) and 15 from controls. The sex, age and topical treatment information for all participants was gathered through a questionnaire and the results for the 20 cases are shown in Table 3.4. No control participants were undergoing chronic topical treatment. Only two participants underwent glaucoma surgery, CASE08 and CASE39.

Table 3.4. Summary of case samples used in all subsequent transcriptome analyses

Sample Label	Age	Sex	Specific Diagnosis	Chronic topical treatment
CASE_06			XFS	None
CASE06_LE-NO XFS	53	M	No XFM	None
CASE_07			XFG	AA, PG, BB, CAI
CASE07-RE-NO XFS	63	M	No XFM	None
CASE_08	45	F	XFG	AA, PG, BB
CASE_09			XFG	AA
CASE09_LE-NO XFS	53	M	No XFM	None
CASE_11	76	F	XFS	None
CASE_12	67	F	XFG	AA, PG, BB
CASE_15			XFG	None
CASE15-RE-XFS	80	M	XFS	Topical Steroid
CASE_17	80	F	XFG	Topical Steroid
CASE_19	69	M	XFG	AA, PG, BB
CASE_20	62	M	XFG	AA, PG, BB
CASE_21	58	M	XFG	PG, BB
CASE22-RE-XFS			XFS	None
CASE_22	84	F	XFG	AA, PG, BB
CASE_25	71	M	XFG	AA, PG, BB
CASE_27	87	F	XFS	None
CASE_28	85	M	XFS	None
CASE_32	73	M	XFG	AA, PG, BB
CASE_33	74	F	XFG	AA, PG, BB
CASE_36	66	M	XFG	AA
CASE_38	74	M	XFG	AA, PG, BB
CASE_39	77	F	XFG	Allergy - off treatment

Discordant samples are bolded. AA – Alpha agonists; BB – Beta Blockers; CAI – Carbonic Anhydrase Inhibitors; PG – Prostaglandin analogues; RE – Right eye; LE – Left eye; No XFM - no exfoliative material seen on slit lamp examination, specifically in individuals with discordant eyes

No significant difference was observed between the distribution of males and females in the control and case groups ($p = 0.22$) (Table 3.5). Additionally, no significant difference in the ages between all participants in the case and control groups was observed ($p = 0.24$) (Table 3.5). The mean age of the males was 6.3 years less than females, on average, for the cases.

Table 3.5. Breakdown of sex and age of participants in the transcriptome study

	Cases (n=20)	**Controls (n=15)**	**p-value**
Female (percentage)	8 (40.0%)	10 (66.7%)	0.22
Male (percentage)	12 (60.0%)	5 (33.3%)	
Mean age (age±SD)	69.9 ± 11.5	68.9 ± 10.1	0.24
Mean Female age (age±SD)	70.2 ± 10.1	72.4 ± 13.0	0.28
Mean Male age (age±SD)	63.9 ± 11.0	68.8 ± 12.6	0.37

n – sample number; SD – standard deviation.

3.2.1.3. Library preparation and quality

Each sample was tagged with a unique nucleotide 'code', enabling disaggregation of the data according to each sample after analyses were completed on the pooled samples. The libraries were constructed from the cDNA and multiple samples were included in each pool. Twelve samples were included in both pools 1 and 2, while eight samples were included in both pools 3 and 4. The average amount of RNA for the RNA-seq libraries ranged from 1341 ng to 2329 ng for the four pools (Table 3.6). Analysis of the pools on the Agilent Bioanalyzer 2100 indicated a wide range of average fragment sizes from 385bp to 876bp, larger than the expected size of ~260bp (Figure 3.1).

Table 3.6. Library concentrations and yields per pool

Pool*	**Concentration of undiluted library (nM)**	**Concentration of undiluted library (ng/µL)**	**Library volume (µL)**	**Available amount of library (ng)**
Pool 1 (n = 12)	349	75	25	1888
Pool 2 (n = 12)	430	93	25	2329
Pool 3 (n = 8)	250	54	25	1354
Pool 4 (n = 8)	248	53	25	1341

*Pool – each pool includes several of the samples combined into one library. n – sample number.

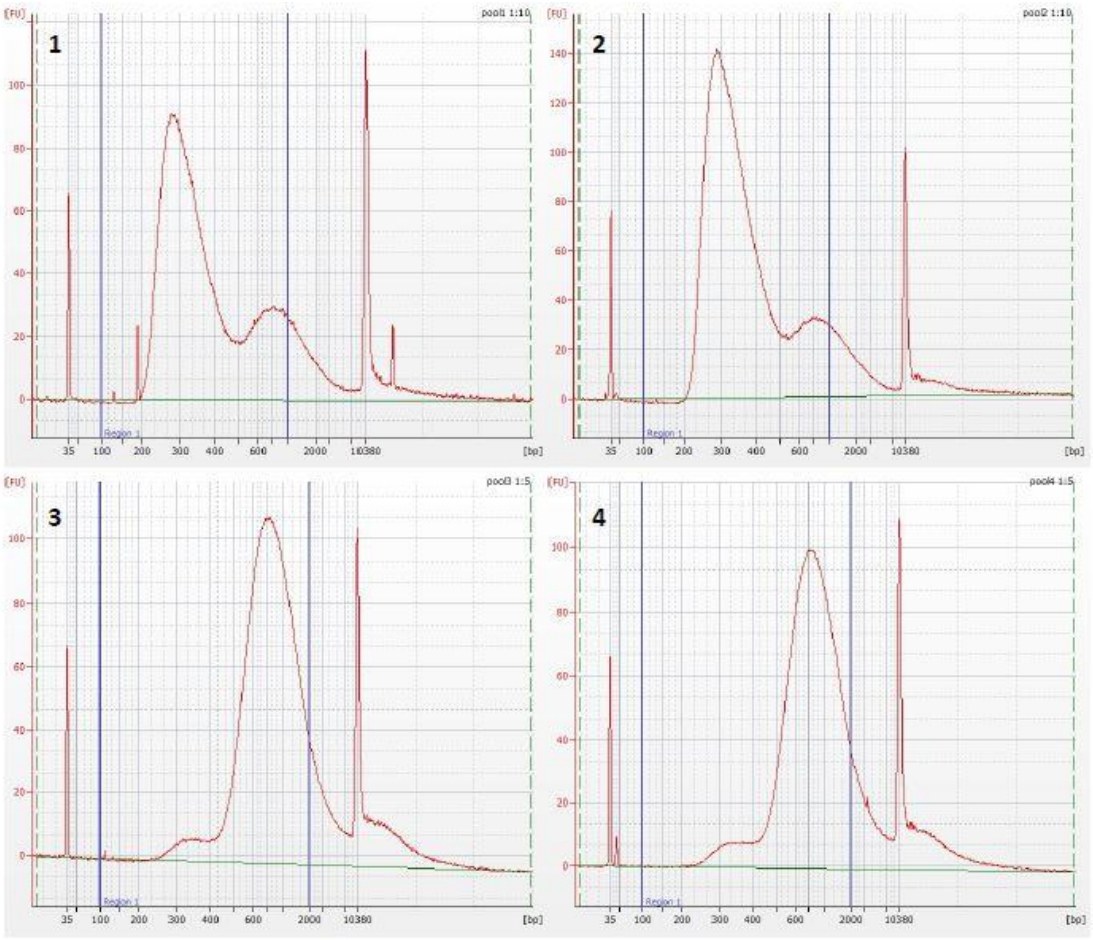


Figure 3.1. BioAnalyser traces showing fragment size distributions. 1) Pool 1, 2) Pool 2, 3) Pool 3, 4) Pool 4. While pools 1 and 2 have more reads of the expected size (~250bp-290bp), the bimodal distribution indicates some larger read sizes. Pools 3 and 4 are more unimodal, but the libraries consist of larger than expected fragment sizes (>600bp). Library construction for Pools 1 and 2 was performed at the same time, while Pools 3 and 4 were constructed in a separate experiment. The x-axis shows read size in base pairs (bp), while the y-axis is the fluorescent units (FU).

The additional purification step enabled sub-selection for shorter fragments through denaturation and bead purification and resulted in denatured libraries reaching approximately 350 bp for each pool (Figure 3.2).

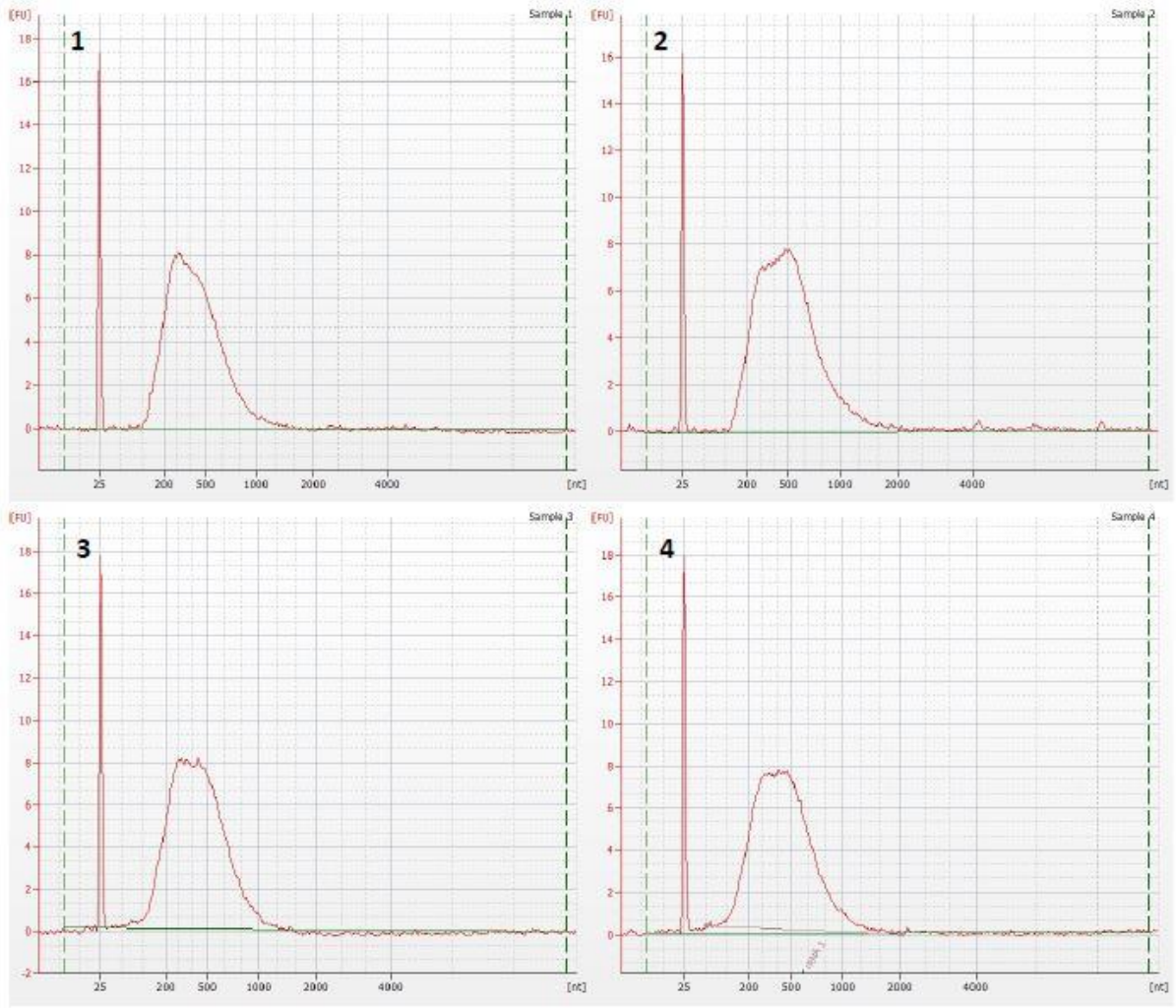


Figure 3.2. BioAnalyser traces showing fragment size distributions following additional quality control performed by CPGR. 1) Pool 1, 2) Pool 2, 3) Pool 3, 4) Pool 4. Following the denaturation step, all four pools show a unimodal distribution with fragment sizes approaching the desired length of 260bp. The x-axis shows read size in base pairs (bp), while the y-axis is the fluorescent units (FU).

3.2.1.4. RNA-sequencing quality control output

Quality Metrics obtained from Illumina's Sequence Analysis viewer were used to assess the success of the runs. The cluster densities of three of the pools; pool 1 (173.5 K/mm^2), 3 (193.8) and 4 (183.0) fell within the optimal range, while pool 2 (155.8) was slightly under-clustered. The average Q-score greater than or equal to 30 was 93.6% (61.0 Gbp) of bases called for pool 1, 92.3% (54.1 Gbp) for pool 2, 93.0% (66.8 Gbp) for pool 3 and 93.8% (64.2 Gbp) for pool 4.

FastQC was used as an independent tool to Illumina's Sequence Analysis viewer, and assessed the quality of each individual sample. All the sequenced samples passed this quality assessment and have quality scores above 30 (Figure 3.3) across the read length. The greatest N percentage

across the samples in this study was 0.11%, indicating very few ambiguous calls. The range in percentage of unique reads compared to duplicate reads varied widely across the samples, from 30.7% to 48.9% unique reads. The total number of unique reads also varied across the samples (lowest read count of 7 726 036 and highest read count of 22 435 408).

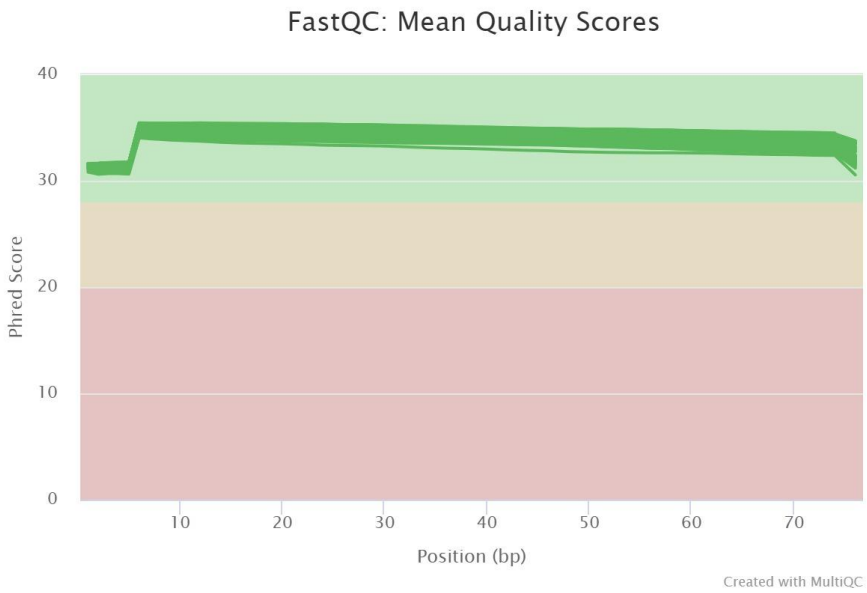


Figure 3.3. Mean quality scores (Phred scores) for each sample from FastQC. This figure illustrates the average Phred score for each of the samples sequenced. It provides a score across the length of each read (~75 bases due to the 2x75bp paired end sequencing approach). While the scores are consistently above 30, there is slight score reduction at the beginning and ends of the reads. Samples are represented by the green lines.

3.2.1.5. RNA-seq: Read count and alignment scores

Following quality control, STAR was used for read alignment. The percentage of uniquely mapped reads ranged from 83.5% to 93.3% between samples (Appendix E), with the remaining reads either mapping to multiple loci or unmapped due to short length.

HTSeq and featureCounts were then used to count the reads mapped to features. The range in percentage of assigned reads counted for HTSeq was 83.1% to 88.3% (Figure 3.4), with the remaining reads generally falling into the ambiguous, not unique or mapped to no feature categories. The range in percentage of assigned reads counted for featureCounts was 43% to 56.9%, decreased in comparison to HTSeq with the remaining reads generally falling into the unassigned category according to fragment length or due to multi-mapping (Appendix F).

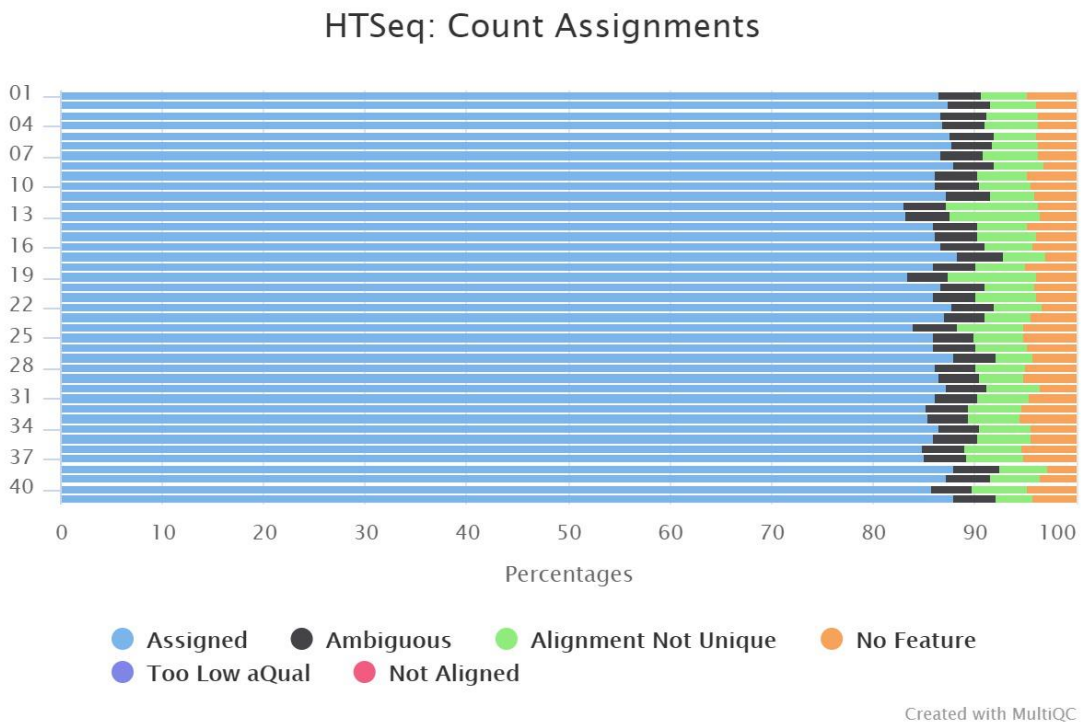


Figure 3.4. Per sample percentage of mapped reads assigned with different levels of certainty according to HTSeq. Each horizontal bar represents a sample. Over 80% of reads were assigned to a gene feature.

No samples were excluded based on the quality metrics, as all samples were of sufficient quality to proceed with the downstream analyses.

3.2.2. Differential Gene Expression Analyses

The different comparisons for DGE analyses are outlined in Figure 3.5. All DGE analyses were performed using DESeq2. The parameters inputted for DESeq2 for all comparisons were as follows: FDR-adjusted p-value < 0.05 and Log2 Fold Change threshold of ±0.5 (~1.5-fold change), adjusted for sex and age. Only genes achieving significance for differential expression with these parameters are shown in the results.

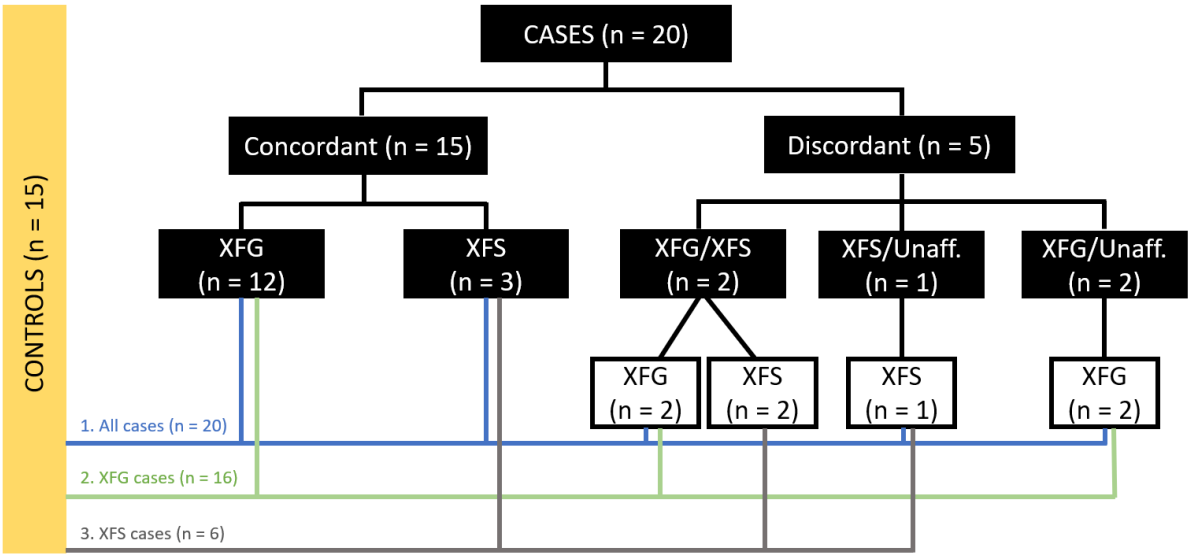


Figure 3.5. Workflow of the different comparisons for differential gene expression analysis. The 20 cases are divided into concordant or discordant cases, and then according to diagnosis (XFG, XFS or unaffected). The number of samples for each group are shown. Three comparisons are shown in the figure, with all cases (blue) compared to the control group; XFG cases (green) compared to the control group; and XFS cases (grey) compared to the control group.

3.2.2.1. XFG/XFS (Case) vs control DGE analysis

The final count for the main comparison included 20 cases (one affected eye per individual) and 15 controls (unaffected individuals) (Table 3.3). Four cases presented with exfoliation syndrome and the remaining 16 presented with exfoliation glaucoma, all but two of whom were undergoing chronic topical treatment. For this comparison, the XFS-affected eye (where the other eye was normal) or the XFG-affected eye from the XFS/XFG discordant cases was included, while the sample from the other eye was excluded to reduce any bias from intra-individual samples.

For the case and control comparison, 81 genes were found to be differentially expressed (Appendix G). Only one gene, SSX family member 1 (*SSX1*), was identified as down-regulated, while the remaining 80 were up-regulated (Figure 3.6).

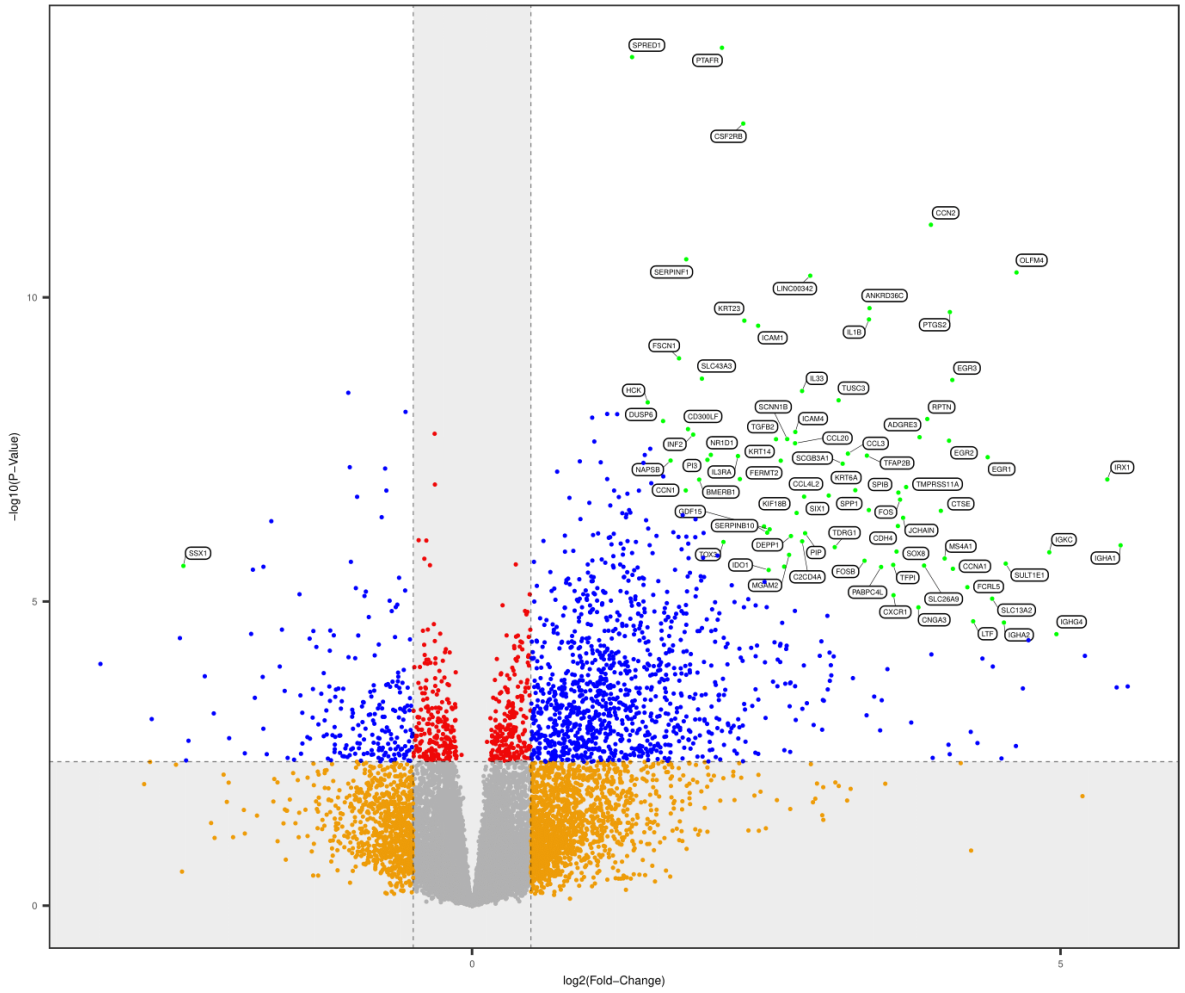


Figure 3.6. Volcano plot showing significantly up-regulated or down-regulated genes identified in the case and control comparison. Genes are mapped according to their statistical significance (p-value) and the magnitude of change in gene expression (Log2FC). Green: Significant genes after filtering; blue: genes that met both the Log2FC and significance thresholds; red: genes that met the significance threshold only; orange: genes that met the Log2FC threshold only; and grey: genes that did not meet the Log2FC and significance thresholds. Log2FC of ±0.5 (~1.5-fold change); FDR-adjusted p-value < 0.05.

A PCA of normalised RNA-seq read counts was also performed (Figure 3.7). The PCA plot illustrates the clustering of samples based on the similarity of the normalised read counts for the DEGs. PC1 accounted for 22% of variance and appeared to separate the samples according to whether they were cases or controls. The controls cluster more tightly according to PC1, while the cases are spread across the figure. PC2 accounted for 13% of variance and separated the samples according to sex.

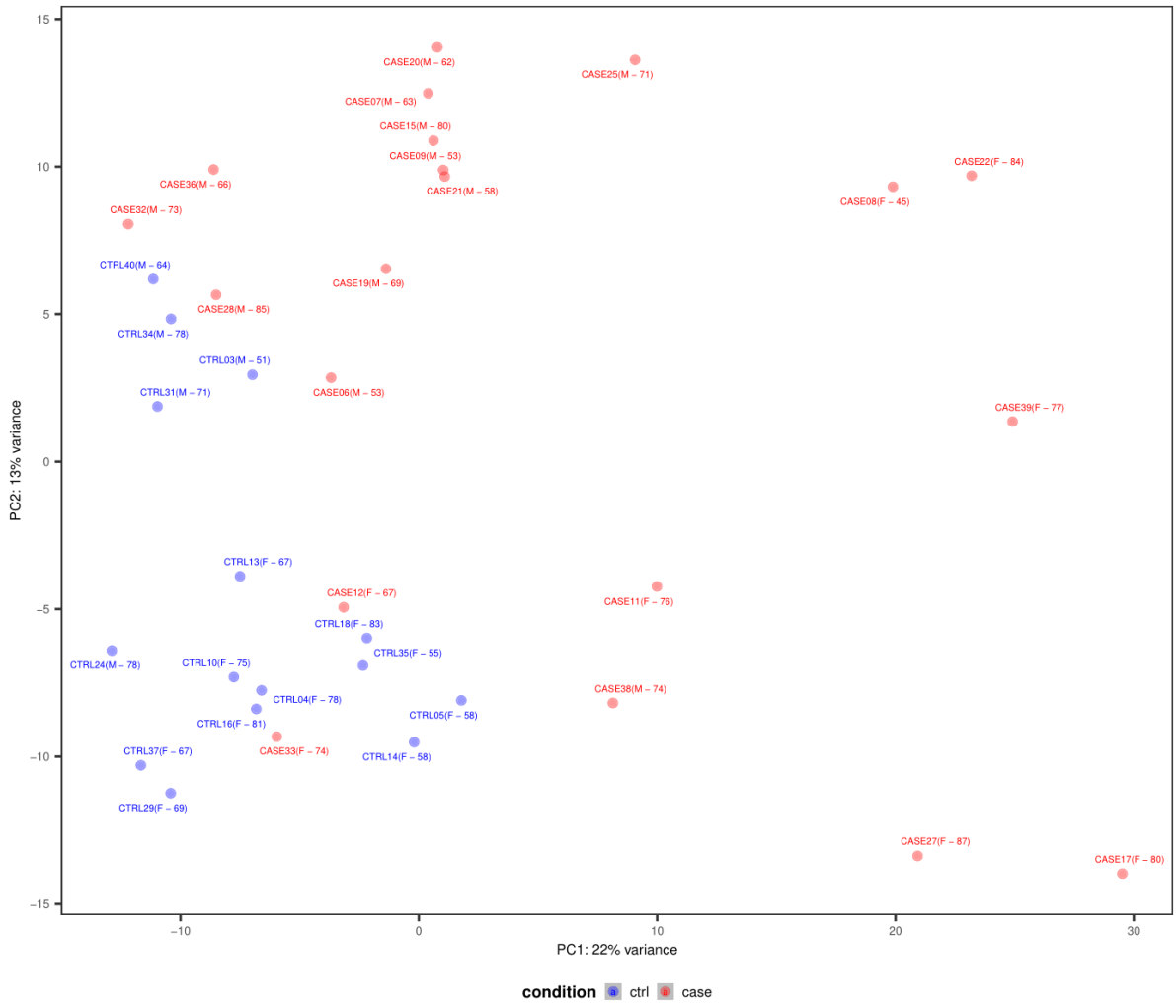


Figure 3.7. PCA plot showing clustering of the case and control samples. Cases are in red, while control samples (ctrl) are shown in blue. Both age and sex are indicated in the brackets in the labelling of each participant. For each group, sample to sample distances (within-and between-case or control status) are illustrated on the first two principal components. PC1 accounts for 22% of variance, while PC2 accounts for 13% of variance.

