What constitutes a healthy eye?

Investigations of genetic and environmental influences in eye disease

Seyhan Yazar

BMed Sci, MOrth

This thesis is presented for the degree of Doctor of Philosophy of Seyhan Yazar of the University of Western Australia.

Centre for Ophthalmology and Visual Science

2015
Declaration

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.
Summary

With its unique tissue composition and easy-to-measure physiological characteristics (ocular biometry), the eye serves as a window for discovering the genetic and environmental factors that determine its structure and status in health and disease. Revolutionary findings through multiple omic technologies have already allowed us to collect vast amounts of ocular data at an individual level (Chapter 2). A major bottleneck of biological discovery using this immense volume of data is emerging at the computational level. Cloud computing is confirmed to be an efficient and cost-effective alternative for analysis of large genomic datasets in small and medium-sized research laboratories with limited computational capacity (Chapter 3). This finding is promising especially given that data collection and analysis are the basis for individualised ophthalmic care, which requires tailoring of treatment and preventative measures according to individual genomic characteristics and environmental influences. The work required for individualising ophthalmic care has many steps and can be described in layers similar to a geographic information system. The bottom layer of this model consists of identifying genetic factors that cultivate ocular disease. Next layers contain various studies that build a bridge between a genotype and a phenotype. One of these layers is determining environmental exposures from conception onwards. The objective of the present thesis is to determine the genetic and environmental risk factors relating to eye disease.

One method to achieve this goal is pooling data from large population-based cohorts longitudinally and data mining through various techniques. With this in mind, the 20-year follow-up the Western Australian Pregnancy Cohort (Raine) Eye Health Study was
conceived to collect multi-scale data to characterize ocular biometric parameters and determine the baseline prevalence and risk factors of common eye diseases in a young adult population born in Western Australia. Dissecting components of complex ocular diseases relies on phenotypic variation that necessitates accurate phenotyping in assembly of such large volume of data. Therefore, standardized methodology and practical guidelines were described at the beginning of this work to utilize in current and future eye research (Chapter 4).

In the 20-year-old individuals, the prevalence of most eye diseases was low, except myopia. This was in line with the increased prevalence of myopia worldwide, highest in East-Asian countries. Although simple myopia itself is a non-sight threatening condition and can be corrected with spectacles, contact lenses and refractive surgery, high myopia is commonly associated with debilitating eye diseases such as glaucoma, maculopathies and retinal detachment. Hence, any preventative measure or treatment modality will have benefits not only at individual level but also for the ocular health care system.

Myopia is due to imbalance of ocular biometry, mainly axial length and corneal curvature. However, higher levels of monochromatic aberrations were found to be associated with increased severity of myopia in young adults. This suggests that while some biometry has a major role in myopia occurrence, other parameters have a small and substantial effect (Chapter 5).

Disproportionate ocular biometry in myopia is a consequence of interactions between the genes and the environmental risk factors an individual is exposed to at various stages of life including the perinatal period and early childhood. We confirmed that
paediatric has no effect on levels of myopia in young adulthood (Chapter 6). Time spent outdoors during childhood is suggested to be protective against myopia development, although the exact mechanism of this effect remains unknown. One of the challenges of studies investigating these mechanisms is that ocular sun exposure cannot be measured directly and quantitatively. Therefore biomarkers representing this measurement need to be identified and utilised instead. Conjunctival UV autofluorescence (CUVAF) can be an ideal biomarker for this purpose. However, the following characteristics must be considered when used: [1] prevalence of CUVAF increases with increasing latitude (toward equator); [2] although it is largely environmental, genes also play a significant role in its development. Notably a single polymorphism (SNP) in solute carrier family 1, member 2 (SLC1A5) gene is associated with CUVAF, suggesting that some individuals have higher risk than the others because of their genetic composition (Chapter 7). Vitamin D pathways were hypothesised to be involved in myopia development. Myopic individuals were found to have lower levels of serum vitamin D. Although this association could be evidence for an underlying mechanism, it could be simply a biomarker of whole-body sun exposure (Chapter 8).

Although the relationship of corneal curvature and axial length is the main driver of the refractive status and myopia development in childhood, uncorrected corneal astigmatism may inhibit emmetropisation and cause myopia during ocular growth. Hence dissecting the genetics of corneal curvature and corneal astigmatism, which are highly heritable, may help us to understand uncoordinated growth of refractive components in myopia. A limited number of genetic studies of corneal astigmatism and corneal curvature have been performed. The only published genome-wide association study (GWAS) for corneal astigmatism was from a Singaporean Asian population. We were unable to replicate the reported association of corneal astigmatism and the platelet-
derived growth factor receptor alpha (PDGFRA) locus in Australians of Northern European ancestry. The same variant has also been implicated in corneal curvature in the same Singaporean population. We searched for the same variant in our cohorts and were able to replicate previous findings (Chapter 9).

In this research, a number of important associations with myopia are identified utilising mainly a cohort of healthy young adults. Although translation of these findings requires extensive efforts and completion of multiple stages, the milestones achieved in other complex diseases give us the hope that similar advancements are feasible for myopia. However, now, the next phase of studies should focus on replication studies in other cohorts to validate these findings.
Preface

This thesis is presented in ten chapters and consists of the following components: a general introduction (Chapter 1), a review chapter on large and complex datasets (big data) in ophthalmology (Chapter 2), a computational experiment chapter on analysis of genomic datasets (Chapter 3), a methodology chapter describing the main cohort study utilised in this thesis project (Chapter 4), five analysis chapters exploring various genetic and environmental factors involved in ocular health and disease (Chapter 5-9) and a general discussion (Chapter 10). The computational experiment chapter as well as the five analysis chapters are presented as a series of papers and include background, methods, results and discussion sections. Each chapter is formatted to allow for continuity and to comply with the University of Western Australia Graduate Research School guidelines.
Contents

Declaration ........................................................................................................................................ ii
Summary ........................................................................................................................................ iii
Preface ........................................................................................................................................... vii
Contents ......................................................................................................................................... viii
Acknowledgements ................................................................................................................... ix
Statement of contribution ........................................................................................................... xiii
Publications arising from this thesis ........................................................................................ xiv
Abbreviations ............................................................................................................................. xviii
Chapter 1 Introduction ............................................................................................................... 1
Chapter 2 Omics and Big Data in Ophthalmology: Current State and Future Play ............... 8
Chapter 3 Benchmarking Undedicated Cloud Computing Providers for Analysis of Genomic Datasets ................................................................. 35
Chapter 4 Raine Eye Health Study (REHS) .............................................................................. 50
Chapter 5 Comparison of Monochromatic Aberrations in Young Adults with Different Visual Acuity and Refractive Errors .............................................................................. 72
Chapter 6 Early Anaesthesia Exposure and the Effect on Visual Acuity, Refractive Error, and Retinal Nerve Fiber Layer Thickness of Young Adults .............................................. 89
Chapter 7 Genetic and Environmental Factors in Conjunctival UV Autofluorescence 102
Chapter 8 Association of Myopia and Vitamin D Status in Young Adults ........................... 118
Chapter 9 Genome-Wide Association Studies of Corneal Astigmatism and Curvature in Australians with Northern European Ancestry ......................................................... 138
Chapter 10 Discussion .............................................................................................................. 165
Bibliography ............................................................................................................................. 178
Appendices ............................................................................................................................... 209
Acknowledgements

In a doctoral research by definition the student is the only person working on the precise topic of the thesis thus this journey is often considered a solitary experience. I hope this acknowledgement section will prove the opposite. It is my pleasure to thank all those who made this experience joyful and enlightening.

First and foremost, I owe my deepest gratitude to both my supervisors Professor David A. Mackey and Associate Professor Alex W. Hewitt. Without their continuous enthusiasm, encouragement and support, this thesis could not have been completed. Their guidance complemented each other so well that I felt supported throughout all aspects of my journey. Both have many characteristics as scientists and clinicians that I truly admire and hope to match one day in my career.

This thesis was mainly based on the data collected during the Raine Eye Health Study (REHS). Being the coordinator of the study and the clinical examiner of the participants it was always at the forefront of my mind that this study could never proceed if not for goodwill of our participants. So I would like to acknowledge the hundreds of participants and their parents who voluntarily committed their time to this study.

Maintaining and tracing a cohort over many years and undertaking data collection at follow-ups requires an immense team effort. I would like to thank the Raine Study Manager Ms Jenny A. Mountain for her leadership and efforts to retain the cohort over a decade and helping me with all the administrative work. I thank the Cohort Coordinator Mrs Diane Wood and all the Raine research assistants for their help during
data collection. I am also grateful to the Raine Study Data Manager Mrs Angela Jacques
who attended to my data requests immediately.

The eye team of the REHS was equally large and devoted during our examination days.
Particularly, I would like to thank Dr Hannah Forward, Dr Charlotte M. McKnight, Dr
Alex X. Tan, Mrs Alla Soloshenko and Mrs Sandra K. Oates for their contributions to
the ocular examinations of the participants.

I am grateful to Associate Professor Craig E. Pennell for allowing us to utilise the Raine
Study genotypic data and Mr Wei Ang for helping out with its management and
analysis. I also thank the Raine Study Scientific Director, Professor Peter Eastwood and
the Associate Scientific Director, Professor Leon Straker for their support.

I would like to express my sincere thanks to Associate Professor Stuart MacGregor for
hosting me at the QIMR Berghofer Medical Research Institute for a week at the start of
my PhD. This gave me the opportunity to understand the basic statistical techniques in
genetic analysis. I am also thankful to Mr Aniket Mishra and Mr Gabriel Cuellar-
Partida for sharing their knowledge which resulted in several joint publications.

The Raine Study investigators bring expertise from 25 distinct areas of research. Some
of the work included in this thesis established new collaborations between various
Raine investigators. I am very thankful to Professor Wendy H. Oddy and Dr Lucinda J.
Black who provided the vitamin D data and valuable comments during its analysis. I am
indebted to Professor Robyn M. Lucas for her constructive comments and suggestions
during the preparation of our manuscript. She allowed to me step outside of my comfort
zone and challenged me to be more a rigorous scientist. Similarly, I am grateful to
Professor Britta S. von Ungern-Sternberg and Dr Caleb Ing for providing me the early anaesthetic exposure data and sharing their knowledge in paediatric anaesthesiology.

I am grateful to Professor Nicholas G. Martin and Professor Grant W. Montgomery for permission to use the data from the Brisbane Adult Twin Studies (BATS). Mrs Lisa S. Kearns was the research orthoptist who dedicated most of her weekends to examine the BATS and Twins Eye Study Tasmania participants. I would like to thank her for her commitment to our group and hard work.

I would like to thank Professor Minas T. Coroneo for providing us with the camera system utilised in measuring ocular sun exposure as part of our examination protocol and sharing his knowledge in ophthalmohelioses when interpreting the results involving this dataset.

It was a pleasure to meet Professor Chris Hammond and Professor Terri Young during their one-month stay in Australia as the Raine Visiting Professors. Discussions with them on complex ocular diseases have been very illuminating and I thank them for their teaching.

I appreciate the feedback offered by Dr Justin C. Sherwin during the preparation of two manuscripts included in this thesis, particularly for providing his expert knowledge in epidemiological studies.

I am grateful to Mr George E.C. Gooden (aka G2) for this patience during our discussions on the cloud computing project. With his help, I have learnt a lot about computer science in the past two years. He has also been a great friend in the office.
I am thankful to Mrs Hilary Salisbury for helping me with all the administrative work and organising my conference travels. She has also been a great supporter of mine from the very beginning.

Past and present many wonderful colleagues and friends have made life fun while working towards this thesis. A special warm thank you to Dr Maria Franchina for her enormous everyday help in work and life, never ending words of encouragement and wise advice throughout this journey. I am particularly grateful to Ms Lisa Booth for being there at difficult times as well as for all the funny and great moments we have had at lunchtimes and coffee breaks. Also my thanks to Mrs Tracey Dickens, Dr Anne-Marie Yardley, Ms Kate Hanman, Ms Emily Huynh, Dr Karim Johnson and Dr Joos Meyer for all their support and friendship.

I am greatly indebted to Mrs Lori Bonertz for her proofreading which has profoundly improved the composition of this thesis.

This work was supported by scholarship funding from The University of Western Australia and a Raine Study PhD Top-Up Scholarship for which I am extremely grateful.

Last, but certainly not least, I would like to thank and dedicate this work to my parents, Nazmi and Suheyla and my siblings Reyhan, Orhan and Ferihan for all their love, support and trust in me while I pursued my dreams. Despite our geographical distance, they were always nearby and always caring.
Statement of contribution

This thesis incorporates the publications generated from joint research projects undertaken in collaboration with a number clinicians and scientists under the supervision of Professor David A Mackey and Associate Professor Alex W Hewitt.

The collaboration is covered in Chapters 3 through 9 of the thesis. Each chapter includes at least one publication I have contributed as a first author (Chapter 3, 4, 5, 6, 8) or a joint first author (Chapter 7 & 9) according to Vancouver Protocol. Detailed contributions of each author are explained in the following paragraphs.

Computational experiments in the Chapter 3 publication were conceived and designed by Assoc Prof Hewitt and me. George GE Gooden and I performed the experiments and analysed the data. Prof Mackey and Assoc Prof Hewitt provided the resources required. All the authors contributed to the writing of the manuscript.

In the Chapter 4 publication, Prof Mackey, Assoc Prof Hewitt and I developed the key ideas and the primary structure of the manuscript. The contribution of other co-authors was primarily through the provision of feedback and revision for intellectual content of the manuscript.

Data utilised in the Chapter 5 publication were collected by all the authors. Assoc Prof Hewitt and I developed the analysis plan. I completed the statistical analysis. All co-authors contributed to the preparation of the manuscript.
Prof Britta von Ungern-Sternberg and Dr Caleb Ing collected part of the data included in Chapter 6 publication around 15 years ago. Prof Mackey, Assoc Prof Hewitt, and Dr Hannah Forward were involved in collection of ocular data. Prof Mackey and I generated the hypothesis. I designed the analysis plan, executed statistical analysis and interpreted the data. All the authors contributed to the writing of the manuscript.

The publication included in Chapter 7 was conceptualised and designed by me, Gabriel Cuellar-Partida, Assoc Prof Hewitt, Assoc Prof Stuart MacGregor and Prof Mackey. Dr Charlotte McKnight and Miss Quach-Thanissorn completed the analysis of the images. Mr Cuellar-Partida and Assoc Prof MacGregor performed the statistical analysis. Prof Mackey, Prof Coroneo, Assoc Prof Pennell, Assoc Prof MacGregor and Mrs Mountain provided administrative, technical and material support. All authors contributed to interpretation of the data and critical revision of the manuscript.

The publication in Chapter 8 was conceived and designed by Prof Mackey, Assoc Prof Hewitt and me. Prof Wendy Oddy, Dr Lucinda Black, Mrs Mountain, Prof Mackey, Assoc Prof Hewitt and Dr McKnight were involved in data collection. All the authors contributed to the writing of the manuscript.

Chapter 9 includes two publications. In both publications, Aniket Mishra completed the genetic analysis and all the authors were involved in preparation of the manuscript. The major publication (Title: Interrogation of the platelet-derived growth factor receptor alpha locus and corneal astigmatism in Australians of Northern European ancestry: Results of a genome-wide association study) included in Chapter 8 was developed, designed and interpreted by me, Assoc Prof Hewitt and Prof Mackey.
Publications arising from this thesis


### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>25(OH)D₃</td>
<td>25-hydroxyvitamin D</td>
</tr>
<tr>
<td>AAU</td>
<td>acute anterior uveitis</td>
</tr>
<tr>
<td>ACTH</td>
<td>adrenocorticotropic hormone</td>
</tr>
<tr>
<td>AIC</td>
<td>akaika information criterion</td>
</tr>
<tr>
<td>AMD</td>
<td>age-related macular degeneration</td>
</tr>
<tr>
<td>AO</td>
<td>adaptive optics</td>
</tr>
<tr>
<td>ATR</td>
<td>against-the-rule</td>
</tr>
<tr>
<td>BATS</td>
<td>Brisbane Adolescent Twin Study</td>
</tr>
<tr>
<td>BD2K</td>
<td>Big Data to Knowledge</td>
</tr>
<tr>
<td>BioShare</td>
<td>Biobank Standardisation and Harmonisation for Research Excellence</td>
</tr>
<tr>
<td>CA</td>
<td>corneal astigmatism</td>
</tr>
<tr>
<td>CC</td>
<td>corneal curvature</td>
</tr>
<tr>
<td>CCT</td>
<td>central corneal thickness</td>
</tr>
<tr>
<td>CDVA</td>
<td>corrected distance visual acuity</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variant</td>
</tr>
<tr>
<td>COV</td>
<td>coefficient of variation</td>
</tr>
<tr>
<td>CPU</td>
<td>central processing unit</td>
</tr>
<tr>
<td>CS</td>
<td>contrast sensitivity</td>
</tr>
<tr>
<td>CU</td>
<td>chronic non-infectious uveitis</td>
</tr>
<tr>
<td>CUVAF</td>
<td>conjunctival ultraviolet autofluorescence</td>
</tr>
<tr>
<td>D</td>
<td>diopter</td>
</tr>
<tr>
<td>DEXA</td>
<td>Dual Energy X-ray Absorptiometry</td>
</tr>
<tr>
<td>DZ</td>
<td>dizygotic</td>
</tr>
<tr>
<td>EHR</td>
<td>electronic health record</td>
</tr>
<tr>
<td>EMR</td>
<td>Elastic MapReduce</td>
</tr>
<tr>
<td>ENCODE</td>
<td>Encyclopaedia of DNA Elements</td>
</tr>
<tr>
<td>EST</td>
<td>expressed sequence tags</td>
</tr>
<tr>
<td>eyeGENE</td>
<td>National Ophthalmic Genotyping and Phenotyping Network</td>
</tr>
<tr>
<td>GCE</td>
<td>Google Compute Engine</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information system</td>
</tr>
<tr>
<td>GO</td>
<td>gene ontology</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>HEALS</td>
<td>Health and Environment-wide Association based on Large population Surveys</td>
</tr>
<tr>
<td>HELIX</td>
<td>Human Early-Life Exposure</td>
</tr>
<tr>
<td>HGP</td>
<td>Human Genome Project</td>
</tr>
<tr>
<td>HOA</td>
<td>higher-order aberrations</td>
</tr>
<tr>
<td>HPA</td>
<td>hypothalamic-pituitary-adrenal</td>
</tr>
<tr>
<td>HPP</td>
<td>Human Proteome Project</td>
</tr>
<tr>
<td>HWE</td>
<td>Hardy-Weinberg equilibrium</td>
</tr>
<tr>
<td>I/O</td>
<td>input/output</td>
</tr>
<tr>
<td>IOL</td>
<td>intraocular lens</td>
</tr>
<tr>
<td>IQR</td>
<td>interquartile range</td>
</tr>
<tr>
<td>IRIS</td>
<td>Intelligent Research in Sight</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>IRSAD</td>
<td>index of relative advantage and disadvantage</td>
</tr>
<tr>
<td>JSON API</td>
<td>JavaScript Object Notation Application Programming Interface</td>
</tr>
<tr>
<td>KEMH</td>
<td>King Edward Memorial Hospital</td>
</tr>
<tr>
<td>LC-MS</td>
<td>liquid chromatography mass spectrometer</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography tandem mass spectrometer</td>
</tr>
<tr>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>LIU</td>
<td>lens-induced uveitis</td>
</tr>
<tr>
<td>LOA</td>
<td>lower-order aberrations</td>
</tr>
<tr>
<td>LogMAR</td>
<td>Logarithm of the minimum angle of resolution</td>
</tr>
<tr>
<td>MAF</td>
<td>minor allele frequency</td>
</tr>
<tr>
<td>MMP</td>
<td>matrix metalloproteinases</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MSE</td>
<td>mean spherical equivalent</td>
</tr>
<tr>
<td>MudPIT</td>
<td>multi-dimensional protein identification technology</td>
</tr>
<tr>
<td>MZ</td>
<td>monozygotic</td>
</tr>
<tr>
<td>NEI</td>
<td>National Eye Institute</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>NMDA</td>
<td>N-methyl-d-aspartate</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OCT</td>
<td>optical coherence tomography</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>PC</td>
<td>principal component</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDR</td>
<td>proliferative diabetic retinopathy</td>
</tr>
<tr>
<td>PEX</td>
<td>pseudoexfoliation</td>
</tr>
<tr>
<td>POAG</td>
<td>primary open angle glaucoma</td>
</tr>
<tr>
<td>PVR</td>
<td>proliferative vitreoretinopathy</td>
</tr>
<tr>
<td>Q-Q</td>
<td>quantile-quantile</td>
</tr>
<tr>
<td>QC</td>
<td>quality control</td>
</tr>
<tr>
<td>RA</td>
<td>research assistant</td>
</tr>
<tr>
<td>REHS</td>
<td>Raine Eye Health Study</td>
</tr>
<tr>
<td>RGC</td>
<td>retinal ganglion cell</td>
</tr>
<tr>
<td>RNFL</td>
<td>retinal nerve fiber layer</td>
</tr>
<tr>
<td>ROP</td>
<td>retinopathy of prematurity</td>
</tr>
<tr>
<td>RPE</td>
<td>retinal pigment epithelium</td>
</tr>
<tr>
<td>RPH</td>
<td>Royal Perth Hospital</td>
</tr>
<tr>
<td>RRD</td>
<td>rhegmatogenous retinal detachment</td>
</tr>
<tr>
<td>SAGE</td>
<td>serial analysis of gene expression</td>
</tr>
<tr>
<td>SCORM</td>
<td>Singapore Cohort Study of Risk Factors for Myopia</td>
</tr>
<tr>
<td>SE</td>
<td>spherical equivalence</td>
</tr>
<tr>
<td>SEIFA</td>
<td>socioeconomic index for areas</td>
</tr>
<tr>
<td>SIMES</td>
<td>Singapore Malay Eye Study</td>
</tr>
<tr>
<td>SINDI</td>
<td>Singapore Indian Eye Study</td>
</tr>
<tr>
<td>SLO</td>
<td>scanning laser ophthalmoscopy</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SP2</td>
<td>Singapore Prospective Cohort Study</td>
</tr>
<tr>
<td>SSH</td>
<td>Secure Shell</td>
</tr>
<tr>
<td>T3DB</td>
<td>Toxin-Toxin Target Database</td>
</tr>
<tr>
<td>TEST</td>
<td>Twin Eye Study in Tasmania</td>
</tr>
<tr>
<td>TKI</td>
<td>Telethon Kids Institute</td>
</tr>
<tr>
<td>TM</td>
<td>trabecular meshwork</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>V1</td>
<td>visual cortex</td>
</tr>
<tr>
<td>VCSEL</td>
<td>vertical-cavity surface emitting laser</td>
</tr>
<tr>
<td>VEGAS</td>
<td>versatile gene-based association study</td>
</tr>
<tr>
<td>WIO</td>
<td>waiting for disk input/output</td>
</tr>
<tr>
<td>WTR</td>
<td>with-the-rule</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Visual impairment continues to be one of the leading causes of morbidity and poor quality of life globally. According to latest World Health Organization estimates, 285 million people suffer from visual impairment worldwide. While approximately 246 million of these have low vision, 39 million are estimated to be blind, with a presenting visual acuity worse than 3/60. The majority of eye diseases are found in the elderly and children. Research focussing on these age groups is critical in understanding disease manifestation and progression and for comparing treatment modalities to support policy decisions and evidence-based medicine. However, the underlying causes of the origins of disease may remain unknown without investigating normative data from a general population of young adults. Therefore, findings from an eye health study on young adults can add significantly to our current knowledge of the components of a healthy or disease-free eye.

It is clear that ocular disorders culminate from genetic and environmental factors and their interactions. Approximately 90% of genes identified in the human genome are expressed in one or more of the eye tissues and cell types, making the study of the human eye and related disease a wonderful window for novel genetic discovery. Uncovering genes that predispose individuals to disease should allow for early predictive risk profiling, diagnosis and treatment. Identifying genes through genetic testing will allow more accurate genetic counselling for the affected individuals and their family members. For example, in some types of Usher syndrome, retinal changes
associated with the disease are not present at early stages of life. A precise diagnosis may allow a timely cochlear implantation, which will facilitate normal development of speech.\(^3\) Additionally, genetic testing may help at-risk relatives of the affected individual to modify their career and lifestyle choices. Some individuals may opt for prenatal screening and consider termination of pregnancy, as is possible for retinoblastoma, Leber congenital amaurosis and choroideremia. Although, there is no approved gene therapy or genetic treatment for ocular diseases at the moment, recent studies are promising. For example, a gene therapy vector carrying the \(RPE65\) gene was safely administered to 12 patients with congenital blindness due to \(RPE65\) associated Leber congenital amaurosis. Improvements in retinal and visual functions were observed. In their recent work, the same group showed that second administration of the vector to the contralateral eye was also safe and efficacious.\(^4\) Glaucoma is one of the five leading causes of blindness yet there is no standardized diagnosis definition for the disease. Identification of genes underlying this heterogeneous disease may help to identify susceptible individuals and facilitate early treatment before considerable visual field loss occurs.

Environmental insults also contribute to disease manifestation. For example, congenital cataract is a multifactorial condition and preventing intrauterine rubella infection decreases its incidence.\(^5\) In babies with retinopathy of prematurity (ROP), reduction in the neonatal environmental oxygen alters the retinal blood vessels and leads to retinal detachment, myopia, strabismus and poor visual outcomes. Moreover, it was shown that even without clinical presence of ROP, premature babies may have long-term retinal vascular anomalies of tortuosity and branching.\(^6\) Cigarette smoking during pregnancy is a well-known risk factor for strabismus in the infant.\(^7\) Environmental effects are not limited to the \textit{in utero} and neonatal periods; experiences in childhood and adolescence
can also contribute to adult phenotype. For example, cumulative insults such as long-term ultraviolet (UV) light exposure or corticosteroid treatment can cause cataract in adults.\textsuperscript{8}

Identification of genetic and environmental factors alone is inadequate for determining those at risk of disease for whom there are modifiable factors. There is an increasing need to understand gene-environment interactions. Individuals carrying age-related macular degeneration (AMD) risk genes, the complement factor H (\textit{CFH}) and \textit{LOC387715}, and who smoked were shown to be at higher risk of having the disease.\textsuperscript{9,10} This interaction suggests an environmental intervention is possible for genetically high-risk individuals. High UV exposure has been associated with pterygium\textsuperscript{11}, cataracts\textsuperscript{12} and AMD\textsuperscript{13}. Recently, outdoor activity has been suggested as a protective factor for myopia. Thus one may suppose that myopes should increase sunlight exposure. On the other hand, in addition to osteoporosis, lack of sunlight and consequently vitamin D deficiency has been associated with several cancers.\textsuperscript{14} Therefore, an optimal amount of sun exposure should be determined, especially for genetically high-risk individuals.

Genetic and environmental factors do not always cause overt disease but may induce observable variation between individuals in physiology, morphology and disease susceptibility within the population. By analysing the data from a young healthy population we can understand the variations in ocular biometry and identify the genetic and environmental factors contributing to their distribution. The major ocular traits that were investigated during this thesis project are included in Table 1-1.
<table>
<thead>
<tr>
<th>Ocular traits</th>
<th>Related ocular conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monochromatic aberrations</td>
<td>Myopia</td>
</tr>
<tr>
<td>Corneal curvature</td>
<td></td>
</tr>
<tr>
<td>Spherical error</td>
<td>Astigmatism</td>
</tr>
<tr>
<td>Cylindrical error</td>
<td></td>
</tr>
</tbody>
</table>
An endophenotype is a heritable biomarker or characteristic that is generally associated with a disease but does not manifest itself as a defined symptom of the disease.\textsuperscript{15} For example, myopia is associated with increased axial length and corneal curvature.\textsuperscript{16} Glaucoma, which is associated with an increase in the cup-to-disc ratio, is characteristically associated with raised intraocular pressure.\textsuperscript{17} Reduced central corneal thickness is a risk factor for glaucoma progression.\textsuperscript{18} Some of these traits and their related ocular diseases are highly heritable.\textsuperscript{19,20} This suggests that a substantial proportion of the genetic variation in the various traits and disorders are due to genetic factors.

It is well established that diseases that are associated with quantitative traits that have a statistical normal distribution will have multiple genes that contribute small effect sizes on the phenotype.\textsuperscript{21} Investigating continuous traits in healthy individuals, for instance through genome-wide approaches, may allow us to identify the genetic loci and the associated environmental factors for disease.

The main dataset that was used in this thesis includes eye examinations from a total of 1344 young adults who participated at the 20-year follow-up of the Western Australian Pregnancy Cohort (Raine) Study. Each examination record included approximately 338 different data variables that are used to define the ocular phenotype of each individual. In addition to this comprehensive phenotypic data, genotypic data were available for most of the cohort. Planning and execution of this structured data collection process is described in Chapter 4 in detail.

A single nucleotide polymorphism (SNP) occurs when the DNA is altered at a single position among individuals. It is the most common genetic variation in the human
A genome-wide association study (GWAS) is an examination of large numbers of SNPs from different individuals to identify any association with a disease or a trait. Commonly, two groups of participants are used in a GWAS: individuals with disease being studied and individuals from the same population without the disease. First, each individual’s complete DNA is genotyped. Genotyping can be done by various methods, which vary in experimental time and cost. A DNA microarray (also known as DNA chip) is a tool that searches the known variants (alleles) one at a time. After millions of SNPs are read through the DNA microarray for each individual, these are aligned to each other and surveyed for the most common variation. If a certain SNP is found to be significantly more frequent in individuals with the disease than in those without the disease, then this SNP is called to be “associated” with the disease. The associated SNPs are considered to point to a region of genome that is involved with the risk of disease.

DNA samples from majority of the Raine Study participants had been genotyped with the Illumina Human660W-quad DNA analysis bead chip. This array contains more than 658,000 markers with a median spacing of 2.5kb. Each chip assays four samples and only 200ng of sample DNA is required per sample. The content of this kit was based on HapMap release 23 data. It also contains 110,000 markers that specifically target regions of common copy number variants (CNV) that were recurrently identified in a high-density CNV discovery project across multiple populations. The CNV regions target the least stable 6% of the genome that is the most likely to contain medically relevant CNV regions such as segmental duplications and unSNPable regions lacking SNPs.
Ophthalmic GWAS studies have been successful in combatting the scepticism that GWAS initially attracted because of their assumption that common genetic variation plays a large role in explaining the heritable variation of common disease. The association of the CFH gene and HtrA serine peptidase 1 (HTRA1) with AMD,\textsuperscript{24-26} the lysyl oxidase-like 1 (LOXL1) gene with pseudoexfoliation syndrome\textsuperscript{27} and the transcription factor 4 (TCF4) gene with Fuchs endothelial dystrophy\textsuperscript{28} were all reported from GWAS.

Generally, a case-control design is used in a GWAS to compare data between the healthy group and the disease-affected group. When results from a large number of affected individuals are not available, quantitative phenotypic data can be analysed instead.\textsuperscript{29} This alternative approach was utilised in Chapters 7 and 9.
Chapter 2

Oomics and Big Data in Ophthalmology: Current State and Future Play

Although identification of the genetic and environmental risk factors is an integral part of understanding the architecture of complex diseases, studying these in isolation does not allow adequate appreciation of phenotypic variation. A systems biology approach is needed to investigate the gene-environment interactions and observe these in a functional context. In this chapter, we give an overview of application of this approach in ophthalmology.

Introduction

The art of medicine and ophthalmology are changing. Introduction of “omics” technologies in the late 20th century revolutionised our approach to understanding complex biological systems. Rather than more traditional reductionism, a holistic approach that focuses on interaction of biological systems is now applied in biological research. The core principle underlying this approach is the fact that each of us is biologically unique and no two individuals mirror each other completely, not even monozygotic twins. The rapid developments in the omic technologies and computational frameworks enable us to study each individual in detail. Today, we can collect vast amounts of data for any individual and assemble an integrative map of an individual’s biological content similar to the construction of a geographic information system (GIS). Ultimately, understanding the role of each biological phenomenon and
relationships between them via this system approach will provide individualised medicine and ophthalmic care.

With its unique tissue composition and easy-to-measure physiological characteristics (ocular biometry), the eye serves as a window for discovering factors that determine its structure and status in health and disease. The field of ophthalmology has the potential to be the first medical speciality to achieve individualised care. In this paper, we review and summarise the research from omics assessment and big data analysis that is yielding novel insights into understanding many ophthalmic diseases.

In the first part of this review, we consolidate and synthesise information from various omics fields including genomics, epigenomics, transcriptomics, proteomics and metabolomics. We hope this will allow ophthalmic researchers to navigate up-to-date research and consider use of available databases for future work. Then, we explore many large-scale data-rich fields that are emerging beyond these traditional omics. For example, we discuss how big data from imaging devices, scanners, electronic health records (EHRs) and social media will advance ophthalmic care and allow implementation of “P4 medicine” or healthcare that is predictive, preventive, personalized and participatory. This will require combining and interpreting omics and big data from many sources. Currently available datasets are scattered across various institutions and linkage of these has both technical and social challenges. In the final section of the review, we discuss these challenges in the ophthalmic care setting and propose resolutions that can be adapted by ophthalmic researchers.
Traditional –omic Studies

Genomics

The advent of DNA sequencing technologies and computing power led to the birth of the genomics era in which it became possible to study structure, function and expression of genes. A key milestone in this era was the completion of Human Genome Project (HGP) in 2001. By extensive data sharing and collaboration, a consortium of 20 centres from six countries determined the first sequence of human genome. Following this significant achievement, the very large International HapMap Project was established and catalogued genetic similarities and differences in humans. This project found that the genome between two individuals only differs by approximately 1.0%. These variations occur at various genetic levels, including DNA. Single nucleotide polymorphisms (SNPs) in which a given base (A, G, C, T) is changed, and copy number variations (CNVs) in which a repeated segment of the DNA (e.g. CCTCCTCCT) varies in number, are two types of variation in DNA. SNPs are the more common of the two and occur every 1000 bases. In theory, there are approximately 11 million common SNPs in the human genome of 3.2 billion basepairs. Genome-wide association (GWA) studies are statistical approaches that have revolutionised the field of human quantitative genetics. The basic idea underlying a GWA study is relatively simple: the study population is divided in two groups based on the presence or lack of a certain physical trait or disease, then each of the 11 million known SNPs is considered in an attempt to identify those that are statistically correlated with that binary outcome. The first successful GWA study was published in March 2005 and reported an association of age-related macular degeneration (AMD) with a variation in the gene for CFH, which produces a protein involved in regulating inflammation. This successful work encouraged researchers to search for disease-causing genetic variants in other ocular
diseases. Recently several loci associated with primary open-angle glaucoma (POAG), primary angle-closure glaucoma, refractive error and related ocular biometric markers including axial length and arterial microcirculation were identified.\textsuperscript{37}

The potential to derive meaningful knowledge from the human genome rapidly advanced with “second generation” or more commonly known next-generation sequencing (NGS) technologies, and prompted a reduction in that cost of sequencing at an unpredicted pace. Today, it is possible to sequence a whole human genome for $1000 compared to a $28.8 million cost in 2004.\textsuperscript{30} The high-throughput capacity of the NGS not only enabled scientists to sequence an entire human genome\textsuperscript{38}, but also helped to determine the DNA sequence of tumour cells obtained from cancer patients.\textsuperscript{39} Novel mutations have been identified for many monogenic disorders. For example, we know that there are more than 200 genes involved in inherited retinal and vitreoretinal diseases. (RetNet, https://sph.uth.edu/retnet/) A growing number of laboratories are now embracing NGS and offering genetic testing ranging from multigene disease-specific panels to entire exomes and the complete genomes.\textsuperscript{40,41} Recent evidence shows that compared to the 24% mutation identification rate of traditional genetic testing, the NGS approach can readily detect disease-causing variants in at least half of the individuals with inherited retinal disease.\textsuperscript{42} Similarly the utility of NGS in congenital cataract diagnosis has been reported.\textsuperscript{43} Clinical tests for 25 POAG genes including myocilin (\textit{MYOC}), paired-like homeodomain 2 (\textit{PITX2}), forkhead box C1 (\textit{FOXC1}), paired box 6 (\textit{PAX6}), cytochrome P450, family 1, subfamily B, polypeptide 1 (\textit{CYP1B1}), latent transforming growth factor beta binding protein 2 (\textit{LTBP2}) and their variants have also been offered. Identification of a disease-causing mutation in one of these known genes helps with tailoring clinical care and genetic counselling. It is also envisaged that it may prevent or limit vision loss through early intervention.\textsuperscript{44} Thus, family screening is a
crucial part of the clinical approach to complex diseases such as glaucoma, and families are appreciative of this level of care. In a large family in which MYOC mutations segregated, 26 of 27 individuals tested had a positive attitude towards predictive testing and genetic counselling. Moreover, although no current reliable testing is available, identification of disease-associated SNPs promises prediction, diagnosis and new avenues for therapy of complex ocular diseases such as AMD.

