# *Genetic Study on Ocular Quantitative Traits of Glaucoma and Myopia in Indian population*

**THESIS**

Submitted in partial fulfillment

of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

By


## FERDINAMARIE SHARMILA P


Under the Supervision of

**Dr. S. Sripriya**



**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**2014**

## BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

# CERTIFICATE

This is to certify that the thesis entitled **"Genetic Study on Ocular Quantitative traits of Glaucoma and Myopia in Indian population"** submitted by **Ms.P.Ferdinamarie Sharmila** ID No **2009PHXF018P** for the award of Ph.D. degree of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor

Name in capital letters : **Dr.S.Sripriya**

Designation : **Principal Scientist & Associate Professor**

**SNONGC Department of Genetics and Molecular Biology**

**Vision Research Foundation**

**Sankara Nethralaya, Chennai- 600006**

**Tamilnadu India.**

Date:

**Acknowledgements**

First and Foremost I submit by thanks and gratefulness to **THE ALMIGHTY** for providing me with the strength and perseverance to complete this important phase of my career.

I thank Padmabhushan **Dr. S.S. Badrinath**, Chairman Emeritus, Sankara Nethralaya for giving me the opportunity to be a part of this esteemed institution and to carry out my doctoral work.

I thank **Mr. R. Rajagopal**, President VRF, for permitting me pursue my PhD under VRF. I also thank **Dr. S. Meenakshi**, Director of Academics, **Prof. H.N. Madhavan,** Director of Microbiology for permitting me to take up this doctoral program under DLPD programme of BITS, Pilani and also for providing necessary infrastructure and resources to accomplish my research work. I also extend my gratitude to **Dr. Lingam Vijaya**, Director of Glaucoma dept for her kind support. My sincere thanks to **Dr. Ronnie J George,** Research Director for his continued support. I extend my gratitude to **Dr. Rama Rajagopal**, Research Advisor for her kind support.

I thank **Dr. A.J. Pandian**, HOD department of Genetics and Molecular Biology for his enthusiastic support.

I am grateful to **Dr. S. Sripriya**, my guide, for her valuable advice, constructive criticism and extensive discussions around my work. I owe my heartfelt thanks for her constant care and motivation through difficult times.

I extend my sense of gratitude and appreciation to **Prof. G. Kumaramanickavel**, my former HOD, whose wisdom, knowledge and commitment to the highest standards inspires and motivates me.

My sincere thanks to **Chennai Wellington Corporate Foundation, Department of Biotechnology (DBT)** and **Indian Council of Medical Research (ICMR)** Government of India - New Delhi for providing financial assistance in the form of Senior Research Fellowship.

I extend my sense of gratitude to **Prof. B.N.Jain,** Vice chancellor and Director BITS**, Prof. Sanjay Kumar Verma**, Dean ARD BITS and **Prof. Hemant R. Jadev**, Associate Dean ARD BITS for allowing me to get associated with this renowned institute for

both my Masters (MS MLT) and PhD. My Sincere thanks to **Dr. P.R. Deepa**, BITS co-ordinator for her support and valuable advice. My sincere thanks to the DAC members **Dr.Angshuman Sarkar** and **Dr.Utpal Roy** for their insights and evaluation of my thesis. I also extend my sincere thanks to **Mr.Navin**, and **Mr.Dhineshkumar** of PhD monitoring.

My sincere thanks to **Dr. Janey Wiggs**, Associate Chief Ophthalmology Clinical Research and Associate Director Ocular Genomics Institute for making me a part of INDO-US project. I extend my thanks and gratitude to **Dr. Baojian Fan**, Instructor in Ophthalmology, Massachusetts Eye and Ear Infirmary for his initiative in Indo-US project.

I thank **Dr. K. Lily Therese**, VIBS co-ordinator for her continued support and encouragement. My extended sense of gratitude to **Drs. K.N. Sulochana, N. Angayarkanni, Dorien Gracious, S. Krishnakumar, J. Malathi, B. Mahalakshmi, K. Coral, J. Subbulakshmi, V.Umashankar, Nivedita Chatterjee** for evaluating my practice lectures.

I also thank **Mr. S. Narayan,** Manager VRF for mobilizing funds required for research work. I also thank the VIBS staff members for co-ordinating all the documentations.

My sincere thanks to my colleagues **Dr. N Soumittra**, **Mr. T. Arokiaswamy**, **Mrs. Srilekha, Mr.Ramprabhu, Mrs. Srikrupa, Mr. Malaichamy, Mrs. Vinita Kumari, Ms. Bhavana roa, Ms. D. Sudha, Ms. K. Sudha**, **Ms Karthiyayene**, **Mr. JayaPrakash**, **Ms. Jyothi**, **Mr. Babu** for their sincere encouragement and inspiration; especially in extending their help in various ways, despite the enormous work pressures we were facing together.

My special appreciation goes to **Mrs. Gomathy Narayanan**, and **Mrs. Sathyapriya**, for their continued friendship, encouragement and several rounds of discussions on various aspects of science and personal issues which has helped me in various situations.

Heartfelt gratitude to **Mrs. Rashima Ashokan** optometrist for help in clinical data management, **Ms.Sacikala**, **Mr.Rajesh** for all the technical help they have extended

without any hesitance and **Mr. Madhan** for identifying and motivating patients to participate in the study.

My gratitude and heartfelt thanks to all **Patients** and their family members for willingly coming forward to provide their DNA for my research.

My love and appreciation to **Mr. Marie Antony** my husband, without whom this effort would have been worth nothing. Your love, support and constant patience have taught me so much about sacrifice, discipline and compromise – even if there were times when you said "I told you so".

My heartfelt thanks and respect to my In-Laws for all their encouragement and profound understanding.

My hearty wishes and thanks to my lovely sisters **Ms. Esther** and **Ms. Priscilla**.

And finally I dedicate this thesis work to my parents **Mr. Philomenadin**, and **Mrs. Sagayamarie** with whose blessings, support and encouragement all this was possible.

## Abstract

### Aim

Endophenotypes are measurable or quantifiable clinical traits (QTs) which are highly inherited. Identification of gene loci for these precursor risk traits has potential significance in risk prediction and disease management. Ocular QTs like optic disc area (ODA), vertical cup disc ratio (VCDR) etc., are implicated in eye diseases like glaucoma, myopia and others in different populations. The current study was undertaken to evaluate the association of known QT loci with endophenotypes like AL, CCT, IOP and VCDR.

### Methodology

We included three different sample sets namely (a) multigenerational consanguineous families without any history of ocular disorders (b) POAG and (c) myopia. The samples were recruited from the Chennai glaucoma study, INDO-US study on QTs and Myopia study after detailed clinical evaluation. Heparinised blood sample was collected from these patients after informed consent for DNA extraction and subsequent genotype analysis.

The association of SNPs flanking *ATOH7* (rs1900004, rs3858145, rs17231602, rs4746741) and *ZNF469* (rs77330683) with optic disc parameters was analysed in large consanguineous south Indian pedigrees. The individuals were genotyped by direct sequencing. The effect size (β) and *P* value were obtained using the QFAM procedure in PLINK, which performs a simple linear regression of phenotype on genotype, and then uses a special permutation procedure to correct for family structure.

The distribution of SNPs rs1900004, rs3858145 (*ATOH7*), rs10483727 (*SIX1/SIX6),* rs1063192 (*CDKN2B*), rs9607469 (*CARD10*), rs1192415 (*CDC7*) in POAG (cases= 97) and controls= 360) and *LOXL1* (rs3825942 and rs1048661) in Pseudoexfoliation (cases=114 and controls=135) were analysed for putative association with IOP, AL, CCT and VCDR. A model-based (additive, recessive, dominant) Chi square and linear regression analysis was performed to analyse the statistical significance; while the interaction analysis was performed by logistic regression methods.

Mutation screening of the eye development genes involved in AL regulation (*MFRP, VSX2* and *PAX6)* was done in a total of 189 samples with (N= 98) and without (N= 91) myopia. Genotype / haplotype / QT analysis was performed using PLINK, Haploview and Thesias

softwares. Bioinformatic analysis was performed using FASTSNP, SIFT, Polyphen, CAIcal, MUPro. EasyLinkage – FastLink package were used for linkage analysis.

**Results**

All the SNPs compiled with HWE. In multigenerational consanguineous families, QFAM analysis showed significant association of rs1900004, rs3858145 and rs4746741 with optic disc area, (*P*= 0.02, 0.013, 0.015 respectively) and rs4746741 with cup area (*P*= 0.026).

In POAG group, nominal significance (*P*<0.05) was observed for QTs like i) VCDR with SNPs rs1900004 (*ATOH7*); rs1192415 (*CDC7)*; rs10483727 (*SIX1/SIX6),* rs9607469 (*CARD10);* ii) CCT with rs1192415*;* iii) IOP with rs1900004 and iv) AL with rs1900004 and rs1063192 (*CDKN2B).* We were able to replicate previously known interactions between *ATOH7-SIX6* and *SIX6-CDKN2B* along with few novel interactions between *ATOH7 -CDC7* and *SIX6* with genes including *CARD10* and *CDC7.*

Fifteen variations were observed in the *MFRP* gene of which, rs36015759 (c.492C>T, T164T) in exon 5 was distributed at a high frequency in the controls and significantly associated with a low risk for myopia (*P*=4.10*$e^{-07}$ OR <1.0). An increased frequency for the coding haplotype block [CG**T**CGG] harbouring rs36015759 was observed in controls (31%) than cases (8%) that also correlated with a decreased mean AL (-1.35085; *P*=0.000444) by Thesias analysis. The 'T' allele of rs36015759 was predicted to abolish the binding site for splicing enhancer (SRp40) by FASTSNP analysis. None of the variations in genes *VSX2* and 3' UTR of *PAX6* showed significant association with myopia /AL. Exclusion mapping of known myopia / its QTLs were done by genome wide approach in 2 myopia families. We have also identified possible myopia regions by linkage to chrs 3q27.1 (LOD: 1.4), and 17p13.2 (LOD: 1.7) for family A and chr 2q36.3 with homozygosity mapping in family B.

**Conclusion**

The data generated in the current study has helped in generating knowledge repository of the allele frequencies for the ocular QTLs in Indian population. The study also replicated the associations of candidate genes like *ATOH7, SIX1/SIX6, CARD10,* with endophenotypes like VCDR, and IOP in Indian population. Additionally novel associations were identified between genes and QTs like *SIX1/SIX6*-CCT, *CDC7*–CCT, *MFRP*-myopia and chromosomal regions 3q27.1, 17p13.2 and 2q36.3 with myopia in

Indian population. These associations have to be replicated and functionally validated to understand the role of these genes in disease pathology. This would help in identification of risk individuals at an earlier stage especially in case of age related disorders (myopia and glaucoma) where appropriate treatment will help in preventing/ reducing the morbidity involved.

## Table of Contents


| Chapter | Title | Page no |
|---|---|---|
| | **List of tables** | 3 |
| | **List of figures** | 6 |
| | **List of abbreviations/symbols** | 8 |
| **1** | **Genetic study designs of complex diseases** | 11 |
| | Organization of thesis | 22 |
| **2** | **Genetic mapping of endophenotypes of Glaucoma** | 23 |
| | Introduction to Primary open angle glaucoma, Pseudoexfoliation glaucoma, and its endophenotypes. | |
| **2.1** | **Analysis of association between optic disc parameters and SNPs near ATOH7 and ZNF469 in normal consanguineous families** | 32 |
| | Objective | |
| | Methodology | |
| | Results | |
| | Discussion | |
| **2.2** | **Genetic association study for the endophenotypes of glaucoma with SNPs near ATOH7, CDC7, CDKN2B, CARD10 and SIX1 in POAG** | 39 |
| | Objective | |
| | Methodology | |
| | Results | |
| | Discussion | |
| **2.3** | **Genetic association study of glaucoma endophenotypes with SNPs of LOXL1 in XFS/XFG** | 49 |
| | Objective | |
| | Methodology | |
| | Results | |
| | Discussion | |
| **3** | **Genetic Mapping endophenotypes of myopia** | 54 |
| | Introduction to myopia and its endophenotypes | |
| **3.1** | **Genetic analysis of MFRP, VSX2 and PAX6 for its association with AL in myopia** | 62 |
| | Objective | |
| | Methodology | |
| | Results | |
| | Discussion | |
| **3.2** | **Exclusion mapping of MYP locus in myopia families** | 79 |
| | Objective | |
| | Methodology | |
| | Results | |

| | | |
|---|---|---|
| | Discussion | |
| **4** | **Digest** | |
| | Summary | 92 |
| | Specific contribution | 94 |
| | Future scope | 94 |
| | **References** | 96 |
| | **Appendices** | 110 |
| | **List of publications and presentations** | 129 |
| | **Brief Biography of the Candidate** | 132 |
| | **Brief Biography of the Supervisor** | 133 |

## List of Tables


| Table no | Title | Page No |
|---|---|---|
| Table 2.1 | Summary of the genetic association of rs1048661, rs3825942 in exon 1 of *LOXL1* gene with XFS/XFG | 29 |
| Table 2.1.1 | Primers flanking the SNPs rs4746741, rs1900004, rs3858145, rs17231602 and rs77330683 | 32 |
| Table 2.1.2 | Genotype distribution for the SNPs rs1900004, rs3858145, rs17231602, rs4746741 and rs77330683 | 34 |
| Table 2.1.3 | Association analysis of *ATOH7* and *ZNF469* in Indian consanguineous families with i) disc area ii) cup area iii) cup/disc ratio and iv) CCT | 36 |
| Table 2.2.1 | Primer sequences, product size and Tm for PCR amplification of rs1900004, rs3858145, rs10483727, rs1192415, rs1063192, and rs9607469 | 39 |
| Table 2.2.2 | Demographic and clinical features of the study subjects | 40 |
| Table 2.2.3 | Analysis of minor allele frequencies and $\chi^2$ of the SNPs between POAG cases (HTG, NTG, combined) and controls | 41 |
| Table 2.2.4 | Multiple regression analysis for the SNPs with AL, CCT, and VCDR as dependant variables in the whole cohort | 42 |
| Table 2.2.5 | Association of the SNPs with VCDR and CCT in controls and POAG (HTG/NTG/combined) | 43 |
| Table 2.2.6 | Coding of genotypes in dominant and recessive model analysis of interaction between genes. | 44 |
| Table 2.2.7 | Analysis for interaction between SNPs in HTG/NTG by dominant and recessive models | 45 |
| Table 2.3.1 | Primer sequences, product size and Tm for PCR amplification of *LOXL1* exon1 | 49 |
| Table 2.3.2 | Demographic details of the subjects included in the study | 50 |
| Table 2.3.3 | Distribution of genotype and allelic frequency among the cases and controls | 51 |
| Table 2.3.4 | Linear regression analysis of rs1048661 and rs3825942 | 51 |

| | | |
|---|---|---|
| | with the endophenotypes in the whole cohort | |
| Table 2.3.5 | Linear regression analysis of rs1048661 and rs3825942 with the endophenotypes in cases and controls | 51 |
| Table 3.1 | List of loci mapped for myopia worldwide | 56 |
| Table 3.1.1 | Primer sequences, product size and $T_m$ for PCR amplification of *MFRP* | 62 |
| Table 3.1.2 | Primer sequences, product size and Tm for PCR amplification of *VSX2* | 63 |
| Table 3.1.3 | Primer sequences, product size and Tm for PCR amplification of the 3'UTR region of *PAX6* | 63 |
| Table 3.1.4 | Demographic data of the study subjects | 64 |
| Table 3.1.5 | Genotype and allele frequency distribution of *MFRP* variants in myopia patients | 65 |
| Table 3.1.6 | Distribution of *MFRP* haplotypes and chi-square *P* values as analyzed by Haploview v4.1 | 67 |
| Table 3.1.7 | Summary of exhaustive haplotype analyses based on omnibus test of sliding windows of all possible sizes across the 15 SNPs of the *MFRP* gene under this study | 68 |
| Table 3.1.8 | Details of haplotype analysis for the 6 window showing the most significant results among the all possible sliding windows | 69 |
| Table 3.1.9 | Frequency distribution of *MFRP* haplotypes and their interaction with AL as generated by Thesias software v3.1 | 70 |
| Table 3.1.10 | Bioinformatic analysis by Polyphen, SIFT and SNAP for non synonymous variations of *MFRP* to determine the pathogenicity | 70 |
| Table 3.1.11 | FastSNP and MUpro analysis for the putative effect of *MFRP* gene variants on its function and protein stability. | 71 |
| Table 3.1.12 | Genotype and allele frequency distribution of *VSX2* variants detected and investigated in myopia patients. | 72 |
| Table 3.1.13 | Association of *VSX2* haplotypes with myopia as analyzed by a. Haploview and b. Thesias softwares | 73 |
| Table 3.1.14 | Genotype and allele frequency distribution of *PAX6* | 74 |

| | | |
|---|---|---|
| | variants in 3'UTR region | |
| Table 3.2.1 | Comparison of LOD scores between the current study and reported literature for all the reported myopia loci | 82 |
| Table 3.2.2 | List of markers in chrs 20, 17 and 3 and the respective LOD scores suggestive of linkage (-2 to +2 ) at maximum recombination | 83 |
| Table 3.2.3 | Comparison of genotypes of SNPs in 250K region between II1 and II3 individuals | 87 |
| Table 3.2.4 | Region on chr2q36.3 narrowed down using homozygosity mapping. | 88 |
| Table 4.1 | Complete list of associations of SNPs and QTs seen in the present study | 93 |

## List of Figures


| Figure No | Title | Page No |
|---|---|---|
| Figure 1.1 | Interaction between genetic variants and environmental factors in (A) single gene and (B) complex disorders | 11 |
| Figure 1.2 | Distribution of various causes of Visual Impairment | 12 |
| Figure 1.3 | Flowchart depicting the different methods employed in mapping genes | 12 |
| Figure 1.4 | Pictorial representation of crossing over during meiosis | 13 |
| Figure 1.5 | Homozygosity mapping of recessive disease genes | 16 |
| Figure 1.6 | SNPs, Haplotype and Tagged SNPs | 17 |
| Figure 1.6 | Flowchart of thesis organization | 22 |
| Figure 2.1 | Diagrammatic representation of formation of optic nerve cupping in POAG | 24 |
| Figure 2.2 | Photographic image of: A) Normal pupil B) Deposits of pseudoexfoliative material around the pupil edge. | 28 |
| Figure 2.3 | Pathway depicting the development of glaucoma from the PEX deposits | 28 |
| Figure 2.1.1 | Pedigrees of the 4 families included in the study | 34 |
| Figure 2.1.2 | Electrophoretogram of *ATOH7* and *ZNF469* genomic sequence showing the variations in i) rs1900004, ii) rs3858145, iii) rs17231602, iv) rs4746741, and v)rs77330683 | 35 |
| Figure 2.2.1 | Electrophoretogram of i) rs1900004, ii) rs3858145, iii) rs1192415, iv) rs1063192 and v) rs9607469 | 41 |
| Figure 2.3.1 | Electrophoretogram of a) rs1048661, b) rs3825942 | 50 |
| Figure 3.1 | Comparison of normal and myopic eye structure | 54 |
| Figure 3.2 | Different mechanisms leading to myopia development | 55 |
| Figure 3.3 | Schematic representation of *MFRP* protein domains | 59 |
| Figure 3.4 | Schematic representation of *VSX2* protein domains | 60 |
| Figure 3.1.1 | Electrophorogram representing the 15 *MFRP* variants ( homozygous and heterozygous forms )in *MFRP*, seen in the present study. A-rs199473710, B-rs883247, C-rs79836575, D-rs3814762, E-rs199473708, F-rs36015759, G-rs2510143, H-rs61736238, I-rs10892350, J-rs2509388, K-rs35885438, L-rs185451482, M-rs199473709, N-rs199473711, O-rs11217241 | 66 |

| | | |
|---|---|---|
| Figure 3.1.2 | Haplotype block generated by Haploview software v4.1 for the 15 variations of *MFRP* | 67 |
| Figure 3.1.3 | ESE binding sites as predicted using ESE Finder for rs36015759, wild and variant allele. | 71 |
| Figure 3.1.4 | Electrophorogram representing the 6 variants (homozygous and heterozygous forms) in *VSX2*, seen in the present study. A-rs35435463, B-rs192712847, C-rs137872696, D-rs7595981, E-62006815, F-rs199473712 | 72 |
| Figure 3.1.5 | Haplotype block generated by Haploview software v4.1 for 4 coding variations (rs62006815, rs199473712, rs75395981, rs137872696) observed in exon 5 of *VSX2* gene. | 73 |
| Figure 3.1.6 | Electrophorogram of P1 region showing the variation rs11407950 a) with forward primer, b) with reverse primer | 714 |
| Figure 3.2.1 | Pedigrees recruited in the current study | 79 |
| Figure 3.2.2 | Plot of total LOD scores across different chrs from linkage analysis using FastLink | 82 |
| Figure 3.2.3a | Graph depicting the LOD score of various SNPs in chr 20, the arrow shows rs1569608 (LOD-1.689) | 84 |
| Figure 3.2.3b | Haplotype segregation of SNPs around rs1569608 | 84 |
| Figure 3.2.4a | Graph depicting the LOD score of various SNPs in chr 17, the arrow shows rs724809 (LOD: 1.7778) | 85 |
| Figure 3.2.4b | Haplotype segregation of SNPs around rs724809 | 85 |
| Figure 3.2.5a | Graph depicting the LOD score of various SNPs in chr 3, the arrow shows rs1829515 (LOD-1.4) | 86 |
| Figure 3.2.5b | Haplotype segregation of SNPs around rs1829515 | 86 |

## List of Abbreviations/Symbols

| | |
|---|---|
| .CEL | Cell intensity file |
| µl | Micro litre |
| AL | Axial length |
| *ATOH7* | Atonal Homolog 7 (Dorsophila) |
| bp | Base pairs |
| *CARD10* | Caspase Recruitement Domain family member 10 |
| CCT | Central corneal thickness |
| *CDKN2B* | Cyclin Dependent Kinase inhibitor 2B |
| CDR | Cup-to-disc ratio |
| CHGR | Center for Human Genetic Research |
| CHR | Chr |
| *CHX10* | ceh-10 homeo domain containing homolog |
| cM | CentiMorgan |
| EDTA | Ethylene diamine tetra acetic acid |
| ELOVL5 | Elongation of long-chain fatty acids family member5 |
| ESE | Exonic splicing enhancer |
| FastSNP | A functional analysis and selection tool for SNP in lerge scale association study |
| FASTSNP | Function Analysis and Selection Tool for Single Nucleotide Polymorphisms |
| GCOS | GeneChip operating software |
| GTYPE | GeneChip genotyping analysis software |
| GWAS | Genome wide association studies |
| HOX | Homeobox protein |
| HTG | High tension glaucoma |
| IOP | Intra Ocular Pressure |
| Kda | KiloDalton |

| | |
|---|---|
| LD | Linkage Disequilibrium |
| LOD | Logrithimic of Odds |
| *MFRP* | Membrane frizzled related protein |
| MGH | Massachusetts General Hospital |
| *Mitf* | Microphthalmia transcription factor |
| mM | Milli Molar |
| ng | Nano Gram |
| NTG | Normal tension glaucoma |
| NGS | Next generation sequencing |
| ODA | Optic disc area |
| ONH | Optic nerve head |
| PEX/PEXG | Pseudoexfoliation syndrome / glaucoma |
| POAG | Primary Open Angle Glaucoma |
| POLYPHEN | Polymorphism Phenotyping |
| QC | Quality control |
| QT | Quantitative trait |
| QTL | Quantitative trait loci |
| RB | Recovery buffer |
| RE | Refractive error |
| RNFL | Retinal nerve fibre thickness |
| RPC | Retinal progenitor cells |
| RPE | Retinal pigment epithelium |
| SIFT | Sorting tolerant from intolerant |
| *SIX1* | Sine Oculis Homeobox homolog 1 |
| *SIX6* | Sine Oculis Homeobox homolog 6 |
| SMA | Smooth muscle actin |
| SNAP | SNP Annotation and proxy search |

| | |
|---|---|
| SR | Seriene arginine rich proteins |
| SRBD1 | S1 RNA binding domain1 |
| TBE | Tris EDTA buffer |
| TdT | Termianl deoxynucleotidyl Transferase buffer |
| TE | Tris buffer |
| TGF | Transforming growth factor |
| THESIAS | Testing Haplotype Effects In Association studies |
| TM | Trabecular meshwork |
| VCDR | Vertical Cup to Disc Ratio |
| *VSX2* | Visual system homeobox |
| WHO | World health organization |

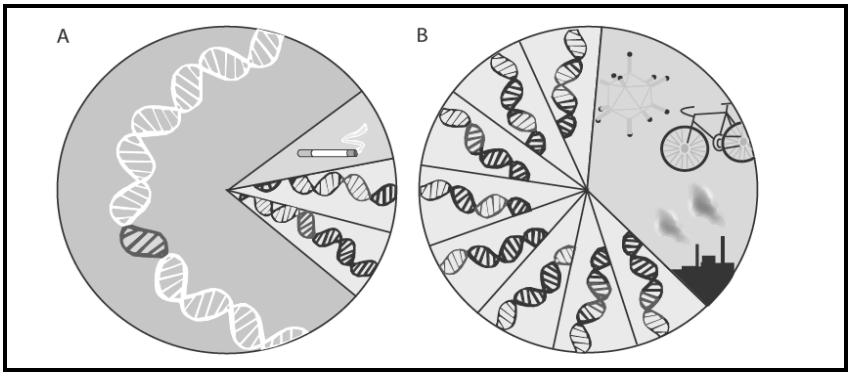
# Chapter 1

## Genetic study designs for complex diseases

### Introduction

The past two decades in modern genetics had made a great impact in the field of single gene disorders which are rare and often exhibit characteristic Mendelian inheritance patterns (autosomal dominant, recessive, X linked). "Complex Diseases" are the other spectrum of genetic disorders that arises from the interplay of genetic and environment factors (Fig: 1.1). The late onset of disease manifestation, epistasis (the effect of one gene modified by one or more genes) and pleiotrophy (multiple phenotypic effects from different alleles of the same gene) are some of the features which pose difficulty in dissecting the genetics of complex diseases. The risk predisposing locus determines the individual's susceptibility/resistance for the disease and inturn is dependent on various interacting factors (both genetic and environment).

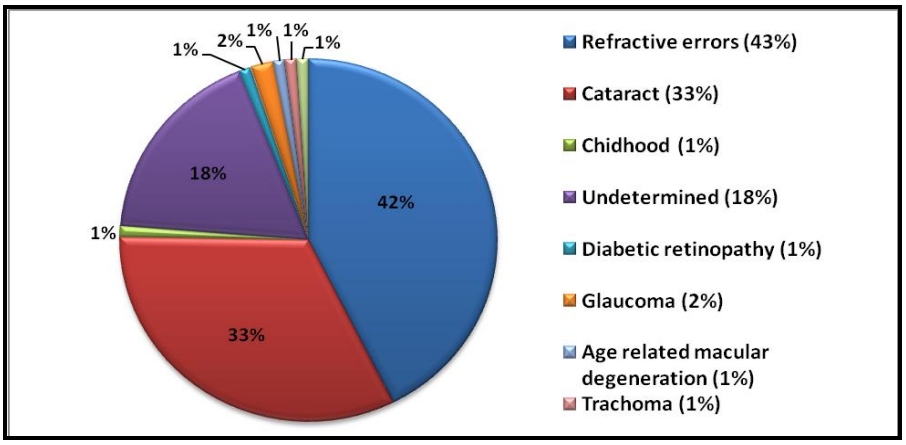
**Fig 1.1: Interaction of genetic variants and environmental factors as observed in (A) single gene and (B) complex disorders [1]**




Worldwide, the incidence of ocular diseases such as glaucoma, cataract, age related macular degeneration, diabetic retinopathy, myopia etc with complex inheritance pattern are increasing and expected to further raise by twofold in 2020 [2]. According to the statistics provided by WHO, out of 285 million visually impaired people, 39 million are blind, 246 have low vision, and about 90% of the world's visually impaired live in developing countries including India, China, Japan etc. The major causes of visual impairment include uncorrected refractive errors such as myopia, hyperopia etc (42%) [2]. In India, the common causes of low vision, ranks in the descending order of prevalence as refractive errors (68%) cataract (22%), glaucoma (2%) and others (Fig: 1.2) [3]. Since the

growing crisis of visual impairment causes huge economical and social burden to the society discovering the genetic etiology could throw more light on the causative genes / possible mechanisms which has potential therapeutic implications. In addition these details can also help in carrier testing for the afflicted family members, disease management and counselling.

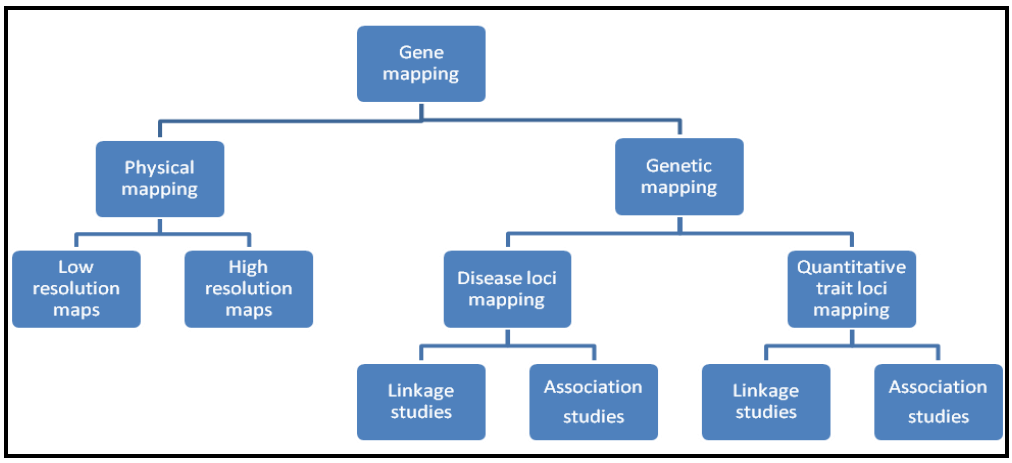
**Fig 1.2: Distribution of various causes of Visual Impairment [3]**




## Mapping of the disease loci/ genes

"Gene mapping" refers to mapping of genes to specific locations on chromosomes using markers that include detectable phenotypes (blood group, eye colour), protein markers (enzymes), DNA based markers (microsatellite markers, restriction fragment length polymorphisms (RFLPs), or single nucleotide polymorphisms (SNPs) and copy number variations) and others.

**Fig 1.3: Flowchart depicting the different methods employed in mapping genes**



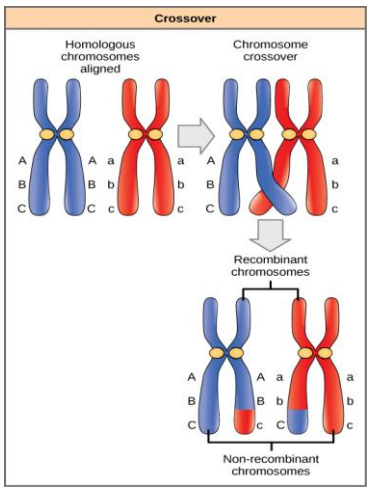
Gene mapping can be either physical or genetic mapping (Fig: 1.3)

I. **Physical mapping**: This method assigns the chromosomal loci for a gene based on the physical distance between the disease gene and marker. Different techniques are used in this method which provides various levels of resolution.
   i) Lower resolution maps are constructed based on the cytogenetic bands observed in the metaphase stage.
   ii) Higher resolution maps represent the distance between genes in actual physical measurements. The maps are generated by using techniques like somatic cell hybridization, radiation hybrid mapping, gene dosage mapping using patient cells, fluorescent *in situ* hybridization (FISH) [4, 5].

Physical mapping provides the actual distance between markers in terms of number of base pairs (bp).

II. **Genetic Mapping**: In this method, the genetic distance between two loci located on the same chromosome is defined based on the number of crossover events averaged from meiosis in patient's families (Fig: 1.4). The location of the disease gene with respect to the marker is expressed in terms of the recombination frequency (cM units).

**Fig 1.4: Pictorial representation of crossing over during meiosis [6]**




Genetic mapping can be further classified as

i) ***Disease gene mapping***: The end phenotype of the disease is used to classify the subjects as affected/unaffected. This stategy has remained the method of choice for several decades.

ii) ***Quantitative trait mapping***: In this method the precursor clinical traits that determine/define the individual's risk for the disease are used to stratify the disease

affection status. These clinical measurable parameters are called quantitative traits (QTs) eg blood pressure, cholesterol level etc and they are highly heritable. The ocular QTs include intra ocular pressure (IOP), axial length (AL), central corneal thickness (CCT), anterior chamber depth (ACD), refractive error (RE), retinal nerve fibre length (RNFL), optic disc parameters etc; these are altered in diseases like glaucoma, myopia, etc [7].

**Gene mapping strategies is moving towards QTL mapping**:

High heritability estimates of QTs make it advantageous to map the associated genes / loci that are important risk factors for common complex human diseases. Identification of the genes and understanding the subsequent molecular events responsible for these precursor QTs may help in localising the genetic susceptibility factors. Treating an underlying risk factor at the early stage could significantly reduce the morbidity associated with the condition for example IOP is a QT and elevated levels of the same pose high risk of glaucomatous damage to the optic nerve; lowering of IOP decreases the degree of optic nerve degeneration and thus offered as a treatment regime in these patients [8, 9]. The advantages of QTL mapping are being appreciated mostly in complex disorders; where disease gene mapping approaches often failed to identify the susceptible loci or the causative genes. Some examples of QT mapped and the corresponding diseases are i) modulating QT wave elongation measured by ECG in cardiac arrhythmia [10], ii) IgE levels in asthma [11], iii) serum uric acid level in gout [12], iv) lipid levels in coronary heart disease [13] and others. QTL mapping are simpler when compared to the identification of susceptibility genes/loci for complex diseases, as they offer the advantages of:

Precise definition of phenotype: The cases are defined precisely with the measurable triats, especially in age related disorders where people at risk may not manifest the disease phenotype at a younger age.

Reduction in number of loci mapped: It is much easier to map the QTs of complex diseases when compared to that of the disease due to high heritability estimates and less number of loci.

Reduced Genetic complexity: Since we map loci for traits and not the complete phenotype, it drastically reduces genetic complexity involved [14] for example it is likely that a genome scan for AL will yield fewer loci than the genome scan for myopia that includes AL, RE, environmental factors etc.

**Study designs for genetic mapping:**

Linkage and association based study designs are the two different methods employed in genetic mapping .

1. **Linkage mapping:** Identification of genetic loci is based on the distance between the DNA markers which is a measure in terms of frequency of recombination ($\theta$). The value of $\theta$ ranges from '0' (completely linked) to 0.5 (completely unlinked) and the genetic distance is measured as cM units; 1cM is defined as the genetic length between two loci at recombination frequency of 1%. The degree of recombination frequency is expressed in terms of LOD (Logarithm of the odds) score:

$$\text{LOD score} = z = \log \frac{\text{Probability that the two loci are linked}}{\text{Probability that the two loci are not linked}}$$

LOD score > 3 indicate complete linkage, -2 to+2 suggestive of linkage and values < -2 indicate recombination. Physical and genetic map distance are correlated as 1 cM= 1 Mb (1Mb= $10^6$ bp).

Two types of linkage mapping include i) parametric and ii) non parametric. In parametric linkage analysis, the exact mode of inheritance is a pre requisite and often performed for Mendelian disorders.

Nonparametric models are independent of inheritance patterns and analysed based on:

1. *Identical By Descent (IBD):* This state is defined as the presence of identical alleles in two or more individuals within a family and are inherited from a common ancestor [15].
2. *Identical By State (IBS):* Alleles at a particular locus on homologous chrs are said to be in IBS if they have identical nucleotide segments in this region. They differ from IBD on the basis that these similar nucleotide sequences can occur as a result of mutation and not necessarily from a common ancestor.