A heatmap of all 81 genes is presented in Figure 3.8, illustrating two main clusters. At the highest level, there is one cluster composed solely of cases, while the second cluster is composed of both cases and controls. The top 10 DEGs, based on p-value, are shown in Table 3.7, while the rest are listed in the Appendix (Appendix G).

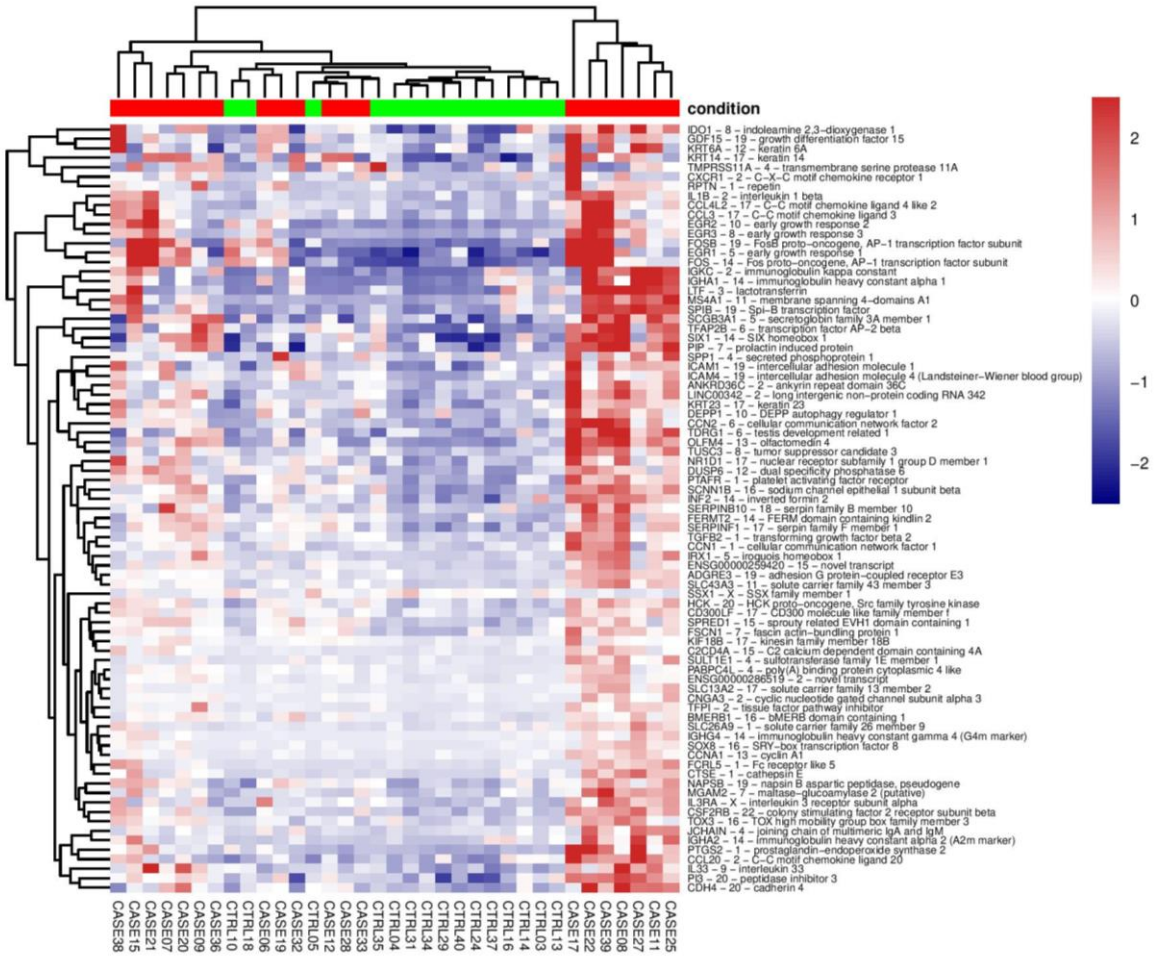


Figure 3.8. Heatmap for data visualisation of gene clustering according to the gene expression signals in the case control comparison. Each cell of the matrix represents the gene expression signal, from down-regulated (navy) to no change (white) to up-regulated (red). Control samples are shown in green, while cases are red (condition).

Table 3.7. Top 10 DEGs identified in the case and control comparison.

Gene ID	Description	Chr	Log2FC	P-adj
OLFM4	olfactomedin 4	13	4.63	2.60E-05
CCN2	cellular communication network factor 2	6	3.9	2.60E-05
PTAFR	platelet activating factor receptor	1	2.12	2.60E-05
CSF2RB	colony stimulating factor 2 receptor subunit beta	22	2.3	3.59E-05
PTGS2	prostaglandin-endoperoxide synthase 2	1	4.06	9.05E-05
ANKRD36C	ankyrin repeat domain 36C	2	3.38	1.55E-04
LINC00342	long intergenic non-protein coding RNA 342	2	2.87	1.55E-04
IL1B	interleukin 1 beta	2	3.37	1.72E-04
EGR3	early growth response 3	8	4.08	3.61E-04
ICAM1	intercellular adhesion molecule 1	19	2.43	1.13E-03

Chr - chromosome, Log2FC - Log2 fold change, P-adj - FDR-adjusted p-value. Log2FC of ±0.5 (~1.5-fold change); FDR-adjusted p-value < 0.05.

Using the raw read-count values for 81 identified DEGs, gene-set enrichment was completed. The databases selected for the annotation of the 81 genes were the GO databases (Molecular Function 2018, Biological Process 2018 and Cellular Component 2018). Plots of significant terms were then created for each database (Figure 3.9) and are further shown in Table 3.8 Multiple Biological Process and Molecular Function GO terms were identified based on the DEGs, however only one Cellular Component GO term, “Tertiary granule (GO:0070820)”, was found to be enriched.

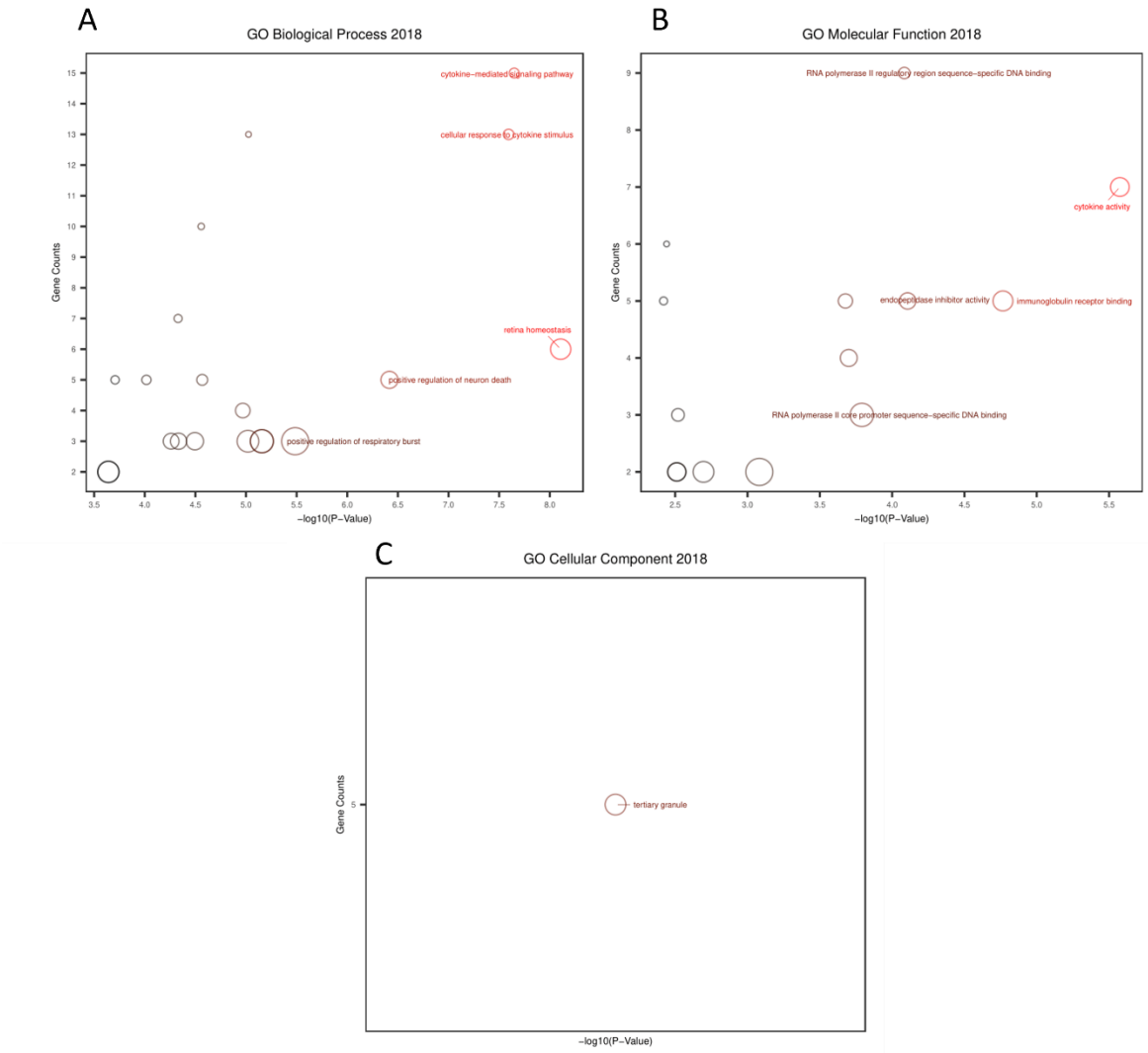


Figure 3.9. Significance plots for the GO terms in the gene-set enrichment analysis of the 81 DEGs for the case and control comparison. Gene-set enrichment was completed using the GO databases. A) GO Biological Process 2018, B) GO Molecular Function 2018, C) GO Cellular Component. Each plot shows the number of genes per pathway on the y-axis and the -log10(p-value) on the x-axis. The colours associated with each term represent the significance, from black (less significant) to red (highly significant). The number of genes for each term is indicated by the size of the circle. Only the top 5 terms are shown for each plot.

Table 3.8. GO terms and genes identified based on the 81 DEGs from the case-control comparison

GO Terms	Genes	P-adj
GO Biological Processes		
Retina homeostasis (GO:0001895)	*IGKC; PIP; IGHA1; IGHA2; JCHAIN; LTF*	7.81E-09
Cytokine-mediated Signalling pathway (GO:0019221)	*EGR1; IL33; CCL20; PTAFR; CSF2RB; FOS; PTGS2; ICAM1; IGHG4; HCK; IL1B; IL3RA; FSCN1; CCL3; CD300LF*	2.25E-08
Cellular response to cytokine stimulus (GO:0071345)	*EGR1; CCL20; PTAFR; CSF2RB; FOS; PTGS2; ICAM1; IGHG4; HCK; IL1B; IL3RA; FSCN1; CCL3*	2.55E-08
Positive regulation of neuron death (GO:1901216)	*EGR1; TFAP2B; TGFB2; CCL3; FOS*	3.83E-07
Positive regulation of respiratory burst (GO:0060267)	*IGHA1; IGHA2; JCHAIN*	3.27E-06
GO Molecular Function		
Cytokine activity (GO:0005125)	*IL33; TGFB2; CCL20; GDF15; IL1B; CCL3; SCGB3A1*	2.66E-06
Immunoglobulin receptor binding (GO:0034987)	*IGHG4; IGKC; IGHA1; IGHA2; JCHAIN*	1.71E-05
Endopeptidase inhibitor activity (GO:0004866)	*SERPINB10; SERPINF1; PI3; TFPI; LTF*	7.82E-05
RNA polymerase II regulatory region sequence-specific DNA binding (GO:0000977)	*SPIB; TFAP2B; EGR1; EGR2; SIX1; FOSB; SOX8; FOS; NR1D1*	8.24E-05
RNA polymerase II core promoter sequence-specific DNA binding (GO:0000979)	*TFAP2B; SOX8; FOS*	0.0002
GO Cellular Component		
Tertiary granule (GO:0070820)	*SERPINB10; ADGRE3; PTAFR; OLFM4; LTF*	0.0005

P-adj- FDR-adjusted p-value

Gene-set enrichment was also completed for three pathway databases, WikiPathways 2016, KEGG 2016 and Reactome 2016. Plots of significant terms were then created for each pathway

(Figure 3.10) and are further shown in Table 3.9. Significantly enriched pathways were identified by all three databases.

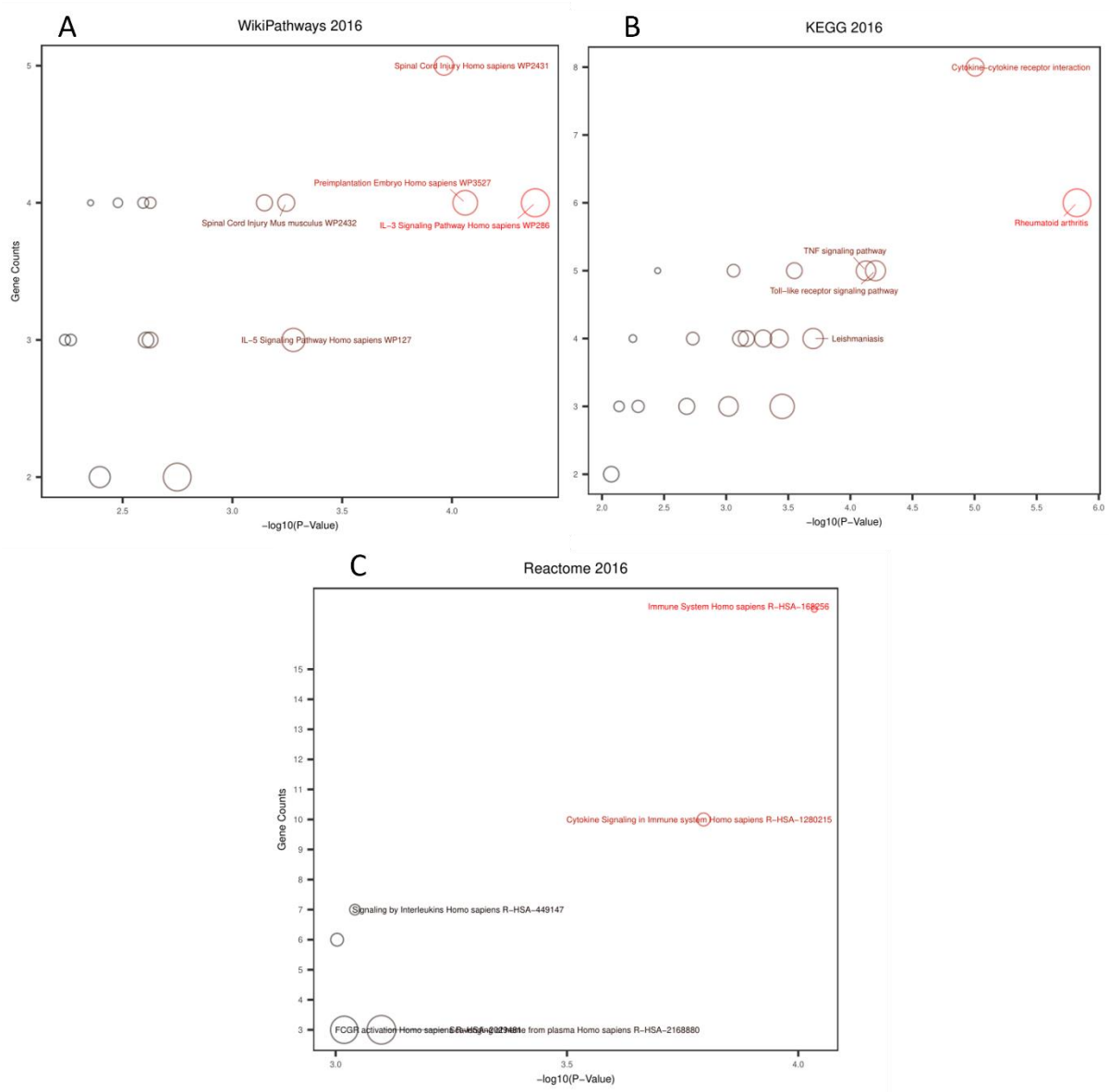


Figure 3.10. Significance plots for the pathways in the gene-set enrichment analysis of the 81 DEGs for the case and control comparison. Gene-set enrichment was completed using three pathways. A) WikiPathways 2016, B) KEGG 2016, C) Reactome 2016. Each plot shows the number of genes per pathway on the y-axis and the -log10(p-value) on the x-axis. The colours associated with each term represent the significance, from black (less significant) to red (highly significant). The number of genes for each term is indicated by the size of the circle. Only the top 5 terms are shown for each plot.

Table 3.9. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 81 DEGs from the case-control comparison

Pathways	**Genes**	**P-adj**
WikiPathways 2016		
IL-3 Signalling Pathway Homo sapiens WP286	*HCK; IL3RA; CSF2RB; FOS*	0.005
Preimplantation Embryo Homo sapiens WP3527	*TFAP2B; EGR1; FOSB; SOX8*	0.005
Spinal Cord Injury Homo sapiens WP2431	*EGR1; IL1B; FOS; PTGS2; ICAM1*	0.005
IL-5 Signalling Pathway Homo sapiens WP127	*SPRED1; CSF2RB; FOS*	0.015
KEGG 2016		
Rheumatoid arthritis Homo sapiens hsa05323	*TGFB2; CCL20; IL1B; CCL3; FOS; ICAM1*	0.0002
Cytokine-cytokine receptor interaction Homo sapiens hsa04060	*TGFB2; CXCR1; CCL20; CCL4L2; IL1B; IL3RA; CCL3; CSF2RB*	0.0005
Toll-like receptor Signalling pathway Homo sapiens hsa04620	*CCL4L2; IL1B; SPP1; CCL3; FOS*	0.002
TNF Signalling pathway Homo sapiens hsa04668	*CCL20; IL1B; FOS; PTGS2; ICAM1*	0.002
Leishmaniasis Homo sapiens hsa05140	*TGFB2;IL1B;FOS;PTGS2*	0.004
Reactome 2016		
Immune System Homo sapiens R-HSA-168256	*EGR1; IL33; PTAFR; ICAM4; CSF2RB; FOS; DUSP6; ICAM1; IGHG4; HCK; SPRED1; IGKC; IL1B; IL3RA; CTSE; CD300LF; LTF*	0.0209
Cytokine Signalling in Immune system Homo sapiens R-HSA-1280215	*HCK; EGR1; IL33; SPRED1; IL1B; IL3RA; PTAFR; CSF2RB; DUSP6; ICAM1*	0.0209
Scavenging of heme from plasma Homo sapiens R-HSA-2168880	*IGKC; IGHA1; IGHA2*	0.0432
Signalling by Interleukins Homo sapiens R-HSA-449147	*HCK; IL33; SPRED1; IL1B; IL3RA; CSF2RB; DUSP6*	0.0432
FCGR activation Homo sapiens R-HSA-2029481	*IGHG4; HCK; IGKC*	0.0432

P-adj- FDR-adjusted p-value.

3.2.2.2. XFG cases vs control group DGE analysis

To account for any differences in gene expression between individuals with XFG and those with XFS, the cases were separated into individuals with XFG and XFS and then individually compared to the control groups. Of the 20 cases, 16 individuals presented with XFG and were used in this comparison.

For the XFG case and control comparison, 85 genes were found to be differentially expressed (Appendix H). Six genes were identified as down-regulated, while the remaining 79 were up-regulated (Figure 3.11).

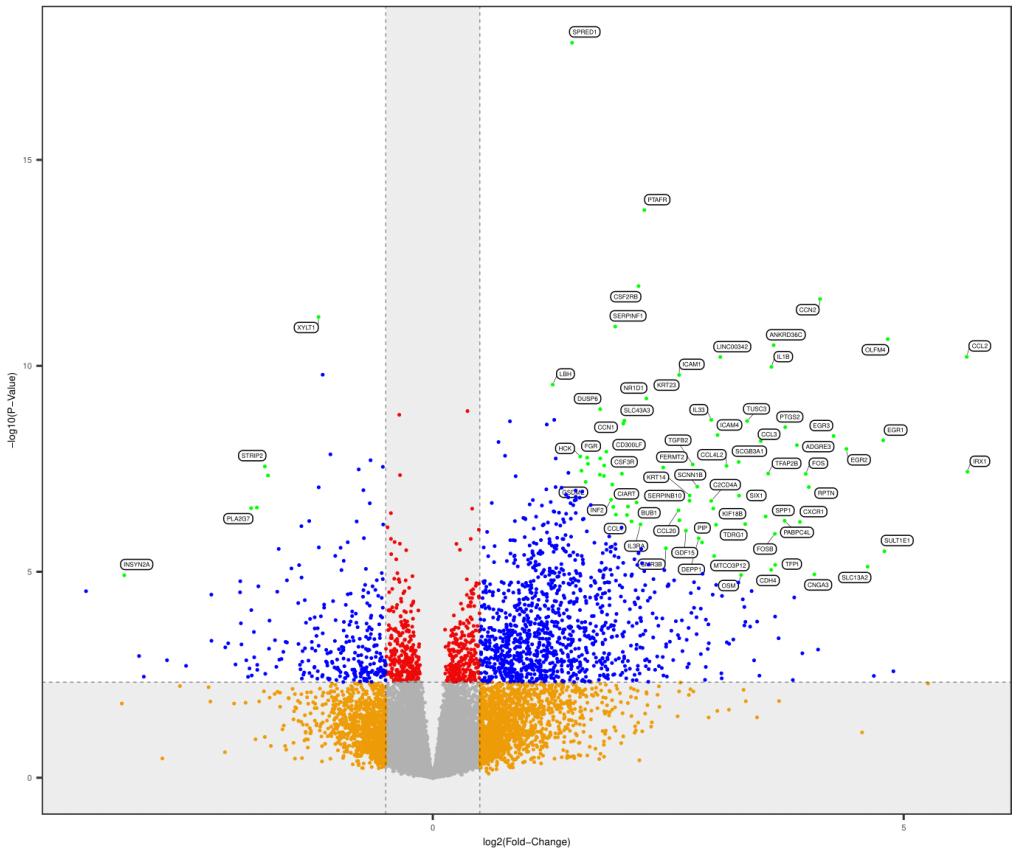


Figure 3.11. Volcano plot showing significantly up-regulated or down-regulated genes identified in the XFG and control comparison. Genes are mapped according to their statistical significance (p-value) and the magnitude of change in gene expression (Log2FC). Green: Significant genes after filtering; blue: genes that met both the Log2FC and significance thresholds; red: genes that met the significance threshold only; orange: genes that met the Log2FC threshold only; and grey: genes that did not meet the Log2FC and significance thresholds. Log2FC of ±0.5 (~1.5-fold change); FDR-adjusted p-value < 0.05.

The PCA for this comparison did not differ significantly from the case and control comparison, with the same percentage of variance explained by PC2, while PC1 accounted for 23% of variance (Figure 3.12).

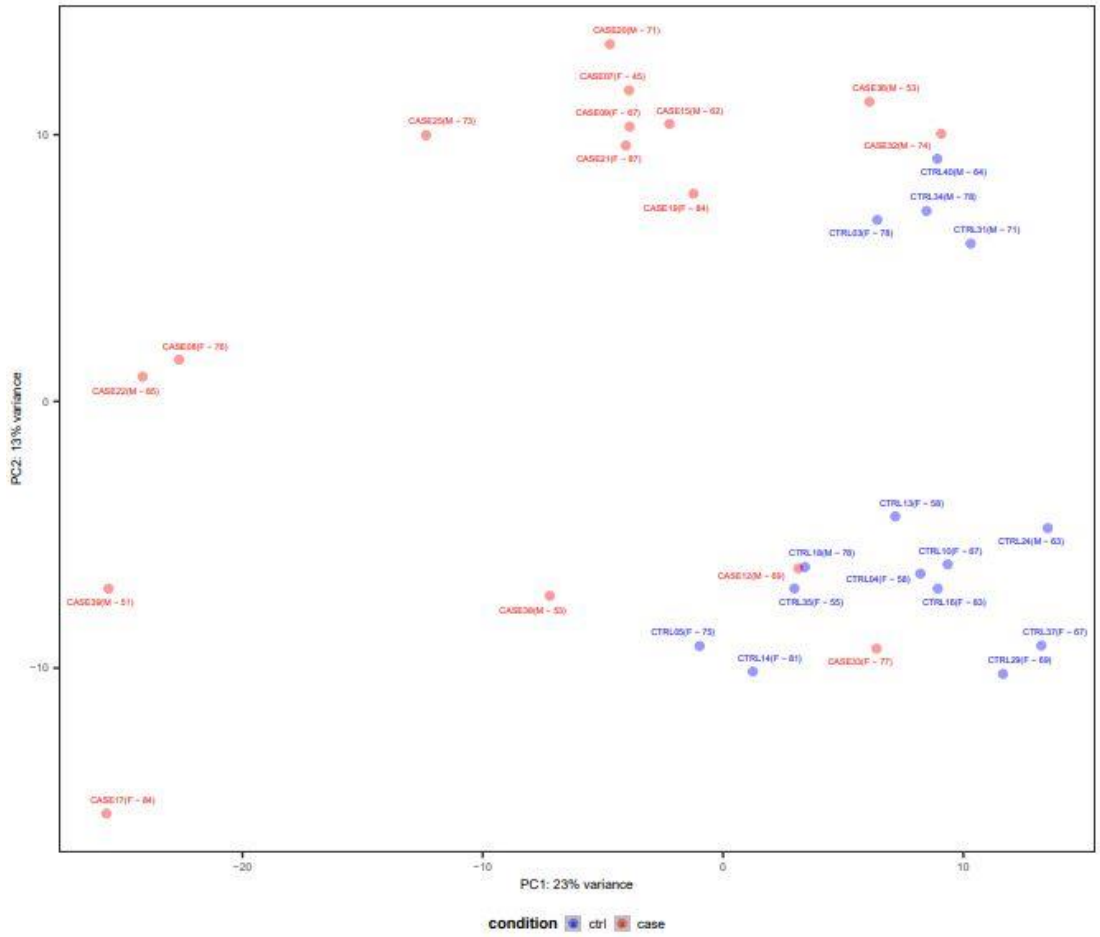


Figure 3.12. PCA plot showing clustering of the XFG-case and control samples. Cases are in red, while control samples (ctrl) are shown in blue. Both age and sex are indicated in the brackets in the labelling of each participant. For each group, sample to sample distances (within-and between-case or control status) are illustrated on the first two principal components. PC1 accounts for 23% of variance, while PC2 accounts for 13% of variance.

A heatmap was generated for all 85 genes (Figure 3.13). There are two main clusters, one composed solely of XFG-case samples, while the second cluster is further separated into a case-only cluster and the other mostly control samples, with four cases, CASE36, CASE33, CASE32 and CASE12, interspersed among the control samples. This heatmap also has additional labels at the top to explore potential confounding factors including of topical treatment, sex, and surgery. There is no distinct clustering pattern based on topical treatment among the cases, or with sex in the control group. However, the cases do demonstrate clustering according to sex,

with almost all male cases clustering into one high-level group. The two participants, both female, who underwent surgery also cluster together, in a high-level group that also includes another affected female.

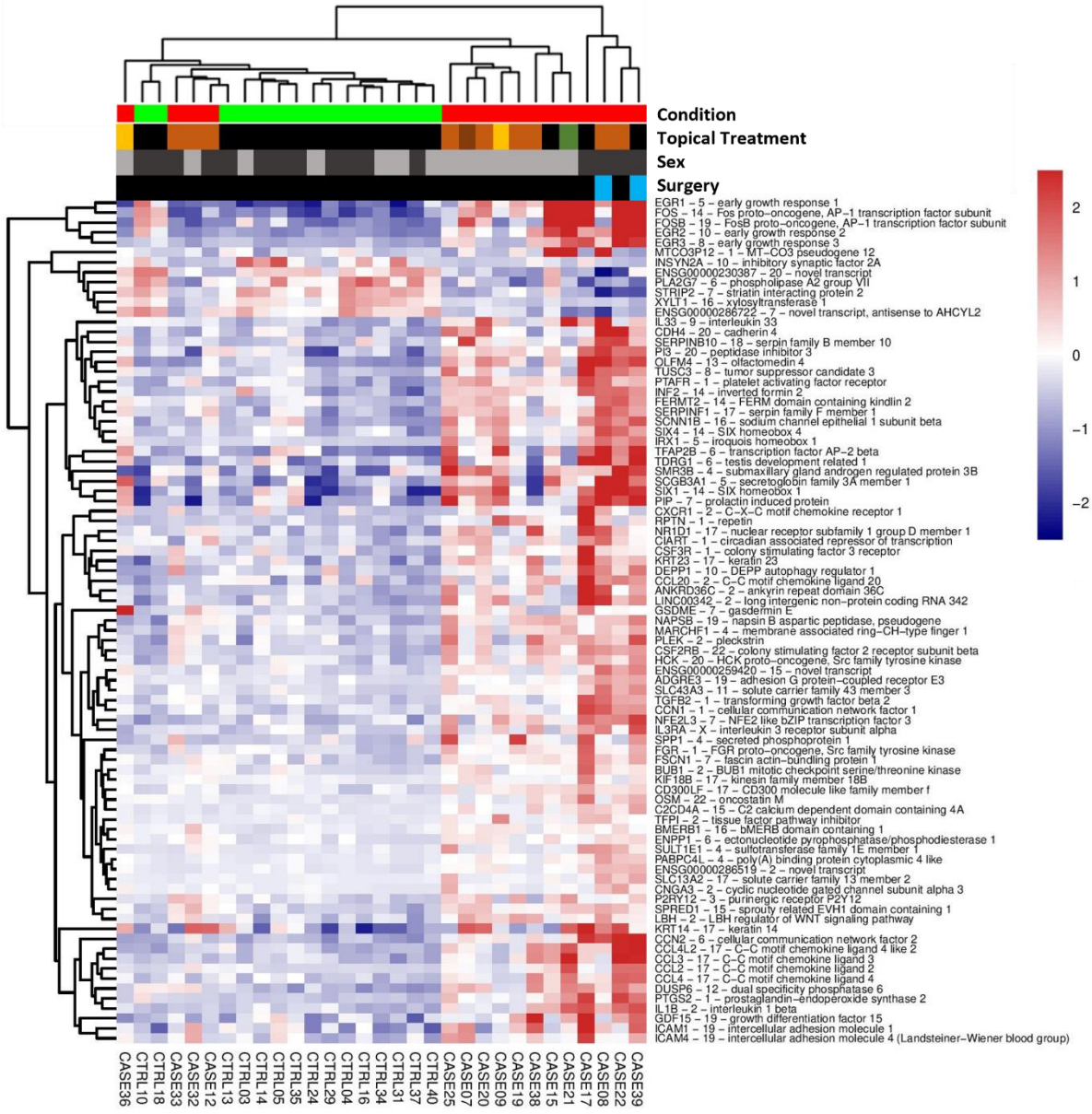


Figure 3.13. Heatmap for data visualisation of gene clustering according to the gene expression signals in the XFG control comparison. Each cell of the matrix represents the gene expression signal, from down-regulated (navy) to no change (white) to up-regulated (red). Control samples are shown in green, while cases are red (condition). Additional labels: Topical treatment: Alpha agonists only (Yellow), Prostaglandin analogues and Beta Blockers (Green), Alpha agonists, Prostaglandin analogues and Beta Blockers (Light Brown), Alpha agonists, Prostaglandin analogues, Beta Blockers and Carbonic Anhydrase Inhibitors (Dark Brown), No chronic topical treatment (Black); Sex: Male (Grey) and Female (Dark Grey); Surgery: Yes (Blue) and No (Black).