In 2003, the National Eye Institute (NEI) established a national-wide collaborative resource, the National Ophthalmic Genotyping and Phenotyping Network (eyeGENE). This partnership brought together multiple stakeholders including government, ophthalmic care providers, diagnostic laboratories, vision researchers and private industry, and permitted patients, clinicians and researchers access to the genotype and phenotype data of ophthalmic patients. As of end of 2014, there were ten distinct studies using the data generated by eyeGENE network to interrogate genetic causes of ophthalmic diseases (www.nei.nih.gov/eyegene).

**Epigenomics**

The word “epigenetics” means “upon” or “above” the genomics. It refers to the external modifications including DNA methylation, histone modifications and nucleosome positioning that tightly regulate gene activity. Epigenetics refers to the study of these heritable modifications in our genome that occur without altering the DNA or genetic code. Initially it was thought that a person’s epigenotype is determined in the early developmental years and remained much the same throughout lifetime. Then it was observed that global methylation profiles decrease with age. This suggests that exposure to certain environmental factors early in life causes changes which may predispose to disease in adulthood.
Epigenetics is one of the fast growing fields of biology and has been the centre of studies focusing on dissecting the correlation of genotype and phenotype in development and disease. However, investigations of epigenetic modifications in ocular diseases began much later than other complex diseases. The Encyclopaedia of DNA Elements (ENCOD#E) was established to unearth the information beyond the DNA sequence, including epigenetic modifications in the human genome.\textsuperscript{51} Despite such work, only a handful of studies on epigenetic mechanisms of complex ocular diseases can be found. In these studies, epigenetic modifications are mainly implicated in retinal diseases and ocular cancers. For instance, Hunter and colleagues\textsuperscript{52} compared the retinal pigment epithelium and choroid DNA methylation levels in patients with AMD and age-matched controls by using bisulfate pyrosequencing. They found evidence of hypermethylation in the promoter region of the glutathione S-transferase isoform mu 1 (\textit{GSTM1}) gene, which is closely located to two downstream genes (\textit{GSTM1} and \textit{GSTM5}) that were reduced in expression. Wei and colleague’s\textsuperscript{53} twins and siblings design revealed decreased levels of methylation in the promoter region of the interleukin 17 receptor C (\textit{IL17RC}) gene of AMD patients compared to age-matched controls. Moreover, they demonstrated increased \textit{IL17RC} expression in peripheral blood, retina and choroid of patients with choroidal neovascularization and geographic atrophy and proposed methylation of \textit{IL17RC} as a biomarker of AMD. However, these findings could not be validated in a subsequent study. Oliver and colleagues\textsuperscript{54} found no evidence of hypomethylation in the \textit{IL17RC} gene promoter in the peripheral blood samples acquired from independent AMD cohorts and studied through three distinct differential methylation techniques. This suggests that, similar to large GWA studies, epigenetic associations need to be confirmed across multiple independent cohorts before being proposed as a clinical biomarker in disease identification.
Epigenetics holds promising implications for cancer research since its first discovery in 1983. There are already several FDA-approved epigenetic therapies and many more are in the preclinical and clinical trial stages. The complex epigenetic landscape of retinoblastoma is widely appreciated and documented by multiple studies. In a study that set out to investigate the rapid tumorigenesis following retinoblastoma 1 (RB1) gene inactivation, authors ascertained that the expression of spleen tyrosine kinase (SYK) proto-oncogene was highly deregulated in the aberrant methylation environment. They proposed the SYK gene as a new potential therapeutic target for retinoblastoma. As noted by Murphree and Triche, although the findings of this study are very promising, Zhang’s interpretation overlooks the fact that the role of non-coding genome is yet to be discovered. Uveal melanoma is another ocular cancer type where epigenetic mechanisms have been well studied. Many previous studies demonstrated increased methylation in the promoter region of the Ras association domain family 1 (RASSF1) gene. In contrast, in a study investigating seven CpG island promoter regions including RASSF1, Moulin and colleagues identified aberrated methylation levels in the human telomerase reverse transcriptase (hTERT) gene. Despite this discrepancy in epigenetic investigations of uveal melanoma, in a recent study, the FDA-approved demethylation agent 5-aza-20-deoxycytidine (5-Aza) for myelodysplastic syndrome was proposed as a therapeutic agent for uveal melanoma.
Transcriptomics

Transcriptomics, or genome-wide expression profiling, aims to catalogue the complete set of RNA transcripts produced by the genome or a specific cell. Various methods have been used for transcriptomic analyses including hybridisation-based microarrays and Sanger sequencing-based methods.\textsuperscript{67-69} However, it was the introduction of high-throughput technologies that transformed the transcriptomic analysis. They enabled direct sequencing of complementary DNAs, and mapping these reads to the genome (RNA-sequencing) transformed transcriptomic analysis.\textsuperscript{70}

A considerable amount of literature has been published on transcriptomic analyses of human ocular tissues. Reference transcriptomes of many ocular tissues including retina, cornea, lacrimal gland, and trabecular meshwork were assembled similar to the generation of reference human genomes. Among these, the retinal transcriptome is the most studied. Approximately 40\% of the complete retina transcriptome is estimated to be the transcripts of the retinal pigment epithelium (RPE).\textsuperscript{71} Several groups have attempted to generate a comprehensive list of genes expressed in the retina. Initially, generation of expressed sequence tags (ESTs) was very common.\textsuperscript{72-75} In addition to direct sequencing, serial analysis of gene expression (SAGE) is a popular approach for profiling expression in human retina. Sharon et al.\textsuperscript{76} identified a total of 320,998 tags from four SAGE libraries that were prepared from the peripheral retina, macula and RPE of two donors’ eyes. However only a small portion of these genes were characterised as unique transcripts in retina (9.9\%) and RPE (19.4\%). Rickman et al.\textsuperscript{77} used a similar approach and identified expression of 722,576 tags from five human donors. Only 12,394 of the gene clusters overlapped with Sharon et al.’s findings suggesting that a SAGE library created from single donors may have major expression differences. Moreover, Booji et al.\textsuperscript{78} reported 1,976 genes with high inter-individual
variability in six human RPE samples. RNA-sequencing is the most recent method to assess gene expression quantitatively. Farkas and colleagues completed the first characterisation of annotated retinal transcriptome using three normal adult human retinal DNA samples. A total of 79,915 novel alternative splicing events, including 29,887 novel exons, 21,757 3’ and 5’ alternate splice sites, 28,271 exon skipping events as well as 116 potential novel genes were identified in this study. Thus it extended the unprecedented diversity of the human retina transcripts. Li and colleagues performed a whole transcriptome expression analysis on eight disease-free human retina and RPE/choroid/sclera samples by RNA-sequencing. Interestingly, differentially expressed genes were not only identified between the layers of the retina but also between the macular and peripheral regions of a layer. For example, the authors described more than 2051 differentially expressed genes between peripheral and macular retina, and 926 genes between the peripheral and macular RPE/choroid/sclera. This indicates that the complexity of gene expression in the retina is greater than what has been proposed. Thus it also reflects the complexity of the retinal tissue and its unique molecular functions.

The NEIBank created by National Institute of Health is a model resource, which provides a comprehensive overview of current knowledge of the transcriptional repertoires of eye tissues and their relation to pathology. As part of this project, the EyeSAGE database was created through integrating large-scale macula expression data from various technologies including SAGE, longSAGE and cDNA microarrays to provide a better and more accurate reference retinal transcriptome. Although many other groups aimed to generate similar resources, currently the majority of these websites are not accessible.
Generating these databases encourages other researchers to combine and analyse the available datasets. For the first time, Schultz et al.\textsuperscript{82} compiled 500,000 publications and constructed a reference retina transcriptome from 13,037 non-redundant gene transcripts. This compilation reflected up to 90\% of the mammalian retinal expression. In the same way, Ziesel et al.\textsuperscript{83} generated a fovea-macular transcriptome that comprised 6,056 genes using six donor eyes, and combined with the previously published data the transcriptome expanded to 9,197 genes.

As introduced above, previous studies assessing mRNA levels in normal human retina, RPE, and choroid, revealed tissue-specific molecular signatures and differences between macular and extramacular transcript expression. The first systematic transcriptional profiling included comparing the gene expression of 37 AMD eyes with 31 normal eyes.\textsuperscript{84} All the donors were categorised into seven AMD classes (normal, preclinical, subclinical, dry, geographic atrophy [GA], choroidal neovascularisation [CNV] and GA/CNV). Expression profiling was performed for both the neural retina and the combined RPE/choroid layers of the eye using an oligonucleotide microarray platform. These experiments uncovered distinct molecular signatures for each assigned AMD class - termed disease modules - as well as a set of differentially regulated genes shared by all classes. In both the RPE/choroid and retina, genes elevated across all AMD grades were enriched for regulators of cell-mediated immunity. In the RPE/choroid, these included immunoglobulin genes, a number of cytokines, and several CD antigens associated with T cell activation. These findings provide strong support for the concept that the microenvironment in AMD eyes is pro-inflammatory, with
increased numbers and/or activities of leukocytes that may be responsible for injury to resident RPE or choroidal cells. Future studies to differentiate which transcripts are translated directly, or are under post-transcriptional regulation and to identify isoforms are needed.

Although the retinal transcriptome is the most studied ocular tissue to date, several small studies were conducted to describe the gene expression in other tissues. For example, Tomarev and colleagues\textsuperscript{85} constructed gene expression patterns in the human trabecular meshwork (TM) by EST analysis. Liu and colleagues\textsuperscript{86} reported the first detailed description of TM transcriptome using longSAGE approach. The same group compared the gene expression in the TM of POAG patients with publicly available control TM transcriptomic datasets and suggested that endocytic and exosome pathways may be involved in the pathogenesis of POAG.\textsuperscript{87} Pseudoexfoliation (PEX) syndrome is a major risk factor for secondary open-angle glaucoma. Comparison of the differentially expressed genes in anterior segment tissues from patients with and without PEX syndrome revealed that PEX syndrome is mainly related to extracellular matrix metabolism and cellular stress.\textsuperscript{88}

**Proteomics**

Proteomics is the large-scale study of proteins, including their structure and function, within a cell/system/organism. While the genome comprises fewer than 25,000 genes that code the basic biological functions of the proteins, the proteome is estimated to comprise over 1 million proteins with post-translational modifications such as phosphorylation, glycosylation, and acylation.\textsuperscript{89}

Although proteomics has come relatively late to the field of ophthalmology and vision science compared with other areas of medicine\textsuperscript{90}, many studies have now been
conducted to characterise the human proteome in ocular disease and the process. Here, we describe the comprehensive studies that were performed to determine protein expression in human healthy ocular tissues.

The human tear proteome has been characterised using various approaches over the years. de Souza and colleagues conducted the first comprehensive normal tear proteome analysis using a mass spectrometry (MS) - based approach in-gel/in-solution digestion, and identified 491 proteins in a closed-eye tear sample. Of these, 200 proteins were classified as intracellular while 68 were found to be intracellular. However, a recent study by Zhou et al. characterised 1543 unique proteins in a sample pool of tears from four healthy individuals. 239 of these proteins overlapped with the previous set isolated by de Souza et al. The most common proteins in the tear are lysozyme, lactoferrin, lipocalin (LCN-1), sIgA, lacritin, and proline-rich proteins. These comprise more than 90% of the total amount of tear proteins and due to their higher abundance, proteins in low concentrations becomes less detectable. The tear collection method - glass capillary tube vs Schmirmer’s strips, or open-eye vs closed eyes - can also affect the results. The tear protein composition is dynamic, and changes in disease status.

In the initial proteome analysis of corneal tissue, Karring et al. described 138 distinct proteins in 12 disease-free donor corneas. New innovations in MS instrumentation and tissue separation methods have facilitated characterisation of additional proteins. Using a shotgun nano liquid chromatography tandem mass spectrometry (LC–MS/MS) strategy and an LTQ Orbitrap mass spectrometer, Galiacy et al. reported 2070 proteins in five healthy corneas extracted from donors with choroidal melanoma. In a subsequent study, Dyrlund et al. described 3250 proteins in ten donor corneas. Of these, 2737 were found in the epithelium, 1679 in the stroma, and 880 in the endothelial layer.
Access to normal aqueous humour is challenging due to known ethical considerations as obtaining aqueous sample is an invasive procedure. Thus, the aqueous humour proteome is often characterised in small samples collected from individuals undergoing elective cataract surgery who otherwise have no ocular disease. For instance, Richardson et al.\textsuperscript{96} collected samples from 12 patients. To remove the masking caused by the abundant proteins, they depleted the albumin and immunoglobulin G proteins from each sample and fractioned the pooled samples into two groups: albumin-bound and albumin-depleted. Both fractions were then analysed with multi-dimensional protein identification technology (MudPIT). This experiment resulted in annotation of 50 and 12 high-confidence proteins in the depleted and bound aqueous humour fractions, respectively. Using MudPIT technology in undepleted aqueous humour samples, Escoffier et al.\textsuperscript{97} identified 71 unique proteins, which were primarily extracellular. Chowdhury\textsuperscript{98} and colleagues on the other hand, had compared the proteomic composition of 155 patient samples using multiple techniques. Of the 155 samples, 85 were divided into three age-, sex- and protein concentration-matched groups. Samples in each group then depleted from the six most abundant proteins (albumin, transferrin, antitrypsin, haploglobin, IgG, and IgA). A total of 355 proteins were identified in the pool of depleted aqueous humour samples using gel-based liquid chromatography mass spectrometer (LC-MS). Various other methods including protein array, Western blot and ELISA were applied to the remaining 70 samples. In a similar approach, Bennett et al.\textsuperscript{99} depleted albumin and IgG proteins from 10 aqueous humour samples and analysed using multiple proteomic analysis techniques. This strategy identified 198 unique proteins across the entire study.
In the only study on human ciliary body proteome currently published, a total of 2815 proteins were characterised in three samples analysed through LC-MS/MS technique. While a large number of them were common to the ciliary body and plasma, only ~200 were characterised both in ciliary body and aqueous humour.

Approximately 30 to 35 percent of the human lens is protein and most of the rest is water. Soluble α-, β-, and γ-crystallins compose 90% of the structural proteins within in the lens. They are extremely long-lived with almost no protein turnover. This provides the unique opportunity to investigate the post-transcriptional modifications encountered within the lens as a result of disease and ageing.

Similar to aqueous humour, it is difficult to sample vitreous humour for ethical reasons. Hence, the proteomic profile of vitreous humour is often investigated using “surrogate normal tissues” which are collected during vitrectomy, mainly due to macular hole. The most comprehensive of these studies delineated 1205 proteins, 682 of which were novel in the vitreous humour proteome.

The human retina proteome has mainly been described in comparative proteomic studies of a specific disease, such as AMD and diabetic retinopathy. West and colleagues led the preliminary work in establishing a human RPE protein database and characterised 278 proteins in 42 healthy RPE samples. Recently, Zhang et al. reported the proteome of normal human retina in the first hypothesis-free study. Using the LC-MS/MS technique, they identified a total of 3436 nonredundant proteins in five donor retinas. Among these proteins, there were 20 unambiguous isoforms, of which 8 had no previous description in the literature. Moreover, Skeie and Mahajan characterised a mean of 4403 unique proteins in three anatomical structures (fovea,
macula and periphery) of choroid-RPE complex. 671 of these proteins were previously implicated in oxidative stress, inflammation, and the complement cascade pathways of retinal diseases.

Recently, the Human Eye Proteome Project was established to facilitate proteomic studies of the ocular tissues in collaboration with the HUPO Human Proteome Project (HPP). The working group has provided a provisional catalogue of 4842 nonredundant proteins identified using mass spectrometry. Future studies will be conducted in accordance with the stringent guidelines for protein identification and characterization established by the HPP.

**Metabolomics**

Metabolites reflect biochemical activity within living systems in response to pathophysiological stimuli or genetic modifications. Metabonomics and metabolomics are two related fields involving the study of metabolites. Metabonomics refers to the study of changes in patterns of metabolites through time in complex systems whereas metabolomics aims to characterise and quantify each of the metabolites within a complex system sample. Although one uses a more analytical approach and the other focuses on modelling, both terms are used interchangeably.

There are two main techniques used for metabolomics; nuclear magnetic resonance (NMR) spectroscopy and MS. Both techniques have advantages and disadvantages. For example, MS is more sensitive than NMR but it requires a physical or chemical treatment of the sample prior to the analysis. However NMR has one advantage over MS. With the Magic Angle Spinning NMR technique, it is possible to study intact tissues, and extract structural and functional information about the molecules at atomic resolution.
NMR spectroscopy can be explained in three sequential steps. First, a spectrum of all hydrogen-containing metabolites above the concentration specified is collected from a typical biological fluid. Then individual spectra for each metabolite are identified and structures are assigned to each spectrum. The profiles of disease and control groups are compared through pattern-recognition techniques. Once the metabolites are identified, they can be used as biomarkers in detecting disease, drug effects and other stimuli.\(^{111}\) One big challenge of this process is that the number of metabolites in any given system is unpredictable in contrast to human genome with a known number of genes. To address this issue, multiple database sources such as METLIN\(^ {112}\) and the Human Metabolome Database\(^ {113}\) were developed to characterise, identify, and quantify metabolites in human body. Other databases employed in metabolomics can be found in the thorough review of Go and colleagues.\(^ {114}\) Another solution is to examine a defined set of metabolites through a targeted approach. Nevertheless, in contrast to an untargeted or global experimental design, a targeted approach limits the number of metabolite discovery and introduces the probability of bias.\(^ {109}\)

Earlier work on ocular metabolomics focused on animal models. An overview of these studies was performed by Midelfart et al.\(^ {115}\) Metabolomic analysis of human ocular tissue was reported much later. *In vivo* and *in vitro* NMR spectroscopy on human vitreous revealed that lactate is the main metabolite in vitreous.\(^ {116,117}\) Investigations of the vitreous from patients with a variety of vitreoretinal diseases including chronic non-infectious uveitis (CU) and lens-induced uveitis (LIU) revealed that two forms of uveitis can be discriminated based on the metabolic profiles, with a sensitivity of 78% and specificity of 85%. Moreover, combining genetic information improved this segregation sensitivity and specificity to more than 90%. In a recent study, 33 differential metabolites were identified in acute anterior uveitis (AAU).\(^ {118}\) Furthermore,
patients with proliferative diabetic retinopathy (PDR) were found to have higher levels of lactate and glucose, and lower levels of galactitol and ascorbic acid in vitreous compared to non-diabetic patients.\textsuperscript{119}

Metabolomic studies of human retinal physiology and pathology have only been published in the last two years. Osborne and colleagues\textsuperscript{120} identified a panel of individual metabolites that differ between neovascular AMD cases and controls. Pathway analysis revealed that these metabolites are involved in metabolic pathways such as tyrosine and urea metabolism that are known to be associated with AMD pathophysiology. The vitreous samples acquired from rhegmatogenous retinal detachment (RRD) and proliferative vitreoretinopathy (PVR) patients had a total of 31 different metabolites than vitreous samples from disease-free individuals. Of these 31 metabolites, while 26 regulated in the same direction, five biomarkers differed in regulation direction between RRD and PVR.\textsuperscript{121}

The tear film is the most readily accessible body fluid. The aqueous layer of the tear film is made of water, electrolytes and other small molecules. Previously, 41 metabolites in tears were identified through conventional targeted metabolomic analyses. Chen and colleagues\textsuperscript{122} characterised 16 of the previously known metabolites as well as 44 new metabolites, through LS-MS/MS untargeted metabolomic analysis. In another study, amino acid profiles in the tear film were found to be different than those in aqueous humour and plasma.\textsuperscript{123} The same study demonstrated that some amino acid levels in the tear film were altered in ocular surface diseases.

Lipids are a subset of metabolites and the study of lipids or lipidomics has been a specialised area of research under metabolomics owing to lipids’ unique structure and
function that requires distinct analysis methods. Although lipids are a less well-studied class of metabolites compared to proteins, there are numerous lipidomics publications on both human tear film and Meibomian gland, which is the main source of lipids in the tear film. However because different experimental techniques have been used, there is no agreement on the global lipidomes of the two structures. Therefore, instead of summarising the current data, we will refer readers to one of the most recent comprehensive reviews on the topic.124

Keratoconus is the leading cause of corneal transplantation. Targeted analysis of 250 metabolites using LC-MS/MS showed that several metabolic products of oxidative stress are upregulated in human keratoconic eyes compared to normal corneal keratoctyes and fibroblasts.125 However, an earlier work did not show any difference in metabolic markers between the keratoconic and control corneal tissues.126

Although there has been a long-standing interest in metabolomics, ocular metabolomics is still in its infancy. Improved experimental techniques and widespread implementation of untargeted metabolomic approaches will eventually solve the challenges encountered today. Studies integrating metabolomics with genomics will allow us to gain a true understanding of ocular diseases at a systems level.127,128

Emerging -omic Studies and Big Data Resources

Microbiomics

Microbiomics is a rapidly expanding new area of research. The advent of non-cultivatable molecular techniques such as polymerase chain reaction (PCR) and 16S ribosomal DNA sequencing provided a better understanding of human microbiota. These methods have been recently implemented in ocular microbiome studies.
The cultivable microbiome of ocular surface has been well elucidated,\textsuperscript{129} however, whether the ocular surface has a resident microbiota similar to that of other mucous membranes remains a controversy.\textsuperscript{130} While this debate continues on, preliminary results from the Ocular Microbiome Project (http://www.microbiota.org/cgi-bin/ocular/index.cgi) show that there are in fact a wide range of microorganisms residing within the healthy cornea.\textsuperscript{131} This supports the evidence gathered in a deep sequencing study of conjunctival ribosomal DNA from four volunteers. In this study, Dong and colleagues\textsuperscript{132} highlight the diversity of normal ocular microbiota and draw our attention to the presence of unclassified or novel bacteria. A number of comparison studies have been performed to determine whether the ocular surface microbiota has a role in the development of dry eye,\textsuperscript{133} bacterial conjunctivitis,\textsuperscript{134} trachoma,\textsuperscript{86} and Behcet syndrome.\textsuperscript{135}

Previously, local inflammatory responses were linked to the development of glaucoma. Although inflammatory responses outside of the central nervous system were found to be related to neurodegenerative diseases such as Alzheimer’s disease, no association was detected with glaucoma. Interestingly, in a recent study Astafurov and colleagues\textsuperscript{136} identified higher bacterial oral counts in mouthwash specimens from patients with glaucoma compared to control subjects. Furthermore, they showed that administration of low-dose bacterial lipopolysaccharides to glaucoma-induced animal models resulted in acceleration of axonal degeneration and neuronal loss. Replication of these findings and further probing of the mechanisms involved in this relationship may provide new therapeutic avenues for glaucoma and other neurodegenerative diseases as well as identification of a biomarker of disease.
Whether there is a microbomic signature in normal and diseased state ocular tissue is currently unknown. The preliminary studies mentioned here suggest that there are indeed non-culturable microbiome communities affecting ocular health. Further cross-sectional and longitudinal studies are necessary to elucidate these findings and characterise the existing microbiota in ocular tissues.

**Exposomics**

Exposomics is a new concept of screening the environmental risk factors for disease in conjunction with functional genomics to allow us to explore gene-environment interactions. Various projects have been established to collect detailed information on environmental exposures from conception to death. The NIH established the Human Exposome Project (http://humanexposomeproject.com/) that will be analogous to the HGP and to define the exposome. European counterparts of these initiatives are Health and Environment-wide Associations based on Large population Surveys (HEALS) (http://www.heals-eu.eu/) and EXPOsOMICS (http://www.exposomicsproject.eu/). The Human Early-Life Exposure (HELIX) project combines measures from six ongoing birth cohort studies within Europe.\(^{137}\) This project particularly aims to identify early-life exposure to multiple environmental factors and link these with various omics biomarkers and childhood health outcomes by applying new exposure assessment methods. Moreover, the Public Health Exposome Project was established to investigate the effect of cumulative lifelong environmental factors for six health areas in four different domains (http://communitymappingforhealthequity.org/public-health-exposome-data/). The Toxin-Toxin Target Database (T3DB) was designed to collection information on the toxic exposome. Currently, the database includes 3,673 toxins (pollutants, pesticides, drugs and food toxins) that are linked to 2,086 corresponding toxin target entries.\(^{138}\) This resource can be used in clinical metabolism, predicting toxicity and toxin targets as well as educating individuals on toxicology.\(^{139}\) All these
projects are collecting data that will likely be useful in ophthalmology and vision research. Ophthalmology researchers should help define biomarkers and develop reliable measurement tools.

Decreasing exposure to polycyclic aromatic hydrocarbons is a good example highlighting the importance of exposomics. These organic compounds are metabolised by the cytochrome P450 enzyme that is encoded by the CYP1B1 gene, mutations in which are known to cause primary congenital glaucoma. Cigarette smoking is a well-investigated modifiable risk factor and extensive health messages are already in place. Sun exposure, in particular UV light, is another environmental factor that is associated with multiple ocular diseases including pterygia and cortical cataract. Children who spend more time outdoors are significantly less myopic. This finding is supported by other studies. Increased outdoor activity has therefore been proposed as a potential protective mechanism to reduce the incidence of myopia and is currently being investigated by many groups. A challenge of these studies and exposomics is the isolation and assessment of each environment influence. Our group is currently investigating the optimal levels of sun exposure required to protect against myopia, through the benefits of increased outdoor activity, without increasing risk of excessive sun exposure-related systemic and ocular disease. In this matter, one of our main goals is to characterise an objective biomarker of ocular sun exposure and make it available to the research community and for clinical use.

**Image-omics**

Of all the omic profiling, the storage, access and analysis of the patient-specific image-ome is probably the one that most directly affects all practicing eye care providers. Several imaging technologies have developed over the last two decades allowing clinicians to investigate various structures of the eye in detail without the limitations of
traditional ophthalmoscopy. Spectral domain optical coherence tomography (OCT), ultra-wide field retinal imaging based scanning laser ophthalmoscopy (SLO) systems and confocal microscopy are now commonly used in the clinics. Adaptive optics (AO) is a newer technology in imaging. The correction of the aberrations with AO has made it possible to image the normal and diseased retina of the living eye at microscopic resolution. This technology has also been translated into other fields of medicine. Professor Eric Betzig won the 2014 Nobel Prize in Chemistry for the lattice light-sheet microscopy that he developed by combining AO strategies that astronomers and ophthalmologists use, to cancel out similar distortions in the images.

Next-generation ophthalmic imaging technology aims to collate multiple imaging modalities into a single instrument. The prototype of ophthalmic swept source/Fourier domain OCT, using vertical-cavity surface emitting laser (VCSEL) light source technology, integrates ultrahigh speed retinal and choroidal imaging, long depth range anterior segment imaging, and ultralong depth range full eye imaging. If fully developed and commercialised, this technology will reduce the patient assessment time as well as the clinical space requirements.

Ocular imaging has long played an important role in documentation and clinical diagnosis, progression and treatment, and it has always been personalised. However, these images are often stored in databases and file formats that have limited interoperability. Once this challenge is addressed, integration of image-based phenotyping will be an invaluable tool in personalised ophthalmic care.
Digital technology and the internet are now integral parts of our life. We are in the era of power computing for the masses. Health apps in hand-held devices and wearable fitness trackers such as Fitbit and Nike Fuelband trackers are already very popular sources of medical information. Individuals can upload data from exercise machines in their gym, or blood glucose levels from their meter, or their dietary intake can be monitored through shoppers’ loyalty cards. Furthermore, crowdsourcing organisations such as “PatientsLikeMe” (www.patientslikeme.com) and “CureTogether” (www.curetogether.com) allow individuals to compare their treatment, symptoms, and experiences with similar groups of patients. Microsoft has set up a web-based platform (www.healthvault.com) to store and maintain health and fitness information. This platform addresses both individuals and healthcare professionals. An individual can also share a part (some data types) or the whole of his/her health record with another interested individual such as a doctor, a spouse or a parent. In March, the American Academy of Ophthalmology launched the Intelligent Research in Sight (IRIS) registry (www.aao.org/iris-registry/). Ophthalmologists can use the registry as a centralized data repository and reporting tool to gather patient data from electronic health records in order to perform statistical analysis. To date the database contains information on more than 10 million patient visits, and is being used by 5000 ophthalmologists in the United States. These data may not have a direct impact on our understanding of ophthalmic diseases, but will have value when prioritising the treatment options based on expected outcome, timing, cost, risk, and impact on daily life.
Major Challenges in Integration of -omics and Big Data in Ophthalmology

Various datasets we have discussed throughout this paper have tremendous potential to improve our understanding of ophthalmic disease and change clinical diagnostics, treatment, and management. However, this translation requires much effort and teamwork. Many communities acknowledge big data challenges in three categories: volume, variety and velocity.¹⁵⁰

Volume denotes the problem of the infrastructure necessary for storage and computation. The magnitude of this problem can be appreciated from the storage space required for a single integrated personal omics profile of 6.7 terabytes³⁰, which is equivalent to 1.8 x 10⁷ OCT images. Another bottleneck is the computational power that is necessary to analyse this volume of data. Sometimes an analysis may last for hours or days, and hence demands a special computational framework. At this time cloud computing is an attractive solution. It offers flexibility and easy access to big computational resources, especially for small research groups that do not have such capacity in-house.¹⁵¹

The second major challenge is the variety of the data types and their sources. Currently, all data are scattered across the institutions and are not easily accessible. Recently Weber et al.³² tabulated a model integrating various types of data at the individual level. Using this approach, EHRs provide in depth information about an individual’s health status and form the main piece of information in the model. However, this model lacks the time dimension as EHRs are collected repeatedly over many years. Linkage of these datasets seems easy on paper but it brings its own challenges such as data quality and
heterogeneity, protection of confidentiality, and privacy of both the patients/participants and providers. This is also applicable to large studies that pursue various omic mechanisms in the eye disease process. For example, there are 19 different eye disease focussed consortia around the world, some which are country-based, while others are international (Table 2-1). Each of these cohorts collects data on disease-specific biomarkers and clinical information. Even the data harmonisation can sometimes be a problem within these groups as there is no data standardisation. A “one-fits-all” approach is not only impossible but also may limit new discoveries. Hence formulation and adherence of a minimum information guideline may grant ease in data analysis and interpretation, though receiving maximum efficiency from this approach is dependent on compliance of investigators. In comparison with other fields of medicine, phenotyping in ophthalmology is straightforward. Most ocular biomarkers can be quantified through advanced instruments. To provide the best data available, ophthalmic researchers should document methods well and recognise the differences between the technologies. This also necessitates collaboration with industry stakeholders and partnership in advancement of technologies, which recalls the problem of velocity.

Just as the volume and variety of big data are expanding, the pace at which data are gathered, sorted and analysed is also increasing. Chained to this, there is a growing demand on automation of steps in analytical pipelines that bring us to the development of new methods and software for analysing and integrating biomedical data. Certain principles have to be followed to generate more efficient and reliable research tools and to increase productivity. Other challenges include training of researchers who can utilise the big biomedical data effectively, and the privacy concerns relating to the hosting of data sets on publicly accessible servers, as well as issues related to storage of
Table 2-1 | Major Eye Disease Consortia around the World

<table>
<thead>
<tr>
<th>Consortium Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Congenital Stationary Night Blindness Consortium</td>
</tr>
<tr>
<td>2. Consortium for Refractive Error and Myopia</td>
</tr>
<tr>
<td>3. European Consortium for the Early Treatment of Diabetic Retinopathy</td>
</tr>
<tr>
<td>4. European Retinal Disease Consortium</td>
</tr>
<tr>
<td>5. Fuchs Endothelial Corneal Dystrophy Genetics Consortium</td>
</tr>
<tr>
<td>6. International AMD Gene Consortium</td>
</tr>
<tr>
<td>8. International Contact Lens Prescribing Survey Consortium</td>
</tr>
<tr>
<td>9. International Eye Disease Consortium</td>
</tr>
<tr>
<td>10. International Glaucoma Genetics Consortium</td>
</tr>
<tr>
<td>11. Neighborhood</td>
</tr>
<tr>
<td>12. RP1 Consortium</td>
</tr>
<tr>
<td>13. The International ABCR Screening Consortium</td>
</tr>
<tr>
<td>14. The Orbital Disease Consortium</td>
</tr>
<tr>
<td>15. The Retinoschisis Consortium</td>
</tr>
<tr>
<td>16. The Three Continent AMD Consortium</td>
</tr>
<tr>
<td>17. UK Biobank Eye and Vision Consortium</td>
</tr>
<tr>
<td>18. UNITE Human Ocular Inflammation Consortium</td>
</tr>
<tr>
<td>19. Writing Committee for the Cone Disorders Study Group Consortium</td>
</tr>
</tbody>
</table>
data from individuals. Initiatives such as Biobank Standardisation and Harmonisation for Research Excellence in the European Union (BioShare), and its US counterpart Big Data to Knowledge (BD2K), promise assembly of a landscape where these challenges are diminished.  