**Advantages of using consanguineous families for mapping QT**

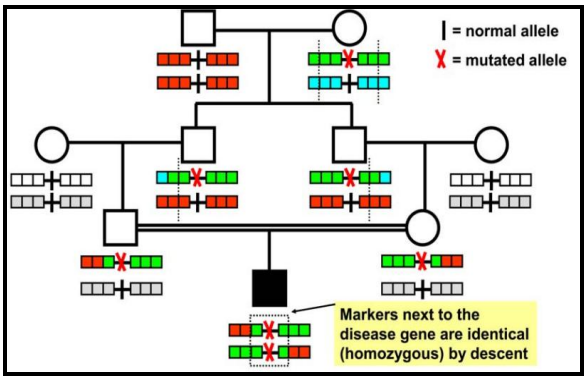
In recent years these properties have been extended to genetic mapping of complex diseases such as homozygosity mapping in families and GWAS. The major challenge in GWAS is to detect alleles of small effect and it demands larger sample size, and efficient study designing. This can be achieved by using consanguineous families especially in case of QTL mapping wherein members from consanguineous families which demonstrate broader variation in magnitude of a QT as there exits greater chance that siblings will inherit two copies of the same allele, or set of alleles contributing to the trait/disease

(assuming additive effects of the allele(s)) and hence these kind of consanguineous families are highly resourceful. These families can help us in discovering rare genetic variants whose implications in complex diseases are being discovered [16]. Founder effect in the inbred families, also reduces the population gene pool, thus minimizing the underlying genetic heterogeneity for a trait. It is also being demonstrated that inbreeding loops can greatly increase the statistical power to genetically map a QT. Additionally, increased prevalence of complex traits in inbred communities supports the hypothesis that predisposing factors can be over-represented in these populations. Earlier studies have also shown the success of this strategy to identify human QTL for IgE levels in asthmatics [17], cholesterol and low-density lipoprotein (LDL) [18], and fasting serum-insulin levels [19].

**Homozygosity mapping – Method of choice in autosomal recessive families**

Homozygosity mapping (HM), is a gene mapping strategy for autosomal recessive diseases, based on IBD. The figure (Fig: 1.5) below explains the strategy of HM. The pedigree shows an individual born out of consanguinity (shaded), has inherited the disease in autosomal recessive pattern. Different marker alleles are represented by different colours. Although for each parent-offspring succession there is a possibility of crossing over (dotted line) to occur in the parents gametes; there is a high likelihood for non recombination of consecutive markers surrounding the mutation and hence will be identical (homozygous) by descent [20].

**Fig 1.5: Homozygosity mapping of recessive disease genes [20]**

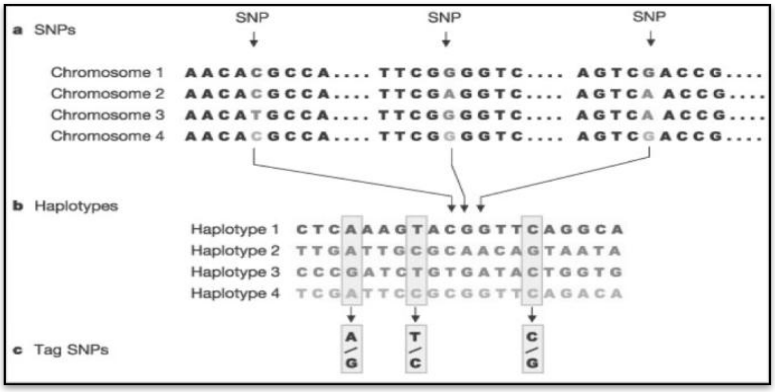



2. **Genetic Association studies:** The non random association between alleles at two or more loci within a population suggests common ancestor and this phenomenon called linkage disequilibrium (LD) have been used for genetic mapping of disease loci. The extent of LD in populations is expected to decrease with both time and recombination

distance between the markers. Association studies are based on LD and tests the distribution frequency of variation between cases and controls. The higher frequency of the disease variant in cases compared to the controls considers an allele to be risk predisposing for the disease, while it's termed protective if it is distributed more in controls. Association based studies are often performed for genetic variants within candidate genes or across the whole genome. The genetic alteration can lie in the regulatory or exonic regions with an effect on the protein expression/function or in LD with a candidate gene (tagged SNP approach).

Tagged SNP approach is based on the concept that SNPs in strong LD tend to inherit together. Thus genotyping one representative SNP provides the genotype of all SNPs in LD which in turn offers a rapid /reliable option of genotyping the whole genome. The process of selecting these SNPs in block is called "Tagging" and the SNPs are called "Tagged SNPs". Such tagging offers the advantage of analysing only small subsample of a study population [21, 22], making the strategy efficient as well as economical. The figure (Fig: 1.5) below depicts the SNPs, haplotype and tagged SNPs.

**Fig 1.6: SNPs, haplotype and Tagged SNPs ject [23]**




a) **SNPs** - Short stretch of DNA from four versions of the same chr region in different people. Each SNP has two possible alleles; the first SNP in panel a has the alleles C and T. B) **Haplotypes** - Shown here are the observed genotypes for 20 SNPs that extend across 6,000 bases of DNA. Only the variable bases are shown, including the three SNPs that are shown in panel a. C) **Tag SNPs** Genotyping just the three tag SNPs out of the 20 SNPs is sufficient to identify these four haplotypes uniquely.

The process of choosing the tagged SNPs has been made easier with the help of the Hap-Map project which has completed scrutinising the complete genome for common variants and providing the tag SNPs. Several algorithms have been used to identify and choose these SNPs, some of them include [22, 24]: Expectation maximisation algorithm (EM),

PHASE, LD select, Haploblock finder, Pairwise algorithm, Tagger, SNP Picker, HapTagger, Tagging SNP wizard etc.These tagged SNPs help in evaluating the association of whole genome with diseases. Several studies using this method have successfully identified the genetic loci for complex diseases such as thalassemias, sickle cell disease, various cancers etc [25]

**Requisite for gene mapping**:

Requirement for gene mapping by linkage / association studies basically includes a preparatory process, which consists of selecting the study subjects - pedigrees/cohort of cases and controls, comprehensive clinical examination, selection of markers and the analysis protocol, followed by selection of appropriate analysis programs with different algorithms.

I. **Study subjects**: Choosing pedigrees for linkage analysis is important as they must be informative to provide a significant LOD score. In an autosomal dominant disease the gametes of the affected parent (or non-penetrant gene carrier parent) provide linkage information to the LOD score while the unaffected non-gene carrier parent provides genetic information that helps interpret the transmission of the disease and marker loci to the offspring from the affected parent. In autosomal recessive disease, meiotic events (gametes) from both parents can contribute to the LOD score if they are heterozygous gene carriers. In case of association studies the cases must be clearly defined and demarcated from the controls.

II. **Detailed clinical examination of the family members:** An accurate documentation of the various clinical parameters facilitates a clear interpretation of the results obtained.

III. **Marker information**: Large ranges of genetic markers which serve as reference in mapping the disease locus can be employed for genetic analysis. In recent years the dense set of SNPs covering the whole genome seems to offer the most likely choice for gene mapping as we can get information from the whole genome in one stretch and hence the dramatic reduction in cost and time.

IV. **Analysis programs**: The data obtained from mapping markers has to be analyzed for linkage/association. This is done with the help of various linkage analysis programs/ statistical tools for association, which reduces the complexity involved while performing the analysis with multiple markers and using larger number of subjects. The various programs commonly used for linkage are GENEHUNTER, ALLERGO, and MERLIN. To facilitate analysis of large volumes of data generated by GWAS/NGS,

various softwares are used that prunes the data to remove various inconsistencies, such as Mendelian errors, compilation to Hardy Weinberg equilibrium, status of minor allele frequencies etc. Hence deciphering the proper pipeline for analysis of huge volumes of data obtained helps us to identify rare variants along with the common variants and serves as a valuable tool in transferring the knowledge from the scientific experiments to clinical practices.

**QTs studied in the present study**

**Axial length**

Axial length (AL) of the eye is an important refractive parameter which represents the length between the anterior and posterior poles of the eye. It is determined by various parameters like the anterior chamber depth, lens thickness, and posterior chamber depth. AL increases rapidly at infant stage, and then slows down in adult stage and decreases in old age. The shift ranges from 16.8 to 23.6 [26]. This increase with respect to development causes shift towards myopia which is offset by the combined coordination of the lens, cornea and optic cup formation. A 1-mm elongation of AL without these compensations is equivalent to a myopia shift (AL>28mm) of –2 or –2.5 dpt[27], thereby establishing the fact that the components of the visual system are in close interaction with one another during the entire maturation process. Heritability of axial length ranges from 0.89 to 0.94 [28]. Genome wide studies have linked AL to chromosome 2 with a LOD score of 2.64[29]. Another classical genome wide twin study of monozygotic and dizygotic twins reared together was done by Zhu et al which shows the highest LOD score of 3.40 (genomewide P = 0.0004), on chromosome 5q (at 98 centimorgans [cM]). Additional regions with suggestive multipoint LOD scores were also identified on chromosome 6 (LOD scores, 2.13 at 76 cM and 2.05 at 83 cM), chromosome 10 (LOD score, 2.03 at 131 cM), and chromosome 14 (LOD score, 2.84 at 97 cM)[30]. Gwas using myopic cases and controls mapped new loci at 11q24 [31]. The candidate gene studies have included several genes like *PAX6* [32, 33], *MFRP,VSX2, COL1A1* [34, 35]., other metalloproteinases, *CYP1B1*[36], *TGF*β [37]*EGR1*[38]etc. Various other genes do play an important role in maintaining the AL of the eye which include transcription factors like SOX2, MITF, TGFβ, RAX, OTX2 etc, and plenty of other genes in the choroid and retina which are yet to studied. Largest meat analysis with pariticipants from CREAM- Consortium for Refractive Error and Myopia identified eight genome-wide significant loci for AL

(*RSPO1, C3orf26, LAMA2, GJD2, ZNRF3, CD55, MIP,* and*ALPPL2*)[39] and confirmed one previously reported AL locus (*ZC3H11B*)[40].

**Intraocular pressure**

Aqueous humor formed by the ciliary body is drained by the trabecular outflow pathways, which includes the trabecular meshwork, the juxtacanalicular connective tissue, the endothelial lining of Schlemm's canal, and the collecting channels and the aqueous veins [41]. IOP depends on the rate at which the aqueous humor is drained out and in normal cases the rate of production equals the rate of removal. When the rate of removal is less than the production it starts to accumulate leading to increase in IOP. Thus elevated IOP is the major risk factor of glaucoma which affects the retinal ganglion cells leading to axonal degeneration, abnormal optic nerve head appearance and finally blindness. Jac C Charlesworth et al mapped 10q22 as the loci for maximum IOP [42] . Interestingly the same region has been associated with systemic hypertension in Japanese population. This substantiates the studies which show that both are associated. Expression QTl analyses in UK twins has identified a new locus FAM125B - a membrane complex involved in vesicular trafficking process [43]. Another GWAS in Australian cohort and later replicated in UK cohort showed evidence of association with 7p21 near to GLCCI1 and ICA1 [44]. Study by Van Koolwijk et al revealed association with GAS7 and TMCO1 in chromosomes 17p13.1 and 1q24.1 respectively [45]. Recent GWAS conducted in subjects of European ancestry identified significant association with IOP at TMCO1 (rs7518099-G, $p = 8.0 \times 10\text{-}8$) and focused analyses of five loci previously reported for IOP and/or POAG, i.e., TMCO1, CDKN2B-AS1, GAS7, CAV1/CAV2, and SIX1/SIX6, also revealed associations thus confirming the involvement of common variants in multiple genomic regions in regulating IOP and/or glaucoma risk [46].

**Central corneal thickness**

CCT has been associated with genes like *ZNF469* [47], *COL5A, COL8A2* [48, 49], *FOXO1,FNDC3B* [47] and *AKAP13* [50]in both Caucasian and Asian cohorts. GWAS in participants of Gutenberg Health and Rotterdam Study groups showed association with *COL5A1* and *ZNF469 [51]*. Quantitative analysis of CCT and a subsequent analysis of POAG, revealed SNPs in two cell adhesion molecules, NTM and CNTNAP4, that may increase POAG susceptibility in a subset of cases [52]. Analysis genes controlling CCT in multi ethinic Indians show strong evidence of association at four novel loci: *IBTK* on

chromosome 6q14.1; *CHSY1* on chromosome 15q26.3; and intergenic regions on chromosomes 7q11.2 and 9p23 [50]. Study in Caucasians identified a missense mutation in *COL8A1* associated with thin CCT [53]. Four pedigrees with early-onset glaucoma phenotypes secondary to segmental chromosomal duplications or deletions encompassing FOXC1 and 18 individuals from 9 FOXC2 mutation pedigrees was associated with increased CCT [54].

**Optic disc parameters- (Vertical cup to disc ratio, optic disc area)**

The major determinants of optic nerve head are size of the optic disc and vertical cup to disc ratio. Genome wide study for these two parameters was conducted in Rotterdam Study population which yielded significant loci for optic disc area on chromosome 1p22 (CDC7), and 10q21.3-q22.1 (ATOH7) and for VCDR on chromosome 9p21 (CDKN2B) and 14q22.3-q23 (SIX1) [55].Several studies followed this which includes study in 4445 Singaporean individuals and in Australian twin cohort which also showed evidence of association of optic disc area to ATOH7, and CDC7 along with other novel genes CARD10 (Asian cohort) and RFTN1 (Australian cohort) [56, 57]. Association studies in US Caucasian population associated *CDKNB2*, *SIX1/SIX6*, and *ATOH7* genes with VCDR suggesting that increased optic disc area can significantly contribute to POAG risk when coupled with risk factors controlling VCDR [58]. Three polymorphisms rs1900004 (within 10kb of *ATOH7*) [56, 57], rs10483727 (near *SIX1-SIX6*) and rs1063192 (in 3'UTR region of *CDKN2B*) [55] have been strongly associated with optic disc parameters such as ODA, optic cup area and VCDR in various populations [59]. The other genes include *CDC7*/*TGFBR3* (rs1192415) and *CARD10* (rs9607469) which have been associated with ODA in Asian cohort [56].

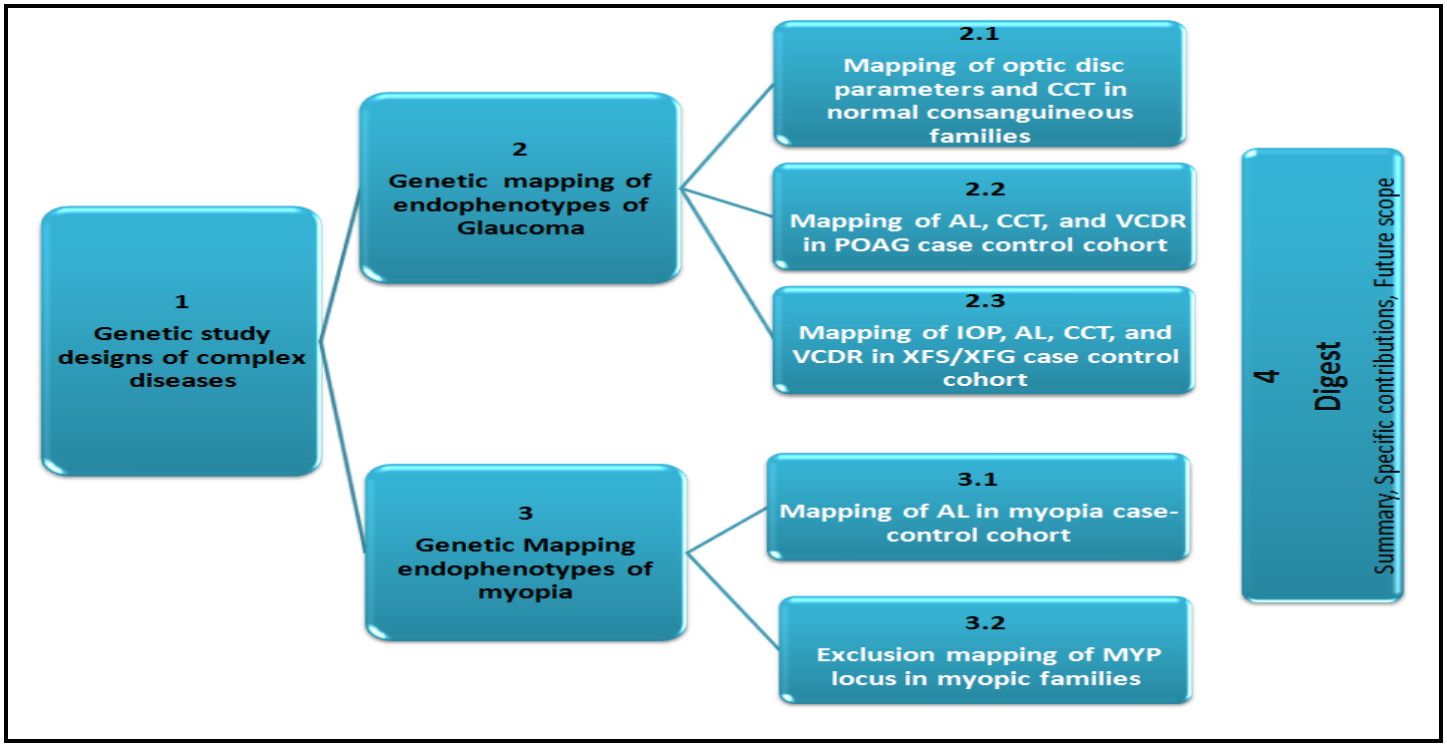
**Need for the current study**

In India, research insights pertaining to genetics of myopia and glaucoma have been minimal and are based on candidate gene screening, familial linkage analyses etc. Worldwide mapping genes for these complex disorders have advanced using the QT approach and have identified the loci as well as few genes which have been proven functionally to be involved in determination of QTs. This approach was taken as an initiative to map loci or evaluate the association of genes with myopia and glaucoma by QT based approach.

**Organization of the thesis**:

This thesis documents QT based gene mapping details for complex eye disorders such as myopia and glaucoma. Chapters two and three are introductory to these two disorders followed by approaches like candidate gene, linkage, association using tagged SNPs etc (Fig1.7)

**Fig 1.7: Flowchart depicting the organization of the thesis**



## Chapter 2
## Genetic mapping of endophenotypes of glaucoma

### Introduction

Glaucoma is one of the major causes of irreversible blindness; ranking as the second leading cause in the world. A recent estimate shows that, globally, about 67 million people are affected with glaucoma; among them 3.1 million are blind [60]. In India about 12 million are affected with this disease contributing to about 16.7% of the total blindness in the country [61]. Most patients experience a painless progressive loss of peripheral vision and stay undiagnosed until the end stage where there is complete loss of central vision.

### Classification of glaucoma

Glaucoma can be classified based on

i) **Etiology**:

- Primary: The common form seen in population, caused by retention of aqueous fluid due to defective drainage system that raises the IOP (>21mmHg).
- Secondary: Occurs as secondary complications due to use of drugs like corticosteroid, or eye disease like Pseudoexfoliation, uveitis or systemic diseases (eg: Sarcoidosis).

ii) **Angle configuration**:

Open Angle glaucoma (POAG): This is the most common form of glaucoma which occurs either due to raised (High tension, >21 mm Hg) or normal (Normal tension< 21 mm Hg) IOP. This type of glaucoma presents with optic nerve damage and narrowed side vision.

- Closed Angle glaucoma (PACG): This type is caused due to synechiae formed by contact between iris and trabecular meshwork (TM). The synechiae hinders the proper outflow of aqueous humor leading to increased IOP. The disease manifests suddenly with severe pain and redness of the eye.

iii) **Age of onset**:

- Congenital: Characterised by onset from birth, and results as a defect in the improper formation of drainage channels for the aqueous humor.
- Late onset: Occurs at later decades (>40 yrs), which mostly involves abnormalities in anterior chamber angle formation.

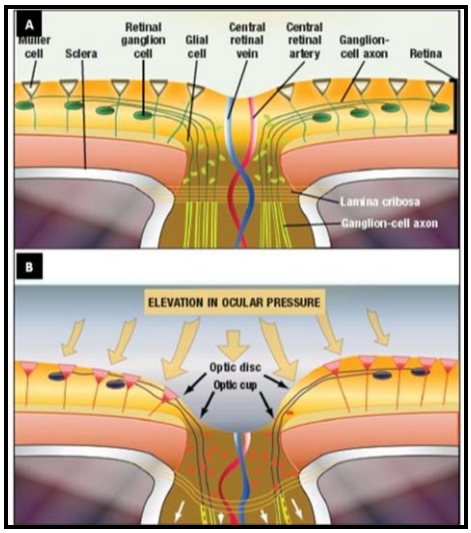
The major risk factors of glaucoma include elevated IOP, positive family history, ethnicity, age, usage of drugs like corticosteroids etc. It has been observed that about 60% of the patients have positive family history and the risk increases about 10 fold for the first-degree relatives of affected individuals [62, 63]. People of African ancestry have increased

susceptibility for POAG (about 6% in individuals of age >40yrs) when compared with the Caucasians (1%); while PACG is more prevalent in Asians [64, 65]. Pseudoexfoliation glaucoma (XFG) is the most common secondary open angle glaucoma, its prevalence in India ranges from 0.69 to 6% [66-69]when compared to 0.2 to 30% [70, 71] in other populations . In the current study we have included 2 types of glaucoma: POAG and XFG to study the major QTs for glaucoma.

**Primary open angle glaucoma (POAG)**:

The disease occurs as a consequence of abnormal outflow of aqueous humor through the trabecular meshwork (TM), a connective tissue, lining the outflow pathway at the iridocornel angle of eye's anterior chamber thereby leading to elevated IOP. The pathogenic event that follows includes loss of retinal ganglion cell axons along with supporting glia and vasculature. This event leads to thinning of the neuroretinal rim of optic nerve thereby widening the optic-nerve cup (optic–nerve cupping). In NTG, the intracranial pressure is decreased whereas the IOP is within normal range; but, affects the net translaminar pressure and hence damages the optic nerve head (ONH) (Fig: 2.1).

**Fig 2.1: Diagrammatic representation of formation of optic nerve cupping in POAG [72]**




A. Normal structure of the optic nerve head and retina. B. Glaucomatous condition due to increased IOP causing damage to the lamina cribosa and formation of cup.

Studies have elucidated the role of both genetic and environmental factors which act together to precipitate this disorder. The complex nature of the disease makes it difficult to decipher the pathophysiology of the disorder [73]. It has been estimated that the risk of

development of POAG is about 22% for a relative of a patient when compared to a relative of a normal person (2-3%) [63].

In the pursuit of finding the contributing genetic variants for POAG different study models have been used. Linkage and association studies has linked 33 chromosomal loci for the disease [74], of which 15 have been designated as GLC1A to GLC1O ('GLC' – glaucoma) by the HUGO genome nomenclature committee (www.gene.ucl.ac.uk/nomenclature) [75]. Mutations in myocilin (*MYOC*- GLC1A) [76], optineurin (*OPTN* -GLC1E) [77], *WDR36*- GLC1G [78], neurotrophin-4 (*NTF4*) [79] and *CYP1B1* [80] genes has been reported across various populations. However mutations in these genes explain only 5% of POAG pathology and are unable to explain the heritability component in the disease. Alternative approaches focuses on the heritable and measurable clinical risk factors (QTs) of POAG. This method is advantageous when compared to the classical methods where the disease as a whole is considered as cases and thus overlooks the individual effect of the contributing clinical risk factors thus affecting the stratification of cases.

***Quantitative traits in POAG***

QTs, also called as endophenotypes, are measurable clinical parameters that are represented as the product of polygenic effects and act as risk predictor of the disease. The major QTs of POAG include IOP, CCT, VCDR, etc. And their heritable estimates are 0.94, 0.68, [81] and 0.48-0.8 [82] respectively. The non-genetic factors (age, fasting glucose etc) account for only a small proportion (0.13) of variance, in all three traits [82].

In POAG, GWAS has identified significant associations near genes like *TMCO1*, *CDKN2B-AS1*, *GAS7*, *CAV1/CAV2*, and *SIX1/SIX6*, *ATOH7*, and *CARD10* with its endophenotypes such as IOP, CCT, VCDR, etc [46, 83]. SNPs in *CDKNB2*, *SIX1/SIX6*, and *ATOH7* genes were not associated with POAG when analysed as a disease trait, but when analysed for its QTs they were significantly associated with VCDR [58]. Similarly another study from Japanese population showed association of *CDKN2B* (rs1063192) and *ATOH7* (rs1900004) as non–IOP-related genetic risk factors for NTG and rs1547014 (*CHEK2*) as genetic risk factor for HTG [59]. Collagen related genes are also associated with all types of glaucoma. Studies on mice have shown that collagen VIII forms an important component of the Descemet's membrane and lack of activation of these genes (*COL8A1* and *COL8A2)* leads to thinning of the corneal layer and eventually glaucoma.

Work by Desronvi et al in Caucasian population identified a missense mutation in *COL8A2* gene in POAG patients [53].

The optic nerve head (ONH) remains as the main target of damage in glaucoma. The major clinical determinants of ONH are optic disc area (ODA) and VCDR. Genome wide study for these two parameters yielded significant LOD score to chr 1p22 (*CDC7*), 10q21.3-q22.1 (*ATOH7*) and to chr 9p21 (*CDKN2B*) and 14q22.3-q23 (*SIX1*) for ODA and VCDR respectively [55]. Later several studies replicated this significant association along with other novel genes *CARD10* (Asian cohort) and *RFTN1* (Australian cohort) [56, 57].

In India, low prevalence of mutations in genes like Myocilin [84, 85], *CYP1B1*, *OPTN* [86, 87] and *NTF4* [88] have been implicated for POAG. The Chennai glaucoma study, aimed at understanding the molecular genetics of the disease prevalence in homogenous cohort has shown 2% of mutations in these genes. The importance for analysis of other measurable and heritable endophenotypes has thus become an important factor in understanding of the genetics of glaucoma. This lead to a collaborative project between INDIA and United States of America in which analysis of QTs in normal consanguineous families was one of the main objectives.

***ATOH7*- atonal homolog of drosophila gene (chr 10q21.3):** ATOH is the atonal homolog of drosophila; named for its chordotonal stretch receptor mutant phenotype [89]. It is a basic proneural helix loop helix (bHLH) transcription factor expressed in the early retinal progenitors. This has been shown to be essential for the initiation of neurogenesis and production of retinal ganglion cells (RGCs) [90]. Defect in the *ATOH7* gene affects the RGC number, fate and differentiation of the other retinal cells [91, 92]. GWAS studies in European, Australian, Caucasians and Indian population have shown association of rs1900004, rs3858145, rs17231602 and rs4746741 with ODA [55-57]. Kamron et al have documented that homozygous mutations (p.E49V and p.P18RfsX69) in this gene play a role in the development of both the anterior and posterior chamber of the eye and its importance in retinal vascular development and hyaloids regression [93].

***SIX1/SIX6* gene (chr 14q23)*:*** These homeoprotein members belongs to SIX/ *sine oculis* family of homeobox transcription factors. The *SIX1*/*SIX6* gene complex has been significantly associated with POAG in several GWAS. It was observed that *SIX6* missense variants deregulated its expression and affected its normal function in controlling the retinal progenitor cell proliferation, during eye development; hence influencing

susceptibility to POAG [94]. Functionally this protein is involved in the regulation of retinal progenitor cell proliferation during eye development. Animal models have shown retinal hypoplasia, reduction in eye size and reduction in eye volume of the optic nerve [95-97].

***CDC7/TGFBR3*** **gene (chr 1p22):** Transforming growth factor (TGF) is a multifunctional cytokine which modulates the development of many tissue and repair processes by binding to its receptor TGFBR3 [98]. This gene has been implicated in POAG development by its association with VCDR (rs1063192) in GWAS. Functionally it is known to interact with *CDKN2A*, another gene known to be associated with VCDR.

***CDNK2B*** **gene (chr 9p21):** This gene is a member of the family of cyclin-dependent kinase (CDK) inhibitors which plays a role in cell cycle regulation, influencing the proliferation/differentiation balance. GWAS has shown association to VCDR in various populations [58, 98, 99]. The protein is involved in pathogenesis of optic nerve degeneration and POAG.

***CARD10*** **gene (chr22q 13.1):** This protein is involved in the regulation of caspase activation and apoptosis. It is activates NF-kappaB, a well characterized transcription factor involved in signalling pathway [100]. Variations in this gene have been proposed to influence the ODA. Association between *CARD10* and ODA [56] has been reported in Asian and Singaporean cohort.

**Zinc finger nuclease (*ZNF469*) gene (chr 16q24):** The protein functions either as a nuclear transcription factor or as extra-nuclear regulatory molecule which is involved in synthesis, organisation and maintenance of collagen fibres [101]. Mutations in this gene have been associated with Brittle corneal syndrome. SNPs near/in *ZNF469*, (MIM 612078) and fork head transcription factor (*FOXO1*, MIM 136533) [102] have been associated with CCT, an important endophenotype of POAG in European and Asian cohort. This observation supports strongly that *ZNF469* is a QTL for CCT [48]. CCT acts as an independent risk factor for POAG apart from other factors; hence it has been proposed as an intermediate trait or endophenotype for POAG [103]. These purported functions makes this gene an ideal candidate to be studied in reference to its influence on CCT and hence glaucoma.

**Pseudoexfoliative glaucoma (XFG)**:

Exfoliation syndrome (XFS) is a systemic disease and one of the most commonly identified cause of OAG [104]. The resulting glaucoma is termed as 'Exfoliative glaucoma' (XFG) which is identified by abnormal deposition of microscopic fibrils in ocular tissues (TM) (Fig: 2.2) [105]. These PEX aggregates decreases the vascular contractivity and ocular blood flow (Fig: 2.3). Though the exact pathophysiology is not known, several researchers have attributed the cause as excessive production and abnormal cross-linking of the elastic fibres in the extracellular matrix [106].

**Fig 2.2 Photographic image of: A) Normal pupil B) Deposits of pseudoexfoliative material around the pupil edge.**

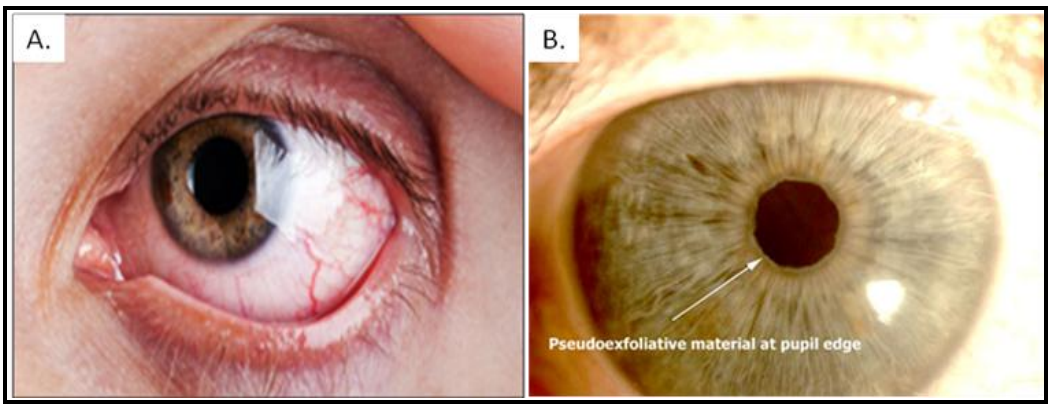
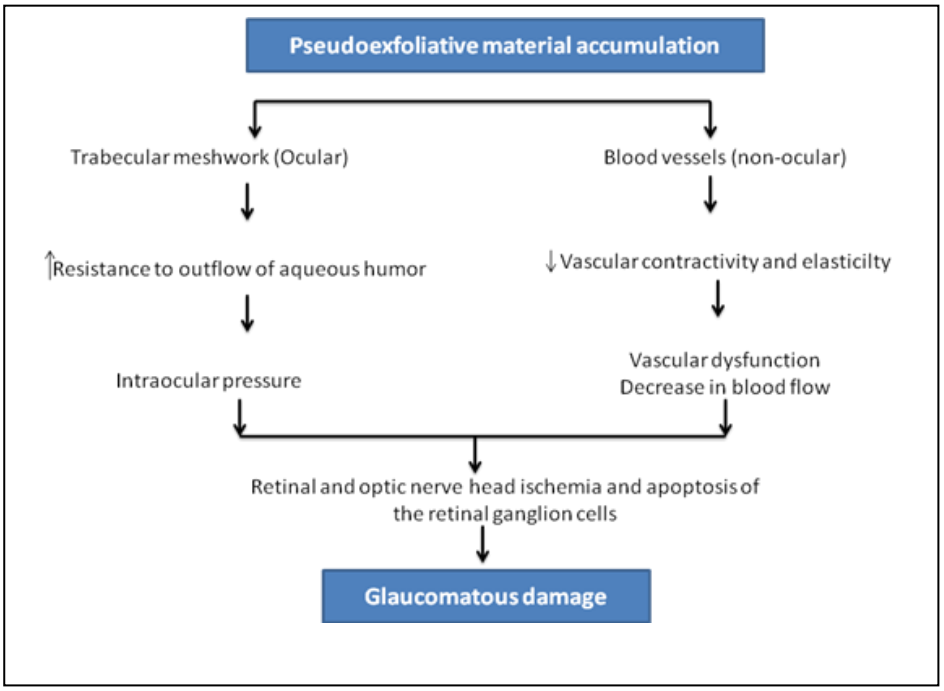



**Fig 2.3 Pathway depicting the development of glaucoma from the PEX deposits [107]**



**Prevalence and genetics of XFG**:

The prevalence of XFS/XFG varies from 0.2%-30% in various populations [70]. The prevalence of this disease varies in different places of India. The Andhra Pradesh Eye disease study revealed a prevalence rate of 5.5% [66], while in Arvind Comprehensive eye survey it was as high as 7.5% [67]. Other studies from the south and central India showed a prevalence rate of 13% [68] and 4.0±1.5 [69] respectively.

Familial studies have proven the role of genetics in XFG, however the inheritance pattern is unclear which may be due to late onset/incomplete penetrance. Several studies have associated SNPs in *lysyl oxidase-like 1*(*LOX1*) gene with XFS/XFG following the initial identification of *LOXL1* SNPs rs2165241, rs1048661, rs3825942 in PEX in Iceland, Sweden [108], US Caucasian [109], Australia [110], India [111], Italy, Germany [112] , Europea [113], China [114], Japan [115], south Africa, and Saudi-Arabia [116] (Table: 2.1)

**Table 2.1: Summary of the genetic association of rs1048661, rs3825942 in exon 1 of *LOXL1* gene with XFS/XFG [70]**

| Studied population | rs1048661 'G' allele | | Significant association | rs3825942 'G' allele | | Significant association | References |
|---|---|---|---|---|---|---|---|
| | Case | Control | | Case | Control | | |
| Icelandic | 0.781 | 0.651 | Yes | 0.984 | 0.847 | Yes | Thorleifsson et al. (2007) |
| Swedish | 0.834 | 0.682 | Yes | 0.995 | 0.879 | Yes | Thorleifsson et al. (2007) |
| American | 0.819 | 0.600 | Yes | 0.986 | 0.880 | Yes | Fingert et al. (2007) |
| American | 0.787 | 0.665 | Yes | 0.939 | 0.844 | Yes | Challa et al. (2008) |
| American | NA | NA | NA | 1.000 | 0.856 | Yes | Yang et al. (2008) |
| American | 0.843 | 0.703 | Yes | 0.959 | 0.798 | Yes | Aragon-Martin et al. (2008) |
| American | 0.829 | 0.719 | No | 0.988 | 0.795 | Yes | Fan et al. (2008) |
| Australian | 0.78 | 0.660 | Yes | 0.95 | 0.84 | Yes | Hewitt et al. (2008) |
| Austrian | 0.841 | 0.671 | Yes | 0.994 | 0.817 | Yes | Mossbock et al. (2008) |
| Germany | 0.844 | 0.660 | Yes | 0.992 | 0.856 | Yes | Wolf et al. (2010) |
| Germany | 0.818 | 0.644 | Yes | 0.951 | 0.857 | Yes | Pasutto et al. (2008) |
| Finnish | 0.825 | 0.683 | Yes | 0.968 | 0.823 | Yes | Lemmela et al. (2009) |
| Italian | 0.825 | 0.693 | Yes | 1.000 | 0.821 | Yes | Pasutto et al. (2008) |
| Saudi Arabian | 0.876 | 0.762 | Yes | 0.968 | 0.817 | Yes | Abu-Amero et al. (2010) |
| Indian | 0.721 | 0.634 | No | 0.923 | 0.742 | Yes | Ramprasad et al. (2008) |
| Chinese | 0.542 | 0.444 | No | 0.992 | 0.918 | Yes | Lee et al. (2009) |
| Chinese | 0.110 | 0.480 | Yes | 1.000 | 0.900 | Yes | Chen et al. (2009) |
| Japanese | 0.036 | 0.493 | Yes | 1.000 | 0.877 | Yes | Fuse et al. (2008) |
| Japanese | 0.008 | 0.460 | Yes | 1.000 | 0.857 | Yes | Hayashi et al. (2008) |
| Japanese | 0.006 | 0.450 | Yes | 0.994 | 0.853 | Yes | Mabuchi et al. (2008) |
| Japanese | 0.005 | 0.474 | Yes | 0.995 | 0.850 | Yes | Mori et al. (2008) |
| Japanese | 0.005 | 0.497 | Yes | 0.986 | 0.863 | Yes | Ozaki et al. (2008) |
| Japanese | 0.005 | 0.554 | Yes | 0.993 | 0.806 | Yes | Tanito et al. (2008) |
| South African | 0.990 | 0.810 | Yes | 0.130 | 0.620 | Yes | Williams et al. (2010) |
| South African | 1.000 | 0.883 | Yes | 0.140 | 0.617 | Yes | Rautenbach et al. (2011) |

NA: not available; *LOXL1*: lysyl oxidase-like 1; XFS: exfoliation syndrome; XFG: exfoliation glaucoma.