The top 10 DEGs are shown in Table 3.10, while the remaining genes are listed in the Appendix (Appendix). For this comparison, the expression of *OLFM4* was again identified as the most significantly altered. Seven of the top ten DEGs were also within the top ten DEGs identified in the case-control comparison, with only *CCL2*, *SPRED1* and *EGR1* novel to the top ten within this comparison.

Table 3.10. Top 10 DEGs identified in the XFG and control comparison.

Gene ID	Description	Chr	Log2FC	P-adj
OLFM4	olfactomedin 4	13	4.83	1.22E-05
CCL2	C-C motif chemokine ligand 2	17	5.67	1.22E-05
CCN2	cellular communication network factor 2	6	4.11	1.22E-05
PTAFR	platelet activating factor receptor	1	2.25	1.22E-05
SPRED1	sprouty related EVH1 domain containing 1	15	1.48	2.34E-05
ANKRD36C	ankyrin repeat domain 36C	2	3.62	3.52E-05
IL1B	interleukin 1 beta	2	3.6	7.75E-05
CSF2RB	colony stimulating factor 2 receptor subunit beta	22	2.18	1.01E-04
LINC00342	long intergenic non-protein coding RNA 342	2	3.05	1.01E-04
EGR1	early growth response 1	5	4.78	4.07E-04

Chr- chromosome, Log2FC- Log2 fold change, P-adj- adjusted p-value.

Underlined genes indicate those identified within the top ten DEGs in the case-control comparison. Log2FC of ±0.5 (~1.5-fold change); FDR-adjusted p-value < 0.05.

Plots of the significant terms were then created for each database for the XFG case and control comparison (Figure 3.14). The results from the GO databases are shown in Table 3.11. Biological Process and Molecular Function GO terms were identified based on the DEGs, however no Cellular Component GO terms were found to be enriched.

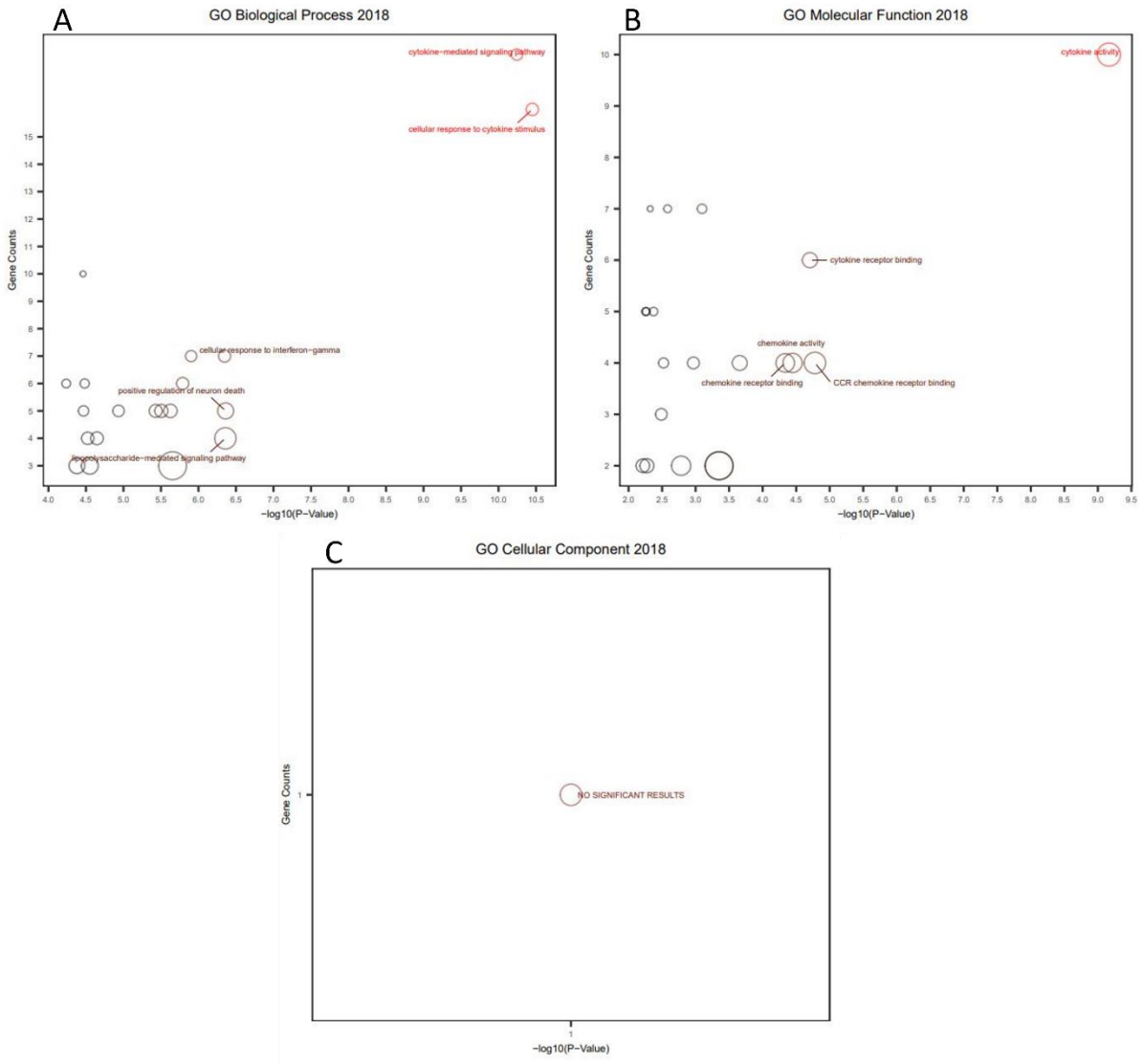


Figure 3.14. Significance plots for the GO terms in the gene-set enrichment analysis of the 85 DEGs for the XFG and control comparison. Gene-set enrichment was completed using the GO databases. A) GO Biological Process 2018, B) GO Molecular Function 2018, C) GO Cellular Component. Each plot shows the number of genes per pathway on the y-axis and the -log10(p-value) on the x-axis. The colours associated with each term represent the significance, from black (less significant) to red (highly significant). The number of genes for each term is indicated by the size of the circle. Only the top 5 terms are shown for each plot.

Table 3.11. GO terms and genes identified based on the 86 DEGs from the XFG case-control comparison

GO Terms	Genes	Adj. P
GO Biological Processes		
Cellular response to cytokine stimulus (GO:0071345)	*EGR1; CSF3R; CCL20; PTAFR; OSM; CSF2RB; FOS; PTGS2; ICAM1; HCK; IL1B; IL3RA; CCL4; FSCN1; CCL3; CCL2*	2.61E-08
Cytokine-mediated Signalling pathway (GO:0019221)	*EGR1; IL33; CSF3R; CCL20; PTAFR; OSM; CSF2RB; FOS; PTGS2; ICAM1; HCK; IL1B; IL3RA; CCL4; FSCN1; CCL3; CCL2; CD300LF*	2.61E-08
Positive regulation of neuron death (GO:1901216)	*EGR1; TFAP2B; TGFB2; CCL3; FOS*	8.24E-05
Lipopolysaccharide-mediated Signalling pathway (GO:0031663)	*HCK; IL1B; CCL3; CCL2*	8.24E-05
Cellular response to interferon-gamma (GO:0071346)	*HCK; CCL20; PTAFR; CCL4; CCL3; CCL2; ICAM1*	8.24E-05
GO Molecular Function		
Cytokine activity (GO:0005125)	*IL33; TGFB2; CCL20; GDF15; OSM; IL1B; CCL4; CCL3; CCL2; SCGB3A1*	1.10E-07
CCR chemokine receptor binding (GO:0048020)	*CCL20; CCL4; CCL3; CCL2*	0.0011
Cytokine receptor binding (GO:0005126)	*SPRED1; TGFB2; GDF15; IL1B; OSM; CD300LF*	0.0011
Chemokine activity (GO:0008009)	*CCL20; CCL4; CCL3; CCL2*	0.0014
Chemokine receptor binding (GO:0042379)	*CCL20; CCL4; CCL3; CCL2*	0.0015
GO Cellular Component		
No significant results	*N/A*	N/A

P-adj- FDR-adjusted p-value; N/A - not applicable (no genes).

Plots of the significant terms were then created for each pathway for the XFG case and control comparison (Figure 3.15). The results from the three pathway databases are shown in Table 3.12. Significantly enriched pathways were identified by all three databases.

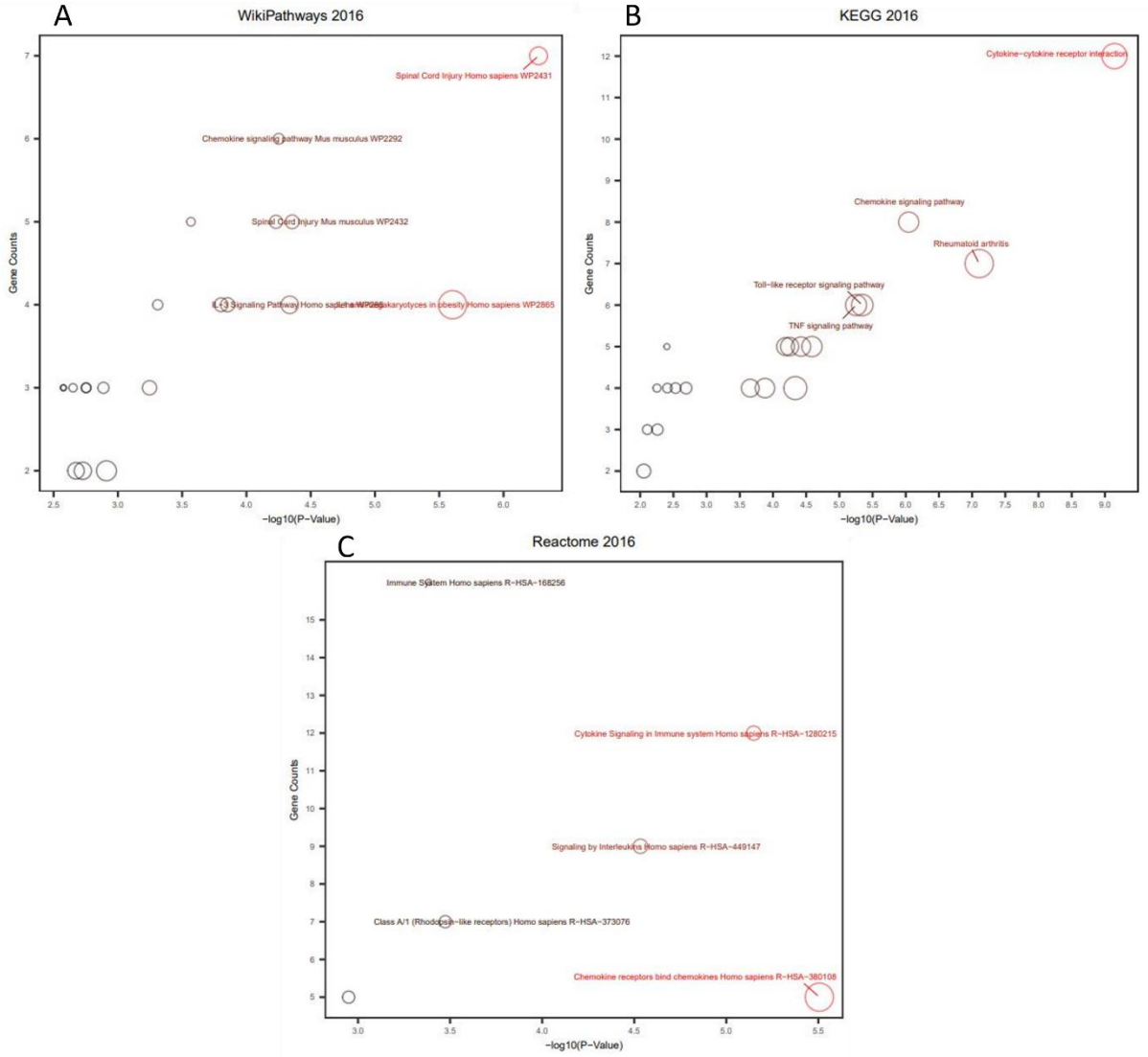


Figure 3.15. Significance plots for the pathways in the gene-set enrichment analysis of the 85 DEGs for the XFG and control comparison. Gene-set enrichment was completed using three pathways. A) WikiPathways 2016, B) KEGG 2016, C) Reactome 2016. Each plot shows the number of genes per pathway on the y-axis and the -log10(p-value) on the x-axis. The colours associated with each term represent the significance, from black (less significant) to red (highly significant). The number of genes for each term is indicated by the size of the circle. Only the top 5 terms are shown for each plot.

Table 3.12. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 86 DEGs from the XFG case-control comparison

Pathways	Genes	P-Adj.
WikiPathways 2016		
Spinal Cord Injury Homo sapiens WP2431	*EGR1; IL1B; XYLT1; CCL2; FOS; PTGS2; ICAM1*	6.94E-05
IL1 and megakaryotyces in obesity Homo sapiens WP2865	*IL1B; CCL2; PLA2G7; ICAM1*	0.0002
IL-3 Signalling Pathway Homo sapiens WP286	*HCK; IL3RA; CSF2RB; FOS*	0.0013
Toll-like Receptor Signalling Pathway Homo sapiens WP75	*IL1B; CCL4; SPP1; CCL3; FOS*	0.0013
Oncostatin M Signalling Pathway Homo sapiens WP2374	*EGR1; OSM; CCL2; FOS*	0.0026
KEGG 2016		
Cytokine-cytokine receptor interaction Homo sapiens hsa04060	*TGFB2; CSF3R; CXCR1; CCL20; CCL4L2; IL1B; IL3RA; CCL4; CCL3; OSM; CCL2; CSF2RB*	7.89E-08
Rheumatoid arthritis Homo sapiens hsa05323	*TGFB2; CCL20; IL1B; CCL3; CCL2; FOS; ICAM1*	4.30E-06
Chemokine Signalling pathway Homo sapiens hsa04062	*FGR; HCK; CXCR1; CCL20; CCL4L2; CCL4; CCL3; CCL2*	3.31E-05
Toll-like receptor Signalling pathway Homo sapiens hsa04620	*CCL4L2; IL1B; CCL4; SPP1; CCL3; FOS*	0.0001
TNF Signalling pathway Homo sapiens hsa04668	*CCL20; IL1B; CCL2; FOS; PTGS2; ICAM1*	0.0001
Reactome 2016		
Chemokine receptors bind chemokines Homo sapiens R-HSA-380108	*CXCR1; CCL20; CCL3; CCL4; CCL2*	0.0007
Cytokine Signalling in Immune system Homo sapiens R-HSA-1280215	*HCK; EGR1; IL33; SPRED1; OSM; CSF3R; IL1B; IL3RA; PTAFR; CSF2RB; DUSP6; ICAM1*	0.0008
Signalling by Interleukins Homo sapiens R-HSA-449147	*HCK; IL33; SPRED1; OSM; CSF3R; IL1B; IL3RA; CSF2RB; DUSP6*	0.0021
Class A/1 (Rhodopsin-like receptors) Homo sapiens R-HSA-373076	*P2RY12; CXCR1; CCL20; PTAFR; CCL4; CCL3; CCL2*	0.0182
Immune System Homo sapiens R-HSA-168256	*EGR1; IL33; CSF3R; PTAFR; ICAM4; CSF2RB; FOS; DUSP6; ICAM1; FGR; HCK; SPRED1; IL1B; IL3RA; CTSE; CD300LF*	0.0182

P-adj- FDR-adjusted p-value.

There was a high overlap in the GO terms and pathways identified for both the case-control comparison and the XFG-case and control comparison, which is expected given some similarity in the identified genes. However, the associated genes differed for these terms and pathways.

3.2.2.3. XFS cases vs control group DGE analysis

An additional comparison of the XFS cases to the control group was carried out. This comparison may highlight genes or pathways that are solely linked to XFS rather than the progression to XFG. However, this comparison does lose power due to only four of the cases presenting with XFS only. Therefore, this comparison included the XFS sample taken from two individuals who presented with XFG in one eye and XFS in the second eye. Therefore, six samples were used in this XFS case group to compare to the control group.

For the XFS case and control comparison, 36 genes were found to be differentially expressed, with the top ten shown in Table 3.13 and the remaining in the appendix (Appendix I). Only one DEG was found to be down-regulated, with 35 up-regulated DEGs (Figure 3.16).

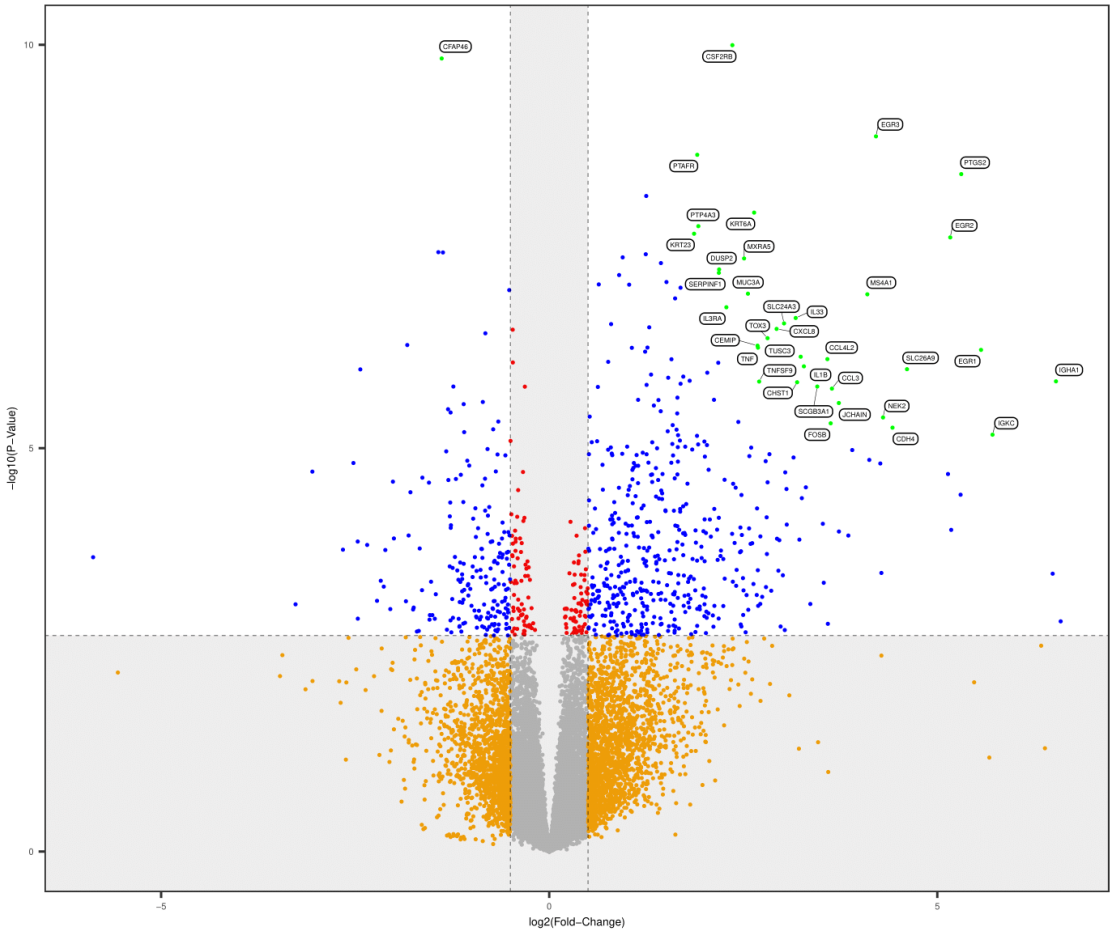


Figure 3.16. Volcano plot showing significantly up-regulated or down-regulated genes identified in the XFS and control comparison. Volcano plot showing significantly up-regulated or down-regulated genes identified in the XFS and control comparison. Genes are mapped according to their statistical significance (p-value) and the magnitude of change in gene expression (Log2FC). Green: Significant genes after filtering; blue: genes that met both the Log2FC and significance thresholds; red: genes that met the significance threshold only; orange: genes that met the Log2FC threshold only; and grey: genes that did not meet the Log2FC and significance thresholds. Log2FC of ±0.5 (~1.5-fold change); FDR-adjusted p-value < 0.05.

In the PCA plot, it is observed that most of the controls are again clustered more tightly together, while the XFS cases show more spread (Figure 3.17). However, there is clearer separation of the XFS cases and controls by PC2 compared to the previous comparisons. PC2 accounts for 16% of the variance, while PC1 accounts for 23% of the variance.

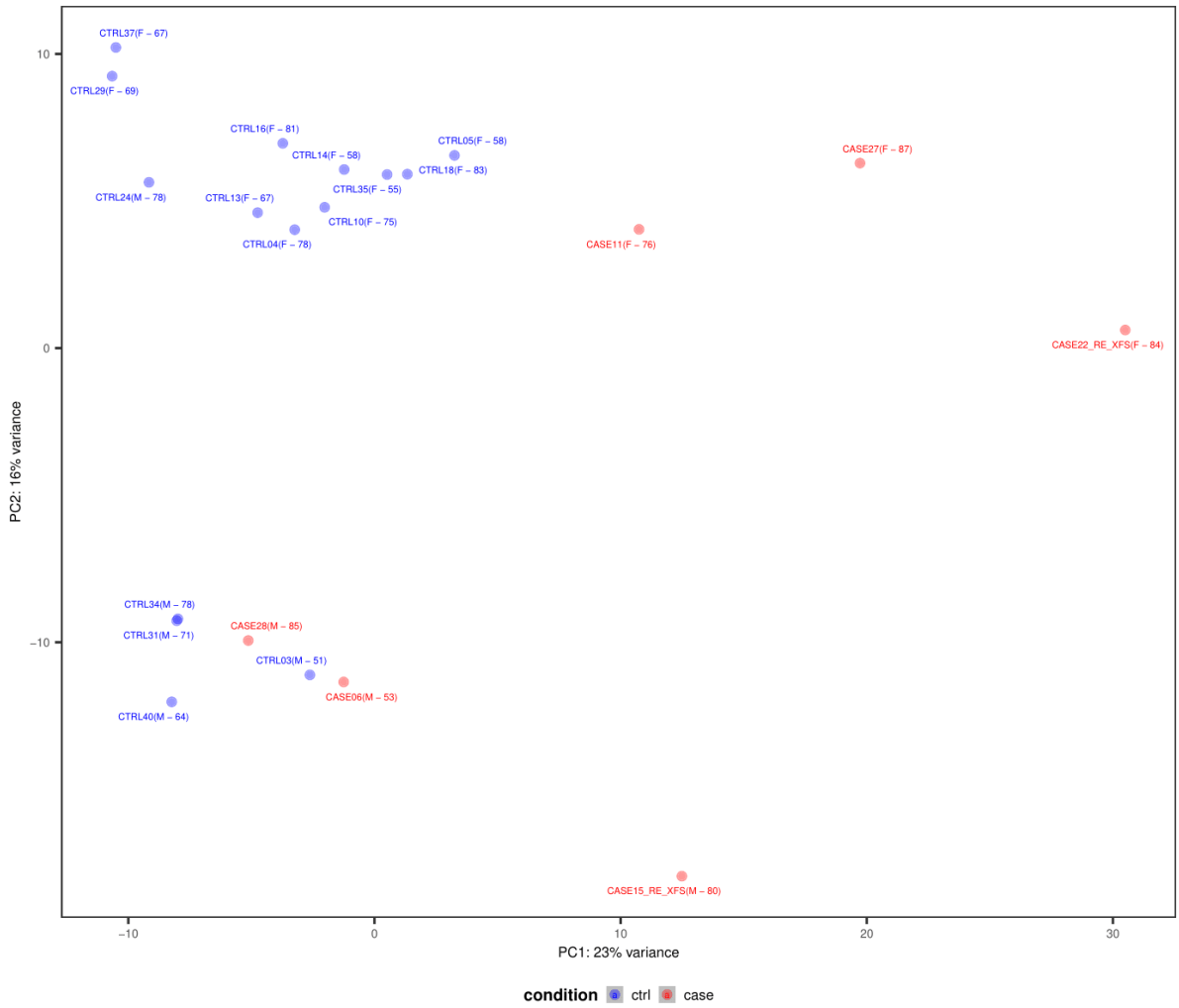


Figure 3.17. PCA plot showing clustering of the XFS and control samples. Cases are in red, while control samples (ctrl) are shown in blue. Both age and sex are indicated in the brackets in the labelling of each participant. For each group, sample to sample distances (within-and between-case or control status) are illustrated on the first two principal components. PC1 accounts for 23% of variance, while PC2 accounts for 16% of variance.

A heatmap was generated for all 36 genes (Figure 3.18). There are three main clusters, with the two high-level clusters solely including cases. The highest split is for the two samples from the discordant individuals, CASE15_RE_XFS and CASE22_RE_XFS. The third main cluster contains all the controls and only two cases, CASE28 and CASE06. CASE06 is 53 years of age, young in comparison to the mean age of 69.9 ± 11.5, therefore disease progression may not be as advanced as others potentially reflected in the underlying gene expression, however the same is not true for CASE28, who is 85 years of age.

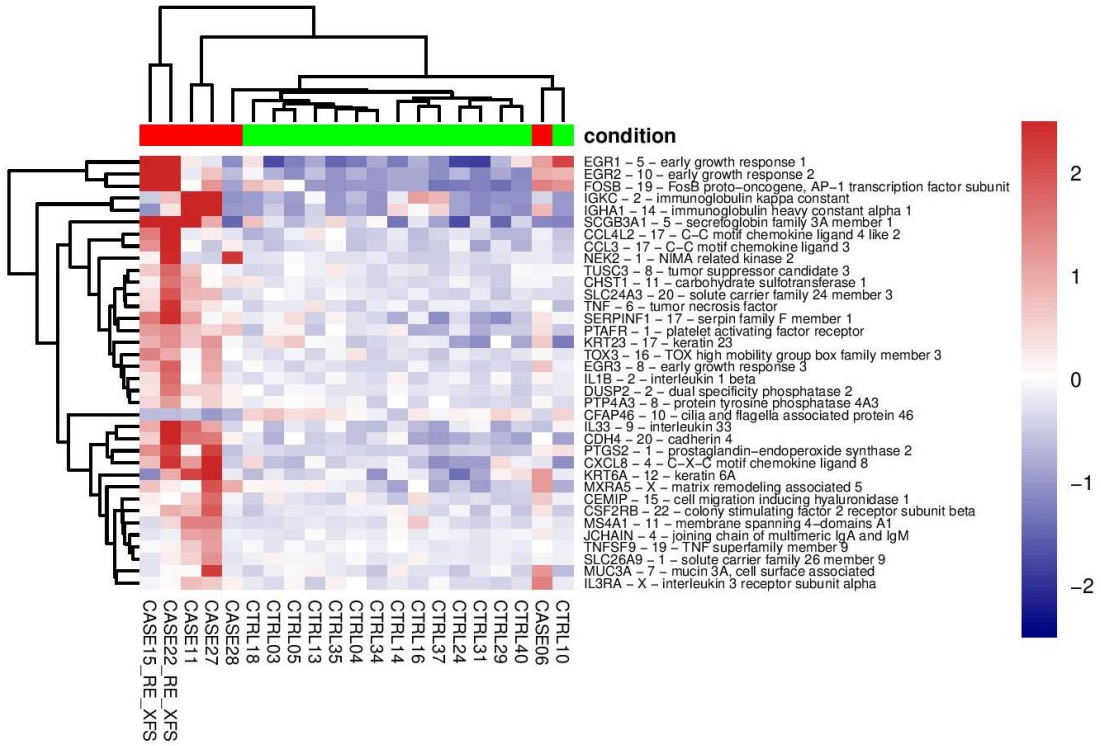


Figure 3.18. Heatmap for data visualisation of gene clustering according to the gene expression signals in the XFS control comparison. Each cell of the matrix represents the gene expression signal, from down-regulated (navy) to no change (white) to up-regulated (red). Control samples are shown in green, while cases are red (condition).

OLFM4, the top DEG identified in the previous comparisons, was not identified as significantly altered between the XFS cases and the control group (Table 3.6). The top DEG in this comparison is *PTGS2*. Only the top three DEGs, *PTGS2*, *EGR3* and *CSF2RB*, were identified as within the top ten DEGs in the main case-control comparison (Table 3.13).

Table 3.13. Top ten DEGs identified in the XFS and control comparison.

Gene ID	Description	Chr	Log2FC	P-adj
PTGS2	prostaglandin-endoperoxide synthase 2	1	5.31	9.89E-04
EGR3	early growth response 3	8	4.21	9.89E-04
CSF2RB	colony stimulating factor 2 receptor subunit beta	22	2.36	2.33E-03
EGR2	early growth response 2	10	5.17	2.37E-03
MS4A1	membrane spanning 4-domains A1	11	4.1	1.29E-02
KRT6A	keratin 6A	12	2.64	1.29E-02
EGR1	early growth response 1	5	5.56	1.61E-02
IGHA1	immunoglobulin heavy constant alpha 1	14	6.53	2.21E-02
MXRA5	matrix remodeling associated 5	X	2.51	2.28E-02
IL33	interleukin 33	9	3.17	2.28E-02

Chr- chromosome, Log2FC- Log2 fold change, P-adj- adjusted p-value.

Underlined genes indicate those identified within the top ten DEGs in the case-control comparison. Log2FC of ±0.5 (~1.5-fold change); FDR-adjusted p-value < 0.05.