Conclusions

Medicine and ophthalmology are evolving in a new direction at an unpredicted pace through the use of high throughput technologies and big data sources. The final form of this evolution is personalised medicine and ophthalmic care through individualised health profiling. In this review, we aimed to give a panoramic overview of the most up-to-date research in various omics fields and acquaint ophthalmic researchers with the non-biological data, which is essential in the systemic approach required for personalised ophthalmic care. As mentioned by many cautious minds, we recognise this approach is still at its infancy. However the work described in this paper clearly shows that the seeds for it to flourish are already embedded. The challenges we face in this evolution are at an extent that cannot be belittled. Yet the solutions to these challenges are not impossible either. Transformation and translation of big data into personalised medical and ophthalmic care requires clinicians, scientists and the non-scientific community to work together, with the support of government agencies and the private sector, including information system and pharmaceutical companies. In time, through such partnerships, solutions will emerge and personalised ophthalmic care will become a reality.
Chapter 3

Benchmarking Undedicated Cloud Computing Providers for Analysis of Genomic Datasets*

As discussed in Chapter 2, processing and analysis of datasets requires high computation power that is not available in every laboratory. One solution to this new challenge is cloud computing, which allows researchers to analyse and share data and analysis tools without having infrastructure on site. Multiple providers offer this service; some are dedicated to genomic research but they are most costly. In this chapter, we assess performance of two undedicated but well-established cloud computing providers in order to help researchers make decisions when using these services.

Background

Through the application of high-throughput sequencing, there has been a dramatic increase in the availability of large-scale genomic datasets.156 With sequencing costs decreasing, small and medium-sized laboratories can now easily amass many gigabytes of data. Given this dramatic increase in the volume of data generated, researchers are being forced to seek efficient and cost-effective measures for computational analysis.157 Cloud computing offers a dynamic means whereby small and medium-sized

* A glossary explaining some of the terminology used in this chapter can be found in Appendix 1.4.
laboratories can rapidly adjust their computational capacity, without concern about its physical structure or ongoing maintenance.\textsuperscript{158-161} However, transitioning to a cloud environment presents unique strategic decisions,\textsuperscript{162} and although a number of general benchmarking results are available (http://serverbear.com/benchmarks/cloud; https://cloudharmony.com/), there has been a paucity of comparisons of cloud computing services specifically for genomic research.

We undertook a performance comparison on two established cloud computing services: Amazon Web Services Elastic MapReduce (EMR) on Amazon EC2 instances, and Google Compute Engine (GCE). Paired-end sequence reads of publicly available genomic datasets (\textit{Escherichia coli} CC102 strain and a Han Chinese male genome) were analysed using Crossbow, a genetic annotation tool, on Hadoop-based platforms with equivalent system specifications.\textsuperscript{38,163,164} A standard analytical pipeline was run simultaneously on both platforms multiple times. The performance metrics of both platforms were recorded using Ganglia, an open-source high-performance computing monitoring system.\textsuperscript{165}

\textbf{Methods}

\textbf{Datasets and Analytical Pipeline}

We benchmarked two independent cloud platforms by a single job that completed read alignment and variant calling stages of next-generation sequencing analysis simultaneously. To investigate the impact of data size on undedicated cluster performance, one small (\textit{E. coli} CC102 strain (3 GB SRA file; Accession: SRX003267) and one large (a Han Chinese male genome (142 GB Fastq files; Accession: ERA000005) publicly available genomic dataset were selected for analysis.\textsuperscript{38,163} For each job in this experiment, a parallel workflow was designed using Crossbow. This
workflow included the following four steps: (1) Download and conversion of files; (2) Short read alignment with Bowtie; (3) SNP call with SOAPsnp; and (4) Combination of the results (Figure 3-1). Crossbow was the preferred genetic
Figure 3-1 | Analytical pipeline demarcating each step required to complete the Crossbow job in the cloud
annotation tool in this experiment, as it has built-in support for running via Amazon’s EMR and Hadoop clusters.  

**Cluster construction and architecture**

Instances were simultaneously established on Amazon’s EMR (http://aws.amazon.com/ec2/) and GCE (http://cloud.google.com/products/compute-engine.html). Undedicated clusters were optimized by selecting computational nodes as suggested for Crossbow. Nodes with equivalent specifications were selected for each system (Table 3-1), these being c1.xlarge node in EMR and the closest specification node n1-highcpu-8 in GCE. For the *E. coli* genome, two nodes (one master and one slave) were used on each platform. For the human genome, the cluster was built with 40 nodes (one master and 39 slaves). As GCE did not provide any included storage for each instance, a 128GB drive (the default storage quota provided by GCE) was added for each node. This was at the additional cost of $0.04/GB/Month or $0.000056/GB/Hour (Jan to June 2014).

Each cluster was run using Apache Hadoop, an open-source implementation of the MapReduce algorithm. MapReduce was used to organise distributed servers, manage the communication between servers, and provide fault tolerance allowing tasks to be performed in parallel.

To explore the effect of network activity differences between the platforms, each job was run simultaneously; same day (including weekdays and weekends) and same time. Detailed description of the set up and scripts to run the jobs can be found in Appendix 1.1 and 1.2.
Table 3-1 | Specification of used computational nodes for each system

<table>
<thead>
<tr>
<th></th>
<th>Virtual Cores</th>
<th>Memory (GB)</th>
<th>Included Storage (GB)</th>
<th>Price (USD/Hour)(^)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Elastic Compute Cloud (EC2) + Elastic MapReduce (EMR) [c1.xlarge]</td>
<td>8</td>
<td>7</td>
<td>4x420</td>
<td>$0.640</td>
</tr>
<tr>
<td>Google Compute Engine [n1-highcpu-8]</td>
<td>8</td>
<td>7.2</td>
<td>0#</td>
<td>$0.352</td>
</tr>
</tbody>
</table>

\(^\) Date accessed: April to June 2014; prior to this period, pricing was $0.700 and $0.520 in Amazon and Google respectively.

# for each instance we added the minimum storage quota of 128GB.
**Cluster Monitoring**

In both EMR and GCE, multiple components of cloud infrastructure including central processing unit (CPU) utilisation, memory usage, and network speeds were monitored and recorded for each node using Ganglia. The default setting of Ganglia for distributing incoming requests is multicast mode; however, since EMR and GCE environments do not currently support multicast Ganglia, it was configured in unicast mode (Figure 3-2). The metric output files constructed in .rrd format were converted into .csv format with a Perl script (Appendix 1.3). For comparison between performance and costs between platforms, the Student t-test was undertaken using the statistical software R (R Foundation for Statistical Computing version 3.0.2; http://www.r-project.org/). In the analysis, cost of each run was calculated using current pricing (June 10\textsuperscript{th} 2014); however, all E. coli runs and one human genome run were performed prior to a recent decrease in price on both platforms.
Figure 3-2 | Directions and types of network transfers in our cloud-computing model. There are a variety of different network transfers between the nodes for each of the services in use in our model. Hadoop requires a bidirectional transmission of data between the master node and the slave nodes. This is required to coordinate the parallel processing of the cluster and to allow for data transfer between nodes. Ganglia uses a unidirectional connection from the slave nodes to the master node to transfer the recorded metrics for storage and visualization. The persistent storage (provided by Amazon S3 (Simple Storage Service) or Google Storage, or an alternative method such as an FTP server), is accessed via the master node. The master node uses it to download input files for Crossbow, such as the manifest file and the reference Jar, and to use for persistent storage of the results of the Crossbow job as the instances destroy their storage on termination. Our local computer can also access the persistent storage via the Internet to allow access to upload the input files or to download the results. The local computer needs to access the master node to initiate Crossbow. In EMR, this is replaced by a web interface and a JavaScript Object Notation Application Programming Interface (JSON API). In GCE, the user is required to remotely log in via Secure Shell (SSH) to commence the job.
Results

Wall-clock time for complete mapping and SNP calling differed by 52.9% (95%CI: 27.5-78.2) and 53.5% (95%CI: 34.4-72.6) for E. coli and human genome alignment and variant calling, respectively, with GCE being more efficient than EMR (Figure 3-3 & 3-4). Table 3-2 displays the key metrics for data analysis using both services. The proportion of CPU usage by Crossbow differed between platforms when aligning and SNP calling each genome, with GCE having better utilisation as the genome size increased. There was considerably more free memory on GCE for the smaller E. coli dataset and on EMR for larger human genome runs. The CPU idle percentage, the percentage of time during which the CPU was idle without waiting for disk input/output (I/O), was greater on EMR for the human genome, while CPU waiting for I/O (WIO) was considerably lower on the same platform. The CPU idle and CPU WIO percentages were both significantly higher on EMR for the E. coli genome. The cost of running this Crossbow pipeline on EMR and GCE also differed significantly (p<0.001), with the costs on EMR 257.3% (95%CI: 211.5-303.1) and 173.9% (95%CI: 134.6-213.1) more expensive than GCE for E. coli and human assemblies, respectively. For ~36x coverage of a human genome, at a current sequencing cost of ~US$1000, the median cost for computation on GCE was US$29.81 (range: US$28.86 to US$45.99), whilst on EMR with a fixed hourly rate it was US$69.60 (range: US$69.60 to US$92.80).

Although runtime variability was inevitable and present in both platforms when assembling each genome, GCE had a considerably greater variability with the larger human genome compared to EMR (coefficient of variation (COV)EMR=4.48% vs COVGCE=16.72%). We identified a single outlier in run time on GCE during the human genome analysis. This occurred due to the virtual cluster having a slower average
network connection (1.55MB/s compared to the average of the other GCE clusters of 2.02MB/s) and a higher CPU WIO percentage than the average for the other GCE runs (9.56% versus 3.52%). The variation in cluster performance likely reflects an increase in network congestion amongst GCE servers.

Runtime predictably is an important issue in undedicated cloud computing. The existing workload of the cloud at the time of service usage is one of the main determinants of variability in runtime of undedicated services. In our benchmarking, EMR was more consistent, though slower, in overall wall-clock time compared to GCE. This may suggest that GCE is more susceptible to server congestion than EMR, although service usage data are difficult to obtain.
Figure 3-3 | Comparison of undedicated cloud performance of Amazon Web Services Elastic MapReduce (EMR) on Amazon EC2 instances (panels a & c) versus Google Compute Engine (GCE)(panels b & d) for *E.coli* genome alignment and variant calling. In this two node cluster the total CPU percent for CPU idle (a and b) and waiting for disk input/output (c and d) is displayed. Note the shorter wall clock times for complete analysis on GCE compared to EMR.
Figure 3-4 | Comparison of undedicated cloud computing performances. The panel includes results of Amazon Web Services Elastic MapReduce (EMR) on Amazon EC2 instances (panels a & c) versus Google Compute Engine (GCE)(panels b & d) for human genome alignment and variant calling. In this 40 node cluster the total CPU percent for CPU idle (a and b) and waiting for disk input/output (c and d) is displayed. Note the greater consistency in performance of Crossbow, though generally longer wall clock times for complete analysis, on EMR compared to GCE.
Table 3-2 | **Comparison of performance metrics for genomic alignment and SNP calling.** All times are presented as hr:min:sec and remaining metrics are shown as mean ± standard deviation. * Calculated by paired t-test.

<table>
<thead>
<tr>
<th>Metric</th>
<th>E.coli Genome</th>
<th>Human Genome</th>
<th>p-value*</th>
<th>E.coli Genome</th>
<th>Human Genome</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EMR (n=10)</td>
<td>GCE (n=10)</td>
<td></td>
<td>EMR (n=10)</td>
<td>GCE (n=10)</td>
<td></td>
</tr>
<tr>
<td>Wall clock time (mean)</td>
<td>0:46:30</td>
<td>0:31:50</td>
<td>&lt;0.001</td>
<td>2:58:24</td>
<td>2:14:12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pre-processing short reads time (mean)</td>
<td>0:14:37</td>
<td>0:12:46</td>
<td>0.109</td>
<td>0:07:29</td>
<td>0:06:23</td>
<td>0.116</td>
</tr>
<tr>
<td>Alignment with Bowtie time (mean)</td>
<td>0:07:04</td>
<td>0:05:03</td>
<td>&lt;0.001</td>
<td>1:51:06</td>
<td>1:15:07</td>
<td>0.003</td>
</tr>
<tr>
<td>Calling SNPs with SOAPsnp time (mean)</td>
<td>0:05:05</td>
<td>0:02:51</td>
<td>&lt;0.001</td>
<td>0:35:31</td>
<td>0:29:31</td>
<td>0.033</td>
</tr>
<tr>
<td>Post-processing time (mean)</td>
<td>0:04:51</td>
<td>0:00:57</td>
<td>&lt;0.001</td>
<td>0:01:23</td>
<td>0:01:03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CPU user (mean %)</td>
<td>17.44 ± 1.30</td>
<td>22.31 ± 3.14</td>
<td>&lt;0.001</td>
<td>43.80 ± 1.87</td>
<td>58.05 ± 6.20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CPU idle (mean %)</td>
<td>72.75 ± 1.23</td>
<td>65.76 ± 4.63</td>
<td>&lt;0.001</td>
<td>47.48 ± 2.30</td>
<td>22.17 ± 3.14</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CPU wio (mean %)</td>
<td>3.88 ± 1.06</td>
<td>0.70 ± 0.16</td>
<td>&lt;0.001</td>
<td>1.86 ± 0.19</td>
<td>4.54 ± 1.82</td>
<td>0.001</td>
</tr>
<tr>
<td>Bytes in (MB/sec)</td>
<td>1.15 ± 0.09</td>
<td>2.12 ± 0.42</td>
<td>&lt;0.001</td>
<td>1.58 ± 0.07</td>
<td>2.00 ± 0.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Memory free (GB)</td>
<td>2.19 ± 0.13</td>
<td>6.17 ± 0.42</td>
<td>&lt;0.001</td>
<td>0.91 ± 0.07</td>
<td>0.70 ± 0.03</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Discussion

Our findings confirm that cloud computing is an efficient and potentially cost-effective alternative for analysis of large genomic datasets. Cloud computing offers a dynamic, economical, and versatile solution for large-scale computational analysis. There have been a number of recent advances in bioinformatic methods utilising cloud resources, and our results suggest that a standard genomic alignment is generally faster in GCE compared to EMR. The time differences identified could be attributed to the hardware used by the Google and Amazon for their cloud services. Amazon offers a 2.0 GHz Intel Xeon Sandy Bridge CPU, whilst Google uses a 2.6 GHz Intel Xeon Sandy Bridge CPU. This clock speed variability is considered the main contributing factor to the difference between the two undedicated platforms. It must also be noted that the resource requirements of Ganglia may have had a small impact on completion times.

There are a number of technical differences between GCE and EMR, which are important to consider when running standard bioinformatic pipelines. Running Crossbow on Amazon Web Services was simplified by an established support service, which provides an interface for establishing and running Hadoop clusters. In contrast, there is currently no built-in support for GCE in Crossbow. The current process to run a Crossbow job on GCE requires users to complete various steps such as installing and configuring the required software on each node in the cluster, transferring input data onto the Hadoop Distributed File System (HDFS), downloading results from the HDFS, and terminating the cluster on completion. All of these steps are automatically performed by Crossbow on EMR. Python scripts offering similar functionality for GCE
that Crossbow provides for EMR were created and are available

While our findings confirm that cloud computing is an attractive alternative to the
limitations imposed by the local environment, it is noteworthy that better performance
metrics and lower cost were found with GCE compared to its established counterpart,
Amazon’s EMR. As experienced during this benchmarking, both platforms make
regular price adjustments to their services. This clearly indicates the competition in the
market, and the necessity of developing genomic tools compatible with multiple
platforms to avoid domination of one. Currently, another major limitation of these
services remains at the initial transfer of large datasets onto the hosted cloud
platform.\textsuperscript{171} To circumvent this in the future, sequencing service providers are likely to
directly deposit data to a designated cloud service provider, thereby eliminating the
need for the user to double handle the data transfer.\textsuperscript{172} Once this issue is resolved, it is
foreseen that demand for these services is likely to increase considerably, given the low
cost, broad flexibility and good customer support for cloud services.\textsuperscript{172} The
development of additional tools specific to genomic analysis in the cloud, which offer
flexibility in choice of providers, is clearly required.
Chapter 4

Raine Eye Health Study (REHS)

Prospective cohort studies are a source of potentially tightly organized structured data. This chapter describes the design and methodology utilized to collect both phenotypic and genotypic data from more than 1300 young adults as part of the eye follow-up of an ongoing cohort study. This cross-sectional study is also the core project of this thesis and encompasses the datasets utilised in the following chapters.

Background

Visual impairment is one of the leading causes of morbidity and poor quality of life. Many population studies in ophthalmology have been conducted targeting older adults (aged over 40 years) or paediatric groups (typically 15 years and younger). To date, population-based data on prevalence of ocular disease and distribution of ocular biometry in young adults (i.e. aged between 20 and 40 years) have not been a major research or public health priority, and therefore remain generally undefined. There is a need for precise, up-to-date data on this age group to help guide evidence-based practice and public health resource allocation. Although 82% of the world’s blind are aged 50 and older, younger age groups affected by blindness and visual impairment should be a high priority due to the potential burden of many years ahead living with reduced vision.

Identification of the prevalence and population-attributable risk of diseases is critical to support policy decisions that will provide better health services to those in greatest
need. However, it is also helpful to understand the underlying risk factors and their associations to prevent disease manifestation and progression, thereby reducing their societal impact. Birth cohort studies are valuable for understanding such exposure-disease relationships in a defined population. For example, using a life course epidemiological approach, the 1958 British Birth Cohort Study found that myopia in later life is influenced by prenatal and early life biological, social, and lifestyle influences. Additional advantages of cohort studies include: (1) estimation of distribution, prevalence and incidence of disease in the reference population; (2) identification of risk factor trends over time; and (3) evaluation of relationships among the available variables through both hypothesis generating and testing approaches.

When a combination of genetic and phenotypic information is available in a birth cohort study, it may then be possible to determine whether genetic susceptibility to disease can be offset or exacerbated by particular life-course and socioeconomic trajectories.

**Delivery and Development of the Western Australian Pregnancy Cohort (Raine) Study**

The 20-year review of the Western Australian Pregnancy Cohort (Raine) Study investigated ophthalmic health and established the Raine Eye Health Study (REHS). The Raine Study is one of the largest ongoing prospective cohort studies of pregnancy, childhood, adolescence, and young adulthood. From 1989 to 1991, 2900 pregnant women were recruited at 16–18 weeks’ gestation into a study at King Edward Memorial Hospital (KEMH), Perth, Western Australia. The original study was a randomized clinical trial investigating whether the use of intensive ultrasound and Doppler studies alter pregnancy outcome, in terms of days of neonatal stay and the rate of preterm birth. An obstetrician at KEMH led the trial as a hospital-based research project. Subsequently, a specific research group was founded to incorporate the newborns of the
recruited families into a cohort study. The aim of this cohort study was to determine how events during pregnancy and childhood influence health in later life. The study was initially funded by the Raine Medical Research Foundation at the University of Western Australia. Over the past 24 years the study maintained its existence through successful grant applications to various funding bodies, including The Australian National Health and Medical Research Council. The cohort was evaluated in detail during childhood (at 1, 2, 3, 5, 8 and 10 years), and adolescence (at 14 and 17 years). Trained nurses and research assistants collected information across a variety of specialties including cardiovascular, reproductive, gastroenterology and respiratory medicine under the guidance of investigators from a wide range of disciplines. Each follow-up included one or more questionnaires, as well as physical and psychometric assessments. Figure 4-1 illustrates the broad categories of measurements completed in each follow-up phase. A core set of measurements has been repeated in each follow-up. These include: (1) social and physical activity at 5, 8, 10, 14 and 17 years; (2) detailed dietary assessment at 8, 10, 14 and 17 years; (3) health services utilization at 1, 2, 3, 5, 8, 10, 14 and 17 years; and (4) school assessments at 8 and 10 years. The physical activity measures included fitness and cardiovascular endurance and the International Physical Activity Questionnaire. Similarly, detailed assessment of diet included food frequency questionnaires and the Australian Commonwealth Scientific Investigation and Research Organisation nutritional assessments.

In addition to these general measurements, Raine cohort data include some specialized measurements. For example, fetal biometric data measured by obstetric ultrasound at 18–20 weeks gestation are available on the entire cohort. Serial ultrasound and Doppler assessment of fetal growth and the umbilico-placental circulation at 24, 28, 34 and 38 weeks’ gestation were performed in approximately half of the cohort. 177,178
Figure 4-1 | Broad categories of assessments in the Raine cohort follow-ups
Accurate measures of gestation in the cohort allow separation of the influence of birth weight from gestational age. The Raine cohort has detailed assessment of basal and stress-induced hypothalamic-pituitary-adrenal (HPA) - axis activity as part of ongoing investigations into the role of the HPA-axis in developmental programming. At 17 years of age, awakening (fasting) salivary cortisol level was measured on 3 consecutive days and serum cortisol and adrenocorticotrophic hormone (ACTH) were measured on the third day. Additionally, at 18 years of age, 1137 participants of the Raine study completed the Trier Social Stress Test (TSST).179

The Raine Study has proven to be a valuable scientific resource with some of the major findings to date including: (1) infants who are breastfed longer than 6 months tend to have better mental health at 6 and 8 years of age180; (2) a high quality breakfast that include foods from three different healthy food groups is associated with better mental health in teenagers181; and (3) children whose mothers were stressed or socially disadvantaged during pregnancy generally had a higher risk of developing behavioural and emotional problems.182 Detailed information on the key findings have been outlined by McKnight and colleagues.183

Raine participants underwent a comprehensive ocular examination for the first time at the 20-year follow-up. The major objectives of this eye health study were: (1) to determine the prevalence of ocular conditions such as refractive error, strabismus, amblyopia, pterygium, and keratoconus in young adults; (2) to document the population distribution of disease-related endophenotypes (i.e. central corneal thickness, axial length) in young adults; (3) to determine genetic and environmental factors that influence ocular biometry and predispose to ocular diseases; (4) to establish an ocular baseline data for a population cohort that can be followed through later adult life; (5) to
understand the association between early life factors (including maternal factors, preconceptual/perinatal factors, social, biological, and lifestyle factors) and eye disorders and traits. This paper presents the REHS study methodology including its recruitment process and examination procedures. It also describes the baseline prevalence of ophthalmic disease in a young adult population.

Methods

Ethics Approval
The 20-year review of the Raine Study cohort obtained ethics approval from the Human Research Ethics Committee at the University of Western Australia. The REHS was conducted in accordance with the Declaration of Helsinki and informed consent was obtained from all participants. Previous ethics approvals for other aspects of data collection were completed for each of the earlier examinations.

Recruitment of Participants
More than 3000 pregnant women attending the public antenatal clinic at KEMH or nearby private practices between May 1989 and November 1991 were invited to participate in the ultrasound imaging study. The selection criteria for enrolment were gestational age between 16 and 20 weeks, adequate level of English proficiency to understand the study implications, expected delivery at KEMH, and intention of residence in Western Australia in the future, enabling follow-ups. From the 2900 enrolled pregnancies, 2834 singletons, 64 sets of twins and two sets of triplets were born. Of the 2834 singletons, 1415 were randomized to an intensive ultrasound group and 1419 were randomized to a regular ultrasound group. A total of 2804 mothers remained in the study with 2868 newborns recruited for the cohort follow-up. Of these, 2135 participants were “active” (i.e. previously gave permission to be contacted for
Enrolment for Ophthalmic Examination

All active members of the original birth cohort were invited to attend at 20 years of age. There were no inclusion or exclusion criteria. Booking and assessment procedures are summarized in Figure 4-2. The Raine Study administrative staff contacted participants by phone and invited participants to attend the follow-up assessment. Prior to the appointment, each participant was mailed a detailed information sheet and follow-up questionnaires, which could also be completed with assistance on the appointment day. In addition, directions, a map, contact telephone numbers, information on public transport and parking were provided. All participants received a reminder call 2 days prior to their appointment and a reminder text message on the day. All information on attending participants was verified for accuracy against the restricted access central Raine Study Database by the study coordinators. While eye research was the focus of the 20-year follow-up, data on core longitudinal variables including anthropomorphic measurements, socio-economic status, and demographics were also collected from the cohort. There was ongoing follow-up of nutritional intake, exercise habits, and cardiovascular disease. The cohort also underwent a Dual Energy X-ray Absorptiometry (DEXA) scan and a transient elastography (a liver scan). Heart rate, diastolic and systolic blood pressures were measured. Male participants were also invited to participate in a fertility substudy. The data collected for continuing cohort assessment and substudies will be utilized to investigate potential novel risk factors for abnormal ocular biometric parameters.

The ophthalmic examinations and physical assessments were conducted at the Lions Eye Institute and Sir Charles Gairdner Hospital, Perth, Western Australia. A team consisting of an ophthalmologist, ophthalmology trainees, medical students, orthoptists,
Figure 4-2 | Summary of the booking and examination process.
ophthalmic assistants, and Raine Study research assistants (RAs) performed the eye examinations and physical assessments. A study manual, which included detailed information on the ocular examinations, was provided to all examiners to aid data collection and standardize the examination protocol and recordings. Certified orthoptists conducted the extraocular motility assessment.

Ocular Examination

All the examination equipment was available at the research site. The eye examination protocol (Table 4-1) was arranged into 12 stations. A detailed explanation of each station can be found in the Appendix. The participants rotated through the examination stations accompanied by a research assistant. Participants followed the order from station 1 to 12. If there was a delay at any station, the participant completed a different station with consideration of the requirement for cycloplegia.

Definitions Utilized in the Prevalence of Ophthalmic Diseases

Colour photographs of the nasal and temporal conjunctiva in both eyes were assessed for the presence or absence of pterygium. Pterygium was defined as a fibrovascular conjunctival lesion with characteristic appearance extending to or across the limbus. Participants having a mean spherical equivalent (sum of spherical error and half of cylindrical error) of both eyes less than or equal to -3 diopters were considered to have myopia. This cut-off was used instead of conventional -0.5D as it better defines myopia as a disease. Keratoconus was defined by the presence of unilateral high irregular corneal astigmatism, vogt’s striae, Fleischer ring, or retinoscopic scissoring on slit-lamp biomicroscopy. The data on other ophthalmic diseases were obtained from participants’ previous medical questionnaires. Prevalence rate is based on the findings per person and not per eye.
| Station 1 | Pre-cycloplegia Autorefraction (Nidek ARK-510A, NIDEK Co.Ltd, Japan)  
| Station 2 | Best corrected vision tests  
| Station 3 | Orthoptic Binocular vision function tests  
| Station 4 | Eye Photography  
| Station 5 | IOP Measurement (Icare TAO1i Tonometer, Icare Finland Oy, Helsinki, Finland)  
| Station 6 | Measurement of Corneal Higher Order Aberrations (Zywave II Wavefront Aberrometer, Bausch&Lomb, Inc., Rochester, NY)  
| Station 7 | Ocular Biometry (IOLMAster (V.5), Carl Zeiss Meditec AG, Jena, Germany)  
| Station 8 | Anterior Segment Tomography (Oculus Pentacam, Oculus Optikgerate GmbH, Wetzlar, Germany)  
| Station 9 | 60° Optic Disc centred colour/ Fovea centred colour/ Red Free Disc centred monochrome (Canon CF-60DSi and CF-60UVi Digital Fundus Camera, USA)  
| Station 10 | Optical Coherence Tomography (Spectralis HRA+OCT, Heidelberg Engineering, Heidelberg, Germany)  
| Station 11 | Post-cycloplegia Autorefraction (Nidek ARK-510A, NIDEK Co.Ltd, Japan)  
| Station 12 | Retinal Tomography (Heidelberg Retina Tomograph 3, Heidelberg Engineering, Heidelberg, Germany) |
Questionnaires

Each participant completed a follow-up and a medical questionnaire, which included detailed information on sociodemographic data, ocular history, family history of ocular disease, general medical history, and environmental risk factors with a focus on UV exposure. In addition, participants completed an extensive food frequency questionnaire that had been validated previously.184

DNA Sampling

DNA samples from previous assessments and consents for GWAS studies were available for most participants. If a DNA sample was not previously available, participant consent was sought to obtain DNA from a blood or saliva sample. As part of the 20-year cohort review, the participants were asked to provide a fasting blood sample, which was collected on a separate day by the Raine Study phlebotomist at the participants’ house. Appointments for blood sampling were made at the eye examination appointment to introduce participants to the phlebotomist and reduce the psychological stress of blood sampling. A 43 ml sample was drawn from the cubital fossa vein and delivered to the Royal Perth Hospital (RPH), Perth where the blood analyses were performed. If required, DNA was extracted and sent for storage at KEMH. Surplus blood was stored at -80°C in freezers located at the RPH. Blood test results were sent to participants, with advice to consult their general practitioner regarding any results outside the normal range. Genome-wide genotyping was performed using 250 nanograms of DNA on Illumina 660 Quad Arrays. Genotype quality control and calling were undertaken on the Illumina Bead Array Reader at the Centre for Applied Genomics (Toronto, Ontario, Canada).
Data Entry and Statistical Analysis

The eye examination and physical assessment data were entered into a password-protected database created in Microsoft File Maker Pro (Version 7). All data entry was checked by a second person and validation checks performed by the Raine Study Data Manager. Raine Study research assistants coded, scanned, and verified questionnaires. Where available, crude eye examination data from the instruments were exported into the database.

All phenotype data are stored on secure servers at the Telethon Kids Institute (TKI). All genotype data were stored on the Pawsey Supercomputing Centre supported by the Western Australian Government and the Australian Federal Government. We conducted univariate and multivariate genetic analysis utilizing the PLINK software package.\textsuperscript{192}

National and international collaborations are encouraged in the Raine Study. The Raine Study Executive Committee members are the custodians of the Raine Study data and biological samples. Interested investigators are required to seek approval from the Executive Committee for proposed projects. Access policies, data dictionaries, copies of questionnaires and assessment protocols for each follow-up are available to registered researchers on the Raine Study website (www.rainestudy.org.au).

Study Power for Genetic Discovery

Power calculations for genetic analysis were performed using QUANTO.\textsuperscript{193} For all calculations an additive model with 2 degrees of freedom was used and the marker allele was assumed to be in high linkage disequilibrium with the cause variant ($r^2=1$). Given our population-based design, we had greater power to detect quantitative trait loci of modest effect size (Figure 4-3).
Figure 4-3 | The power of this study to detect a quantitative trait locus conferring 1.5–3.5% of the continuous trait’s variance at the genome-wide threshold ($a = 5 \times 10^{-8}$).
Results

From the original cohort of 2868 live births, 37 (1.3%) participants were deceased, 182 (6.3%) were lost to follow-up and 514 (17.9%) participants had withdrawn from the study by the time of the 20-year follow-up. Of the remaining 2135 active Raine Study members, 1743 individuals (81.6%) verbally agreed to participate in the 20-year follow-up. Of these 1743 participants, 1344 (77.1%) were examined in the 24-month period from March 2010 to February 2012. Participation in the 20-year follow-up is shown in Figure 4-4. Among the 1344 examined participants, 119 attended only one of the last three follow-ups and 32 of participants had attended none (Figure 4-5).

Of the 1344 participants, 690 (51.3%) were males and 654 (48.7%) were females. The ethnic background of participants is shown in Table 4-2. A total of 1214 (90.3%) had a Caucasian mother and 1210 (90.0%) a Caucasian father. For 1149 (85.5%) participants, both parents were Caucasian.

In terms of occupation 37.1% of participants were full-time students and 5.7% were part-time students, 21.9% were in full-time employment and 26.3% were working part-time, whilst 21.9% studied and worked at the same time. A total of 63.5% of participants had completed the final year of high school or its equivalent.

Baseline Prevalence of Ophthalmic Disease

Prevalence of ophthalmic disease in the cohort is shown in Table 4-3. The most common ophthalmic condition was myopia. A total of 5.5% participants had a spherical equivalent of less than or equal to -3 diopters. No participant had been diagnosed with retinopathy of prematurity, retinal dystrophy or glaucoma.
Figure 4-4 | Participation of active cohort members in the Raine Eye Health Study assessment.
Figure 4-5 | The number of Raine Study participants who attended physical examination in the last four follow-ups (year 10, 14, 17 and 20). A total of 1344 individuals attended eye examinations at the 20-year follow-up. Within this group, some individuals who had participated in none or only some of the previous Raine follow-ups were reengaged.
<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mother’s ethnicity (%)</th>
<th>Father’s ethnicity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>1214 (90.3)</td>
<td>1210 (90.0)</td>
</tr>
<tr>
<td>Chinese</td>
<td>54 (4.0)</td>
<td>36 (2.7)</td>
</tr>
<tr>
<td>Indian</td>
<td>36 (2.7)</td>
<td>39 (2.9)</td>
</tr>
<tr>
<td>Indigenous Australians and Torres Islanders</td>
<td>9 (0.7)</td>
<td>12 (0.9)</td>
</tr>
<tr>
<td>Polynesian</td>
<td>10 (0.7)</td>
<td>11 (0.8)</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>5 (0.4)</td>
<td>6 (0.5)</td>
</tr>
<tr>
<td>Not stated/ Unknown</td>
<td>16 (1.2)</td>
<td>30 (2.2)</td>
</tr>
</tbody>
</table>
Table 4-3 | Baseline prevalence of ophthalmic disease in REHS

<table>
<thead>
<tr>
<th>Ophthalmic disease</th>
<th>Number of affected participants (n=1344)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myopia (≤-3D)*</td>
<td>74</td>
<td>5.5</td>
</tr>
<tr>
<td>Pterygium*</td>
<td>15</td>
<td>1.22</td>
</tr>
<tr>
<td>Cataract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congenital cataract</td>
<td>3</td>
<td>0.22</td>
</tr>
<tr>
<td>Traumatic cataract</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>Keratoconus*</td>
<td>2</td>
<td>0.15</td>
</tr>
<tr>
<td>Uveitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idiopathic acute anterior uveitis</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>Juvenile idiopathic arthritis-related anterior uveitis</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>Herpes simplex keratitis</td>
<td>1</td>
<td>0.07</td>
</tr>
</tbody>
</table>

* See text for clinical definitions.
Discussion

The Raine Study cohort is unique in allowing prospectively collected antenatal, childhood, and adolescent data to be correlated with outcomes of an ophthalmic examination. Detailed phenotypic data have been collected prospectively over the last 20 years, resulting in a database of over 16,000 variables on each member of the cohort. This provides the possibility of studying multiple exposures and multiple outcomes at many life stages, allowing identification of potential associations and risk factors in ocular disease processes. Identification of these associations and risk factors may help to improve clinical practice and, possibly, prevent ocular disease.