***LOXL1* (chr 15q24.1)**

*LOXL1* gene codes for lysyl oxidase, an enzyme that catalyzes the deamination of lysine residues of tropo-elastin and essential for synthesis and maintenance of elastic fibres. Expression studies have localized the protein to extra cellular matrix region and also in

cells undergoing epithelial-mesenchymal transition, malignant lesions of the breast and lung, thus emphasising its role in connective tissue remodelling during dynamic processes such as tissue injury, fibrosis, cancer, and development [117, 118]. In the eye the expression is at its maximum in the iris; which is the one of the first components to be involved in XFG/XFS, followed by other regions such as cornea, lens, ciliary body, lamina cribosa and optic nerve [110, 119, 120].

At the molecular level *LOXL1* functions by polymerising tropo-elastin monomers into growing elastic polymers [121], which activates the C-terminal domain, a catalytic domain responsible for deaminating the lysine residues of tropo-elastin and allowing it to join the growing elastic polymer [122]. Immunohistochemistry studies have shown the presence of *LOXL1* and its substrate tropo-elastin in the PEX deposits.

Two nonsynonymous coding SNPs rs1048661 and rs3825942 have been associated with XFS/XFG in various populations and association status of these alleles varies in different population. For example in Chinese population they have shown to be of risk [114] whereas in a recent meta analysis including the Caucasians and Asians [123]. Recent study in Greece population as the first to study the association of these SNPs with an endophenotype IOP, but did not find any association [124].

Studies on the expression of *LOXL1* and its effect on XFG/XFS have showed varied results [125, 126]. Few studies have functionally associated the risk allele of rs1048661 with reduced *LOXL1* expression [127]. In contrast, the risk allele of rs3825942, had no effect on expression levels, but is suggested to have functional consequences on enzyme activation and/or substrate binding [108]. *LOXL1* co-localizes with clusterin, a highly efficient extracellular chaperone, that recognizes misfolded and abnormally aggregated proteins in the extracellular space, and which has been previously shown to be involved in the accumulation of PEX material in anterior segment tissues of XF eyes [128]. Studies have shown that *LOXL1*-knockout mice exhibit multiple abnormalities like lax skin, diverticula, enlarged air spaces and prolapse of the rectum and pelvis [129]. The female mice have normal tissue initially but no new elastic tissues are formed and deposited after pregnancy and birth; thus specifying that this protein is necessary for targeted renewal of elastic tissue.

**Organisation of the chapters**: It focuses on the association of endophenotypes of POAG with SNPs rs1900004 and rs3858145 (*ATOH7*), rs10483727 (*SIX1/SIX6*), rs1192415 (*CDC7*), rs1063192 (*CDNK2B*) and rs9607469 (*CARD10*) in Indian population.

The current study was executed with the main aim of analysing the genetic association of endophenotypes of glaucoma in two different models such as normal consanguineous families and disease (case-control) models. Consanguineous pedigree model was used based on the hypothesis that inbred families facilitate the gene identification and overcome the limitation of large sample size; thereby improves the power of the study. Disease models; POAG and XFG/XFS were also used to evaluate the association of SNPs in candidate gene for ocular QTs.

## Chapter 2.1

## Analysis of association between optic disc parameters and SNPs near *ATOH7* and *ZNF469* in normal consanguineous families

### Objectives

1. To identify and recruit consanguineous families from south Indian cohort
2. To perform genotyping of SNPs rs1900004, rs3858145, rs17231602, rs4746741 near *ATOH7* and rs77330683 near *ZNF469* genes.
3. To perform statistical analysis for association for the SNPs with the optic disc parameters like ODA, VCDR, and cup area.

### Methodology

**Sample recruitment and clinical examination:** (Appendice 1) Analysis was done based on the global measurements of the right eye of individuals

### Laboratory methods

1. **Blood collection**, **DNA extraction- Nucleospin technique, and DNA Quantification** (Appendices- 2 to 4)
2. **PCR Amplification**: Amplification of the genomic region flanking the SNPs rs4746741, rs1900004, rs3858145, rs17231602 at a genetic distance of 69.65 -69.68 cM on 10q21 was performed using primers as described earlier [57]; and rs77330683 near *ZNF469* gene was performed using primers designed with Primer 3 software (table 2.1.1) [130] after checking for self complimentary and specificity using Oligocalc [131]. Each reaction mixture was prepared for PCR with corresponding annealing temperatures as given in table 2.1.1(Appendice-5and 6) and subjected for direct sequencing in *ABI Prism 3100 AVAN*T genetic analyzer (Applied Biosystems, California) (Appendice-8 and 9).

**Table 2.1.1: Primers flanking the SNPs rs4746741, rs1900004, rs3858145, rs17231602 and rs77330683**

| **SNP** | **Gene** | **Primers 5'-3'** | **Amplified product length (bp)** | **Annealing temperature (**˚C) |
|---|---|---|---|---|
| rs1900004 | *ATOH7* | FP: ACCAAATGCTCACCCAGTGT<br>RP: GCCTGCAGATTGCTAGATCC | 367 | 60° |

| | | | | |
|---|---|---|---|---|
| rs3858145 | *ATOH7* | FP:TGCAAAGCAAAGTGTTTACCT<br>RP:GCCACCAACAGTCCTCACTT | 398 | 61° |
| rs17231602 | *ATOH7* | FP: GCACAGCTGTCCTTTTGTCC<br>RP: TGCTATTTTTGGCATTGCAT | 399 | 61° |
| rs4746741 | *ATOH7* | FP:TCATGTTGGTCAGGCTGTTC<br>RP:GCAGGACTCTAAGTTATCCATG | 359 | 61° |
| rs77330683 | ZNF469 | FP: GGCTCCGTCTTTAAGCACAA<br>RP: GGAATGGTGTTCGTCCTGTC | 154 | 65˚C to 60˚C |

## I. Data Analysis

All the analysis was based on the global measurements of the right eye of individuals in four South Indian families. The effect size (beta) and *P* value are obtained using the QFAM procedure in PLINK, which performs a simple linear regression of phenotype on genotype, and then uses a special permutation procedure to correct for family structure. The input data included .ped and .map files. The .ped file contains the following data in columns: family id, individual id, father id, mother id, gender, disease status, and the genotypes of the 5 SNPs. Map file includes chr id, db SNP id, genetic distance (morgans), Base-pair position (bp units). The following commands were given to execute the QFAM analysis:

1. To open the file containing ped and map

   D:\Plink>plink --file family

2. To create binary files

   D:\Plink>plink --file family --make-bed --out family

3. To perform QFAM

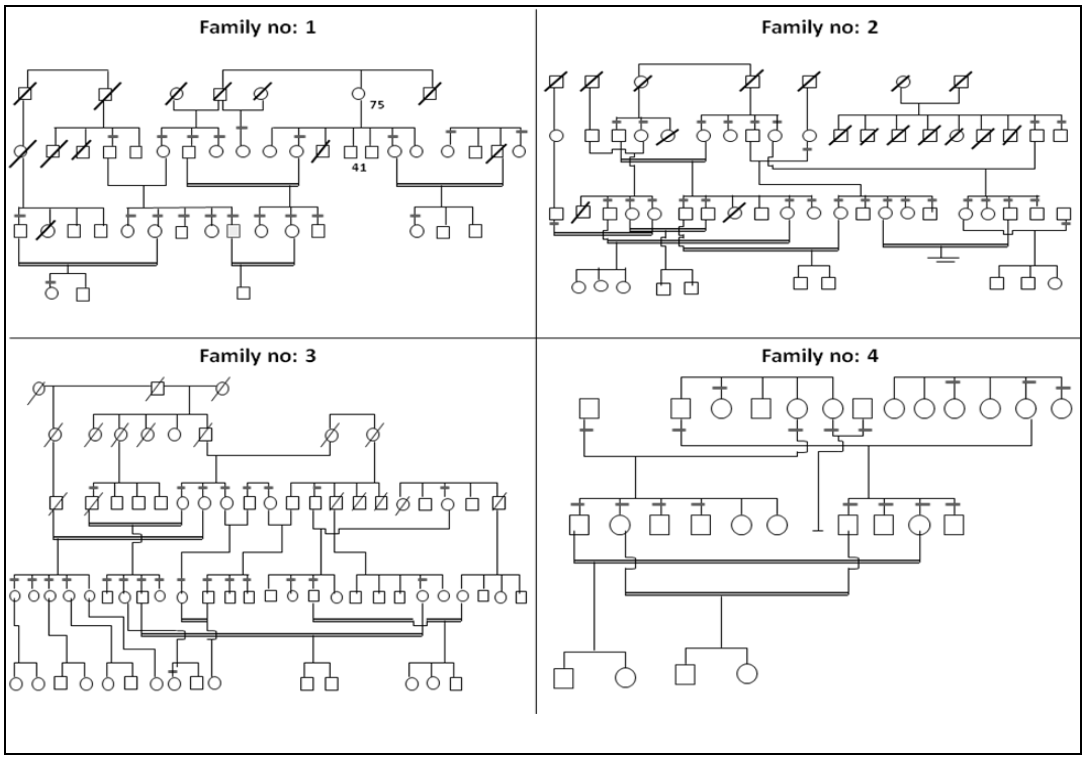
   D:\Plink>plink --bfile family --qfam-total --mperm 1,00,000

**Results:**

Four large normal consanguineous families were identified (exhibiting First and second degrees of consanguinity) and recruited from the Chennai Glaucoma study [132] (Fig

2.1.1) and 83 individuals were recruited from these 4 families after complete ophthalmic examination. The power for this study was extrapolated based on the earlier observations using SOLAR software. The number of pedigrees to give a statistically significant power of >80% depended on the heritability estimates for a given trait. Sample size of 4-9 pedigrees resulted in a statistically significant power to detect association for QTs with heritability of 0.3 and above respectively [133]. The QT traits measured in the current study that included optic disc area, cup area, cup to disc ratio and CCT have an heritability estimate of 0.4-0.9; all followed normal distribution.

**Fig 2.1.1: Pedigrees of the 4 families included in the study [The individuals who participated in the study are marked with a bar on top]**



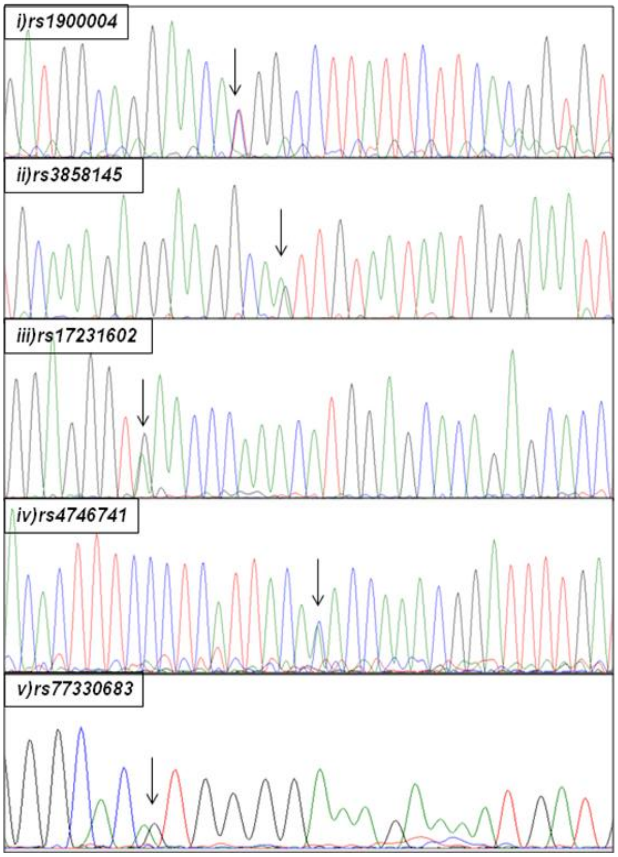

The frequency distribution and electrophorogram of the genotypes is given in table 2.1.2 and fig 2.1.2

**Table 2.1.2: Genotype distribution for the SNPs rs1900004, rs3858145, rs17231602, rs4746741 and rs77330683.**

| Family no | rs1900004 | | | rs3858145 | | | rs17231602 | | | rs4746741 | | | rs77330683 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CC | CT | TT | GG | GA | AA | GG | GA | AA | CC | CA | AA | TT | CT | CC |
| 1 (n=20) | 7 | 13 | 0 | 1 | 13 | 6 | 18 | 2 | 0 | 0 | 9 | 11 | 15 | 5 | 0 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 (n=27) | 15 | 10 | 2 | 2 | 16 | 9 | 9 | 16 | 2 | 0 | 0 | 27 | 19 | 6 | 2 |
| 3 (n=19) | 5 | 11 | 3 | 3 | 12 | 4 | 13 | 6 | 0 | 2 | 5 | 12 | 18 | 1 | 0 |
| 4 (n=17) | 13 | 4 | 0 | 0 | 4 | 13 | 17 | 0 | 0 | 0 | 3 | 14 | 16 | 0 | 1 |

**Fig 2.1.2: Electrophoretogram of *ATOH7* and *ZNF469* genomic sequence showing the variations in i) rs1900004, ii) rs3858145, iii) rs17231602, iv) rs4746741, and v)rs77330683**




## Statistical analysis

QFAM analysis showed significant association for rs4746741, rs1900004 and rs3858145 with cup disc area (*P* = 0.025, 0.012, 0.016 respectively; table 2.1.3 (I)), and rs4746741 with cup area (*P* = 0.03; table 2.1.3 (II)). We did not find any significant association for any of the 5 SNPs with cup-to-disc ratio and CCT.

**Table 2.1.3: Association analysis of *ATOH7* and *ZNF469* in Indian consanguineous families with**

**i) Disc area**

| SNP | Chr | bp (build 36.3) | Nearest gene | Minor allele | Beta | *P* (1,00,000 permutations) |
|---|---|---|---|---|---|---|
| rs4746741 | 10 | 69659516 | *ATOH7* | C | -0.51 | 0.025* |

| | | | | | | |
|---|---|---|---|---|---|---|
| rs1900004 | 10 | 69670887 | *ATOH7* | T | -0.41 | 0.012* |
| rs3858145 | 10 | 69681844 | *ATOH7* | G | -0.39 | 0.016* |
| rs17231602 | 10 | 69682295 | *ATOH7* | A | 0.13 | 0.51 |
| rs77330683 | 16 | 88485006 | *ZNF469* | C | -0.08 | 0.79 |

*** p<0.05, Beta: this value represents the regression coefficient, which represents the direction of effect (similarly for the other tables)**

**ii) Cup area**

| SNP | Chr | Bp (build 36.3) | Nearest gene | Minor allele | Beta | P(1,00,000 permutations) |
|---|---|---|---|---|---|---|
| rs4746741 | 10 | 69659516 | *ATOH7* | C | -0.19 | 0.03* |
| rs1900004 | 10 | 69670887 | *ATOH7* | T | -0.07 | 0.36 |
| rs3858145 | 10 | 69681844 | *ATOH7* | G | -0.09 | 0.17 |
| rs17231602 | 10 | 69682295 | *ATOH7* | A | 0.12 | 0.17 |
| rs77330683 | 16 | 88485006 | *ZNF469* | C | -0.08 | 0.52 |

**iii) Cup/disc ratio**

| SNP | Chr | Bp (build 36.3) | Nearest gene | Minor allele | Beta | P(1,00,000 permutations) |
|---|---|---|---|---|---|---|
| rs4746741 | 10 | 69659516 | *ATOH7* | C | -0.01 | 0.72 |
| rs1900004 | 10 | 69670887 | *ATOH7* | T | 0.03 | 0.35 |
| rs3858145 | 10 | 69681844 | *ATOH7* | G | 0.008 | 0.82 |
| rs17231602 | 10 | 69682295 | *ATOH7* | A | 0.01 | 0.64 |
| rs77330683 | 16 | 88485006 | *ZNF469* | C | -0.06 | 0.25 |

**iv) CCT**

| SNP | Chr | Bp (build 36.3) | Nearest gene | Minor allele | Beta | P(1,00,000 permutations) |
|---|---|---|---|---|---|---|
| rs4746741 | 10 | 69659516 | *ATOH7* | C | 22.19 | 0.22 |
| rs1900004 | 10 | 69670887 | *ATOH7* | T | -0.03 | 0.99 |
| rs3858145 | 10 | 69681844 | *ATOH7* | G | -19 | 0.21 |
| rs17231602 | 10 | 69682295 | *ATOH7* | A | -24.99 | 0.19 |
| rs77330683 | 16 | 88485006 | *ZNF469* | C | 8.82 | 0.22 |

## Discussion:

The molecular events in multifactorial disorders are due to complex interaction of multiple genetic factors thus a greatest challenge for mapping the genes for these diseases. While family based linkage studies are successful in identifying genes of large effect in Mendelian diseases (eg: cystic fibrosis and Huntington's disease) the complex disorders

always suffered the limitation of uninformative pedigree size. On the other hand the effect of association based studies have been able to identify several genes involved in various pathways for common diseases such as type 2 diabetes [134], cancers etc wherein their effect are considerably small, and almost always contribute to a less than 2-fold increase in risk; however the main prerequisite to achieve such statistical significance is large sample size [135].

The present study is performed with an aim to test the hypothesis that large consanguineous families could help overcome the limitations of sample size. We have used this novel approach for mapping the reported loci for optic disc parameters in 4 consanguineous pedigrees with 83 individuals. QFAM analysis revealed significant association for rs4746741 (C>A), rs1900004 (C>T) and rs3858145 (A>G), with disc area; rs4746741 with cup area.

GWAS results have shown significant association for SNPs in *ATOH7* with ODA and *ZNF469* with CCT among different populations. In Rotterdam Study I and II (population of 7360 subjects) a significant association ($P = 2.67\times10^{-33}$) for SNP rs1900004 (within 10 kb of the *ATOH7* gene) was observed. Similarly, a replication study in Australian, Dutch and Netherland cohort showed significant association of rs3858145 in *ATOH7* gene with mean disc area [55, 57]. In the present study we have replicated the association of rs1900004 with decreased disc area while rs3858145 showed an association in the opposite direction (decreased) when compared to the previous studies (increased disc area). However the association of SNP rs77330683 in *ZNF469* with CCT as observed across different population is not replicated in this study.

Using a cohort of only 83 individuals we were able to replicate the association with disc area thereby reducing the time and the cost involved in mapping genetic loci for complex disorders. QTL mapping studies in consanguineous families rather than normal families adds an additional advantage of sharing of regions of homozygosity (increases up to 6% in the offspring of first cousin marriages) thereby reducing the genetic complexity involved [136]. Contemporary attention on consanguineous families continues to be largely focused on the expression and identification of rare autosomal recessive alleles which are the major causative agents in complex disorders. The recent example includes a GWAS conducted in behavioural disorders which indicated the potential contribution of thousands of alleles of very small effect in schizophrenia and bipolar disorder, with significant genetic overlap

between the two disease states [137, 138]. Homozygosity mapping in autism [139] and a case-control study of bipolar disorder type 1 in consanguineous progeny [140] both implicated the causal expression of rare recessive genes as well.

This is the first report by a family based approach replicating the strong association of *ATOH7* gene in Indian population for rs4746741, rs1900004 and rs3858145 with cup disc area (*p* value = 0.02, 0.013, 0.015) and rs4746741 with cup area (*p* value = 0.026). We did not find any significant association between the 5 SNPs and QTs cup disc ratio and CCT. Pertaining to the limitations, this strategy of using consanguineous families to obtain good power is possible only in the presence of multiple consanguineous marriages in the family. Inability to obtain significant association with CCT and cup to disc ratio in the current study can be attributed to the usage of limited number of SNPs. To conclude, we were able to prove our hypothesis that large consanguineous pedigrees provide sufficient power to detect genetic variants responsible for QTs.

## Chapter 2.2

## Genetic association study for the endophenotypes of glaucoma with SNPs near *ATOH7, CDC7, CDKN2B, CARD10* and *SIX1* in POAG

**Objective:**

1. To identify and recruit POAG patients and controls
2. To perform genotyping of rs1900004, rs3858145, rs10483727, rs1192415, rs1063192 and rs9607469
3. To perform the analysis for association of the SNPs with the disease and its endophenotypes including VCDR, CCT, AL and ACD.

**Methodology:**

**I. Sample recruitment and clinical examination:** (Appendice 1)

**II. Laboratory methods**

1. **Blood collection**, **DNA extraction- Nucleospin technique, and Quantification of the DNA extracted** (Appendices- 2 to 4)
2. **SNP Genotyping**: SNPs rs1900004 and rs3858145 (about 10kb at the 5' upstream of *ATOH7* on 10q21.3-22.1), rs10483727 (within 40kb of *SIX1/SIX6* gene complex on 14q22-23), rs1192415 (within 117kb of *CDC7*on 1p22), rs1063192 (within *CDNK2B* on 9p21), and rs9607469 (at the 5' upstream of *CARD10* on 22q13.1) were included in the present study. Amplification of these SNPs was done using primers given in table 2.2.1 at the corresponding temperatures (Appendices- 5 to 8).

**Table 2.2.1: Primer sequences, product size and $T_m$ for PCR amplification of rs1900004, rs3858145, rs10483727, rs1192415, rs1063192, and rs9607469**

| SNP | Genotyping method | Primers 5'-3' | Annealing temperature (˚C) | Amplified product length (bp) |
|---|---|---|---|---|
| rs1900004 | RFLP-*NlaIII* | FP: ACCAAATGCTCACCCAGTGT<br>RP:GCCTGCAGATTGCTAGATCC | 60 | 366 |
| rs3858145 | RFLP-*MfeI* | FP:TGCAAAGCAAAGTGTTTACCT<br>RP: AAGTGAGGACTGTTGGTGGC | 61 | 398 |
| rs10483727 | RFLP-*NdeI* | FP: ATGGGAG CCTACCTCTT GCT<br>RP:CATGGGGGTTAAAAACGAGA | 61.3-54.3 | 230 |
| rs1192415 | Direct Sequencing | FP: GGCCTTAAGATCTTCACAAACAAA<br>RP: CGAGGTGGCGGACTTCTC | 62.9-55.9 | 249 |
| rs1063192 | ASPCR | FW:TCTTTCCTAATGACAAAC<br>FM:TCTTTCCTAATGACAAAT<br>CR:AAACAGCTAAACCTGTCTGC | 43 | 241 |
| rs9607469 | RFLP *Hha I* | FP: GAGAACAGCCTCCTCCTGTG<br>RP: TTGGGATGGAAGTAGGCAAC | 65 | 239 |

Genotyping of SNPs were done by i) PCR based RFLP for SNPs rs1900004 (*NlaIII*), rs3858145 (*MfeI*), rs10483727(*NdeI*), and rs9607469 (*HhaI*) (Appendice 7) (ii) Allele specific PCR for rs1063192 as described by Mabuchi et al [59] and (iii) direct sequencing for rs1192415 in ABI Prism® *Avant 3100* genetic analyzer.

III. **Statistical analysis:** Analyses of allele/ genotype frequencies for association were performed by $\chi^2$ test. All SNPs were analyzed for deviation from Hardy-Weinberg equilibrium (HWE). Association of endophenotypes CCT, VCDR, and AL with the SNPs was performed using linear regression analysis in SPSS software v22. The effect size (β ± SE; conveys the magnitude of the relationship of the SNPs with the endophenotypes in the study) was calculated for the 3 models with the codes 0,1,2 / 0,1,1 / 0,0,1 representing homozygous major , heterozygous and homozygous minor alleles for the additive/dominant/ recessive models respectively in linear regression analysis. Logistic regression analysis was used to analyse the gene-gene interaction among HTG and NTG with 2 models (dominant and recessive) after adjustment for age and gender; wherein the SNPs were used as explanatory variables.

**Results**:

A total of 97 POAG patients (NTG=31; HTG=66) and 371 controls were included in the current study. Table 2.2.2 describes the demographic and clinical features of the study subjects. All the endophenotypes were significantly distributed between the cases and controls (*P* value = <0.0001)

**Table 2.2.2: Demographic and clinical features of the study subjects:**

| | | | age (y)* | | IOP (mmHg) | VCDR | CCT (µm) | AL (mm) | ACD (mm) |
|---|---|---|---|---|---|---|---|---|---|
| | **n** | **male** | **range** | **mean±SD** | **mean±SD** | **mean±SD** | **mean±SD** | **mean±SD** | **mean±SD** |
| **Control** | 371 | 161 | 42-84 | 59.1 ± 8.7 | 14.5 ± 3.2 | 0.4 ± 0.2 | 511.5 ± 31.7 | 22.59 ± 0.8 | 2.8 ± 0.3 |
| **HTG** | 66 | 45 | 45-89 | 64.42 ± 10.36 | 28.3 ± 6.5 | 0.7 ± 0.2 | 528.9 ± 30.9 | 23.9 ± 1.8 | 2.6 ± 0.8 |
| **NTG** | 31 | 24 | 44-74 | 59.39 ± 8 | 17.87± 2.7 | 0.77 ± 0.2 | 518.87 ± 30.1 | 24.04 ± 1.9 | 2.9 ± 0.5 |

*the age at recruitment for controls or age of disease onset for POAG patients are shown

SD-standard deviation, IOP- intraocular pressure, VCDR-vertical cup to disc ratio, CCT- central corneal thickness, AL- axial length, ACD- anterior chamber depth.

## Analysis for association of the SNPs with POAG:

All the six SNPs genotyped in the present study did not show any deviation from HWE (all *P* >0.001). The distribution of genotypes (additive/dominant/recessive) and minor allele

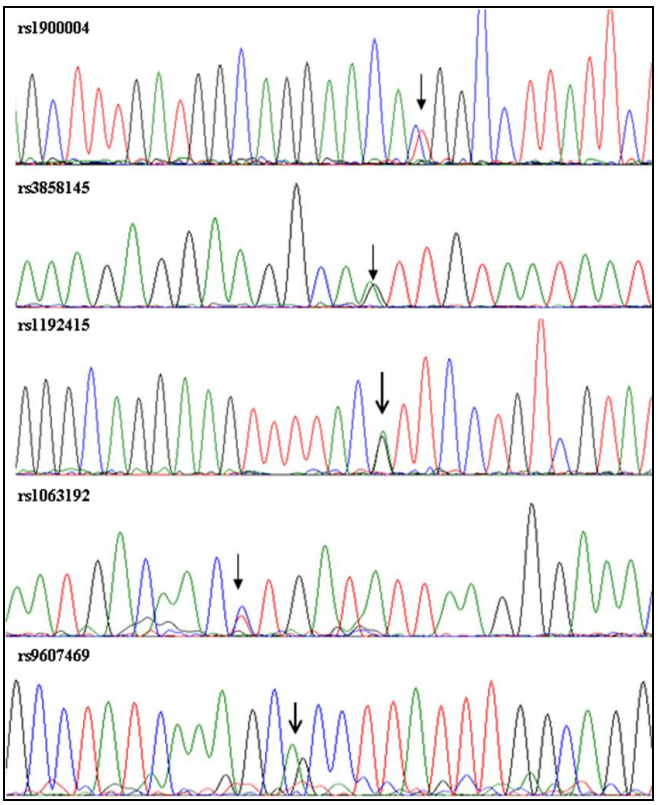
frequencies of all the 6 SNPs did not show any significant difference between the cases and controls ($\chi^2$ analysis) (Table 2.2.3).

**Table 2.2.3: Analysis of minor allele frequencies and $\chi^2$ of the SNPs between POAG cases (HTG, NTG, combined) and controls**

| SNP | Near gene | M/m | Additive model | | | | | | | Domiant model | | Recessive model | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Controls | | Cases | | Genotypic distribution P* | Allelic distribution P* | OR | OR | *P** | OR | *P** |
| | | | N | (%) | N | (%) | | | | | | | |
| rs1900004 | *ATOH7* | C/T | 225 | 30.3 | 62 | 31.9 | 0.9 | 0.75 | 1.1 | 1.1 | 0.7 | 1.1 | 0.8 |
| rs3858145 | *ATOH7* | A/G | 303 | 40.8 | 88 | 45.3 | 0.5 | 0.46 | 1.2 | 1.2 | 0.3 | 1.3 | 0.4 |
| rs10483727 | *SIX1* | T/C | 304 | 40.9 | 82 | 42.2 | 0.1 | 0.11 | 1.3 | 1.6 | 0.1 | 1.5 | 0.2 |
| rs1063192 | *CDKN2B* | T/C | 171 | 23.0 | 47 | 24.2 | 0.8 | 0.95 | 1.0 | 1.0 | 0.9 | 1.2 | 0.7 |
| rs9607469 | *CARD10* | G/A | 217 | 29.2 | 52 | 26.8 | 0.7 | 0.67 | 0.8 | 0.8 | 0.4 | 0.9 | 0.8 |
| rs1192415 | *CDC7* | A/G | 233 | 31.4 | 74 | 38.1 | 0.1 | 0.25 | 1.3 | 1.3 | 0.2 | 1.8 | 0.1 |

M- Major allele, m-minor allele; * *P* values are derived from $\chi^2$ tests; significance: P<0.05

**Fig 2.2.1:** Fig 2.2.1: Electrophoretogram of the variations in i) rs1900004, ii) rs3858145, iii) rs1192415, iv) rs1063192 and v) rs9607469



**Analysis of association of SNPs with IOP, AL, CCT and VCDR in the whole cohort:**

Analysis of association of SNPs with IOP, AL, CCT, and VCDR were performed after adjustment for age and gender; IOP was adjusted as an additional covariate for CCT and vice versa. The Bonferroni corrected *P* value was set to <=0.008 (0.05/6) (Table 2.2.4). SNP i) rs1192415 (*CDC7)* was associated with decreased CCT (*P*= 0.021 (β±SE = -10.9±4.7); recessive model) and increased VCDR (*P*= 0.027 (β±SE = 0.05±0.02); recessive model), ii) rs1900004 (*ATOH7)* was associated with decreased VCDR (*P*= 0.02 (β±SE= -0.03±0.01); additive, *P*= 0.029 (β±SE=-0.09±0.04); recessive model), and increased AL (*P*= 0.028 (β±SE=-0.83±0.4)). However none of these *P* values survived Bonferroni correction.

**Table 2.2.4: Multiple regression analysis for the SNPs with AL, CCT, VCDR and ACD as dependant variables in the whole cohort.**

| **SNP** | **Additive model** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **AL** | | | **CCT** | | | **VCDR** | | | **IOP** | | |
| | **(n=109)** | | | **(n=454)** | | | **(n=426)** | | | **(n=448)** | | |
| | β | SE | ***P**** | β | SE | ***P**** | β | SE | ***P**** | β | SE | ***P**** |
| **rs1900004** | 0.31 | 0.165 | 0.07 | 0.9 | 2.7 | 0.74 | -0.03 | 0.01 | **0.02** $^{\Omega}$ | -0.014 | 0.011 | 0.19 |
| **rs3858145** | 0.13 | 0.157 | 0.38 | -3.3 | 2.5 | 0.19 | 0.023 | 0.012 | 0.06 | 0.002 | 0.01 | 0.88 |
| **rs10483727** | -0.26 | 0.162 | 0.11 | -1.3 | 2.3 | 0.58 | 0.015 | 0.011 | 0.17 | 0.014 | 0.009 | 0.12 |
| **rs1063192** | -0.18 | 0.158 | 0.24 | -1.6 | 2.4 | 0.94 | -0.013 | 0.011 | 0.25 | 0 | 0.009 | 0.97 |
| **rs9607469** | -0.13 | 0.141 | 0.38 | -1.17 | 2.3 | 0.61 | -0.007 | 0.011 | 0.5 | -0.004 | 0.009 | 0.67 |
| **rs1192415** | 0.018 | 0.151 | 0.91 | -2.96 | 2.2 | 0.18 | 0.014 | 0.011 | 0.21 | 0 | 0.009 | 0.97 |
| | **Dominant model** | | | | | | | | | | | |
| | β | SE | ***P**** | β | SE | ***P**** | β | SE | ***P**** | β | SE | ***P**** |
| **rs1900004** | 0.204 | 0.213 | 0.34 | -0.13 | 3.4 | 0.97 | -0.028 | 0.017 | 0.1 | -0.019 | 0.014 | 0.17 |
| **rs3858145** | 0.36 | 0.241 | 0.13 | -2.8 | 3.6 | 0.43 | 0.022 | 0.018 | 0.22 | -0.001 | 0.014 | 0.97 |
| **rs10483727** | -0.207 | 0.215 | 0.34 | -0.98 | 3.17 | 0.76 | 0.026 | 0.015 | 0.09 | 0.023 | 0.012 | 0.06 |
| **rs1063192** | -0.225 | 0.207 | 0.28 | -2.7 | 3.02 | 0.36 | -0.017 | 0.015 | 0.26 | -0.004 | 0.012 | 0.75 |
| **rs9607469** | -0.262 | 0.202 | 0.19 | -0.92 | 2.9 | 0.76 | -0.015 | 0.014 | 0.29 | -0.006 | 0.012 | 0.61 |
| **rs1192415** | 0.047 | 0.205 | 0.82 | -1.16 | 2.9 | 0.69 | 0.006 | 0.014 | 0.7 | 0 | 0.012 | 0.97 |
| | **Recessive model** | | | | | | | | | | | |
| | β | SE | ***P**** | β | SE | ***P**** | β | SE | ***P**** | β | SE | ***P**** |
| **rs1900004** | 0.83 | 0.4 | **0.028** $^{\Omega}$ | 3.08 | 5.6 | 0.59 | -0.06 | 0.03 | **0.029** $^{\Omega}$ | -0.012 | 0.023 | 0.59 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **rs3858145** | **-0.089** | **0.27** | 0.75 | -7.36 | 4.3 | 0.09 | 0.034 | 0.021 | 0.1 | 9.6*e-6 | 0.018 | 1 |
| **rs10483727** | **-0.548** | **0.339** | 0.11 | -2.41 | 4.2 | 0.57 | 0.005 | 0.021 | 0.8 | 0.007 | 0.017 | 0.67 |
| **rs1063192** | -0.457 | 0.368 | 0.22 | 9.2 | 5.6 | 0.1 | -0.005 | 0.028 | 0.86 | 0.011 | 0.023 | 0.63 |
| **rs9607469** | -0.069 | 0.303 | 0.82 | -2.4 | 5.09 | 0.64 | -0.001 | 0.026 | 0.97 | -0.005 | 0.021 | 0.8 |
| **rs1192415** | -0.079 | 0.326 | 0.81 | -10.92 | 4.7 | **0.021**$^{\Omega}$ | 0.051 | 0.02 | **0.027**$^{\Omega}$ | 0.003 | 0.019 | 0.87 |

β = standard regression coefficient; SE = standard error; *Obtained from linear regression, adjusted for age and gender. Bonferroni corrected *P* value=0.008 (0.05/6)

## Genetic Association of the SNPs with endophenotypes IOP, CCT and VCDR in POAG and controls:

The analysis of endophenotypes (IOP, CCT, AL and VCDR) for potential association with the 6 SNPs was performed in POAG (HTG/NTG/combined) and controls using SPSS v17 (Table 2.2.5) after adjustment for age and gender; IOP was adjusted as an additional covariate for CCT and vice versa. The minor allele 'C' of rs1063192 (*CDKN2B*) showed association with decreased AL (additive model) in controls (*P*= 0.012 (β±SE = -0.388±0.15)). The minor allele 'C' of rs10483727 (*SIX1/SIX6)* was associated with decreased VCDR in combined cases and NTG (recessive model) (*P* = 0.014, β±SE =-0.062±0.025, *P* = 0.017, β±SE = -0.102±0.04 respectively). The minor allele 'A' of rs9607469 (*CARD10)* was associated with decreased VCDR in HTG (recessive model) (*P* = 0.022, β±SE = -0.141±0.06). The minor allele 'T' of rs1900004 (*ATOH7)* was associated with decreased VCDR in controls (additive; *P* = 0.013, β±SE = -0.028±0.02) and decreased IOP (additive; *P*=0.042, β±SE =-0.042±0.02). However none of these significant *P* values passed Bonferroni correction.