Plots of the significant terms were then also generated for each database for the XFS case and control comparison (Figure 3.19). The results from the GO databases are shown in Table 3.14. Biological Process and Molecular Function GO terms were identified based on the DEGs, however no Cellular Component GO terms were found to be enriched.

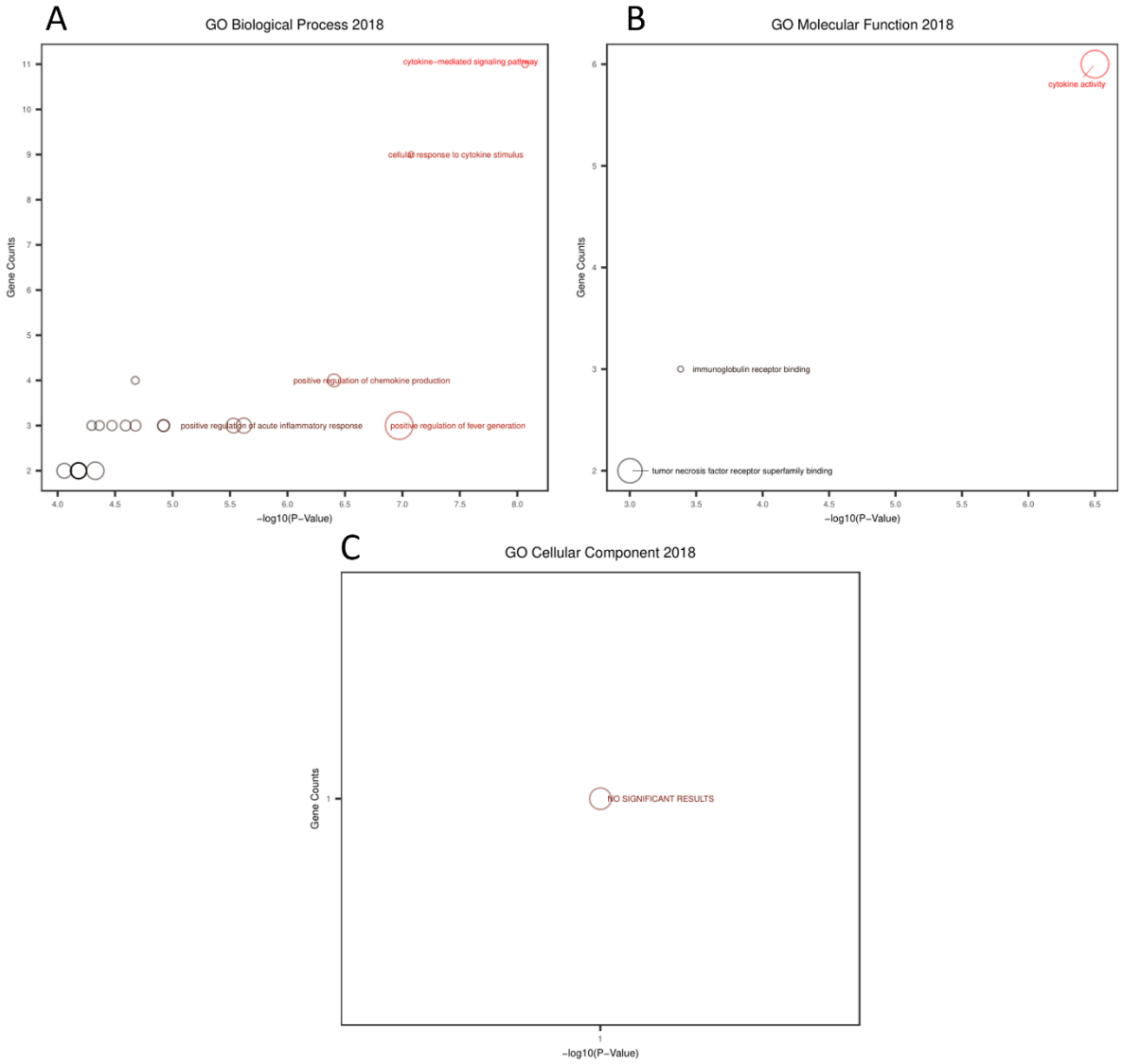


Figure 3.19. Significance plots for the GO terms in the gene-set enrichment analysis of the 36 DEGs for the XFS and control comparison. Gene-set enrichment was completed using the GO databases. A) GO Biological Process 2018, B) GO Molecular Function 2018, C) GO Cellular Component. Each plot shows the number of genes per pathway on the y-axis and the -log10(p-value) on the x-axis. The colours associated with each term represent the significance, from black (less significant) to red (highly significant). The number of genes for each term is indicated by the size of the circle. Only the top 5 terms are shown for each plot.

Table 3.14. GO terms and genes identified based on the 36 DEGs from the XFS case-control comparison

GO Terms	Genes	P-Adj.
GO Biological Processes		
cytokine-mediated Signalling pathway (GO:0019221)	*EGR1; IL33; CXCL8; IL1B; PTAFR; IL3RA; CCL3; TNFSF9; CSF2RB; PTGS2; TNF*	5.33E-06
cellular response to cytokine stimulus (GO:0071345)	*EGR1; CXCL8; IL1B; PTAFR; IL3RA; CCL3; CSF2RB; PTGS2; TNF*	2.22E-05
positive regulation of fever generation (GO:0031622)	*IL1B; PTGS2; TNF*	2.22E-05
positive regulation of chemokine production (GO:0032722)	*EGR1; IL33; IL1B; TNF*	6.14E-05
positive regulation of acute inflammatory response (GO:0002675)	*IL1B; PTGS2; TNF*	0.0003
GO Molecular Function		
cytokine activity (GO:0005125)	*IL33; CXCL8; IL1B; CCL3; SCGB3A1; TNF*	3.38E-05
immunoglobulin receptor binding (GO:0034987)	*IGKC; IGHA1; JCHAIN*	0.0222
tumor necrosis factor receptor superfamily binding (GO:0032813)	*TNFSF9; TNF*	0.0400
GO Cellular Component		
No significant results	*N/A*	N/A

P-adj- FDR-adjusted p-value; N/A – not applicable (no genes).

Plots of the significant terms were then also generated for each pathway for the XFS case and control comparison (Figure 3.20). The results from the three pathway databases are shown in Table 3.15. Multiple significantly enriched pathways were identified by WikiPathways 2016 and KEGG 2016, however only one, “Cytokine Signalling in Immune system Homo sapiens R-has-1280215”, was identified by Reactome 2016.

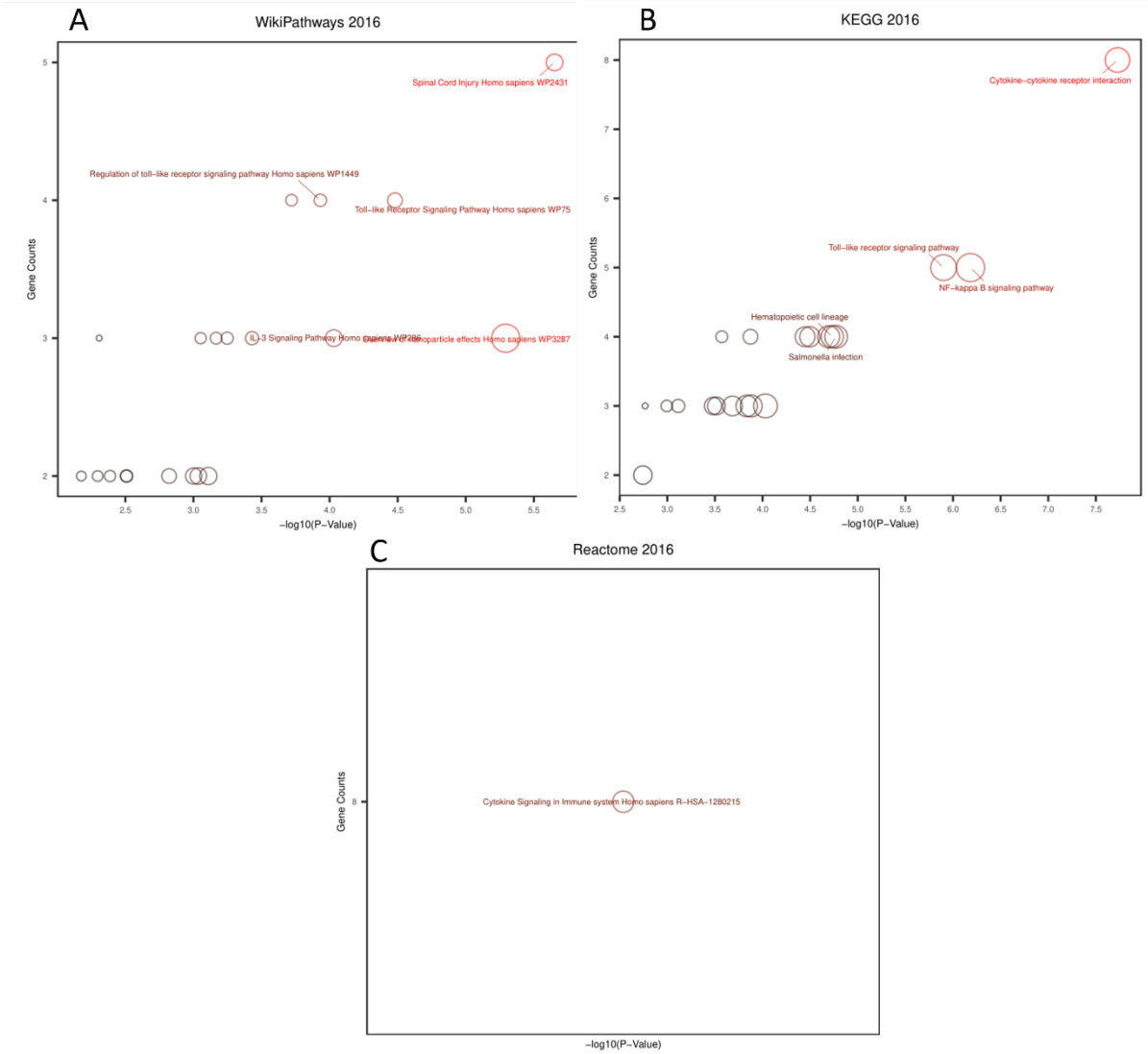


Figure 3.20. Significance plots for the pathways in the gene-set enrichment analysis of the 36 DEGs for the XFS and control comparison. Gene-set enrichment was completed using three pathways. A) WikiPathways 2016, B) KEGG 2016, C) Reactome 2016. Each plot shows the number of genes per pathway on the y-axis and the -log10(p-value) on the x-axis. The colours associated with each term represent the significance, from black (less significant) to red (highly significant). The number of genes for each term is indicated by the size of the circle. Only the top 5 terms are shown for each plot.

Table 3.15. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 36 DEGs from the XFS case-control comparison

Pathways	Genes	P-Adj.
WikiPathways 2016		
Spinal Cord Injury Homo sapiens WP2431	*EGR1; CXCL8; IL1B; PTGS2; TNF*	0.0002
Overview of nanoparticle effects Homo sapiens WP3287	*CXCL8; PTGS2; TNF*	0.0002
Toll-like Receptor Signalling Pathway Homo sapiens WP75	*CXCL8; IL1B; CCL3; TNF*	0.0011
IL-3 Signalling Pathway Homo sapiens WP286	*CXCL8; IL3RA; CSF2RB*	0.0023
Regulation of toll-like receptor Signalling pathway Homo sapiens WP1449	*CXCL8; IL1B; CCL3; TNF*	0.0023
KEGG 2016		
Cytokine-cytokine receptor interaction Homo sapiens hsa04060	*CXCL8; CCL4L2; IL1B; IL3RA; CCL3; TNFSF9; CSF2RB; TNF*	1.61E-06
Toll-like receptor Signalling pathway Homo sapiens hsa04620	*CXCL8; CCL4L2; IL1B; CCL3; TNF*	3.61E-05
Hematopoietic cell lineage Homo sapiens hsa04640	*IL1B; IL3RA; TNF; MS4A1*	0.0003
Rheumatoid arthritis Homo sapiens hsa05323	*CXCL8; IL1B; CCL3; TNF*	0.0003
Reactome 2016		
Cytokine Signalling in Immune system Homo sapiens R-HSA-1280215	*EGR1; IL33; DUSP2; IL1B; PTAFR; IL3RA; CSF2RB; TNF*	0.0021

P-adj- FDR-adjusted p-value.

Even with fewer DEGs identified, similar pathways were found to be associated with the XFS-case group. Additionally, a large portion of these pathways fell within the immune system and cytokine associated pathways. Interestingly, “Rheumatoid arthritis Homo sapiens hsa05323” was identified through KEGG for all three comparisons.

3.2.2.4. Comparison of DEGs identified in the three case-control DGE analyses

Of all the DEGs identified in the comparisons, 17 genes were identified in all three (Figure 3.21); these are *CCL3, CCL4L2, CDH4, CSF2RB, EGR1, EGR2, EGR3, FOSB, IL1B, IL33, IL3RA, KRT23, PTAFR, PTGS2, SCGB3A1, SERPINF1* and *TUSC3*.

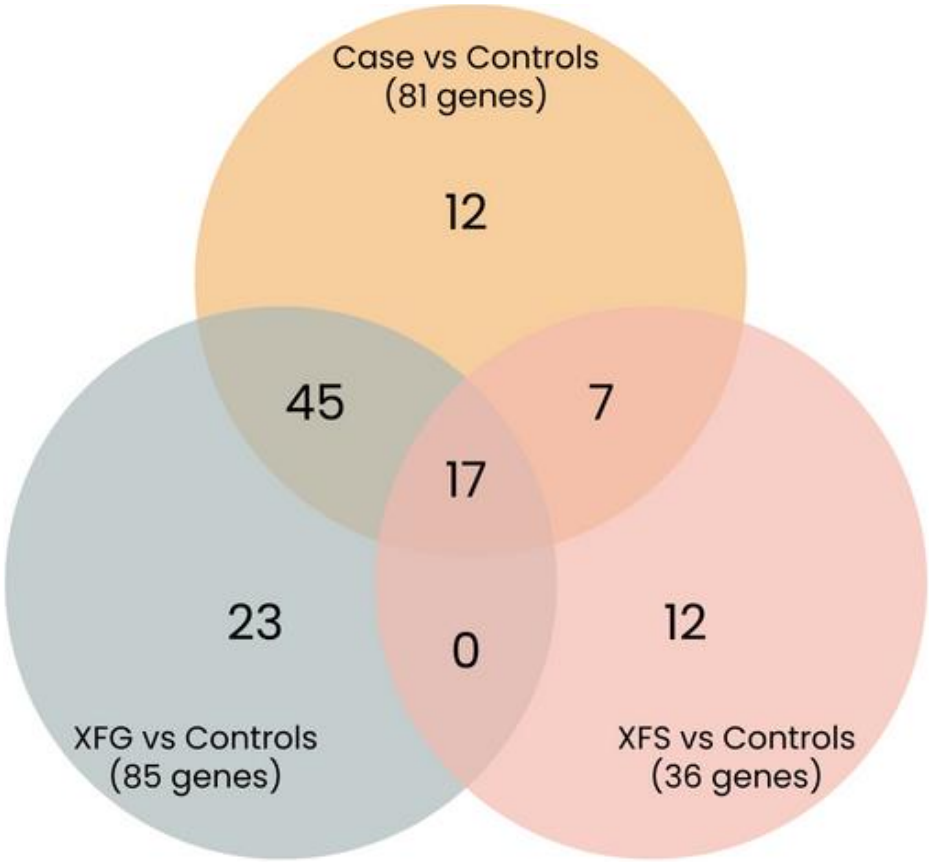


Figure 3.21. Venn diagram showing the overlap of DEGs identified in the 3 comparisons. The case-control comparison, including all 20 samples, is illustrated by an orange circle (81 DEGs identified), the XFG case-control comparison (16 cases) is indicated by a blue circle (85 DEGs identified), and the XFS case-control comparison (6 cases) is indicated by a pink circle (36 DEGs identified). The DEGs found in all three comparisons are in the centre of the diagram.

Interestingly, there were 62 genes (45 + 17) that were differentially expressed in both the full comparison and the XFG comparison. There are 24 genes (17 + 7) that overlap between the main comparison and XFS only. There are 68 genes (45 + 23) that are unique to XFG, while there are 19 genes (12 + 7) unique to XFS. These genes, along with the 17 identified in all comparisons were explored further.

The 17 DEGs found in all three comparisons were all upregulated. Multiple Biological Process GO terms and one Molecular Function GO term, "Cytokine activity (GO:0005125)", were identified based on these 17 DEGs, however no Cellular Component GO terms were found to be enriched (Table 3.16). The GO terms enriched based on only the 17 DEGs are mainly involved in the immune system.

Table 3.16. GO terms and genes identified based on the 17 DEGs identified in all three comparisons

GO Terms	Genes	P-adj.
GO Biological Processes		
Cellular response to organic substance (GO:0071310)	*EGR1; EGR2, FOSB, SERPINF1, IL1B, EGR3, IL3RA, PTGS2, IL33, CSF2RB, PTAFR, CCL3, CCL4L2*	4.41E-06
Cytokine-mediated Signalling pathway (GO:0019221)	*EGR1, IL1B, IL3RA, IL33, CSF2RB, CCL3, CCL4L2*	3.73E-05
Regulation of neuroinflammatory response (GO:0150077)	*IL1B, PTGS2, IL33, CCL3*	6.16E-05
Response to chemical (GO:0042221)	*EGR1, EGR2, FOSB, SERPINF1, IL1B, EGR3, IL3RA, PTGS2, IL33, CSF2RB, PTAFR, CCL3, CCL4L2, CDH4*	6.16E-05
Positive regulation of multicellular organismal process (GO:0051240)	*EGR1, EGR2, SERPINF1, IL1B, EGR3, PTGS2, IL33, PTAFR, CCL3, CDH4*	6.16E-05
GO Molecular Function		
Cytokine activity (GO:0005125)	*IL1B, SCGB3A1, IL33, CCL3, CCL4L2*	0.0001
GO Cellular Component		
No significant results	*N/A*	N/A

P-adj- FDR-adjusted p-value; N/A - not applicable (no genes).

Multiple significantly enriched pathways were identified by all three pathway databases (Table 3.17). The pathways enriched are mainly involved in the immune system and pro-fibrotic responses.

Table 3.17. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways and genes identified based on the 17 DEGs identified in all three comparisons

Pathways	Genes	P-adj.
WikiPathways 2016		
Network map of SARS-CoV-2 Signalling pathway WP5115	*EGR1, IL1B, PTGS2, IL33, CCL3*	9.70E-05
Overview of proinflammatory and profibrotic mediators WP5095	*IL1B, IL33, CCL3, CCL4L2*	0.0001
Cytosolic DNA-sensing pathway WP4655	*IL1B, IL33, CCL4L2*	0.0002
COVID-19 adverse outcome pathway WP4891	*IL1B, CCL3*	0.0002
Spinal cord injury WP2431	*EGR1, IL1B, PTGS2*	0.0003
KEGG 2016		
Cytokine-cytokine receptor interaction hsa04060	*IL1B, IL3RA, IL33, CSF2RB, CCL3, CCL4L2*	2.18E-09
C-type lectin receptor Signalling pathway hsa04625	*EGR2, IL1B, EGR3, PTGS2*	3.81E-09
Cytosolic DNA-sensing pathway hsa04623	*IL1B, IL33, CCL4L2*	3.81E-09
Human cytomegalovirus infection hsa05163	*IL1B, PTGS2, CCL3, CCL4L2*	1.03E-08
IL-17 Signalling pathway hsa04657	*FOSB, IL1B, PTGS2*	1.11E-08
Reactome 2016		
Cytokine Signalling in Immune system HSA-1280215	*EGR1, IL1B, IL3RA, PTGS2, IL33, CSF2RB, PTAFR, CCL3*	0.0022
Signalling by Interleukins HSA-449147	*IL1B, IL3RA, PTGS2, IL33, CSF2RB, PTAFR, CCL3*	0.0028
Interleukin-10 Signalling HSA-6783783	*IL1B, PTGS2, PTAFR, CCL3*	0.0040
NGF-stimulated transcription HSA-9031628	*EGR1, EGR2, FOSB, EGR3*	0.0056
Signal Transduction HSA-162582	*EGR1, EGR2, FOSB, EGR3, IL3RA, IL33, CSF2RB, PTAFR, CCL3*	0.0056

P-adj- FDR-adjusted p-value.

For the pathway analysis for the 68 DEGs unique to XFG, multiple Biological Process and Molecular Function GO terms were identified while only two Cellular Component GO terms were found to be enriched (Table 3.18).

Table 3.18. GO terms and genes identified based on the 68 DEGs identified in XFG cases only

GO Terms	Genes	P-adj.
GO Biological Processes		
Urea homeostasis (GO:0097274)	*SCNN1B, TFAP2B*	0.0217
Olfactory placode formation (GO:0030910)	*SIX1, SIX4*	0.0217
Positive regulation of ureteric bud formation (GO:0072107)	*SIX1, SIX4*	0.0286
Regulation of branch elongation involved in ureteric bud branching (GO:0072095)	*SIX1, SIX4*	0.0286
Trigeminal ganglion development (GO:0061551)	*SIX1, SIX4*	0.0286
GO Molecular Function		
Integrin binding (GO:0005178)	*ICAM4, ICAM1, SPP1, CCN2, CCN1, FERMT2*	0.0016
Cytokine receptor binding (GO:0005126)	*TGFB2, CCL20, CCL2, CCL4, FERMT2, OSM, SPRED1, CF300LF*	0.0046
Cytokine activity (GO:0005125)	*OSM, CCL4, CCL2, TGFB2, CCL20, SPP1, GDF15*	0.0135
Signaling receptor binding (GO:0005102)	*CD300LF, FGR, HCK, MARCHF1, ICAM1, SPRED1, ICAM4, FERMT2, OSM, CCL4, CCL20, GDF15, ENPP1, SPP1, CCL2, TGFB2, CCN2, CCN1*	0.0135
GO Cellular Component		
Extracellular space (GO:0005615)	*XYLT1, SMR3B, PIP, PI3, FSCN1, SLC13A2, TFPI, OLFM4, SERPINB10, ENPP1, GDF15, CCL20, SCNN1B, SPP1, CCL2, CCL4, OSM, ICAM1, TGFB2, CCN2, FGR, PLA2G7*	0.0060
Extracellular region (GO:0005576)	*XYLT1, SMR3B, PIP, PI3, FSCN1, SLC13A2, TFPI, OLFM4, SERPINB10, ENPP1, GDF15, CCL20, SCNN1B, SPP1, CCL2, CCL4, OSM, ICAM1, TGFB2, CCN2, FGR, PLA2G7, PLEK, ADGRE3, KRT14, RPTN, CCN1*	0.0441

P-adj- FDR-adjusted p-value.

The pathways identified by WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways are mostly associated with fibrosis and the immune response (Table 3.19).

Table 3.19. WikiPathways 2016, KEGG 2016 and Reactome 2016 pathways identified based on the 68 DEGs identified in XFG cases only

Pathways	Genes	P-adj.
WikiPathways 2016		
Platelet-mediated interactions with vascular and circulating cells (WP4462)	*TGFB2, ICAM1, CCL2*	0.0161
IL1 and megakaryocytes in obesity (WP2865)	*PLA2G7, ICAM1, CCL2*	0.0161
Immune infiltration in pancreatic cancer (WP5285	*TGFB2, CCL20, CCL2*	0.0161
Lung fibrosis (WP3624)	*CCN2, SPP1, CCL2, CCL4*	0.0161
Oncostatin M signaling pathway (WP2374)	*CCN1, FOS, CCL2, OSM*	0.0292
KEGG 2016		
Rheumatoid arthritis (hsa05323)	*ICAM1, FOS, CCL2, CCL20*	0.0085
Viral protein interaction with cytokine and cytokine receptor (hsa04061)	*CCL2, CCL20, CCL4, CXCR1*	0.0085
TNF signaling pathway (hsa04668)	*ICAM1, FOS, CCL2, CCL20*	0.0159
Chemokine signaling pathway (hsa04062)	*CCL2, CCL20, CCL4, CXCR1, FGR, HCK*	0.0204
Cytokine-cytokine receptor interaction (hsa04060)	*CCL2, CCL20, CCL4, CXCR1, GDF15, OSM, CSF3R*	0.0278
Reactome 2016		
Interleukin-10 signaling (HSA-6783783)	*CCL2, CCL20, CCL4, ICAM1*	0.0027
Chemokine receptors bind chemokines (HSA-380108)	*CCL2, CCL20, CCL4, CXCR1*	0.0045
Interleukin-4 and Interleukin-13 signaling (HSA-6785807)	*ICAM1, FSCN1, OSM, CCL2, FOS*	0.0112
Signaling by Interleukins (HSA-449147)	*DUSP6, FSCHN1, FOS, ICAM1, CCL2, CCL4, CCL20, HCK, OSM, CSF3R*	0.0130
Immune System (HSA-168256)	*PI3, ICAM4, CCL20, OLFM4, CXCR1, ICAM1, CCL2, CCL4, FOS, DUSP6, FSCN1, OSM, CSF3R, FGR, HCK, CD300LF, SERPINB10, ADGRE3, GSDME*	0.0161

P-adj- FDR-adjusted p-value.

For pathway analysis of the 19 DEGs unique to XFS, the only enriched term was the Biological Process GO term, "Humoral immune response (GO:0006959)", with genes *JCHAIN*, *MS4A1*, *TNF*, *CXCL8*, *KRT6A*.

3.2.3. Assessing Potential Confounding Factors (Effect of sex and surgery)

Two potential confounding factors for the case-control comparison are sex and surgery status. Only two participants underwent surgery, both were female, and therefore no conclusions could be drawn regarding gene expression related to surgery.

Due to the separation of cases by sex in Figure 3.7, a DGE analysis was performed to determine if any DEGs from the case-control comparisons were identified based on sex. All samples, both cases and controls, were used in the comparison of samples from males and females, where the read counts for the samples identified as female (n = 18) were compared to those identified as male (n = 17). This resulted in 23 DEGs, two of which were up-regulated, and the rest down-regulated (Appendix J). None of these DEGs overlapped with the previously identified DEGs and were all found on either the Y or the X chromosomes. It is therefore unlikely that sex is a confounder in this study.

3.2.4. Discordant Case analysis

The final comparisons investigated the discordant cases. All samples from the discordant cases were included, whether from individual eyes either affected with XFG or XFS, or unaffected (n = 10), and compared to the controls (n = 15). This resulted in 97 DEGs, with four found to be down-regulated. While *OLFM4* was not identified as a DEG for this comparison, *CCN2* was found to be upregulated. A heatmap was generated for all 97 genes (Figure 3.22). There are two main clusters, with one solely including the two samples from the discordant case, CASE22. The second cluster is further divided into discordant cases and controls, with CASE06 the only discordant case clustering with the control samples. While CASE06 is 53 years of age, young in comparison to the mean age of 69.9 ± 11.5, CASE22 is 84 years of age. For all the discordant cases, both eyes from the same participant clustered together and showed similar expression patterns. This demonstrated that intra-individual variation (irrespective of the diagnosis in the two eyes) is much lower than inter-individual variation.

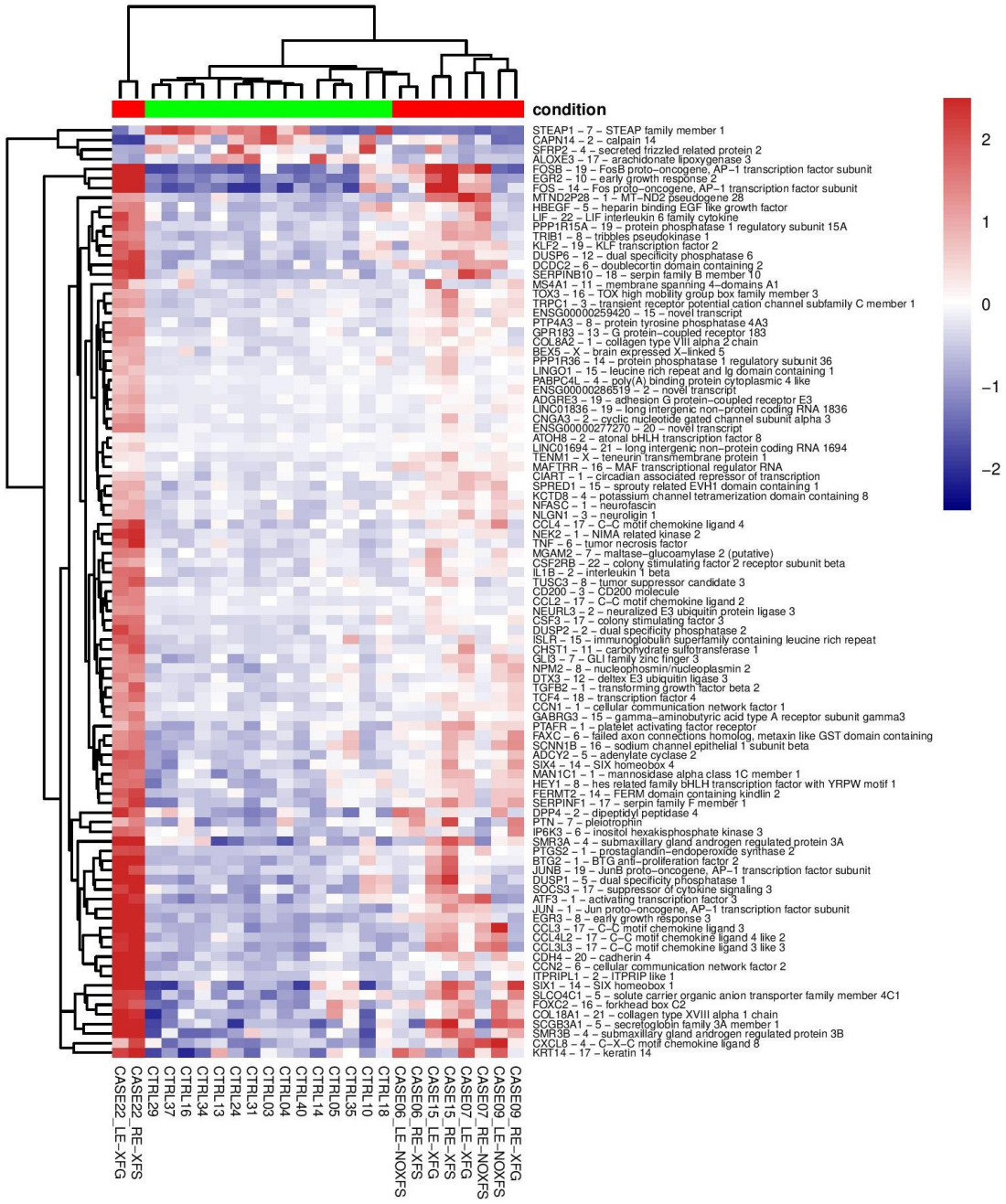


Figure 3.22. Heatmap for data visualisation of gene clustering according to the gene expression signals in the discordant-case control comparison. Each cell of the matrix represents the gene expression signal, from down-regulated (navy) to no change (white) to up-regulated (red). Control samples are shown in green, while discordant cases are red.