We acknowledge that the limits of conventional epidemiology include being unable to confirm or refute causality, reverse causality, residual confounding, and ensuring appropriateness of adjustment for confounders/mediators. However, it is possible to construct a “causal” model that incorporates the interaction of many different risk factors at multiple life stages. Moreover, we will investigate genetic risk factors, which may suggest a causal association (e.g. Mendelian Randomization). However, over its 25 years of existence, the Raine Study has maintained high participation rates with minimal attrition, thus giving accurate prevalence rates. Further possible disadvantages of this prospective cohort study include relatively low numbers of incident cases of some endpoints, and the problems of multiple significance tests conducted at different time points. We also acknowledge that there is a possible selection bias relating to the initial recruitment (i.e. mothers giving birth in a public hospital might be of lower social class or different socio-demographic characteristics than those in a private hospital, or exclusion of non-English speaking mothers).
The REHS provides extensive data on ocular biometry and disease from a young adult, predominantly Caucasian population. Strengths of this follow-up include its large sample size, prospective information on neonatal and early life exposure and a standardized clinical assessment.

Characteristics of the original cohort published earlier\textsuperscript{194} were maintained in the REHS. The REHS had a slightly greater male than female participation rate. This is reflective of state population characteristics, where according to 2010 population estimates, 50.7\% of the Western Australian residents were males.\textsuperscript{195}

Population stratification – a systematic difference in allele frequencies between subpopulations in a population due to different ancestry - can cause spurious results in GWAS. Therefore, it is imperative to understand the population structure and apply statistical methods such as genomic control and structured association to discern and correct these differences.\textsuperscript{196}

The REHS sample is primarily Caucasian. This genetic homogeneity will be advantageous in the analysis of quantitative trait loci through genome-wide association studies. Some may argue that finding variants in such a population will have limited application to the ethnically diverse Australian population. While this may be one limitation of the study, it should be remembered that ethnic differences in the prevalence and severity of disease, and the responses to treatment are still part of the unsolved equation of genetic and environmental factors predisposing to ocular disease. Focusing on a particular population will allow us to obtain more specific results, which could be compared with outcomes from other populations.
With the development of new technologies, clinicians are more reliant on highly specialized equipment to make diagnostic decisions. Many clinicians now use these technologies as a part of their routine patient assessment to understand the development and progression of ocular diseases such as age-related macular degeneration and glaucoma. We used equipment such as the Spectralis HRA/OCT and HRT, which to date have not been used in large population studies. The unique set of results obtained from using this technology will construct baseline data for identification and tracking of disease.

**Strategies for High Retention Rates**

Maintenance of sample size in longitudinal prospective studies is critical to maximize statistical power and reduce bias. The Raine Study has sustained high retention rates although the number of active participants has reduced over the years. Participants move away from the metropolitan area, and go interstate or overseas. There is the loss of contact details through change of address, change of name, and loss of contact details of relatives. Some participants withdraw from the study. The Raine Study tries through various methods to establish a feeling of ownership and belonging amongst participants, and reduce cohort attrition. Raine Study staff are highly trained, regular newsletters are sent out, and participants are involved in study planning.

Participation by active members of the cohort (those not lost to follow up, deceased or withdrawn) in the REHS was high (81.6%). At age 20 participants had a variety of work, study and family commitments, and other issues such as examination periods and school holidays had to be accommodated. The cohort assessment was planned at times that were suited to the needs of the cohort, which included weekends, public holidays and weekdays. Childcare was available for participants with infants and young children. A motivating factor for participation could have been the access to a comprehensive
eye-health examination and the immediate discussion of the individual results with an ophthalmologist.

In this chapter, we aimed to describe standardized methodology and practical guidelines to utilize in future eye research. We provided the baseline prevalence of ophthalmic disease in a young adult population. Western Australian data collection ceased in March 2012, although there is still provision for follow-up of participants in other states of Australia. Further analysis of data will define the prevalence of refractive error, strabismus, corneal dystrophy and amblyopia in a young adult population. Investigation of GWAS and early life data could identify genes and environmental factors that influence ocular biometry and predispose to ocular disease. The 22-year cohort follow-up of the Raine Study commenced in March 2012 and completed December 2014 with a focus of sleeping patterns and asthma in young adults. We anticipate repeating the eye follow-up when the cohort reaches 40 years of age when many chronic, age-related, ocular diseases begin to manifest.
Chapter 5
Comparison of Monochromatic Aberrations in Young Adults with Different Visual Acuity and Refractive Errors

Myopia was the most prevalent ophthalmic disease in the REHS. While this was not surprising given the young age of the cohort, it was interesting to see a lower prevalence of myopia compared to rates reported from elsewhere around the world. Myopia is thought to develop due to an imbalance of ocular biometry, mainly an increase in axial length and a reduction in corneal curvature. Other ocular parameters such as optical aberrations may also play a role in myopia development. In this chapter, we explored the effects of monochromatic optical aberrations on vision and refractive error of REHS participants.

Background

The complex process of seeing begins an optical image formed on the retina. The quality of this image varies depending on the unique optical properties of an individual’s eye. Of these properties, monochromatic aberrations are associated with errors that arise from irregular surfaces of the optical media. They are usually described in terms of deformations of the corresponding image wavefronts. Through the development of wavefront-measuring technology, ocular monochromatic aberrations can be readily measured and classified as lower-order aberrations (LOAs) (sphere and cylinder) or higher-order aberrations (HOAs). In general, higher-order aberrations are
less clinically significant than LOAs; nonetheless, when their effects are reduced through adaptive optics techniques, the retinal image quality, and thus the overall visual performance, can significantly improve.197

The analysis of HOAs is clinically important for wavefront-guided excimer laser refractive surgery as well as for custom intraocular lens (IOL) and contact lens designs. Conventional refractive surgery procedures, such as photorefractive keratectomy198,199 and laser in situ keratomileusis,200-203 increase the amount of HOAs and can cause postoperative visual symptoms, including loss of contrast sensitivity or colour perception as well as the perception of glare, halos, or comet tails around lights. With the introduction of wavefront technology, custom corneal ablations can now correct a considerable amount of pre-existing HOAs while minimizing the amount of surgically induced HOAs.199 It has been shown that the use of spherical IOLs in cataract surgery is associated with increased HOAs.198 Many studies204-207 have assessed the effects of materials and designs in pseudophakic eyes to reduce these postoperative aberrations and improve contrast sensitivity. Aspheric IOLs with negative spherical aberrations have been used to compensate for the positive spherical aberration of the cornea.208,209 Although there is considerable variation in the amount of spherical aberrations depending on pupil size and IOL tilt or decentration, overall aspheric IOL designs have been shown to cancel, reduce, or maintain preoperative spherical aberrations.210-214 Furthermore, custom aspheric IOLs based on the corneal wavefront were found to improve image quality compared with generic aspheric IOLs.215 Measurement of HOAs during contact lens wear and a comparison with pre-existing HOAs allow better contact lens fitting.216,217 In addition, custom HOA-reducing soft contact lenses can improve visual acuity in eyes with keratoconus.218,219
Since their initial description, many approaches have been used to quantify ocular aberrations using wavefront technology. Among these representations, Zernike polynomials are now the most widely adopted because they appear to have the most appropriate mathematic properties for the circular pupil. Each Zernike polynomial or term is a trigonometric function of the corresponding wavefront deviation and is calculated from series of positive or negative coefficients.

Monochromatic aberrations vary greatly between people. Age, ethnicity, and refractive error have been proposed to be important determinants of this variation. Population studies have shown that ocular aberrations increase with age. This finding is consistent with reduced contrast sensitivity with increasing age. Subtle changes in the crystalline lens could also account for the increased aberrations in older people. While several studies report the variation in LOAs (spherocylinder errors) among ethnic groups, remarkably few have evaluated the variation in HOAs based on ethnicity. Carkeet et al. found slightly lower aberrations in Malay children than in their Chinese counterparts. Similarly, Lim et al. reported greater ocular aberrations in South East Asia Chinese than in Caucasian and other Asian populations. Interestingly, Cerviño et al. found no significant difference in total HOAs between Asian people and Caucasian people living in Great Britain.

An increase in total monochromatic aberrations can cause a decrease in visual acuity. Although it was suggested that combined Zernike coefficients across each mode could result in better visual performance despite an increase in the total aberration, we are not aware of any study that directly compared monochromatic HOAs between people with different levels of normal visual acuity. Several studies have assessed the relationship between monochromatic aberrations and refractive errors; however, in
general, results have been inconsistent. As such, the clinical importance of monochromatic aberrations in relation to visual function remains somewhat ambiguous.

In this study, we explored the relationship of monochromatic aberrations with different levels of visual acuity and refractive error. Our study cohort included the REHS healthy young adults who were predominantly Caucasian. Therefore we minimized the important potential confounders, such as age and ethnicity, while evaluating the significance of monochromatic aberrations in visual function.

**Methods**

Study design and participant recruitment were described in Chapter 4.

**Participant Examinations**

Aberrations in both eyes of 1320 participants of the REHS were measured with a Zywave II wavefront aberrometer (Bausch & Lomb) before cycloplegia in natural pupils in a dark room. The participant positioned his or her head in the chinrest and was instructed to look at the centre of a fixation target. The operator centred the pupil at the intersection of the crosshairs on the screen and adjusted the focus until the edge between the pupil and iris came sharply into focus. At this point, the subject was asked to blink once and then keep both eyes wide open. If 3 of the 5 consecutive measurements with lowest repeatability criteria were not successful, the measurement was repeated. Three hundred and three participants (23%) had a pupil smaller than 5.0 mm pupil, which resulted in erroneous measurement. For these participants, repeat measurements were performed after cycloplegia was achieved.
The corrected distance visual acuity (CDVA) at 6 m was measured using a logMAR chart (Test Chart 2000 Xpert, Thomson Software Solutions), which was run on the Windows XP operating system (Dell Wyse P20, Wyse Technologies). Participants were actively encouraged to read down the chart until no letters in a line were identified correctly.

After completion of precycloplegic tests, 1 drop of tropicamide 1.0% and 1 drop of phenylephrine 10.0% were administered to achieve mydriasis. Cycloplegic autorefraction was performed with an ARK-510A device (Nidek Co. Ltd.) 20 minutes after instillation. The spherical equivalent (SE) was calculated by summatting the spherical error and half the cylindrical error.

**Statistical Analysis**

Statistical analyses were performed using SPSS software (version 20.0.0.1, International Business Machine Corp.). The raw data were extracted from the aberrometer in .ATE file format and converted into .CSV file format for the statistical package. The participants’ pupil size ranged from 4.6 to 9.1 mm. The majority of the Zernike coefficients were measured with a 6.0 mm pupil diameter. Therefore, data corresponding to 5.0 mm pupil size were rescaled to the same common 6.0 mm pupil size in Matlab software version 7.11 (Mathworks, Inc., Natick, Massachusetts) using the Schwiegerling method as described by Campbell et al. The root-mean-square wavefront errors were computed for coma-like aberrations, 3rd to 5th Zernike orders, and total HOAs (3rd to 5th order combined). Contact lens wearers were instructed to remove their contact lenses three hours before their appointment, allowing them to be included in this dataset. One of two participants diagnosed with keratoconus was excluded from the analysis. The other participant with keratoconus had an incomplete assessment and was also excluded. The Kolmogorov-Smirnov test was used to assess
normality across each coefficient. Data are presented as the median and interquartile range unless otherwise stated.

Results

The HOA data were available from 2039 eyes; at least 1 eye was measured for each of the 1040 participants. For this study, complete right-eye data with corresponding refractive and visual acuity measurements from 1007 participants were analysed. Table 5-1 displays the participants’ demographics. Pupil sizes were normally distributed, with a mean value of 6.8 ± 0.8 mm (Figure 5-1). Figure 5-2 shows the distribution of individual Zernike coefficients from the 3rd to 5th order. These were not normally distributed (p < .001, Kolmogorov-Smirnov test). The median total HOAs was 0.58 µm (range 0.44 to 0.79 µm). There was no significant association between total HOAs (order 3rd, 4th, and 5th combined) and sex (p = 0.254). Also, there was no significant difference in the median total HOAs of Caucasian participants and non-Caucasian participants (p = 0.093).

Men had better CDVA than women (p > 0.001) (Table 5-2). Table 5-3 shows data separated by eyes with normal vision (CDVA range −0.1 to 0.1 logMAR) and eyes with supernormal vision (CDVA better than −0.1 logMAR). Defocus Z(2,0) was the only individual Zernike term significantly different between these groups (p < 0.001). Coma-like aberrations combined, trefoil-like aberrations combined, and 3rd-, 4th-, and 5th-order aberrations were slightly higher in the group with normal vision than in those with supernormal vision (all p < 0.02).

Overall, the participants were slightly hyperopic (Table 5-2). Table 5-4 shows the right-eye data divided into 3 refractive SE categories: myopic eyes (SE ≤ −0.50 D),
emmetropic eyes (SE range −0.50 D to 0.50 D), and hyperopic eyes (SE ≥ +0.50 D).

When the median value of each individual Zernike term was compared, the tetrafoil Z(4,−4), astigmatism Z(4,−2), and defocus Z(2,0) terms varied significantly across these categories (p < 0.001). Similarly, the median values of combined modes of coma-like aberrations, trefoil-like aberrations, 3rd- to 5th-orders, and total HOAs were also significantly different between the categories. Figure 5-3 shows the correlation between CDVA and refractive error with total HOAs, coma-like modes, and spherical aberration. Although there was a positive correlation between total HOAs and coma-like aberrations with visual acuity, there was a negative correlation with refractive error.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (n)</td>
<td>1007</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>20.05 ± 0.45</td>
</tr>
<tr>
<td>Range</td>
<td>18.3, 22.1</td>
</tr>
<tr>
<td>White ethnicity, n (%)</td>
<td>858 (85.2)</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>530 (52.6)</td>
</tr>
</tbody>
</table>
Figure 5-1 | Distribution of post-cycloplegic pupil sizes (N = 1007).
Figure 5-2 | Frequency of individual Zernike terms. No individual Zernike term was normally distributed because they were significantly kurtotic (P > 0.001).
Table 5-2 | Median visual acuity, refractive error, and total HOAs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median (Interquartile Range)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (N = 1007)</td>
<td>Females (n = 477)</td>
<td>Males (n = 530)</td>
<td>P Value</td>
</tr>
<tr>
<td>CDVA (logMAR)</td>
<td>-0.06 (-0.10, 0.00)</td>
<td>-0.06 (-0.08, -0.02)</td>
<td>-0.08 (-0.12, 0.00)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Spherical error (D)</td>
<td>+0.50 (-0.25, 0.75)</td>
<td>+0.50 (-0.25, +0.75)</td>
<td>+0.50 (0.00, +0.75)</td>
<td>0.043</td>
</tr>
<tr>
<td>Cylindrical error (D)</td>
<td>-0.25 (-0.50, -0.25)</td>
<td>-0.25 (-1.00, -0.25)</td>
<td>-0.50 (-1.00, -0.25)</td>
<td>0.144</td>
</tr>
<tr>
<td>Spherical equivalent (D)</td>
<td>+0.25 (-0.38, 0.63)</td>
<td>+0.13 (-0.44, 0.63)</td>
<td>+0.38 (-0.38, 0.63)</td>
<td>0.052</td>
</tr>
<tr>
<td>Total HOAs (µm)</td>
<td>0.58 (0.44, 0.79)</td>
<td>0.58 (0.45, 0.82)</td>
<td>0.57 (0.43, 0.77)</td>
<td>0.254</td>
</tr>
</tbody>
</table>

CDVA = corrected distance visual acuity; HOAs = higher-order aberrations
Table 5-3 | Comparison of Zernike coefficients between eyes with differing levels of visual acuity.

<table>
<thead>
<tr>
<th>Zernike Term</th>
<th>Eyes with Normal Vision (n = 767)</th>
<th>Eyes with Supernormal Vision (n = 203)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z (2,−2) astigmatism</td>
<td>−0.131 (−0.393, 0.129)</td>
<td>−0.107 (−0.269, 0.103)</td>
<td>0.348</td>
</tr>
<tr>
<td>Z (2,0) defocus</td>
<td>−0.695 (−2.011, 0.004)</td>
<td>−0.173 (−0.812, 0.225)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Z (2,2) astigmatism</td>
<td>0.141 (−0.281, 0.569)</td>
<td>0.166 (−0.164, 0.367)</td>
<td>0.607</td>
</tr>
<tr>
<td>Z (3,−3) trefoil</td>
<td>0.074 (−0.077, 0.248)</td>
<td>0.083 (−0.023, 0.220)</td>
<td>0.228</td>
</tr>
<tr>
<td>Z (3,−1) coma</td>
<td>−0.060 (−0.281, 0.152)</td>
<td>−0.045 (−0.194, 0.114)</td>
<td>0.406</td>
</tr>
<tr>
<td>Z (3,1) coma</td>
<td>−0.131 (−0.276, 0.001)</td>
<td>−0.127 (−0.246, −0.012)</td>
<td>0.861</td>
</tr>
<tr>
<td>Z (3,3) trefoil</td>
<td>−0.080 (−0.212, 0.037)</td>
<td>−0.084 (−0.193, 0.035)</td>
<td>0.847</td>
</tr>
<tr>
<td>Z (4,−4) tetrafoil</td>
<td>−0.051 (−0.110, −0.001)</td>
<td>−0.048 (−0.102, −0.007)</td>
<td>0.677</td>
</tr>
<tr>
<td>Z (4,−2) astigmatism</td>
<td>0.026 (−0.010, 0.064)</td>
<td>0.023 (−0.009, 0.055)</td>
<td>0.394</td>
</tr>
<tr>
<td>Z (4,0) spherical</td>
<td>−0.231 (−0.394, −0.091)</td>
<td>−0.189 (−0.325, −0.072)</td>
<td>0.060</td>
</tr>
<tr>
<td>Z (4,2) astigmatism</td>
<td>0.009 (−0.066, 0.098)</td>
<td>−0.003 (−0.065, 0.065)</td>
<td>0.090</td>
</tr>
<tr>
<td>Z (4,4) tetrafoil</td>
<td>−0.046 (−0.120, 0.014)</td>
<td>−0.033 (−0.111, 0.016)</td>
<td>0.241</td>
</tr>
<tr>
<td>Z (5,−5) pentafoil</td>
<td>0.019 (−0.019, 0.060)</td>
<td>0.025 (−0.011, 0.049)</td>
<td>0.502</td>
</tr>
<tr>
<td>Z (5,−3) trefoil</td>
<td>−0.006 (−0.043, 0.027)</td>
<td>−0.010 (−0.033, 0.027)</td>
<td>0.671</td>
</tr>
<tr>
<td>Z (5,−1) coma</td>
<td>0.010 (−0.029, 0.055)</td>
<td>0.002 (−0.027, 0.047)</td>
<td>0.551</td>
</tr>
<tr>
<td>Z (5,1) coma</td>
<td>−0.019 (−0.048, 0.006)</td>
<td>−0.017 (−0.038, 0.004)</td>
<td>0.261</td>
</tr>
<tr>
<td>Z (5,3) trefoil</td>
<td>−0.007 (−0.033, 0.017)</td>
<td>−0.010 (−0.032, 0.012)</td>
<td>0.271</td>
</tr>
<tr>
<td>Z (5,5) pentafoil</td>
<td>0.007 (−0.026, 0.042)</td>
<td>0.004 (−0.019, 0.032)</td>
<td>0.690</td>
</tr>
<tr>
<td>Coma-like modes combined</td>
<td>0.346 (0.222, 0.515)</td>
<td>0.288 (0.177, 0.419)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Trefoil-like modes combined</td>
<td>0.266 (0.165, 0.415)</td>
<td>0.235 (0.143, 0.366)</td>
<td>0.018</td>
</tr>
<tr>
<td>Order 3</td>
<td>0.456 (0.310, 0.652)</td>
<td>0.384 (0.256, 0.562)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Order 4</td>
<td>0.331 (0.216, 0.478)</td>
<td>0.276 (0.196, 0.404)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Order 5</td>
<td>0.118 (0.078, 0.180)</td>
<td>0.098 (0.070, 0.130)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>All higher orders (3, 4, 5)</td>
<td>0.599 (0.452, 0.828)</td>
<td>0.531 (0.391, 0.712)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Wilcoxon rank-sum test
Table 5-4 | Distribution of Zernike coefficients across 3 refractive groups.

<table>
<thead>
<tr>
<th>Zernike Term</th>
<th>Myopic Eyes (n = 217)</th>
<th>Emmetropic Eyes (n = 476)</th>
<th>Hyperopic Eyes (n = 314)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z (2,−2) astigmatism</td>
<td>−0.064 (−0.335, 0.247)</td>
<td>−0.125 (−0.652, 0.113)</td>
<td>−0.145 (−0.353, 0.090)</td>
<td>0.101</td>
</tr>
<tr>
<td>Z (2,0) defocus</td>
<td>−3.794 (−6.948, −2.259)</td>
<td>−0.494 (−1.223, −0.052)</td>
<td>0.095 (−0.444, 0.655)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Z (2,2) astigmatism</td>
<td>0.123 (−0.498, 0.693)</td>
<td>0.159 (−0.264, 0.505)</td>
<td>0.168 (−0.165, 0.462)</td>
<td>0.646</td>
</tr>
<tr>
<td>Z (3,−3) trefoil</td>
<td>0.068 (−0.109, 0.259)</td>
<td>0.078 (−0.057, 0.228)</td>
<td>0.093 (−0.031, 0.277)</td>
<td>0.339</td>
</tr>
<tr>
<td>Z (3,−1) coma</td>
<td>−0.087 (−0.324, 0.152)</td>
<td>−0.072 (−0.269, 0.143)</td>
<td>−0.023 (−0.197, 0.116)</td>
<td>0.124</td>
</tr>
<tr>
<td>Z (3,1) coma</td>
<td>−0.143 (−0.319, −0.005)</td>
<td>−0.135 (−0.281, −0.003)</td>
<td>−0.111 (−0.237, 0.001)</td>
<td>0.998</td>
</tr>
<tr>
<td>Z (3,3) trefoil</td>
<td>−0.092 (−0.262, 0.033)</td>
<td>−0.078 (−0.188, 0.039)</td>
<td>−0.082 (−0.324, 0.152)</td>
<td>0.292</td>
</tr>
<tr>
<td>Z (4,−4) tetrafoil</td>
<td>−0.071 (−0.138, −0.005)</td>
<td>−0.047 (−0.103, −0.004)</td>
<td>−0.048 (−0.097, −0.002)</td>
<td>0.024</td>
</tr>
<tr>
<td>Z (4,−2) astigmatism</td>
<td>0.035 (−0.003, 0.081)</td>
<td>0.019 (−0.013, 0.059)</td>
<td>0.028 (−0.008, 0.063)</td>
<td>0.029</td>
</tr>
<tr>
<td>Z (4,0) spherical</td>
<td>−0.242 (−0.476, −0.079)</td>
<td>−0.206 (−0.361, −0.073)</td>
<td>−0.239 (−0.381, −0.097)</td>
<td>0.076</td>
</tr>
<tr>
<td>Z (4,2) astigmatism</td>
<td>0.005 (−0.090, 0.104)</td>
<td>0.012 (−0.064, 0.092)</td>
<td>−0.001 (−0.066, 0.079)</td>
<td>0.873</td>
</tr>
<tr>
<td>Z (4,4) tetrafoil</td>
<td>−0.035 (−0.129, 0.033)</td>
<td>−0.043 (−0.120, 0.014)</td>
<td>−0.053 (−0.113, 0.008)</td>
<td>0.649</td>
</tr>
<tr>
<td>Z (5,−5) pentafoil</td>
<td>0.014 (−0.021, 0.067)</td>
<td>0.022 (−0.017, 0.054)</td>
<td>0.022 (−0.016, 0.055)</td>
<td>0.974</td>
</tr>
<tr>
<td>Z (5,−3) trefoil</td>
<td>−0.007 (−0.051, 0.029)</td>
<td>−0.005 (−0.037, 0.027)</td>
<td>−0.007 (−0.035, 0.027)</td>
<td>0.387</td>
</tr>
<tr>
<td>Z (5,−1) coma</td>
<td>0.017 (−0.022, 0.060)</td>
<td>0.006 (−0.032, 0.052)</td>
<td>0.009 (−0.026, 0.051)</td>
<td>0.218</td>
</tr>
<tr>
<td>Z (5,1) coma</td>
<td>−0.013 (−0.048, 0.012)</td>
<td>−0.018 (−0.045, 0.007)</td>
<td>−0.022 (−0.043, −0.003)</td>
<td>0.218</td>
</tr>
<tr>
<td>Z (5,3) trefoil</td>
<td>−0.004 (−0.034, 0.013)</td>
<td>−0.009 (−0.033, 0.017)</td>
<td>−0.007 (−0.032, 0.013)</td>
<td>0.694</td>
</tr>
<tr>
<td>Z (5,5) pentafoil</td>
<td>−0.001 (−0.038, 0.047)</td>
<td>0.005 (−0.025, 0.035)</td>
<td>0.009 (−0.021, 0.037)</td>
<td>0.533</td>
</tr>
<tr>
<td>Coma-like modes combined</td>
<td>0.417 (0.247, 0.595)</td>
<td>0.326 (0.211, 0.462)</td>
<td>0.297 (0.191, 0.438)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Trefoil-like modes combined</td>
<td>0.313 (0.187, 0.478)</td>
<td>0.238 (0.160, 0.388)</td>
<td>0.257 (0.159, 0.387)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Order 3</td>
<td>0.524 (0.360, 0.764)</td>
<td>0.423 (0.290, 0.628)</td>
<td>0.404 (0.294, 0.560)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Order 4</td>
<td>0.369 (0.247, 0.604)</td>
<td>0.309 (0.203, 0.454)</td>
<td>0.309 (0.217, 0.447)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Order 5</td>
<td>0.123 (0.081, 0.208)</td>
<td>0.111 (0.073, 0.163)</td>
<td>0.106 (0.076, 0.145)</td>
<td>0.003</td>
</tr>
<tr>
<td>All higher orders (3, 4, 5)</td>
<td>0.723 (0.507, 1.020)</td>
<td>0.559 (0.428, 0.773)</td>
<td>0.548 (0.445, 0.725)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Kruskal-Wallis test
Figure 5-3 | Correlation of monochromatic aberrations with CVDA and refractive error for 6.0 mm pupils. A: Visual acuity increased with higher amounts of total HOAs (P < .001) and coma-like aberrations (P = .004) but decreased with increasing amounts of spherical aberrations (P = .002) B: Myopia increased with increasing amounts of total HOAs (P < .001) and coma-like aberrations (P < .001) but decreased with increasing spherical aberrations (P < .001) (CDVA = corrected distance visual acuity; SE = spherical equivalent)
Discussion

Despite extensive research, the clinical significance of monochromatic aberrations and their relationship to visual acuity and refractive error are still not fully understood. There is also a paucity of data on the normal reference ranges for monochromatic aberrations in the general population. In this study, we report the distribution of monochromatic aberrations in 1007 young adult participants. Previously, Lim et al.\textsuperscript{227} reported similar magnitudes of total monochromatic aberrations in a small set (n = 70) of Chinese individuals. In contrast, other studies\textsuperscript{226,234-236} have shown lower magnitudes of total monochromatic aberrations. Our findings also indicate greater variation in individual Zernike terms, which is also in agreement with the work of Lim et al.\textsuperscript{227}

It is theoretically possible to achieve supernormal vision through the elimination of all optical aberrations. In an emmetropic eye, the CDVA was calculated to be $-0.4$ to $-0.3$ logMAR with the perfect photoreceptor diameter and packing.\textsuperscript{237} Applegate et al.\textsuperscript{238} suggest the combined Zernike modes cause better visual performance despite the higher magnitudes of HOAs. In alignment with this hypothesis, some of our participants (1%) achieved visual acuity of better than $-0.2$ logMAR regardless of the presence of HOAs. Amesbury and Shallhorn\textsuperscript{239} measured similar amounts of HOAs in naval aviators with and without supernormal vision. They also found that pilots with excellent vision did not have fewer HOAs than normal subjects. These findings could be explained by the balanced equilibrium between the HOAs and other deteriorated optical elements in the eye; however, this requires further study.

Previously, Applegate et al.\textsuperscript{238} found that a more than 0.05 µm increase in monochromatic aberrations caused a reduction in visual acuity on high-contrast and
low-contrast charts. Our comparison of HOAs between participants with normal vision and those with supernormal vision showed that a total magnitude as small as 0.062 μm was sufficient to lead to a letter or more difference on the logMAR visual chart. No single Zernike term varied significantly between the 2 groups; however, each combined mode did, indicating that a very small magnitude of monochromatic aberration (<0.020 μm) does not affect visual acuity. Furthermore, although we found no difference in total HOAs between men and women, men had slightly better visual acuity. Such distribution difference in visual acuity may be attributed to characteristics of genes in the hemizygous X-chromosome in men.240

Refractive errors, in particular myopia, are multifactorial conditions likely influenced by multiple biometric parameters. Monochromatic aberrations were postulated to contribute to the development of refractive errors by causing retinal image blur.241 Although several studies have examined the influence of monochromatic aberrations on refractive status, there has been no consensus. Some studies228-230,242 support the hypothesis that individuals with more myopic eyes would have higher levels of aberrations. Others222,226,243,244 found no differences in aberration characteristics in myopic eyes, emmetropic eyes, and hyperopic eyes. This variation could be attributed to differences in methodologies, population background, and age. The closest study to ours in terms of these characteristics was that by Cheng et al.243 in which no correlation was identified between refractive error and the level of HOAs. We found higher magnitudes of individual order aberrations and total HOAs in myopic than in emmetropic and hyperopic eyes. Similarly, Paquin et al.230 detected greater amounts of HOAs in myopic eyes in a study using a 5.0 mm diameter pupil with spectacle correction in place. We also identified higher magnitudes of coma-like aberrations in myopic patients, as suggested by Paquin et al. Further studies of the association between monochromatic
aberrations and other ocular parameters would help us understand the accurate effect of monochromatic aberrations in the determination of refractive status.

In conclusion, we report the distribution of monochromatic aberrations in a healthy young adult population with predominantly Caucasian ancestry and identified relatively high magnitude of monochromatic aberrations in this population. We found that supernormal vision is attainable even in the presence of monochromatic aberrations. Finally, we detected a positive correlation between the magnitude of monochromatic aberrations and refractive status.
Chapter 6

Early Anaesthesia Exposure and the Effect on Visual Acuity, Refractive Error and Retinal Nerve Fiber Layer Thickness of Young Adults

Refractive error and myopia occurs as a result of an imbalance in ocular biometry. This imbalance is a consequence of interactions between the genes and the environmental risk factors an individual is exposed to at various life stages including the perinatal period and early childhood. In this chapter, we report the possible effects of on the visual acuity and refractive status of young adults who underwent general anaesthesia early in life.

Background

“Primum non nocere – first do no harm” is the overriding principle in medicine. Yet, the question of whether being anesthetised as a child harms brain development remains largely unanswered. While some studies showed children had an increased risk of being diagnosed with subsequent cognitive impairment following anaesthesia exposure in early childhood\textsuperscript{245-249}, others found no evidence for a neurotoxic effect of a single anaesthesia exposure.\textsuperscript{250-252} However, there is some consistent evidence suggesting that repeated anaesthesia exposure is related to cognitive deficit later in life.\textsuperscript{246,253-255}

In the human brain, age-related synaptogenesis in the visual cortex (V1) begins during gestation and progresses in two stages. The first period, which ends at about postnatal
age 8 months, involves rapid synapse production whilst the second is a longer period of synapse elimination that extends past age 3 years. As the synapses are effective at evoking a response, they grow stronger overtime. A synaptic modification is hypothesised to occur depending on the correlation between the pre- and postsynaptic firing. The experience-dependent cortical plasticity studies allowed us to understand that uncorrected firing of action potentials between the pre-synaptic and post-synaptic neurons in V1 weakens synaptic connections.256 These excitatory synaptic transmissions are mainly mediated by NMDA and 5-methyl-4-isoxazolepropionic acid (AMPA) receptors while inhibitory synaptic transmissions are regulated by GABA\textsubscript{A} receptors.257,258 Recently, Hensch and colleagues demonstrated that increasing inhibition throughout the critical period with GABA\textsubscript{A} receptor agonists could lead to a 30% increase in columnar width of the visual cortex, whereas inverse agonists could produce column shrinkage.259

Due to the aforementioned neuronal activities occurring during visual development in humans and also the evidence for cognitive impairment in individuals exposed to anaesthesia, we hypothesised that exposure to these agents in early childhood could potentially impact normal visual development. Standard synaptogenesis in a child’s brain continues until the end of teen years and the visual system is not completely developed and refractive status is not stabilized until the early adult life. This study was set out to investigate whether being anaesthetised at least once during early life had an impact on three surrogate measures of visual function: visual acuity, refractive error, and thickness of retinal nerve fiber layer (RNFL) in young adulthood.
Methods

The Western Australian Pregnancy Cohort (Raine) Study is an ongoing prospective cohort study of pregnancy, childhood, adolescence, and young adulthood in Perth, Western Australia. At the initiation of the study, 2900 pregnant women at 16-18 weeks’ gestation were recruited from the state’s largest public women’s hospital and surrounding private practices for a randomised clinical trial investigating effects of intensive ultrasound imaging and Doppler flow studies in pregnancy outcomes.

Following this study, 2868 offspring born to 2804 of the recruited women have been evaluated in detail during subsequent childhood, adolescent, and young adult follow-ups. From birth, parents were asked to keep detailed diaries of their child(ren)’s medical history. At the 1-, 2-, and 3-year follow-ups, parents were asked to complete questionnaires describing illnesses and medical problems, which were then coded into the International Classification of Diseases, 9th Revision by the research staff. Any child who had a surgical or diagnostic procedure requiring anaesthesia before the age of 3 years was classified into the “exposed” group. The remaining individuals were included in the “non-exposed” group. Exposure to anaesthesia was confirmed by review of the types of procedures recorded in the questionnaire. Individuals who were found to have diagnostic procedures not requiring anaesthesia were classified into the non-exposed group. No direct access was available to medical records perinatally including surgical and anesthetic records.

At the 20-year review of the cohort, participants underwent a comprehensive ocular examination. This examination included assessments of best-corrected visual acuity using LogMAR chart and contrast sensitivity (CS) using a low-contrast letter chart (Test Chart Xpert, Thomson Software Solutions, Herts, UK), measurement of refractive error
by autorefraction (Nidek 510 ARK, NIDEK Co. Ltd, Japan) after administration of cycloplegic drops and an orthoptic examination by a qualified orthoptist. Refractive error was determined as spherical equivalence (SE) by summating the spherical error and half the cylindrical error. During the examination, ocular history including previous ocular surgery and/or diagnostic procedure was recorded, and binocular vision function was assessed. Those who were reported to have strabismus surgery as a child were excluded from the analysis. RNFL thickness measured both globally and in four quadrants around the optic disc using Spectralis spectral-domain optical coherence tomography (Heidelberg Engineering GmbH, Heidelberg, Germany). Unclear scans or scans with low signal strength were removed from the dataset.