**Table 2.2.5: Association of the SNPs with VCDR and CCT in controls and POAG (HTG/NTG/combined)**

| **SNP** | **Minor allele** | **Control (n=371)** | | | **NTG (n=31)** | | | **HTG (n=66)** | | | **Combined (n=97)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **β** | **SE** | ***P**** | **β** | **SE** | ***P**** | **β** | **SE** | ***P**** | **β** | **SE** | ***P**** |
| **CCT** | | | | N= 369 | | | N= 27 | | | N= 63 | | | N= 87 |
| **rs1900004** | T | | | 0.34∞ | | | 0.46∞ | | | 0.31¶ | | | 0.29£ |
| **rs3858145** | G | | | 0.16¶ | | | 0.48£ | | | 0.21£ | | | 0.32∞ |
| **rs10483727** | C | | | 0.21∞ | | | 0.59£ | | | 0.08£ | | | 0.15¶ |
| **rs1063192** | C | | | 0.17∞ | | | 0.11£ | | | 0.08∞ | | | 0.27∞ |
| **rs9607469** | A | | | 0.4£ | | | 0.79¶ | | | 0.53£ | | | 0.27£ |
| **rs1192415** | G | | | 0.11 ∞ | | | 0.65£ | | | 0.38∞ | | | 0.14∞ |
| **VCDR** | | | | N= 363 | | | N= 21 | | | N= 45 | | | N= 66 |
| **rs1900004** | T | **-0.03** | **0.012** | **0.013¶** | | | .39∞ | | | 0.41¶ | | | 0.48¶ |
| **rs3858145** | G | | | 0.18∞ | | | .42£ | | | 0.39£ | | | 0.43¶ |
| **rs10483727** | C | | | 0.27£ | **-0.102** | **0.04** | **0.017**∞ | | | 0.3∞ | **-0.06** | **0.025** | **0.014**∞ |
| **rs1063192** | C | | | 0.19¶ | | | 0.13£ | | | 0.18£ | | | 0.52£ |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **rs9607469** | A | | | 0.27£ | | | 0.4£ | **-0.14** | **0.06** | **0.022**∞ | | | 0.07∞ |
| **rs1192415** | G | | | 0.12∞ | | | 0.16¶ | | | 0.52£ | | | 0.34¶ |
| **IOP** | | | | N= 369 | | | N= 27 | | | N= 54 | | | N=81 |
| **rs1900004** | T | | | 0.2£ | | | 0.23¶ | **-0.04** | **0.02** | **0.042¶** | | | 0.29£ |
| **rs3858145** | G | | | 0..62∞ | | | 0.21¶ | | | 0.2£ | | | 0.27∞ |
| **rs10483727** | C | | | 0.11£ | | | 0.23∞ | | | 0.07∞ | | | 0.75∞ |
| **rs1063192** | C | | | 0.14£ | | | 0.72£ | | | 0.59∞ | | | 0.36£ |
| **rs9607469** | A | | | 0.69£ | | | 0.57£ | | | 0.21£ | | | 0.68∞ |
| **rs1192415** | G | | | 0.21¶ | | | 0.15∞ | | | 0.51£ | | | 0.59£ |
| **AL** | | | | **N= 74** | | | | | | | | | **N= 39** |
| **rs1900004** | T | | | 0.49∞ | | | - | | | - | | | 0.11∞ |
| **rs3858145** | G | | | 0.29£ | | | - | | | - | | | 0.45£ |
| **rs10483727** | C | | | 0.26∞ | | | - | | | - | | | 0.21∞ |
| **rs1063192** | C | **-0.38** | **0.15** | **0.012¶** | | | - | | | - | | | 0.17£ |
| **rs9607469** | A | | | 0.22£ | | | - | | | - | | | 0.46∞ |
| **rs1192415** | G | | | 0.62£ | | | - | | | - | | | 0. 07£ |

β = standard regression coefficient; SE = standard error

£- dominant model, ∞- recessive model, ¶-additive model

*Obtained from linear regression

- AL values were not available for NTG group hence regression was not performed.

## Gene-gene interaction analysis:

Since in POAG, interaction of several genes has been proposed, we carried out logistic regression analysis for studying the interaction between the 6 SNPs (coded as in table 2.2.6) [141]. Significant interaction was observed in dominant model for the explanatory variable 2 (model) in HTG cases between rs10483727-rs1063192 (*SIX6-CDKN2B*) (Table 2.2.7) ("CC/CT"-"TT" (*P*= 0.04, OR=2.5, 95%CI: 1.0-6.1)) and rs10483727-rs9607469 (*SIX6-CARD10*) ("CC/CT"-"GG" (*P*=0.031,OR=3.1(1.1-8.6)). In recessive model, we observed significant interactions (NTG group) between rs3858145 with rs10483727 (*ATOH7-SIX6*) ("GG"-"TT/CT", (*P*=0.037, (OR=2.6(1.06-6.5) explanatory variable 1).

**Table 2.2.6: Coding of genotypes in dominant and recessive model analysis of interaction between genes.**

| **SNPs rs10483727/ rs1063192/ rs9607469/ rs1192415** | **Genotype code*** | **SNPs in *ATOH7* (rs1900004/rs3858145)** | |
|---|---|---|---|
| | | **MM and MW** | **WW** |
| Dominant model | **MM and MW** | 3 | 1 |
| | **WW** | 2 | 0 |
| | | **MM** | **WW and MW** |
| Recessive model | **MM** | 3 | 2 |
| | **WW and MW** | 1 | 0 |

*** MM –mutant homozygote, MW- heterozygote, WW- wild homozygote**

**rs1900004: MM- TT, MW-CT, WW-CC; rs3858145 and rs1192415: MM- GG, MW-GA, WW-AA; rs10483727 and rs1063192: MM- CC, MW-CT, WW-TT; rs9607469: MM- AA, MW-GA, WW-GG**

**Table 2.2.7: Analysis for interaction between SNPs in HTG/NTG by dominant and recessive models.**

| | **Dominant model** | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **HTG** | | | **NTG** | | |
| **SNP1** | **SNP2** | **1** | **2** | **3** | **1** | **2** | **3** |
| rs1900004 | rs10483727 | 0.1 | 0.9 | 0.2 | 0.6 | 0.9 | 0.3 |
| | rs1063192 | 0.1 | 0.1 | 0.7 | 0.7 | 0.4 | 0.3 |
| | rs9607469 | 0.8 | 0.9 | 0.7 | 0.7 | 0.3 | 0.6 |
| | rs1192415 | 0.1 | 0.3 | 0.4 | 0.3 | 0.9 | 0.4 |
| rs3858145 | rs10483727 | 0.4 | 0.7 | 0.3 | 0.8 | 0.3 | 0.2 |
| | rs1063192 | 0.2 | 0.2 | 0.8 | 0.9 | 0.2 | 0.2 |
| | rs9607469 | 0.4 | 0.6 | 0.6 | 0.8 | 0.2 | 0.4 |
| | rs1192415 | 0.2 | 0.6 | 0.3 | 0.9 | 0.3 | 0.2 |
| rs10483727 | rs1063192 | 0.2 | 0.04 (2.5(1.0-6.1)) * | 0.1 | 0.7 | 0.8 | 0.7 |
| | rs9607469 | 0.3 | 0.031 (3.1(1.1-8.6)) * | 0.1 | 0.8 | 0.5 | 0.9 |
| | rs1192415 | 0.8 | 0.4 | 0.1 | 0.1 | 0.4 | 0.9 |
| rs1063192 | rs9607469 | 0.4 | 0.6 | 0.9 | 0.1 | 0.3 | 0.7 |
| | rs1192415 | 0.8 | 0.2 | 0.4 | 0.4 | 0.3 | 0.9 |
| rs9607469 | rs1192415 | 0.9 | 0.2 | 0.7 | 0.4 | 0.2 | 0.6 |
| | **Recessive model** | | | | | | |
| **SNP1** | **SNP2** | **1** | **2** | **3** | **1** | **2** | **3** |
| rs1900004 | rs10483727 | 0.5 | 0.6 | 1.0 | 0.2 | 0.1 | 0.9 |
| | rs1063192 | 0.9 | 0.7 | 0.4 | 0.8 | 0.5 | 0.1 |
| | rs9607469 | 0.6 | 0.7 | 1.0 | 0.3 | 0.8 | 0.9 |
| | rs1192415 | 0.6 | 0.1 | 1.0 | 0.2 | 0.1 | 0.9 |
| rs3858145 | rs10483727 | 0.6 | 0.6 | 0.7 | 0.037 (2.6(1.1-6.5)) | 0.09 | 0.5 |
| **SNP1** | **SNP2** | **1** | **2** | **3** | **1** | **2** | **3** |
| | rs1063192 | 0.7 | 0.5 | 0.9 | 0.1 | 0.7 | 0.4 |
| | rs9607469 | 0.8 | 1.0 | 1.0 | 0.1 | 0.5 | 0.5 |
| | rs1192415 | 1.0 | 0.3 | 0.1 | 0.1 | 0.3 | 0.2 |
| rs10483727 | rs1063192 | 0.5 | 0.2 | 0.9 | 0.1 | 0.6 | 0.9 |
| | rs9607469 | 0.5 | 0.4 | 0.9 | 0.1 | 0.9 | 0.9 |
| | rs1192415 | 0.7 | 0.5 | 0.1 | 0.2 | 0.3 | 0.3 |
| rs1063192 | rs9607469 | 0.4 | 0.7 | 0.9 | 0.9 | 0.7 | 0.9 |
| | rs1192415 | 0.6 | 0.2 | 0.4 | 0.8 | 0.2 | 0.9 |
| rs9607469 | rs1192415 | 0.9 | 0.2 | 0.6 | 0.9 | 0.2 | 0.9 |

*Represents *P* value/ oddsratio (95%CI))

## Discussion

POAG is a complex genetic disorder exhibiting clinical and genetic heterogeneity. Identification of the genetic variants associated with the endophenotypes has been proven as one of the effective method, in disorders with complex inheritance pattern [142]. Literature reveals significant association of these genes with POAG in several populations

like US Caucasians [58], Japanese [59], Rotterdam study cohort I and II [55], Australian and UK twin cohort [57], Afro-Caribbean [143] etc. In contrast, our study did not show significant association for any of these SNPs with POAG which could be possibly attributed to difference in the population. A similar observation has been earlier reported in East Indian cohort [144]. The reduced power (49% by Genetic power calculator [145]) of the current study has limited the probability to detect true associations.

We analyzed for the potential association of SNPs near *ATOH7, CDKN2B, SIX6, CARD10,* and *CDC7* with AL, CCT, IOP and VCDR. The other optic disc parameters like ODA, were not included for the current study due to the unavailability of data for these samples.

*CDKN2B* (p15), a cyclin dependant kinase inhibitor type 2B gene, acts as a tumour suppressor in the retinoblastoma pathway. This gene along with the adjacent genes like *CDKN2A, CDKN2B-AS* is involved in cell signalling pathways and has been associated with POAG [146]. The 'C' allele of rs1063192 was previously associated with a decreased risk for POAG in Afro-Caribbean population [143] and with decreased VCDR [55, 58], while the other allele "T" has been associated with increased VCDR and NTG [59]. We did not observe any significant association with POAG/VCDR. A similar observation was made by Vishal et al in East Indian population, thus suggesting that rs1063192 is not associated with VCDR in Indian population [144]. The 'C' allele of rs1063192 was correlated with decreased AL in controls; which indirectly suggests less risk for POAG in the study population. This is the first report on the association of *CDKN2B* with AL and needs to be replicated for further disease correlations.

*ATOH7* is a transcription factor, required for the genesis of retinal ganglion cells [147, 148]. The ectopic expression of *ATOH7* was shown to increase the number of differentiated retinal ganglion cells in *invitro* models [149, 150]. Mutations in this gene have been associated with autosomal recessive persistent hyperplasia and optic nerve hypoplasia [57, 151]. Earlier GWAS analysis has shown association of SNPs rs1900004 [55], rs3858145 [57] and rs7916697 with ODA and VCDR [56] in various cohorts. In the present study we have analysed 2 SNPs (rs1900004 and rs3858145) in which we were able to replicate the association of rs1900004 with decreased VCDR in whole cohort (additive and recessive models) as well as controls (additive models). Further the 'T" allele of rs1900004 also showed novel associations with decreased i) AL (whole cohort- recessive model) and ii) IOP (HTG- additive model). The 'A' allele of rs3858145 had been

previously associated with increased cup and disc area in Australian and UK cohort [57], but in the current study we did not find any significant association.

*SIX1*/*SIX6* complex belong to the homeoprotein group of proteins involved in development. SNP rs10483727 (T-allele) has earlier been associated with increased VCDR [55, 58, 59, 152], IOP [152, 153] and POAG in several GWAS studies [55, 56, 83, 154, 155]. In the current study, the 'C' allele was observed as minor allele in contrast to that observed in other populations and it was associated with decreased VCDR in NTG and combined cases.

*CDC7* a cell division cycle protein is critical for G1/S transition. Earlier studies have shown association of *CDC7* with increased ODA and POAG in various populations [55, 56, 143]. Interestingly, in the present study we have observed an indirect risk for POAG, as association was observed with subjects having thin CCT and increased VCDR, major risk factors for POAG. This novel association of *CDC7* with CCT has to be replicated for further disease correlation.

*CARD10* belongs to caspase recruitment domain family member 10 which is involved in regulation of caspase activation and apoptosis (via NF-kappaB) in addition to its role in assembling membrane associated signalling complexes. SNP rs9607469 (A-allele) in this gene has been associated with increased ODA [56] by GWAS. A recent meta analysis for VCDR identified SNP rs5756813 near *CARD10* as one of the loci with positive correlation for VCDR [156]. In the current study we have also replicated the association of *CARD10* with VCDR wherein the 'A' allele of rs9607469 is associated with decreased VCDR. We did not identify any association with POAG as seen in other reports.

Gene-gene interactions have long been recognized to be fundamentally important for understanding genetic causes of complex disease traits. Its importance has grown to a great extent especially in complex disorders wherein 2 or more genes interact to precipitate the disorder. Such interactions were earlier observed in POAG as described initially by Fan et al; they reported interaction of MYOC, OPTN and apolipoprotein E (APOE) genes using logistic regression based test and stratified analysis [157]. Another study in Chinese population identified a best 6-factor model for POAG: MYOC IVS2+35A>G, OPTN Met98Lys, OPTN IVS5+38T>G, OPTN IVS8-53T>C, WDR36 IVS5+30C>T, and APOE -491A>T [158] by MDR analysis. Research in Spanish population showed an increased risk for glaucoma with different WDR36 genotypes that also increased significantly in

combination with the P53 RP risk genotype, indicating the existence of a genetic interaction [159]. Moreno et al observed a significant interaction between the SEC14L2/TAP and SLC23A2 polymorphisms which increases POAG risk in those subjects who had both risk genotypes at the same time ($p<0.01$) [160]. The other include interaction of OLFM2-OPTN [159], and ATOH7-RFTN1 [161].

Analysis of potential interaction between the SNPs was carried out using logistic regression analysis. We were able to replicate the association between the genes *ATOH7-SIX6* among NTG as shown by Fan etal [58] and also between genes *SIX6-CDKN2B* among HTG as shown by Iglesias et al [152]. Novel interactions, that were statistically significant was observed between *SIX6-CARD10, SIX6-CDC7* and *ATOH7-CDC7* among HTG patients. Chen et al showed that though individually the SNPs were not associated with POAG, an introduction of interaction term brought significant result [161]. In our study we also did not observe any significant association with POAG but when analysed with interaction effects, the association was ascertained thus substantiating the fact that complex disorders such as glaucoma, result from interactions between several genes. These interactions among genes like *ATOH7, SIX6, CDKN2B* and *CDC7* suggests a putative additive role of the developmental and growth signalling pathways in POAG as hypothesized in many studies [83, 152].

In summary, our results suggest that a probable interaction among the candidate genes for QTs play a major role in determining the individual's susceptibility to POAG, similar to other complex diseases. The current study reports a potential interaction between SNPs near *ATOH7 - CDC7, SIX6-CARD10, SIX6-CDC7* genes with POAG for the first time in the literature. SNPs near the genes (*ATOH7, SIX6, CDKN2B, CARD10,* and *CDC7*) that did not show individual association with the disease however correlated with increased risk when analysed for interactions. We also replicated the association of these genes with QTs as observed in other population. Additionally, we report here a novel association of SNPs near (i) *ATOH7* and *CDKN2B* with AL (ii) *CDC7* with CCT and iii) *ATOH7* with IOP . However as none of these *p*-values (linear regression analysis) survived Bonferroni correction which is a stringent criterion to prevent type one errors, replication and functional studies are necessary for further conclusions.

## Chapter 2.3

## Genetic association study of glaucoma endophenotypes with SNPs of *LOXL1* in XFS/XFG

**Objective**:

1. To identify and recruit of patients with Pseudoexfoliation syndrome/ glaucoma (XFS/XFG).
2. To genotype rs1048661 and rs3825942 in XFS cases and controls from Indian population.
3. To perform association analysis between the genotypes and QTs- IOP, CCT, VCDR and ACD.

**Methodology**:

**I. Patient recruitment:** (Appendice- 1)

**Sample size**: Pseudoexfoliation Cases: 114

Controls: 135

**II. Laboratory protocols:**

1. **Blood collection**, **DNA extraction- Nucleospin technique, Quantification of the DNA extracted** (Appendice- 2-4)
2. **PCR Amplification:** The region of exon 1 of *LOXL1* gene flanking the two SNPs () were amplified using the primers mentioned in table 2.3.1. Each reaction mixture was prepared as described in table 2.3.2. The prepared mix was then subjected to PCR with primer annealing from 58˚C to 55˚C with 0.5˚C decrement for each cycle (Appendice-5 and 6). Genotyping was performed using *ABI 3100Avant* genetic analyser (Appendice-8 and 9 )

**Table 2.3.1 Primer sequences, product size and $T_m$ for PCR amplification of *LOXL1***

| Gene | Primer sequence | Base pairs | $T_m$ ˚C | Product size (bp) |
|---|---|---|---|---|
| *LOXLl-3* | Forward primer 5'- ATT CGG CTT TGG CCA GGT –3'<br>Reverse primer 5'- GAA CTG CTG CGG GTA GGA –3' | 18<br>18 | 58-55 | 178 |

## III. Data Analysis

Genotype and allelic distribution was calculated by $\chi^2$ analysis. Association of the SNPs with the QTs were performed by linear regression analysis in SPSS software v.21.

**Results**:

Table 2.3.2 shows the demographic details of the subjects included in the current study. The mean age of controls and cases were 63 and 67 respectively. The IOP were <21 mmHg in controls while the cases had a range of 8-46mmHg. The mean VCDR and IOP were significantly higher in cases when compared to controls ($P$= 2.2e$^{-16}$, $P$= 1.5e$^{-7}$).

**Table 2.3.2: Demographic details of the subjects included in the study**

| | | Age | Gender %males | IOP mmHg | CCT | VCDR |
|---|---|---|---|---|---|---|
| **Controls** | MIN_MAX | 40-84 | 60% | 8-21 | 436-585 | 0.1-0.7 |
| N=132 | MEAN±SD | 63±7.6 | | 14.3±2.9 | 511±37.1 | 0.4±0.14 |
| **XFS** | MIN_MAX | 50-84 | 61.9% | 9-36 | 435-562 | 0.3-0.9 |
| N=63 | MEAN±SD | 66±7.3 | | 15.8±5.5 | 505±31.9 | 0.6±0.19 |
| **XFG** | MIN_MAX | 45-107 | 60.6% | 8-46 | 403-582 | 0.2-0.9 |
| N=33 | MEAN±SD | 69±12.4 | | 21.8±11.3 | 510.4±34.5 | 0.77±0.2 |
| **XFS+OHT** | MIN_MAX | 48-84 | 62.5% | 10-38 | 460-540 | 0.2-0.8 |
| N=16 | MEAN±SD | 67±10.2 | | 23.2±7.6 | 510.4±24.2 | 0.54±0.2 |
| ***P* value** | NA | 0.00046 | NA | 1.5e$^{-7}$ | 0.62 | 2.2e$^{-16}$ |

Table 2.3.3 and fig 2.3.1 shows the analysis of distribution and electrophorogram of rs1048661 and rs3825942 among XFS cases and controls in the current study. The genotypic and allelic frequencies of both the SNPs were significantly distributed in this cohort with a $P$ value of 0.00024 and 0.0056 for rs1048661 and 0.0000014 and 6.8e$^{-7}$ for rs3825942 respectively. The OR values were < 1 and the minor allele frequencies of rs1048661 and rs3825942 were represented more in controls when compared to cases.

**Fig 2.3.1: Electrophorogram of A. rs1048661 and B. rs3825942**

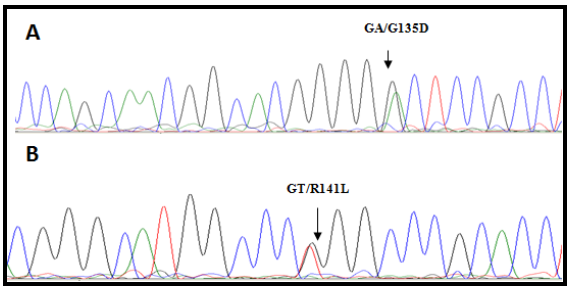


**Table 2.3.3: Distribution of genotype and allelic frequency among the cases and controls**

| SNP | Genotype distribution | | | Allelic distribution | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Cases n=112 HH HM MM | Control=132 HH HM MM | *P* value | Cases H M | MAF | Control H M | MAF | *P* value | OR (95%CI) |
| rs1048661 | 58 45 9 | 36 86 10 | $2.4e^{-4}$ | 161 63 | 0.28 | 158 106 | 0.4 | 0.0056 | 0.58 (0.39-0.85) |
| rs3825942 | 103 8 1 | 87 40 5 | $1.4e^{-6}$ | 214 10 | 0.045 | 214 50 | 0.189 | $6.8e^{-7}$ | 0.2 (0.09-0.4) |

HH –homozygous wild allele, HM – heterozygous, MM- homozygous variant allele; MAF: minor allele frequency.

Table 2.3.4 and 2.3.5 shows the linear regression analysis of the two SNPs with the endophenotypes like IOP, CCT, and VCDR in the whole cohort (cases + controls) and individually. In the whole cohort analysis rs3825942 was significantly associated in both additive and dominant models with IOP $P$=0.006(β=-2.8, SE=1.02), $P$=0.014(β=-2.9,SE=1.2), and VCDR $P$=0.031(β=-0.068,SE=0.03), $P$=0.041(β=0.07,SE=0.04 respectively.

**Table 2.3.4: Linear regression analysis of rs1048661 and rs3825942 with the endophenotypes in the whole cohort**

| **SNP** | **IOP*** (β,SE) | **CCT*** | **VCDR*** (β,SE) |
|---|---|---|---|
| rs1048661 | 0.806 | 0.803 | 0.695 |
| rs3825942 | 0.006(-2.8,1.02) | 0.425 | 0.031(-0.068,0.03) |

* P value

Analysis in cases (XFS, XFG, XFS+XFG) and controls separately revealed significant *P* value for IOP with rs1048661 and rs3825942 in controls ($P$ =0.017(β=-1.56, SE=0.64), dominant model; $P$ =0.051(β=2.7, SE=1.3), recessive model). SNP rs3825942 was also associated with CCT in XFS ($P$ =0.051(β=2.7, SE=1.3), dominant model) and with VCDR in XFS ($P$ =0.025(β=0.42, SE=0.17), recessive model).

**Table 2.3.5: Linear regression analysis of rs1048661 and rs3825942 with the endophenotypes in cases and controls**

| **rs1048661** | | | | | |
|---|---|---|---|---|---|
| | **Controls*** | **XFS*** | **XFG*** | **XFS+OHT*** | **Combined cases*** |
| IOP | **0.017£ (-1.56,0.64)** | 0.102 Ŧ | 0.677£ | 0.145 Ŧ/£ | 0.456 γ |
| CCT | 0.427£ | 0.410 γ | 0.254£ | 0.992 Ŧ/£ | 0.564 γ |
| VCDR | 0.822 γ | 0.076 γ | 0.430 γ | 0.445 Ŧ/£ | 0.054 γ |
| **rs3825942** | | | | | |
| | **Controls*** | **XFS*** | **XFG*** | **XFS+OHT*** | **Combined cases*** |
| IOP | **0.05γ (2.7,1.3)** | 0.431 γ | 0.363£ | 0.320 Ŧ/£ | 0.231 Ŧ |
| CCT | 0.276 γ | 0.061 Ŧ | **0.05£ (2.7,1.3)** | 0.597 Ŧ/£ | 0.210 γ |
| VCDR | 0.247£ | **0.025γ (0.42,0.17)** | 0.383£ | 0.291 Ŧ/£ | 0.083 γ |

£-DOMINANT, γ-RECESSIVE, Ŧ-ADDITIVE

**Discussion:**

Glaucoma is a one of the leading cause of blindness worldwide. XFG is a secondary open angle glaucoma occurring as a consequence of XFS. The prognosis is much worse when compared to that of POAG. Prevalence of XFS/XFG in India ranges from 1.25 -6% (in persons above 40yrs) [67-69]. Hence discovering the predisposing genetic factors of XFG could help in predicting the risk of development of the disease, and facilitating early medical intervention.

Researchers at DECODE Genetic,Inc. in Iceland along with researchers at the University of Iceland and Uppsala University , Sweden were the first to demonstrate that two non-synonymous changes rs1048661 and rs3825942 (R141L and G153D) in exon 1 of the *LOXL1* gene[108] was associated with XFS/XFG. Functionally this gene catalyses the crosslinking of tropo-elastin and fibulin-5. Thus any variation in this gene leads to improper crosslinking and hence deposition of PEX material in the TM which in turn is responsible for increase in IOP.

In a recent normative data and association study among the Chinese cohort, it is observed that PEX was associated with ocular traits such as shallow ACD, short AL which was in accordance with Reykjavik Eye study[162] and South Indian Chennai glaucoma study [68] or Burmese Meiktila Eye study[163]. With this prelude of association of *LOXL1* SNPs with PEX/XFG, we hypothesised that *LOXL1*could be a candidate QTL for XFG. Hence we analysed the association of *LOXL1* SNPs rs3825942 and rs1048661 with QTs like IOP, VCDR AL, ACD, and LT. This is the first study; to our knowledge evaluating the association of these well studied SNPs with quantitative traits of XFG.

In this study with 114 cases and 135 unrelated controls of Indian origin, it was observed that there is allelic association of rs3825942 ($P$=6.8e$^{-7}$) and also rs1048661 ($P$= 0.0056) with XFS. Higher frequency of minor alleles in controls when compared to cases suggests a protective / reduced risk for the disease. Linear regression analysis of the two SNPs with various QTs yielded significant *P* value for rs1048661 with IOP in controls ($P$=0.017, β= -1.56 ± 0.64, dominant model); rs3825942 was significantly associated with i) with IOP in controls ($P$=0.051, β= 2.7 ± 1.3, recessive model) ii) with CCT in XFG cases ($P$=0.051, β= 2.7 ± 1.3, dominant model) and iii) with VCDR in XFS cases ($P$=0.025, β= 0.42 ± 0.17, recessive model).

Analysis of association of rs1048661 and rs3825942 with XFS/G and its endophenotypes revealed significant association of these SNPs with the disease and its endophenotypes IOP, CCT and VCDR. Recent study by Anastasopoulos et al in Greece cohort, carried out a similar study but did not find association of *LOXL1* SNPs with differences in IOP, the possible reason for this could be the difference in the ethinicity, and study design (theirs was a population based study, present is a case control approach) [124].

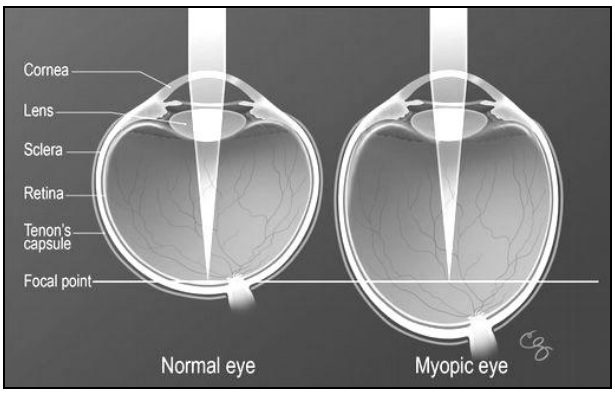
The current study is first of its kind to evaluate association of well known *LOXL1* SNPs with endophenotypes of XFG, and these results might augment the present knowledge on XFS/XFG. There are other studies which state that though the *LOXL1* SNPs has been proven to be associated with PEX, their association with glaucoma still stands to be established[114, 126, 164] thus suggesting that apart from *LOXL1* there could be other genes environmental factors which could attribute to the development of the XFG. Hence further analysis using functional methods and high throughput techniques is required to study the pathophysiology of XFS/XFG.

## Chapter 3
## Genetic Mapping of endophenotypes of myopia

### Introduction

Myopia or nearsightedness is one of the most common human eye disorders with significant global public health concern. Along with cataract, macular degeneration, and vitamin A deficiency, myopia is an important cause of visual impairment worldwide. A cross-sectional study in US based population show the prevalence rate to be high among Asians (19.8%) when compared to African Americans, Hispanics, and Whites (5.2-14.5%) [165]. The prevalence rate is reported to be higher among school children (36.71%) [166]. Chronic high myopes are more susceptible to serious complications such as glaucoma, retinal detachment, choroidal neovascularisation etc, thereby causing permanent visual impairment or even blindness [167-169].

**Fig 3.1: Comparison of normal and myopic eye structure** [170].




Emmetropia is a state of the normal eye, where parallel rays of light are focused on the retina due to the proper growth and maintenance of different eye structures. The major factors contributing to the normal emmetropamization includes the refractive powers of the cornea and lens, AL refractive indices of the aqueous and vitreous; all of which are highly heritable. In addition to this, certain conformational factors such as intrauterine environment, bony orbits and eyelids also influence the eye shape and growth [171].

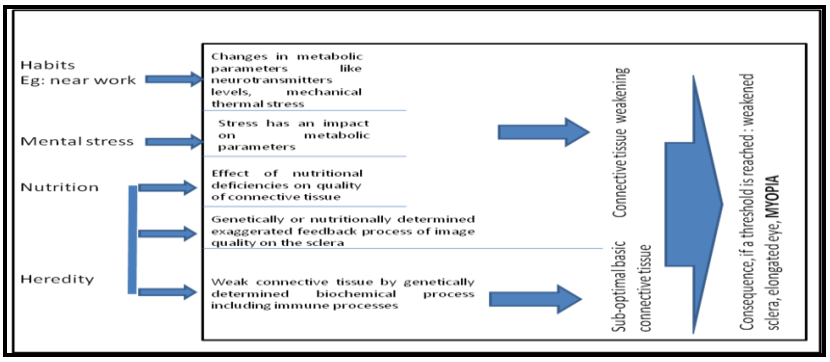
In myopia distant objects are not displayed sharply on the retina, because the light rays from the lens converge even before they reach the retina (Fig: 3.1). This occurs due to increased growth of the eye ball (increased AL) which is attributed to excessive lengthening of the posterior chamber caused by scleral remodelling; however the exact etiology still remains unclear. Substantial resources are required in correcting the refractive error that includes spectacles, contact lenses, and surgical procedures such as LASIK. The market for these optical aids has gone up drastically and was estimated to be

about 755 million dollars/year, [172] causing a huge economic burden to the society. Thus studies are warrentied to identify the cause and hence the prevention or delay in onset of this disorder.

**Mechanisms involved in pathology of myopia**

Myopia results from complex interactions of both genetic and environmental factors. Genetic predisposition for the disease is identified based on several twin / family based studies. The role of environmental factors in precipitating the disorder has been proven by studies from urban cities of Asia including Singapore and Hong Kong, which has high prevalence rate of myopia. Few studies have also proven that childhood exposure to night lighting increases the prevalence of myopia, but conflicting results prevail [173]. Based on these studies scientists have deciphered the interaction of genetics and environment factors (mental stress, near work, nutrition status) in the pathology of myopia through various mechanisms (fig: 3.2). The mechanical stress elicited during accommodation process causes changes in connective tissue leading to remodelling of ocular shape, and biochemically weaken the ocular structure. The mechanical stress can initiate biochemical alterations via a process called mechanotrancduction which describes the cellular processes that translate mechanical stimuli into biochemical signals, thus enabling the cell to adapt to their physical surroundings. In case of myopia, mechanical stress upregulates several genes in the human scleral fibroblasts like parathyroid hormone/its receptor, which in turn triggers activation of protein kinase C that is directly involved in scleral remodelling leading to abnormal increase in eye size [174].

**Fig 3.2: Different mechanisms leading to the myopia development** [171].




**Gene mapping strategies for myopia**

The general strategies followed in mapping genes for myopia includes linkage and association studies. Apart from human studies, animal models have also contributed to the

current understanding on the mechanisms of myopia development. The recent trends in gene mapping for complex disorders like myopia include QTL mapping, as it overcomes the limitations in classical gene mapping strategies that are based on the end phenotype (disease status); hence reduces the problems of genetic heterogeneity, phenotype variations, pedigree size etc.

In myopia, the measurable phenotypes like AL, RE, ACD, LT etc are selected as QTs for gene mapping studies and many chromosomal regions have been identified till date. Twin studies have identified high heritability for ocular biometric determinants of myopia including AL (0.89-0.94), ACD (0.88) and LT (0.94) [28].

*Linkage based methods*

Linkage based methods are more often adopted as the first method for gene mapping studies in complex diseases when the knowledge on the exact disease mechanisms are minimally available. The first chromosomal locus (MYP1) for myopia (Xq28) was mapped by Schwartz et al in a syndromic condition called Bornholm Eye Disease (BED) which included ambylopia, myopia, and deutranopia [175] followed by other studies using linkage analysis (Table 3.1) in large pedigrees or affected sib pairs; of which the chromosomal loci, 18p11, 12q21-23, 7q36, 17q21-22, 4q22-27, 2q37, Xq23-25 and 10q21.2 have been mapped for non-syndromic high myopia, while 22q12, 1q36 has been attributed to common myopia. These linkage studies for myopia have been carried out worldwide in different populations like Chinese, Africans, Caucasians, Asians etc. Study by Ciner et al in 94 African-American and 36 White families showed a significant linkage to chr 20 [176]. Another study in a genetically well defined population of Older Amish families revealed a maximum LOD of 2.04 in chr 8q23 [177]. Yu et al performed linkage analysis using 330 microsatellite markers spanning the genome in an autosomal dominant family with pathological myopia and narrowed on 15q12-13 with LOD of 1.76[178]. Study by Lam et al mapped novel loci on chr 5p15.33-p15.2 with an interval of 17.45 in 3 autosomal dominant myopic pedigrees from China [179].