Notably, the three cases where one eye was unaffected showed that the affected and unaffected eyes had similar gene expression patterns, perhaps suggesting that early DGE patterns in the unaffected eye were already present. However, when the unaffected samples from the three discordant cases, CASE06, CASE07 and CASE09, were compared to the controls, only two genes were identified as significantly differentially expressed, *CXCL8* and *CCL3*.

3.3. PART III: Candidate gene expression (Replication)

A replication study was carried out to determine if a subset of the DEGs were differentially expressed in an independent sample group. Unfortunately, no RNA was left over from the original sample included in the transcriptome study, and therefore validation of the DEGs could not be performed. In addition, two known XFS-associated genes, *LOXL1*, and its isoform, *LOX*, were examined for DGE in this experiment, even though they were not amongst the DEGs.

3.3.1. Participants

Replication of the main findings from this study, that of the read count of eight DEGs, was assessed by comparing gene expression between another set of cases (n = 9) and controls (n = 5) not included in the original study by using Real-Time Quantitative PCR. The breakdown of the samples included in the validation group is shown in Table 3.20. No significant difference in sex or age was observed between the two groups (Table 3.21).

Table 3.20. Summary of case and control samples used in the replication study, with age, sex, diagnosis, topical treatment and surgical history described.

Sample Label	Age	Sex	Specific Diagnosis	Chronic topical treatment
CASE41	61	F	XFG	PG, BB
CASE42	65	F	XFG	AA, PG, BB
CASE47	87	F	XFG	PG, BB
CASE48	77	F	XFS	None
CASE43	72	M	XFG	PG, BB
CASE49	69	M	XFG	AA, PG, CAI
CASE50	70	M	XFG	AA, PG, BB
CASE53	53	M	XFS	None
CASE54	63	M	XFG	AA, PG, CAI
CTRL44	67	M	No XFM	None
CTRL45	58	M	No XFM	None
CTRL51	72	M	No XFM	None
CTRL46	58	F	No XFM	None
CTRL52	71	F	No XFM	None

Table 3.21. Summary of samples used in the replication of RNA sequencing through RT-qPCR

	Cases (n = 9)	**Controls (n = 5)**	**p-value**
Female (percentage)	4 (44.4%)	2 (40.0%)	1.00
Male (percentage)	5 (55.6%)	3 (60.0%)	
Average age (age±SD)	68.6±9.8	65.2±6.8	0.10

3.3.2. Replication of DEGs from transcriptome study

The top two significant DEGs in the main case-control comparison and the XFG-control comparison, *CCN2* and *OLFM4*, were selected for replication. The remaining genes were selected based on literature (*PTGS2*, *TGFB2*, *CDH4* and *RPTN*) or interactions with other identified DEGs (*EGR1* and *CCL3;* Figure 3.23). This enabled higher confidence in the results of other DEGs not selected for replication. Four of these genes, *PTGS2*, *EGR1*, *CDH4*, and *CCL3*, were among those 17 identified in all comparisons.

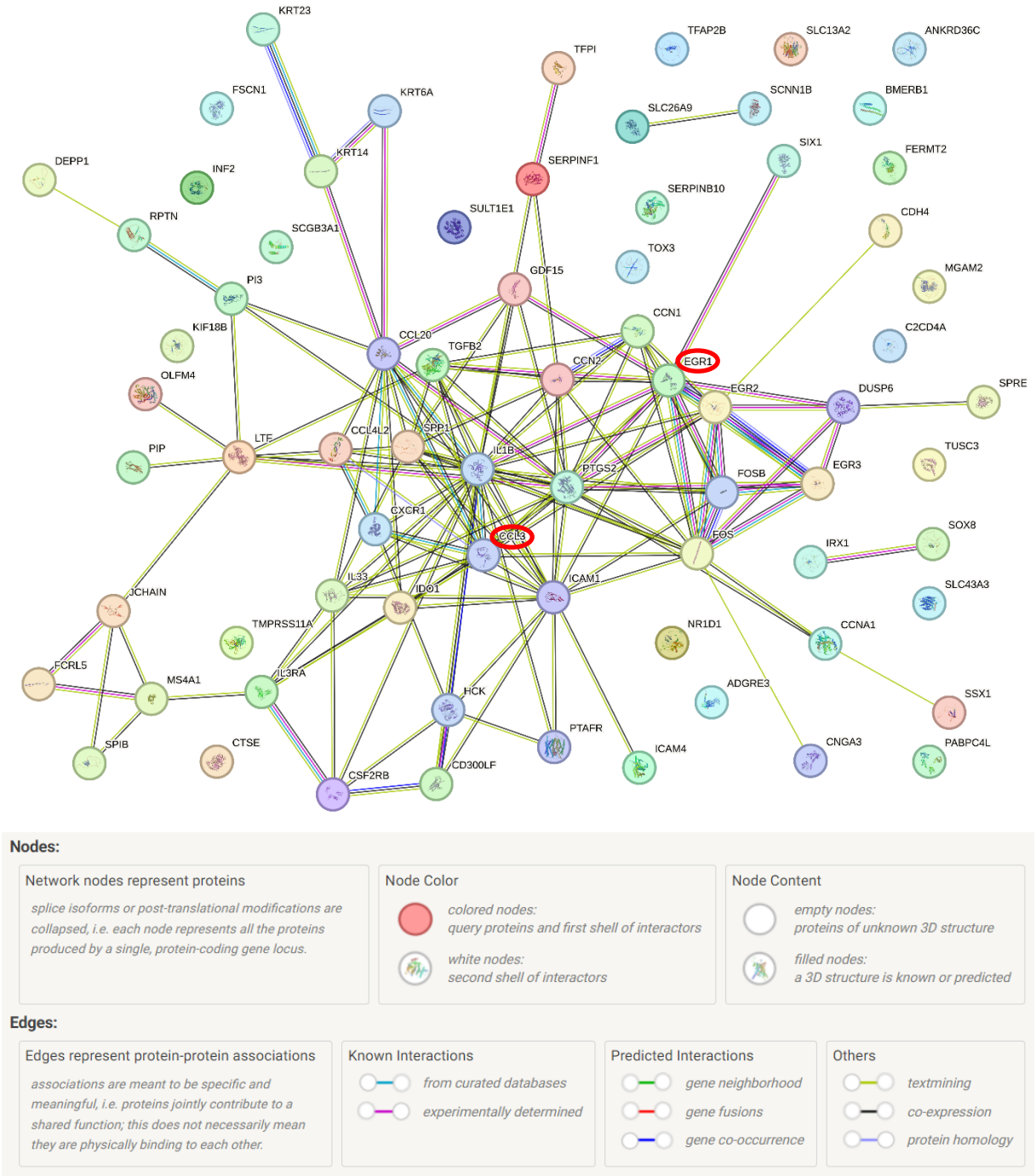


Figure 3.23. STRING Interaction network of all DEGs identified in the case-control comparison. Each circle (node) represents one of the identified DEGs, with different coloured lines representing a type of interaction (in key above). The two DEGs, *EGR1* and *CCL3*, chosen to be included in the replication set are circled in red.

The mean expression and SD of each gene within the groups was calculated and then compared. The DGE between the cases and controls are shown in Figure 3.24. The trend of increased gene expression (up-regulation) in the cases for the genes was observed in these comparisons, which supports the upregulation of these genes observed in the main transcriptome study. However,

only one gene reached the significance threshold for DGE and that was the early growth response one gene, *EGR1* (p=0.03), replicating the finding from the DGE analyses. The wide standard deviation (SD) intervals are the result of the small sample sizes in each of the groups, limiting the power of the replication study.

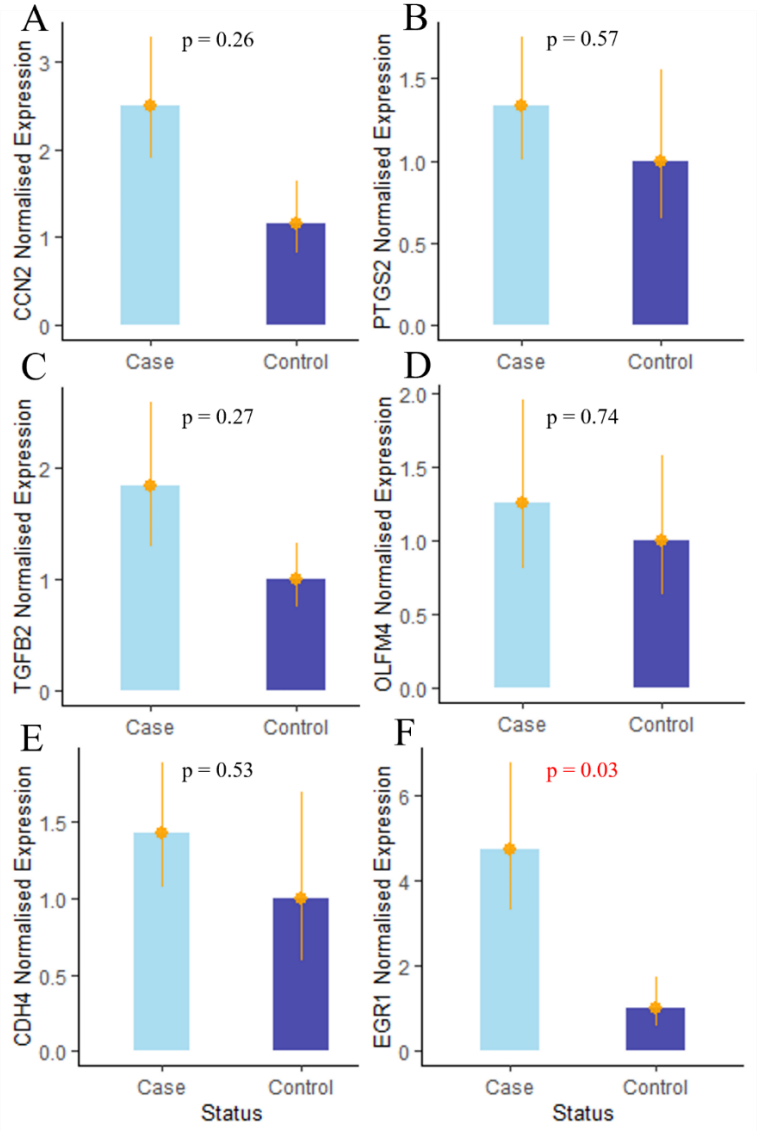


Figure 3.24. The calculated mean gene expression per gene compared between the control samples (n = 5) and the case samples (n = 9). The plots for each gene are shown as follows: (A) CCN2, (B) PTGS2, (C) TGF-B2, (D) OLFM4, (E) CDH4 and (F) EGR1. The significance values are shown above each histogram. Red: $p < 0.05$. Cases (n = 9); controls (n = 5), assay completed in triplicate for each sample.

3.3.3. RT-qPCR of *LOX* and *LOXL1*

The gene *LOXL1* and its isoform, *LOX*, were included in the Real-Time Quantitative PCR. *LOXL1* is the most well-studied gene found in association with exfoliation syndrome, however the comparative expression data of this gene shows different outcomes across studies. The *LOXL1* gene was not identified as differentially expressed in any of the comparisons in Part II of this study. There was no difference in gene expression for *LOXL1* (p=0.74) and *LOX* (p=0.98) between the control and case samples (Figure 3.25), which supports the data from the DGE analyses in the Transcriptome section of this study.

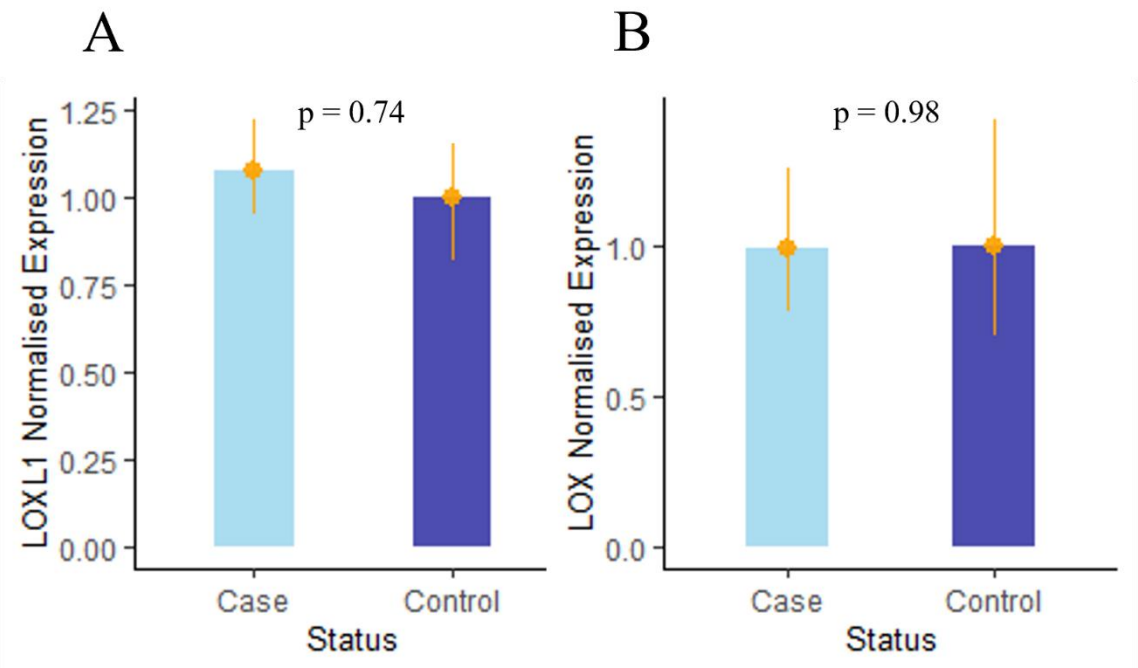


Figure 3.25. The calculated mean gene expression for LOXL1 and LOX compared between the control samples and the case samples. The plots for each gene are shown as follows: (A) LOXL1, (B) LOX. There was no significant difference in expression of LOX or LOXL1 between the cases and controls. The significance values are shown above each histogram. Cases (n = 9); controls (n = 5), assay completed in triplicate for each sample.

In summary, whole transcriptome sequencing using conjunctival cells collected via the EYEPRIM™ impression cytology device yielded reads that passed quality control measures. DGE analysis identified both known and novel DEGs and pathways associated with XFS and XFG. Real-time quantitative PCR showed no significant difference in the expression of LOXL1 and LOX between the cases and controls in this South African cohort.

CHAPTER 4: DISCUSSION

Exfoliation glaucoma (XFG) is the most recognised disease-manifestation of systemic exfoliation syndrome (XFS). It is a condition of public health relevance, as glaucoma is the leading cause of irreversible blindness globally (Pascolini and Mariotti, 2011). XFG presents with faster progression and increased severity than other forms of glaucoma and is associated with poorer outcomes and responses (Konstas *et al.*, 2004). This is particularly impactful in the black South African population due to the higher-than-expected prevalence of XFG in black South Africans compared to other populations of African ancestry (Luntz, 1972; Rotchford and Johnson, 2002; Rotchford *et al.*, 2003; Kaimbo, 2012; Olawoye *et al.*, 2012; Idakwo *et al.*, 2018). This increased prevalence of a more severe form of glaucoma illustrates the need for early detection and diagnosis, which are crucial to reduce or even prevent XFG-related visual loss (Nickells and Pelzel, 2015). The identification of early markers of disease could present an avenue for timely intervention and improved outcome, specifically in individuals where the disease has not yet progressed to glaucoma. The purpose of this study was to examine gene expression in affected individuals, both with XFG and/or XFS, compared to a control group, to identify DEGs. Through the approach of investigating both genetic, and external factors, we hope to provide a clearer understanding of the aetiology of XFS within the South African context.

Identifying DEGs also enables further investigation into potential pathogenic pathways which could help to explain the biology of XFS and XFG. It has been proposed that the accumulation of extracellular matrix (ECM) material throughout the body may result from aberrant synthesis, maintenance, or degradation of the ECM. There are several known pathogenic pathways within XFS literature that may contribute to this proposed mechanism. These include but are not limited to TGF-β1 signalling (Koliakos *et al.*, 2001; Zenkel *et al.*, 2011; Takai, Tanito and Ohira, 2012), MMP and TIMP function (Chen *et al*. 2021; Starikova *et al*. 2021), inflammatory cytokines (Zenkel *et al.*, 2010), and retinoic acid receptor-mediated signalling (Berner *et al.*, 2019). Studies focusing on pathway analyses are lacking in black South African patients with XFS/XFG, and the higher-than-expected prevalence of XFG in black South Africans necessitates further investigation. The validation of potential pathways already identified in the literature, or the identification of new potential contributory pathways will inform future steps regarding the pathophysiology of XFS. I hypothesised that genes or pathways involved in the

development and maintenance of the fibrillar extracellular matrix would be identified as differentially expressed between the XFS/XFG and control groups.

For future work to expand on the current knowledge in the gene expression landscape of XFS and other ocular diseases, a suitable and easily attainable tissue sample is required. Due to the tissue-specific nature of gene expression, most ocular disease studies rely on surgical excisions or post-mortem tissue, which limits patient recruitment options. This study investigated the potential of using conjunctival cells for the purpose of whole transcriptome sequencing. Exfoliation material has been observed in the conjunctiva (Ritch *et al.*, 2010), and conjunctival biopsies have shown evidence of exfoliation material at the earliest stage identified (Oliveira *et al.*, 2006). Therefore, conjunctival cells present an ideal sample tissue, specifically for studies aimed at identifying early markers of disease. However, collection of conjunctival cells through impression cytology for the purpose of whole transcriptome sequencing has not previously been reported, since previous studies have focused only on targeted gene expression approaches (Ganesalingam *et al.*, 2019; Thia and Tong, 2019; Latta *et al.*, 2021). This study therefore investigated the feasibility of utilising the EYEPRIM™ impression cytology device for collecting conjunctival cells for whole transcriptome sequencing. This work is published (Hulley *et al*. 2023 - Appendix K). Transcriptomic work using the appropriate tissue sample would expand avenues of research that were previously limited and allow investigation into the underlying pathways involved in the aetiology of a disease. This combined with investigation into external factors allow for a more in-depth examination of the disease.

The external factors investigated in this study included demographic information, medical history, and lifestyle habits. Caffeine, specifically through coffee consumption, has been shown to increase the risk of developing XFS (Pasquale *et al.*, 2012). Therefore, this study investigated certain common medical comorbidities, such as hypertension and diabetes, as well as lifestyle habits, such as caffeine intake, through coffee and tea consumption, and smoking status, in patients with XFS and/or XFG.

The discussion will include participant recruitment outcomes, explain the feasibility of using EYEPRIM™-harvested conjunctival cells for transcriptome sequencing, and focus on the outcomes of the transcriptome study.

4.1. Participant recruitment outcomes

Eighty-eight participants were recruited for this study, including 39 unaffected controls and 49 patients with XFG and/or XFS. Information regarding age, sex, medical history, and lifestyle habits were used to inform further analyses. Although diabetes mellitus (Wood *et al.*, 2011) and hypertension (Spečkauskas, Tamošiūnas and Jašinskas, 2012) have no known link to exfoliation syndrome, it was important to exclude them as potential confounders by comparing occurrence rates between affected and unaffected participants. There were no significant differences for diabetes mellitus or hypertension between the cases and controls, however there was a higher-than-expected occurrence of hypertension in both cases (51%) and controls (69%) compared to the expected prevalence of 43% in South Africa (Ware *et al.*, 2019). This may be due to the age of participants; however, it is also important to note that the status of hypertension and diabetes were self-reported in the questionnaire.

There was also no statistically significant difference in smoking habits or whether individuals consumed tea or coffee between the cases and controls, however a significant difference was observed for the amount of coffee consumed. Interestingly, higher consumption of coffee was observed in the control group compared to the cases. This is contrary to a previous study which found higher coffee consumption to have a 1.66 increased risk of developing XFS and/or XFG (Pasquale *et al.*, 2012). However, caffeine has been shown to decrease liver fibrosis through reduction of matrix metalloprotease (MMP) secretion (Shan *et al.*, 2022), as well as act as an antifibrotic agent in the lung by inhibiting TGF-β activation in lung epithelial cells (Tatler *et al.*, 2016). Therefore, caffeine may have a protective effect in this group through reduction of TGF-β. The study by Pasquale *et al*., (2012) was a large epidemiological study and was able to obtain information on consumption of caffeine-containing beverages through repeated questionnaires (Pasquale *et al.*, 2012). This allowed for specific measurements of caffeine, which was not possible in this study, whereby the amount of caffeine consumed was estimated based on participant memory.

4.2. Novel use of an impression cytology device to harvest cells for whole transcriptome sequencing

Harvesting the appropriate cell types at a sufficient yield for downstream transcriptome analysis is a challenge. For that reason, we investigated the feasibility of using impression cytology for non-invasive harvesting of conjunctival cells. The main challenge in using the EYEPRIM™ impression cytology device for the purpose of collecting cells for whole transcriptome sequencing was RNA extraction. Initial yields and quality of RNA extracted from the conjunctival cells were poor and not sufficient for library construction. Steps were then taken to improve both the yield and quality of the extracted RNA. One of the steps, which involved sampling a single eye twice, applying the device to two separate areas of the eye, the superior bulbar conjunctiva, and the inferior bulbar conjunctiva, served to maximise the surface area sampled. This is imperative due to exfoliation material presenting in clusters, rather than ubiquitously throughout the tissue (Kivelä, 2018). With the two samples per eye, as well as the additional disruption step using the TissueLyser LT, a five-fold increase in the average RNA yield was observed, yielding sufficient RNA for cDNA library construction and subsequent sequencing.

It is also important to maintain standardized conditions when using the EYEPRIM™ device. For example, the product information states that no topical anaesthetic is required for collection of conjunctival cells, however reports of discomfort upon application necessitated the use of eye drops in this study. All participants in both the control and case group received this topical anaesthetic eye drop, thus limiting the potential of this step acting as a confounding factor and influencing the DGE results.

During library construction, larger than optimum fragment sizes were observed. The fragmentation conditions within the Illumina TruSeq Stranded mRNA LT Sample Prep kit were optimised for high-quality RNA, and the use of low quality or degraded RNA can contribute to poorer outcomes. However, due to the RIN values and Qubit readings indicating high quality starting RNA, the reason for the larger fragment sizes was most likely due to something else. Based on knowledge from the TruSeq Stranded mRNA kit protocol (Illumina, San Diego, California, United States), one potential contributor may have been the step in

which the mRNA is fragmented into small pieces using divalent cations under elevated temperature. Perhaps, if the required temperature was not reached, this step would not have performed at an optimal level. Another potential cause for the apparently longer fragments may have occurred in the cDNA library enrichment and purification step. Excessive PCR cycles within this step may skew the representation of fragments in the library, however the same number of PCR cycles were used in the second round of library construction as the first. Additionally, an incorrect volume of AMPure XP Beads added at any of the purification steps may have skewed the fragment sizes of the library.

The success of the sequencing run was determined using Illumina's Sequence Analysis viewer and FastQC. The metrics evaluated and described in the results support the use of the EYEPRIM™ device for collection of cells due to all 40 samples passing the selected quality thresholds. It was shown that pool 2 from library construction consistently performed the poorest with a cluster density of 155.8 K/mm² falling short of the expected range of 170 K/mm² and 220 K/mm² and reduced data output. However, the percentage of the data exceeding the score of 30 is similar across the four pools, thus indicating that pool 2 retained the high quality of the data despite being under-clustered. There was also a high percentage of duplicate reads, which arise through PCR amplification. TruSeq kits use heat-fragmentation of mRNA and the only amplification is of the sequencing library, therefore all PCR duplicates can be identified through their mapping positions. Observing high rates of read duplicates in RNA-sequencing libraries is common, with previous studies showing duplicates in ranges between 36%-74% (Parekh *et al.*, 2016) comparable to our result of 30.7% to 48.9%. The percentage of duplicate reads can reflect poor library complexity, often caused by low sample input or over-amplification; it may also be due to a high abundance of a small number of genes, leading to overrepresented sequences or it may arise from real biological variation. Literature is unclear on whether it is informative or reductive to remove duplicates from downstream sequencing analysis (Parekh *et al.*, 2016), however in this study, only unique reads were used for the read counts in the gene expression analyses.

Another important factor to note was the difference between the mapping rates of HTSeq and featureCount, particularly noting the assigned reads. The range in percentage of assigned reads

counted for HTSeq was 83.1% to 88.3%, while for feature counts it was 43% to 56.9% across the 40 samples. This may be due to the sequence content, as well as algorithmic differences in interpreting the sequence content (Liao, Smyth and Shi, 2014). The large proportion of duplicate reads and overrepresented sequences may have contributed to the low percentage of assigned reads for featureCounts.

Quantification of RNA expression was performed on cells collected by the EYEPRIM device, however to our knowledge, there is no literature with evidence supporting the use of this device for whole transcriptome sequencing. The quality control data shown above indicates the potential for further studies to make use of this impression cytology device for whole transcriptome sequencing. This would open new avenues of research into ocular diseases that affect the conjunctiva, as this method is relatively non-invasive compared to other methodologies and allows for easy collection of conjunctival cells.

4.3. Differential gene expression to detect genes that are significantly up or down regulated in XFS and/or XFG

When conducting DGE analyses, it is important to take note of any potential confounders. The demographic and medical information obtained through the questionnaire examined the distribution of sex, the use of daily topical treatments by the patients with exfoliation glaucoma, and surgical history.

Male predominance in XFS studies has been observed (Parker, 2009; Idakwo *et al.*, 2018), however this is hypothesised to be due to increased outdoor activities of men compared with women and not underpinned by a biological difference. In our study, there was no statistically significant difference in the number of male and female participants in the case and control groups to account for any biological bias. Additionally, no DEGs identified by comparing all male and female participants overlapped with the DEGs identified in the case-control comparisons. Therefore, it is likely that the disease related DEGs were not linked to sex differences.

Surgical operations may also impact the gene expression levels in the eye by affecting genes involved in the healing process. Mitomycin C is an antimetabolite used in glaucoma surgeries as an antifibrotic agent to promote healing and prevent scarring. Antimetabolites are drugs that interfere with one or more enzymatic reactions necessary for DNA synthesis and therefore could alter gene function and expression (Parker, 2009). Therefore, an unoperated eye was sampled where possible (sampling from the eye that did not undergo surgery) and where this was not possible, the impression sites were in the inferior fornix, as far away as possible from the surgical site. No samples were taken from a surgical site. The effect of past surgeries on the gene expression profiles is uncertain and therefore crucial to limit as a confounding factor. In this study, only two cases previously underwent glaucoma surgery, and therefore it is likely that the disease related DEGs were not linked to surgical history.

There are different forms of topical treatment prescribed for glaucoma, and these may interfere with the gene expression levels observed in the conjunctival cells, as they are directly exposed to the treatment. In this study, all but two of the XFG participants were undergoing chronic topical treatment, while those with XFS were not. Therefore, it would be difficult to distinguish between the effects of topical treatment in the separate XFG-control and XFS-control comparison. However, the overlap of DEGs in the grouped case-control comparison supports the biological relevance of the identified DEGs. Additionally, the clustering of discordant eyes from the same individual, regardless of topical treatment status, indicates that the expression patterns are associated with the systemic disease rather than phenotype of the eye or treatment.

Three DGE analyses were carried out to maximise the knowledge gained of the pathogenesis of XFG/XFS. The first DGE analysis was a case-control comparison, including data from one eye in each case. The case group in this comparison was composed of all 20 cases sequenced, those with XFS and those who had progressed to XFG. The second and third DGE analyses were also case-control comparisons, however the case groups were divided into either XFS cases or XFG cases and compared to the controls separately. Following this, DEGs identified in the different comparisons were noted. While most of the DEGs were identified in at least two comparisons, some of the identified DEGs were only present in the XFG-control comparison or the XFS-control comparison. When integrated in pathway analysis, these DEGs may elucidate the underlying mechanism that leads to the progression of XFS to XFG.

For the three comparisons, 81 genes were found to be differentially expressed in all cases, 85 genes were found to be differentially expressed in XFG cases and 36 genes were found to be differentially expressed for XFS cases. There was a significant overlap in genes between the three comparisons. Seventeen genes were identified as differentially expressed in all comparisons. All DEGs identified as significantly differentially expressed in both the XFG-control comparison and XFS-control comparison were also identified in the main case-control comparison.

The two most significantly altered genes identified in the case-control comparison (where cases were both XFG and XFS) were *OLFM4* and *CCN2/CTGF*. Both DEGs were identified in the XFG case-control comparison but were not identified when only including the XFS cases, suggesting they may only be at play once XFS has progressed to glaucoma or potentially due to chronic topical treatment. Olfactomedin 4 (*OLFM4*) is a member of the olfactomedin family and encodes an antiapoptotic factor that promotes tumour growth and is an extracellular matrix glycoprotein. While there is no recorded link to exfoliation syndrome in the literature, *OLFM4* has previously been implicated in rheumatoid arthritis, specifically joint inflammation (Ren *et al.*, 2021). In fibroblast-like synoviocytes (FLS), inflammation cytokines, TNF-α, and interleukin (IL)-1β, can upregulate *OLFM4* (Ren *et al.*, 2021). Interleukin (IL)-1β was also identified as one of the most significantly altered genes, therefore a potential interplay between the two genes may be contributory. IL-1β is part of the IL1 gene family and has also been implicated in rheumatoid arthritis disease progression (Buchs *et al.*, 2001), while differences in gene expression between XFG cases and controls were not previously observed (Yildirim *et al.*, 2013).

While *OLFM4* has no previous links to XFS or XFG, the connective tissue growth factor (*CCN2*/*CTGF*) has been linked to multiple ocular diseases, including XFG (Ho *et al.*, 2005; Browne *et al.*, 2011). It is a member of the CCN family of matricellular proteins, which exhibit ECM-like structural features. *CCN2/CTGF* is known to be induced by TGF-β2 (Dillinger *et al.*, 2023), which was also identified as up-regulated only in the XFG-cases. The lack of evidence for this gene as a DEG in XFS cases is supported by the literature. *CCN2/CTGF* and its total protein have been shown to be markedly higher in XFG patients, but not XFS patients (Ho *et al.*, 2005; Browne *et al.*, 2011; Demirdöğen *et al.*, 2019). *CCN2/CTGF* was also found

to be a poor predictor of XFG (Demirdöğen *et al.*, 2019) and therefore limits the use of it as a biomarker.