The 20-year follow-up of the Raine Study cohort obtained ethics approval from the Human Research Ethics Committee at the University of Western Australia. The study was conducted in accordance with the Declaration of Helsinki and informed consent was obtained from all participants. Previous ethics approvals were completed for each of the earlier examinations.
**Statistical Analysis**

All variables were assessed for normality and summarized using median (interquartile range [IQR]). Differences between two continuous variables were assessed with the Mann-Whitney U test. Differences between categorical variables were assessed with the chi-squared test. We used mean SE of both eyes of individuals for estimation of prevalence of myopia. Myopia was defined as mean SE < -0.5 diopters (D). Statistical analyses were considered significant at the p < 0.05 level and all P-values were two-tailed. Statistical analyses were performed using the statistical software R version 2.15.1 (R Foundation for Statistical Computing; http://www.r-project.org/).
Results

Ophthalmic and paediatric anaesthesia exposure data were available for 1010 individuals. We excluded 8 individuals who were reported to have strabismus surgery in early life questionnaires and 32 individuals identified with strabismus and other ocular problems at the time of eye examinations. 136 individuals with non-European ancestry had more myopic refractive error compared to their peers with Northern European ancestry (median SE of +0.12 vs +0.31 diopters, p<0.001); thus, they were excluded from the analysis (Figure 6-1). No participants had a history of an ophthalmic disease diagnosis other than strabismus that required surgery in the first 3 years of life.

Of the 834 individuals included in the analysis, 420 (51.3%) were male, and the mean age of the participants at ocular examination was 20.0 ± 0.42 years (range: 18.3 to 22.1). Of the total, 127 (15.2%) participants were exposed to anaesthesia at least once before the age of 3 years. While 24 individuals exposed to anaesthesia twice, only 9 individuals had three or more exposures. No age difference was present between the exposed and non-exposed groups (p=0.99); however, there were more boys in the exposed group compared to non-exposed group (63.0% vs 49.4% males, p=0.005).
2900 pregnant women recruited for the original study.

2868 offspring born to 2804 of the recruited women enrolled into the follow-up studies.

1344 individuals attended eye examinations at the 20-year follow-up.

Complete eye and anaesthesia data were available for 1010 participants.

136 participants without Northern European Ancestry were excluded.

8 participants with previous history of strabismus surgery and 32 individuals identified with strabismus and other ocular problems were excluded.

Remaining 834 participants were included in the analysis.

No exposure to anaesthesia = 707
Exposed to anaesthesia once = 94
Exposed to anaesthesia more than once = 33

Figure 6-1 | Flow chart of study participants.
Twenty-eight participants (3.1%) had a score of less than 80 seconds of arc indicating reduced binocularity. Of those, 24 did not have exposure to anaesthesia in the first 3 years of life. None of participants had amblyopia that is defined as two lines of visual acuity difference. Table 6-1 displays the comparison of exposed versus non-exposed group results for visual acuity and refractive error. Median visual acuity was -0.06 logMAR score in the right eye and -0.08 logMAR score in the left eye of the exposed group. No difference was present in visual acuity of both exposed and non-exposed groups ($p_{\text{right eye}} = 0.625$ and $p_{\text{left eye}} = 0.413$). Similarly, there was no difference in CS of the 2 groups (median CS in both eyes for both groups=1.35 logCS, $p>0.05$). The median SE refractive error of both eyes was +0.44 D (IQR: -0.38, +0.75) in the exposed group and +0.31 D (-0.25, +0.63) in the non-exposed group ($p=0.126$); 18.4% of the exposed group were myopic compared to 19.7% of the non-exposed group ($p = 0.729$).

The mean global RNFL thickness was 101.1 µm ± 13.3 in the exposed group and 100.8 µm ± 10.9 in the non-exposed group ($p=0.830$). As displayed in Table 6-2, there was also no difference in mean RNFL thickness in any of the quadrants between the 2 groups.
Table 6-1 | Comparison of visual acuity and refractive error in participants with and without exposure to anaesthesia as a child

<table>
<thead>
<tr>
<th></th>
<th>Median (Interquartile Range)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposed (n=127)</td>
<td>Non-exposed (n=707)</td>
<td>p-value*</td>
</tr>
<tr>
<td>Visual Acuity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Eye</td>
<td>-0.06 (-0.12, 0)</td>
<td>-0.06 (-0.10, 0)</td>
<td>0.625</td>
</tr>
<tr>
<td>Left Eye</td>
<td>-0.08 (-0.10, 0)</td>
<td>-0.06 (-0.10, 0)</td>
<td>0.413</td>
</tr>
<tr>
<td>Spherical Equivalent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Eye</td>
<td>+0.38 (-0.38, +0.75)</td>
<td>+0.38 (-0.25, +0.63)</td>
<td>0.091</td>
</tr>
<tr>
<td>Left Eye</td>
<td>+0.38 (-0.25, +0.75)</td>
<td>+0.38 (-0.25, +0.63)</td>
<td>0.215</td>
</tr>
</tbody>
</table>

* Wilcoxon signed-rank test.
Table 6-2 | Comparison of retinal nerve fiber layer thickness in eyes of individuals exposed to anaesthesia versus those who were not.

<table>
<thead>
<tr>
<th>Quadrant</th>
<th>Exposed group (mean ± SD)</th>
<th>Non-exposed group (mean ± SD)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferior</td>
<td>128.24 ± 17.78</td>
<td>127.21 ± 17.36</td>
<td>0.556</td>
</tr>
<tr>
<td>Superior</td>
<td>123.28 ± 15.92</td>
<td>124.76 ± 15.3</td>
<td>0.350</td>
</tr>
<tr>
<td>Nasal</td>
<td>81.11 ± 18.65</td>
<td>79.92 ± 15.79</td>
<td>0.515</td>
</tr>
<tr>
<td>Temporal</td>
<td>70.39 ± 9.29</td>
<td>71.02 ± 11.05</td>
<td>0.510</td>
</tr>
<tr>
<td>Global</td>
<td>101.05 ± 13.25</td>
<td>100.78 ± 10.86</td>
<td>0.830</td>
</tr>
</tbody>
</table>

* Two sample t-test.
Discussion

The results of this study suggest that exposure to anaesthesia in early childhood is not associated with reduced visual acuity or increased myopia in young adulthood. Previously, it was shown that the offspring of mothers who were exposed to anaesthesia during pregnancy exhibited longer “looking times” at the visual stimuli and had different visual pattern preferences compared to an unexposed control group.\textsuperscript{260} We compared best-corrected visual acuity of the individuals who were exposed to anaesthesia early in life and found no difference compared to their non-exposed peers. Visual acuity is an outcome from the proper functioning of the cornea, lens, retina, optic nerve (axon and myelination), and higher cortical aspects from parietal/temporal/occipital lobe; however, the term 'vision' has broader applications highly dependent on higher cortical functioning (proprioception, movement, color, face/letter/object recognition) that can be dysfunctional despite a normal visual acuity. For example, a patient with temporal-occipital lobe injuries can have normal visual acuity but “poor vision”. For this reason, although visual acuity cannot be generalized to visual development and vision completely, such is a strong surrogate that measures the development of various structures within the visual pathway. It is, therefore, likely that the evidence of no association between the visual acuity and anaesthesia exposure may also account for the hypothesis that early anaesthesia has no effect on vision.

Emmetropisation of the eye highly correlates with the eye growth and is mediated by a local feedback loop within the sensory retina\textsuperscript{261} without involvement of the central nervous system (CNS). Some evidence suggests that through involvement of various amacrine cells, via the choroid and sclera, the CNS could have at least a modulatory
influence on eye growth and emmetropisation.\textsuperscript{261,262} Moreover disconnecting the eye from the CNS by cutting the optic nerve alters the regulatory process in chick models, suggesting that an intact optic nerve and healthy CNS appear to be essential for emmetropia and normal eye growth.\textsuperscript{263} Deprivation of vision due uncorrected myopia, cataracts, and corneal dystrophy can be associated with poor emmetropisation in children. Kugelberg et al.\textsuperscript{264} found children with unilateral congenital cataract have shorter axial length in the affected eye compared to unaffected eye. Despite this evidence, similar to visual acuity, no difference was present in refractive error of the exposed and non-exposed groups.

During normal synaptogenesis, a small percentage of neurons (<1\%) that are redundant are eliminated through apoptosis.\textsuperscript{265} Yet it was surprising when Ikonomidou and colleagues\textsuperscript{266} first reported that blockade of glutamate NMDA receptors, or the excessive activation of \(\gamma\)-aminobutyric acid A receptors could result in apoptotic neuronal degeneration in developing rat brains particularly during synaptogenesis. Retinal ganglion cells (RGCs) are the last chain of the neurons that link the eye to the brain. Given that RNFL axons originates from the RGC bodies, we postulated that a possible reduction in RNFL thickness may indicate RGC loss in young adults who were exposed to anaesthesia early in life. However, our findings did not support this hypothesis.

The generalizability of these results is subject to certain limitations. The outcomes we studied were secondary measures for the cohort study, and that the sample size might be lacking. Moreover, our data were underpowered to investigate the effect of repeated measures on visual acuity, refractive error and RNFL thickness. Because of lack of access to medical records, we were unable to identify the anaesthetic agents used and
the dosage and duration of the anaesthesia. During the period when the Raine Study participants were exposed to anaesthesia (1990-1994), there was likely initially a predominance of halothane and enflurane followed by increasing numbers of patients receiving isoflurane or sevoflurane. Halothane depresses the CNS by blocking the effects of the excitatory neurotransmitter, glutamic acid, at the NMDA receptors. Similar to the neurotoxic effects caused by other inhalation anaesthetic agents, significant reduction in synaptic density was observed in rat models after chronic exposure to halothane. It must be noted that some of the anaesthetic agents such as halothane and enflurane are no longer widely used and may have greater neurotoxic effects than the agents used currently. Also, we were unable to eliminate all other illness and surgeries that may have affected visual acuity and other measures from early childhood to young adulthood. Finally, results of this study are limited to individuals with Northern European ancestry.

In conclusion, despite concerns raised by the literature regarding the need for interplay between the central nervous system and the eye in modulating eye growth, our findings suggest that exposure to anaesthesia at least once in early childhood likely has no impact on three important proxies of visual function: visual acuity, refractive error and RNFL thinning. To our knowledge, this is the first clinical study reporting evidence from a longitudinal study of exposed infants. Therefore, replication cohorts are necessary to validate this epidemiological finding.
Chapter 7

Genetic and Environmental Factors in Conjunctival UV Autofluorescence

Time spent outdoors is proposed to be protective against myopia development in multiple epidemiological studies. Yet there is no accepted objective method to measure sun exposure. Conjunctival ultraviolet autofluorescence (CUVAF) photography was developed to detect and characterise pre-clinical sunlight-induced ocular damage. It has excellent potential as an objective biomarker of sun exposure. In this study, we investigated sources of variation in CUVAF in relation to its potential clinical relevance.

Background

Excessive sun exposure particularly ultraviolet-light (UV) increases the risk of many ocular diseases including pterygium,\textsuperscript{11} cortical cataract,\textsuperscript{12} ocular surface squamous neoplasia,\textsuperscript{269} climatic droplet keratopathy\textsuperscript{270} and eyelid malignancy\textsuperscript{271}. Despite early work suggesting sun exposure has a role in the pathogenesis of age-related macular degeneration\textsuperscript{13} and ocular melanoma,\textsuperscript{272} these associations remain inconclusive. In recent years, a considerable number of epidemiological studies have reported that increased time spent outdoors is associated with lower rates of myopia in children, suggesting that sunlight brightness or UV-light may have a beneficial effect.\textsuperscript{143} These conflicting reports on effects of sun exposure require a better understanding of mechanisms underlying ocular sun damage and related eye diseases.
A challenge of studying ophthalmohelioses (sun-related ocular diseases) is the difficulty of assessing sun exposure. The usual method of determining an individual’s sun exposure is by self-reported questionnaire, which is subject to recall errors. Often questions are designed to assess whole-body sun exposure rather than ocular sun exposure, thus accuracy of these measures in ocular diseases is arbitrary. CUVAF photography was developed to detect precursors of ocular sun damage using a technique similar to UV fluorescence in the detection of UV exposure-related dermatologic diseases. Previous studies have reported an association of CUVAF with the presence of pterygia and shown increasing total area of CUVAF is associated with increasing prevalence of pterygium. Time spent outdoors correlates highly with the area of CUVAF, which suggests that CUVAF could be regarded as an objective measure of sun damage corresponding to amount of time spent outdoors and could help characterize local sun exposure.

Multiple biological mechanisms have been proposed to explain the cause of detected CUVAF in other tissues. These mechanisms include alterations of collagen cross-linking or changes in cell metabolites, such as reduced nicotinamide adenine dinucleotide, or derivatives of amino acids, such as tryptophan.

CUVAF can be an ideal biomarker of ophthalmohelioses once its characteristics are defined better. In the present study, we sought to determine whether there is a genetic predisposition to variation in CUVAF area exists in three Australian cohorts; Twin Eye Study in Tasmania (TEST), Brisbane Adolescent Twin Study (BATS), and Raine Study. However, given that sun exposure is highly dependent on geographical location, we investigate the effect of latitudinal differences on CUVAF distribution. After this analysis, we explored the contribution of genes to CUVAF variation through a classical
twin study and a GWAS.

Methods

Participants

This study included two twin and one singleton cohort, all of Northern European ancestry and all from Australia. Twin pairs were identified from two existing cohorts: TEST and BATS. Methods of these studies were described in detail previously. In brief, a total of 487 twin pairs (200 monozygotic [MZ], 287 dizygotic [DZ] twin pairs) were recruited in the TEST through several overlapping methods, including use of national twin registry and existing state-wide studies. A total of 2443 individuals who were enrolled into BATS were invited to participate in the twin eye study. Among the 1199 individuals agreed to participate, there were 185 MZ and 278 DZ twin pairs. Approval for twin studies was obtained from the Human Research Ethics Committees of the University of Tasmania, Royal Victorian Eye and Ear Hospital, and QIMR Berghofer Medical Research Institute. The singleton cohort consisted of the REHS participants (please see Chapter 4 for details). All 1344 participants from the 20-year follow-up were included in this analysis. Comparison between the individuals who did and did not participate in the 20-year follow-up has been presented in Chapter 8.

Quantitative analysis of CUVAF

A camera system developed by Coroneo and colleagues was used to obtain CUVAF images for each participant. The camera system included a height adjustable table equipped with subject head-rest, camera positioning assembly, digital single-lens reflex camera (Nikon D100 (Nikon, Melville, New York, USA)), 105 mm f/2.8 Micro Nikkor (Nikkor, Melville, New York, USA) lens, and filtered electronic flash. The nasal and temporal regions of both eyes were photographed at a magnification of x0.94 in
total darkness. All images were saved in RGB format at the D100 settings of JPEG Fine (compression; 1:4) and large resolution (3000x2000 pixels). The area of fluorescence in square millimetres (mm²) for each photograph was determined using Adobe Photoshop CS4 Extend (Adobe Systems Inc., San Jose, California, USA). Reliability of CUVAF as a biomarker of sunlight exposure has been validated previously.281

Questionnaire
As part of the REHS, participants were asked to complete questionnaires regarding their socio-economic status, medical history, and sun exposure. In relation to sun exposure, participants were asked to estimate time spent outdoors, with four possible responses to the question “In the summer, when not working at your job or at school, what part of the day do you spend outside?” Responses were ‘none’, ‘<¼ of the day, approximately half of the day’ and ‘>¾ of the day’. ‘None’ and ‘<¼ of the day’ groups were combined due to low numbers in the ‘none’ category. Only socio-economic status and medical history questionnaires were available for TEST and BATS cohorts.

Study analysis was divided into three main components: (1) comparison of CUVAF areas between the TEST and BATS cohorts to identify the effect of latitude; (2) a classical twin study using the TEST and BATS cohorts to estimate heritability of CUVAF; (3) a meta-GWAS analysis of CUVAF to identify common variants associated with this measurement by pooling data from all three cohorts.

Analytical Approach for Classical Twin Study
The classical twin model based on the multivariable linear structural equation was applied using OpenMx package in the statistical software R version 2.15.1 (R Foundation for Statistical Computing; http://www.r-project.org/). This model assumes the phenotypic variation observed between the MZ and DZ twins are due to variation in
additive genetic (A), common environmental (C), and unique environmental (E) effects.

To determine the heritability of CUVAF, deterioration in the model fit was assessed by dropping each component in a hierarchical order from the full model. We then compared each of the nested sub-models with the full model by chi-squared tests. We used the Akaike information criterion to determine the best fitting model in which variation was explained by as a few variables as possible. Before the model fitting analyses, CUVAF was adjusted for age and sex.

**Genotyping and quality control**

The TEST and BATS participants were genotyped using the Illumina Human 660W-Quad bead chip. A total of 1903 individuals from the REHS (some did not participate in the eye study) were genotyped in two different batches: 1593 individuals were genotyped in 2009 using the Human 660W-Quad bead chip and a further 310 individuals were genotyped in 2012 using the Illumina Human-OmniExpress bead chip.

As part of quality control (QC), the data were filtered by single nucleotide polymorphism (SNP) call rate <0.95, a Hardy-Weinberg equilibrium (HWE) p-value < $10^{-6}$ and a minor allele frequency (MAF) >0.01. To exclude population outliers, a principal component (PC) analysis was carried out using SNPs with genotyping rate >0.98. Identical SNPs with the 1000 Genome panel were identified for the PC analysis. All the samples beyond 6 SDs from PC1 and PC2 of the 1000 Genomes Project British population were excluded. Individuals with identity-by-descent estimate > 0.24 with another participant were also removed from the analysis.
**Genotype imputation**

The TEST and BATS cohorts were imputed against the August 4, 2010 version of the publicly released 1000 Genomes Project European genotyping using MACH software. Likewise, the Raine Study cohort was imputed against the November 23, 2010 version of the 1000 Genome Project European genotyping using MACH. We included SNPs with a MAF >0.01 and MaCH Rsq of greater than 0.3.

**Genome-wide Association (GWA) Studies of CUVAF**

The GWAS of twin cohorts and the Raine Study were conducted separately. Associations of 7,773,124 SNPs (439,454 genotyped) of 295 families from the TEST and BATS cohorts were performed using MERLIN with addition of age, sex and latitude as covariates in a linear model. For the Raine Study cohort, a linear regression model in R with a PLINK interface was used to determine associations between 9,131,795 SNPs (561,216 genotyped) and CUVAF. In this cohort, reported time spent outdoors showed a correlation with CUVAF ($r=0.19$, $p<0.001$). Hence, reported time outdoors was included as a covariate with age and sex for the 661 individuals who remained in the analysis. An inverse variance-weighted meta-analysis with common SNPs imputed in both cohorts ($n = 5,003,381$) was conducted using METAL. A pathway analysis was performed by combining SNP p values obtained from the REHS and TEST/BATS cohort analyses in the Versatile Gene-based Association Study (VEGAS) analysis tool.
Results

After quality control, 590 participants of 295 families from TEST/BATS and a total of 661 unrelated participants from the REHS had complete data available and were included in the present study. Characteristics of these three groups are given in Table 7-1. The age range varied between the cohorts, with the mean (range) age being 12 (5-51), 19 (13-28) and 20 (18-22) years in the TEST, BATS and REHS, respectively. The TEST and BATS cohorts included more female participants (55% and 57%, respectively), whereas the REHS cohort included more male participants (52%). Sex and age were correlated with CUVAF at r = -0.09 (p=0.001) and r = 0.07 (p=0.013) respectively in the pool of three cohorts.

Effect of latitude in distribution of CUVAF

We compared areas of CUVAF in the two twin cohorts based on their geographical locations. Of the 590 individuals, 146 were from Tasmania (Hobart latitude, 42.88° S) and 444 were from Queensland (Brisbane latitude, 27.47° S). The median CUVAF area was greater in individuals from Queensland (45.41 mm², interquartile range [IQR]: 26.77, 68.50) compared with the individuals from Tasmania (28.74 mm², IQR: 15.01, 42.34) (p<0.001). To ensure that this difference did not result from a confounding effect of a difference in age and sex distribution within the two twin cohorts, we adjusted CUVAF for age and sex before the comparison. The difference remained, with median CUVAF area being 43.36 mm² (IQR: 26.54, 66.69) in individuals from Queensland and 30.90 mm² (IQR: 18.96, 47.31) in individuals from Tasmania (p<0.001). Moreover, a similar difference was present when the analysis was restricted to younger twin pairs (10-20 years) (BATS: 47.43 mm² [IQR: 27.92, 66.4] vs TEST: 37.53 mm² [IQR: 23.64, 48.53]; p=0.006).
CUVAF heritability

Of the 295 twins pairs included in the analysis, 150 (50.8%) were MZ twins. The pairwise correlation coefficient of CUVAF was 0.88 for MZ twins and 0.70 for DZ twins. The slightly higher correlation of MZ twins suggests a stronger common environmental contribution for the phenotype variance compared with the genetic contribution under a classical twin model. This observation was confirmed by univariate model fitting. The best-fit model was ACE model (A indicates additive genetic effects; C, common environment; E, unique environment effects) adjusted by age and sex. With this model, we estimated the variation explained by the additive genetic component to be 0.37 (95% confidence interval [CI], 0.22-0.56) while the common environment component explained 0.5 (95%CI, 0.29-0.71) of the variability of the trait.

Genome-wide association (GWA)

A genome-wide significant locus rs1060043 at (p=3.193x10^-8) and suggestive loci are shown in Figure 7-1 and summarized in Table 7-2. The effect size of the CUVAF increasing allele was 11.34 mm² per copy. Figure 7-2 shows the region around the rs1060043 locus. The top 10 CUVAF-associated genes obtained from the gene-based test using VEGAS and SNP meta-analysis p-value estimates are displayed in Table 7-3.
Table 7-1: Demographic characteristics of Conjunctival UV autofluorescence (CUVAF) study participants

<table>
<thead>
<tr>
<th></th>
<th>TEST</th>
<th>BATS</th>
<th>REHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>146</td>
<td>444</td>
<td>661</td>
</tr>
<tr>
<td>Number of families</td>
<td>73</td>
<td>222</td>
<td>661</td>
</tr>
<tr>
<td>Mean age in years (range)</td>
<td>12 (5-51)</td>
<td>19 (13-28)</td>
<td>20 (18-22)</td>
</tr>
<tr>
<td>Number of MZ vs DZ twins</td>
<td>26/47</td>
<td>124/98</td>
<td>-</td>
</tr>
<tr>
<td>Gender (%females)</td>
<td>55%</td>
<td>57%</td>
<td>48%</td>
</tr>
<tr>
<td>Median CUVAF (IQR)</td>
<td>28.7 (15.0,42.3)</td>
<td>45.4 (26.7,68.5)</td>
<td>44.2 (20.3,69.8)</td>
</tr>
</tbody>
</table>
Figure 7-1 | Manhattan Plot of the Meta-analysis Association p values for Conjunctival UV Autofluorescence. Single nucleotide polymorphisms (SNPs) are plotted based on chromosomal position vs the logarithm of the p values. Red line denotes genome-wide significance (p $< 5 \times 10^{-8}$). The SNPs above the blue line represent the suggestive loci.
Table 7-2 | Top five loci associated with conjunctival UV autofluorescence (CUVAF).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Closest Locus</th>
<th>A1/A2</th>
<th>TEST/BATS</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect</td>
<td>SE</td>
<td>p-value</td>
<td>Effect</td>
</tr>
<tr>
<td>rs1060043</td>
<td>19</td>
<td>SLC1A5</td>
<td>A/G</td>
<td>7.32</td>
<td>2.70</td>
<td>0.006</td>
<td>16.71</td>
</tr>
<tr>
<td>rs1558253</td>
<td>17</td>
<td>SPAG9/NME1</td>
<td>T/G</td>
<td>-20.89</td>
<td>3.89</td>
<td>8.47x10^{-8}</td>
<td>-8.92</td>
</tr>
<tr>
<td>rs990320</td>
<td>3</td>
<td>C3orf58</td>
<td>T/C</td>
<td>-6.97</td>
<td>2.02</td>
<td>0.00058</td>
<td>-7.13</td>
</tr>
<tr>
<td>rs7309814</td>
<td>12</td>
<td>HDAC7</td>
<td>C/G</td>
<td>16.89</td>
<td>4.56</td>
<td>0.00021</td>
<td>12.91</td>
</tr>
<tr>
<td>rs1213</td>
<td>9</td>
<td>MSANTD3</td>
<td>T/C</td>
<td>-34.68</td>
<td>10.91</td>
<td>0.0014</td>
<td>-35.53</td>
</tr>
</tbody>
</table>

|          |     |               |       |           |          |          |           |          |          |           |          |          |
Figure 7-2 | Association of variants at the \textit{SLC1A5} locus. P values (-log10) of SNP association with conjunctival UV autofluorescence in the meta-analysis are plotted against their positions at the \textit{SLC1A5} locus. SNPs are coloured to display their linkage disequilibrium (LD) with rs1060043.
Table 7-3 | VEGAS pathway analysis results for the ten most significant genes associated with conjunctival UV autofluorescence (CUVAF).

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Gene</th>
<th>Number of SNPs</th>
<th>Start Position</th>
<th>Stop Position</th>
<th>Test Statistic</th>
<th>p-value</th>
<th>Best-SNP</th>
<th>SNP p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>IQCF3</td>
<td>32</td>
<td>51837608</td>
<td>51839916</td>
<td>260.975</td>
<td>7.80x10^{-5}</td>
<td>rs9836804</td>
<td>6.77x10^{-6}</td>
</tr>
<tr>
<td>8</td>
<td>PXMP3</td>
<td>110</td>
<td>78055048</td>
<td>78075079</td>
<td>518.863</td>
<td>7.90x10^{-5}</td>
<td>rs7008266</td>
<td>8.26x10^{-6}</td>
</tr>
<tr>
<td>10</td>
<td>ARMELT1</td>
<td>81</td>
<td>14901256</td>
<td>14919989</td>
<td>354.384</td>
<td>1.37x10^{-4}</td>
<td>rs2688849</td>
<td>1.02x10^{-5}</td>
</tr>
<tr>
<td>14</td>
<td>TRMT5</td>
<td>62</td>
<td>60507919</td>
<td>60517535</td>
<td>285.471</td>
<td>1.77x10^{-4}</td>
<td>rs10129952</td>
<td>5.68x10^{-4}</td>
</tr>
<tr>
<td>9</td>
<td>FANCC</td>
<td>152</td>
<td>96901156</td>
<td>97119812</td>
<td>943.041</td>
<td>1.78x10^{-4}</td>
<td>rs4647558</td>
<td>3.57x10^{-5}</td>
</tr>
<tr>
<td>10</td>
<td>HSPA14</td>
<td>73</td>
<td>14920266</td>
<td>14953746</td>
<td>339.968</td>
<td>1.79x10^{-4}</td>
<td>rs2688849</td>
<td>1.02x10^{-5}</td>
</tr>
<tr>
<td>3</td>
<td>C3orf58</td>
<td>105</td>
<td>145173602</td>
<td>145193895</td>
<td>1003.223</td>
<td>2.18x10^{-4}</td>
<td>rs1075113</td>
<td>3.37x10^{-6}</td>
</tr>
<tr>
<td>16</td>
<td>SNX20</td>
<td>72</td>
<td>49264386</td>
<td>49272667</td>
<td>455.891</td>
<td>2.60x10^{-4}</td>
<td>rs6500327</td>
<td>4.40x10^{-5}</td>
</tr>
<tr>
<td>10</td>
<td>SLIT1</td>
<td>277</td>
<td>98747784</td>
<td>98935673</td>
<td>1073.202</td>
<td>2.65x10^{-4}</td>
<td>rs2636813</td>
<td>1.13x10^{-5}</td>
</tr>
<tr>
<td>1</td>
<td>CD1A</td>
<td>63</td>
<td>156490550</td>
<td>156494682</td>
<td>433.266</td>
<td>4.28x10^{-4}</td>
<td>rs614164</td>
<td>2.91x10^{-4}</td>
</tr>
</tbody>
</table>
Discussion

A strong relationship between CUVAF and sun-related ocular damage has been reported previously, suggesting that CUVAF could serve as a useful biomarker of ophthalmohelioses. In this study, we investigated the genetic characteristics of CUVAF. Given the possible confounding effect of geographical location of CUVAF, we initially explored the levels of CUVAF over two geographical regions defined by latitude in two ethnically homogeneous twin cohorts of European ancestry and identified smaller areas of CUVAF in individuals from a region of lower ambient UV radiation (Tasmania). Although previous studies report that individuals from a region of higher ambient UV radiation (Brisbane) spend less time outdoors compared with other regions of Australia, including Tasmania, the intensity of UV exposure in Tasmania is lower. The finding of greater areas of CUVAF in Brisbane is consistent with previous work by Wlodarczyk et al., who reported Queensland had double the rate of surgery for pterygia per 100,000 when compared with Tasmania. Thus, pterygia may well be a sensitive indicator of UV exposure because the cornea focuses peripheral incident light approximately 20-fold onto the usual limbal location of pterygia.

We assessed heritability of CUVAF and have found that the additive genetic effects are responsible for variance of as much as 0.37 in detected CUVAF areas, indicating that genes are a significant contributor to variation in CUVAF. This present finding corroborates earlier evidence showing that the tendency to develop pterygia may be inherited. Hecht identified 11 early-onset pterygia cases in two generations resident mainly in the Midwestern United States without known extreme environmental insult and suggested a genetic-environmental model for pterygia more than two decades ago. Susceptibility to the development of pterygia is also increased in genetic conditions...
with abnormal DNA repair mechanisms,\textsuperscript{273,291,292} such as xeroderma pigmentosum,\textsuperscript{293,294} porphyria cutanea tarda,\textsuperscript{292} polymorphous light eruption and possibly Cockayne syndrome.\textsuperscript{294}

To further understand the genetic contribution to the development of CUVAF, we conducted a GWAS in both twin cohorts and the REHS cohort. The meta-analysis of GWAS allowed the identification of a significant association of rs1060043, which is located 800 base pairs upstream of the solute carrier Family 1 (Neutral Amino Acid Transporter), Member 5 (\textit{SLC1A5}) gene on chromosome 19q13. \textit{SLC1A5} is a peptide transporter gene expressed in retinal Muller cells and also serves as an effluxer of D-serine agonist in N-methyl-D-aspartate receptor sites.\textsuperscript{295} Many of the genes that belong to the \textit{SLC1} gene family and SLC families have been detected in human cornea, rabbit cornea, and corneal epithelium cells (\textit{SLC1A4}, \textit{SLC6A14}, \textit{SLC7A5}).\textsuperscript{296-298} Variants in \textit{SLC45A2} and \textit{SLC24A4} influence pigmentation traits, including iris color.\textsuperscript{299} The particular SNP identified in this study gives rise to a synonymous codon that is highly conserved in zebrafish and among multiple mammalian species, including rhesus monkeys, chimpanzees, cattle, and dogs, suggesting that this gene has a critical function in mammals. The only locus in the best VEGAS pathway result was \textit{C3orf58}. This gene and none of the other genes identified in the gene-based analysis had an ocular function.

The present study was designed to investigate whether genetic and environmental factors play a role in the development of CUVAF. This investigation had 3 important results. First, individuals living in areas with higher levels of UV radiation are more likely to have increased areas of CUVAF. Secondly, although CUVAF was caused primarily by environmental factors, genetic factors also play a role in its development. Finally, we detected a susceptibility locus related to CUVAF. Although the study
successfully demonstrated these findings, certain limitations in terms of its design and sample size must be acknowledged. For example, the environment of older twins varies, possibly owing to relocation, compared with young twin pairs growing up together. Therefore, the inclusion of older adult twin pairs may have caused a probable increased discrepancy in environmental exposures that may have diluted the apparent relationship between exposure and CUVAF. On the other hand, when the analysis was restricted to younger twins, the effect of latitude on CUVAF remained the same. Thus this finding indicated that the effect of older individuals was minimal on the representation of the young twin pairs in the present study. Moreover, single GWASs are commonly underpowered. Our twin and singleton discovery cohorts were limited in sample size, which resulted in the detection of inconsistent signals in individual cohort analysis. This issue was overcome by performing a meta-analysis that resulted in more reliable outcomes.

The associated SNP, rs1060043, can serve as a pointer to the region of the human genome where the disease-causing problem resides. For this reason, we need to take additional steps, such as sequencing DNA base pairs in that particular region of the genome, and to identify the exact genetic change involved in the disease. Then, we can determine expression levels of nearby genes as a function of genotype at each locus (eQTL) and replicate candidate susceptibility genes in other populations. These work will lead to construction of in vivo and in vitro models to study functional mechanisms.

Overall, the present findings add to a growing body of literature contributing to the understanding of CUVAF development. Further research investigating the role of genetics and the environment would assist in identifying individuals who are predisposed to ocular sun damage to recommend personalised health messages.
Chapter 8

Association of Myopia and Vitamin D Status in Young Adults

As previously discussed time spent outdoors has been found to be protective against developing myopia. However, we do not know the mechanism underlying this relationship. Some researchers propose that vitamin D may play a role in myopia, as outdoor sun exposure is the main source of vitamin D production in the human body. In this chapter I discuss our investigation of the association between serum vitamin D levels and myopia in young adults.

Background

Worldwide, including Australia, the prevalence of myopia has been increasing. Myopia prevalence varies across populations of different regions, ethnicities and age groups. In some East Asian countries, myopia is an epidemic, with as many as 80% of children estimated to be myopic. In two Australian schoolchildren cohorts who were aged 6 and 12 years at the baseline examination, the incidence of myopia was reported to be 2.2% and 4.1% over a 5- to 6-year period. Mild myopia is a relatively benign disorder and blurred vision due to elongation of the eye can be corrected with spectacles, contact lenses or laser refractive surgery. However, individuals with severe myopia are at increased risk of visual impairment and blindness due to associated conditions such as retinal detachment, retinal degeneration and choroidal...
Myopia is also associated with increased risk of age-related eye diseases including cataract and glaucoma.\textsuperscript{305,306}

Several possible mechanisms have been proposed for the development of refractive error. One of the earliest of these hypothesised that vitamin D may have a role in the development of myopia. In the 1930s and 1940s, researchers investigated the association of myopia with vitamin D status both experimentally and clinically.\textsuperscript{307} Recently, Mutti and Marks proposed that decreasing population-level vitamin D status (measured by the concentration in blood of 25-hydroxyvitamin D (25(OH)D)) may be associated with the rising prevalence of myopia. In a small cohort of 22 participants, after adjustment for age and dietary intake, myopes had lower 25(OH)D\textsubscript{3} concentrations than non-myopes.\textsuperscript{308} More recently, low serum 25(OH)D levels were found to be associated with higher myopia prevalence in a large Korean population.\textsuperscript{309}

Epidemiological studies have identified a range of potential environmental risk factors for the development of myopia.\textsuperscript{310} The rapid increase in myopia prevalence in East Asian populations points to environmental or lifestyle factors sufficient to exert an effect in a short time period. Within these factors, decreasing time spent outdoors has been identified as a potential explanatory lifestyle behaviour. In the last decade, a number of observational studies have investigated the hypothesis that greater time spent outdoors is protective against myopia.\textsuperscript{142,311-313} This is supported by findings from a recent meta-analysis of cross-sectional studies that demonstrated an inverse association between time spent outdoors and myopia prevalence.\textsuperscript{143} These findings have been substantiated in prospective population-based studies and randomised controlled trials.\textsuperscript{314,315}
In addition to the evidence of a well-grounded environmental contribution to risk, some variation in myopia and refractive error is accounted for by genetic factors. Interestingly, one of the candidate genes identified in family-based studies is the vitamin D receptor (VDR). Polymorphisms within this gene were associated with low-to-moderate myopia in Caucasians, and a polymorphism in the VDR gene start codon (Fok1) was associated with high myopia in Indians. However, these studies identified different risk alleles within the VDR, and the VDR (and related) gene polymorphisms for which there is evidence of functional effects are not those that show an association. These inconsistencies do cast some doubt on a causal role of polymorphisms in the VDR, and replication studies are needed.