**Table 3.1: List of loci mapped for myopia worldwide [180]**

| Myopia locus | Inheritance /QTL | Location | LOD score | Ethnicity of subjects |
|---|---|---|---|---|
| MYP1 | XR | Xq28 | 4.8 | Danish |
| MYP2 | AD | 18p11.31 | 9.59 | American and Chinese |
| MYP3 | AD | 12q21-23 | 3.85 | German/Italian and UK |

| | | | | |
|---|---|---|---|---|
| MYP5 | AD | 17q21-22 | 3.17 | English/ Canadian |
| MYP6 | AD | 22q12 | 3.54 | Ashkenazi Jewish families and American of European / |
| MYP7 | QTL | 11p13 | 6.10 | UK population |
| MYP8 | QTL | 3q26 | 3.7 | UK population |
| MYP9 | QTL | 4q12 | 3.3 | UK population |
| MYP10 | AD & QTL | 8q23 | 4.1 | UK population |
| MYP11 | AD | 4q22-27 | 3.11 | Han Chinese population |
| MYP12 | AD | 2q37.1 | 4.75 | US family of northern Europe |
| MYP13 | XR | Xq23-25 | 2.75 | Chinese |
| MYP14 | QTL | 1p36 | 9.54 | Ashkenazi Jewish |
| MYP15 | AD | 10q21.1 | 3.22 | Hutterite family |
| MYP16 | AD | 5p15.33-p15.2 | 4.81 | Chinese family |
| MYP17 | AD | 7p15 | 4.07 | Algerian and French |
| MYP18 | AR | 14q22.1-q24.2 | 2.19 | Chinese family |
| MYP19 | AD | 5p15.1-p13.3 | 3.02 | Chinese family |
| MYP20 | AD | 13q12.12 | p = 7.75 x 10 (-5) | Han Chinese population |
| MYP21 | AD | 1p22.2 | 3.19 | Han Chinese family |
| MYP22 | AD | 4q35.1 | Exome sequencing | Chinese family |

Till date 21 myopia loci have been identified out of which five MYP loci (MYP7 –MYP10 and MYP 14) were identified through QT linkage analysis for refractive error measurements (SE and SPH) and MYP7-MYP10 loci were identified in 221 dizygotic twin pairs with maximum LOD score of 6.1 at 40cM on chr 11p13. A regression based linkage analysis on Ashkenazi Jewish families identified a region about 49.1cM on chr 1p36 with a maximum score of 9.5 for ocular refraction[181]. Another important endophenotype of myopia includes AL. Genome wide studies have linked AL to chr 2 with a LOD score of 2.64 [29]. Another classical genome wide twin study done by Zhu et al showed the highest LOD score of 3.40, on chr 5q (at 98 cM). Additional regions with suggestive multipoint LOD scores were also identified on chr 6 (LOD scores, 2.13 at 76 cM and 2.05 at 83 cM), chr 10 (LOD score, 2.03 at 131 cM), and chr 14 (LOD score, 2.84 at 97 cM) [30]. Linkage to 21 loci for myopia across different populations proves its genetic heterogeneity.

*Genetic association studies in myopia and its QTs*

Association studies in myopia are performed with functional / non functional SNPs in candidate genes including genes involved in scleral/extracellular matrix remodelling, retinal cell proliferation and maintenance, collagen remodelling etc [182]. SNPs in candidate genes like *TGFβ* [183], insulin like growth factor I [184], *COL1A1* [185] early growth response factor 1[186], *PAX6* [187], matrix metalloproteinase genes [188], hepatocyte growth factor [189] *MFRP* [190], *VSX2* [191] etc have been studied for the association with myopia/hyperopia in different population. QTL studies / meta analysis of

3 GWAS conducted in Chinese children observed genome wide significance for AL at 1q41 harbouring the zinc finger 11B pseudogene *ZC3H11B*[40]. Corneal curvature (CC), another QT of myopia, has been associated with 2 genes *PDGFRA* and *TRIM29* in Australian and Asian cohorts [192].

Consortium for Refractive Error and Myopia (CREAM) is the largest study on GWAS meta-analysis pertaining to refractive error [182, 193] . It includes data from different parts of the world including Asia, Europe, United States and Australia, and covers about 32 GWAS. The GWAS meta-analysis was done in three stages and included a total of 45,758 individuals [182]. They identified a total of 24 novel loci for refractive error and the risk of getting the disease increased by 10 fold in individuals who carried more number of risk alleles, thus depicting the clinical significance of these newly identified loci. Among those associated include genes involved in:

1. Neurotransmission – (GRIA4, GJD2, RASGRF1, A2BP1),
2. Ion transport, channel activity and maintenance of membrane potential (KCNQ5, CD55, CACNA1D, KCNJ2, CHRNG,MYO1D),
3. Retinoic acid metabolism (RDH5, RORB, CYP26A1),
4. Extracellular matrix remodelling (LAMA2 and BMP2),
5. Eye development (SIX6 and PRSS56)

**Animal models for myopia**

Animal models too have provided great insight into the parameters/ pathways that are involved in the development of myopia. The ability to produce an environmental change mimicking the real scenario and then measuring the ocular changes as they develop is particularly important in separating casual relationships from changes that develop as a consequence of induced myopia. Linkage analysis in mice has mapped 2 loci (*Eye1* and *Eye2*) for eyeball size and weight. [194]. Insight into the different mechanisms like role of (i) *GABA* (γ aminobutyric acid) in refractive changes [195],(ii) TGF-β isoforms in scleral remodelling [196], (iii) altered fibrillogenesis and mechanical expansion [197], iv) regulation of AL of by membrane frizzled protein (*MFRP*) [198] are from various animal models like chick, tree shrews and monkeys.

Studies pertaining to genetics of myopia from Indian subcontinent are minimal, which includes association studies in candidate gene like *TGFβ* [199], *Fok1*[200], and linkage studies [201] etc. Hence the current study was conducted to gain deeper knowledge on the genetics of myopia and its QTs in our cohort.

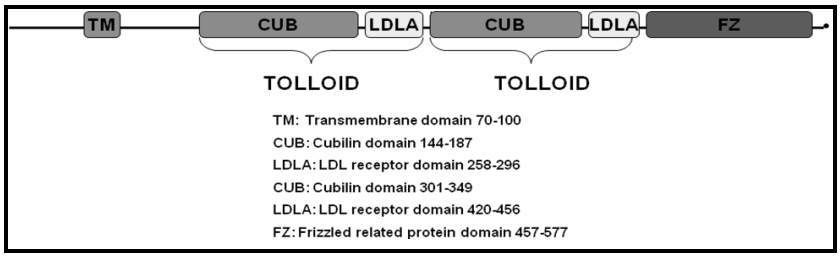
**Candidate genes analyzed in the current study for the possible association with myopia and its QT.**

The candidate genes included in the current study have been selected based on their regulatory role in eye development. The three genes *MFRP*, *VSX2*, and *PAX6* are developmental genes which play a role in determination of eye size as they are expressed early during fetal growth and vital for the proper formation of eye structures as shown by several animal studies.

***MFRP***

*MFRP* has been identified from previous studies as "small eye" gene which is associated with conditions like non-syndromic microphthalmia, nanophthlamia, extreme hyperopia etc which are ocular diseases with altered AL and thus proposed as putative candidate gene for AL[191, 202]. The gene is located at 11q23 and codes for a type 2 trans-membrane protein expressed in the retinal pigment epithelium (RPE), ciliary bodies and adult brain. The protein consists of 579 amino acids which include a transmembrane domain, and an extracellular region with tandem repeats of two cubilin (CUB) domains, low-density lipoprotein receptor related sequence (LdLa), and a cysteine-rich domain (CRD) (Fig 3.3). The presence of the frizzled domain at the 'C' termini suggests that this protein could act as a regulator of the WNT signalling pathway during eye development [203]. The cysteine rich domain is the binding motif for Wingless, a member of the WNT family of proteins.

**Fig 3.3: Schematic representation of *MFRP* protein domains**



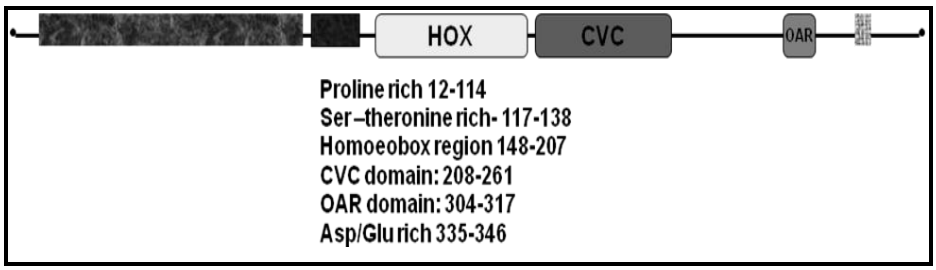

*MFRP* expression has been shown to increase in the E11 stage and coincides with the development of the pigment layer in the presumptive retina and then at birth when the

ciliary body, TM and iris mature (P8) and reduces immediately within a week's time [204] thus suggesting its probable role in the development of the anterior segment and hence in the determination of eye size [204]. Mutational studies in *MFRP* have been correlated with severe hyperopia and nanophthalmos. Study by Sudin et al demonstrated a shift towards myopia in patients with heterozygous null mutation [205]. Thus suggesting a regulatory role of this gene in AL and hypothesized as a good candidate gene for studying AL related disorders [203].

***VSX2/CHX10***

*VSX2/CHX10*, belongs to the paired-like class of homeobox (HOX) proteins, which encodes a homeodomain-containing transcription factor that is expressed by RPCs since the early stages of retinal development and during optic cup formation [206]. It is the vertebrate homolog of nematode Ceh 10 containing homeobox. The human protein is 39KDa and contains 361 amino acids. The gene is located on chr 14q band 24.3. The protein contains four regions: an octapeptide (FGIQEILG), a homeodomain, an OAR domain (from Otp, Aristaless and Rax), and a CVC domain (Chx10, Vsx1 and Ceh 10) (Fig 3.4). The homeodomain contains 50 aminoacids and has a helix-turn-helix motif that mediates DNA binding [207].

**Fig 3.4: Schematic representation of *VSX2* protein domains**




*CHX10* and *PAX6* are co-expressed in RPCs during embryonic development. While *PAX6* maintains the multipotency, *CHX10* is important for the proliferation of RPCs, thus mutations in this gene lead to reduction in number of RPCs which manifests as "small eye" (reduction in AL) [208, 209]. The protein repress the expression of cyclin dependant kinase inhibitor $p27^{kip1}$ and Microphthalmia transcription factor (Mitf) which in turn prevents RPCs from exiting cell cycle enabling proliferation and maintaining the neuroretinal identity [210]. In adults its expression is restricted to the outer part of the inner nuclear layer particularly in the bipolar cells.

Burmeister et al. has shown that mutation in the homeodomain region of this protein causes microphthalmos (small eye) in ocular retardation (Or) mouse [211] . Human studies have identified mutations in the DNA-recognition helix of the homeodomain (R200Q and R200P) disrupting the DNA binding activity of the protein and deletion of homeobox. Identification of the *CHX10* consensus DNA-binding sequence (TAATTAGC) showed that R200Q and R200P mutations severely disrupt its function.

***PAX6***

*PAX6* gene (11p13) is a major transcription factor which contains both paired and homeo domains. It is a homolog of the gene *eyeless* of drosophila [212] It is expressed in all ocular tissues during the 6-9 weeks of human embryo development [213] The function of this gene mainly involves the differentiation of horizontal, amacrine cells in the inner nuclear layer [214] and also the corneal cells, TM (all involving the anterior chamber), and the schelemm's canal [215]. In humans mutations present in this gene leads to aniridia, Peters' anomaly, iris aplasia, ectopia of the pupil, anophthalmia and retardation in eye and nose development in newborns [216]. Studies on mice models with severe phenotypes (small-eye) showed mutations in this gene and also *Mitf* (Micropthalmia associated transcription factor) which severely affects the AL and various other quantitative parameters of eye including ACD, cornea and LT [32]. Recent studies in chick model have shown that *PAX6* and *SOX2* co-bind to a regulatory element driving lens induction and any failure in this process could be responsible for microphthalmia in patients with mutations in these genes [33]. Thus it is one of the most important candidate gene to be considered in various disorders with either increase or decrease in AL.

**Organization of chapters in the current study**:

We have taken two approaches to map the genes / loci associated with myopia and its QTs.

*Chapter 3.1* details the association study carried out in candidate genes *MFRP*, *VSX2* and *PAX6* with myopia and its endophenotype AL in group of cases and controls from Indian population.

*Chapter 3.2* details a Genome-wide approach to map loci for myopia in 2 families using Affymetrix 10K and 250K Array chips.

## Chapter 3.1

## Genetic analysis of *MFRP, VSX2* and *PAX6* for its association with AL in myopia

**Objectives:**

1. To screen the coding regions of *MFRP, VSX2* and 3'UTR of *PAX6* gene in myopia patients and unrelated healthy controls from India
2. To estimate the genotype and allele frequencies and determine the genetic association with myopia
3. To analyze the potential association of the SNPs and haplotype with AL and myopia.
4. To perform bioinformatic analysis for investigating the putative effect of variations on the protein function.

**Materials and Methods**

**I. Sample recruitment and clinical examination** (Appendice 1)

Sample size**:** Myopia: N= 98, Normal controls-:N= 91

**II. Lab protocols**:

**1. Blood collection**, **DNA extraction- Nucleospin technique, Quantification of the DNA extracted** (Appendice 2-4)

**2. PCR Amplification:** The *MFRP* and *VSX2* genes were amplified using the primers given below [217, 218] (Table 3.1.1 and 3.1.2) and genotyped by direct sequencing in ABI Prism 3100 AVANT genetic analyzer (Applied Biosystems, California). An 850 bp length of 3'UTR (NCBI build: 31811300 to 31810450bp) in *PAX6* gene with a regulatory role on expression was predicted by Target Scan software and scanned with overlapping primers (Table 3.1.3) followed by genotyping using direct sequencing technology. Each reaction mixture was prepared as described in appendice 5 and subjected to PCR. Corresponding annealing temperature are given in table 3.1.1, 3.1.2, 3.1.3.

**Table 3.1.1: Primer sequences, product size and $T_m$ for PCR amplification of *MFRP*.**

| Exon | Primer sequence (5'-3') | Annealing Temperature (˚C) | Product length (bp) |
|---|---|---|---|
| 1,2 | F-CTTTTCCCTCGGGTAGTAGA<br>R-GTTGGCAGGTGGGGTTTTGAA | 59° | 494 |
| 3,4 | F-AGGACAGCATGAGGAATACC<br>R-AACCCCACCCCGTCATCTTG | 57° | 511 |
| 5 | F-GAGAAGGGCCCAAGATGACG<br>R-TCCCTGCCACTCCCTGATTC | 64° | 419 |
| 6,7 | F-TTGGGGGTTGAGAAAATAGG<br>R-CTGGGCCAAAGAATGACTGA | 63-58° | 601 |
| 8 | F-ATCACCGCCAGCCCTATTG<br>R-CATCCCCCGTCTGCTTGAT | 58° | 281 |

| | | | |
|---|---|---|---|
| 9 | F-GGTGCCCGGGATGAGACAG<br>R-CCGGGGGTGGCAGACAGT | 58° | 267 |
| 10,11 | F-GCAGTGCCCCCTCAGTCAGC<br>R-GTGGGCACCCAGCCTGCTC | 69-64° | 510 |
| 12,13 | F-CAGGCCACAGAGCCAGTGAG<br>R-AGCCCTGACCGGCAAAAGAG | 63° | 548 |

**Table 3.1.2: Primer sequences, product size and $T_m$ for PCR amplification of *VSX2* exons**

| Exon | Primer sequence (5'-3') | Annealing Temperature (˚C) | Product |
|---|---|---|---|
| 1 | FP : GGCACCTGGGACCAACTTCGC<br>RP : CCCCAGTCAGGCAGCAGCGC | 54 | 501 |
| 2 | FP : GTTTCGGGCACAGCGGAGCGC<br>RP : GACCTCAGATCCGTTGTCGGCG | 55 | 256 |
| 3 | FP : GTTCTGTCTTGTCTGAGACAGGC<br>RP : AACCCTGGAGTGGTAGATGTCAC | 56 | 260 |
| 4 | FP : GAAATCTTCACTCCAAGCCTACAAG<br>RP : AGCCCGCTCTCCTCACCGCAG | 54 | 323 |
| 5 | FP : GGGAGTAAGGCTTTCTGCTCGTC<br>RP : GAGTCTTGGGGCTCCGGCATCTG | 56 | 480 |

**Table 3.1.3: Primer sequences, product size and $T_m$ for PCR amplification of the 3'UTR region of *PAX6* gene**

| 3'UTR region | Primer sequence (5'-3') | Annealing Temperature (˚C) | Product |
|---|---|---|---|
| P1 | FP: ACTGGGGAAGGAATGGACTT<br>RP: TCTAGTGCATGTTGTTCCAGGT | 57.7ºC | 369 |
| P2 | FP: ATCGAATACTTCTACCCATCTGTT<br>RP: TGATTGGACCGTGAACAGTAA | 56.4ºC | 437 |
| P3 | FP: TGTCTTCCCTAGAAATCCTCAGA<br>RP: TCAGGATGTAAAACAGAGTGTGTG | 58.7ºC | 434 |

## III. Data Analysis

**Genotype and Haplotype analysis**: The genotypes were checked for HWE and statistical significance by PLINK [219] and Fisher's Exact test [220]. Pairwise LD was computed in Haploview program, version 4.1[221]. The parameters for LD computation include the correlation coefficient ($r^2$), haplotype estimation using accelerated EM algorithm similar to the partition/ligation method described in Qin et al, 2002. Logistic regression was performed for haplotype association using the sliding window method in PLINK; wherein for a given window size within a gene under consideration, the test was performed for all window sizes shifting one SNP at a time toward the 3' end of the gene. This technique has been proven to be more powerful especially for identifying SNPs lying in low LD regions [222]. Interaction of the haplotype with QT like AL was analyzed using the THESIAS software v3.1 [223] that calculates the significance of associations based on the maximum likelihood model and linked to the SEM algorithm that can perform the simultaneous estimation of haplotype frequencies and their associated effects on the phenotypes of interest.

**Bioinformatic analysis:** The pathogenic effect of the SNPs on the function of the protein was predicted using bioinformatic tools like SIFT, Polyphen [224, 225] etc. The prediction of function of variations at transcriptional level, pre-mRNA splicing, protein structure, etc. were analyzed using the Function Analysis and Selection Tool for Single Nucleotide Polymorphisms (FASTSNP) [226]. MUPro was used to test the protein stability changes due to single site mutations. Since the prediction was based on the sequence itself rather than the tertiary structure, [227] the proteins with no information on the tertiary structure (eg: *MFRP*) can also be analyzed by this software. Additionally the ΔRSCU (Relative Synonymous Codon Usage) was also calculated using CAIcal software[228] to measure the effect of synonymous variation on the translation kinetics (Appendice 7)

## Results

### 1. Demographic data

Table 3.1.4 shows the demographic data of the study participants. The mean age of cases (with myopia) (26.2±6.2) was not significantly different from the controls (without myopia) (26.5±5). Refractive error was significantly higher (p value=$2.27e^{-52}$) in the cases (-9.84±3.84) than the control group (±). AL was also significantly higher in cases (26.99±1.86mm) than the controls (22.95±0.7mm) with a *P* value of $1.45e^{-21}$ and no significant difference was observed for ACD.

**Table 3.1.4: Demographic data**

| Parameters | Controls | Cases | *P* value |
|---|---|---|---|
| Number | 91 | 98 | |
| Sex(Male/Female) | 40/53 | 57/45 | 0.08 |
| Age(mean±SD,Years) | 26.5±5 | 26.2±6.2 | 0.41 |
| Refractive degree(mean±SD,D) | ± | -9.84±3.84 | $2.27e^{-52}$ |
| Axial length(mean±SD,mm) | 22.95±0.7 | 26.99±1.86 | $1.45e^{-21}$ |
| Anterior chamber depth(mean±SD, mm) | 3.45±0.21 | 3.56±0.25 | 0.13 |

±: </= 0.5

### 2. MFRP Results

Fifteen variations (5 novel SNPs: c.55-15_17dup GTA, c.290C>T, c.975+18T>C, c.1124+11C>G, c.1255+33_34del) were observed in the *MFRP* gene (Table 3.1.5). We observed significant difference for the genotype (*P*= $4.10*e^{-07}$) and allele frequencies (*P*= $4.5*e^{-07}$) of SNP rs36015759 (c.492C>T, T164T) in the controls when compared to myopic cases with OR <1.0, CI: 95% (Table 3.1.5).

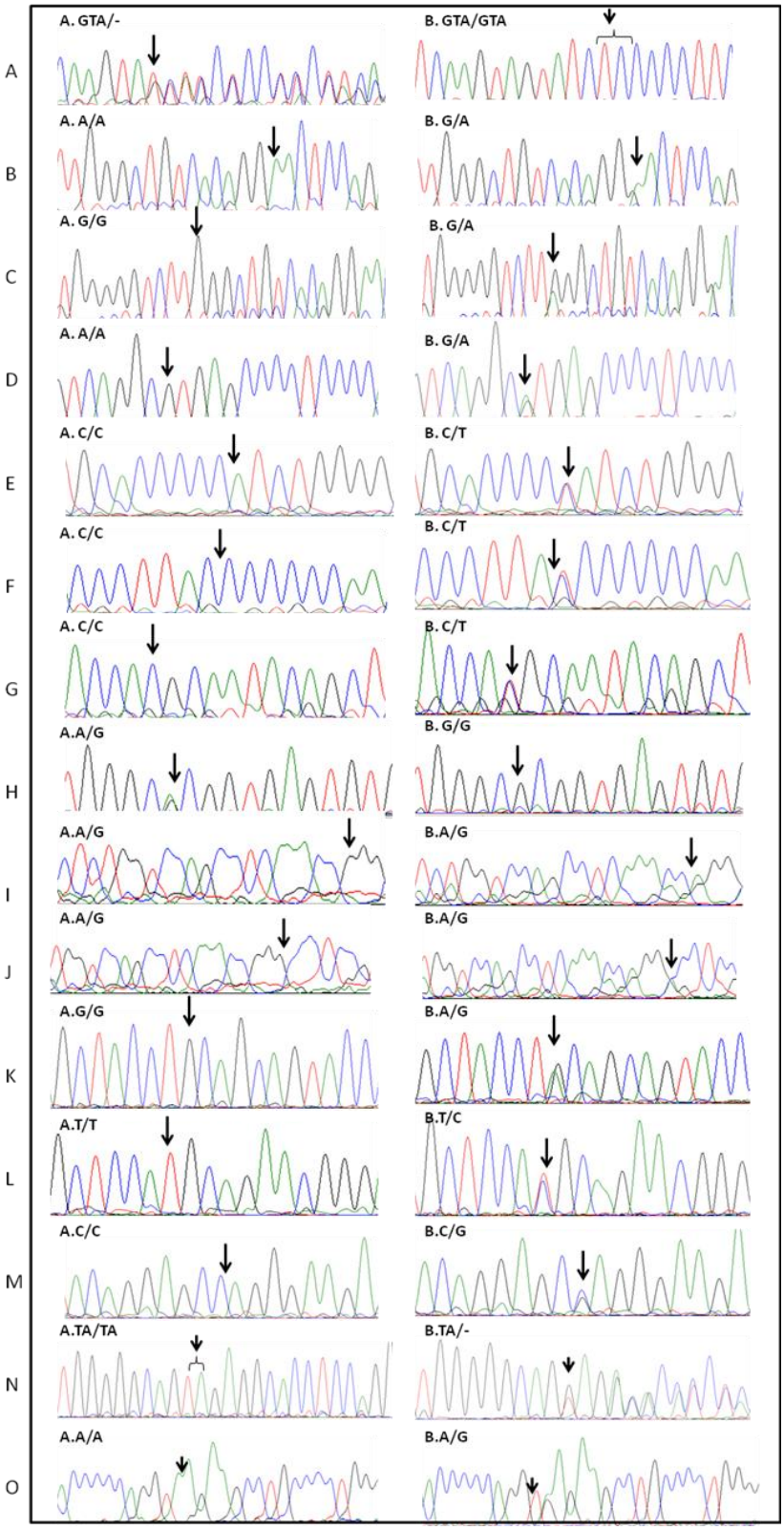
**Table 3.1.5: Genotype and allele frequency distribution of *MFRP* variants in myopia patients**

| Variation (rs ID)/ position | cDNA position | Amino acid change | Allelic distribution | | | | | Genotype distribution | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Cases (N=98) h | m | Controls (N=91) h | m | *P* value | Cases (N=98) hh | hm | mm | Controls (N=91) hh | hm | mm | *P* value |
| rs883247 5'UTR | c.-30G>A | - | 47 | 149 | 59 | 123 | 0.08 | 7 | 33 | 58 | 11 | 37 | 43 | 0.21 |
| rs79836575 5'UTR | c.-43G>A | - | 191 | 5 | 175 | 7 | 0.5 | 94 | 3 | 1 | 84 | 7 | 0 | 0.2 |
| Novel, rs199473710 intron 1 | c.55-15_17dupGTA | - | 189 | 7 | 175 | 7 | 1 | 94 | 1 | 3 | 84 | 7 | 0 | 0.2 |
| Novel, rs199473708 Exon 4 | c.290C>T | P97L | 194 | 2 | 182 | 0 | 0.49 | 96 | 2 | 0 | 91 | 0 | 0 | 0.49 |
| rs3814762 Exon 4 | c.406G>A | V136M | 158 | 38 | 155 | 27 | 0.27 | 65 | 28 | 5 | 64 | 27 | 0 | 0.1 |
| **rs36015759 Exon5 ©** | **c.492C>T** | **T164T** | **173** | **23** | **122** | **60** | **4.5*e$^{-07}$** | **76** | **21** | **1** | **37** | **48** | **6** | **4.10*e$^{-07}$** |
| rs2510143 Exon5 | c.540T>C | H180H | 193 | 3 | 177 | 5 | 0.32 | 95 | 3 | 0 | 86 | 5 | 0 | 0.48 |
| rs61736238 Exon6 | c.770G>A | R257H | 195 | 1 | 182 | 0 | 1 | 97 | 1 | 0 | 91 | 0 | 0 | 0.99 |
| rs10892350 Intron7 | c.898+86G>A | - | 88 | 108 | 91 | 91 | 0.35 | 22 | 44 | 32 | 26 | 39 | 26 | 0.6 |
| rs2509388 Intron7 | c.898+89C>G | - | 88 | 108 | 92 | 90 | 0.3 | 22 | 44 | 32 | 26 | 40 | 25 | 0.5 |
| rs35885438 Exon 8 | c.954G>A | L318L | 166 | 30 | 167 | 15 | **0.03** | 71 | 24 | 3 | 77 | 13 | 1 | 0.11 |
| Novel, rs185451482 intron 8 | c.975+18T>C | - | 195 | 1 | 179 | 3 | 0.3 | 97 | 1 | 0 | 88 | 3 | 0 | 0.35 |
| Novel, rs199473709 Intron 9 | c.1124+11C>G | - | 191 | 5 | 177 | 5 | 1 | 94 | 3 | 1 | 86 | 5 | 0 | 0.48 |
| Novel, rs199473711 Intron 10 | c.1255+33_34del | - | 195 | 1 | 182 | 0 | 1 | 97 | 1 | 0 | 91 | 0 | 0 | 1 |
| rs11217241 intron11 | c.1387+3G>A | - | 100 | 96 | 92 | 90 | 1 | 27 | 46 | 25 | 23 | 46 | 22 | 0.89 |

Allele and genotype frequencies of the sequence variations within *MFRP* gene in the current study are shown. The distributions did not deviate from those predicted by the Hardy–Weinberg equilibrium. Proportions of groups were compared by the Fisher's exact test. The criterion for statistical significance was $p \leq 0.05$. DNA changes are documented based on cDNA sequences with +1 corresponding to the A of the ATG translation initiation codon in reference to NM_031433.2 (*MFRP*) sequences. (h – ancestral allele, m- mutant allele).

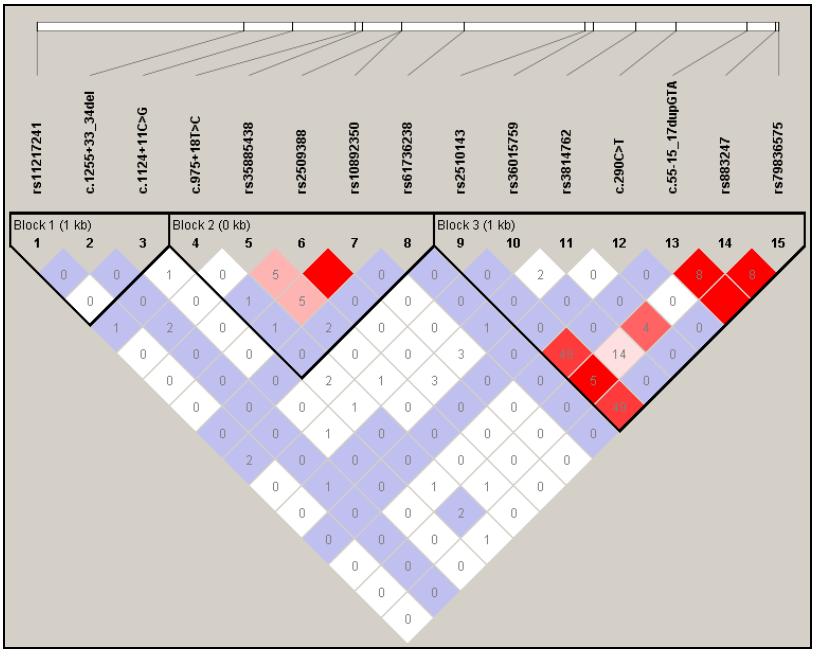
© rs36015759: hh- CC, hm – CT, mm- TT.

**Fig 3.1.1: Electrophorogram representing the 15 variants (** homozygous and heterozygous forms )in *MFRP*, seen in the present study. A-rs199473710, B-rs883247, C-rs79836575, D-rs3814762, E-rs199473708, F-rs36015759, G-rs2510143, H-rs61736238, I-rs10892350, J-rs2509388, K-rs35885438, L-rs185451482, M-rs199473709, N-rs199473711, O-rs11217241



HAPLOVIEW v4.1 analysis showed 3 haplotype blocks for the *MFRP* variants (fig: 3.1.1) that were corrected for multiple comparisons (15000 permutations; table 3.1.6). There was no significant difference in the distribution for blocks 1 and 2 between cases and controls. Two haplotypes (CCGCAAG and CTGCAAG with wild type (C) and variant (T) allele for rs36015759) in block 3 showed significant difference in distribution between cases (1.3%) and controls (13%) ($P_{perm}$ = 0.04 and 0.0001 respectively; table 3.1.6).

**Fig 3.1.2: Haplotype block generated by Haploview software v4.1 for the 15 variations of *MFRP***




Order of the SNPs for the respective blocks is: rs11217241, rs99473711, rs199473709 (Block 1); rs185451482, rs35885438, rs2509388, rs10892350 rs61736238 (Block 2); rs2510143, rs36015759, rs3814762, rs199473708, rs199473710, rs883247, rs79836575 (Block 3). The haplotypes were generated using Solid Spine algorithm and the LD measures are indicated as the correlation coefficient ($r^2$). The shades of gray indicate the magnitudes of the $r^2$ values.

**Table 3.1.6: Distribution of *MFRP* Haplotypes and chi-square *P* values as analyzed by Haploview v4.1**

| Block | Haplotype | Freq. | Case, Control Frequencies | Chi Square | $P_{asym}$ | $P_{Perm}$ |
|---|---|---|---|---|---|---|
| Block 1 | GAC | 0.497 | 0.492, 0.502 | 0.04 | 0.8407 | 1 |
| | AAC | 0.474 | 0.478, 0.471 | 0.02 | 0.8885 | 1 |
| | AAG | 0.018 | 0.012, 0.024 | 0.766 | 0.3814 | 1 |
| Block 2 | TGGAG | 0.491 | 0.516, 0.465 | 0.979 | 0.3225 | 0.9 |
| | TGCGG | 0.379 | 0.326, 0.436 | 4.868 | 0.0274 | 0.1 |
| | TACGG | 0.092 | 0.118, 0.064 | 3.301 | 0.0692 | 0.5 |
| | TAGAG | 0.024 | 0.030, 0.018 | 0.523 | 0.4697 | 1 |
| | CGGAG | 0.011 | 0.005, 0.016 | 1.168 | 0.2799 | 0.9 |
| Block 3 | CCGCAAG | 0.488 | 0.557, 0.415 | 7.625 | 0.0058 | 0.04 |

| Block | Haplotype | Freq. | Case, Control Frequencies | Chi Square | $P_{asym}$ | $P_{Perm}$ |
|---|---|---|---|---|---|---|
| | CCACAAG | 0.144 | 0.180, 0.106 | 4.141 | 0.0419 | 0.3 |
| | CTGCAGG | 0.128 | 0.099, 0.160 | 3.231 | 0.0723 | 0.5 |
| | CCGCAGG | 0.106 | 0.108, 0.104 | 0.01 | 0.9202 | 1 |
| | **CTGCAAG** | **0.07** | **0.013, 0.132** | **20.512** | **$5.93*e^{-06}$** | **0.0001** |
| | TCGCCGA | 0.018 | 0.015, 0.022 | 0.228 | 0.6333 | 1 |
| | CTACAAG | 0.015 | 0.006, 0.024 | 2.01 | 0.1563 | 0.8 |

Order of the SNPs for the respective blocks is: rs11217241, rs99473711, rs199473709 (Block 1); rs185451482, rs35885438, rs2509388, rs10892350 rs61736238 (Block 2); rs2510143, rs36015759, rs3814762, rs199473708, rs199473710, rs883247, rs79836575 (Block 3). $P_{asym}$ asymptotic P value, $P_{Perm}$ empirical P value for 15,000 permutations.

In addition, we also examined haplotypes using the sliding window strategy in PLINK and generated a total of 120 windows, of which 59 showed a significant *P* value (omnibus $P_{emp}$ <0.05). For easiness in representing the SNPs in haplotype they are named as S1, S2….S15. The omnibus test result was significant for those SW consisting of rs36015759 (S6, Table 3.1.7) in the haplotype. The importance of rs36015759 (S6) was even more obvious when we probed more into the details of the most significant group of SW which was of size upto 6 SNPs per window. In window S1-S2-S3-S4-S5-S6, the most significant haplotype was observed with the first five SNPs in the wild type and the last SNP rs36015759 (S6) was either the wild allele 'C' or the variant 'T' ($P_{emp}=6.67*e^{-.05}$ and $P_{asym}=0.004$ respectively). Interestingly the 'T' allele haplotype was more represented in controls (13%) whereas the haplotype with 'C' allele was seen more in cases (56%) (Table: 3.1.8).

**Table 3.1.7: Summary of Exhaustive Haplotype Analyses Based on Omnibus Tests for Sliding Windows of All Possible Sizes Across the 15 SNPs of the *MFRP* gene for the subjects under this study**

| **Sliding window** | | **SW with omnibus test $P_{emp}$ <0.05** | | | **Most significant results** | | |
|---|---|---|---|---|---|---|---|
| **SNPs no** | **No. of SWs** | **No. of SWs** | **First SW** | **Last SW** | **SW** | $P_{asym}^{a}$ | $P_{emp}^{a}$ |
| 2 | 14 | 2 | S5-S6 | S6-S7 | S6-S7 | $1.95*e^{-06}$ | $6.67*e^{-05}$ |
| 3 | 13 | 5 | S4-S6 | S6-S8 | S6-S8 | $2.23*e^{-06}$ | $6.67*e^{-05}$ |
| 4 | 12 | 4 | S3-S6 | S6-S9 | S4-S7 | $5.37*e^{-05}$ | $6.67*e^{-05}$ |
| 5 | 11 | 5 | S2-S6 | S6-S10 | S4-S8 | $5.37*e^{-05}$ | $6.67*e^{-05}$ |
| 6 | 10 | 6 | S1-S6 | S6-S11 | S1-S6 | $8.59*e^{-06}$ | $6.67*e^{-05}$ |

| SNPs no | No. of SWs | No. of SWs | First SW | Last SW | SW | $P_{asym}{}^{a}$ | $P_{emp}{}^{a}$ |
|---|---|---|---|---|---|---|---|
| 7 | 9 | 6 | S1-S7 | S6-S12 | S2-S8 | $7.04*e^{-05}$ | $6.67*e^{-05}$ |
| 8 | 8 | 6 | S1-S8 | S6-S13 | S6-S13 | $4.73*e^{-05}$ | $6.67*e^{-05}$ |
| 9 | 7 | 5 | S1-S9 | S6-S14 | S6-S14 | 0.00021 | $6.67*e^{-05}$ |
| 10 | 6 | 6 | S1-S10 | S6-S15 | S4-S13 | 0.00018 | $6.67*e^{-05}$ |
| 11 | 5 | 5 | S1-S11 | S5-S15 | S4-S14 | 0.00022 | $6.67*e^{-05}$ |
| 12 | 4 | 4 | S1-S12 | S4-S15 | S3-S14 | 0.00041 | 0.000533 |
| 13 | 3 | 3 | S1-S13 | S3-S15 | S2-S14 | 0.00042 | 0.001333 |
| 14 | 2 | 2 | S1-S14 | S2-S15 | S1-S14 | 0.00061 | $6.67*e^{-05}$ |
| 15 | 1 | 1 | S1-S15 | S1-S15 | S1-S15 | 0.00745 | 0.0056 |

For easiness in representing the SNPs in haplotype they are named as S1, S2….S15. The SW is indicated as Sa-Sb, where 'a' is the first SNP and 'b' the last SNP of the SW. For example, S2-S4 refers to the SW S2-S3-S4. Multiple comparisons were corrected by running 15,000 permutations. The minimum P value achievable with 15,000 permutations is 6.67*e-05. For each fixed-size SW, the most significant results are shown in the 3 rightmost columns. Of the 120 sliding windows tested the S5-S6 ranks the first and S6-S8 the second in providing evidence for association. a Abbreviation Pasym asymptomatic P value; Pemp empirical P value; SW- Sliding window.