Both *OLFM4*, an ECM glycoprotein, and *CCN2/CTGF*, exhibiting ECM-like structural features, support our hypothesis detailing the involvement of genes in the development and maintenance of the fibrillar ECM, however they are only present upon progression to glaucoma, and therefore are unlikely to be potential biomarkers of early disease. Additionally, when comparing the samples from the discordant eyes to the control group, *CCN2* but not *OLFM4* was identified as a DEG. This supports the biological relevance of *CCN2* in XFS/XFG as it was most likely not identified because of topical treatment. However, the lack of *OLFM4* shows potential for the up regulation of *OLFM4* to be a result of topical treatment, which requires further investigation.

The XFS cases were not undergoing topical treatment, therefore their inclusion in the various comparisons shed further light on biologically relevant DEGs. The following 17 genes were found to be significantly altered in both XFG and XFS cases: *CCL3, CCL4L2, CDH4, CSF2RB, EGR1, EGR2, EGR3, FOSB, IL1B, IL33, IL3RA, KRT23, PTAFR, PTGS2, SCGB3A1, SERPINF1* and *TUSC3*. Several of these genes are associated with inflammatory pathways, including genes encoding cytokines *CCL3, CCL4L2, IL1B* and *IL33*, and cytokine receptors *CSF2RB* and *IL3RA*. The platelet-activating factor receptor (*PTAFR*) is associated with the accumulation of mRNA for TGF-β, collagen IV and α-smooth muscle actin and deposition of extracellular fibrils (Latchoumycandane *et al.*, 2015), all linked to exfoliation material deposition (Zenkel *et al.*, 2011). While *EGR1* has not yet been firmly linked to XFS or XFG; *EGR2*, *FOSB*, and *EGR3* have all previously been implicated in XFS (Mullany *et al.*, 2022). Both *EGR2* and *EGR3* directly interact with *EGR1*, and therefore support the involvement of *EGR1* as well.

The most likely DEGs that could potentially be used as biomarkers would be those identified solely in the XFS cases. The development of glaucoma is accompanied by damage to ocular tissues and nerves and therefore DEGs identified in XFG cases may result from this damage. Biomarkers should enable early detection, therefore the DEGs present in XFS cases only could

be indicative of early disease, prior to the development of glaucoma and its associated damage. Identification of potential biomarkers could lead to earlier intervention and monitoring of XFS and thus potentially limit progression to XFG. The DEGs identified in XFS cases only are *TNF*, *TNFSF9*, *CEMIP*, *CFAP46*, *CHST1*, *CXCL8*, *DUSP2*, *MUC3A*, *MXRA5*, *NEK2*, *PTP4A3* and *SLC24A3*.

The most likely biomarker candidate is tumor necrosis factor (*TNF*), a pro-inflammatory cytokine belonging to the TNF superfamily. It has a critical role in microglial inflammation, which are immune cells in the retina implicated in glaucoma (Pulukool *et al.*, 2022). Microglial inflammation is implicated in neurodegeneration and therefore glaucoma development. Additionally, *TNF* affects the expression of both *MMP* and tissue inhibitors of MMPS (*TIMPs*) (Sorkhabi *et al.*, 2013). *TNF* is also known to upregulate *OLFM4* (Ren *et al.*, 2021), however *OLFM4* was not identified as upregulated in the XFS cases. Therefore, *TNF*, potentially through the activation of *OLFM4*, alteration of *MMP* and *TIMP* expression and its role in microglial inflammation, may contribute to the progression of XFS to glaucoma. *TNF* would be a good candidate for further study to determine its potential as a biomarker of early disease.

Another potential biomarker is TNF superfamily member 9 (*TNFSF9*), belonging to the TNF ligand family. *TNFSF9* is known to induce endothelial cell apoptosis and regulates the release of cytokines and transforming growth factor-β (TGF-β) (Wu *et al.*, 2021). *TNFSF9* has also been suggested as a potential prognostic biomarker for pancreatic cancer and is correlated with immune infiltrates (Wu, Wang and Jiang, 2021). Pancreatic cancer is associated with a dense ECM, thus *TNFSF9* may contribute to the dysregulation of the ECM in a similar manner in XFS (Wu, Wang and Jiang, 2021). *TNFSF9* has also been shown to be stimulated by ultraviolet light radiation (Pisarchik *et al*., 2004), therefore may be implicated in increased XFS risk due to increased UV exposure for prolonged periods (Stein *et al.*, 2011).

The XFS-associated DEG, cell migration inducing hyaluronidase 1 (*CEMIP*) has also been suggested as a biomarker for cancer. *CEMIP* encodes a cell migration-inducing protein that when upregulated can facilitate ferroptosis resistance during ECM detachment in prostate cancer. Ferroptosis a nonapoptotic type of cell death based on iron-dependent accumulation of reactive oxygen species (Liu *et al*. 2022). *CEMIP* is also involved in the depolymerization of

the extracellular matrix component hyaluronic acid (HA) (Domanegg, Sleeman and Schmaus, 2022). HA is a glycosaminoglycan (GAG), a critical component of the ECM that regulates normal structural integrity and development (Garantziotis and Savani, 2019). The spatial localisation of HA is critical to normal tissue physiology, as well as the organisation of elastic fibres (Hahn *et al.*, 2006). Increased HA concentration in the aqueous humour of the eye has been correlated with XFG (Gartaganis *et al.*, 2001), while decreased HA concentration was identified in the trabecular meshwork in primary open angle glaucoma patients (Navajas *et al.*, 2005). Therefore, upregulation of *CEMIP* may impact the role of HA thus disrupting normal structural integrity of the ECM contributing to the increased deposition of extracellular material.

Three of the other DEGs, *NEK2*, *PTP4A3* and *MXRA5*, have associations with the extracellular matrix. NIMA related kinase 2 (*NEK2*) has been implicated in the regulation of the actin cytoskeleton (Emery *et al.*, 2009), protein tyrosine phosphatase 4A3 (*PTP4A3*) is a regulator of cell adhesion structures to the extracellular matrix (Foy *et al.*, 2017), while matrix remodeling associated 5 (*MXRA5*) is a secreted glycoprotein involved in cell adhesion and ECM remodeling and implicated in pancreatic cancer (Peng *et al.*, 2023). None of these genes have been implicated in XFS, and therefore require further study into their possible contributions to its development. While DEGs are biologically informative to understanding XFG and XFS disease progression, functional and pathway analyses provide a clearer understanding of the underpinnings to the mechanisms behind the syndrome.

4.4. Known XFS/XFG-associated pathways

Several pathways have been implicated in XFS and XFG development. These include TGF-β1 signalling (Koliakos et al. 2001; Schlötzer-Schrehardt, Körtje, and Erb 2001; Takai, Tanito, and Ohira 2012; Zenkel et al. 2011), *MMP* and *TIMP* function (Chen *et al*. 2021; Starikova *et al*. 2021), inflammatory cytokines (Zenkel *et al.*, 2010), and retinoic acid receptor-mediated signalling (Berner *et al.*, 2019).

The results of this study support literature identifying inflammatory cytokines. There were several different GO terms and pathway results specifically associated with inflammatory

cytokines or the immune system in general. It has been hypothesised that inflammatory cytokines, which include chemokines, contribute to XFS development through the deposition and remodeling of the ECM, with increased cytokine activity upregulating elastin, fibrillin-1, fibulin-2 and TGF-β1, which are all ECM molecules and involved in its modulation (Zenkel *et al.*, 2011; Mullany *et al.*, 2022). Inflammatory cytokines are also known to produce pro-fibrotic effects, with fibrosis identified as a known clinical characteristic of XFS (Tomczyk-Socha *et al.*, 2023).

Several GO terms, such as "cytokine-mediated signalling pathway (GO:0019221)", "cytokine activity (GO:0005125)" and the "cellular response to cytokine stimulus (GO:0071345)" were identified in all three case-control comparisons. These terms directly link the DEGs identified to inflammatory cytokines, directly supporting their involvement in XFS development. Increased expression of certain pro-inflammatory cytokines, such as *IL-6* and *IL-8*, was only previously observed in early stage XFS (Zenkel *et al.*, 2010). We did not identify either *IL-6* or *IL-8* in our study, however interleukin 1 beta (*IL-1β*) was identified in both XFS and XFG cases, thus implicating pro-inflammatory cytokines in late stage XFS as well, even once progression to XFG has occurred.

There were several other pathways identified that do not directly refer to cytokine activity, however, were enriched based on cytokine and chemokine genes. This includes the KEGG pathway "Rheumatoid arthritis Homo sapiens hsa05323" which was enriched for all three case comparisons. Rheumatoid arthritis is an autoimmune and inflammatory disease, and genes involved in rheumatoid arthritis have previously been identified in an XFS/XFG cohort (Hirbo *et al.*, 2023). Hirbo *et al*. (2023) also highlighted the importance of the inflammatory response in the aetiology of XFS, identifying various connective tissue and inflammatory genes. Additionally, WikiPathways identified the pathway "Spinal Cord Injury Homo sapiens WP2431" for all three case comparisons, indicating an enrichment of genes involved in spinal cord injury (SCI). Inflammation is a fundamental response mechanism during the second phase of SCI (Freyermuth-Trujillo *et al.*, 2022), therefore genes involved in inflammatory responses were identified, including the pro-inflammatory cytokine, *IL-1β*, and the chemokine, *CCL2*, which is modulated by proinflammatory cytokines (Gruber *et al.*, 2015).

Another pathway linked to proinflammatory cytokines is the KEGG pathway "TNF Signalling pathway Homo sapiens hsa04668" which was enriched in the DEGs from XFG cases, while the GO molecular function term "XFS tumor necrosis factor receptor superfamily binding (GO:0032813)" was enriched in XFS cases. These further implicate TNF Signalling in XFS and XFG. As mentioned previously, tumor necrosis factor (*TNF*) and TNF superfamily member 9 (*TNFSF9*) were both identified as DEGs specific to the XFS cases, while the link to TNF Signalling for the XFG cases resulted from TNF-associated genes, *CCL20*, *IL-1β*, *CCL2*, *FOS*, *PTGS2* and *ICAM1*. The increased expression of *TNF* in XFS cases again suggests a potential role of *TNF* in the progression to glaucoma.

One pathway, potentially only identified due to proinflammatory cytokines and chemokines, was the KEGG pathway "Toll-like receptor Signalling pathway Homo sapiens hsa04620", and WikiPathways pathway "Toll-like receptor Signalling pathway Homo sapiens WP75" for both XFG and XFS cases. WikiPathways also identified "Regulation of toll-like receptor Signalling pathway Homo sapiens WP1449" for XFS cases only. Toll-like receptors (TLRs) are germline-encoded pattern-recognition receptors (PRRs), which are involved in the innate immune system (Kawasaki and Kawai, 2014). TLRs are expressed in innate immune cells such as dendritic cells (DCs) and macrophages, but are also expressed in non-immune cells, such as fibroblast cells and epithelial cells (Kawasaki and Kawai, 2014). While the TLR Signalling pathway has not been implicated as a whole in XFS previously, polymorphisms in the Toll-like receptor 4 gene (*TLR4*) were shown to be associated with XFG risk in a Japanese cohort (Takano *et al.*, 2012). *TLR4* was not identified as a DEG in this study. This pathway may have been identified due to general pro-inflammatory cytokines and chemokines, which are not unique to TLR receptor Signalling.

Another interesting enriched immune-linked pathway was the KEGG pathway "C-type lectin receptor Signalling pathway (hsa04625)". This pathway was identified when only the 17 genes found in all three comparisons were included in the pathway analysis. The C-type lectins are characterised as extracellular lectins and are secreted into the extracellular matrix and mediate cell adhesion, cell signalling, glycoprotein clearance as well as pathogen recognition (Jung *et al.*, 2018). Cell surface adhesion molecule genes were identified as generally down-regulated

in XFS (Mullany *et al.*, 2022), while the genes associated with this pathway enrichment, *EGR2*, *IL-1β*, *EGR3* and *PTGS2*, were upregulated in our study.

When combining all the above GO terms and pathways, the main link between them is the enrichment based on different pro-inflammatory cytokines and chemokines which are involved in the maintenance and development of the ECM. Therefore, this study supports the hypothesis of the involvement of genes or pathways that modulate the fibrillar ECM and its constituents. No other known potential pathogenic pathway was as strongly supported as pro-inflammatory cytokines. However, with regard to TGF-β1 signalling, our findings support a recent study where only *TGF-β2*, and not *TGF-β1* or any latent transforming growth factor beta proteins (LTBP), was found to be differentially expressed (Mullany *et al.*, 2022). Additionally, while no *MMP* and *TIMP* genes were characterised as DEGs in our study, several other identified DEGs, such as *TNF*, are known mediators in *MMP* and *TIMP* function (Sorkhabi *et al.*, 2013). In addition to providing supporting evidence of known pathways, we propose a novel potential contributory pathway based on the results of this study.

4.5. Identification of a novel potential contributory pathway unique to our South African cohort

Although the pathway discussed below has links to the immune system and proinflammatory cytokines, there is sufficient uniqueness in the potential underlying mechanism to consider it as a proposed novel pathway in need of further investigation. The potential pathway proposed involves megakaryocytes and neutrophils, both involved in immune system activity.

The WikiPathways pathway "IL1 and megakaryocytes in obesity Homo sapiens WP2865" was identified in XFG cases only, predominantly due to the DEG, phospholipase A2 group VII (*PLA2G7*). *PLA2G7* is a macrophage-induced gene, involved in neutrophil activation and is significantly related to interstitial fibrosis (Wang *et al.*, 2022). *PLA2G7 also* plays a crucial role in the differentiation of smooth muscle cells and megakaryocytes (Xiao *et al.*, 2012). Megakaryocytes are large cells of the bone marrow that are responsible for the production of platelets (Malara *et al.*, 2015), and are involved in mediating fibrosis (Papadantonakis, Matsuura and Ravid, 2012). The protein lysyl oxidase (*LOX*) has been implicated in the

proliferation of megakaryocytes (Papadantonakis, Matsuura and Ravid, 2012). The pathway link to obesity may be due to the characterisation of obesity by low-grade inflammation (Benova and Tencerova, 2020). The potential activation of neutrophils in XFG is additionally supported by the only GO cellular component term identified, "tertiary granule (GO:0070820)". A tertiary granule is a type of secretory granule comprising neutrophils, which are highly mobile, short-lived white blood cells. Neutrophils are critical inflammatory cells that are involved in tissue damage in a range of diseases and disorders (Lacy, 2006). They respond to chemotactic signals and induce inflammation through receptor-mediated respiratory bursts. The tertiary granules contain lactoferrin (*LTF*) and matrix metalloprotease 9 (*MMP9*). Lactoferrin was identified as a DEG in the case-control comparison, while no matrix metalloproteases were identified in our study. Lactoferrin is involved in the modulation of ocular inflammatory responses and normal cell growth critical for maintaining normal ocular surface health (Tamhane *et al.*, 2019), therefore any aberration in normal functioning may contribute to an abnormal inflammatory response. Additionally, WikiPathways 2016 identified "Platelet-mediated interactions with vascular and circulating cells (WP4462)" as being significantly enriched in three DEGs unique to XFG: *TGFB2, ICAM1*, *CCL2*. When the endothelium is under stress, firm adhesion or binding of platelets to the damaged site will occur through ICAM-1 in a fibrinogen-dependent manner (Koupenova *et al.*, 2018). The platelets form aggregates with neutrophils contributing to the innate immune response. Therefore, the proposed pathway may contribute to XFG through upregulation of *PLA2G7* and several cytokines thereby increasing activation of neutrophils and subsequent build-up of lactoferrin and potentially MMP9, as well as through an abnormal inflammatory response.

Several pathways were identified in this study, some unique to XFS cases and some unique to XFG cases. These pathways and terms inform the complex nature of exfoliation syndrome, while offering both support to previously recognized pathogenic pathways as well as novel pathways which may contribute to the aetiology of XFS and progression to XFG.

4.6. Samples from discordant eyes show similar expression patterns

The comparison of discordant samples to controls yielded 97 DEGs. The increase in the number of DEGs identified using these samples may be due to the compounding of gene expression because two samples from the same participants were used. This is supported by the clustering of the discordant samples from the same participant together. This suggests that the DGE for the discordant eye is related to the systemic condition rather than a specific phenotype in the specific eye or to treatment. However, the gene expression patterns may be influenced by age and therefore advancement of the disease, as CASE06, one of the youngest participants, repeatedly clustered with the control group, while CASE22, one of the oldest, clustered separate to other cases as well as the control. CASE22 also demonstrated increased up-regulation of DEGs.

The two DEGs identified in the unaffected eye to control comparison, *CXCL8* and *CCL3*, are both chemokine ligands. This indicates that immune system dysregulation associated with XFS and XFG occurs prior to visible exfoliation material, and therefore supports the contributory role of inflammatory cytokines. *CXCL8* (Litman-Zawadzka *et al.*, 2018) and *CCL3* (Ntanasis-Stathopoulos, Fotiou and Terpos, 2020) are both potential biomarkers in tumours and cancer, however this may be challenging in XFS/XFG due to the immune response to other cellular stimuli in the eye. Therefore, further investigation with an increased number of unaffected eyes in discordant cases would be beneficial.

4.7. RT-qPCR: Replication and investigation into *LOXL1* and *LOX* gene expression

Due to insufficient RNA for the participants who were used in the transcriptome study, an independent group of cases and controls was used for a replication study. While only *EGR1* was found to be significantly upregulated in the cases for the replication study, the trend of increased expression was observed for all replicated targets. The reduced sample size and therefore reduced power may have contributed to the lack of significance.

The *LOXL1* and *LOX* genes were also included in the RT-qPCR experiment, but not for replication purposes. These genes were included due to preexisting literature on the involvement of *LOXL1* in XFS development. While there have been several studies which focus on specific genotypes of *LOXL1* risk alleles in South African individuals (Williams *et al.*, 2010; Hauser *et al.*, 2015), there is no record of *LOXL1* expression data from a South African cohort. Therefore, we investigated the expression of *LOXL1* and its isoform in a small cohort of South African patients with XFG and/or XFS. No statistically significant change in gene expression was observed for either *LOXL1* or *LOX*. This expression pattern was also observed in a recent study where no differential expression of *LOXL1* was observed (Mullany *et al.*, 2022). Although *LOXL1* is the most well-known associated gene in XFS, specific risk alleles are no longer considered causative factors in XFS development due to the allele flipping phenomena observed in different ancestries (Williams *et al.*, 2010) and results from functional studies (Kim and Kim, 2012). The specific role of *LOXL1* in XFS is unclear, but potentially may be due to a contributory relationship in other biologically relevant genes and pathways.

4.8. Study Limitations/Challenges

The main study limitation was sample size, specifically for the XFS case-control comparison and the replication study. Another challenge in data interpretation could be due to XFS/XFG being an age-related condition. Although the age of participants within the case and control groups was not significantly different, there was a wide range in age (45 – 87 years), which could account for the gene expression variation specifically seen among the case participants due to potential differing stages of XFS/XFG. A smaller age range may reduce the observed interpersonal variation.

In this study, we maximised the surface area of the eye that was sampled to increase the probability of collecting cells with exfoliation material, thus sampling both the superior and inferior bulbar conjunctiva. However, this would make it difficult to distinguish the topical variations in the conjunctival cells and potentially limit the scope of studies focusing on regional conjunctival conditions. Although single-cell sequencing was out of the scope of this study it would be an important approach for future studies.

The impact of environmental factors such as UVR exposure could also not be accurately ascertained based on the questionnaire, and caffeine intake was not recorded with specific measurements, limiting the scope of the investigation of external factors.

Additionally, although the significantly enriched pathway analyses identified have support in the literature based on the underlying mechanism of the disease, these pathways are candidates only and require further research.

4.9. Future Studies

For studies focusing on regional conjunctival conditions, single cell analyses would enable different cell types from the different regions of the eye to be investigated, thus removing the limitation observed in this study. Further intrapersonal variation among the sample donors could also be examined in future studies. This was examined in a limited scope in this study through the discordant cases, but a larger sample size accounting for the variation in presentation of discordant cases would be informative in highlighting intrapersonal variation. This could also potentially aid in explaining the high interpersonal variation observed for the cases. There was a large range in the read counts of the cases, shown in several of the PCA plots, where the controls clustered together while cases were spread throughout the plot. Perhaps, sampling both eyes of affected individuals, for discordant and concordant cases, and completing DGE analyses between the eyes could shed more light on this variation.

There is also room for investigating additional contributing factors further. This could be explored through a longitudinal study, with specific measurements of caffeine consumption rather than estimates based on number of cups of coffee consumed per day. A detailed history of medical issues may make the study more translatable and could be applied in a precision medicine capacity. Investigation into ultraviolet light radiation exposure, which has previously been linked to the development of XFS (Stein *et al.*, 2011), could also be carried out if information regarding working and living conditions were gathered and examined over a period of time. These factors may have relevance on why black South Africans have an increased prevalence of XFS compared to other populations of African ancestry.

Additionally, combining whole genome with the transcriptome data may identify expression quantitative trait loci (eQTL) and contribute to the complex literature on the combined phenotypic effect of genetic variants. The identification of associated eQTLs could inform which genetic variants affect the gene expression alterations observed in the cases and may also then guide future studies.

CHAPTER 5: CONCLUSIONS

The use of the EYEPRIM device is a less invasive collection method for studies focusing on diseases that impact the conjunctiva, thereby enabling studies that do not rely on post-mortem tissues or surgical excisions. The fact that no surgical excision is required for transcriptome studies will also aid in the selection and sampling of a control group. This is the first documented study showing the use of this device for whole transcriptome sequencing.

This study provides support for pre-existing pathogenic pathways and DEGs, while also providing novel associations that further shed light on the aetiology of XFS and XFG. The identification of both genes and pathways associated with pro-inflammatory cytokines supports the literature and hypothesis that genes involved in ECM modulation would be identified.

This study also demonstrates that there is no change in the gene expression of *LOXL1* in a South African cohort, even though *LOXL1* is the most well-known XFS/XFG-associated gene. This is the first documented study investigating gene expression data in the context of XFS and XFG in South Africa.

Bibliography

Aboobakar, I.F. *et al.* (2017) 'Major review : Exfoliation syndrome ; advances in disease genetics , molecular biology , and epidemiology', *Experimental Eye Research*, 154, pp. 88–103. Available at: https://doi.org/10.1016/j.exer.2016.11.011.

Van Acker, S.I. *et al.* (2019) 'Selecting Appropriate Reference Genes for Quantitative Real-Time Polymerase Chain Reaction Studies in Isolated and Cultured Ocular Surface Epithelia', *Scientific Reports*, 9(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-019-56054-1.

Allingham, R.R. *et al.* (2001) 'Pseudoexfoliation syndrome in Icelandic families', *British Journal of Ophthalmology*, 85(6), pp. 702–707. Available at: https://doi.org/10.1136/bjo.85.6.702.

Andrews, S. (2010) 'FastQC: A Quality Control Tool for High Throughput Sequence Data', *Online* [Preprint]. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arpino, V., Brock, M. and Gill, S.E. (2015) 'The role of TIMPs in regulation of extracellular matrix proteolysis', *Matrix Biology*, 44–46, pp. 247–254. Available at: https://doi.org/10.1016/j.matbio.2015.03.005.

Ashworth Briggs, E.L. *et al.* (2015) 'TIMP1, TIMP2, and TIMP4 are increased in aqueous humor from primary open angle glaucoma patients', *Molecular Vision*, 21(March), pp. 1162–1172.

Aung, T. *et al.* (2015) 'A common variant mapping to CACNA1A is associated with susceptibility to exfoliation syndrome', *Nature Genetics*, 47(4), pp. 387–392. Available at: https://doi.org/10.1038/ng.3226.

Aung, T. *et al.* (2017) 'Genetic association study of exfoliation syndrome identifies a protective rare variant at LOXL1 and five new susceptibility loci', *Nature Genetics*, 49(7), pp. 993–1004. Available at: https://doi.org/10.1038/ng.3875.

Benova, A. and Tencerova, M. (2020) 'Obesity-Induced Changes in Bone Marrow Homeostasis', *Frontiers in Endocrinology*, 11(May), pp. 1–15. Available at: https://doi.org/10.3389/fendo.2020.00294.

Berner, D. *et al.* (2019) 'The protective variant rs7173049 at LOXL1 locus impacts on retinoic acid signaling pathway in pseudoexfoliation syndrome', *Human Molecular Genetics*, 28(15), pp. 2531–2548. Available at: https://doi.org/10.1093/hmg/ddz075.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. Available at: https://doi.org/10.1093/bioinformatics/btu170.

Browne, J.G. *et al.* (2011) 'Connective tissue growth factor is increased in pseudoexfoliation glaucoma', *Investigative Ophthalmology and Visual Science*, 52(6), pp. 3660–3666. Available at: https://doi.org/10.1167/iovs.10-5209.

Buchs, N. *et al.* (2001) 'IL-1B and IL-1Ra gene polymorphisms and disease severity in rheumatoid arthritis: Interaction with their plasma levels', *Genes and Immunity*, 2(4), pp. 222–228. Available at: https://doi.org/10.1038/sj.gene.6363766.

Budenz, D.L. *et al.* (2013) 'Prevalence of glaucoma in an urban west african population: The tema eye survey', *JAMA Ophthalmology*, 131(5), pp. 651–658. Available at: https://doi.org/10.1001/jamaophthalmol.2013.1686.

Buhrmann, R.R. *et al.* (2000) 'Prevalence of glaucoma in an African population', *Investigative*

Ophthalmology and Visual Science, 41(1), pp. 40–48.

Bustin, S.A. *et al.* (2009) 'The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments', *Clinical Chemistry*, 55(4), pp. 611–622. Available at: https://doi.org/10.1373/clinchem.2008.112797.

Chaqour, B. and Karrasch, C. (2020) 'Eyeing the extracellular matrix in vascular development and microvascular diseases and bridging the divide between vascular mechanics and function', *International Journal of Molecular Sciences*, 21(10), pp. 1–22. Available at: https://doi.org/10.3390/ijms21103487.

Chen, E.Y. *et al.* (2013) 'Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool', *BMC Bioinformatics*, 14(128). Available at: http://www.biomedcentral.com/1471-2105/14/128.

Chen, Y. *et al.* (2021) 'Higher TGF-β1, TGF-β2, MMP-2, and TIMP-1 Levels in the Aqueous Humor of Patients with Acute Primary Angle Closure', *Ophthalmic Research*, 64(1), pp. 62–67. Available at: https://doi.org/10.1159/000507762.

Chiam, N. *et al.* (2018) 'Social, health and ocular factors associated with primary open-angle glaucoma amongst Chinese Singaporeans', *Clinical and Experimental Ophthalmology*, 46(1), pp. 25–34. Available at: https://doi.org/10.1111/ceo.13008.

Davis, R.E. and Schuman, J.S. (2013) 'Pseudoexfoliation syndrome: don't brush it off', *British Journal of Ophthalmology*, 97(9), pp. 1091–1092. Available at: https://doi.org/10.1136/bjophthalmol-2012-302417.

Demirdöğen, B.C. *et al.* (2019) 'Evaluation of tear and aqueous humor level, and genetic variants of connective tissue growth factor as biomarkers for early detection of pseudoexfoliation syndrome/glaucoma', *Experimental Eye Research*, 189(October). Available at: https://doi.org/10.1016/j.exer.2019.107837.

Dillinger, A.E. *et al.* (2023) 'CCN2/CTGF tip the balance of growth factors towards TGF-β2 in primary open-angle glaucoma', *Frontiers in Molecular Biosciences*, 10(May), pp. 1–18. Available at: https://doi.org/10.3389/fmolb.2023.1045411.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. Available at: https://doi.org/10.1093/bioinformatics/bts635.

Domanegg, K., Sleeman, J.P. and Schmaus, A. (2022) 'CEMIP, a Promising Biomarker That Promotes the Progression and Metastasis of Colorectal and Other Types of Cancer', *Cancers*, 14(20), pp. 1–21. Available at: https://doi.org/10.3390/cancers14205093.

Emery, L.A. *et al.* (2009) 'Early dysregulation of cell adhesion and extracellular matrix pathways in breast cancer progression', *American Journal of Pathology*, 175(3), pp. 1292–1302. Available at: https://doi.org/10.2353/ajpath.2009.090115.

Fang, Z. and Cui, X. (2011) 'Design and validation issues in RNA-seq experiments', *Briefings in Bioinformatics*, 12(3), pp. 280–287. Available at: https://doi.org/10.1093/bib/bbr004.

Foy, M. *et al.* (2017) 'PRL-3/PTP4A3 phosphatase regulates integrin β1 in adhesion structures during migration of human ocular melanoma cells', *Experimental Cell Research*, 353(2), pp. 88–99. Available at: https://doi.org/10.1016/j.yexcr.2017.03.012.

Freyermuth-Trujillo, X. *et al.* (2022) 'Inflammation: A Target for Treatment in Spinal Cord Injury', *Cells*, 11(17), pp. 1–34. Available at: https://doi.org/10.3390/cells11172692.

Ganesalingam, K. *et al.* (2019) 'Use of a Purpose-Built Impression Cytology Device for Gene Expression Quantification at the Ocular Surface Using Quantitative PCR and Droplet Digital PCR',

Cornea, 38(1), pp. 127–133. Available at: https://doi.org/10.1097/ICO.0000000000001792.

Garantziotis, S. and Savani, R.C. (2019) 'Hyaluronan biology: A complex balancing act of structure, function, location and context.', *Matrix biology : journal of the International Society for Matrix Biology*, 78–79(1), pp. 1–10. Available at: https://doi.org/10.1016/j.matbio.2019.02.002.

Gartaganis, S.P. *et al.* (2001) 'Increased aqueous humor basic fibroblast growth factor and hyaluronan levels in relation to the exfoliation syndrome and exfoliative glaucoma', *Acta Ophthalmologica Scandinavica*, 79(6), pp. 572–575. Available at: https://doi.org/10.1034/j.1600-0420.2001.790605.x.

Van Ginderdeuren, R. *et al.* (2016) 'The conjunctiva in normal tension glaucoma patients is thinner than in primary open-angle glaucoma patients: A comparative histologic study', *Journal of Glaucoma*, 25(5), pp. e546–e549. Available at: https://doi.org/10.1097/IJG.0000000000000388.

Gottanka, J. *et al.* (1997) 'Correlation of pseudoexfoliative material and optic nerve damage in pseudoexfoliation syndrome', *Investigative Ophthalmology and Visual Science*, 38(12), pp. 2435–2446.

Gottanka, J. *et al.* (2004) 'Effects of TGF-β2 in Perfused Human Eyes', *Investigative Ophthalmology and Visual Science*, 45(1), pp. 153–158. Available at: https://doi.org/10.1167/iovs.03-0796.