In many populations the main source of vitamin D is endogenous synthesis following sun exposure of the skin. Vitamin D deficiency is reportedly widespread, and population 25(OH)D levels have been decreasing over time, possibly due to behavioural changes to decrease sun exposure. Taken together, the environmental and genetic associations and the correlative temporal pattern provide highly suggestive evidence myopia risk is linked to vitamin D-related factors.

Previous refractive error studies did not take into account individual ocular and non-ocular sun exposure when exploring the relationship of myopia and vitamin D levels. The purpose of our current study was to examine the association between serum 25(OH)D$_3$ concentrations and the prevalence of myopia, adjusting for potential confounders including a marker of ocular sun exposure, in a large cohort of young adults of mainly Northern European ancestry but with a subset of East Asian ancestry.
Methods

Study Participants

The study comprised participants who were enrolled in the 20-year follow-up of the Western Australian Pregnancy Cohort (Raine) Study conducted between March 2010 and April 2012. Chapter 4 provides detailed Raine Study Eye follow-up methodology.

Questionnaire

Each participant completed a questionnaire providing sociodemographic data, and information on current education status (studying part- or full-time) and parental myopia, i.e. whether one or both parents were myopic or short-sighted. Individuals were asked to report their time spent outdoors and had four possible responses to the question “In the summer, when not working at your job or at school, what part of the day do you spend outside?”: none, < ¼ of the day, approximately half of the day and > ¾ of the day. “None” and “<¼ of the day” groups were combined due to low numbers in the “none” category.

Assessment of myopia and ocular sun exposure

As part of a comprehensive eye examination, post-cycloplegic autorefraction was measured using the Nidek ARK-510A (NIDEK Co.Ltd, Japan) autorefractor. The mean of three consecutive measurements was recorded for each eye. Myopia was defined as mean spherical equivalent (MSE, sum of spherical error and half of cylindrical error) of both eyes ≤ -0.5 diopters (D). This definition was adopted due to it is widely used and reliability has been validated in young individuals. MSE of two eyes was calculated for each participant to determine the prevalence of myopia. We used a camera system developed by Coroneo and colleagues to derive a score for a biomarker of ocular sun-exposure by measuring conjunctival UV autofluorescence (CUVAF). The area of
fluorescence in mm$^2$ for each photograph was determined using Adobe Photoshop CS4 Extend (Adobe Systems Inc., San Jose, California, USA). Total ocular sun exposure of individuals was determined as the summed area of the CUVAF in four photographs (left and right eyes, nasal and temporal conjunctiva) of each individual. The reliability of CUVAF as a biomarker of sub-acute sunlight exposure has been previously validated.\(^{281}\)

### Assessment of serum 25(OH)D₃ concentrations

Participants provided a fasting blood sample for analysis of serum 25(OH)D₃ concentration at the age 20 years follow-up. Venous blood samples were taken from an antecubital vein after an overnight fast and samples were stored at -80°C until analysed using liquid chromatography-tandem mass spectrometry (RDDT, Victoria, Australia), according to published methodology.\(^{322}\) The inter-assay coefficients of variation ranged from 5.8% to 9.2%, at 28.2 and 180.8 nmol/L of 25(OH)D₃, respectively.

### Statistical analysis

A comparison of participants completing the 20-year follow-up with those who were part of the original cohort but did not attend an eye examination was performed using the data from the year 1 follow-up to examine the sociodemographic characteristics between two groups. These characteristics included sex, ethnicity, family structure (sole parent vs couple families), income levels and Socioeconomic Index for Areas (SEIFA) Index of Relative Advantage and Disadvantage (IRSAD) of parents/carers. For the latter, a higher score reflects higher relative socioeconomic advantage (www.abs.gov.au).

The fasting blood samples were collected year-round between March 2010 and April 2012. Samples were analysed using isotope-dilution liquid chromatography-tandem mass spectrometry (RDDT, Victoria, Australia)\(^{322}\). The seasonal component was
removed (deseasonalised) according to published methodology\textsuperscript{323} by fitting a sinusoidal model to serum 25(OH)D\textsubscript{3} concentrations incorporating the month the blood sample was taken. Serum 25(OH)D\textsubscript{2} (oral vitamin D intake) concentrations were detectable in just 10 participants, and were at levels below 7 nmol/L, therefore only deseasonalised serum 25(OH)D\textsubscript{3} concentrations were included in the analysis.

MSE and total 25(OH)D\textsubscript{3} concentrations were not normally distributed (evidence from Kolmogorov-Smirnov test, $p<0.001$ and non-linear quantile-quantile (Q-Q) plots) thus summary data are presented as medians and interquartile ranges (IQRs). The difference in serum 25(OH)D\textsubscript{3} concentrations between myopic and non-myopic participants was assessed using the Mann-Whitney U test.

We defined vitamin D status as being sufficient when concentrations of serum 25(OH)D\textsubscript{3} were $\geq 75$ nmol/L, insufficient when they were 50-74.9 nmol/L and deficient when they were $<50$ nmol/L.\textsuperscript{324} Differences between categorical variables were assessed with chi-squared tests. We used a chi-square test for trend to assess a possible dose-response relationship with myopia prevalence across categories of vitamin D status. A simple linear regression model was generated to describe the relationship between levels of 25(OH)D\textsubscript{3} and MSE using the least squares method. We used simple logistic regression to estimate the odds ratios (OR) and confidence intervals (95%CI) of myopia prevalence in relation to each covariate, testing for trend by replacing categorical predictors with a single predictor, taking category rank scores. A $p$-value of $<0.05$ was considered statistically significant. A multivariable logistic regression model was constructed to assess the association between myopia prevalence and 25(OH)D\textsubscript{3} concentration (or vitamin D status) while adjusting for age, sex and other covariates identified as being significant in univariable analysis. Separate multivariable models
containing either total CUVAF or time spent outdoors were constructed due to the expected collinearity between these.

Two subgroups were created based on self-reported ethnicity, Australians with Northern European ancestry and East Asians, and the above analyses repeated.

Statistical analyses were performed using the statistical software R version 2.15.1 (R Foundation for Statistical Computing; http://www.r-project.org/).

**Results**

Compared with individuals from the original cohort who did not participate in the 20–year follow-up, participants who attended eye examinations were more likely to be born into couple families (84% vs 61%) and families with a combined income of more than US$ 23,500 (59% vs 38%) at a time when the average income of a full-time worker in Australia was approximately US$27,500. Similarly the mean IRSAD score was higher for parents/carers of the participants from the 20-year follow-up (1039±89) compared to parents/carers of their peers who were not examined (1001±86, p<0.001). There was no significant difference in sex and ethnicity between the two groups.

Of 1344 participants who attended an eye examination, 198 (14.7%) participants did not have a 25(OH)D level measurement and 200 (14.9%) participants had incomplete clinical examination or questionnaire data. Serum 25(OH)D₃ concentration and potential confounders including age, sex, ethnicity, parental myopia, education and ocular sun exposure were available for 946 participants (70.4%); just over half (n= 480; 50.7%) of these were female. Only 837 participants had data for time outdoors and potential confounders. The mean (± standard deviation) age was 20.0±0.4 years (range 18.3 to
22.1 years) and the majority (n=798, 84.4%) of the participants had Northern European ancestry. Of the 946 participants, 837 (88.5%) reported their time spent outdoors during summer. Of these, 406 (48.5%) spent less than a quarter of an average summer day outside, 332 (39.7%) spent approximately half of their day outside and 99 (11.8%) spent the majority of their day outside. Serum 25(OH)D$_3$ concentration was lower in males (70.9 nmol/L; IQR=56.1 to 84.8) compared to females (71.7 nmol/L [IQR=58.6 to 85.2]; p=0.015) and East Asian individuals (n=60, 6.3%) had lower serum 25(OH)D$_3$ concentrations compared to their peers with Northern European ancestry (55.3 nmol/L [IQR=42.4 to 70.1] vs 73.0 nmol/L [IQR=59.6 to 87.9], p<0.001). Serum 25(OH)D$_3$ concentration increased with increasing CUVAF (Figure 8-1a) and time spent outdoors ($\beta$ estimate =9.0nmol/L increase per one category of time outdoors; standard error=1.2, p-trend<0.001).

Over one-fifth of participants (n=221 (23.4%)) had myopia (MSE $\leq$ -0.5 D). Median MSE was -1.56D (IQR= -3.19 to -0.88) and +0.44D (IQR= +0.13 to +0.75) in the myopia and non-myopia groups, respectively (p<0.001). The demographic data for myopic and non-myopic individuals are displayed in Table 8-1. Serum 25(OH)D$_3$ concentrations were significantly lower in the myopic compared to the non-myopic participants: median of 67.6 nmol (IQR=52.3 to 79.6) and 72.5 nmol (IQR=59.4 to 87.2), respectively (p=0.003). Figure 8-1b shows the positive association between serum 25(OH)D$_3$ concentrations and MSE. The prevalence of myopia decreased in association with higher 25(OH)D$_3$ concentration (OR=0.88, 95%CI 0.82 - 0.94, per 10nmol/L increase) and across categories of increasing vitamin D status (chi-square for linear trend=19.63, p<0.001). Presence of myopia was also positively associated with currently studying and parental myopia and inversely associated with higher CUVAF.
area, Northern European ethnicity and greater time spent outdoors, in univariable analyses (Table 8-2).

Table 8-3 shows the results of the multivariable logistic regression models. In the model adjusted for age and sex only, the odds of being myopic decreased with increasing 25(OH)D₃ concentration (OR: 0.88, 95%CI: 0.82 - 0.94, per 10 nmol/L increase, p<0.001), and increased across categories of decreasing vitamin D status (OR: 2.67, 95%CI: 1.72 – 4.11 for vitamin D deficiency vs sufficiency). The significant association with 25(OH)D₃ as a continuous variable was retained in the multivariable model adjusted for age, sex, ethnicity, parental myopia, education and CUVAF (OR: 0.91, 95%CI: 0.85 - 0.98 per 10nmol/L increase, p=0.013). Across categories of vitamin D status the odds of being myopic was significantly increased only in the comparison of vitamin D deficiency to vitamin D sufficiency in the fully adjusted model but there was evidence of a significant trend (OR: 1.42, 95%CI: 1.12 - 1.79 per category increase, p for trend<0.001).

In the sub-group analysis including only participants with North European background, the myopic group had a significantly lower median 25(OH)D₃ concentration compared to the non-myopic group (Figure 8-2). The odds of having myopia decreased significantly with increasing 25(OH)D₃ concentration in the age- and sex-adjusted model. This effect was no longer statistically significant after adjustment for the other factors in the fully adjusted model, although the strength of the association was very similar. Nevertheless, vitamin D deficiency was associated with an increased risk of myopia compared to vitamin D sufficiency in the fully adjusted model, with evidence of a trend across the categories (OR:1.35, 95%CI:1.04 - 1.75, p for trend=0.024).
Figure 8-1 | Simple linear regression equations of 25(OH)D₃ concentration (nmol/L) with (A) Ocular Sun Exposure (Conjunctival UV Autofluorescence) and (B) Refractive Error in young adults (● =Northern Europeans, ○ =East Asians).
<table>
<thead>
<tr>
<th>Sex (male)</th>
<th>Myopic participants, n=221 (%)</th>
<th>Non-myopic participants, n=725 (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>102 (53.8)</td>
<td>364 (49.7)</td>
<td>0.328</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern European ancestry</td>
<td>170 (76.9)</td>
<td>628 (86.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>East Asian ancestry</td>
<td>29 (13.1)</td>
<td>31 (4.3)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>22 (10.0)</td>
<td>66 (9.1)</td>
<td></td>
</tr>
<tr>
<td>Education (currently studying)</td>
<td>166 (75.1)</td>
<td>430 (59.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Parental myopia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neither parent</td>
<td>125 (56.6)</td>
<td>559 (77.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>One parent</td>
<td>65 (29.4)</td>
<td>126 (17.4)</td>
<td></td>
</tr>
<tr>
<td>Both parents</td>
<td>31 (14.0)</td>
<td>40 (5.5)</td>
<td></td>
</tr>
<tr>
<td>Time spent outdoors during summer*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than ¼ of day</td>
<td>115 (56.9)</td>
<td>291 (45.8)</td>
<td></td>
</tr>
<tr>
<td>½ day</td>
<td>69 (34.2)</td>
<td>263 (41.4)</td>
<td>0.020</td>
</tr>
<tr>
<td>more than ¾ of day</td>
<td>18 (8.9)</td>
<td>81 (12.8)</td>
<td></td>
</tr>
<tr>
<td>Conjunctival UV Autofluorescence**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1\textsuperscript{st} quartile</td>
<td>72 (32.6)</td>
<td>165 (22.8)</td>
<td></td>
</tr>
<tr>
<td>2\textsuperscript{nd} quartile</td>
<td>73 (33.0)</td>
<td>163 (22.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3\textsuperscript{rd} quartile</td>
<td>40 (18.1)</td>
<td>196 (27.0)</td>
<td></td>
</tr>
<tr>
<td>4\textsuperscript{th} quartile</td>
<td>36 (16.3)</td>
<td>201 (27.7)</td>
<td></td>
</tr>
<tr>
<td>Vitamin D status***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficient</td>
<td>51 (31.7)</td>
<td>90 (12.4)</td>
<td></td>
</tr>
<tr>
<td>Insufficient</td>
<td>100 (45.2)</td>
<td>309 (42.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Optimal</td>
<td>70 (31.8)</td>
<td>326 (45.0)</td>
<td></td>
</tr>
</tbody>
</table>

*Numbers of participants may not equal to 946, due to missing data in the time spent outdoors.

** Conjunctival UV autofluorescence was divided into quartiles; 1\textsuperscript{st} quartile: 0 to 20.80; 2\textsuperscript{nd} quartile: 20.81 to 45.10; 3\textsuperscript{rd} quartile: 45.11 to 70.40; 4\textsuperscript{th} quartile: 70.41 to 180.0 and is used here as a marker of ocular sun exposure.

*** Vitamin D deficiency is defined as a deaseasonalised serum 25(OH)D\textsubscript{3} concentration below 50 nmol/L and vitamin D insufficiency as deaseasonalised serum 25(OH)D\textsubscript{3} concentration of 50-74.9 nmol/L.
Vitamin D deficiency is defined as a deaseasonalised serum 25(OH)D3 concentration below 50 nmol/L and vitamin D insufficiency as deaseasonalised serum 25(OH)D3 concentration of 50 - 74.9 nmol/L. CUVAF was divided into quartiles; 1st quartile: 0 to 20.80; 2nd quartile: 20.81 to 45.10; 3rd quartile: 45.11 to 70.40; 4th quartile: 70.41 to 180.0. Only 837 participants included in this univariate analysis due to missing data in the time spent outdoors.

Table 8-2 | Univariable Logistic Regression Analysis of Associations with Myopia

<table>
<thead>
<tr>
<th>Covariate</th>
<th>All participants (n=946)</th>
<th>Northern European participants (n=798)</th>
<th>East Asian participants (n=60)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>p-value</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Age, per year</td>
<td>1.25 (0.90 - 1.74)</td>
<td>0.180</td>
<td>1.26 (0.85 - 1.85)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female Reference</td>
<td></td>
<td>Male 1.17 (0.87- 1.59)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.07 (0.76 - 1.51)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>Reference</td>
<td></td>
<td>Northern European 0.51 (0.35 - 0.76)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.51 (0.35 - 0.76)</td>
</tr>
<tr>
<td>25(OH)D3 level, per 10 nmol/L increase</td>
<td>0.88 (0.82 – 0.94)</td>
<td>&lt;0.001</td>
<td>0.92 (0.85 - 0.99)</td>
</tr>
<tr>
<td>Vitamin D status*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insufficient</td>
<td>1.51 (1.07 – 2.13)</td>
<td>&lt;0.001</td>
<td>1.57 (1.09 – 2.28)</td>
</tr>
<tr>
<td>Deficient</td>
<td>2.63 (1.71 – 4.05)</td>
<td>0.019</td>
<td>2.03 (1.19 – 3.42)</td>
</tr>
<tr>
<td>CUVAF, per mm²</td>
<td>0.99 (0.98 - 0.99)</td>
<td>&lt;0.001</td>
<td>0.98 (0.98 - 0.99)</td>
</tr>
<tr>
<td>1st quartile</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>1.03 (0.69 - 1.52)</td>
<td>0.896</td>
<td>1.02 (0.65 - 1.59)</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>0.47 (0.30 - 0.72)</td>
<td>&lt;0.001</td>
<td>0.55 (0.34 - 0.89)</td>
</tr>
<tr>
<td>4th quartile</td>
<td>0.41 (0.26 - 0.64)</td>
<td>&lt;0.001</td>
<td>0.45 (0.27 - 0.75)</td>
</tr>
<tr>
<td>Time Outdoors in Summer***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than ¼ of day</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>½ day</td>
<td>0.66 (0.47 - 0.93)</td>
<td>0.019</td>
<td>0.66 (0.45 - 0.98)</td>
</tr>
<tr>
<td>more than ¾ of day</td>
<td>0.56 (0.31 - 0.96)</td>
<td>0.042</td>
<td>0.64 (0.34 - 1.13)</td>
</tr>
<tr>
<td>Parental Myopia</td>
<td>One parent affected 2.31 (1.61 - 3.29)</td>
<td>0.001</td>
<td>2.44 (1.64 - 3.61)</td>
</tr>
<tr>
<td></td>
<td>Both parents affected 3.47 (2.08 - 5.75)</td>
<td>&lt;0.001</td>
<td>3.01 (1.68 - 5.30)</td>
</tr>
<tr>
<td>Education</td>
<td>Not studying Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Currently studying 2.07 (1.48 - 2.93)</td>
<td>&lt;0.001</td>
<td>1.84 (1.27 - 2.68)</td>
</tr>
</tbody>
</table>

*Vitamin D deficiency is defined as a deaseasonalised serum 25(OH)D3 concentration below 50 nmol/L and vitamin D insufficiency as deaseasonalised serum 25(OH)D3 concentration of 50-74.9 nmol/L. **CUVAF was divided into quartiles; 1st quartile: 0 to 20.80; 2nd quartile: 20.81 to 45.10; 3rd quartile: 45.11 to 70.40; 4th quartile: 70.41 to 180.0. *** Only 837 participants included in this univariate analysis due to missing data in the time spent outdoors.
**Table 8-3 | Adjusted Multivariable Logistic Regression Analysis of the association between 25(OH)D₃ Levels and presence of Myopia**

<table>
<thead>
<tr>
<th></th>
<th>Age and Sex Adjusted Model</th>
<th>Multivariable Model*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95%CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>25(OH)D₃ (per 10 nmol/L)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participants</td>
<td>0.88 (0.82 - 0.94)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Northern European participants</td>
<td>0.92 (0.85 - 0.99)</td>
<td>0.037</td>
</tr>
<tr>
<td>East Asian participants</td>
<td>0.66 (0.46 – 0.88)</td>
<td>0.009</td>
</tr>
<tr>
<td>Vitamin D status**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficient</td>
<td>2.67 (1.72 – 4.11)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insufficient</td>
<td>1.51 (1.07 – 2.13)</td>
<td>0.019</td>
</tr>
<tr>
<td>Sufficient</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td>Northern European participants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficient</td>
<td>2.01 (1.17 – 3.39)</td>
<td>0.010</td>
</tr>
<tr>
<td>Insufficient</td>
<td>1.57 (1.09 – 2.28)</td>
<td>0.017</td>
</tr>
<tr>
<td>Sufficient</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td>East Asian participants***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficient</td>
<td>6.24 (1.07 – 55.27)</td>
<td>0.060</td>
</tr>
<tr>
<td>Insufficient</td>
<td>3.59 (0.65 – 29.71)</td>
<td>0.173</td>
</tr>
<tr>
<td>Sufficient</td>
<td>Ref</td>
<td></td>
</tr>
</tbody>
</table>

*Multivariate model adjusted for age, sex, ethnicity (where applicable), parental myopia, CUVAF (mm²), education
** Vitamin D deficiency is defined as a deaseasonalised serum 25(OH)D₃ concentration below 50 nmol/L and vitamin D insufficiency as deaseasonalised serum 25(OH)D₃ concentration of 50-74.9 nmol/L.
***The sample size was too small to undertake further adjustment in this subgroup.
Figure 8-2 | Comparison of 25(OH)D₃ Levels (nmol/L) Between Myopic and Non-myopic Participants with Northern European and East Asian Ancestries.
Among the 60 participants with East Asian background, 48.3% (n=29) were myopic. Serum 25(OH)D₃ concentration was significantly lower in the myopic, compared to non-myopic East Asian participants (53.0 nmol/L [IQR: 11.2 to 94.8] vs 61.0 nmol/L [IQR:13.8 to 108.3], p=0.035). Similarly, higher 25(OH)D₃ concentration was associated with a decreased odds of having myopia in the age- and sex-adjusted model as well as fully adjusted model (Table 8-3) but the trend across categories of vitamin D status was not statistically significant (p=0.078), probably because of the small sample size. The point estimates suggest that the odds of being myopic were increased 6-fold in association with vitamin D deficiency, compared to sufficiency. We were unable to undertake further adjustment in this subgroup because of the small sample size.

In analyses of both the full sample and the Northern European sub-group, CUVAF, parental myopia and education remained significant in the adjusted multivariable regression model for myopia that also included serum 25(OH)D₃ concentration (all p \leq 0.001).

Separate models constructed to adjust for time spent outdoors provided similar results to those adjusting for CUVAF. For example, in the full sample, the odds of being myopic decreased with increasing 25(OH)D₃ concentrations (adjusted OR: 0.89, 95%CI: 0.83 - 0.97 per 10 nmol/L increase, p=0.008). Similarly, the adjusted odds of having myopia were 0.93 (95%CI: 0.85 - 1.01) and 0.56 (95%CI: 0.33 – 0.83) with every 10 nmol/L increase in 25(OH)D₃ concentrations in Northern European (p=0.091) and East Asian subsets (p=0.011), respectively. The odds of being myopic were significantly increased in the comparison of vitamin D deficiency to vitamin D sufficiency in the fully adjusted model (OR: 2.14, 95%CI: 1.27 – 3.58, p=0.008). In individuals with Northern European ancestry
(n=707), the odds of being myopic were not significantly different between those who were vitamin D deficient and sufficient (OR: 1.78; 95%CI: 0.96 - 3.27, p=0.065).

East Asian ethnicity was associated with both lower 25(OH)D₃ levels and greater prevalence of myopia. We therefore examined these associations in more detail, by comparing a model containing only ethnicity as a risk factor, with a second model that included ethnicity and serum 25(OH)D₃ concentrations. Addition of serum 25(OH)D₃ concentration improved the model fit and East Asian ethnicity remained significant with a Akaike information criterion (AIC) difference of 7.8 (p<0.001) and the effect of ethnicity slightly reduced (β estimate of -0.664 to -0.518).

Finally, a binary variable for serum 25(OH)D concentration was generated using a cut-off point of < 75 nmol/L vs. ≥ 75 nmol/L. We used this to explore the determinants of vitamin D insufficiency by setting this variable as the dependent parameter and using a model that included factors, which were associated with vitamin D concentrations as independent variables. Presence of myopia was associated with increased odds of being vitamin D insufficient in this model (Adjusted OR: 1.63; 95%CI: 1.18-2.27; p=0.003).
Discussion

In this study of young adults at 20 years of age, lower serum 25(OH)D₃ concentrations were associated with a higher prevalence of myopia as previously identified in Caucasian teens\textsuperscript{308,28} and Korean adolescents.\textsuperscript{309} This association could be evidence of an underlying biochemical mechanism between serum 25(OH)D concentrations and myopia, and explain previous findings that greater time spent outdoors is associated with reduced risk of myopia development. Alternatively, the 25(OH)D₃ concentration could be simply a biomarker of sun exposure, with some other non-vitamin D element being a protective factor. Although it might be suggested that our results reflect reverse causality, whereby myopic young adults prefer to spend more time indoors and thus have lower self-reported time outdoors and lower 25(OH)D₃ levels, this explanation is not supported by the findings from the study of Jones-Jordan et al.\textsuperscript{325} In that prospective study, sports/outdoor activities were decreased in myopic subjects three years before onset, thus pointing to a causal relationship between outdoor exposure and myopia development. Hence, the results of the present study should not be interpreted in the sense of reflecting reverse causality. Our results are consistent with previous findings of environmental or demographic risk factors for myopia including Asian ethnicity, history of parental myopia, higher education, lower levels of CUVAF and less time spent outdoors.\textsuperscript{326,327}

Country of origin, genetic traits and cultural behaviour are important factors determining serum 25(OH)D₃ concentrations.\textsuperscript{328} A higher risk of vitamin D deficiency in individuals with Asian, Middle-Eastern and African origins is well-described.\textsuperscript{329-333} Lower serum 25(OH)D levels in dark-skinned individuals are likely due to both behavioural factors\textsuperscript{334}
and decreased efficiency of vitamin D production in darker-skinned individuals. Lower 25(OH)D₃ concentrations among East Asian individuals have been reported previously. Therefore, the markedly lower 25(OH)D₃ concentrations in participants with East Asian compared to Northern European ancestry in our cohort was not unexpected. However, it was interesting to note that the East Asian group also had a higher prevalence of myopia. Further probing of this finding led us to identify that adjusting for 25(OH)D concentration accounted for part of the association between ethnicity and prevalence of myopia. Hence we suggest that lower vitamin D status is one factor that mediates the difference in myopia prevalence between various ethnicities.

Many mechanisms have been postulated to explain the apparent protective effect of time spent outdoors for myopia development. Given that the association between near work and myopia is weak and inconsistent, substitution of outdoor activities for near work does not appear to be the important factor nor does participating in sport per se. In animal models, emmetropisation is an active process by which optical defocus adjusts the rate of axial elongation during growth and development of the eye. Therefore, it was proposed that improved retinal image quality during distance viewing with a smaller pupil size and accommodative errors may inhibit ocular growth, thus decreasing the risk of development of myopia. However, evidence from animal models did not support this hypothesis. Another hypothesis was that greater light intensities outside may alter the release of dopamine, known to inhibit ocular growth, in the retina and this has been tested in chick models. In support of this hypothesis, high ambient lighting was found to retard development of myopia in chick and rhesus monkey studies.
Matrix metalloproteinases (MMP) have been implicated in the scleral remodeling of experimental myopia\textsuperscript{344} and in the development of simple myopia.\textsuperscript{345} Recently an inverse correlation between blood MMP9 and 25(OH)D levels has been found in submariners.\textsuperscript{346} A possible explanation for this might be that suboptimal serum 25(OH)D levels (from either sun avoidance\textsuperscript{347} or Western diets that are typically low in vitamin D\textsuperscript{348}) may modulate blood and perhaps tissue levels of MMP with a downstream effect on scleral morphology and refraction.

This study is unique in having a large sample size and an objective measure of ocular sun exposure that correlates well with time spent outdoors. The only method to determine vitamin D deficiency or sufficiency in an individual is to measure circulating 25(OH)D concentration. In this study, serum 25(OH)D concentrations were measured as the common pathway in vitamin D metabolism for both dietary and sun-induced vitamin D. Moreover, this study contains a wealth of supporting data reinforcing the identified association between myopia and vitamin D status. One caveat of this study that must be acknowledged is that this study has cross-sectional data collection from a pregnancy cohort and refractive error was measured at only one point in time; therefore causality cannot be inferred. Furthermore, the highest incidence of myopia occurs in children aged 5-15 years. As no data on 25(OH)D concentrations were available from younger time points at the time of this analysis, we have made the assumption that 25(OH)D levels in young adults are consistent with those in younger years during which myopia may have developed. Information on time spent outdoors and other potential risk factors including ethnicity, education and parental myopia was extracted from self-reported questionnaires and is thus subject to recall bias, although this is likely to be non-differential across the groups defined by having
myopia or not, so would have resulted in a bias toward null findings, i.e. our findings may be an underestimate of the true associations. It is also possible that the findings in participants are different from those who have not participated, given that more than 50% of the original cohort were lost to follow-up, and a relatively high proportion (30%) of participants in the twenty-year follow-up had incomplete data. Further investigations are necessary to validate the findings from this cohort and to assess effects of population differences.

In conclusion, findings from this study suggest that there could be a biological association between the risk of myopia and reduced 25(OH)D₃ concentrations within different populations. However it is important to bear in mind that the 25(OH)D₃ could be acting as a proxy for ocular sun exposure, with the latter the important factor. Therefore, future studies prospectively investigating the effects of 25(OH)D₃ concentrations and ocular sun exposure in the development of refractive error are warranted.
Chapter 9

Genome-Wide Association Studies of Corneal Astigmatism and Curvature in Australians with Northern European Ancestry

The clear uniform curvature of the healthy cornea focuses light on the retina. Abnormalities of corneal curvature result in refractive errors and various corneal disorders. For example, myopia is related to steeper cornea, and irregular or toric shape of the cornea causes astigmatism. Additionally, corneal diseases such as keratoconus, pterygium and Fuchs’ endothelial dystrophy cause changes in corneal structure. Better understanding the variation in the corneal structure and the associated genetic and environmental factors will allow us to identify the cellular and molecular pathways that may help in preventing and treating corneal disease and myopia. So, in this chapter we report the genetic variations of corneal astigmatism and corneal curvature in individuals with Northern European ancestry.

Background

The majority of light refraction occurs at the air-tear film/cornea interface, as light enters the eye. Corneal astigmatism occurs when the light rays entering the eye are not focused into a single point on the retina and results in blurred vision. It is typically categorised into regular and irregular astigmatism. Regular astigmatism presents when the horizontal and vertical meridians are at a right angle. This type of astigmatism is further classified into
with-the-rule (WTR) and against-the-rule (ATR) depending on the axis of the astigmatism.

WTR astigmatism is more prevalent in the younger populations but a transition to ATR astigmatism is observed with aging, probably due to corneal change. Regular astigmatism is often corrected with spherocylindrical lenses. Irregular astigmatism occurs when the two meridians of the cornea are not at right angles to each other. Unlike regular astigmatism, irregular astigmatism is difficult to measure and cannot be corrected with glasses. Often, patients use rigid contact lenses or have refractive surgery.349

Reported prevalence of corneal astigmatism varies across populations. Approximately, 40% of participants from a Singapore Chinese population were astigmatic (as defined by cylindrical autorefraction readings >0.5 diopter [D]).350,351 In separate studies using identical definitions, more than 50% of rural Asian Indian and Persian populations were found to have astigmatism.352,353 Interestingly, with marginally higher astigmatic groupings (either ≥0.75D or ≥1.00D), the age-adjusted prevalence of astigmatism was reported to be just over 35% in Caucasian populations from Australia and the United States.354,355

Despite much work, the aetiology of astigmatism remains poorly understood. Nonetheless, genetic and environmental factors have been suggested to have important roles in its development. Using a classical twin study, Hammond and colleagues356 reported that dominant and additive genetic effects accounted for approximately 46% to 79% of the phenotypic variance in corneal astigmatism. Similarly, Dirani et al.357 found a heritability of 50% and 60% in men and women, respectively, while Grjibovski and colleagues358 calculated an overall heritability of 63% for corneal astigmatism.
Corneal curvature (CC) is another important biometric feature, which has been recently interrogated at the genetic level. Keratoconus is a disease characterized by a conical-shaped cornea and irregular astigmatism. Patients with this corneal condition often experience vision distortion, multiple images, and sensitivity to light. In addition to keratoconus, corneal irregularities are also associated with refractive error and Marfan syndrome.

Variation in corneal curvature is dependent on ethnic background, geographical, as well as environmental conditions, age, and stature. CC is highly heritable, with previous studies revealing heritability estimates ranging between 60% and 95%. Improved understanding of the genetic architecture of this biometric trait will aid in determining the molecular mechanisms of blinding eye disorders, and contribute to our ocular development and evolutionary biology knowledge.

A limited number of genome-wide association studies (GWASs) investigating corneal parameters have been conducted. Since central corneal thickness (CCT) has been found to be one of the most heritable human traits, the best studied corneal trait by GWAS to date is CCT. In the first published GWAS for CCT, the zinc finger 469 locus on chromosome 16q24 was identified in Australian and UK twin cohorts, and subsequently confirmed in other populations. Additional quantitative trait loci for CCT have been identified on chromosomes 1p34 COL8A2, 6q14 IBTK, 9q34 COL5A1, and 15q26 CHSYI genes, and two intergenic regions on chromosomes 7q11 and 9p23. More recently, multiple genes associated with CCT and keratoconus were identified in a large meta-analysis encompassing multiple cohorts.
The only published GWAS for CC is from a Singaporean Asian population, in which the significant associations of single nucleotide polymorphisms (SNPs) in FK506 binding protein-rapamycin complex-associated protein 1 (FRAP1) and platelet-derived growth factor receptor alpha (PDGFRA) genes with corneal curvature were reported. In a separate study, Fan and colleagues found that this locus was also associated with CA in a Singaporean Asian population. As is the case with other quantitative traits and complex disease, CA and CC are likely to be determined by many genes, with ever larger GWASs likely to lead to the identification of additional associated SNPs. Furthermore, it is unknown whether genes found to be significantly associated with CA and CC in Asian population would be relevant to other racial groups. In previous studies, ethnic and environmental backgrounds constitute an important determinant of CC and CA. Thus, we aimed to test whether the PDGFRA and FRAP1 genes found to be associated with CA and CC in a Singaporean Asian population also cause CA and determine the CC in Australians of Northern European ancestry. We also present results from a genome-wide meta-analysis for CA and CC in more than 2,700 people.