**Table 3.1.8: Details of haplotype analysis for the 6 window showing the most significant results among the all possible sliding windows.**

| Haplotype | Frequency in | | OR | $P^{asym}$ | $P^{emp}$ |
|---|---|---|---|---|---|
| **rs883247, rs79836575, rs199473710, rs199473708, rs3814762, rs36015759** | **Cases** | **Controls** | | | |
| **AG1CAT** | **0.006457** | **0.02612** | **0.0283** | **0.044** | **0.0172** |
| GG1CGT | 0.09962 | 0.1683 | 0.516 | 0.047 | 0.04453 |
| **AG1CGT** | **0.01285** | **0.138** | **0.0135** | **0.00042** | **$6.67*e^{-05}$** |
| AG1CAC | 0.1818 | 0.1059 | 1.96 | 0.0364 | 0.033 |
| GA2CGC | 0.02585 | 0.03565 | 0.749 | 0.614 | 0.6369 |
| GG1CGC | 0.1093 | 0.1073 | 1.03 | 0.933 | 0.9363 |
| AG1CGC | 0.5642 | 0.4187 | 1.89 | 0.00413 | 0.004 |

Haplotypes were indicated in ACGT format. The 1 in the third position indicates the normal while the 2 indicates GTA duplication.

Haplotype blocks that included only the 6 coding SNPs of the order rs199473708, rs3814762, **rs36015759**, rs2510143, rs61736238, rs35885438 (Table 3.1.9) were generated by THESIAS and analyzed for interaction with AL after adjusting for age and sex as covariates. The frequency distribution for the block CG**T** (rs36015759) CGG was high in controls (31%; *P* = 0.000001, OR: 0.18240 95%CI: 0.09244 - 035988) and showed

significant *P* value for decreased mean AL (*P*= 0.000444; difference in mean value = 0.10320; Table 3.1.9).

**Table 3.1.9: Frequency distribution of *MFRP* haplotypes and their interaction with AL as generated by Thesias software v3.1**

| ***MFRP* Haplotype (P97L,V136M, T164T, H180H,** R**257H, L318L)** | | | **Frequency distribution** | | **Haplotype interaction with AL** | |
|---|---|---|---|---|---|---|
| | **Frequency in Controls (N=91)** | **Frequency in Cases (N=98)** | ***P* value** | **OR ( 95% CI)** | ***P* value** | **Difference in mean value** |
| CGCCGG | 0.437 | 0.550 | Intercept | | | |
| CGCCGA | 0.068 | 0.110 | 0.427961 | 1.43337 (0.58850 - | 0.978239 | 0.01693 (-1.19965 - 1.23351) |
| CGCTGG | 0.022 | 0.020 | - | | | |
| CGTCGG | 0.315 | 0.088 | **0.000001** | **0.18240 (0.09244 - 0.35988)** | **0.000444** | **-1.35085 (-2.10460 - - 0.59711)** |
| CGTCGA | 0.006 | 0.016 | - | | | |
| CACCGG | 0.130 | 0.169 | 0.692164 | 1.16495 (0.54707 - 2.48068) | 0.813546 | 0.10320 (-0.75438 - 0.96078) |
| CACCGA | 0.008 | 0.018 | - | - | | |
| CACCAG | 0 | 0.005 | - | - | | |
| CATCGG | 0.011 | 0.009 | - | - | | |
| TGCCGG | 0 | 0.010 | - | - | | |

Haplotypes are indicated in ACGT format. The data are boldfaced if their corresponding haplotypes show significant P value (<0.05). All data were adjusted for gender and age.

Polyphen/ SIFT / SNAP analysis of the *MFRP* variants did not show any significant changes (Table 3.1.10). FastSNP analysis revealed an alternative splicing regulatory region (either change in Transcription factor (TF) binding or Exonic splicing enhancer (ESE) site) for c.290C>T, c.975+18T>C, rs11217241, rs35885438, rs61736238 and rs36015759 (Table 3.1.11). SNPs c.290C>T, rs3814762, rs61736238 showed a decrease in protein stability by MUpro analysis (Table 3.1.11). Except for rs36015759 all other SNPs did not show any significant association with myopia. The significant SNP rs36015759 showed a loss of ESE binding site for SRp40 protein [229] (Fig: 3.1.2).

**Table 3.1.10: Bioinformatic analysis by Polyphen, SIFT and SNAP for non synonymous variations of *MFRP* to determine the pathogenicity**

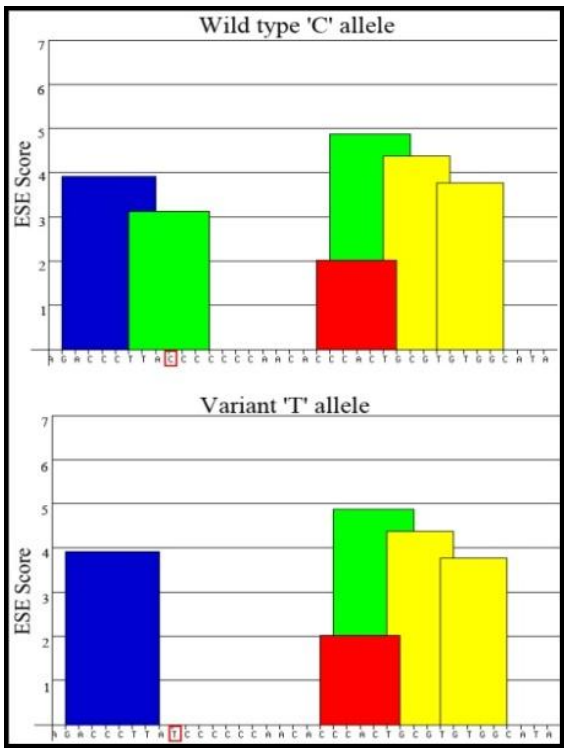
| **Variations** | **Aminoacid change** | **Polyphen** | **SIFT** | **SNAP** |
|---|---|---|---|---|
| rs199473708 | P97L | Benign | Tolerated (1) | Neutral |
| rs3814762 | V136M | Benign | Tolerated (0.4) | Neutral |
| rs61736238 | R257H | Benign | Tolerated (0.6) | Neutral |

**Table 3.1.11: FastSNP and MUpro analysis for the putative effect of *MFRP* gene variants on its function and protein stability**

| **Variations (rs ID)** | **FastSNP** | | **MUpro** (confidence score) |
|---|---|---|---|
| | **Functional effect** | **Alternative splicing regulatory** | |
| rs199473708* | - | 2 new transcription factors for C M00126 GATA-1 , M00128 GATA-1 and 2 new ESE binding sites. | Decrease (-0.13) |
| rs185451482* | - | 2 new sites at 21 and 16 for ESE | - |
| rs199473709* | - | No variation predicted | - |
| rs11217241 | Splicing regulation with risk of 3-4 | Change (loss) in the TF binding region | - |
| rs883247 | Promoter/regulatory region with a risk of 1-3 | _ | - |
| rs35885438 | Sense/synonymous; Splicing regulation risk 2-3 | Loss of ESE binding site at 16, 19, 18 | - |
| rs3814762 | Missense (conservative), risk: 2-3 | No change in protein structure but it is a missense leading to a conservative change | Decrease (-0.55) |
| rs61736238 | - | Loss of TF binding region and new site at position 18 | Decrease (-0.49) |
| rs 36015759 | Splicing regulation risk 2-3 | Loss of ESE binding site for SRp40 | - |

* **Novel SNPs**

**Fig 3.1.3: ESE Binding sites as predicted using ESE Finder for rs36015759, wild and variant allele.**




ESE binding sites for wild type (upper panel) and variant (c.492C>T; lower panel) of MFRP cDNA (NM_031433.2) shown as red, blue, green and yellow boxes. The binding site for SRp40 protein is lost for the variant (2nd green bar, lower panel).

## 3. VSX2 Results

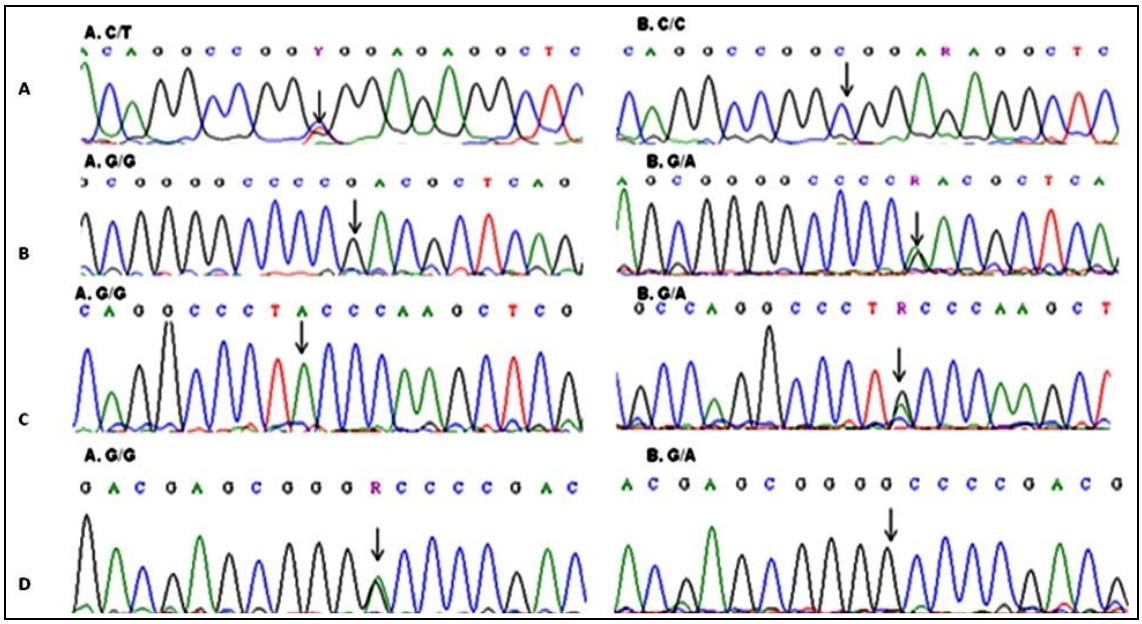
Direct sequencing revealed no variations in exons 1, 2, 3 and 4. There were 4 variations in exon 5 of which 3 were novel (rs199473712, c.866G>A, G289D; rs75395981, c.871G>A, D291N; rs137872696, c.1046C>T, A349V) and one known variation (rs62006815, c.831G>A, p.Leu277=) respectively (Fig: 3.1.3). All these variations in *VSX2* were equally distributed between cases and controls (Table 3.1.12).

**Table 3.1.12**: **Genotype and allele frequency distribution of *VSX2* variants in exon 5 detected and investigated in myopia patients.**

| **Variation** | **cDNA position** | **Amino acid change** | **Allelic distribution** | | | | ***P* value** | **Genotype distribution** | | | | | | ***P* value** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Cases N=98** | | **Controls N=91** | | | **Cases N=98** | | | **Controls N=91** | | | |
| | | | **h** | **m** | **h** | **m** | | **hh** | **hm** | **mm** | **hh** | **hm** | **mm** | |
| rs62006815 Synonymous | c.831 G>A | L277L | 194 | 2 | 182 | 0 | 0.49 | 96 | 2 | **0** | 91 | 0 | 0 | 0.49 |
| rs199473712 missense | c.866 G>A | G289D | 194 | 2 | 182 | 0 | 0.49 | 96 | 2 | 0 | 91 | 0 | 0 | 0.49 |
| rs75395981 missense | c.871 G>A | D291N | 194 | 2 | 180 | 2 | 1 | 96 | 2 | 0 | 89 | 2 | 0 | 0.99 |
| rs137872696 Non Synonymous | c.1046 C>T | A349V | 195 | 1 | 182 | 0 | 1 | 97 | 1 | 0 | 91 | 0 | 0 | 1 |

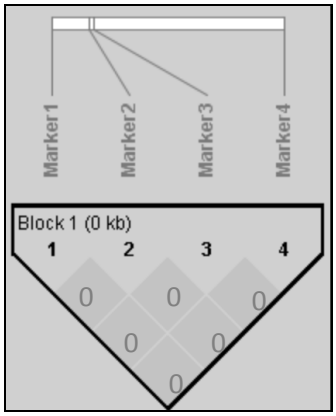
Allele and genotype frequencies of the sequence variations within *VSX2* gene in the current study are shown. The distributions did not deviate from those predicted by the Hardy–Weinberg equilibrium. Proportions of groups were compared by the Fisher's exact test. The criterion for statistical significance was p≤0.05. DNA changes are documented based on cDNA sequences with +1 corresponding to the A of the ATG translation initiation codon in reference to NM_182894.2 (*VSX2*) sequence. (h – ancestral allele, m- mutant allele)

**Fig 3.1.4: Electrophorogram representing the 4 variants** (homozygous and heterozygous forms) in *VSX2*, seen in the present study.A-rs137872696, B-rs7595981, C-62006815, D-rs199473712



Two haplotypes (1 block) were generated by Haploview (Fig 3.1.4, Table 3.1.13a) and five by Thesias softwares (Table 3.1.13b). Haplotype analysis also did not reveal any significant association with myopia or interaction with its endophenotype AL.

**Fig 3.1.5: Haplotype block generated by Haploview software v4.1 for 4 coding variations (rs62006815, rs199473712, rs75395981, rs137872696) observed in exon 5 of *VSX2* gene.**




Markers 1-4 correspond to rs62006815, rs199473712, rs75395981, rs137872696 respectively. The haplotypes were generated using Solid Spine algorithm and the LD measures are indicated as the correlation coefficient ($r^2$). The shades of gray indicate the magnitudes of the $r^2$ values.

**Table 3.1.13 A: Association of *VSX2* haplotypes with myopia as analyzed by Haploview v4.2**

| Haplotype | χ 2 value | # Permutation *P*-value |
|---|---|---|
| GGGC | 1.873 | 0.6040 |
| GGAC | 0.002 | 1.0000 |

#1000 permutations performed.

**Table 3.1.13 B: association of *VSX2* haplotypes with myopia as analyzed by Thesias v3.1**

| Haplotype | Frequency in Controls | Frequency in Cases | Haplotype effect *P* value | OR | Haplotype effect on AL *P* value | OR |
|---|---|---|---|---|---|---|
| GGGC | 0.989 | 0.969 | Intercept | | Intercept | |
| GGGT | 0 | 0.005 | NA | NA | NA | NA |
| GGAC | 0.011 | 0.010 | 0.921 | 0.904 [0.125-6.548] | 0.065 | 2.028 [-0.124 - 4.180] |
| GAGC | 0 | 0.010 | NA | NA | NA | NA |
| AGGC | 0 | 0.005 | NA | NA | NA | NA |

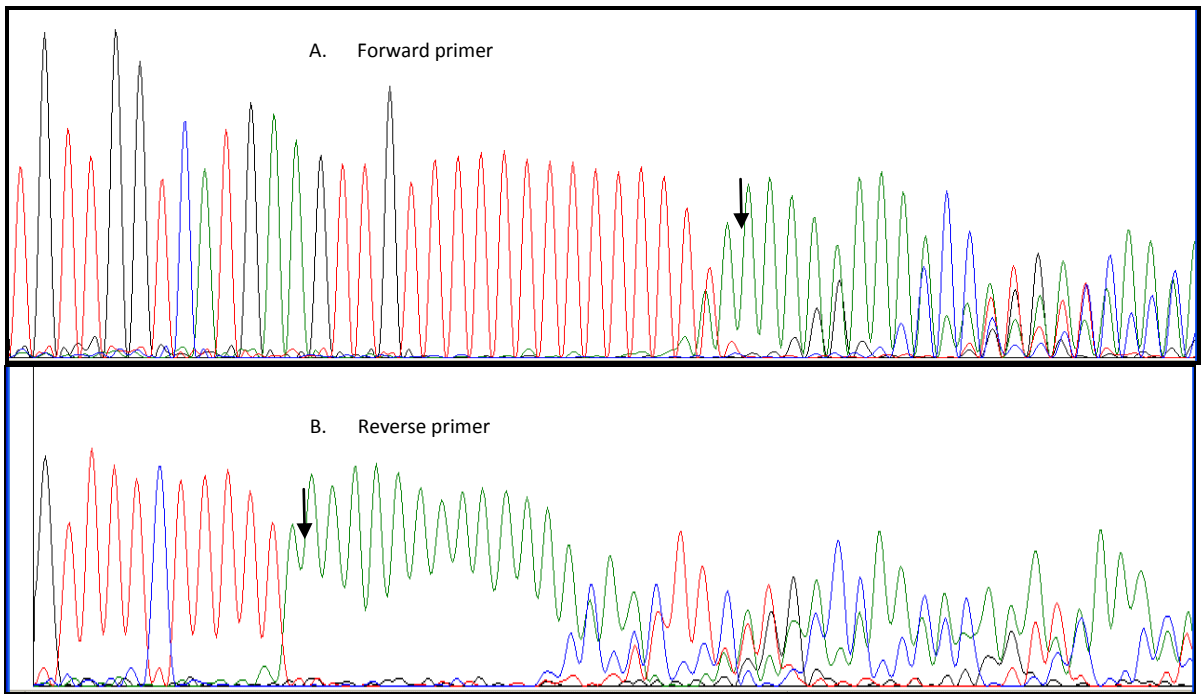
### 3. *PAX6* Results:

Of the 100 cases PCR amplified and genotyped we did not observe any variations for regions P2 and P3. In P1 we observed a known SNP, rs11407950, *NG_008679.1:g.33384_33385insT* in 25 cases and 5 controls (Table 3.1.14, Fig: 3.1.5).

**Table 3.1.14: Genotype and allele frequency distribution of *PAX6* variants in 3'UTR region**

| Region amplified | Cases | Controls | Variations |
|---|---|---|---|
| P1 | 25 | 5 | rs11407950 |
| P2 | 100 | - | Nil |
| P3 | 100 | - | Nil |

**Fig 3.1.6: Electrophorogram of P1 region showing the variation rs11407950 a) with forward primer, b) with reverse primer**




## DISCUSSION

The coordinated growth changes of various components of the eye are required for achieving and maintaining emmetropia of which AL remains as one of the major determinants [230]. Emmetropisation is a process in the postnatal stage of development that regulates the eye growth, whereby the plane of the retina coincides with the focal point of the eye's optical system. An alteration in the visual stimuli initiates a signaling cascade originating in the sensory retina, traversing the retinal pigment epithelium and the vascular

choroid, and ultimately regulating eye growth through scleral remodeling [231-233]. Alterations in the sequence of these biochemical events might disrupt this finely tuned mechanism and lead to refractive errors. Hence, genes like, *MFRP, VSX2* that plays important role in these complex signaling pathways could be studied for the genetic association with diseases manifesting altered AL.

In the current study, we have attempted a comprehensive genotype-haplotype analysis of *MFRP* and *VSX2* gene variants and 3'UTR of *PAX6* for their putative association with myopia and its biometric determinants (AL), first time in Indian population. Fifteen variations were observed in the *MFRP* gene, of which, rs36015759 (T164T) in exon 5 showed negative association with myopia ($P$ =4.1 $e^{-07}$) and decreased AL ($P$= 0.000444, difference in mean value=-1.35085). We did not observe any significant association with *VSX2* variants and analysis of PAX6 3'UTR revealed a single SNP rs11407950 and no other significant variation.

***MFRP***

*MFRP* gene expression studies revealed that the protein is being produced in E11 stage which coincides with the development of the presumptive retina followed by significant increase at birth suggesting a critical role in the development of the anterior segment that eventually determines the eye size or AL [204]. Another rationale for its candidature in AL regulation is the presence of the frizzled domain at the 'C' termini suggesting that it could act as a regulator of the WNT signaling pathway during eye development [203]. Mutational studies in *MFRP* gene have shown its association with hyperopia, nanophthalmos, and microphthalmos [205] [234] and also has been correlated with shift towards both hyperopia and myopia. In a study by Sundin et al, heterozygotes for 1143insC frameshift mutation in *MFRP* exhibited an increased AL [202] Similar observations has been made in other genes like PRSS56 (small eye gene;[182] Hepatocyte growth factor [189], insulin like growth factor [235]. SNPs in these genes are associated with both myopia and hyperopia with either increase or decreased AL.

In the current study we tested for the possible interaction of the MFRP variants with biometric indices like AL by haplotype analysis using THESIAS software. Haplotype CGTCGG, bearing the 'T' allele of rs36015759 was represented at an higher frequency in controls (31%) when compared to cases (8%) and also associated with decreased mean AL

(P= 0.000444, difference in mean value=-1.35085). Since the SNP was represented in a low LD region, we adopted a sliding window (SW) strategy in PLINK analysis which allows an effective evaluation of haplotypic effects for such SNPs [236]. 120 SWs were tested for the 15 MFRP variants identified in the current study. SWs with SNPs of the order, rs883247-rs79836575-rs199473710-rs199473708–rs3814762–rs36015759, ranked first followed by rs36015759-rs2510143-rs61736238 providing evidence for the significant association of rs36015759 ('T'allele), in non myopia controls. This region has been suggested as a mutation hotspot and a deletion of C allele at c.492 position, terminating the protein prematurely after 25 residues, was observed in an Indian family [237]. In recent years, haplotype analysis is being preferred and suggested to have a potential utility in genetic association studies [238].

The synonymous SNP rs36015759 results in loss of binding site for SRp40 protein to the exonic splicing enhancer region for the 'T' allele (FAST SNP analysis; Fig 2). Such concept of improper splicing has been studied earlier in neurofibromatos gene (*NF1*) and functionally proven to result in an non functional protein by skipping of the exon. Similarly a synonymous variation of T255T in *ASB10* gene affected the exon splicing enhancer site and altered the mRNA in lymphoblasts and was associated with primary open angle glaucoma [239]. The haplotype (C**T**GCGG) bearing the **T** allele of rs36015759 could have a similar effect on the expression /function of the gene by loss of splice site enhancer region. Recent review on significance of synonymous variations, suggests a putative effect of such SNPs via improper splicing accuracy, translation fidelity, mRNA structure or protein folding [240] and has given a formulae for measuring the effect of synonymous variations on the translation kinetics, ΔRSCU (Relative synonymous codon usage). In our study ΔRSCU is 0.28 between wild type (TAC=1.14) and variant (TAT=0.86) codon of rs36015759 calculated using CAIcal software[228]. A positive ΔRSCU implies increased translation kinetics, thus providing a putative functional relevance for this SNP.

***VSX2***

In the present study association of *VSX2* with myopia was taken up as studies have established its role in the development of the optic cup and determination of AL. VSX2 protein is a homeodomain class of protein which includes a conserved CVC domain. Most of these proteins are transcription factors. They play an important role in the proliferation

and maintenance of the retinal progenitor cells (RPCs) number. Mutations in the homeodomain / CVC domain have been associated with various developmental syndromes characterized by "small eyes". Study in two consanguineous families with non-syndromic congenital bilateral microphthalmia revealed a change of aminoacid arginine at 227 position to tryptophan [209]; and a recent case study showed a substitution at 223 from alanine to glycine [241] which segregated in a homozygous fashion in microphthalmos. The reason for the development of this phenotype has been attributed to the function of the protein.

In the present study we observed a total of 6 variations, of which rs35435463 was in the HOX domain and rs192712847 in CVC domain and others were present outside the domains. None of the 6 SNPs obtained showed a significant *P* value, which could be attributed to the small sample size.

***PAX6***

*PAX6* belongs to a group of transcription factors containing paired and homeobox DNA-binding domains. It is considered to be the master gene in development of the eye, as it plays an important role during lens induction and retina differentiation [242]. Expression based studies have associated lower levels of this protein with myopia. In RPE cells, a decreased expression of this protein inversely correlated with cell proliferation and direct relation with scleral proliferation [243]; interestingly, this scenario had been earlier observed in experimental animals after inducing myopia [244, 245].

Association of *PAX6* gene variants with myopia, based on GWS, was first demonstrated in Caucasian population [246-248]. The difference in the clinical definition of myopia, sample size, attributed to the non-replication of the statistical significance for the *PAX6* gene variants in different populations. However, redefining the clinical status for myopia based on AL and SPE solved the discrepancies found in these association studies, as pointed out by Zayats et al [249-251].

Recent discovery of miRNA's role in regulation of gene expression have helped scientists in unraveling the reasons behind the susceptibility to several genetic disorders. Study by Chen et al proved the involvement of microRNA-328 in regulating the expression of *PAX6* by binding to rs662702 [243]. Target scan (for the analysis of miRNA binding sequence motifs) results revealed many miRNA binding sites in the 3'UTR of *PAX6*. Hence, in the

current study we screened the 3'UTR (NCBI build: 31811300 to 31810450bp) of *PAX6* gene by direct sequencing to identify the reported / novel variations in miRNA binding sites for disease correlation. However we did not identify any SNPs in the miRNA binding regions except rs11407950 that was present in the non binding region. The rs11407950 has been previously noted by Devin's group in Baltimore, but no functional significance has been reported till date.

**Conclusion**

The results of the current study show that the haplotype (CTGCGG) harboring the T allele of rs36015759 in *MFRP* gene is significantly associated with decreased AL across various statistical methods and multiple correction. The results substantiate the role of *MFRP* in AL development. However, the exact role of *MFRP* in AL development needs to be validated by (i) replication of the results in hyperopia, a condition with decreased axial length (ii) functional analysis, to decipher the pathway which involves these genes in the development of myopia and (iii) identification of the interacting partners with potential role in AL determination.

## Chapter 3.2

## Exclusion mapping of MYP locus in myopia families

**Objectives:**

1. To perform whole genome SNP genotyping using Affymetrix Genechip in families with history of myopia.

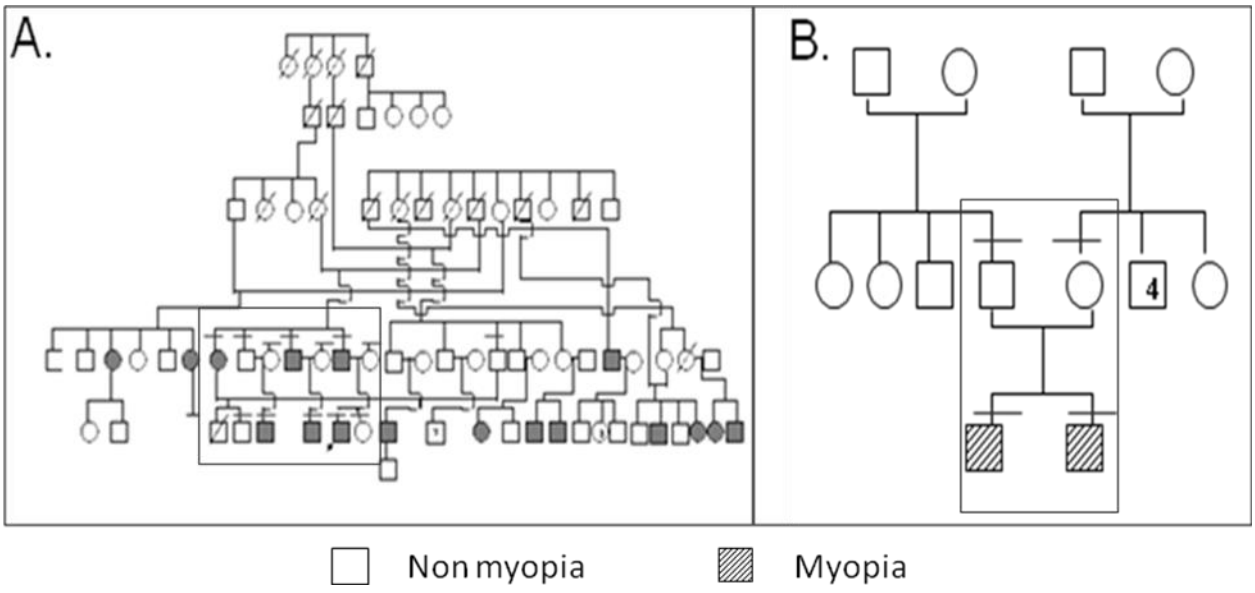
2. To localize the putative chromosomal region after exclusion mapping of the known candidate regions/genes for myopia.

**Materials and Methods:**

**I. Sample recruitment and clinical examination**: (Appendice 1)

**Sample size:** Two families were recruited in the current study (Fig 3.2.1) after detailed ophthalmic examination.

**Fig 3.2.1: Pedigrees recruited in the current study**




Pedigree members of families A and B included under the current study are represented within the square.

**Lab protocols**:

1. **Blood collection**, **DNA extraction- Nucleospin technique, Quantification of the DNA extracted** (Appendice 2-4)

2. **Genotyping using Affymetrix array:** The detailed protocol followed for genotyping using Affymetrix array is given in Appendice 11 and flowchart for the overall steps involved are as mentioned below:

Genomic DNA of 50µg/ µl per sample is digested with restriction enzyme.

↓

Ligation with adaptor

↓

PCR to amplify the fragments

↓

Purification of the amplified products using vacuum manifold.

↓

Quantification of the PCR product and normalization with elution buffer

↓

Fragmentation of the PCR product

↓

Labeling of the fragments with Terminal Deoxynucleotidyl Transferase (TDT)

↓

Hybridization of the labeled products to the affymetrix chip.

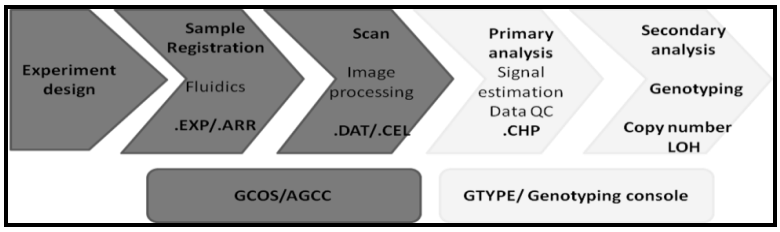
↓

Washing, staining and scanning of the chip

↓

Data analysis

**Data analysis:**




After loading the hybridized chip in Affymetrix fluidics workstation followed by staining and washing, the arrays were scanned in GeneChip®Scanner 3000. Each chip was registered in the GCOS/AGCC software to generate a * .EXP/.ARR file, that has the details of the experiments. After scanning the array, the resulting data were stored as *.DAT (raw tiff image of the hybridised chip) and *.CEL (contains the intensity/position values) files followed by primary and secondary analysis as explained below:

i) Primary analysis: The *.cel files were primarily analysed to assess the experiment quality (QC); the recommended QC cutoff for 10K is >91% and >98% for 250K.

*.CEL file images represent the spatial distribution of the probe intensities on a chip, based on which the quality of signals are analysed and stored as *.CHP files.

ii) Secondary analysis: This step involves specific data analysis such as genotype analysis/ loss of heterozygosity/CNVs. The genotype data can be imported in specific format eg: plink /linkage (ped, map files) for secondary analyses in other softwares like PLINK, MERLIN, etc.

**Linkage analysis for 10K data:**

Linkage analysis was performed using EasyLinkage- FastLink (v4.1) package to calculate two point LOD score at different values of θ assuming autosomal recessive mode of inheritance. The input files included i) *Pedigree file* (.pro): contains the pedigree details in linkage format, ii) *SNP file (.snp):* contains the genotype of the individuals, iii) *loop file (.dat)*: this file contains information of the consanguinity loop, in this case the generation wherein consanguinity was introduced (Appendice 12). Parametric linkage analysis was performed with the following parameters namely, equal male to female recombination rate, 100% penetrance, disease allele frequency of 0.0001 and equal allele frequencies of the genotyped markers. The genotype data were further pruned for Mendelian errors and HWE compliance. Exclusion mapping was performed for the known chromosomal loci for myopia and its QTs. The SNP details for the 10K array markers were retrieved from NetAffyx and the corresponding microsatellite / SNPs for known MYP loci were compared and excluded.

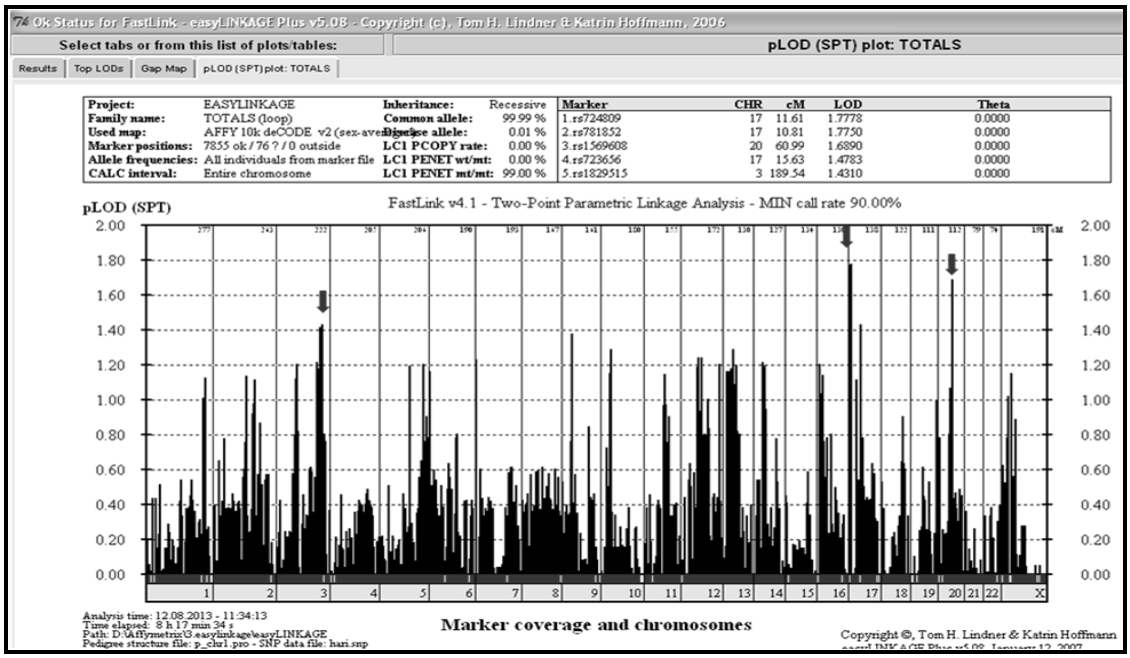
**Homozygosity mapping data:** The genotype data (250K array) of the unaffected and affected members of family B were compared for overlapping regions of homozygosity. Microsoft Office Excel 2007 was used for this purpose. The genotypes were sorted according to chr position in separate sheets and the regions with homozygous blocks (homozygous genotypes for a minimum of 15 consecutive SNPs) in the affected while the being heterozygotes in the unaffected were taken into record.

**Results**

I. **Analysis of 10K Affymetrix array data in family A:** Thirteen members were recruited from family A (affected- 5, unaffected- 8) and were genotyped in Affymetrix 10K array. The QC analysis of genotype data showed >91% for all samples except one (II2, <91% call rate). After QC analysis, genotype data was available for 10204 SNPs, out of which

1425 were uninformative and 965 had low call rate. Excluding all these 7814 SNPs were available for analysis. Pedcheck results further excluded 64 SNPs for Mendelian errors and all the markers were in HWE. Two point linkage analyses were performed for the informative markers by Fastlink and fig 3.2.2. Represents the LOD scores (vertical bars) for all the markers.

**Fig 3.2.2: Plot of total LOD scores across different chrs from linkage analysis using FastLink**




X axis: Maker coverage per chr for the entire genome, Y axis: Parametric LOD score. The arrows represent the first 3 highest LOD scores obtained at rs724809 (chr 17, LOD-1.7778), rs1569608 (chr 20, LOD-1.689) and rs1829515 (chr 3, LOD-1.4) respectively. All markers, with LOD score greater than 0 are represented as vertical bars.