Green, C.M. *et al.* (2007) 'How significant is a family history of glaucoma? Experience from the glaucoma inheritance study in Tasmania', *Clinical and Experimental Ophthalmology*, 35(9), pp. 793–799. Available at: https://doi.org/10.1111/j.1442-9071.2007.01612.x.

Gruber, H.E. *et al.* (2015) 'Proinflammatory cytokines modulate the chemokine CCL2 (MCP-1) in human annulus cells in vitro: CCL2 expression and production', *Experimental and Molecular Pathology*, 98(1), pp. 102–105. Available at: https://doi.org/10.1016/j.yexmp.2014.12.002.

Hahn, M.S. *et al.* (2006) 'Quantitative and Comparative Studies of the Vocal Fold Extracellular Matrix I: Elastic Fibers and Hyaluronic Acid', *Annals of Otology, Rhinology & Laryngology*, 115(2), pp. 156–164. Available at: https://doi.org/10.1177/000348940611500213.

Hampton, J.R. *et al.* (1975) 'Relative Contributions of History-taking, Physical Examination, and Laboratory Investigation to Diagnosis and Management of Medical Outpatients', *British Medical Journal*, 2(5969), pp. 486–489. Available at: https://doi.org/10.1136/bmj.2.5969.486.

Hauser, M.A. *et al.* (2015) 'Genetic variants and cellular stressors associated with exfoliation syndrome modulate promoter activity of a lncRNA within the LOXL1 locus', *Human Molecular Genetics*, 24(22), pp. 6552–6563. Available at: https://doi.org/10.1093/hmg/ddv347.

Hirbo, J.B. *et al.* (2023) 'Analysis of genetically determined gene expression suggests role of inflammatory processes in exfoliation syndrome', *BMC Genomics*, 24(1), pp. 1–17. Available at: https://doi.org/10.1186/s12864-023-09179-7.

Ho, S.L. *et al.* (2005) 'Elevated aqueous humour tissue inhibitor of matrix metalloproteinase-1 and connective tissue growth factor in pseudoexfoliation syndrome', *British Journal of Ophthalmology*, 89(2), pp. 169–173. Available at: https://doi.org/10.1136/bjo.2004.044685.

Hulley, M. *et al.* (2023) 'Non-invasive harvesting of conjunctival cells for whole transcriptome sequencing', *Experimental Eye Research*, 234(July), p. 109613. Available at: https://doi.org/10.1016/j.exer.2023.109613.

Idakwo, U. *et al.* (2018) 'Exfoliation syndrome in Northern Nigeria', *Clinical Ophthalmology*, 12, pp. 271–277. Available at: https://doi.org/10.2147/OPTH.S153298.

Jung, S.Y. *et al.* (2018) 'Expression, distribution, and role of C-type lectin receptors in the human and animal middle ear and eustachian tube: A review', *Molecules*, 23(4). Available at:

https://doi.org/10.3390/molecules23040734.

Kaimbo, D. (2012) 'Pseudoexfoliation syndrome in Congolese patientsLe syndrome de pseudoexfoliation capsulaire chez les patients congolais', *Journal Français d'Ophtalmologie*, 35(1), pp. 40–45. Available at: https://doi.org/10.1016/j.jfo.2011.04.018.

Kang, J.H. *et al.* (2014) 'A prospective study of folate, vitamin B6, and vitamin B12 intake in relation to exfoliation glaucoma or suspected exfoliation glaucoma', *JAMA Ophthalmology*, 132(5), pp. 549–559. Available at: https://doi.org/10.1001/jamaophthalmol.2014.100.

Kapetanakis, V. V *et al.* (2016) 'Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): a systematic review and meta-analysis', *British Journal of Ophthalmology*, 100, pp. 86–93. Available at: https://doi.org/10.1136/bjophthalmol-2015-307223.

Kawasaki, T. and Kawai, T. (2014) 'Toll-like receptor signaling pathways', *Frontiers in Immunology*, 5(SEP), pp. 1–8. Available at: https://doi.org/10.3389/fimmu.2014.00461.

Kelly, E. *et al.* (2019) 'Effects of an Aging Population and Racial Demographics on Eye Disease Prevalence: Projections for Georgia through 2050', *American Journal of Ophthalmology*, 210, pp. 35–40. Available at: https://doi.org/10.1016/j.ajo.2019.10.028.

Kim, S. and Kim, Y. (2012) 'Variations in LOXL1 associated with exfoliation glaucoma do not affect amine oxidase activity', *Molecular Vision*, 18(January), pp. 265–270.

Kivelä, T.T. (2018) 'Histopathology of Exfoliation Syndrome', *Journal of Glaucoma*, 27(7), pp. S38–S43. Available at: https://doi.org/10.1097/IJG.0000000000000947.

Koliakos, G.G. *et al.* (2001) 'Transforming and insulin-like growth factors in the aqueous humour of patients with exfoliation syndrome', *Graefe's Archive for Clinical and Experimental Ophthalmology*, 239(7), pp. 482–487. Available at: https://doi.org/10.1007/s004170100287.

Konstas, A.G.P. *et al.* (2004) 'Factors Associated with Long-term Progression or Stability in Exfoliation Glaucoma', *Archives of Ophthalmology*, 122(1), pp. 29–33. Available at: https://doi.org/10.1001/archopht.122.1.29.

Koupenova, M. *et al.* (2018) 'Circulating platelets as mediators of immunity, inflammation, and thrombosis', *Circulation Research*, 122(2), pp. 337–351. Available at: https://doi.org/10.1161/CIRCRESAHA.117.310795.

Kreft, D. *et al.* (2019) 'Prevalence , incidence , and risk factors of primary open-angle glaucoma - a cohort study based on longitudinal data from a German public health insurance', *BMC Public Health*, 19(851), pp. 1–14. Available at: https://doi.org/10.1186/s12889-019-6935-6.

Kuleshov, M. V. *et al.* (2016) 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic Acids Research*, 44(1), pp. W90–W97. Available at: https://doi.org/10.1093/nar/gkw377.

Kyari, F. *et al.* (2015) 'A Population-based survey of the prevalence and types of glaucoma in Nigeria : results from the Nigeria National Blindness and Visual Impairment Survey', *BMC Ophthalmology*, 15(176), pp. 1–15. Available at: https://doi.org/10.1186/s12886-015-0160-6.

Lacy, P. (2006) 'Mechanisms of degranulation in neutrophils', *Allergy, Asthma and Clinical Immunology*, 2(3), pp. 98–108. Available at: https://doi.org/10.2310/7480.2006.00012.

Latchoumycandane, C. *et al.* (2015) 'Inflammatory PAF receptor signaling initiates hedgehog signaling and kidney fibrogenesis during ethanol consumption', *PLoS ONE*, 10(12), pp. 1–22. Available at: https://doi.org/10.1371/journal.pone.0145691.

Latta, L. *et al.* (2021) 'Abnormal neovascular and proliferative conjunctival phenotype in limbal stem cell deficiency is associated with altered microRNA and gene expression modulated by PAX6 mutational status in congenital aniridia', *Ocular Surface*, 19(October 2019), pp. 115–127. Available at: https://doi.org/10.1016/j.jtos.2020.04.014.

Leske, M.C. *et al.* (2003) 'Factors for glaucoma progression and the effect of treatment', *Archives of Ophthalmology*, pp. 48–56.

Li, Z. *et al.* (2021) 'Association of Rare CYP39A1 Variants With Exfoliation Syndrome Involving the Anterior Chamber of the Eye', *JAMA*, 325(8), pp. 753–764. Available at: https://doi.org/10.1001/jama.2021.0507.

Liao, Y., Smyth, G.K. and Shi, W. (2014) 'FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30(7), pp. 923–930. Available at: https://doi.org/10.1093/bioinformatics/btt656.

Litman-Zawadzka, A. *et al.* (2018) 'Serum chemokine CXCL8 as a better biomarker for diagnosis and prediction of pancreatic cancer than its specific receptor CXCR2, C-reactive protein, and classic tumor markers CA 19-9 and CEA', *Polish Archives of Internal Medicine*, 128(9), pp. 524–531. Available at: https://doi.org/10.20452/pamw.4307.

Liu, B. *et al.* (2022) 'CEMIP promotes extracellular matrix-detached prostate cancer cell survival by inhibiting ferroptosis', *Cancer Science*, 113(6), pp. 2056–2070. Available at: https://doi.org/10.1111/cas.15356.

Liu, Y. *et al.* (2013) 'Gene expression profile in human trabecular meshwork from patients with primary open-angle glaucoma', *Investigative Ophthalmology and Visual Science*, 54(9), pp. 6382–6389. Available at: https://doi.org/10.1167/iovs.13-12128.

López-Miguel, A. *et al.* (2017) 'RNA Collection from Human Conjunctival Epithelial Cells Obtained with a New Device for Impression Cytology', *Cornea*, 36(1), pp. 59–63. Available at: https://doi.org/10.1097/ICO.0000000000000977.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–67. Available at: https://doi.org/10.1186/s13059-014-0550-8.

Luntz, M.H. (1972) 'Prevalence of pseudo-exfoliation syndrome in an urban South African clinic population', *American Journal of Ophthalmology*, 74(4), pp. 581–587. Available at: https://doi.org/10.1016/0002-9394(72)90816-1.

Malara, A. *et al.* (2015) 'The secret life of a megakaryocyte: emerging roles in bone marrow homeostasis control', *Cellular and Molecular Life Sciences*, 72(8), pp. 1517–1536. Available at: https://doi.org/10.1007/s00018-014-1813-y.

Marchand, M. *et al.* (2019) 'Extracellular matrix scaffolding in angiogenesis and capillary homeostasis', *Seminars in Cell and Developmental Biology*, 89(July), pp. 147–156. Available at: https://doi.org/10.1016/j.semcdb.2018.08.007.

Mpangase, P.T. *et al.* (2021) 'Nf-Rnaseqcount: a Nextflow Pipeline for Obtaining Raw Read Counts From Rna-Seq Data', *South African Computer Journal*, 33(2), pp. 1–16. Available at: https://doi.org/10.18489/SACJ.V33I2.830.

Mullany, S. *et al.* (2022) 'RNA Sequencing of Lens Capsular Epithelium Implicates Novel Pathways in Pseudoexfoliation Syndrome', *Investigative ophthalmology & visual science*, 63(3), p. 26. Available at: https://doi.org/10.1167/iovs.63.3.26.

Navajas, E.V. *et al.* (2005) 'Concentration of hyaluronic acid in primary open-angle glaucoma

aqueous humor', *Experimental Eye Research*, 80(6), pp. 853–857. Available at: https://doi.org/10.1016/j.exer.2004.12.016.

Nian, M. *et al.* (2004) 'Inflammatory cytokines and postmyocardial infarction remodeling', *Circulation Research*, 94(12), pp. 1543–1553. Available at: https://doi.org/10.1161/01.RES.0000130526.20854.fa.

Nickells, R.W. and Pelzel, H.R. (2015) 'Tools and resources for analyzing gene expression changes in glaucomatous neurodegeneration', *Experimental Eye Research*, 141, pp. 99–110. Available at: https://doi.org/10.1016/j.exer.2015.05.009.Tools.

Ntanasis-Stathopoulos, I., Fotiou, D. and Terpos, E. (2020) 'CCL3 Signaling in the Tumor Microenvironment', *Advances in Experimental Medicine and Biology*, 1231, pp. 13–21. Available at: https://doi.org/10.1007/978-3-030-36667-4_2.

O'Brien, J.M. *et al.* (2018) 'Family History in the Primary Open-Angle African American Glaucoma Genetics Study Cohort', *American Journal of Ophthalmology*, 192, pp. 239–247. Available at: https://doi.org/10.1016/j.ajo.2018.03.014.

Olawoye, O.O. *et al.* (2012) 'Exfoliation Syndrome in Nigeria', *Middle East African Journal of Ophthalmology*, 19(4), pp. 402–405. Available at: https://doi.org/10.4103/0974-9233.102759.

Oliveira, C. *et al.* (2006) 'Early diagnosis of exfoliation syndrome in the offspring of affected patients', *Acta Ophthalmologica Scandinavica*, 84(4), pp. 512–515. Available at: https://doi.org/10.1111/j.1600-0420.2006.00670.x.

Papadantonakis, N., Matsuura, S. and Ravid, K. (2012) 'Megakaryocyte pathology and bone marrow fibrosis: The lysyl oxidase connection', *Blood*, 120(9), pp. 1774–1781. Available at: https://doi.org/10.1182/blood-2012-02-402594.

Parekh, S. *et al.* (2016) 'The impact of amplification on differential expression analyses by RNA-seq', *Scientific Reports*, 6(April), pp. 1–11. Available at: https://doi.org/10.1038/srep25533.

Parker, W.B. (2009) 'Enzymology of purine and pyrimidine antimetabolites used in the treatment of cancer.', *Chemical reviews*, 109(7), pp. 2880–93. Available at: https://doi.org/10.1021/cr900028p.

Pascolini, D. and Mariotti, S.P. (2011) 'Global estimates of visual impairment: 2010', *British Journal of Ophthalmology*, 96(5), pp. 614–618. Available at: https://doi.org/10.1136/bjophthalmol-2011-300539.

Pasquale, L.R. *et al.* (2012) 'The relationship between caffeine and coffee consumption and exfoliation glaucoma or glaucoma suspect: A prospective study in two cohorts', *Investigative Ophthalmology and Visual Science*, pp. 6427–6433. Available at: https://doi.org/10.1167/iovs.12-10085.

Peng, S. qing *et al.* (2023) 'Identification of matrix-remodeling associated 5 as a possible molecular oncotarget of pancreatic cancer', *Cell Death and Disease*, 14(2), pp. 1–14. Available at: https://doi.org/10.1038/s41419-023-05684-5.

Pflugfelder, S.C. and de Paiva, C.S. (2017) 'The Pathophysiology of Dry Eye Disease', *Ophthalmology*, 124(11), pp. S4–S13. Available at: https://doi.org/10.1016/j.ophtha.2017.07.010.

Porter, K.M., Epstein, D.L. and Liton, P.B. (2012) 'Up-regulated expression of extracellular matrix remodeling genes in phagocytically challenged trabecular meshwork cells', *PLoS ONE*, 7(4), pp. 7–9. Available at: https://doi.org/10.1371/journal.pone.0034792.

Pulukool, S.K. *et al.* (2022) 'Elevated ATP, cytokines and potential microglial inflammation distinguish exfoliation glaucoma from exfoliation syndrome', *Cytokine*, 151(July 2021), p. 155807.

Available at: https://doi.org/10.1016/j.cyto.2022.155807.

Putri, G.H. *et al.* (2022) 'Analysing high-throughput sequencing data in Python with HTSeq 2.0', *Bioinformatics*, 38(10), pp. 2943–2945. Available at: https://doi.org/10.1093/bioinformatics/btac166.

Ren, X. *et al.* (2021) 'Quantitative Proteomic Analysis of Synovial Tissue Reveals That Upregulated OLFM4 Aggravates Inflammation in Rheumatoid Arthritis', *Journal of Proteome Research*, 20(10), pp. 4746–4757. Available at: https://doi.org/10.1021/acs.jproteome.1c00399.

Ribeiro Vitorino, T. *et al.* (2023) 'MMP-2 and its implications on cardiac function and structure: Interplay with inflammation in hypertension', *Biochemical Pharmacology*, 215(July), p. 115684. Available at: https://doi.org/10.1016/j.bcp.2023.115684.

Ritch, R. (2008) 'The management of exfoliative glaucoma', *Progress in Brain Research*, pp. 211–224. Available at: https://doi.org/10.1016/S0079-6123(08)01115-1.

Ritch, R. *et al.* (2010) 'Association of exfoliation syndrome and central retinal vein occlusion: An ultrastructural analysis', *Acta Ophthalmologica*, 88(1), pp. 91–95. Available at: https://doi.org/10.1111/j.1755-3768.2009.01578.x.

Ritch, R. (2014) 'Ocular and systemic manifestations of exfoliation syndrome', *Journal of Glaucoma*, 23(8), pp. S1–S8. Available at: https://doi.org/10.1097/IJG.0000000000000119.

Ritch, R. and Schlötzer-Schrehardt, U. (2001) 'Exfoliation Syndrome', *Survey of Ophthalmology*, 45(4), pp. 265–315. Available at: https://doi.org/10.1016/S0039-6257(00)00196-X.

Roodnat, A.W. *et al.* (2022) 'Genome-Wide RNA Sequencing of Human Trabecular Meshwork Cells Treated with TGF-β1: Relevance to Pseudoexfoliation Glaucoma', *Biomolecules*, 12(11). Available at: https://doi.org/10.3390/biom12111693.

Rotchford, A.P. *et al.* (2003) 'Temba glaucoma study: A population-based cross-sectional survey in urban South Africa', *Ophthalmology*, 110(2), pp. 376–382. Available at: https://doi.org/10.1016/S0161-6420(02)01568-3.

Rotchford, A.P. and Johnson, G.J. (2002) 'Glaucoma in Zulus: A Population-Based Cross-sectional Survey in a Rural District in South Africa', *Archives of Ophthalmology*, 120(4), pp. 471–478. Available at: https://doi.org/10.1001/archopht.120.4.471.

Schlötzer-Schrehardt, U. (2009) 'Molecular pathology of pseudoexfoliation syndrome/glaucoma - New insights from LOXL1 gene associations', *Experimental Eye Research*, 88(4), pp. 776–785. Available at: https://doi.org/10.1016/j.exer.2008.08.012.

Schlötzer-Schrehardt, U., Körtje, K.H. and Erb, C. (2001) 'Energy-filtering transmission electron microscopy (EFTEM) in the elemental analysis of pseudoexfoliative material', *Current Eye Research*, 22(2), pp. 154–162. Available at: https://doi.org/10.1076/ceyr.22.2.154.5522.

Shan, L. *et al.* (2022) 'Caffeine in liver diseases: Pharmacology and toxicology', *Frontiers in Pharmacology*, 13(October), pp. 1–13. Available at: https://doi.org/10.3389/fphar.2022.1030173.

Shibata, N. *et al.* (2020) 'Relative gene expression analysis of human pterygium tissues and UV radiation-evoked gene expression patterns in corneal and conjunctival cells', *Experimental Eye Research*, 199(August), p. 108194. Available at: https://doi.org/10.1016/j.exer.2020.108194.

Sorkhabi, R. *et al.* (2013) 'High-sensitivity C-reactive protein and tumor necrosis factor alpha in pseudoexfoliation syndrome', *Oman Medical Journal*, 28(1), pp. 16–19. Available at: https://doi.org/10.5001/omj.2013.04.

Soucy, J.R. *et al.* (2023) 'Retinal ganglion cell repopulation for vision restoration in optic

neuropathy : a roadmap from the RReSTORe Consortium', *Molecular Neurodegeneration*, 18(64), pp. 1–46. Available at: https://doi.org/10.1186/s13024-023-00655-y.

Spečkauskas, M., Tamošiūnas, A. and Jašinskas, V. (2012) 'Association of ocular pseudoexfoliation syndrome with ischaemic heart disease, arterial hypertension and diabetes mellitus.', *Acta ophthalmologica*, 90(6), pp. e470-5. Available at: https://doi.org/10.1111/j.1755-3768.2012.02439.x.

Starikova, D. *et al.* (2021) 'Novel Data about Association of the Functionally Significant Polymorphisms of the MMP9 Gene with Exfoliation Glaucoma in the Caucasian Population of Central Russia', *Ophthalmic Research*, 64(3), pp. 458–464. Available at: https://doi.org/10.1159/000512507.

Stein, J.D. *et al.* (2011) 'Geographic and climatic factors associated with exfoliation syndrome', *Archives of Ophthalmology*, 129(8), pp. 1053–1060. Available at: https://doi.org/10.1001/archophthalmol.2011.191.

Szklarczyk, D. *et al.* (2021) 'Erratum: The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets (Nucleic Acids Research (2021) 49 (D605–D612) DOI: 10.1093/nar/gkaa1074)', *Nucleic Acids Research*, 49(18), p. 10800. Available at: https://doi.org/10.1093/nar/gkab835.

Takai, Y., Tanito, M. and Ohira, A. (2012) 'Multiplex cytokine analysis of aqueous humor in eyes with primary open-angle glaucoma, exfoliation glaucoma, and cataract', *Investigative Ophthalmology and Visual Science*, 53(1), pp. 241–247. Available at: https://doi.org/10.1167/iovs.11-8434.

Takano, Y. *et al.* (2012) 'Association of Toll-like Receptor 4 Gene Polymorphisms in Japanese Subjects With Primary Open-Angle, Normal-Tension, and Exfoliation Glaucoma', *American Journal of Ophthalmology*, 154(5), pp. 825-832.e1. Available at: https://doi.org/10.1016/j.ajo.2012.03.050.

Tamhane, M. *et al.* (2019) 'Review of Biomarkers in Ocular Matrices: Challenges and Opportunities', *Pharmaceutical Research*, 36(3). Available at: https://doi.org/10.1007/s11095-019-2569-8.

Tarkkanen, A. and Kivelä, T. (2002) 'John G. Lindberg and the discovery of exfoliation syndrome', *Acta Ophthalmologica Scandinavica*, 80(2), pp. 151–154. Available at: https://doi.org/10.1034/j.1600-0420.2002.800206.x.

Tatler, A.L. *et al.* (2016) 'Caffeine inhibits TGFβ activation in epithelial cells, interrupts fibroblast responses to TGFβ, and reduces established fibrosis in ex vivo precision-cut lung slices', *Thorax*, 71(6), pp. 565–567. Available at: https://doi.org/10.1136/thoraxjnl-2015-208215.

Team, R.C. (2021) 'A language and environment for statistical computing', *R Foundation for Statistical Computing, Vienna* [Preprint]. Available at: https://www.r-project.org/.

Thia, Z.-Z. and Tong, L. (2019) 'Update on the role of impression cytology in ocular surface disease', *Taiwan J Ophthalmol*, 9, pp. 141–149.

Thorleifsson, G. *et al.* (2007) 'Common Sequence Variants in the LOXL1 Gene Confer Susceptibility to Exfoliation Glaucoma', *Science*, 317(2004), pp. 1397–1401.

Tian, L. *et al.* (2015) 'Transcriptome of the human retina, retinal pigmented epithelium and choroid', *Genomics*. Elsevier Inc., pp. 253–264. Available at: https://doi.org/10.1016/j.ygeno.2015.01.008.

Tomczyk-Socha, M. *et al.* (2023) 'Pseudoexfoliation Syndrome—Clinical Characteristics of Most Common Cause of Secondary Glaucoma', *Journal of Clinical Medicine*, 12(10). Available at: https://doi.org/10.3390/jcm12103580.

Tong, L. *et al.* (2009) 'Comparison of gene expression profiles of conjunctival cell lines with primary cultured conjunctival epithelial cells and human conjunctival tissue', *Gene Expression*, 14(5), pp.

265–278. Available at: https://doi.org/10.3727/105221609788681231.

Traboulsi, E.I. and Sarfarazi, M. (2008) ‘The Use of Microarray Technology in Deciphering the Cause of Genetic Eye Diseases: LOXL1 and Exfoliation Syndrome’, *American Journal of Ophthalmology*, 145(3), pp. 391–393. Available at: https://doi.org/10.1016/j.ajo.2007.12.002.

Waisberg, E. and Micieli, J.A. (2021) ‘Neuro-Ophthalmological Optic Nerve Cupping: An Overview’, *Eye and Brain*, 13, pp. 255–268. Available at: https://doi.org/10.2147/EB.S272343.

Wang, J. *et al.* (2022) ‘Characterizing cellular heterogeneity in fibrotic hypersensitivity pneumonitis by single-cell transcriptional analysis’, *Cell Death Discovery*, 8(38). Available at: https://doi.org/10.1038/s41420-022-00831-x.

Ware, L.J. *et al.* (2019) ‘Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2)’, *Journal of Human Hypertension*, 33(2), pp. 157–166. Available at: https://doi.org/10.1038/s41371-018-0125-3.

Weinreb, R.N. *et al.* (2020) ‘Matrix Metalloproteinases and Glaucoma Treatment’, *Journal of Ocular Pharmacology and Therapeutics*, 36(4), pp. 208–228. Available at: https://doi.org/10.1089/jop.2019.0146.

Weinreb, R.N., Aung, T. and Medeiros, F.A. (2014) ‘The pathophysiology and treatment of glaucoma: A review’, *JAMA*, pp. 1901–1911. Available at: https://doi.org/10.1001/jama.2014.3192.

Weinreb, R.N. and Khaw, P.T. (2004) ‘Primary open-angle glaucoma’, *The Lancet*, 363(9422), pp. 1711–1720. Available at: https://doi.org/https://doi.org/10.1016/S0140-6736(04)16257-0.

Wickham, H. (2016) ‘ggplot2: Elegant Graphics for Data Analysis’, *Springer-Verlag New York* [Preprint]. Available at: https://ggplot2.tidyverse.org.

Wiggs, J.L. and Pasquale, L.R. (2017) ‘Genetics of glaucoma’, *Human Molecular Genetics*, 26(May), pp. 21–27. Available at: https://doi.org/10.1093/hmg/ddx184.

Williams, S.E.I. *et al.* (2010) ‘Major LOXL1 risk allele is reversed in exfoliation glaucoma in a black South African population’, *Molecular Vision*, 16(March), pp. 705–712.

Winer, J. *et al.* (1999) ‘Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro’, *Analytical Biochemistry*, 270(1), pp. 41–49. Available at: https://doi.org/10.1006/abio.1999.4085.

Wolf, S.T. and Kenney, W.L. (2019) ‘The vitamin D-folate hypothesis in human vascular health’, *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, 317(3), pp. R491–R501. Available at: https://doi.org/10.1152/AJPREGU.00136.2019.

Wood, S.D. *et al.* (2011) ‘The relationship between diabetes mellitus and exfoliation syndrome in a United States Veterans Affairs population: A case-control study’, *Journal of Glaucoma*, 20(5), pp. 278–281. Available at: https://doi.org/10.1097/IJG.0b013e3181e3d483.

Wu, J. *et al.* (2021) ‘TNFSF9 promotes metastasis of pancreatic cancer through Wnt/Snail signaling and M2 polarization of macrophages.’, *Aging*, 13(17), pp. 21571–21586. Available at: https://doi.org/10.18632/aging.203497.

Wu, J., Wang, Y. and Jiang, Z. (2021) ‘TNFSF9 Is a Prognostic Biomarker and Correlated with Immune Infiltrates in Pancreatic Cancer’, *Journal of Gastrointestinal Cancer*, 52, pp. 150–159.

Xiang, M. *et al.* (2019) ‘Comparative transcriptome analysis of human conjunctiva between normal and conjunctivochalasis persons by RNA sequencing’, *Experimental Eye Research*, 184(October 2018), pp. 38–47. Available at: https://doi.org/10.1016/j.exer.2019.04.005.

Xiao, Q. *et al.* (2012) 'Nrf3-Pla2g7 interaction plays an essential role in smooth muscle differentiation from stem cells', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 32(3), pp. 730–744. Available at: https://doi.org/10.1161/ATVBAHA.111.243188.

Xie, Z. *et al.* (2021) 'Gene Set Knowledge Discovery with Enrichr.', *Current protocols*, 1(3), p. e90. Available at: https://doi.org/10.1002/cpz1.90.

Yasuda, M. *et al.* (2014) 'Retinal transcriptome profiling at transcription start sites: A cap analysis of gene expression early after axonal injury', *BMC Genomics*, 15(1), pp. 1–15. Available at: https://doi.org/10.1186/1471-2164-15-982.

Yildirim, Z. *et al.* (2013) 'The role of the cytokines in the pathogenesis of pseudoexfoliation syndrome', *International Journal of Ophthalmology*, 6(1), pp. 50–53. Available at: https://doi.org/10.3980/j.issn.2222-3959.2013.01.10.

Yue, B. (2014) 'Biology of the extracellular matrix: An overview', *Journal of Glaucoma*, 23(8), pp. S20–S23. Available at: https://doi.org/10.1097/IJG.0000000000000108.

Yurchenco, P.D. (2011) 'Basement membranes: Cell scaffoldings and signaling platforms', *Cold Spring Harbor Perspectives in Biology*, 3(2), pp. 1–27. Available at: https://doi.org/10.1101/cshperspect.a004911.

Zenkel, M. *et al.* (2010) 'Proinflammatory cytokines are involved in the initiation of the abnormal matrix process in pseudoexfoliation syndrome/glaucoma', *American Journal of Pathology*, 176(6), pp. 2868–2879. Available at: https://doi.org/10.2353/ajpath.2010.090914.

Zenkel, M. *et al.* (2011) 'Regulation of Lysyl Oxidase-like 1 (LOXL1) and elastin-related genes by pathogenic factors associated with pseudoexfoliation syndrome', *Investigative Ophthalmology and Visual Science*, 52(11), pp. 8488–8495. Available at: https://doi.org/10.1167/iovs.11-8361.

Zenkel, M. and Schlötzer-Schrehardt, U. (2014) 'Expression and regulation of LOXL1 and elastin-related genes in eyes with exfoliation syndrome', *Journal of Glaucoma*, 23(8), pp. S48–S50. Available at: https://doi.org/10.1097/IJG.0000000000000120.

Zhao, S. *et al.* (2015) 'Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap', *BMC Genomics*, 16(1), pp. 1–14. Available at: https://doi.org/10.1186/s12864-015-1876-7.

APPENDICES

Appendix A

Ethics clearance certificate for "Differentially Expressed Genes in Exfoliation Syndrome, Exfoliation Glaucoma and Primary Open-Angle Glaucoma in Black South Africans" (M160701):

R14/49 Dr Susan Williams et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M160701

NAME: (Principal Investigator)	Dr Susan Williams et al
DEPARTMENT:	Neurosciences, Division of Ophthalmology Charlotte Maxeke Johannesburg Academic Hospital Sydney Brenner Institute for Molecular Bioscience
PROJECT TITLE:	Differentially Expressed Genes in Exfoliation Syndrome, Exfoliation Glaucoma and Primary Open-Angle Glaucoma in Black South Africans
DATE CONSIDERED:	29/07/2016
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	09/12/2016

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

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in July and will therefore be due in the month of July each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date 13/12/2016

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B

Ethics clearance certificate for “Gene Identification in Exfoliation Syndrome and Exfoliative Glaucoma” (M160706):

R14/49 Dr Susan E.I. Williams et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M160706

NAME: (Principal Investigator)	Dr Susan E.I. Williams et al
DEPARTMENT:	Ophthalmology Genome Istitute of Singapore
PROJECT TITLE:	Gene Identification in Exfoliation Syndrome and Exfoliative Glaucoma
DATE CONSIDERED:	29/07/2016
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	
APPROVED BY:	Professor P Cleaton-Jones,Chairperson, HREC (Medical)
DATE OF APPROVAL:	12/08/2016

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

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 10004, 10th floor, Senate House/3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in July and will therefore be due in the month of July each year.