**Methods**

**Ethical approval**

This study was conducted in accordance with the Declaration of Helsinki, and informed consent was obtained from all adult participants and at least one parent of the child participants before examination. Approval for this study was obtained from the Human Research Ethics Committees of the University of Western Australia, University of Tasmania, Royal Victorian Eye and Ear Hospital, and Queensland Institute of Medical Research.
Sample populations, data collection and analysis

Two Australian cohorts of Northern European ancestry were included in this study. In both studies, corneal astigmatism was calculated as the absolute difference between horizontal and vertical keratometry readings. An inverse normal transformation was applied to the average corneal astigmatism of both eyes and used for analysis. Participants who had a pterygium or had previously undergone ocular surgery were excluded from analysis.

The first cohort comprised of participants who are enrolled in the REHS (Chapter 4). As part of the examination, corneal curvature was measured in horizontal and vertical meridians with the IOLMaster V:5 (Carl Zeiss Meditec AG, Jena, Germany). Three consecutive measurements of corneal curvature within 0.3D, within each meridian, were recorded with careful alignment and focus (Appendix 2.1). DNA samples and consent for GWASs were available from the previous assessments. Genotype data were generated using the genome-wide Illumina 660 Quad Array at the Centre for Applied Genomics (Toronto, Ontario, Canada). As part of quality control, we investigated any individuals who were related with \( \pi >0.1875 \) (second- or third-degree relatives) and excluded individuals with a higher proportion of missing data. We also excluded people who had a high degree of missing genotyping data (>3%). The data were filtered for a Hardy–Weinberg equilibrium \( p \) value \( >5.7 \times 10^{-7} \), single nucleotide polymorphism (SNP) call rate >95%, and a minor allele frequency >0.01. We conducted PC analysis and constructed the first five principal components for a subset of 42,888 SNPs that were not in linkage disequilibrium (LD) with each other using the EIGENSTRAT program. We also performed the GWAS imputation of 22 autosomes in the MACH v1.0.16 software using the CEU samples from HapMap phase2 build 36 release 22. A linear regression model in R with a PLINK
interface was used to determine associations between SNPs and corneal astigmatism. The model was adjusted for age, sex, and the first two principal components that accounted for the population stratification.

The second cohort comprised participants from the Twins Eye Study in Tasmania (TEST) and the Brisbane Adolescent Twin Study (BATS). In both studies, corneal curvature was measured using a Humphrey-598 Automatic Refractor/Keratometer (Carl Zeiss Meditec, Inc., Miami, FL), and there was no significant difference between the measurements of the right and left eyes (Student t test, p value=0.24). In the BATS and the TEST, DNA was obtained from either saliva or peripheral blood samples. Blood was collected in tubes containing ethylenediaminetetraacetic acid and saliva samples were collected using an Oragene saliva kit (DNA Genotek, Inc., Kanata, ON, Canada). The extracted DNA from these samples was genotyped on the Illumina HumanHap 610W Quad Arrays (Illumina Inc., San Diego, CA). The majority of the BATS samples were genotyped at deCODE Genetics (Sturlugata 8; Reykjavik, Iceland) as part of a larger project. All TEST samples and a small proportion of the BATS samples (50) were genotyped at the Centre for Inherited Disease Research (CIDR; Baltimore, MD). As outlined previously, genotype data were excluded if they did not satisfy a Hardy–Weinberg equilibrium test p value $\geq 10^{-6}$, SNP call rate $>95\%$, Illumina BeadStudio GenCall score $\geq 0.7$, or a minor allele frequency $\geq 1\%$. Ancestral outliers were corrected with PC analysis using the “smartpca” program from v3.0 of EIGENSOFT. The Australian twin data were compared with all populations in HapMap phase 3 and a collection of five other GenomEU-TWIN populations.
the outliers were identified and filtered, only PC1 (the difference between the African population and others) and PC2 (the difference between the East Asian population and others) with the highest eigenvalues were used. We calculated the mean and standard deviation of the ancestral relation of the collective European population for reference PC1 and PC2 scores. Any individual who fell away from the mean by >6 times the standard deviation on PC1 and PC2 were removed. Considering the sensitivity of imputation toward missingness and SNP distribution, we conducted imputation using 469,117 common SNPs from the genotyping data present in HapMap CEU I+II data (release 22, build 36). This imputation was performed using the MACH v1.0.16b and Minimac packages, which generated association statistics for 2,543,887. These SNPs further underwent quality control with the following criteria: Hardy–Weinberg equilibrium test p value ≥10^{-6}, a minor allele frequency ≥1%, and Rsq score >0.3. A total of 2,428,106 SNPs passed the filtering step and were used for further analysis. The association of these SNPs with corneal astigmatism was tested using the –fastAssoc option in MERLIN. The association model was adjusted for age and sex.

In Table 9-1, the quality control details of the genotyping in both studies are outlined. The PC analysis of both population structures is shown in Appendix 3.1.

**Genetic Power Calculation**

GWAS power (type 1 error cutoff of = 5x10^{-8}) and replication power for two SNP test (cutoff = 0.05/2) was calculated using the Genetic Power Calculator program.

**Joint cohort analysis**

Meta-analysis of the data from two cohorts was conducted using the β-coefficients method.
of the METAL program. Only the common SNPs imputed in both cohorts (n about 2.5 million) were included in the meta-analysis. Regional associations were generated using SNAP.

Pathway analysis was undertaken using Pathway-VEGAS, an extension of the recently developed gene-based analysis tool VEGAS program. We selected pathways from the Gene Ontology (GO) database if the pathway size ranged in 10 to 1,000 genes, which resulted in 4,628 for further analysis.

To perform pathway analysis with Pathway-VEGAS, we first conducted a gene-based test on the summary data generated from the meta-analysis. To include most regulatory effects, each gene region was defined as being 50 kb up- and downstream of a gene. VEGAS calculated the gene-based test statistics by incorporating the effects of all SNPs in the gene region by correcting the linkage disequilibrium between the SNPs through a simulation approach for the multivariate normal distribution. Since the participants in our sample are European descendants, we used the linkage disequilibrium pattern from the HapMap2 CEU reference sample. Pathway p values were calculated by summing the $X^2$ test statistics of the respective gene derived from the VEGAS p values. These summarized p values were compared with 500,000 simulations where the summarized $X^2$ test statistics of randomly drawn genes depending on the pathway size to calculate the empirical p values of the pathway. To avoid adverse effects due to clustered genes, we considered only one gene from each cluster of genes, chosen randomly, and dropped others if the distance between them was <500 kb.
Table 9-1 | Quality control details of genotyping in both cohorts.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Raine</th>
<th>BATS/TEST</th>
<th>BATS/TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyping Centre</td>
<td>Centre for Applied Genomics (Toronto, Ontario, Canada)</td>
<td>deCODE (Iceland)</td>
<td>CIDR (USA)</td>
</tr>
<tr>
<td>Chip</td>
<td>Illumina 660K</td>
<td>Illumina 610K</td>
<td>Illumina 610K</td>
</tr>
<tr>
<td># genotyped SNPs (as supplied)</td>
<td>657366</td>
<td>592392</td>
<td>589296</td>
</tr>
<tr>
<td>mean GenCall &lt; 0.7</td>
<td>95876</td>
<td>47418</td>
<td>36877</td>
</tr>
<tr>
<td>&gt;5% missing</td>
<td>1843 (97719)</td>
<td>8447 (47950)</td>
<td>12455 (37499)</td>
</tr>
<tr>
<td>p(HWE) &lt; 10^-6</td>
<td>919 (98449)</td>
<td>2841 (49616)</td>
<td>15474 (51646)</td>
</tr>
<tr>
<td>MAF&lt;0.01(or monomorphic)</td>
<td>23370 (121734)</td>
<td>33347 (69632)</td>
<td>28607 (67969)</td>
</tr>
<tr>
<td># SNPs left (%)</td>
<td>535632</td>
<td>529379</td>
<td>531042</td>
</tr>
<tr>
<td>% genotyped SNPs</td>
<td>81.48%</td>
<td>89.36%</td>
<td>90.11%</td>
</tr>
<tr>
<td>Dropout rate due to QIMR SNP QC</td>
<td>18.52%</td>
<td>10.64%</td>
<td>9.89%</td>
</tr>
</tbody>
</table>
Results

After quality control, a total of 1,013 (51.3% male) unrelated individuals from the REHS and 1,788 (56.7% female) individuals of 857 twin families who were recruited through the TEST and the BATS were included in the genome-wide analyses of corneal astigmatism and corneal curvature. Demographic and phenotypic characteristics of these cohorts are shown in Table 9-2.

Corneal Astigmatism

No loci in the TEST/BATS or Raine populations attained genome-wide significance \( (p < 5 \times 10^{-8}) \) for corneal astigmatism. Additionally, following meta-analysis on >2.5M overlapping genotyped and imputed SNPs, no locus reached the level of genome-wide significance (Figure 9-2 and Figure 9-3). Eleven loci had a nominal threshold of suggestive significance \( (p < 1 \times 10^{-5}) \). Table 9-3 shows details regarding the ten most significant SNPs following the meta-analysis.

To identify genes associated with any known pathways, we tested the VEGAS results using pathways defined in the GO database. In our analysis, the top-ranking pathways were segmentation (GO:0035282) and embryonic pattern specification (GO:0009880; Table 9-4). Genes involved in differentiation of mesoderm (mesogenin 1 \([MSGN1]\), mesenchyme homeobox 1 \([MEOXI]\), mesenchyme homeobox 2 \([MEOX2]\), teratocarcinoma-derived growth factor 1 \([TDGF1]\)) and anterior and posterior axis formation (homeobox D8 \([HOXD8]\), homeobox A2 \([HOXA2]\), homeobox B6 \([HOXB6]\)) were common in both pathways.
<table>
<thead>
<tr>
<th></th>
<th>REHS</th>
<th>TEST/BATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>1013</td>
<td>1788</td>
</tr>
<tr>
<td>Number of families</td>
<td>1013</td>
<td>857</td>
</tr>
<tr>
<td>Mean age in years (Range)</td>
<td>20.0 (18 to 22)</td>
<td>22.2 (5 to 90)</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>493 (48.7)</td>
<td>1014 (56.7)</td>
</tr>
<tr>
<td>Mean corneal astigmatism in D</td>
<td>0.77 (0.46; 0.08 to 5.16)</td>
<td>0.76 (0.57; 0-9)</td>
</tr>
<tr>
<td>(SD; range)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean corneal curvature in mm</td>
<td>7.63 (0.24; 6.77 to 8.41)</td>
<td>7.72 (0.24;7.04 to 8.61)</td>
</tr>
<tr>
<td>(SD; range)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D=diopters, mm = millimeters
Figure 9-2 | Quantile-quantile (Q-Q) plot for age and sex-adjusted genome-wide association of corneal astigmatism.
Figure 9-3 | Manhattan plot of meta-analysis results for corneal astigmatism. The association of single nucleotide polymorphism and corneal astigmatism (age and sex adjusted) are plotted for each chromosome.
<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>bp</th>
<th>Allele</th>
<th>Effect</th>
<th>SE</th>
<th>p-value</th>
<th>Effect</th>
<th>SE</th>
<th>p-value</th>
<th>Beta</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1151008</td>
<td>12</td>
<td>31988627</td>
<td>G</td>
<td>-0.138</td>
<td>0.046</td>
<td>3.07x10^-3</td>
<td>-0.150</td>
<td>0.038</td>
<td>7.50x10^-5</td>
<td>-0.144</td>
<td>0.030</td>
<td>1.37x10^-6</td>
</tr>
<tr>
<td>rs1164064</td>
<td>3</td>
<td>110945090</td>
<td>A</td>
<td>0.212</td>
<td>0.043</td>
<td>9.91x10^-7</td>
<td>0.065</td>
<td>0.037</td>
<td>7.50x10^-2</td>
<td>0.134</td>
<td>0.028</td>
<td>1.86x10^-6</td>
</tr>
<tr>
<td>rs11841001</td>
<td>13</td>
<td>75147104</td>
<td>A</td>
<td>0.185</td>
<td>0.072</td>
<td>1.10x10^-2</td>
<td>0.195</td>
<td>0.057</td>
<td>6.30x10^-4</td>
<td>0.208</td>
<td>0.045</td>
<td>4.31x10^-6</td>
</tr>
<tr>
<td>rs7651778</td>
<td>3</td>
<td>157819249</td>
<td>C</td>
<td>0.096</td>
<td>0.043</td>
<td>2.64x10^-2</td>
<td>0.139</td>
<td>0.036</td>
<td>9.30x10^-5</td>
<td>0.126</td>
<td>0.028</td>
<td>4.76x10^-6</td>
</tr>
<tr>
<td>rs11859036</td>
<td>16</td>
<td>78863052</td>
<td>A</td>
<td>0.145</td>
<td>0.043</td>
<td>8.74x10^-4</td>
<td>0.107</td>
<td>0.037</td>
<td>3.60x10^-3</td>
<td>0.128</td>
<td>0.029</td>
<td>7.03x10^-6</td>
</tr>
<tr>
<td>rs438465</td>
<td>6</td>
<td>169562306</td>
<td>C</td>
<td>-0.211</td>
<td>0.058</td>
<td>2.70x10^-4</td>
<td>-0.143</td>
<td>0.051</td>
<td>5.50x10^-3</td>
<td>-0.173</td>
<td>0.039</td>
<td>7.22x10^-6</td>
</tr>
<tr>
<td>rs979976</td>
<td>2</td>
<td>137485172</td>
<td>A</td>
<td>0.163</td>
<td>0.047</td>
<td>4.81x10^-4</td>
<td>0.102</td>
<td>0.038</td>
<td>7.20x10^-3</td>
<td>0.134</td>
<td>0.030</td>
<td>7.52x10^-6</td>
</tr>
<tr>
<td>rs4805442</td>
<td>19</td>
<td>34780233</td>
<td>A</td>
<td>-0.192</td>
<td>0.060</td>
<td>1.48x10^-3</td>
<td>-0.141</td>
<td>0.050</td>
<td>4.40x10^-3</td>
<td>-0.168</td>
<td>0.038</td>
<td>1.18x10^-5</td>
</tr>
<tr>
<td>rs10079889</td>
<td>5</td>
<td>33011247</td>
<td>A</td>
<td>0.165</td>
<td>0.052</td>
<td>1.50x10^-3</td>
<td>0.124</td>
<td>0.043</td>
<td>4.30x10^-3</td>
<td>0.145</td>
<td>0.033</td>
<td>1.38x10^-5</td>
</tr>
<tr>
<td>rs2116538</td>
<td>2</td>
<td>137485890</td>
<td>A</td>
<td>-0.162</td>
<td>0.052</td>
<td>1.92x10^-3</td>
<td>-0.135</td>
<td>0.044</td>
<td>2.10x10^-3</td>
<td>-0.146</td>
<td>0.034</td>
<td>1.63x10^-5</td>
</tr>
</tbody>
</table>
* p-values displayed first as uncorrected for number of pathways tested, with value after Bonferroni correction in parenthesis.

Table 9-4 | VEGAS Pathway Analysis results from gene-based meta-analysis. Biological pathways implicated with development of corneal astigmatism.

<table>
<thead>
<tr>
<th>GO ID</th>
<th>GO Term</th>
<th>p-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0035282</td>
<td>Segmentation</td>
<td>2.0x10^-6 (0.009)</td>
<td><strong>PSEN2, WNT3A, MSGN1, TCF7L1, ZEB2, HOXD8, TDGF1, RBPJ, LEF1, SFRP2, TBX18, DLL1, MEOX2, HOXA2, NKL3-1, SFRP1, PRKDC, MLLT3, ROR2, BMI1, EGR2, ATM, FRS2, MYF5, TBX3, PCDH8, PSEN1, MESP2, RPRGRIP1L, ACD, DVL2, HES7, TCAP, KAT2A, MEOX1, HOXB6, AXIN2, MIB1, DLL3, TCF15, PAX1, POBFUT1, MAFB, EP300</strong></td>
</tr>
<tr>
<td>GO:0009880</td>
<td>Embryonic pattern specification</td>
<td>4.8x10^-5 (0.222)</td>
<td><strong>DISP1, MSGN1, TCF7L1, HOXD8, SATB2, CTNNB1, TDGF1, CXXC4, SMAD1, FGF10, SMAD5, DLL1, MEOX2, HOXA2, MLLT3, BMI1, NODAL, FRAT1, ZBTB16, FRS2, TBX3, SMAD6, RPRGRIP1L, DVL2, LHX1, MEOX1, HOXB6, SMAD2, MAFB, BMP7, SIM2</strong></td>
</tr>
<tr>
<td>GO:0007379</td>
<td>Segment specification</td>
<td>1.4x10^-4 (0.648)</td>
<td><strong>MSGN1, DLL1, MEOX2, HOXA2, MLLT3, BMI1, RPRGRIP1L, DVL2, MEOX1, MAFB</strong></td>
</tr>
</tbody>
</table>
We found no evidence for replication of the *PDGFRA* locus (Figure 9-4). In our cohorts, the previously reported top SNP in this region (rs7677751) was not significantly associated with corneal astigmatism (beta = -0.0423, standard deviation error=0.0423; p=0.32). The minor allele frequency of rs7677751 was 0.133 and 0.123 in the Raine study and the TEST/BATS, respectively. The top SNP at this locus was rs6821576 (p=0.003).
Figure 9-4 | Locus-specific plots of the most significant single nucleotide polymorphisms (SNPs). These plots display the most significant result in the meta-analysis GWAS (the locus identified by Fan et al. Chr 4q12 (A) and two loci identified in this study Chr 12p11 (B) and 3q13 (C)). SNPs are plotted as the $-\log_{10}$ of the p-value.
Corneal Curvature

With consideration of combined sample size, the GWAS power to detect the variant that could explain at least 1% variation of CC is very low (power 1/4 30%), whereas replication power for significant replication cutoff 0.05/2 is >99%. Thus, our primary emphasis was on replication of FRAP1 and PDGFRA genes, reported genome-wide association with CC in Asians, in an Australian population of Northern European ancestry. We tested whether the most significant SNP in each gene was associated in our European ancestry samples. We found that SNP rs2114039 in PDGFRA was associated with corneal curvature (p = 0.0045). Figure 9-5 shows the recombination profile between the SNPs and PDGFRA locus. The effect size of the trait increasing allele was 0.02275 mm per copy of the T allele. The effect sizes across different studies are shown in Figure 9-6A(2A). However, SNP rs6540964 in gene FRAP1 was not associated with curvature (p = 0.298); effect sizes are shown in Figure 9-6B.

Although our study does not have enough power to detect a genome wide-associated variant, we report the initial findings. The independent association test on BATS/TEST and REHS data yielded a best SNP rs4552334 (p = 2.5 x 10^{-6}) and rs11930632 (p = 2.47 x 10^{-6}), respectively. We followed this analysis by meta-analysis of outcomes from the 1,704,858 SNPs common in both studies.

The 25 most significant SNPs from the meta-analysis are shown in Appendix 3.2. As expected, meta-analysis did not reveal any genome wide-associated SNPs: the most associated genotyped SNP rs2444240 had a p value of 3.658 x 10^{-7} at 120.040 Mb (build 37) on chromosome 11. The nearest gene to four top SNPs on chromosome 11 is TRIM29 on the chromosome 11q23.3 region (NCBI build 37), which spans only 26 kb.
The recombination profile between the SNPs and \textit{TRIM29} locus is shown in Figure 9-7. These SNPs had a slightly stronger signal in the REHS, as shown in Table 9-6.
Figure 9-5 | Association of corneal curvature variants at the PDGFRA locus. The top SNP rs2114039 has a p value $4.549 \times 10^{-3}$. The red shading shows the degree of linkage disequilibrium between rs2444240 and neighboring SNPs. The light blue line displays the rate of recombination with scale on right-hand axis.

**Table 1.** Demographic Details of Study Participants

<table>
<thead>
<tr>
<th>TEST and BATS Raine Study</th>
<th>Number of subjects</th>
<th>Number of families</th>
<th>Mean age, y</th>
<th>Range of age</th>
<th>Sex (% female)</th>
<th>Mean corneal curvature, mm (SD)</th>
<th>Range of corneal curvature, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1788</td>
<td>857</td>
<td>22.2</td>
<td>5 to 90</td>
<td>1014 (56.7)</td>
<td>7.63 (0.24)</td>
<td>6.77 to 8.41</td>
</tr>
<tr>
<td></td>
<td>1013</td>
<td>1013</td>
<td>20.0</td>
<td>18 to 22</td>
<td>497 (49.1)</td>
<td>7.72 (0.24)</td>
<td>7.04 to 8.61</td>
</tr>
</tbody>
</table>
Figure 9-6 | Forrest plot, showing effect size distribution in different studies for (A) PDGFRA SNP rs2114039 and (B) FRAP1 SNP rs6540964. The top four studies are reported by corneal curvature studies in an Asian population. The last two studies demonstrate the effect size distribution for SNPs found in our analysis of corneal curvature GWAS on Australians with Northern European Ancestry.
**Figure 9-7 | Association of variants at the TRIM29 locus.** The top SNP rs2444240 has a p value $1.28 \times 10^{-3}$. The red shading shows the degree of linkage disequilibrium between rs2444240 and neighboring SNPs. This SNP is 31kb upstream to the TRIM29 gene. The light blue line displays the rate of recombination with scale on right hand axis.
Table 9-6 | Top four SNPs and their association results

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr</th>
<th>Coordinate (build 36)</th>
<th>Nearest Gene</th>
<th>Alleles*</th>
<th>BATS/TEST Effect</th>
<th>SE</th>
<th>BATS/TEST p</th>
<th>RAINE Effect</th>
<th>SE</th>
<th>RAINE p</th>
<th>Weighted Effect</th>
<th>SE</th>
<th>Meta analysis p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2444240</td>
<td>11</td>
<td>119545652</td>
<td>TRIM29</td>
<td>T/G</td>
<td>0.030631</td>
<td>0.009544</td>
<td>1.28×10⁻³</td>
<td>0.044733</td>
<td>0.011125</td>
<td>5.71×10⁻⁵</td>
<td>-0.0364</td>
<td>0.0071</td>
<td>3.66×10⁻⁷</td>
</tr>
<tr>
<td>rs494965</td>
<td>11</td>
<td>119558860</td>
<td>TRIM29</td>
<td>T/C</td>
<td>-0.0309</td>
<td>0.009392</td>
<td>1.12×10⁻³</td>
<td>-0.04317</td>
<td>0.010944</td>
<td>9.35×10⁻⁵</td>
<td>0.0365</td>
<td>0.0073</td>
<td>4.50×10⁻⁷</td>
</tr>
<tr>
<td>rs470606</td>
<td>11</td>
<td>119533412</td>
<td>TRIM29</td>
<td>T/G</td>
<td>0.029068</td>
<td>0.009541</td>
<td>2.2×10⁻³</td>
<td>0.045626</td>
<td>0.011163</td>
<td>4.41×10⁻⁵</td>
<td>-0.0359</td>
<td>0.0071</td>
<td>5.66×10⁻⁷</td>
</tr>
<tr>
<td>rs470373</td>
<td>11</td>
<td>119531481</td>
<td>TRIM29</td>
<td>T/C</td>
<td>0.029068</td>
<td>0.009541</td>
<td>2.2×10⁻³</td>
<td>-0.04492</td>
<td>0.010922</td>
<td>4.61×10⁻⁵</td>
<td>0.0362</td>
<td>0.0073</td>
<td>5.90×10⁻⁷</td>
</tr>
</tbody>
</table>

*The first letter denotes the effect allele for specified SNP.
Discussion

Understanding the molecular mechanism of cornea-related disease is useful for developing novel corrective or therapeutic strategies. Recently, through meta-analysis of five Singaporean cohorts, Fan and colleagues reported a statistically significant association between a variant (rs7677751) at the \( PDGFRA \) locus on chromosome 4q12 and corneal astigmatism. In our present study, we found no strong evidence for transfer of risk for corneal astigmatism of this locus in two Australian cohorts of Northern European ancestry. Our results suggest there are underlying genetic differences between populations, which may account for differing development and prevalence of corneal astigmatism.

Although no SNP in our study was significantly associated with corneal astigmatism at the genome-wide level, we identified several putative loci, which reached a suggestive level of significance \( (p<1\times10^{-5}) \). Our strongest signal on meta-analysis (rs1151008) was located on chromosome 12p11. This SNP is approximately 100 kb centromeric to the antagonist of the mitotic exit network 1 homolog (\( AMN1 \)) gene. \( AMN1 \) has been shown to be important in resetting the cell cycle, and the \( AMN1 \) domain of the lysine-specific demethylase 2A gene appears to inhibit keratinocyte growth in vitro.

Our second strongest-association (rs1164064) was on chromosome 3q13, near the developmental pluripotency associated 4 (\( DPPA4 \)) and \( DPPA2 \) genes. These genes have important roles in stem cell generation and are rapidly down regulated during cellular or fetal differentiation. Given that the cornea is of ectodermal origin, \( DPPA4 \) regulates differentiation of embryonic stem cells into a primitive ectoderm lineage.
Following gene-based pathway analysis, we found that genes involved with segmentation and embryonic pattern specification were associated with the development of corneal astigmatism. In vertebrates, the periocular mesenchymal cells migrate into the cornea giving rise to cornea stroma during embryogenesis. Interestingly, the mesogenin 1 (MSGN1), mesenchyme homeobox 1 (MEOX1), mesenchyme homeobox 2 (MEOX2), and teratocarcinoma-derived growth factor 1 (TDGF1) genes identified in our pathway analysis are involved in differentiation of mesoderm. Additionally, some of the genes in these pathways are part of the homeobox (HOX) family, which included the HOX8.1 gene that was demonstrated to be expressed during murine ocular development.

It is somewhat surprising that, despite our reasonable power, we were unable to replicate the association of corneal astigmatism and the rs7677751 variant. We also failed to identify any locus associated with this trait at the genome-wide significance level. Our results suggest that in dissecting the genetic architecture of corneal astigmatism in people of Northern European ancestry, no major single locus will predominate, similar to other complex quantitative traits. Clearly, larger, better-powered cohorts are required to intimately dissect the genetic etiology of this biometric trait.

In the second part of this analysis, we have conducted a genome-wide association study on corneal curvature of the same study population. Our sample showed that SNPs near PDGFRA, which were very recently shown to play a role in CC in Asians, also play a role in samples of Northern European Ancestry. The allele frequency at rs2114039 is similar (~0.3) across the range of European and Asian ancestry samples in the
HapMap3 data. The variance explained by rs2114039 in our European ancestry sample for BATS/TEST and REHS was 0.6% and 0.1%, respectively, which is somewhat lower than that seen in the Asian studies (1.8%, 11.1%, 4.9%, and 7.5% for SP2, SIMES, SINDI, and SCORM study, respectively). However, the estimates of variance explained in the Asian studies are probably an overestimate of the true effect size due to this SNP being selected as one of the most significant results from a genome-wide scan.\textsuperscript{394} An estimate of the effect size in an independent Asian population would allow us to determine if the effect size at this SNP truly differs in the two populations.

Although our study did not identify any genome-wide significant loci as we expected, meta-analyzing across two Australian studies led to a more significant top SNP, with rs2444240 in tripartite motif containing 29 (\textit{TRIM29}) achieving a p value of $3.658 \times 10^{-7}$. Larger sample sizes are required to unambiguously identify novel loci. With reference to past literature, we found involvement of some of the genes near or within the top 25 SNPs with some related traits.\textsuperscript{395-398} A gene expression profile of human keratoconus suggests significant expression of the \textit{TRIM29} gene.\textsuperscript{395} This differential expression of the \textit{TRIM29} gene could be responsible for the conical shape of the cornea. Apart from this, a linkage study on Ashkenazi Jewish families also supports a possible role of \textit{TRIM29} in variation of CC by their report of linkage between 11q23 loci and myopia.\textsuperscript{396} In addition, a gene expression study on focal loss of retinal ganglion cells suggests significant down regulation of the B-cell CLL/lymphoma 11B (\textit{BCL11B}) gene.\textsuperscript{398} A study of myopia in chickens has suggested that ganglion cell destruction may cause corneal flattening in this model.\textsuperscript{397} Further study needs to be done to investigate the possible role of \textit{TRIM29} and \textit{BCL11B} genes in determining CC.
In summary, while we found no strong evidence for replication or transferability of the previously reported association between the rs7677751 variant, at the PDGFRA locus, and corneal astigmatism in our Australian cohorts of Northern European ancestry, our study of CC showed that the SNP in PDGFRA, recently implicated in this trait in Asians, also underlies trait variation in Australians of Northern European ancestry. The other gene reported to be associated with CC in Asian populations, FRAP1, did not show a significant effect in our samples. Although our study was underpowered to detect novel loci, we found some evidence that SNPs near TRIM29 may play a role in determining CC. Our findings of SNPs at TRIM29 and other regions should be replicated in further studies, with meta-analyses likely to prove important in further dissecting this important endophenotype.
Chapter 10

Discussion

Introduction

The eye is one of the most complex organs in the human body. It allows us to change our focus from the screen in front of us to a distant object on the horizon in a third of a second. Indeed, the eye was once considered the prime example of irreducible complexity, that is the inability to function without any of the components, and therefore suggested that it was not evolved naturally from a primitive form.\textsuperscript{399} Due to this complexity, the goal of identifying the constituents of a healthy eye is similar to in effort to putting together a jigsaw puzzle with an unknown number of pieces. Hence it requires a systematic approach similar to constructing a geographic information system, which provides the capability to combine various data into a composite data layer that becomes a base layer in a database. When such a model is applied to the human eye, gene and environment are the base layers of the “human eye” database. Having this model in mind, this thesis was undertaken to identify some of the genetic and environmental elements that comprise a healthy human eye. The utility of cloud computing was discussed as a means of analysing large genomic datasets. A good tool will have no power without accurate data. This thesis also defines various stages of a well-designed cross-sectional cohort study, and describes the collection of ocular phenotypes and genotypes in detail to help other researchers.
Summary of key findings*

Tools for big data analysis

Identification of genetic and environmental patterns in human complex diseases requires an extraordinary effort. As discussed in chapter 2, big data analysis is a key tool for success in this work. However, it has its own many challenges in various areas and the computational framework is one of them. Cloud computing offers a dynamic means for researchers to access large computing resources without concern about the physical structure or its maintenance. However, choosing an appropriate service can be difficult, especially for researchers with limited computational knowledge. Although there is a vast amount of benchmarking information produced by companies and individuals devoted to help the scientific community, there is a lack of basic performance comparison of undedicated cloud computing services for genomic research (Box 10-1). In chapter 3, we released our cost-effective comparison of two commercially available cloud platforms (EC2 and GCE), which can be utilized as a basis when these services are being considered. Additionally, we presented available ready-to-use scripts for establishing Hadoop instances with Ganglia monitoring. These scripts can be found online and applied to other platforms with some modifications.

* An infograph is attached to the end of this section to give an overall picture how the findings from this thesis fit into the greater myopia research.
Box 10-1 | Why Consider Undedicated Cloud Computing Services for Analysis of Genomic Datasets?

**Low cost:** Users can access on demand therefore no need to finance hardware and maintain it on the site.

**Elasticity:** Computing resources can be scaled up and down easily depending on the task.

**Flexibility:** Users can tailor the computing resources precisely according to needs when and wherever required without working around the permissions that have been given on a local environment.

**Support:** Satisfaction of users is a priority of cloud computing providers. Users can easily get timely help and support; however, inter-departmental issues within in a local centre can take time to resolve.
Phenotyping in complex diseases

Unlike Mendelian disorders, complex diseases are heterogeneous syndromes with multiple phenotypic subtypes. Each of these subtypes may have distinct genetic or environmental factors that cause its manifestation and progression. Therefore, a thorough examination and description of this heterogeneity is the first step in research. The REHS design and methodology was explained in Chapter 4 in order to allow comparison with other cohort and cross-sectional studies. Moreover, a description of the prevalence of ocular disease in young adults was previously shown to be absent in the literature. Closing this gap, the study also establishes the prevalence of eye disorders in a large sample of predominantly Caucasian young Australian adults. This baseline information revealed myopia as the most common disorder in this age group.

Although much broader terms were used in the introductory chapters of this thesis aiming to look at factors determining various eye diseases, because of the aforementioned myopia prevalence in otherwise healthy eyes of this young age group, a number of chapters dedicated to myopia characteristics in association with other factors. Hence, I would like to acknowledge that this thesis has a tendency towards explaining environmental and factors in relation to myopia.

Ocular biometry in myopia

Although myopia was known to Aristotle, it was not until the 17th century that first Kepler, and then Newton, concluded that light was brought to a focus in front of the retina of the myopic eye. 18th century researchers decided that the myopic eye was longer than the normal eye. Today it is well accepted that myopia occurs as a result of an imbalance of the ocular biometry during the emmetropisation process. While axial length weighs as the major endophenotype on one arm of the balance, other ocular
biometry provides the optical power necessary to counteract on the opposite arm of this balance. The driving force of this balance is the retinal image quality. Higher order aberrations (HOAs) are small imperfections of the optical media, which distort the retinal image quality and consequently are involved in the emmetropisation feedback mechanism. HOAs vary with age and ethnicity hence it is important to have a well-defined population to describe the normal distribution. In Chapter 5, we investigated the variation of HOAs in age- and ethnicity-controlled REHS cohort. In this population, the HOAs were marginally higher than previously reported values. Moreover, there was an overall difference in monochromatic aberrations between differing vision and refractive groups. Yet, supernormal vision was attainable in the presence of HOAs. This suggests that the effects of HOAs can be disregarded in a system where other biometry is in equilibrium. Previous studies had no agreement regarding the correlation between monochromatic aberrations and refractive error. In the REHS, HOAs were associated with an increased severity of myopia. These findings support the view that rather than being a precursor for myopia development, HOAs are components of the imbalanced ocular biometry due to axial length elongation.402

**Genes and Environment in Myopia Development**

The dichotomy of genes versus environment argument has dictated research on the aetiology of myopia for decades. The heritability of myopia was calculated to be 90% (95%CI: 81%-95%) in twin studies and an additive genetic effect was shown to be responsible for up to 86% of the variance in refractive error as a continuous spectrum between myopia and hyperopia.356 Parental myopia is consistently associated with myopia in offspring. For example, the prevalence of myopia was found to be 30-40% in children with two myopic parents, 20-25% in children with one myopic parent, and < 10% in children whose parents did not have myopia.403,404 Family studies have allowed
us to identify a number of genetic loci for non-syndromic high myopia and a number of genes particularly associated with connective tissue growth and extracellular matrix reorganisation were reported in the candidate gene association studies. Furthermore, genome-wide association studies (GWAS) have identified several genetic loci for refractive error and high myopia.