***Exclusion mapping*** was performed for the known myopia or QTL locus. Table 3.2.1 shows the comparison of LOD scores obtained in the current study with the reported literature. The results did not show any significant LOD score suggestive of linkage thus excluding all the known MYP loci.

**Table 3.2.1: Comparison of LOD scores between the current study and reported literature for all the reported myopia loci**

| Myopia locus | Inheritance /QTL | Location | LOD score in previous studies | LOD score in current study | SNP id |
|---|---|---|---|---|---|
| MYP1 | XR | Xq28 | 4.8 | 0 | rs725796 |
| MYP2 | AD | 18p11.31 | 9.59 | 0.37 | rs976655 |
| MYP3 | AD | 12q21-23 | 3.85 | 1.2 | rs724201 |
| MYP5 | AD | 17q21-22 | 3.17 | 0.5 | rs2680398 |

| | | | | | |
|---|---|---|---|---|---|
| MYP6 | AD | 22q12 | 3.54 | 0.38 | rs20037 |
| MYP7 | QTL | 11p13 | 6.10 | 0.1 | rs353599 |
| MYP8 | QTL | 3q26 | 3.7 | 0.35 | rs1920116 |
| MYP9 | QTL | 4q12 | 3.3 | 0.2 | rs1488939 |
| MYP10 | AD & QTL | 8q23 | 4.1 | 0.41 | rs1073913 |
| MYP11 | AD | 4q22-27 | 3.11 | 0.42 | rs2010003 |
| MYP12 | AD | 2q37.1 | 4.75 | 0.18 | rs726017 |
| MYP13 | XR | Xq23-25 | 2.75 | 0 | rs4131866 |
| MYP14 | QTL | 1p36 | 9.54 | 0.43 | rs728611 |
| MYP15 | AD | 10q21.1 | 3.22 | 0.15 | rs1459990 |
| MYP16 | AD | 5p15.33-p15.2 | 4.81 | 0.16 | rs464221 |
| MYP17 | AD | 7p15 | 4.07 | 0.04 | rs1023534 |
| MYP18 | AR | 14q22.1-q24.2 | 2.19 | 1.2 | rs1112330 |
| MYP19 | AD | 5p15.1-p13.3 | 3.02 | 0.21 | rs959643 |
| MYP20 | AD | 13q12.12 | p = 7.75 x 10 (-5) | 0.02 | rs521600 |
| MYP21 | AD | 1p22.2 | 3.19 | 0.18 | rs2390795 |
| MYP22 | AD | 4q35.1 | | 0.18 | rs726466 |

Linkage analysis of the markers after excluding the known myopia loci, showed LOD score suggestive of linkage (-2 to +2) for markers in 3 chromosomal regions (chr 20, 17, and 3) (table 3.2.2).

**Table 3.2.2: List of markers in chrs 20, 17 and 3 and the respective LOD scores suggestive of linkage (-2 to +2) at maximum recombination**

| Chr No | Probe id | rs ID | Marker distance | LOD score | θ |
|---|---|---|---|---|---|
| 20 | SNPA-1518500 | rs910760 | 59.47 | 0 | 0.5 |
| | SNPA-1513114 | rs1569608 | 60.99 | 1.689 | 0 |
| | SNPA-1510071 | rs718092 | 61.06 | 0.2463 | 0.15 |
| | SNPA-1510607 | rs718093 | 61.06 | 0.2387 | 0.15 |
| 17 | SNP_A-1515824 | rs1367950 | 10.09 | 0.0891 | 0 |
| | SNP_A-1517492 | rs781852 | 10.81 | 1.775 | 0 |
| | SNP_A-1508477 | rs724809 | 11.61 | 1.7778 | 0 |
| | SNP_A-1510969 | rs723656 | 15.63 | 1.4783 | 0 |
| 3 | SNP_A-1513554 | rs952770 | 188.82 | 0.4488 | 0 |
| | SNP_A-1514266 | rs262980 | 189.35 | 1.2146 | 0 |
| | SNP_A-1514321 | rs262982 | 189.35 | 1.2032 | 0 |
| | SNP_A-1517816 | rs262981 | 189.35 | 1.2032 | 0 |
| | SNP_A-1517448 | rs1829515 | 189.54 | 1.431 | 0 |
| | SNP_A-1513585 | rs2872249 | 189.58 | 1.2996 | 0 |
| | SNP_A-1518235 | rs727272 | 190.94 | 0.3621 | 0 |

## *Chr 20*

**Fig 3.2.3a: Graph depicting the LOD score of various SNPs in chr 20, the arrow shows rs1569608, (LOD-1.689)**

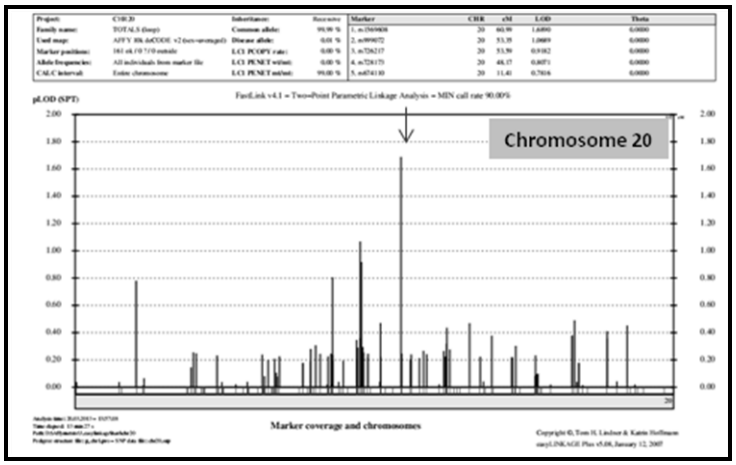
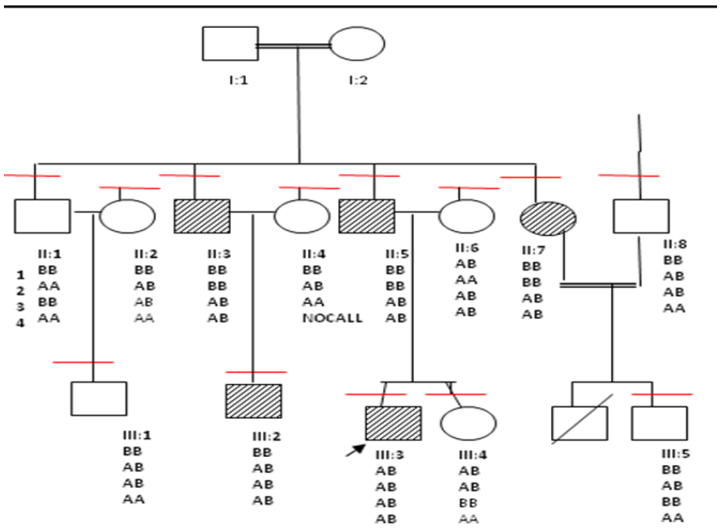



**Fig 3.2.3b: Haplotype segregation of SNPs around rs1569608**




| POS_cM | LOD_MAX | THETA_AT_MAX | rS ID | A/B |
|---|---|---|---|---|
| 59.47 | 0 | 0.5 | rs910760 | A/C |
| 60.99 | 1.6 | 0 | rs1569608 | C/T |
| 61.06 | 0.2 | 0.15 | rs718092 | C/G |
| 61.06 | 0.2 | 0.15 | rs718093 | A/C |

A maximum LOD score of 1.6 was observed for rs1569608 (table 3.2.2) (fig 3.2.3a), but the flanking SNPs rs910760, rs718092, rs718093 did not reveal evidence of linkage at θ max was =0.5 and 1.5 respectively. Segregation analysis showed rs1569608 to be non

segregating among the affected in third generation (III2 and III3) (fig 3.2.3b); hence this region was excluded.

## ***Chr 17***

**Fig 3.2.4a: graph depicting the LOD score of various SNPs in chr 17, the arrow shows rs724809 (LOD: 1.7778)**

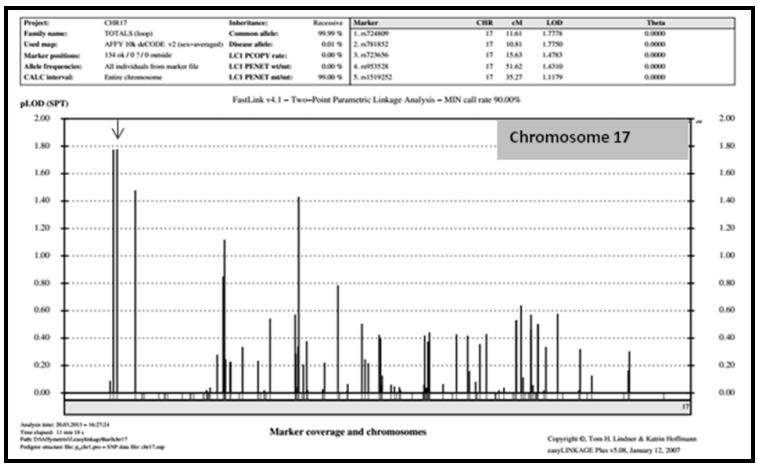
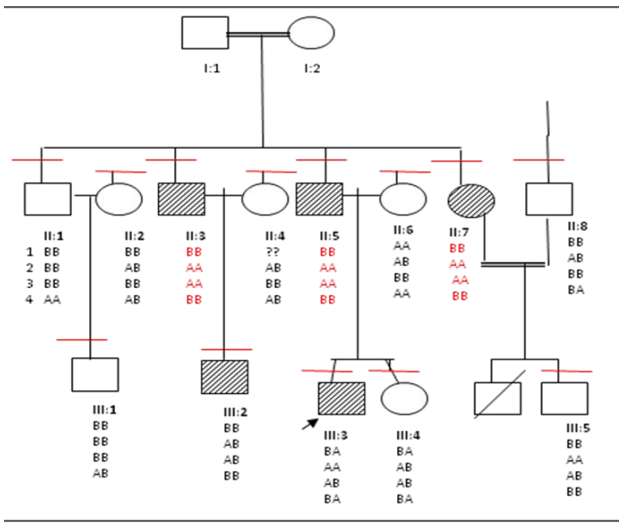



**Fig 3.2.4b: Haplotype segregation of SNPs around rs724809**




| POS_cM | LOD_MAX | THETA_AT_MAX | rS ID | A/B |
|---|---|---|---|---|
| 10.09 | 0.0891 | 0 | rs1367950 | C/T |
| 10.81 | 1.775 | 0 | rs781852 | C/T |
| 11.61 | 1.7778 | 0 | rs724809 | A/G |
| 15.63 | 1.4783 | 0 | rs723656 | C/T |

The next region with significant LOD score of 1.7778 was obtained for SNP rs72480 in chr17, (table 3.2.2 and fig 3.2.4a). Information from flanking SNPs rs1367950, rs781852,

rs723656 showed segregation of genotypes only among affected in second generation (II3, II5, II7- affected and II1-unaffected), but not among the third generation (III2, III3 – affected) (fig 3.2.4b).

## ***Chr 3***

**Fig 3.2.5a Graph depicting the various SNPs in chr 3, the arrow shows rs1829515 (LOD-1.4)**

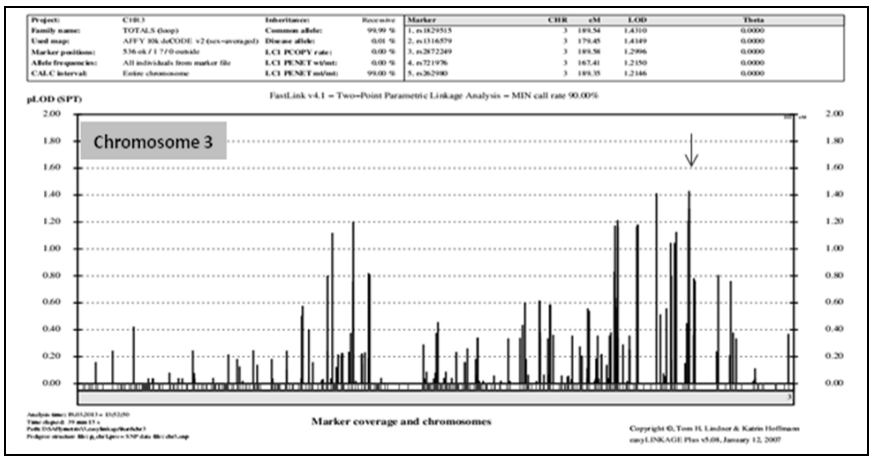
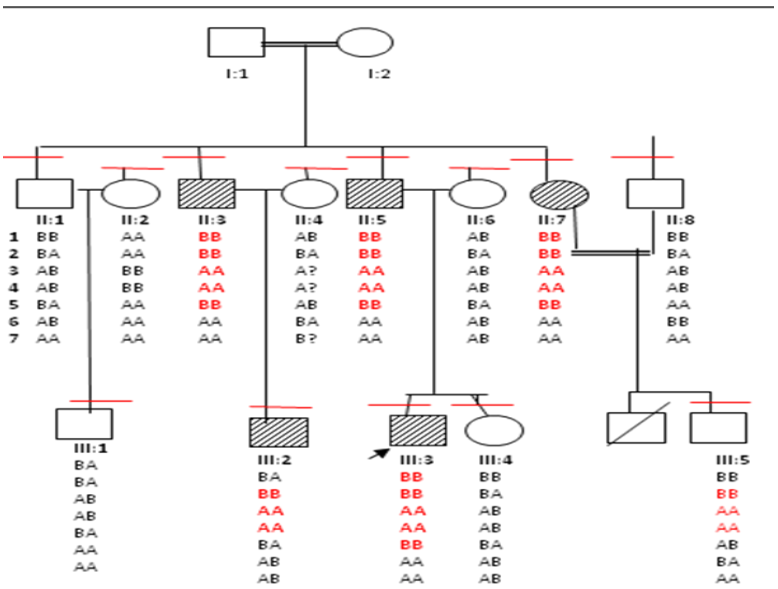



**Fig 3.2.5b: Haplotype segregation of SNPs around rs1829515**




| Pos cM | LOD score | θ | rs ID | A/B |
|---|---|---|---|---|
| 188.82 | 0.4488 | 0 | rs952770 | A/G |
| 189.35 | 1.2146 | 0 | rs262980 | C/T |
| 189.35 | 1.2032 | 0 | rs262982 | C/T |
| 189.35 | 1.2032 | 0 | rs262981 | C/T |
| 189.54 | 1.431 | 0 | rs1829515 | C/T |
| 189.58 | 1.2996 | 0 | rs2872249 | A/T |
| 190.94 | 0.3621 | 0 | rs727272 | C/T |

The third region with significant LOD score was in chr3 rs1829515 (LOD-1.4). Except for III2, all others showed segregation of genotypes for the flanking SNPs rs262980, rs262982, rs262981, rs2872249, and rs727272 (fig 3.2.4).

Additionally, genotyping was performed for 2 members (one unaffected (II1) and affected (II3) in Affymetrix 250K array with an aim to compare the alleles by loss homozygosity. The results (table 3.2.3) however did not suggest any significant loss of homozygosity.

**Table 3.2.3: Comparison of genotypes of SNPs in 250K between II1 and II3 individuals**

| CHR | 10K probe ID | 250K probe ID | II3 | II1 | rsID |
|---|---|---|---|---|---|
| 3 | SNP_A-1513554 to SNP_A-1517816 | SNP_A-4209885 | BB | BB | rs10513794 |
| | | SNP_A-2112100 | AA | AA | rs16857887 |
| | | SNP_A-4237229 | AA | AA | rs2293031 |
| | | SNP_A-2179589 | NoCall | AA | rs11917105 |
| | | SNP_A-1818250 | BB | BB | rs11920784 |
| | | SNP_A-2128661 | BB | BB | rs9882822 |
| | | SNP_A-2104235 | AA | AA | rs10937144 |
| | | SNP_A-1867808 | AA | AA | rs9826007 |
| | | SNP_A-1838407 | AA | AB | rs7619431 |
| | | SNP_A-2219396 | AB | AB | rs262990 |
| | | SNP_A-2189424 | AA | AB | rs263029 |
| | | SNP_A-1876442 | AA | AB | rs263028 |
| | SNP_A-1517448 to SNP_A-1513585 | SNP_A-2284634 | AB | AB | rs9844777 |
| 17 | SNP_A-1517492 | SNP_A-2218912 | AB | AB | rs4790161 |
| | | SNP_A-2116835 | AA | AA | rs781861 |
| | | SNP_A-1834018 | AB | NoCall | rs2176324 |
| | | SNP_A-1866792 | AA | BB | rs781850 |
| | | SNP_A-1794240 | BB | AB | rs781848 |
| | | SNP_A-1799197 | AA | BB | rs781845 |
| | | SNP_A-1813317 | AA | AA | rs781844 |
| | | SNP_A-2297332 | AA | AA | rs781843 |
| | | SNP_A-1806809 | BB | BB | rs2271717 |
| | | SNP_A-1834019 | BB | NoCall | rs781841 |
| | | SNP_A-2173004 | AB | AA | rs16953678 |
| | | SNP_A-4203100 | BB | BB | rs17176065 |
| | | SNP_A-2182419 | AB | BB | rs12947597 |
| | | SNP_A-1834049 | AB | AA | rs10521129 |
| | | SNP_A-1834051 | AB | AA | rs9890881 |
| | | SNP_A-2100419 | BB | AA | rs1377806 |
| | | SNP_A-4232085 | NoCall | AA | rs4467121 |
| | | SNP_A-4205424 | NoCall | NoCall | rs7219998 |
| | | SNP_A-2054241 | BB | BB | rs7220352 |
| | | SNP_A-2277341 | AB | BB | rs7218574 |

| | | | | | |
|---|---|---|---|---|---|
| | | SNP_A-1872298 | BB | BB | rs11078484 |
| | | SNP_A-2243392 | BB | AA | rs17176322 |
| | SNP_A-1508477 | SNP_A-4236325 | BB | BB | rs7211542 |
| | | SNP_A-4220963 | NoCall | NoCall | rs12601020 |
| | | SNP_A-1948729 | BB | AA | rs12601491 |
| | | SNP_A-4203247 | AB | BB | rs11078512 |
| | | SNP_A-2106028 | AA | BB | rs1038121 |
| | | SNP_A-4239978 | NoCall | NoCall | rs1038118 |
| | | SNP_A-2059428 | AB | AA | rs11654457 |
| **CHR** | **10K probe ID** | **250K probe ID** | **II3** | **II1** | **rsID** |
| **17** | SNP_A-1510969 | SNP_A-2079853 | AA | AA | rs7502609 |
| 20 | SNP_A-1513114 | SNP_A-4229341 | AA | AA | rs6101483 |
| | | SNP_A-4206419 | AA | BB | rs4812381 |
| | | SNP_A-2235856 | NoCall | BB | rs16987952 |
| | | SNP_A-1968320 | BB | AA | rs2206437 |
| | | SNP_A-1849446 | AA | AA | rs6101486 |
| | | SNP_A-1792359 | AA | BB | rs932385 |
| | | SNP_A-2092103 | BB | AB | rs959415 |
| | | SNP_A-2221686 | AB | AB | rs17188356 |
| | | SNP_A-1935148 | AA | BB | rs6101503 |
| | | SNP_A-1789415 | BB | BB | rs6093170 |
| | | SNP_A-4200417 | BB | BB | rs7263998 |
| | | SNP_A-2066303 | AB | AB | rs16987984 |

The yellow shaded cells represent regions compiling with loss of homozygosity rule.

**Family B - Analysis by Homozygosity mapping**:

The study subjects in family B were only four; hence linkage analysis was not performed. Therefore Homozygosity mapping (HM) was used which is an efficient method of choice for autosomal recessive disorders. There were 38 homozygous regions in family B in which homozygosity mapping was attempted. Homozygous regions were defined as genomic regions with homozygous genotype for 15 or more SNPs consecutively in the proband and his affected sib while the parents were heterozygous for the same. The presence of ROH in candidate regions for myopia and its QTs like AL, RE, SPH etc was then analysed. The analysis revealed a homozygous block of 151.38KB (227471488 - 227622869) in chr 2q36.3 (table 3.2.4) shared between the 2 affected sibs and the parents were observed to be heterozygous for the same.

**Table 3.2.4: Region on chr2q36.3 narrowed down using homozygosity mapping.**

| Affy SNP ID | Father | Mother | Son1 | Son2 | SNP ID | CHR | Base pair |
|---|---|---|---|---|---|---|---|
| SNP_A-2242377 | AB | AB | AA | AA | rs10197757 | 2 | 227471488 |
| SNP_A-1882954 | NoCall | AB | BB | BB | rs4675117 | 2 | 227478038 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SNP_A-1806794 | AB | AB | BB | BB | rs6723774 | 2 | 227478836 |
| SNP_A-2087965 | AB | AB | AA | AA | rs17416354 | 2 | 227479088 |
| SNP_A-1967228 | AB | AB | AA | AA | rs3820928 | 2 | 227481710 |
| SNP_A-1785281 | AB | AB | BB | BB | rs4675133 | 2 | 227551254 |
| SNP_A-2135803 | BB | BB | BB | BB | rs13031337 | 2 | 227554669 |
| SNP_A-2108273 | AB | AB | BB | BB | rs4461255 | 2 | 227554768 |
| SNP_A-2137601 | AB | AB | AA | AA | rs4281898 | 2 | 227563737 |
| SNP_A-2100293 | AB | AB | BB | BB | rs1054413 | 2 | 227575963 |
| SNP_A-1957222 | AB | AB | BB | BB | rs4423583 | 2 | 227583201 |
| SNP_A-1784011 | AB | AB | BB | BB | rs6740108 | 2 | 227597803 |
| SNP_A-2081066 | AB | AB | AA | AA | rs1949807 | 2 | 227600105 |
| SNP_A-4202457 | AB | AB | BB | BB | rs2272198 | 2 | 227601261 |
| SNP_A-2264837 | AB | AB | AA | AA | rs10176882 | 2 | 227604676 |
| **Affy SNP ID** | **Father** | **Mother** | **Son1** | **Son2** | **SNP ID** | **CHR** | **Base pair** |
| SNP_A-2208061 | AB | AB | BB | BB | rs13423714 | 2 | 227604896 |
| SNP_A-2039294 | AA | AA | AA | AA | rs3754846 | 2 | 227619905 |
| SNP_A-4202110 | AB | AB | AA | AA | rs1317770 | 2 | 227619969 |
| SNP_A-1826760 | AB | AB | AA | AA | rs10190621 | 2 | 227622728 |
| SNP_A-2071997 | AB | AB | AA | AA | rs10178202 | 2 | 227622805 |
| SNP_A-2121686 | AB | AB | BB | BB | rs2272202 | 2 | 227622869 |

## DISCUSSION

Whole genome study using a high throughput platform to cover the entire genome is one of the greatest achievements to the scientific society. The key aspect in the efficiency of GWS lies in the indirect association, where only a subset of SNPs is genotyped, taking advantage of the LD between SNPs. With respect to mapping loci for myopia using this technology several studies have been conducted in various populations, adding more knowledge to the molecular basis of the disease.

Several GWAS for myopia have been conducted which includes: UK twin cohort (15q25, rs8027411, P value=$7.91*10^{-8}$)[252], Dutch population (15q14) [253], Chinese and Japanese (*CTNND2*, 5p15) [254], Han Chinese population (BI480957, MYP11 locus) [255], (rs9318086, 13q12.12) [256]. Few GWAS for the endophenotypes of myopia have resulted in successful identification of loci which includes 1q41 (AL) [40], *PDGFRA* and *TRIM29* [192] (CC), MYP7-MYP10 (SE and SPH) [246].

Linkage analysis in dizygotic twins / high myopic families has identified several MYP loci (MYP7 –MYP10 and MYP 14) [246]. A regression based linkage analysis on Ashkenazi Jewish families identified a region about 49.1cM on chr 1p36 with a maximum score of 9.5 for ocular refraction[181]. Other linkage based studies in Chinese (14q22.1-q24.2; [257],

15q12-13; [178], 5p15.33-p15.2; [179]), African-American and White families (chr 20) [176], Older Amish families (chr 8q23) [177].

In the present study we identified two families with high myopia and performed high throughput genotyping using 10K and 250K Affymetrix gene chip technology. In family A, 13 members were included (5 affected and 8 unaffected) and 2 point linkage analysis excluded linkage to known candidate regions of myopia and its QTs. Significant LOD score suggestive of linkage was observed in 3 chromosomal regions 3q27.1 (LOD max of 1.2 at $\theta$ =0), chr 17p13.3 (LOD max of 1.7 $\theta$ =0) and chr 20q12 (LOD max of 1.6 at $\theta$ =0). Chr 20 was ruled out of linkage as the flanking SNPs did not reach suggestive LOD score at $\theta$ =0. Chromosomal regions at 3 and 17 showed evidence of linkage but the haplotype did not segregate among the affected of third generation, thus resulting in inconclusive report. The possible reason for this could be small size of only 13 members; presence of consanguinity, and usage of less dense markers (10K array). Thus inclusion of the expanded family members and using more dense markers or exome sequencing might increase the LOD score and help in narrowing down the causative gene. The candidate genes present in chr 3 included *MCF2L2, YEATS2, ABCC5, PARL, MAGEF1;* of these *ABCC5* is a prospective gene to be evaluated as its shown to be expressed in retina [258] and has also been associated with ACD [259]. In chr 17 the candidate genes included *ZZEF1, WSCD1,* and *NLRP1* but none of them have been earlier studied in eye related tissues. Targeted resequencing of these chromosomal regions in chrs 17 and 3 might provide more evidence to the fact whether they are truly linked with the disease loci.

The second family was genotyped using 250K Affymetrix array chip, and analyzed using homozygosity mapping technique. We observed a homozygous block in chr 2q36.3, a region which had been earlier mapped to myopia by Paluru et al in an US family [260]. This region falls within the broad region described by them but away from the narrowed down region. It contains *COL4A4* a possible candidate gene in myopia. The gene mediates remodelling of extracellular matrix collagen. Variations in this gene have been associated with a syndrome called Alprot syndrome (OMIM: 203780) [261], which is characterized by haematuria, deafness and the ocular component involves myopia. The probands in the current study do not exhibit these symptoms, which was confirmed by a general questionnaire but further clinical evaluation is warrentied. Additional evaluation of the gene for variations can yield a better clue into its actual role in the development of the disease.

Thus evaluation of the two pedigrees of myopia for candidate loci/genes yielded 3 regions on chr3, chr7 (family 1) and chr2q36.3 (family2). The results obtained from the current study can be viewed as an initial step in identification of novel loci / gene for myopia in our cohort. Further probing into these regions of interest by fine mapping or Exome sequencing can yield a better insight into the variations in the candidate genes and their role in the development of the disease.

## Chapter 4

## DIGEST

### Summary

The present study was undertaken with an aim to identify the association of known QTLs for ocular traits namely AL, CCT, IOP, and VCDR in Indian population. To accomplish this aim subjects were included from 3 cohorts namely (a) multigeneration consanguineous families without any history of ocular disorders, (b) POAG, and (c) myopia. The QTL for (i) AL, CCT, IOP and VCDR were studied in cohorts a and b (ii) AL was studied in the myopia cohort. The SNPs of candidate QTLs, earlier identified in different population, were genotyped and analysed for statistical significance.

A novel approach of mapping QTs in consanguineous families was undertaken to exploit the advantages of both QTL mapping as well as high representation of alleles in consanguineous families. By this approach we were able to replicate the association of rs4746741, rs1900004 and rs3858145 all present near *ATOH7* gene with disc area and rs4746741 with cup area.

In POAG, a case-control study design was adopted and SNPs rs1900004, rs3858145 (*ATOH7),* rs10483727 (*SIX1/SIX6),* rs1192415 (*CDC7),* rs1063192 (*CDNK2B)* and rs9607469 (*CARD10)* earlier associated with glaucoma endophenotypes (VCDR, CCT and IOP) were analysed. Significant association was obtained for i) rs1192415 (CDC7) with decreased CCT and increased VCDR ii) rs1900004 (ATOH7) with decreased VCDR and increased AL iii) rs1063192 (CDKN2B) with decreased AL iv) rs10483727 (SIX1/SIX6) with decreased VCDR and v) rs9607469 (CARD10) with decreased VCDR. Gene-gene interaction studies revealed significant interaction between SIX6-CDKN2B, SIX6-CARD10 and ATOH7-SIX6.

In POAG, a case-control study design was adopted and SNPs rs1900004, rs3858145 (*ATOH7),* rs10483727 (*SIX1/SIX6),* rs1192415 (*CDC7),* rs1063192 (*CDNK2B)* and rs9607469 (*CARD10)* earlier associated with glaucoma endophenotypes (VCDR, CCT and IOP) were analysed. Significant association was obtained for (i) rs1900004, rs3858145, rs10483727 and rs9607469 with VCDR (ii) rs10483727 with IOP, iii) rs10483727 and rs1192415 with CCT and (iv) rs1063192 with AL. Interaction studies revealed a significant interaction between genotypes of *ATOH7* (rs1900004) and *CDC7* (rs1192415) in HTG.

We also observed an earlier reported interaction between *ATOH7* (rs1900004, CC/CT) and *SIX1/SIX6* (rs10483727, CC) in NTG.

Study of association of rs1048661 and rs3825942 in exon 1 of *LOXL1* was done in a cohort of XFS/XFG cases and controls. We obtained a significant association of both the SNPs with the disease. When evaluating for its endophenotypes in the whole cohort we observed significant association of rs3825942 with IOP and VCDR. Analysis in cases and controls separately revealed significant association of IOP with rs1048661 and rs3825942 (controls). The rs3825942 was also associated with CCT and VCDR in XFS and XFG respectively.

In myopia, 3 candidate genes implicated in AL development (M*FRP, VSX2,* 3' region of *PAX6)* were screened by direct sequencing. Out of the fifteen variations observed in *MFRP* gene, rs36015759 was significantly associated with low risk for myopia. Haplotype analysis indicated an increased frequency for the coding haplotype block [CGTCGG] including rs36015759 in controls (decreased AL) (31%) than cases (increased AL) (8%). Further bioinformatic analysis (FASTSNP) revealed that the 'T' allele of rs36015759 abolished the binding site for splicing enhancer (SRp40). In addition to this, array based whole genome analysis was performed in 2 myopic families. Through linkage analysis and homozygosity mapping we excluded known QTLs of myopia as well as other MYP locus. Further evaluation of these two myopia pedigrees revealed 3 possible candidate regions on chr3q27.1, chr17p13.2 (family A) and chr2q36.3 (family B). In family A though LOD score was suggestive of linkage the haplotype segregation was inconclusive as they did not segregate among the third generation and in family B further validation of chr2q36.3 locus is necessary with exome sequencing and inclusion other family members.

**Table 4.1: Complete list of associations of SNPs and QTs seen in the present study in comparison with previous reports**

| rs id | Gene | QT | Effect | Samples | Case-control cohort | Reference | Direction of effect in comparison to previous |
|---|---|---|---|---|---|---|---|
| rs4746741 | *ATOH7* | Disc area | Decreased | Consanguineous families | NA | Novel | NA |
| rs4746741 | *ATOH7* | Cup area | Decreased | Consanguineous families | NA | Novel | NA |
| rs1900004 | *ATOH7* | AL | Increased | Whole cohort | POAG | Novel | NA |
| rs1900004 | *ATOH7* | IOP | Decreased | HTG | POAG | Novel | NA |
| rs1192415 | *CDC7* | CCT | Decreased | Whole cohort | POAG | Novel | NA |
| rs1063192 | *CDKN2B* | AL | Decreased | Controls | POAG | Novel | NA |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs1048661 | *LOXL1* | IOP | Decreased | Controls | XFS/XFG | Novel | NA |
| rs3825942 | *LOXL1* | IOP | Decreased | Whole cohort | XFS/XFG | Novel | NA |
| rs3825942 | *LOXL1* | IOP | Increased | Controls | XFS/XFG | Novel | NA |
| rs3825942 | *LOXL1* | CCT | Increased | XFG | XFS/XFG | Novel | NA |
| rs3825942 | *LOXL1* | VCDR | Decreased | Whole cohort | XFS/XFG | Novel | NA |
| rs3825942 | *LOXL1* | VCDR | Increased | XFS | XFS/XFG | Novel | NA |
| rs36015759 | *MFRP* | AL | Decreased | | Myopia | Novel | NA |
| rs3858145 | *ATOH7* | Discar | Decreased | Consanguineous families | NA | Increased | Opposite |
| rs10483727 | *SIX1/SIX6* | VCDR | Decreased | NTG | POAG | Ozel 2013 | Opposite allele |
| rs10483727 | *SIX1/SIX6* | VCDR | Decreased | combined cases | POAG | Ramdas 2010, and 2011 | Opposite allele |
| rs1900004 | *ATOH7* | Discar | Decreased | Consanguineous families | NA | Ramdas et al 2010 | Similar |
| rs1192415 | *CDC7* | VCDR | Increased | Whole cohort | POAG | Ramdas et al 2010 | Similar |
| rs9607469 | *CARD10* | VCDR | Decreased | HTG | POAG | Khor 2011 | Similar |
| rs1900004 | *ATOH7* | VCDR | Decreased | Whole cohort, controls | POAG | Ramdas et al 2010 | Similar |

## Specific contributions:

The data generated from the present study has helped in the identification/replication of associations between the genes and the disease and/ its endophenotypes in Indian population. Some of the specific contributions include:

1. This is the first study in Indian population to evaluate the association of variations in/near genes like (i) *ATOH7*, *SIX1/SIX6, CARD10,* and *CDC7* with POAG ii) *VSX2* and *MFRP* with myopia and iii) to adopt genome wide technique for mapping loci in myopic families.
2. First study in the literature to evaluate association of SNPs rs3825942 and rs1048661 in *LOXL1* gene with the endophenotypes of glaucoma.
3. First study to observe significant interaction between genotypes of *ATOH7* (rs1900004) and *CDC7* (rs1192415) in HTG.
4. Negative association of rs36015759 (*MFRP)* with myopia was reported for the first time.

## Future scope

Results of the present study throw light on the genetic pathology of these two complex disorders glaucoma (POAG and XFG) and myopia in Indian scenario. Further

substantiation at the cellular level is warrentied to specifically identify the pathway behind the development of the disease. The thought that all these variations identified in the current study can be mere markers and lie in LD with the true candidate gene should not be neglected. The limitations include low sample size and hence replication in another cohort will help in substantiating the results obtained. Apart from these genes several other genes also needs to be evaluated to understand the complex genetic nature of these disorders. High throughput experiments like whole exome and next generation sequencing might hold the key to unlock the mysteries of these diseases. The newly expanding fields of metabolomics, miRNA, and epistasis might help in deciphering the susceptible nature of the individuals. Thus studying genetics of these two disorders helps in identification of individuals at risk and provide them with appropriate management modalities to prevent/ manage the disorder. Since environment also plays an important role in the development of myopia and glaucoma interaction between them should also be considered significant.

## References:


1. Manolio, T.A., L.D. Brooks, and F.S. Collins, *A HapMap harvest of insights into the genetics of common disease.* J Clin Invest, 2008. **118**(5): p. 1590-605.

2. WHO *Visual impairment and blindness*.

3. Vijaya, L., et al., *Prevalence and causes of low vision and blindness in an urban population: The Chennai Glaucoma Study.* Indian J Ophthalmol, 2013.

4. Bruce Birren, E.D.G., *Genome analysis - A laboratory manual*. MApping Genomes, ed. Eric.Green. Vol. 4. 1997: Cold Spring Harbor Laboratory Press.

5. Tom Strachan, A.P.R., *Human molecular Genetics*. Vol. 3. 2003: Garland Science.

6. OpenStax, c., *Biology*, in *Laws of inheritance*. 2014.

7. Khawaja, A.P., et al., *Corneal biomechanical properties and glaucoma-related quantitative traits in the EPIC-Norfolk Eye Study.* Invest Ophthalmol Vis Sci, 2014. **55**(1): p. 117-24.

8. Kass, M.A., et al., *The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma.* Arch Ophthalmol, 2002. **120**(6): p. 701-13; discussion 829-30.

9. Leske, M.C., et al., *Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial.* Arch Ophthalmol, 2003. **121**(1): p. 48-56.