Principal Investigator Signature

Date 13/08/2016

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix C

Ethics clearance certificate for this study (M170883):

R14/49 Miss Michaella Hulley

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M170883

NAME: (Principal Investigator)	Miss Michaella Hulley
DEPARTMENT:	Human Genetics Charlotte Maxeke Johannesburg Academic Hospital National Health Laboratory Services
PROJECT TITLE:	Differential Gene Expression in Exfoliation Syndrome and Exfoliation Glaucoma in the Conjunctiva of Black South Africans
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under Primary Study (M160701)
SUPERVISOR:	Dr Susan Williams
APPROVED BY:	Professor P. Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	22/09/2017

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

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 10004,10th floor, Senate House/3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand. I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. in this case, the study was initially review August and will therefore be due in the month of August each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 18/10/2017

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix D

Information Sheet:

Participant Code:		
Date:	BP=	HGT=
Date of birth:	Age:	Sex:
Place of birth:	Home language:	
Self-identified ethnicity:		
Place of schooling:		
Highest level of education:		
Occupation:		
Place of occupation:		
Type of abode:		
Place of abode:		
Family history of eye problems:		
Medical history:		
Diabetic? (if yes do they use tablets or insulin. When was the diagnosis made?)		
Hypertensive? (if yes what medication are they using. When was the diagnosis made?)		
Other illnesses?		
Smoker: Y / N	..... pack years	
Coffee: Y / N	..... cups per day	
Tea: Y / N	..... cups per day	
Coke: Y / N (or other caffeinated beverage)	..... cans per day	
Alcohol: Y / N	..... units per day	
Other information:		

Participant code:

Diagnosis (circle): Control / XFS / XFG / POAG

Date of diagnosis:……..

IOP at diagnosis: RE:……. LE:…….

Ophthalmic treatment:……

……

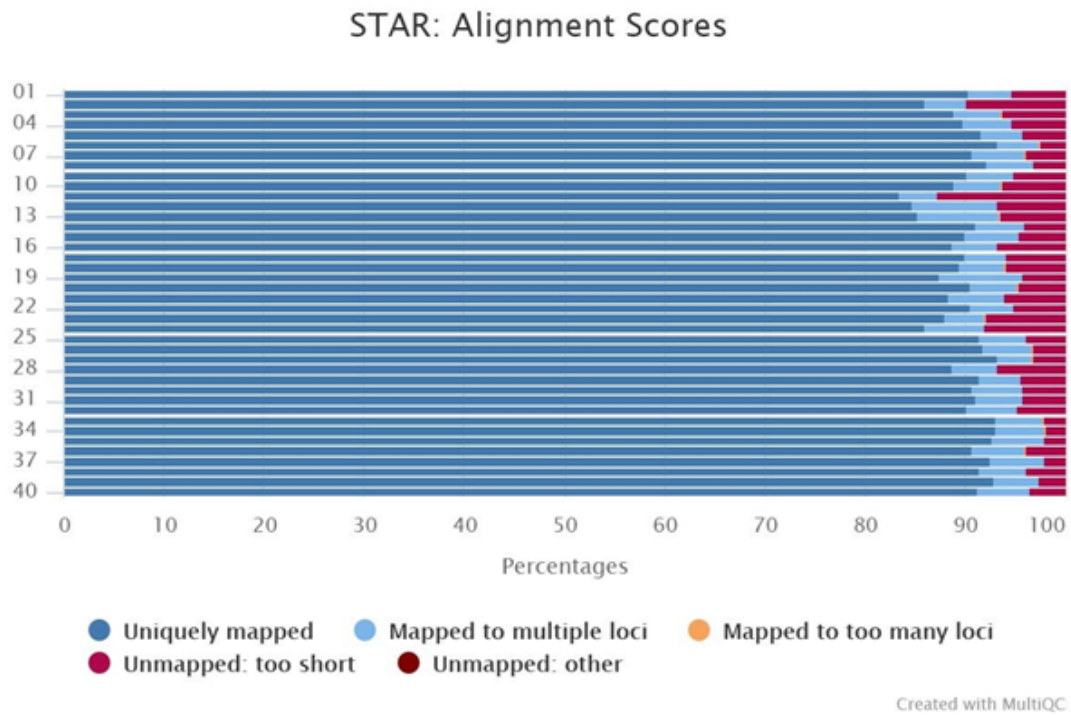
Previous ophthalmic surgery:……

……

Examination:

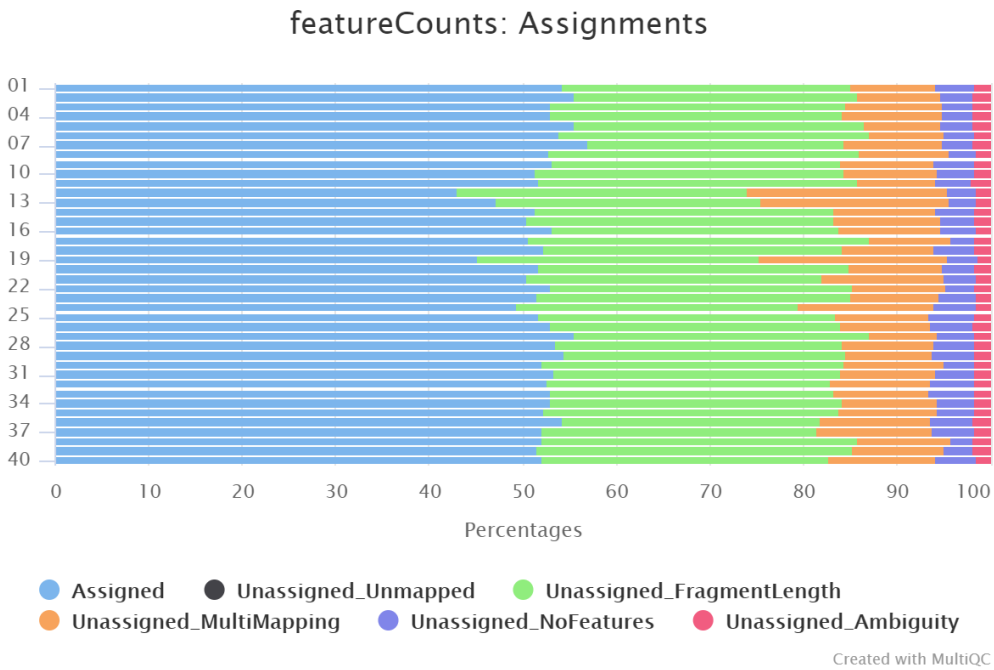
Right Eye		**Left Eye**
	GDx NFI	
	GDx TSNIT	
	Oculus MD	
	Visual Acuity	
	Conjunctiva	
	Cornea	
	GAT	
	Gonioscopy	
	CCT	
	Anterior chamber	
	Pupil	
	Lens	
	Vitreous	
	Fundus	
	Vertical disc diameter	
	Cupping	
	Diagram if required	

Appendix E



Appendix E Figure. Per sample percentage breakdown of reads aligned based on aligned categories (STAR). Each horizontal bar represents a sample. Over 80% of reads were uniquely mapped for all samples.

Appendix F



Appendix F Figure. Per sample percentage of mapped reads assigned with different levels of certainty according to featureCounts. Each horizontal bar represents a sample. Over 50% of reads were assigned to a gene feature.

Appendix G

Appendix G Table. List of the 81 genes returned as differentially expressed by DESeq2 for the full case and control comparison

GeneID	Description	Chr	Log2FC	P-adj
OLFM4	olfactomedin 4	13	4.63	2.60E-05
CCN2	cellular communication network factor 2	6	3.9	2.60E-05
PTAFR	platelet activating factor receptor	1	2.12	2.60E-05
CSF2RB	colony stimulating factor 2 receptor subunit beta	22	2.3	3.59E-05
PTGS2	prostaglandin-endoperoxide synthase 2	1	4.06	9.05E-05
ANKRD36C	ankyrin repeat domain 36C	2	3.38	1.55E-04
LINC00342	long intergenic non-protein coding RNA 342	2	2.87	1.55E-04
IL1B	interleukin 1 beta	2	3.37	1.72E-04
EGR3	early growth response 3	8	4.08	3.61E-04
ICAM1	intercellular adhesion molecule 1	19	2.43	1.13E-03
RPTN	repetin	1	3.87	1.13E-03
KRT23	keratin 23	17	2.31	1.19E-03
TUSC3	tumor suppressor candidate 3	8	3.11	1.37E-03
EGR1	early growth response 1	5	4.38	1.37E-03
EGR2	early growth response 2	10	4.05	1.37E-03
ADGRE3	adhesion G protein-coupled receptor E3	19	3.8	1.37E-03
SERPINF1	serpin family F member 1	17	1.82	1.37E-03
IL33	interleukin 33	9	2.8	1.37E-03
SPRED1	sprouty related EVH1 domain containing 1	15	1.36	1.37E-03
IRX1	iroquois homeobox 1	5	5.4	1.37E-03
TFAP2B	transcription factor AP-2 beta	6	3.35	2.93E-03
CCL3	C-C motif chemokine ligand 3	17	3.19	3.23E-03
ICAM4	intercellular adhesion molecule 4 (Landsteiner-Wiener blood group)	19	2.75	3.46E-03
CCL20	C-C motif chemokine ligand 20	2	2.74	4.02E-03
SCGB3A1	secretoglobin family 3A member 1	5	3.15	4.02E-03
SCNN1B	sodium channel epithelial 1 subunit beta	16	2.68	4.02E-03
TMPRSS11A	transmembrane serine protease 11A	4	3.69	4.02E-03

TGFB2	transforming growth factor beta 2	1	2.58	4.50E-03
SPIB	Spi-B transcription factor	19	3.62	4.50E-03
FOS	Fos proto-oncogene AP-1 transcription factor subunit	14	3.64	5.25E-03
CTSE	cathepsin E	1	3.98	5.26E-03
SLC43A3	solute carrier family 43 member 3	11	1.95	5.43E-03
KRT6A	keratin 6A	12	3.26	5.43E-03
KRT14	keratin 14	17	2.62	5.92E-03
IGHA1	immunoglobulin heavy constant alpha 1	14	5.51	5.92E-03
FSCN1	fascin actin-bundling protein 1	7	1.76	6.98E-03
JCHAIN	joining chain of multimeric IgA and IgM	4	3.66	6.98E-03
SPP1	secreted phosphoprotein 1	4	3.37	7.01E-03
SIX1	SIX homeobox 1	14	3.03	7.01E-03
IGKC	immunoglobulin kappa constant	2	4.9	8.26E-03
CDH4	cadherin 4	20	3.62	8.27E-03
CCL4L2	C-C motif chemokine ligand 4 like 2	17	2.82	8.93E-03
IL3RA	interleukin 3 receptor subunit alpha	X	2.26	9.25E-03
SULT1E1	sulfotransferase family 1E member 1	4	4.53	1.27E-02
KIF18B	kinesin family member 18B	17	2.76	1.39E-02
FERMT2	FERM domain containing kindlin 2	14	2.28	1.40E-02
MS4A1	membrane spanning 4-domains A1	11	4.02	1.40E-02
SOX8	SRY-box transcription factor 8	16	3.61	1.45E-02
NR1D1	nuclear receptor subfamily 1 group D member 1	17	2.03	1.45E-02
INF2	inverted formin 2	14	1.88	1.49E-02
CD300LF	CD300 molecule like family member f	17	1.83	1.52E-02
CCNA1	cyclin A1	13	4.09	1.60E-02
PI3	peptidase inhibitor 3	20	2	1.63E-02
SLC26A9	solute carrier family 26 member 9	1	3.84	1.65E-02
PIP	prolactin induced protein	7	2.83	1.75E-02
TDRG1	testis development related 1	6	3.08	1.83E-02
TFPI	tissue factor pathway inhibitor	2	3.58	1.86E-02

FOSB	FosB proto-oncogene AP-1 transcription factor subunit	19	3.33	1.99E-02
DEPP1	DEPP1 autophagy regulator	10	2.71	2.01E-02
C2CD4A	C2 calcium dependent domain containing 4A	15	2.8	2.01E-02
PABPC4L	poly(A) binding protein cytoplasmic 4 like	4	3.48	2.01E-02
GDF15	growth differentiation factor 15	19	2.53	2.14E-02
FCRL5	Fc receptor like 5	1	4.21	2.14E-02
ENSG00000259420	novel transcript	15	2.48	2.14E-02
SERPINB10	serpin family B member 10	18	2.51	2.35E-02
DUSP6	dual specificity phosphatase 6	12	1.62	2.38E-02
BMERB1	bMERB domain containing 1	16	1.93	2.43E-02
SLC13A2	solute carrier family 13 member 2	17	4.42	2.50E-02
MGAM2	maltase-glucoamylase 2 (putative)	7	2.69	2.92E-02
HCK	HCK proto-oncogene Src family tyrosine kinase	20	1.49	3.07E-02
CXCR1	C-X-C motif chemokine receptor 1	2	3.58	3.52E-02
NAPSB	napsin B aspartic peptidase pseudogene	19	1.69	3.53E-02
CCN1	cellular communication network factor 1	1	1.82	3.93E-02
ENSG00000286519	novel transcript	2	2.65	3.93E-02
CNGA3	cyclic nucleotide gated channel subunit alpha 3	2	3.79	4.10E-02
IGHA2	immunoglobulin heavy constant alpha 2 (A2m marker)	14	4.52	4.40E-02
LTF	lactotransferrin	3	4.26	4.69E-02
TOX3	TOX high mobility group box family member 3	16	2.14	4.77E-02
SSX1	SSX family member 1	X	-2.45	4.77E-02
IDO1	indoleamine 23-dioxygenase 1	8	2.52	4.77E-02
IGHG4	immunoglobulin heavy constant gamma 4 (G4m marker)	14	4.96	4.98E-02

Appendix H

Appendix H Table. List of the 85 genes returned as differentially expressed by DESeq2 for the XFG and control comparison

GeneID	Description	Chr	Log2FC	P-adj
OLFM4	olfactomedin 4	13	4.83	1.22E-05
CCL2	C-C motif chemokine ligand 2	17	5.67	1.22E-05
CCN2	cellular communication network factor 2	6	4.11	1.22E-05
PTAFR	platelet activating factor receptor	1	2.25	1.22E-05
SPRED1	sprouty related EVH1 domain containing 1	15	1.48	2.34E-05
ANKRD36C	ankyrin repeat domain 36C	2	3.62	3.52E-05
IL1B	interleukin 1 beta	2	3.6	7.75E-05
CSF2RB	colony stimulating factor 2 receptor subunit beta	22	2.18	1.01E-04
LINC00342	long intergenic non-protein coding RNA 342	2	3.05	1.01E-04
EGR1	early growth response 1	5	4.78	4.07E-04
ICAM1	intercellular adhesion molecule 1	19	2.62	4.16E-04
EGR3	early growth response 3	8	4.26	4.16E-04
PTGS2	prostaglandin-endoperoxide synthase 2	1	3.74	4.37E-04
TUSC3	tumor suppressor candidate 3	8	3.34	5.23E-04
EGR2	early growth response 2	10	4.39	5.28E-04
SERPINF1	serpin family F member 1	17	1.94	5.92E-04
ADGRE3	adhesion G protein-coupled receptor E3	19	3.87	5.93E-04
IRX1	iroquois homeobox 1	5	5.68	5.93E-04
KRT23	keratin 23	17	2.46	6.24E-04
IL33	interleukin 33	9	2.96	6.40E-04
CCL3	C-C motif chemokine ligand 3	17	3.48	6.54E-04
ICAM4	intercellular adhesion molecule 4 (Landsteiner-Wiener blood group)	19	3.02	9.39E-04
NR1D1	nuclear receptor subfamily 1 group D member 1	17	2.27	1.25E-03
FOS	Fos proto-oncogene AP-1 transcription factor subunit	14	3.96	1.40E-03
SCGB3A1	secretoglobin family 3A member 1	5	3.25	1.75E-03
TFAP2B	transcription factor AP-2 beta	6	3.56	1.87E-03
RPTN	repetin	1	3.99	2.13E-03
CCL4L2	C-C motif chemokine ligand 4 like 2	17	3.12	2.17E-03
TGFB2	transforming growth factor beta 2	1	2.76	3.46E-03
SLC43A3	solute carrier family 43 member 3	11	2.03	4.26E-03
CCN1	cellular communication network factor 1	1	2.02	4.69E-03
SIX1	SIX homeobox 1	14	3.25	5.31E-03
FERMT2	FERM domain containing kindlin 2	14	2.45	6.25E-03
SCNN1B	sodium channel epithelial 1 subunit beta	16	2.81	6.37E-03
DUSP6	dual specificity phosphatase 6	12	1.78	6.90E-03
CXCR1	C-X-C motif chemokine receptor 1	2	3.9	7.72E-03
SPP1	secreted phosphoprotein 1	4	3.53	7.76E-03
C2CD4A	C2 calcium dependent domain containing 4A	15	2.96	7.76E-03
PABPC4L	poly(A) binding protein cytoplasmic 4 like	4	3.74	7.76E-03

KRT14	keratin 14	17	2.73	8.59E-03
KIF18B	kinesin family member 18B	17	2.98	9.64E-03
SERPINB10	serpin family B member 10	18	2.73	1.00E-02
TDRG1	testis development related 1	6	3.32	1.17E-02
FOSB	FosB proto-oncogene AP-1 transcription factor subunit	19	3.63	1.29E-02
SULT1E1	sulfotransferase family 1E member 1	4	4.8	1.34E-02
CD300LF	CD300 molecule like family member f	17	1.84	1.44E-02
CCL20	C-C motif chemokine ligand 20	2	2.61	1.51E-02
PIP	prolactin induced protein	7	3.01	1.51E-02
CSF3R	colony stimulating factor 3 receptor	1	2.01	1.57E-02
FSCN1	fascin actin-bundling protein 1	7	1.78	2.04E-02
ENSG00000259420	novel transcript	15	2.62	2.04E-02
XYLT1	xylosyltransferase 1	16	-1.21	2.06E-02
NAPSB	napsin B aspartic peptidase pseudogene	19	1.82	2.06E-02
SLC13A2	solute carrier family 13 member 2	17	4.62	2.34E-02
STRIP2	striatin interacting protein 2	7	-1.78	2.34E-02
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	2	2.16	2.34E-02
GDF15	growth differentiation factor 15	19	2.69	2.42E-02
CIART	circadian associated repressor of transcription	1	1.91	2.54E-02
DEPP1	DEPP autophagy regulator 1	10	2.82	2.56E-02
BMERB1	bMERB domain containing 1	16	1.82	2.56E-02
SIX4	SIX homeobox 4	14	1.77	2.78E-02
ENSG00000286519	novel transcript	2	2.86	2.78E-02
FGR	FGR proto-oncogene Src family tyrosine kinase	1	1.64	2.83E-02
P2RY12	purinergic receptor P2Y12	3	2.07	2.92E-02
ENSG00000286722	novel transcript antisense to AHCYL2	7	-1.75	2.92E-02
ENPP1	ectonucleotide pyrophosphatase/phosphodiesterase 1	6	1.65	3.09E-02
TFPI	tissue factor pathway inhibitor	2	3.64	3.12E-02
NFE2L3	NFE2 like bZIP transcription factor 3	7	2.06	3.48E-02
HCK	HCK proto-oncogene Src family tyrosine kinase	20	1.57	3.48E-02
CNGA3	cyclic nucleotide gated channel subunit alpha 3	2	4.05	3.48E-02
IL3RA	interleukin 3 receptor subunit alpha	X	2.21	3.48E-02
MTCO3P12	MT-CO3 pseudogene 12	1	2.99	3.48E-02
INF2	inverted formin 2	14	1.89	3.48E-02
LBH	LBH regulator of WNT Signalling pathway	2	1.27	3.49E-02
CDH4	cadherin 4	20	3.59	3.54E-02
PLEK	pleckstrin	2	1.92	3.69E-02
PLA2G7	phospholipase A2 group VII	6	-1.93	3.69E-02
CCL4	C-C motif chemokine ligand 4	17	2.11	3.69E-02
PI3	peptidase inhibitor 3	20	1.94	4.19E-02
MARCHF1	membrane associated ring-CH-type finger 1	4	1.58	4.19E-02
ENSG00000230387	novel transcript	20	-1.87	4.19E-02
SMR3B	submaxillary gland androgen regulated protein 3B	4	2.48	4.41E-02
GSDME	gasdermin E	7	1.62	4.49E-02

OSM	oncostatin M	22	3.27	4.96E-02
INSYN2A	inhibitory synaptic factor 2A	10	-3.28	4.96E-02

Appendix I

Appendix I Table. List of the 36 genes returned as differentially expressed by DESeq2 for the XFS and control comparison

GeneID	Description	Chr	LLog2FC	P-adj
PTGS2	prostaglandin-endoperoxide synthase 2	1	5.31	9.89E-04
EGR3	early growth response 3	8	4.21	9.89E-04
CSF2RB	colony stimulating factor 2 receptor subunit beta	22	2.36	2.33E-03
EGR2	early growth response 2	10	5.17	2.37E-03
MS4A1	membrane spanning 4-domains A1	11	4.1	1.29E-02
KRT6A	keratin 6A	12	2.64	1.29E-02
EGR1	early growth response 1	5	5.56	1.61E-02
IGHA1	immunoglobulin heavy constant alpha 1	14	6.53	2.21E-02
MXRA5	matrix remodeling associated 5	X	2.51	2.28E-02
IL33	interleukin 33	9	3.17	2.28E-02
PTAFR	platelet activating factor receptor	1	1.91	2.28E-02
SLC26A9	solute carrier family 26 member 9	1	4.61	2.28E-02
SLC24A3	solute carrier family 24 member 3	20	3.03	2.82E-02
MUC3A	mucin 3A cell surface associated	7	2.56	2.87E-02
CCL4L2	C-C motif chemokine ligand 4 like 2	17	3.58	2.87E-02
CXCL8	C-X-C motif chemokine ligand 8	4	2.93	2.90E-02
TOX3	TOX high mobility group box family member 3	16	2.81	3.02E-02
TUSC3	tumor suppressor candidate 3	8	3.24	3.02E-02
IL1B	interleukin 1 beta	2	3.28	3.02E-02
SERPINF1	serpin family F member 1	17	2.19	3.02E-02
DUSP2	dual specificity phosphatase 2	2	2.19	3.02E-02
PTP4A3	protein tyrosine phosphatase 4A3	8	1.92	3.02E-02
CEMIP	cell migration inducing hyaluronidase 1	15	2.68	3.05E-02
KRT23	keratin 23	17	1.87	3.05E-02
NEK2	NIMA related kinase 2	1	4.3	3.05E-02
JCHAIN	joining chain of multimeric IgA and IgM	4	3.73	3.05E-02
SCGB3A1	secretoglobin family 3A member 1	5	3.45	3.05E-02
CFAP46	cilia and flagella associated protein 46	10	-1.39	3.05E-02
CHST1	carbohydrate sulfotransferase 1	11	3.19	3.05E-02
IL3RA	interleukin 3 receptor subunit alpha	X	2.28	3.05E-02
IGKC	immunoglobulin kappa constant	2	5.71	3.05E-02
TNF	tumor necrosis factor	6	2.69	3.05E-02
CCL3	C-C motif chemokine ligand 3	17	3.64	3.05E-02
CDH4	cadherin 4	20	4.42	3.32E-02
FOSB	FosB proto-oncogene AP-1 transcription factor subunit	19	3.63	4.71E-02
TNFSF9	TNF superfamily member 9	19	2.7	4.89E-02

Appendix J

Appendix J Table. List of the 23 genes returned as differentially expressed by DESeq2 for the male and female comparison

GeneID	Description	Chr	LLog2FC	P-adj
NLGN4Y	neuroligin 4 Y-linked	Y	-9.67	3.18E-45
ZFY-AS1	ZFY antisense RNA 1	Y	-8.71	9.88E-44
LINC00278	long intergenic non-protein coding RNA 278	Y	-8.32	9.14E-41
KDM5D	lysine demethylase 5D	Y	-9.75	3.42E-39
EIF1AY	eukaryotic translation initiation factor 1A Y-linked	Y	-9.63	9.85E-37
TTTY14	testis-specific transcript	Y	-9.55	3.66E-35
DDX3Y	DEAD-box helicase 3 Y-linked	Y	-9.39	2.26E-32
UTY	ubiquitously transcribed tetratricopeptide repeat containing	Y	-8.77	1.29E-31
ZFY	zinc finger protein Y-linked	Y	-8.65	3.03E-31
TMSB4Y	thymosin beta 4 Y-linked	Y	-7.72	9.01E-30
USP9Y	ubiquitin specific peptidase 9 Y-linked	Y	-8.79	5.56E-28
ANOS2P	anosmin 2	Y	-8.33	6.24E-26
RPS4Y1	ribosomal protein S4 Y-linked 1	Y	-9.48	5.20E-24
PRKY	protein kinase Y-linked (pseudogene)	Y	-7.4	3.76E-18
ENSG00000260197	novel transcript	Y	-5.75	1.29E-16
BCORP1	BCL6 corepressor pseudogene 1	Y	-5.64	4.10E-16
ZNF736P9Y	zinc finger protein 736 pseudogene 9	Y	-5.34	3.89E-12
TTTY10	testis-specific transcript	Y	-5.12	6.24E-10
TBL1Y	transducin beta like 1 Y-linked	Y	-4.9	2.86E-08
ENSG00000278847	novel transcript	Y	-4.75	5.50E-08
TSIX	TSIX transcript	X	4.88	4.76E-06
RFTN1P1	raftlin	Y	-2.45	8.34E-03
XIST	X inactive specific transcript	X	5.11	1.97E-02

Appendix K

Link to the paper below:

https://www.sciencedirect.com/science/article/pii/S0014483523002348

Hulley, MR; Ngcungcu, N; Ramsay, M and Williams, S. 2023. Non-invasive harvesting of conjunctival cells for whole transcriptome sequencing. *Experimental Eye Research*. 234. 109613. Available at: https://doi.org/10.1016/j.exer.2023.109613

Appendix L

Turnitin Plagiarism Report Cover Page and Support Letter from Supervisors:

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 I PO Box 1038, Johannesburg, 2000
[T] +27 11 489 9223 I [M] +27 78 080 8841 I [F] +27 11 489 9226 I [E] human.genetics@nhls.ac.za

07/11/2023

To the Faculty of Health Sciences,

RE: Turnitin Originality Report

The thesis of PhD candidate, Michaella Robyn Hulley (Student number: 1820405) was submitted to Turnitin to check for plagiarism. The Originality Report from Turnitin provided a similarity score of 24%, with 6% coming from a single source. This source is a paper published by the PhD candidate using the results of this study.

We have examined both the report and the full thesis with the similarity overlay and agreed that this is not an indication of plagiarism.

Regards,

Prof Susan Williams | Dr Thandiswa Ngcungcu | Prof Michele Ramsay

This report is intended solely to record the observations and/or opinion of the writer. It does not constitute a medico-legal report

Chairperson: Prof Eric Buch CEO: Dr Karmani Chetty
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X8, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6000/ 0860 00 NHLS(6457) www.nhls.ac.za
Practice number: 5200296

MHulley Thesis .docx

ORIGINALITY REPORT

24%	16%	19%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Michaella Hulley, Thandiswa Ngcungcu, Michele Ramsay, Susan Williams. "Non-invasive harvesting of conjunctival cells for whole transcriptome sequencing", Experimental Eye Research, 2023 Publication	6%
2	Inas F. Aboobakar, William M. Johnson, W. Daniel Stamer, Michael A. Hauser, R. Rand Allingham. "Major review: Exfoliation syndrome; advances in disease genetics, molecular biology, and epidemiology", Experimental Eye Research, 2017 Publication	1%
3	www.ncbi.nlm.nih.gov Internet Source	1%
4	www.biorxiv.org Internet Source	1%
5	Submitted to University of Witwatersrand Student Paper	1%
6	core.ac.uk Internet Source	<1%

Appendix M

Session Information for R (DGE and pathway analyses):

R version 4.2.3 (2023-03-15 ucrt)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 19045)

Matrix products: default

locale:

[1] LC_COLLATE=English_United States.utf8 LC_CTYPE=English_United States.utf8
LC_MONETARY=English_United States.utf8

[4] LC_NUMERIC=C LC_TIME=English_United States.utf8

attached base packages:

[1] stats4 grid stats graphics grDevices utils datasets methods base

other attached packages:

[1] PROPER_1.30	gageData_2.36.0	gage_2.48.0	pathview_1.38.0
[5] org.Hs.eg.db_3.16.0	AnnotationDbi_1.60.2	genefilter_1.80.3	PoiClaClu_1.0.2.1
[9] UpSetR_1.4.0	DESeq2_1.38.3	SummarizedExperiment_1.28.0	Biobase_2.58.0
[13] MatrixGenerics_1.10.0	matrixStats_0.63.0	GenomicRanges_1.50.2	GenomeInfoDb_1.34.9
[17] IRanges_2.32.0	S4Vectors_0.36.2	BiocGenerics_0.44.0	BiocManager_1.30.20
[21] kableExtra_1.3.4	enrichR_3.1	xtable_1.8-4	ggrepel_0.9.3
[25] xml2_1.3.3	gtable_0.3.3	lattice_0.20-45	gridExtra_2.3
[29] stringr_1.5.0	knitr_1.42	dplyr_1.1.3	RColorBrewer_1.1-3
[33] pheatmap_1.0.12	ggplot2_3.4.4	biomaRt_2.54.1	

loaded via a namespace (and not attached):

[1] bitops_1.0-7	bit64_4.0.5	filelock_1.0.2	webshot_0.5.4	progress_1.2.2
[6] httr_1.4.6	Rgraphviz_2.42.0	tools_4.2.3	utf8_1.2.3	R6_2.5.1
[11] DBI_1.1.3	colorspace_2.1-0	withr_2.5.0	tidyselect_1.2.0	prettyunits_1.1.1
[16] bit_4.0.5	curl_5.0.0	compiler_4.2.3	graph_1.76.0	cli_3.6.0
[21] rvest_1.0.3	DelayedArray_0.24.0	labeling_0.4.2	KEGGgraph_1.58.3	scales_1.2.1
[26] rappdirs_0.3.3	systemfonts_1.0.4	digest_0.6.31	rmarkdown_2.21	svglite_2.1.1
[31] XVector_0.38.0	pkgconfig_2.0.3	htmltools_0.5.4	dbplyr_2.3.2	fastmap_1.1.1
[36] rlang_1.1.0	rstudioapi_0.14	RSQLite_2.3.0	farver_2.1.1	generics_0.1.3