In addition to these genetic contributions to risk, epidemiological studies have identified many potential environmental risk factors for the development of myopia. Among these, higher level of education attained appears to be consistently correlated with higher prevalence of myopia whereas although “near work” has been well-researched, the evidence is inconsistent. It is now more accepted that myopia is a complex disease caused by a combination of genetic, environmental and lifestyle factors, most of which remain unidentified.

Environmental exposures during fetal or early life can have adverse effects on fetal growth and development of the infant. In chapter 6, we investigated the possible effects of anaesthesia on the visual acuity and refractive status of young adults who underwent general anaesthesia early in life. Our findings suggest that paediatric anaesthesia is not associated with reduced vision or increased myopia in young adulthood. This supports the view that anaesthesia is a safe procedure and has no risk to visual development.

Probably, it will not be incorrect to say that identification of environmental factors and life-style choices in complex diseases is much harder than analysing a known set of genes. The dynamic nature of environmental data makes it difficult to collect accurately. With the goal of overcoming this, our group proposed CUVAF as a
biomarker of ocular sun exposure in previous publications. In chapter 7, we confirmed that although genes play a role in its development, a large environmental component (63%) to CUVAF exists thus making it a suitable candidate for a sun-exposure biomarker. Furthermore, we identified a SNP (rs1060043) near the SLC1A5 gene as being significantly associated with CUVAF; replication of this finding in future studies is warranted.

Both longitudinal and cross-sectional studies have found an association between time spent outdoors and reduced risk of developing myopia. Several mechanisms including reduced near work, physical activity and brighter light are proposed to explain this relationship but there is no definitive conclusion. Due to vitamin D production from exposure to UV light, vitamin D is postulated to be the link between myopia and time spent outdoors. We investigated this hypothesis in chapter 8 and showed lower 25(OH)D₃ concentrations in myopic participants. The prevalence of myopia was also significantly higher in individuals with vitamin D deficiency compared to the individuals with sufficient levels. The scope of our analysis did not include an answer to the question of whether a higher serum 25(OH)D₃ concentration is protective against myopia, or whether it is acting as a proxy for some other biologically effective consequence of sun exposure. On this note, a recent study in children has suggested that vitamin D is only a biomarker of time spent outdoors. Further evidence both from animal models and clinical studies are necessary to define the role of vitamin D in myopia development.
Genetics of a coexisting ocular disease

Astigmatism is a more prevalent refractive error in infants and young children compared with adults. Often, infantile astigmatism is significantly reduced or resolved within the first two years of life. However, the functional significance or effect of this change in emmetropisation has not been truly understood.\textsuperscript{411} Multiple longitudinal studies have also shown the association between the school-onset myopia and infantile astigmatism.\textsuperscript{411-413} There are two likely mechanisms underlying this relationship: either infantile astigmatism disrupts the focusing ability of the developing eye and causes reduced image quality, or axial elongation during development triggers an imbalance of the optical power and results in corneal and/or lenticular astigmatism along with myopia.\textsuperscript{411} Regardless of the mechanism, corneal astigmatism (CA) and corneal curvature (CC) are part of the ocular biometry equation and therefore must have overlapping genetic factors. Previously, the rs7677751 SNP at the \textit{PDGFRA} gene was found to be associated with CA and CC in people of Asian ancestry.\textsuperscript{363,386} In chapter 9, we sought to replicate these findings and identify other genetic markers of CA and CC in an Australian population of Northern European ancestry. Our data suggested that the \textit{PDGFRA} locus does not transfer a major risk of corneal astigmatism in people of Northern European ancestry. On the other hand, a significant role of the \textit{PDGFRA} locus in determining corneal curvature in the Australian population was confirmed in our study, also highlighting the putative association of the \textit{TRIM29} locus with CC.
This infographic displays how the findings from this thesis contribute to research in myopia development. Phrases in red indicate the investigated areas of research.
Limitations of the thesis

The key findings arising from this thesis are discussed above. However, certain limitations need to be acknowledged.

This thesis has defined the prevalence of ocular disease in a young adult population utilising an existing cohort study (Chapter 4). Although we were able to identify the common ocular problems, we were unable to define the rare ocular diseases in this age group. Moreover, this was the first time the cohort underwent a comprehensive ocular examination, and the possible exposure measures were not assessed at a regular basis throughout the lifetime of the participants. Therefore, we were unable to describe some of the exposures in detail (Chapter 6) and infer causality in the associations we identified (Chapter 8).

One other caveat of cohort studies is potential losses to follow-up. The original Raine cohort consisted a relatively large sample of individuals; however, over the years, for the reasons reviewed in Chapter 4, only half of the cohort (1344 individuals) was recruited in the 20-year follow-up. In addition, data on potential confounding factors was frequently missing. This was particularly a challenge in performing GWA studies (Chapter 7 and 9), as this approach requires large sample sizes for solid conclusions. Combining the REHS data with the twin studies data in meta-analyses provided sufficient power to detect some of the genetic variance in CUVAF, corneal curvature and astigmatism.

Genetic and environmental diversity are often measured in a defined population. Thus the findings of this thesis may not be valid for other populations and may entail further
research.

Finally, I would like to acknowledge that the REHS has established a wealth of ocular biometry data in young adults. Some of these data and their analysis related to myopia were included in the publications lead my colleagues ("publications arising during this thesis"). Others had to be omitted due to limited time frame and funding of this PhD project.


**Future directions**

Some of the limitations discussed above may not be eliminated owing to the nature of the cohort studies. However, some can be addressed in future studies for better understanding of what constitutes a healthy eye. For example, replication studies can be performed in similar populations to validate this thesis’ findings and in different populations to identify possible deviations. As experienced in this work, identification of genetic and environmental factors with small effects sizes necessitates larger samples sizes. In these days, many groups work in partnership to surmount this issue and hence collation of bigger consortia may eventually eradicate this challenge.

In the proposed GIS model of the human eye database, it is not sufficient to define various layers. The next step is to understand the interactions between the layers. Further research should therefore concentrate on investigating gene-gene and gene-environment interactions in myopia development. So far only two studies have reported the increased risk of myopia in individuals at higher genetic susceptibility in combination with higher level of education.\textsuperscript{414,415} For example the link between time outdoors and a genetic predisposition to myopia can be examined next.

Another possible area of research would be to investigate how alterations in the disease-associated genes cause changes in molecular pathways and functional outcomes. Expression studies of the associated genes and their products in multiple animal models can be a starting point for such research.
Concluding remarks

In this thesis, data from a large young, healthy adult population extends our knowledge of variation in, and determinants of, ocular biometry and disease. These outcomes contribute to the GIS-like network map of genetic and environmental factors for the disease-free adult eye. Once this network map is complete, it will serve as a baseline tool for further research in preventing and intervening in ocular diseases.
Bibliography


183


84. Newman AM, Gallo NB, Hancox LS, Miller NJ, Radeke CM, Maloney MA, et


150. 3-D Data Management: Controlling Data Volume, Velocity and Variety [Internet]. Laney, Doug; [cited 2015 Jan 22]. Available from:


179. van Rossum EFC, Federenko IS, Koper JW, Kumsta R, Hellhammer DH.


245. DiMaggio C, Sun LS, Kakavouli A, Byrne MW, Li G. A Retrospective Cohort Study of the Association of Anesthesia and Hernia Repair Surgery With


302. French AN, Morgan IG, Burlutsky G, Mitchell P, Rose KA. Prevalence and 5-to 6-Year Incidence and Progression of Myopia and Hyperopia in Australian


339. Schmid KL, Wildsoet CF. Effects on the compensatory responses to positive and negative lenses of intermittent lens wear and ciliary nerve section in chicks.


204


379. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D.


Appendices

1.1 Uploading data and setting up an Amazon Web Services Elastic MapReduce (EMR) cluster.

Crossbow on EMR supports importing data referenced from the manifest file from HTTP, FTP, HDFS, and S3. We uploaded the input data (available from: ftp://public.genomics.org.cn/BGI/yanhuang/Rawdata2) into a S3 bucket in order to have the maximum potential download speed and ensure that the cluster was not sitting idle while the inputs were downloaded from a remote FTP server. The process used to upload the files from the remote FTP to our S3 bucket was as follows:

1) Create a micro instance on EC2, ensuring the instance has enough storage space for the whole dataset.

2) SSH into the instance using the given hostname,

3) (Optional) Install the screen package, and create a screen session. This allows you to close the SSH connection and allow the processes to continue. This is done via:

   sudo apt-get install screen

   Note: a screen session can be reconnected by:

   screen -r

4) Download the data from the remote FTP server:

   wget -r ftp://public.genomics.org.cn/BGI/yanhuang/Rawdata2

5) Upload the data to the S3 bucket

   s3cmd put --recursive /path/to/data/ s3://<your-bucket>/path/to/save/data
The time taken to download the 142GB human genome data from the FTP server was 10:35:49 and the time taken to upload to the S3 bucket was 2:41:30, giving a total transfer time of 13:17:19.

The Ganglia install scripts provided by Amazon were using an out-of-date version of the software, which did not allow the metric data to be exported. As such these install scripts were updated to use the current version of Ganglia and Ganglia-web (3.6.0 and 3.5.10 respectively) and reinstalled. Ganglia was configured to use unicast mode as EMR prevented use of multicast.

Crossbow provided a faulty file in their S3 bucket that consistently caused the final step on EMR to fail. The file was hosted on our development bucket and Crossbow was modified so it was pointing to the S3 bucket. The link in the source code of Crossbow was altered and should now be corrected in the official Crossbow bucket.
1.2 Scripts for configuration and running jobs on Google Compute Engine (GCE)

Crossbow provides an option that sets up and configures an EMR cluster with the required software. Although no such support is available for GCE, Crossbow supports use of a Hadoop Cluster that can be implemented in GCE. Google have released scripts for creating a Hadoop cluster on the GCE services (available from: https://github.com/GoogleCloudPlatform/solutions-google-compute-engine-cluster-for-hadoop). However, these scripts needed to be modified as we required some additional software to be installed on each node of the cluster. These modifications can be found at https://github.com/hewittlab/Crossbow-GCE-Hadoop. The modifications perform the following additional steps when setting up the cluster:

1) Create Ganglia configuration files and Hadoop configuration files specifically for the hostnames and IP addresses assigned to each of the nodes in the cluster.
2) Upload the configuration files to the Google Storage Bucket.
3) Install the required software on each of the nodes in the cluster, and place the configuration files from step 1 into their appropriate locations.

To run Crossbow on GCE, we initially uploaded the input reference Jar and the input manifest file to the HDFS. Given that these files were to be used for multiple runs and to avoid re-upload these files multiple times this was performed via the Google Storage Bucket. Once completed, the cluster was started. The input files were then downloaded from the bucket, onto the master node and then into the HDFS. Crossbow for Hadoop was then run with the appropriate parameters for the current cluster and input files, with
the output files being copied from the HDFS to the bucket on completion. Each cluster was terminated upon completion of the workflow.

To reduce the repetition of steps required to set up the GCE cluster for each experimental run, an additional script, `run_crossbow.py` was created. While this script simplified the process to a level that was comparable to running Crossbow for EMR on the command line, it also raised several other issues. For example, access permission to the RRD directory on each node was made available to all users as required by Ganglia using `chown` command.

Hadoop is designed to manage multiple failures and prevent a job from completely failing within a cluster. The default configuration of Hadoop stops the cluster immediately in case of a single failure. To prevent random task failures stopping the GCE cluster, we increased the threshold number of failed tasks allowable within the cluster.

When running Hadoop on GCE, a storage disk space had to be allocated to the instance at the time of its creation. Moreover, when creating a disk based off a Debian image, the available disk space was limited to 10GB regardless of the specified disk size. Linux partitions were edited to utilize the full size of the disk. A snapshot of the storage disk was created in order to avoid repartitioning of the disk during node creation. These modifications were updated further in the release version to reduce the setup time.

By default, Google only allows access to a limited number of resources per project. Since use of large cluster recommended in Crossbow runs, we applied for access to
additional number of cores, IP addresses and total disk space in the region of our choice through an online form provided on Google Developers Console.

On January 9th 2014, Google depreciated the v1beta15 application programming interface (API) and disk used for temporary storage was replaced with a persistent disk approach. This change caused problems in our initial cluster setup. Therefore, setup scripts were altered in order to generate a disk for each node prior its creation and to delete these disks upon termination. To point our code to the new API, the cluster creation scripts were modified based on a revised version of the Google’s code. However, later on we rewrote our changes into the updated Google scripts which are now available in our repository.

Finally, it is important to configure the number of Crossbow tasks appropriately to ensure complete node usage. The following arguments should be used with Crossbow on GCE: --cpus (number of cores per instance); and --instances (number of instances). It is important to note that these arguments are applicable to Hadoop in a non-EMR environment.
1.3 Transformation of metric outputs from .RRD to .CSV format.

The program rrd2csv.pl was used to convert RRDs to CSVs. It is available from
https://code.google.com/p/rrd2csv/. Its usage is:

perl rrd2csv.pl -s “17/12/2013 12:00” -e “17/12/2013 13:00” file.rrd

To store the output as a csv, add “> file.csv” to the end (i.e perl rrd2csv.pl -s “17/12/2013 12:00” -e “17/12/2013 13:00” file.rrd > file.csv )

The -s and -e flags are representing the times that one adds the csv, where -s is the start time and -e is the end time. So in the example above, we are exporting from today at 12, to today at 1pm. These times should be when the job ran.
1.4 Glossary

This glossary was partly generated using the review on “Computational solutions to large-scale data management and analysis” by Schadt et al. (2010).

**Bucket**: fundamental storage unit provided to Amazon S3 users to store files. Buckets are containers for your files that are similar conceptually to a root folder on your personal hard drive, but in this case the file storage is hosted on Amazon S3.

**Central processing unit (CPU)**: A term often used interchangeably with the term ‘processor’, the CPU is the component in the computer system that executes the instructions in the program.

**Cloud-based computing**: The abstraction of the underlying hardware architectures (for example, servers, storage and networking) that enable convenient, on-demand network access to a shared pool of computing resources that can be readily provisioned and released.

**Cluster**: Linkage of multiple computers, typically through a fast local area network, to complete a task faster than a single computer could

**Computational node**: The unit of replication in a computer cluster. Typically it consists of a complete computer comprising one or more processors, dynamic random access memory (DRAM) and one or more hard disks.
**Computer Node:** A single computer, often not too different from a standard desktop computer, which is combined with multiple other nodes to create a cluster.

**Core:** Typically used in the context of multi-core processors, which integrate multiple cores into a single processor.

**Instance:** Each virtual machine in cloud based computing.

**Manifest file:** A file that contains a list of files needed for the computation.

**Multicast:** Communication between a single sender and multiple receivers on a network. In a multicast computer network, multiple nodes receive and send information to each other to complete the work.

**Persistent storage:** Any data storage device that retains data after power to that device is shut off.

**Undedicated cluster:** A computing cluster, which has no dedicated task. They are normally used in Cloud Computing environments, allowing customers to perform whichever tasks they require on the cluster.

**Unicast:** Communication between a single sender and a single receiver over a network. In a unicast computer network, one node sends and the other receives the information to process.
Virtual cluster: A computer cluster comprised of Virtual machines.

Virtual machine: A computer, whose hardware has been virtualised by another computer, allowing for multiple operating systems to run independently on a single computer.
2.1 REHS Ethics Approval

Winthrop Professor David Mackey  
Ophthalmology & Visual Science (Centre for)  
MBDP: M517

Dear Professor Mackey

HUMAN RESEARCH ETHICS COMMITTEE– ETHICS APPROVAL

Raine Study 20/21 year cohort review - eye health study

Student(s):

Ethics approval for the above project has been granted by the Human Research Ethics Committee from 28 January 2010 to 28 January 2011 in accordance with the requirements of the National Statement on Ethical Conduct in Human Research (National Statement) and the policies and procedures of The University of Western Australia.

You are reminded of the following requirements:

1. The application and all supporting documentation form the basis of the ethics approval and you must not depart from the research protocol that has been approved.
2. The Human Research Ethics Office must be approached for approval in advance for any requested amendments to the approved research protocol.
3. The Chief Investigator is required to report immediately to the Human Research Ethics Office any adverse or unexpected event or any other event that may impact on the ethics approval for the project.
4. The Chief Investigator must inform the Human Research Ethics Office as soon as practicable if a research project is discontinued before the expected date of completion, providing reasons.

Any conditions of ethics approval that have been imposed are listed below:

Special Conditions

None specified

The University of Western Australia is bound by the National Statement to monitor the progress of all approved projects until completion to ensure continued compliance with ethical standards and requirements.

Please note that the maximum period of ethics approval for this project is five (5) years from the date of this notification. However, ethics approval is conditional upon satisfactory progress reports being received by the designated renewal date for continuation of ethics approval. The Human Research Ethics Office will forward a request for a progress report communication for your completion approximately four weeks before this renewal date.

If your progress report is not then received by the renewal date, your ethics approval will expire, requiring that all activities relating to the project cease immediately.

If you have any queries please do not hesitate to contact the Human Research Ethics Office (HREO) on (08) 6488 3703.

Please ensure that you quote the file reference – RA/4/1/2646 – and the associated project title in all future correspondence.

Yours sincerely

Peter Johnstone  
Manager  
Human Research Ethics

Our Ref: RA/4/1/2646  
29 January 2010
2.2 The REHS 12-Station Protocol

All the examiners were instructed to print out a copy of the results where possible, complete the LEI examination recording form and insert their initials for each station.

Station 1

Pre-cycloplegic Autorefraction: Pre-cycloplegic autorefraction and keratometry were performed with Nidek ARK-510A (NIDEK Co.Ltd, Japan). Participants were instructed to relax their eyes and look at the internal image through the eyepiece and try not to bring it into focus. The measured prescription and keratometry results in diopters were recorded in the examination form.

Vertometry: If the participant had a pair of glasses at the time of examination, CL-200 Computerized Lensometer (Topcon Medical Systems Inc., Oakland, NJ) was used to measure the prescription of the glasses.

Colour vision: Participants with colour vision deficiency were identified with Ishihara’s Test for Color Deficiency (24 Plate Edition, Kanehara Trading Inc., Tokyo, Japan)

Station 2: Best corrected vision

Distance Visual Acuity: Visual acuity was measured using a logarithm of the minimum angle of resolution chart (Test Chart 2000 Xpert, Thomson Software Solutions, UK), which was run on Window XP computer. Except for the larger size, each row of the chart contained five letters and letter size reduced in steps of 0.1
logMAR between one row and the next. The monitor was placed at three meters and viewed via a mirror. Participants were tested monocularly and with glasses or contact lenses if worn. The letters were alternated between the eyes. All participants were pinholed, even in the event that one or both eyes achieved 6/6 vision or better. A matching HOVT card was available for disabled participants. If no letters were identified on the chart, counting fingers, hand movements and perception of light were assessed.

**Contrast Sensitivity:** The traditional method of measuring contrast sensitivity is sine wave grating. It was shown that the low-contrast letter charts could provide valuable information in clinical settings. We test contrast sensitivity by using a low-contrast letter chart (Test Chart Xpert, Thomson Software Solutions, UK). The viewing distance was set to six meters. Participants were tested monocularly and with glasses if worn. Participants were asked to read letters displayed in triplets of decreasing contrast from top to the bottom until they could no longer read two of the three letters. The contrast of the last row read was recorded in units of logarithm of the minimum angle of resolution.

**Vernier Acuity:** We determined the minimum angular separation required to detect that two lines placed end to end were not co-linear using the Test Chart Xpert 2000. The participants were tested monocularly with glasses, if worn. They were instructed to inform the examiner as soon as they see the straight line splitting into two on the screen. The test was repeated three times for each eye. Mean of the three measurements were recorded in arc seconds.
Station 3: Binocular Vision Function

Cover Test/Alternate Cover Test: To determine the presence of heterotropia and heterophoria, a cover test and an alternate cover test were performed at both distance (6m) and near (1/3m). The tests were repeated with and without glasses. Prism bar cover test was performed to measure any observed misalignment.

Ocular Motility: Eye movements were examined in nine-positions of gaze with a pen torch. The participants were instructed to hold their head stationary and follow the light with their eyes. Abnormal motility was scored on +4 (gross overaction) to -4 (gross underaction) with increments of 0.5. Bielshowsky head tilt test was performed when a palsy of the superior oblique muscle was suspected.

Four Diopter Base Out Prism Test: This test is used to detect a micro-strabismus and associated central suppression scotoma if present. Participants were instructed to fix on a small fixation target at near (1/3m). A four-diopter base out prism was placed in front of one eye and a fixation movement to confirm fusion of images was observed. In the event of no movement was observed to overcome prism and restore binocular single vision, the test outcome recorded as a negative that was presence of a central suppression. The test repeated on the other eye. If no movement was observed when prism was placed in front of either eye, the response was recorded as equivocal.

Nystagmus: Any nystagmus observed was recorded with its direction and subtype.

Stereoacuity: Lang II test was completed to assess gross stereopsis or depth perception. The identified images were recorded with corresponding arc seconds. Then the Titmus
circle test was completed to assess a finer grade of stereoacuity. Number of correctly identified circles were recorded.

**Ocular Dominance:** Dominant eye was determined by using the Miles test. Participants were instructed to extend their arms straight in the air and create a triangle with their hands, then frame some letters from the letter chart through that triangle. The examiner covered one eye at a time and instructed participant to report when the letters were no longer in his/her view. The eye framing the letters upon covering of the contralateral eye was recorded as the dominant eye. Also, the participants were questioned on their chirality (handedness).

**Station 4: Eye Photography**

**Eyelid Position:** To determine the symmetry between the eyes and also the parts of around the eyes including eyelids, a photo covering the area from temple to temple was taken with a standard pocket camera (Nikon Coolpix E995, Tokyo, Japan).

**Eye Colour:** To determine presence of pinguecula or pterygia, a colour photo of the iris in primary position, and the medial and lateral conjunctiva in dextroversion and laevoversion were taken using a digital camera (Nikon D100, Tokyo, Japan) fitted with 105 mm f/2.8 Micro Nikkor (Nikkor, Melville, New York, USA) lens.

**Conjunctival UV Auto-fluorescence:** A camera system developed by Coroneo and colleagues (Ooi 2006, 2007) was used to take the conjunctival UV auto-fluorescence photos for each participant. The camera system included a height adjustable table equipped with subject head-rest, camera positioning assembly, digital single-lens reflex
camera (Nikon D100 (Nikon, Melville, New York, USA)), 105 mm f/2.8 Micro Nikkor (Nikkor, Melville, New York, USA) lens, and filtered electronic flash. Both nasal and temporal regions of the eyes were photographed at 0.94 magnification. All images were saved in RGB format at the D100 settings of JPEG Fine (1:4 compression) and large resolution (3,000 x 2,000 pixels).

**Station 5: Dilation**

**Intraocular Pressure (IOP) measurement:** IOP of the eyes were measured with the ICare TAO1i Tonometer (Icare Finland, Oy, Helsinki, Finland). Six consecutive measurements were taken for each eye and mean measurement was recorded.

**Cycloplegia:** If the IOP was less than 22mmHg in each eye, one drop of Tropicamide 1% and one drop Phenylepherine 10% were administered for dilation of pupils. In the event of high pressures, an ophthalmologist checked for the closure of the angle of the anterior chamber of the eye.

**Eyelash measurement:** One upper eyelash from each eye was measured with a ruler and the average of the two measurements was recorded in millimeters. Station 6:

**Station 6: Higher Order Aberrations**

**Zywave II Wavefront Aberrometer:** To detect the higher order aberrations in the optic media, we measured the refractive error by using the Zywave II Wavefront Aberrometer (Bausch&Lomb, Inc., Rochester, NY). The measurements were completed prior to the cycloplegia. Measurements with no readings repeated post-cycloplegia. In
the presence of keratoconus, a secondary measurement was completed with Orbscan II (Bausch&Lomb, Inc., Rochester, NY)

Station 7: Ocular Biometry

IOL Master: Ocular biometric parameters including axial length, corneal curvature, anterior chamber depth and horizontal corneal diameter were measured with the IOLMaster V.5 (Carl Zeiss Meditec AG, Jena, Germany). For AL, five consecutive measurements were taken until the following criteria were satisfied: measurements within ± 0.02mm of each other, good waveform – no double peaks, acceptable signal-to-noise ratio >2.0. Any measurement outside the mentioned criteria was deleted and repeated. During keratometry, three measurements within 0.3D within each meridian with careful alignment and focus were recorded. Next, five consecutive ACD measurements were taken when the fixation point was sharply focused. Finally, three corneal diameter measurements within 0.2mm were recorded for each eye. Right and left eye measurements were within 0.02mm.

Station 8: Corneal Topography

Oculus Pentacam: Anterior segment topography of each dilated eye was taken with the Oculus Pentacam (Optikgerate GmbH, Wetzlar, Germany). Quality of the images were checked and repeated if necessary.

Tomey Specular Microscopy: Endothelial cell count analysis helps to evaluate the endothelial cell density, variation in size (polymegathism), variation in shape (polymorphism), and other corneal factors such as injury, inherent disease, and inflammatory or foreign material. EM-3000 Tomey Specular Microscopy (Tomey
Corp., Nagoya, Japan) was used to complete endothelial cell count analysis on each participant. Unclear screen shots were repeated.

Station 9:

**Fundus Photography:** Three sets of retinal photographs were taken for each eye using the Canon CF-60DSI Digital Fundus Camera (USA). The lens angle was set to 60 degrees and the illumination dial was turned to level 2. First, the participants were instructed to stare directly ahead into the camera and a macula-centered photo was taken. Then, the participants were instructed to look at the fixation light attached to the top of the headrest and the disc-centered photo was taken. Finally, the photo setting on the computer screen was changed from “colour” to “red-free”, red free filter of the camera was removed, and the illumination was increased to level 3 to take the disc-centered red-free photo. If the quality of the images was unclear or there was an artefact, ie eyelashes, the photographs were retaken.

**Stereo-disc Photography:** Optic Disc Photography 3-Dx Fundus Camera (Nidek Co, Japan) was used in this study to examine the anatomy of the optic disc in three-dimension. Participants were appropriately positioned in front of the camera and instructed to look at the internal or external fixation lights during the photography shot. The quality of the image was assessed. When it was unclear, poorly illuminated, or the discs were in different colours, the photograph was retaken.

Station 10:

**Optical coherence tomography (OCT):** Images were captured using a Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany). Volume scans were
obtained for the macula (30° x25°, 31 slices at 240 microns) and optic disc (15°x10°, 49 slices at 30 microns). Retinal nerve fiber layer (High speed, ART16) and choroidal thickness images (2 x horizontal & 2 vertical EDI, High speed, ART100) were also taken.

Station 11: Post dilation autorefraction

Post-cycloplegic autorefraction: Autorefraction and keratometry results were repeated after adequate pupillary dilation.

Station 12:

Heidelberg Retina Tomography: Glaucoma is more common in individuals aged over 40, however it could occur at any age. Confocal scanning laser ophthalmoscope is used in analysis of glaucoma manifestation and progression. A baseline glaucoma analysis was done on each participant using the Heidelberg Retina Tomography (Heidelberg Engineering, Heidelberg, Germany). The participant was appropriately positioned in front of the camera and instructed to stare at the flashing light. The position of the camera was adjusted to illuminate and sharpen the image of the optic disc. Poor images were repeated. Each scan was reviewed at the end, and the mean standard deviation of less than 20 µm was maintained for quality check.

Participant Eye Report (Feedback)

At the end of the examination session, each participant was given a detailed debriefing on the results of the completed examinations, provided with a summary of his/her current ocular health status, and given the opportunity to ask questions. Any newly diagnosed pathology was discussed with the participants, and any individuals requiring
An ophthalmological intervention were referred to the clinical experts in sub-specialties. A feedback report outlining the participant’s ocular status as well as a copy of macular-centred retinal photos were provided to all participants. Later, an analysis of the preliminary signs of UV damage and a copy of the conjunctival UV auto-fluorescence photos were mailed to each participant.
3.1 Principal Component Analysis (PCA) Plots of REHS and TEST/BATS Population Structures
3.2 The 25 Most Significant SNPs from the Meta-Analysis of Corneal Curvature

<table>
<thead>
<tr>
<th>Marker</th>
<th>Allele</th>
<th>Effect</th>
<th>SE</th>
<th>P value</th>
<th>Chr</th>
<th>Pos b36</th>
<th>Pos b37</th>
<th>Gene (within 50KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2444240</td>
<td>T</td>
<td>-0.0364</td>
<td>0.00713</td>
<td>3.66×10^{-7}</td>
<td>11</td>
<td>119545652</td>
<td>120040442</td>
<td>TRIM29</td>
</tr>
<tr>
<td>rs494965</td>
<td>T</td>
<td>0.0366</td>
<td>0.00728</td>
<td>4.50×10^{-7}</td>
<td>11</td>
<td>119558860</td>
<td>120053650</td>
<td>TRIM29</td>
</tr>
<tr>
<td>rs470606</td>
<td>T</td>
<td>-0.0359</td>
<td>0.00713</td>
<td>5.66×10^{-7}</td>
<td>11</td>
<td>119533412</td>
<td>120028202</td>
<td>TRIM29</td>
</tr>
<tr>
<td>rs470373</td>
<td>T</td>
<td>0.0362</td>
<td>0.00728</td>
<td>5.90×10^{-7}</td>
<td>11</td>
<td>119531481</td>
<td>120026271</td>
<td>TRIM29</td>
</tr>
<tr>
<td>rs4936509</td>
<td>T</td>
<td>-0.0335</td>
<td>0.00712</td>
<td>2.82×10^{-6}</td>
<td>11</td>
<td>119546654</td>
<td>120044444</td>
<td>TRIM29</td>
</tr>
<tr>
<td>rs7146198</td>
<td>A</td>
<td>0.0480</td>
<td>0.0104</td>
<td>3.94×10^{-6}</td>
<td>14</td>
<td>69713922</td>
<td>70644169</td>
<td>SLC8A3</td>
</tr>
<tr>
<td>rs740145</td>
<td>C</td>
<td>-0.173</td>
<td>0.0367</td>
<td>4.52×10^{-6}</td>
<td>7</td>
<td>29888049</td>
<td>29921524</td>
<td>WIPF3</td>
</tr>
<tr>
<td>rs2383876</td>
<td>A</td>
<td>-0.0508</td>
<td>0.0111</td>
<td>5.25×10^{-6}</td>
<td>8</td>
<td>74228733</td>
<td>74066179</td>
<td>RPESP</td>
</tr>
<tr>
<td>rs7144262</td>
<td>T</td>
<td>0.0468</td>
<td>0.0103</td>
<td>5.37×10^{-6}</td>
<td>14</td>
<td>69716861</td>
<td>70647108</td>
<td>SLC8A3</td>
</tr>
<tr>
<td>rs2250402</td>
<td>T</td>
<td>0.0592</td>
<td>0.0132</td>
<td>6.20×10^{-6}</td>
<td>15</td>
<td>38109844</td>
<td>40322552</td>
<td>EIF2AK4</td>
</tr>
<tr>
<td>rs1999052</td>
<td>A</td>
<td>0.0464</td>
<td>0.0103</td>
<td>6.24×10^{-6}</td>
<td>14</td>
<td>69716944</td>
<td>70647191</td>
<td>SLC8A3</td>
</tr>
<tr>
<td>rs17093693</td>
<td>T</td>
<td>0.0465</td>
<td>0.0103</td>
<td>6.33×10^{-6}</td>
<td>14</td>
<td>69719545</td>
<td>70649792</td>
<td>SLC8A3</td>
</tr>
<tr>
<td>rs11987235</td>
<td>A</td>
<td>-0.0384</td>
<td>0.00848</td>
<td>6.93×10^{-6}</td>
<td>8</td>
<td>117710791</td>
<td>117641610</td>
<td>EIF3H</td>
</tr>
<tr>
<td>rs2693676</td>
<td>A</td>
<td>-0.0435</td>
<td>0.00962</td>
<td>7.22×10^{-6}</td>
<td>14</td>
<td>98677233</td>
<td>99607480</td>
<td>BCL11B</td>
</tr>
<tr>
<td>rs1273225</td>
<td>T</td>
<td>-0.0434</td>
<td>0.00961</td>
<td>7.29×10^{-6}</td>
<td>14</td>
<td>98678970</td>
<td>99609217</td>
<td>BCL11B</td>
</tr>
<tr>
<td>rs1273226</td>
<td>T</td>
<td>0.0438</td>
<td>0.00981</td>
<td>7.39×10^{-6}</td>
<td>14</td>
<td>98677595</td>
<td>99607842</td>
<td>BCL11B</td>
</tr>
<tr>
<td>rs2834902</td>
<td>A</td>
<td>0.0358</td>
<td>0.00804</td>
<td>7.53×10^{-6}</td>
<td>21</td>
<td>35610398</td>
<td>36688528</td>
<td>-</td>
</tr>
<tr>
<td>rs807443</td>
<td>A</td>
<td>-0.0445</td>
<td>0.00997</td>
<td>8.12×10^{-6}</td>
<td>14</td>
<td>98698257</td>
<td>99628504</td>
<td>BCL11B</td>
</tr>
<tr>
<td>rs2307101</td>
<td>T</td>
<td>-0.0594</td>
<td>0.0132</td>
<td>8.41×10^{-6}</td>
<td>15</td>
<td>38109752</td>
<td>40322460</td>
<td>EIF2AK4</td>
</tr>
<tr>
<td>rs1869463</td>
<td>A</td>
<td>-0.0331</td>
<td>0.00741</td>
<td>8.85×10^{-6}</td>
<td>4</td>
<td>58999889</td>
<td>59305132</td>
<td>-</td>
</tr>
<tr>
<td>rs2007685</td>
<td>T</td>
<td>-0.0331</td>
<td>0.00741</td>
<td>8.86×10^{-6}</td>
<td>4</td>
<td>59004041</td>
<td>59309284</td>
<td>-</td>
</tr>
<tr>
<td>rs17137734</td>
<td>T</td>
<td>0.0892</td>
<td>0.0203</td>
<td>9.17×10^{-6}</td>
<td>15</td>
<td>25090519</td>
<td>27507773</td>
<td>GABRG3</td>
</tr>
<tr>
<td>rs10152133</td>
<td>A</td>
<td>0.0910</td>
<td>0.0205</td>
<td>9.31×10^{-6}</td>
<td>15</td>
<td>25093268</td>
<td>27510522</td>
<td>GABRG3</td>
</tr>
<tr>
<td>rs10152135</td>
<td>T</td>
<td>0.0900</td>
<td>0.0205</td>
<td>9.35×10^{-6}</td>
<td>15</td>
<td>25093283</td>
<td>27510537</td>
<td>GABRG3</td>
</tr>
<tr>
<td>rs10152293</td>
<td>A</td>
<td>-0.0879</td>
<td>0.0196</td>
<td>9.59×10^{-6}</td>
<td>15</td>
<td>25093096</td>
<td>27510350</td>
<td>GABRG3</td>
</tr>
</tbody>
</table>