10. Keating, M., et al., *Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey ras-1 gene.* Science, 1991. **252**(5006): p. 704-6.

11. Zhang, Y., et al., *Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma.* Nat Genet, 2003. **34**(2): p. 181-6.

12. Vitart, V., et al., *SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout.* Nat Genet, 2008. **40**(4): p. 437-42.

13. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts.* Nat Genet, 2009. **41**(1): p. 47-55.

14. Majumder, P.P. and S. Ghosh, *Mapping quantitative trait loci in humans: achievements and limitations.* J Clin Invest, 2005. **115**(6): p. 1419-24.

15. Bishop, D.T. and J.A. Williamson, *The power of identity-by-state methods for linkage analysis.* Am J Hum Genet, 1990. **46**(2): p. 254-65.

16. Verhoeven, K.J., J.L. Jannink, and L.M. McIntyre, *Using mating designs to uncover QTL and the genetic architecture of complex traits.* Heredity (Edinb), 2006. **96**(2): p. 139-49.

17. Alarcon, M. and R.M. Cantor, *Quantitative trait loci mapping of serum IgE in an isolated Hutterite population.* Genet Epidemiol, 2001. **21 Suppl 1**: p. S224-9.

18. Falchi, M., et al., *A genomewide search using an original pairwise sampling approach for large genealogies identifies a new locus for total and low-density lipoprotein cholesterol in two genetically differentiated isolates of Sardinia.* Am J Hum Genet, 2004. **75**(6): p. 1015-31.

19. Abney, M., C. Ober, and M.S. McPeek, *Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites.* Am J Hum Genet, 2002. **70**(4): p. 920-34.

20. Hildebrandt, F., et al., *A systematic approach to mapping recessive disease genes in individuals from outbred populations.* PLoS Genet, 2009. **5**(1): p. e1000353.

21. Thompson, D., et al., *Haplotype tagging single nucleotide polymorphisms and association studies.* Hum Hered, 2003. **56**(1-3): p. 48-55.

22. Zhang, K., et al., *Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies.* Genome Res, 2004. **14**(5): p. 908-16.

23. Project, T.I.H., Nature, 2003. **426**(6968): p. 789-96.

24. Carlson, C.S., et al., *Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium.* Am J Hum Genet, 2004. **74**(1): p. 106-20.

25. Need, A.C. and D.B. Goldstein, *Whole genome association studies in complex diseases: where do we stand?* Dialogues Clin Neurosci, 2010. **12**(1): p. 37-46.

26. Gordon, R.A. and P.B. Donzis, *Refractive development of the human eye.* Arch Ophthalmol, 1985. **103**(6): p. 785-9.

27. Brown, N.P., J.F. Koretz, and A.J. Bron, *The development and maintenance of emmetropia.* Eye (Lond), 1999. **13 ( Pt 1)**: p. 83-92.

28. Lyhne, N., et al., *The importance of genes and environment for ocular refraction and its determiners: a population based study among 20-45 year old twins.* Br J Ophthalmol, 2001. **85**(12): p. 1470-6.

29. Biino, G., et al., *Ocular refraction: heritability and genome-wide search for eye morphometry traits in an isolated Sardinian population.* Hum Genet, 2005. **116**(3): p. 152-9.

30. Zhu, G., et al., *Genetic dissection of myopia: evidence for linkage of ocular axial length to chromosome 5q.* Ophthalmology, 2008. **115**(6): p. 1053-1057 e2.

31. Nakanishi, H., et al., *A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1.* PLoS Genet, 2009. **5**(9): p. e1000660.

32. Puk, O., et al., *Variations of eye size parameters among different strains of mice.* Mamm Genome, 2006. **17**(8): p. 851-7.

33. Bardakjian, T., A. Weiss, and A.S. Schneider, *Anophthalmia / Microphthalmia Overview.* 1993.

34. Sorsby, A. and G.A. Leary, *A longitudinal study of refraction and its components during growth.* Spec Rep Ser Med Res Counc (G B), 1969. **309**: p. 1-41.

35. Wiesel, T.N. and E. Raviola, *Myopia and eye enlargement after neonatal lid fusion in monkeys.* Nature, 1977. **266**(5597): p. 66-8.

36. Chen, H., W.N. Howald, and M.R. Juchau, *Biosynthesis of all-trans-retinoic acid from all-trans-retinol: catalysis of all-trans-retinol oxidation by human P-450 cytochromes.* Drug Metab Dispos, 2000. **28**(3): p. 315-22.

37. Honda, S., et al., *Retinal control on the axial length mediated by transforming growth factor-beta in chick eye.* Invest Ophthalmol Vis Sci, 1996. **37**(12): p. 2519-26.

38. Fischer, A.J., et al., *Light- and focus-dependent expression of the transcription factor ZENK in the chick retina.* Nat Neurosci, 1999. **2**(8): p. 706-12.

39. Cheng, C.Y., et al., *Nine loci for ocular axial length identified through genome-wide association studies, including shared loci with refractive error.* Am J Hum Genet, 2013. **93**(2): p. 264-77.

40. Fan, Q., et al., *Genetic variants on chromosome 1q41 influence ocular axial length and high myopia.* PLoS Genet, 2012. **8**(6): p. e1002753.

41. Seiler, T. and J. Wollensak, *The resistance of the trabecular meshwork to aqueous humor outflow.* Graefes Arch Clin Exp Ophthalmol, 1985. **223**(2): p. 88-91.

42. Charlesworth, J.C., et al., *Linkage to 10q22 for maximum intraocular pressure and 1p32 for maximum cup-to-disc ratio in an extended primary open-angle glaucoma pedigree.* Invest Ophthalmol Vis Sci, 2005. **46**(10): p. 3723-9.

43. Nag, A., et al., *A genome-wide association study of intra-ocular pressure suggests a novel association in the gene FAM125B in the TwinsUK cohort.* Hum Mol Genet, 2014. **23**(12): p. 3343-8.

44. *Genome-wide association study of intraocular pressure identifies the GLCCI1/ICA1 region as a glaucoma susceptibility locus.* Hum Mol Genet, 2013. **22**(22): p. 4653-60.

45. van Koolwijk, L.M., et al., *Common genetic determinants of intraocular pressure and primary open-angle glaucoma.* PLoS Genet, 2012. **8**(5): p. e1002611.

46. Ozel, A.B., et al., *Genome-wide association study and meta-analysis of intraocular pressure.* Hum Genet, 2013.

47. Lu, Y., et al., *Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus.* Nat Genet, 2013. **45**(2): p. 155-63.

48. Vithana, E.N., et al., *Collagen-related genes influence the glaucoma risk factor, central corneal thickness.* Hum Mol Genet, 2011. **20**(4): p. 649-58.

49. Vitart, V., et al., *New loci associated with central cornea thickness include COL5A1, AKAP13 and AVGR8.* Hum Mol Genet, 2010. **19**(21): p. 4304-11.

50. Cornes, B.K., et al., *Identification of four novel variants that influence central corneal thickness in multi-ethnic Asian populations.* Hum Mol Genet, 2012. **21**(2): p. 437-45.

51. Hoehn, R., et al., *Population-based meta-analysis in Caucasians confirms association with COL5A1 and ZNF469 but not COL8A2 with central corneal thickness.* Hum Genet, 2012. **131**(11): p. 1783-93.

52. Ulmer, M., et al., *Genome-wide analysis of central corneal thickness in primary open-angle glaucoma cases in the NEIGHBOR and GLAUGEN consortia.* Invest Ophthalmol Vis Sci, 2012. **53**(8): p. 4468-74.

53. Desronvil, T., et al., *Distribution of COL8A2 and COL8A1 gene variants in Caucasian primary open angle glaucoma patients with thin central corneal thickness.* Mol Vis, 2010. **16**: p. 2185-91.

54. Lehmann, O.J., et al., *Novel anterior segment phenotypes resulting from forkhead gene alterations: evidence for cross-species conservation of function.* Invest Ophthalmol Vis Sci, 2003. **44**(6): p. 2627-33.

55. Ramdas, W.D., et al., *A genome-wide association study of optic disc parameters.* PLoS Genet, 2010. **6**(6): p. e1000978.

56. Khor, C.C., et al., *Genome-wide association studies in Asians confirm the involvement of ATOH7 and TGFBR3, and further identify CARD10 as a novel locus influencing optic disc area.* Hum Mol Genet, 2011. **20**(9): p. 1864-72.

57. Macgregor, S., et al., *Genome-wide association identifies ATOH7 as a major gene determining human optic disc size.* Hum Mol Genet, 2010. **19**(13): p. 2716-24.

58. Fan, B.J., et al., *Genetic variants associated with optic nerve vertical cup-to-disc ratio are risk factors for primary open angle glaucoma in a US Caucasian population.* Invest Ophthalmol Vis Sci, 2011. **52**(3): p. 1788-92.

59. Mabuchi, F., et al., *Association between genetic variants associated with vertical cup-to-disc ratio and phenotypic features of primary open-angle glaucoma.* Ophthalmology, 2012. **119**(9): p. 1819-25.

60. Pascolini, D. and S.P. Mariotti, *Global estimates of visual impairment: 2010.* Br J Ophthalmol, 2012. **96**(5): p. 614-8.

61. Thylefors, B. and A.D. Negrel, *The global impact of glaucoma.* Bull World Health Organ, 1994. **72**(3): p. 323-6.

62. Green, C.M., et al., *How significant is a family history of glaucoma? Experience from the Glaucoma Inheritance Study in Tasmania.* Clin Experiment Ophthalmol, 2007. **35**(9): p. 793-9.

63. Wolfs, R.C., et al., *Genetic risk of primary open-angle glaucoma. Population-based familial aggregation study.* Arch Ophthalmol, 1998. **116**(12): p. 1640-5.

64. Kyari, F., et al., *Epidemiology of glaucoma in sub-saharan Africa: prevalence, incidence and risk factors.* Middle East Afr J Ophthalmol, 2013. **20**(2): p. 111-25.

65. Tham, Y.C., et al., *Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040: A Systematic Review and Meta-Analysis.* Ophthalmology, 2014.

66. Thomas, R., P.K. Nirmalan, and S. Krishnaiah, *Pseudoexfoliation in southern India: the Andhra Pradesh Eye Disease Study.* Invest Ophthalmol Vis Sci, 2005. **46**(4): p. 1170-6.

67. Krishnadas, R., et al., *Pseudoexfoliation in a rural population of southern India: the Aravind Comprehensive Eye Survey.* Am J Ophthalmol, 2003. **135**(6): p. 830-7.

68. Arvind, H., et al., *Pseudoexfoliation in South India.* Br J Ophthalmol, 2003. **87**(11): p. 1321-3.

69. Jonas, J.B., et al., *Pseudoexfoliation: normative data and associations. The central India eye and medical study.* PLoS One, 2013. **8**(10): p. e76770.

70. Whigham, B.T. and R.R. Allingham, *Review: The role of LOXL1 in exfoliation syndrome/glaucoma.* Saudi J Ophthalmol, 2011. **25**(4): p. 347-52.

71. You, Q.S., et al., *Pseudoexfoliation: normative data and associations: the Beijing eye study 2011.* Ophthalmology, 2013. **120**(8): p. 1551-8.

72. Karen K. O'Brien, A.W.Y.C., Catherine A. Opere. *An Overview of Glaucoma Management for Pharmacists*. 2010.

73. Fan, B.J., et al., *Genetic and environmental risk factors for primary open-angle glaucoma.* Chin Med J (Engl), 2004. **117**(5): p. 706-10.

74. Kokotas, H., et al., *Biomarkers in primary open angle glaucoma.* Clin Chem Lab Med, 2012. **50**(12): p. 2107-19.

75. Rao, K.N., S. Nagireddy, and S. Chakrabarti, *Complex genetic mechanisms in glaucoma: an overview.* Indian J Ophthalmol, 2011. **59 Suppl**: p. S31-42.

76. Gong, G., et al., *Genetic dissection of myocilin glaucoma.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R91-102.

77. Rezaie, T., et al., *Adult-onset primary open-angle glaucoma caused by mutations in optineurin.* Science, 2002. **295**(5557): p. 1077-9.

78. Skarie, J.M. and B.A. Link, *The primary open-angle glaucoma gene WDR36 functions in ribosomal RNA processing and interacts with the p53 stress-response pathway.* Hum Mol Genet, 2008. **17**(16): p. 2474-85.

79. Pasutto, F., et al., *Heterozygous NTF4 mutations impairing neurotrophin-4 signaling in patients with primary open-angle glaucoma.* Am J Hum Genet, 2009. **85**(4): p. 447-56.

80. Stoilov, I., A.N. Akarsu, and M. Sarfarazi, *Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (Buphthalmos) in families linked to the GLC3A locus on chromosome 2p21.* Hum Mol Genet, 1997. **6**(4): p. 641-7.

81. Freeman, E.E., et al., *The heritability of glaucoma-related traits corneal hysteresis, central corneal thickness, intraocular pressure, and choroidal blood flow pulsatility.* PLoS One, 2013. **8**(1): p. e55573.

82. van Koolwijk, L.M., et al., *Genetic contributions to glaucoma: heritability of intraocular pressure, retinal nerve fiber layer thickness, and optic disc morphology.* Invest Ophthalmol Vis Sci, 2007. **48**(8): p. 3669-76.

83. Ramdas, W.D., et al., *Common genetic variants associated with open-angle glaucoma.* Hum Mol Genet, 2011. **20**(12): p. 2464-71.

84. Banerjee, D., et al., *Comprehensive analysis of myocilin variants in east Indian POAG patients.* Mol Vis, 2012. **18**: p. 1548-57.

85. Sripriya, S., et al., *Low frequency of myocilin mutations in Indian primary open-angle glaucoma patients.* Clin Genet, 2004. **65**(4): p. 333-7.

86. Kumar, A., et al., *Role of CYP1B1, MYOC, OPTN, and OPTC genes in adult-onset primary open-angle glaucoma: predominance of CYP1B1 mutations in Indian patients.* Mol Vis, 2007. **13**: p. 667-76.

87. Sripriya, S., et al., *OPTN gene: profile of patients with glaucoma from India.* Mol Vis, 2006. **12**: p. 816-20.

88. Rao, K.N., et al., *Variations in NTF4, VAV2, and VAV3 genes are not involved with primary open-angle and primary angle-closure glaucomas in an indian population.* Invest Ophthalmol Vis Sci, 2010. **51**(10): p. 4937-41.

89. Jarman, A.P., et al., *atonal is a proneural gene that directs chordotonal organ formation in the Drosophila peripheral nervous system.* Cell, 1993. **73**(7): p. 1307-21.

90. Brown, N.L., et al., *Molecular characterization and mapping of ATOH7, a human atonal homolog with a predicted role in retinal ganglion cell development.* Mamm Genome, 2002. **13**(2): p. 95-101.

91. Brown, N.L., et al., *Math5 is required for retinal ganglion cell and optic nerve formation.* Development, 2001. **128**(13): p. 2497-508.

92. Wang, S.W., et al., *Requirement for math5 in the development of retinal ganglion cells.* Genes Dev, 2001. **15**(1): p. 24-9.

93. Khan, K., et al., *Next generation sequencing identifies mutations in Atonal homolog 7 (ATOH7) in families with global eye developmental defects.* Hum Mol Genet, 2012. **21**(4): p. 776-83.

94. Carnes, M.U., et al., *Discovery and functional annotation of SIX6 variants in primary open-angle glaucoma.* PLoS Genet, 2014. **10**(5): p. e1004372.

95. Li, X., et al., *Tissue-specific regulation of retinal and pituitary precursor cell proliferation.* Science, 2002. **297**(5584): p. 1180-3.

96. Zuber, M.E., et al., *Giant eyes in Xenopus laevis by overexpression of XOptx2.* Cell, 1999. **98**(3): p. 341-52.

97. Bernier, G., et al., *Expanded retina territory by midbrain transformation upon overexpression of Six6 (Optx2) in Xenopus embryos.* Mech Dev, 2000. **93**(1-2): p. 59-69.

98. Cao, D., et al., *CDKN2B polymorphism is associated with primary open-angle glaucoma (POAG) in the Afro-Caribbean population of Barbados, West Indies.* PLoS One, 2012. **7**(6): p. e39278.

99. Wiggs, J.L., et al., *Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma.* PLoS Genet, 2012. **8**(4): p. e1002654.

100. Ruland, J., et al., *Bcl10 is a positive regulator of antigen receptor-induced activation of NF-kappaB and neural tube closure.* Cell, 2001. **104**(1): p. 33-42.

101. Abu, A., et al., *Deleterious mutations in the Zinc-Finger 469 gene cause brittle cornea syndrome.* Am J Hum Genet, 2008. **82**(5): p. 1217-22.

102. Lu, Y., et al., *Common genetic variants near the Brittle Cornea Syndrome locus ZNF469 influence the blinding disease risk factor central corneal thickness.* PLoS Genet, 2010. **6**(5): p. e1000947.

103. Hong, S., et al., *Central corneal thickness and visual field progression in patients with chronic primary angle-closure glaucoma with low intraocular pressure.* Am J Ophthalmol, 2007. **143**(2): p. 362-3.

104. Ritch, R., *Exfoliation syndrome-the most common identifiable cause of open-angle glaucoma.* J Glaucoma, 1994. **3**(2): p. 176-7.

105. Ritch, R., U. Schlotzer-Schrehardt, and A.G. Konstas, *Why is glaucoma associated with exfoliation syndrome?* Prog Retin Eye Res, 2003. **22**(3): p. 253-75.

106. Schlotzer-Schrehardt, U., *Genetics and genomics of pseudoexfoliation syndrome/glaucoma.* Middle East Afr J Ophthalmol, 2011. **18**(1): p. 30-6.

107. Rita Ehrlich, A.H., Carlos Rospigliosi Lopez, Nisha S Kheradiya *Pseudoexfoliation and Blood Flow Abnormalities*. Jaypee journals.

108. Thorleifsson, G., et al., *Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma.* Science, 2007. **317**(5843): p. 1397-400.

109. Aragon-Martin, J.A., et al., *Evaluation of LOXL1 gene polymorphisms in exfoliation syndrome and exfoliation glaucoma.* Mol Vis, 2008. **14**: p. 533-41.

110. Hewitt, A.W., et al., *Ancestral LOXL1 variants are associated with pseudoexfoliation in Caucasian Australians but with markedly lower penetrance than in Nordic people.* Hum Mol Genet, 2008. **17**(5): p. 710-6.

111. Ramprasad, V.L., et al., *Association of non-synonymous single nucleotide polymorphisms in the LOXL1 gene with pseudoexfoliation syndrome in India.* Mol Vis, 2008. **14**: p. 318-22.

112. Pasutto, F., et al., *Association of LOXL1 common sequence variants in German and Italian patients with pseudoexfoliation syndrome and pseudoexfoliation glaucoma.* Invest Ophthalmol Vis Sci, 2008. **49**(4): p. 1459-63.

113. Mossbock, G., et al., *Lysyl oxidase-like protein 1 (LOXL1) gene polymorphisms and exfoliation glaucoma in a Central European population.* Mol Vis, 2008. **14**: p. 857-61.

114. Chen, L., et al., *Evaluation of LOXL1 polymorphisms in exfoliation syndrome in a Chinese population.* Mol Vis, 2009. **15**: p. 2349-57.

115. Fuse, N., et al., *Evaluation of LOXL1 polymorphisms in eyes with exfoliation glaucoma in Japanese.* Mol Vis, 2008. **14**: p. 1338-43.

116. Abu-Amero, K.K., et al., *Analysis of LOXL1 polymorphisms in a Saudi Arabian population with pseudoexfoliation glaucoma.* Mol Vis, 2010. **16**: p. 2805-10.

117. Decitre, M., et al., *Lysyl oxidase-like protein localizes to sites of de novo fibrinogenesis in fibrosis and in the early stromal reaction of ductal breast carcinomas.* Lab Invest, 1998. **78**(2): p. 143-51.

118. Kiemer, A.K., K. Takeuchi, and M.P. Quinlan, *Identification of genes involved in epithelial-mesenchymal transition and tumor progression.* Oncogene, 2001. **20**(46): p. 6679-88.

119. Urban, Z., et al., *Population differences in elastin maturation in optic nerve head tissue and astrocytes.* Invest Ophthalmol Vis Sci, 2007. **48**(7): p. 3209-15.

120. Ritch, R. and U. Schlotzer-Schrehardt, *Exfoliation syndrome.* Surv Ophthalmol, 2001. **45**(4): p. 265-315.

121. Thomassin, L., et al., *The Pro-regions of lysyl oxidase and lysyl oxidase-like 1 are required for deposition onto elastic fibers.* J Biol Chem, 2005. **280**(52): p. 42848-55.

122. Kagan, H.M. and W. Li, *Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell.* J Cell Biochem, 2003. **88**(4): p. 660-72.

123. Tang, J.Z., et al., *Association between polymorphisms in lysyl oxidase-like 1 and susceptibility to pseudoexfoliation syndrome and pseudoexfoliation glaucoma.* PLoS One, 2014. **9**(3): p. e90331.

124. Anastasopoulos, E., et al., *Association of LOXL1 Polymorphisms With Pseudoexfoliation, Glaucoma, Intraocular Pressure, and Systemic Diseases in a Greek Population. The Thessaloniki Eye Study.* Invest Ophthalmol Vis Sci, 2014. **55**(7): p. 4238-43.

125. Khan, T.T., et al., *LOXL1 expression in lens capsule tissue specimens from individuals with pseudoexfoliation syndrome and glaucoma.* Mol Vis, 2010. **16**: p. 2236-41.

126. Mori, K., et al., *LOXL1 genetic polymorphisms are associated with exfoliation glaucoma in the Japanese population.* Mol Vis, 2008. **14**: p. 1037-40.

127. Schlotzer-Schrehardt, U., et al., *Genotype-correlated expression of lysyl oxidase-like 1 in ocular tissues of patients with pseudoexfoliation syndrome/glaucoma and normal patients.* Am J Pathol, 2008. **173**(6): p. 1724-35.

128. Zenkel, M., et al., *Clusterin deficiency in eyes with pseudoexfoliation syndrome may be implicated in the aggregation and deposition of pseudoexfoliative material.* Invest Ophthalmol Vis Sci, 2006. **47**(5): p. 1982-90.

129. Liu, X., et al., *Elastic fiber homeostasis requires lysyl oxidase-like 1 protein.* Nat Genet, 2004. **36**(2): p. 178-82.

130. Untergasser, A., et al., *Primer3--new capabilities and interfaces.* Nucleic Acids Res, 2012. **40**(15): p. e115.

131. Kibbe, W.A., *OligoCalc: an online oligonucleotide properties calculator.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W43-6.

132. George, R., et al., *The Chennai glaucoma study: prevalence and risk factors for glaucoma in cataract operated eyes in urban Chennai.* Indian J Ophthalmol, 2010. **58**(3): p. 243-5.

133. Healey, P., et al., *The heritability of optic disc parameters: a classic twin study.* Invest Ophthalmol Vis Sci, 2008. **49**(1): p. 77-80.

134. McCarthy, M.I. and E. Zeggini, *Genome-wide association studies in type 2 diabetes.* Curr Diab Rep, 2009. **9**(2): p. 164-71.

135. Frost, R. *Genome-Wide association studies: understanding the genetics of common disease.* 2009. GlaxoSmithKline.

136. Woods, C.G., et al., *Quantification of homozygosity in consanguineous individuals with autosomal recessive disease.* Am J Hum Genet, 2006. **78**(5): p. 889-96.

137. Lichtenstein, P., et al., *Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study.* Lancet, 2009. **373**(9659): p. 234-9.

138. Purcell, S.M., et al., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-52.

139. Morrow, E.M., et al., *Identifying autism loci and genes by tracing recent shared ancestry.* Science, 2008. **321**(5886): p. 218-23.

140. Mansour, H., et al., *Consanguinity associated with increased risk for bipolar I disorder in Egypt.* Am J Med Genet B Neuropsychiatr Genet, 2009. **150B**(6): p. 879-85.

141. Funayama, T., et al., *SNPs and interaction analyses of noelin 2, myocilin, and optineurin genes in Japanese patients with open-angle glaucoma.* Invest Ophthalmol Vis Sci, 2006. **47**(12): p. 5368-75.

142. Ritchie, M.D., et al., *Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer.* Am J Hum Genet, 2001. **69**(1): p. 138-47.

143. Calle, M.L., et al., *Improving strategies for detecting genetic patterns of disease susceptibility in association studies.* Stat Med, 2008. **27**(30): p. 6532-46.

144. RipleyBD, *Pattern recognition and neural networks.* Cambridge: Cambridge University Press, 1996.

145. Purcell, S., S.S. Cherny, and P.C. Sham, *Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits.* Bioinformatics, 2003. **19**(1): p. 149-50.

146. Burdon, K.P., et al., *Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1.* Nat Genet, 2011. **43**(6): p. 574-8.

147. Choi, J.H., J.E. Kim, and B.K. Kaang, *Protein synthesis and degradation are required for the incorporation of modified information into the pre-existing object-location memory.* Mol Brain, 2010. **3**: p. 1.

148. Yang, Z., et al., *Math5 determines the competence state of retinal ganglion cell progenitors.* Dev Biol, 2003. **264**(1): p. 240-54.

149. Yao, J., et al., *Math5 promotes retinal ganglion cell expression patterns in retinal progenitor cells.* Mol Vis, 2007. **13**: p. 1066-72.

150. Song, W.T., X.Y. Zhang, and X.B. Xia, *Atoh7 promotes the differentiation of retinal stem cells derived from Muller cells into retinal ganglion cells by inhibiting Notch signaling.* Stem Cell Res Ther, 2013. **4**(4): p. 94.

151. Prasov, L., et al., *ATOH7 mutations cause autosomal recessive persistent hyperplasia of the primary vitreous.* Hum Mol Genet, 2012. **21**(16): p. 3681-94.

152. Marchini, J., P. Donnelly, and L.R. Cardon, *Genome-wide strategies for detecting multiple loci that influence complex diseases.* Nat Genet, 2005. **37**(4): p. 413-7.

153. Ozel, A.B., et al., *Genome-wide association study and meta-analysis of intraocular pressure.* Hum Genet, 2014. **133**(1): p. 41-57.

154. Iglesias, A.I., et al., *Exome sequencing and functional analyses suggest that SIX6 is a gene involved in an altered proliferation-differentiation balance early in life and optic nerve degeneration at old age.* Hum Mol Genet, 2014. **23**(5): p. 1320-32.

155. Osman, W., et al., *A genome-wide association study in the Japanese population confirms 9p21 and 14q23 as susceptibility loci for primary open angle glaucoma.* Hum Mol Genet, 2012. **21**(12): p. 2836-42.

156. Springelkamp, H., et al., *Meta-analysis of genome-wide association studies identifies novel loci that influence cupping and the glaucomatous process.* Nat Commun, 2014. **5**: p. 4883.

157. Fan, B.J., et al., *SNPs and interaction analyses of myocilin, optineurin, and apolipoprotein E in primary open angle glaucoma patients.* Mol Vis, 2005. **11**: p. 625-31.

158. Jia, L.Y., et al., *Multiple gene polymorphisms analysis revealed a different profile of genetic polymorphisms of primary open-angle glaucoma in northern Chinese.* Mol Vis, 2009. **15**: p. 89-98.

159. Blanco-Marchite, C., et al., *WDR36 and P53 gene variants and susceptibility to primary open-angle glaucoma: analysis of gene-gene interactions.* Invest Ophthalmol Vis Sci, 2011. **52**(11): p. 8467-78.

160. Zanon-Moreno, V., et al., *Effects of polymorphisms in vitamin E-, vitamin C-, and glutathione peroxidase-related genes on serum biomarkers and associations with glaucoma.* Mol Vis, 2013. **19**: p. 231-42.

161. Chen, J.H., et al., *Interactive effects of ATOH7 and RFTN1 in association with adult-onset primary open-angle glaucoma.* Invest Ophthalmol Vis Sci, 2012. **53**(2): p. 779-85.

162. Arnarsson, A., et al., *Pseudoexfoliation in the Reykjavik Eye Study: prevalence and related ophthalmological variables.* Acta Ophthalmol Scand, 2007. **85**(8): p. 822-7.

163. Gross, F.J., D. Tingey, and D.L. Epstein, *Increased prevalence of occludable angles and angle-closure glaucoma in patients with pseudoexfoliation.* Am J Ophthalmol, 1994. **117**(3): p. 333-6.

164. Williams, S.E., et al., *Major LOXL1 risk allele is reversed in exfoliation glaucoma in a black South African population.* Mol Vis, 2010. **16**: p. 705-12.

165. Kleinstein, R.N., et al., *Refractive error and ethnicity in children.* Arch Ophthalmol, 2003. **121**(8): p. 1141-7.

166. Fan, D.S., et al., *Prevalence, incidence, and progression of myopia of school children in Hong Kong.* Invest Ophthalmol Vis Sci, 2004. **45**(4): p. 1071-5.

167. Hochman, M.A., C.M. Seery, and M.A. Zarbin, *Pathophysiology and management of subretinal hemorrhage.* Surv Ophthalmol, 1997. **42**(3): p. 195-213.

168. Banker, A.S. and W.R. Freeman, *Retinal detachment.* Ophthalmol Clin North Am, 2001. **14**(4): p. 695-704.

169. Saw, S.M., et al., *Myopia and associated pathological complications.* Ophthalmic Physiol Opt, 2005. **25**(5): p. 381-91.

170. Spinrad, P., *Axial myopia: A closer look at a promising new therapy.* 2009.

171. EJ, T., *the optical elements of the refractive power of the eye. In: Ridley F, Sorsby A, eds. Mordern trends in Ophthalmology.* 1940, New York: Paul B Hoeber.

172. Zheng, Y.F., et al., *The economic cost of myopia in adults aged over 40 years in Singapore.* Invest Ophthalmol Vis Sci, 2013. **54**(12): p. 7532-7.

173. Saw, S.M., *A synopsis of the prevalence rates and environmental risk factors for myopia.* Clin Exp Optom, 2003. **86**(5): p. 289-94.

174. Jaalouk, D.E. and J. Lammerding, *Mechanotransduction gone awry.* Nat Rev Mol Cell Biol, 2009. **10**(1): p. 63-73.

175. Schwartz, M., M. Haim, and D. Skarsholm, *X-linked myopia: Bornholm eye disease. Linkage to DNA markers on the distal part of Xq.* Clin Genet, 1990. **38**(4): p. 281-6.

176. Ciner, E., et al., *Genome-wide scan of African-American and white families for linkage to myopia.* Am J Ophthalmol, 2009. **147**(3): p. 512-517 e2.

177. Stambolian, D., et al., *Genome-wide scan for myopia in the Old Order Amish.* Am J Ophthalmol, 2005. **140**(3): p. 469-76.

178. Yu, Z.Q., et al., *[A genome-wide screening for pathological myopia suggests a novel locus on chromosome 15q12 - 13].* Zhonghua Yan Ke Za Zhi, 2007. **43**(3): p. 233-8.

179. Lam, C.Y., et al., *A genome-wide scan maps a novel high myopia locus to 5p15.* Invest Ophthalmol Vis Sci, 2008. **49**(9): p. 3768-78.

180. Tang, W.C., M.K. Yap, and S.P. Yip, *A review of current approaches to identifying human genes involved in myopia.* Clin Exp Optom, 2008. **91**(1): p. 4-22.

181. Wojciechowski, R., et al., *Genomewide scan in Ashkenazi Jewish families demonstrates evidence of linkage of ocular refraction to a QTL on chromosome 1p36.* Hum Genet, 2006. **119**(4): p. 389-99.

182. Kiefer, A.K., et al., *Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia.* PLoS Genet, 2013. **9**(2): p. e1003299.

183. Ahmed, I., et al., *TGIF1 is a Potential Candidate Gene for High Myopia in Ethnic Kashmiri Population.* Curr Eye Res, 2013.

184. Yoshida, M., et al., *Association study of IGF1 polymorphisms with susceptibility to high myopia in a Japanese population.* Clin Ophthalmol, 2013. **7**: p. 2057-62.

185. Zhang, D., et al., *An association study of the COL1A1 gene and high myopia in a Han Chinese population.* Mol Vis, 2011. **17**: p. 3379-83.

186. Schippert, R., et al., *Relative axial myopia in Egr-1 (ZENK) knockout mice.* Invest Ophthalmol Vis Sci, 2007. **48**(1): p. 11-7.

187. Miyake, M., et al., *Association of paired box 6 with high myopia in Japanese.* Mol Vis, 2012. **18**: p. 2726-35.

188. Wojciechowski, R., et al., *Matrix metalloproteinases and educational attainment in refractive error: evidence of gene-environment interactions in the Age-Related Eye Disease Study.* Ophthalmology, 2013. **120**(2): p. 298-305.

189. Chen, J.H., et al., *Endophenotyping reveals differential phenotype-genotype correlations between myopia-associated polymorphisms and eye biometric parameters.* Mol Vis, 2012. **18**: p. 765-78.

190. Metlapally, R., et al., *Common MFRP sequence variants are not associated with moderate to high hyperopia, isolated microphthalmia, and high myopia.* Mol Vis, 2008. **14**: p. 387-93.

191. Aung, T., et al., *Molecular analysis of CHX10 and MFRP in Chinese subjects with primary angle closure glaucoma and short axial length eyes.* Mol Vis, 2008. **14**: p. 1313-8.

192. Mishra, A., et al., *Genetic variants near PDGFRA are associated with corneal curvature in Australians.* Invest Ophthalmol Vis Sci, 2012. **53**(11): p. 7131-6.

193. Verhoeven, V.J., et al., *Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia.* Nat Genet, 2013. **45**(3): p. 314-8.

194. Zhou, G. and R.W. Williams, *Eye1 and Eye2: gene loci that modulate eye size, lens weight, and retinal area in the mouse.* Invest Ophthalmol Vis Sci, 1999. **40**(5): p. 817-25.

195. Saltarelli, D., et al., *Susceptibility to form-deprivation myopia in chicks is not altered by an early experience of axial myopia.* Optom Vis Sci, 2004. **81**(2): p. 119-26.

196. Jobling, A.I., et al., *Regulation of scleral cell contraction by transforming growth factor-beta and stress: competing roles in myopic eye growth.* J Biol Chem, 2009. **284**(4): p. 2072-9.

197. Funata, M. and T. Tokoro, *Scleral change in experimentally myopic monkeys.* Graefes Arch Clin Exp Ophthalmol, 1990. **228**(2): p. 174-9.

198. Young, T.L., R. Metlapally, and A.E. Shay, *Complex trait genetics of refractive error.* Arch Ophthalmol, 2007. **125**(1): p. 38-48.

199. Rasool, S., et al., *Contribution of TGFbeta1 codon 10 polymorphism to high myopia in an ethnic Kashmiri population from India.* Biochem Genet, 2013. **51**(3-4): p. 323-33.

200. Annamaneni, S., et al., *Association of vitamin D receptor gene start codon (Fok1) polymorphism with high myopia.* Oman J Ophthalmol, 2011. **4**(2): p. 57-62.

201. Ratnamala, U., et al., *Refinement of the X-linked nonsyndromic high-grade myopia locus MYP1 on Xq28 and exclusion of 13 known positional candidate genes by direct sequencing.* Invest Ophthalmol Vis Sci, 2011. **52**(9): p. 6814-9.

202. Sundin, O.H., et al., *Developmental basis of nanophthalmos: MFRP Is required for both prenatal ocular growth and postnatal emmetropization.* Ophthalmic Genet, 2008. **29**(1): p. 1-9.

203. Katoh, M., *Molecular cloning and characterization of MFRP, a novel gene encoding a membrane-type Frizzled-related protein.* Biochem Biophys Res Commun, 2001. **282**(1): p. 116-23.

204. Mandal, M.N., et al., *Spatial and temporal expression of MFRP and its interaction with CTRP5.* Invest Ophthalmol Vis Sci, 2006. **47**(12): p. 5514-21.

205. Sundin, O.H., et al., *Extreme hyperopia is the result of null mutations in MFRP, which encodes a Frizzled-related protein.* Proc Natl Acad Sci U S A, 2005. **102**(27): p. 9553-8.

206. Liu, I.S., et al., *Developmental expression of a novel murine homeobox gene (Chx10): evidence for roles in determination of the neuroretina and inner nuclear layer.* Neuron, 1994. **13**(2): p. 377-93.

207. Liang, L. and J.H. Sandell, *Focus on molecules: homeobox protein Chx10.* Exp Eye Res, 2008. **86**(4): p. 541-2.

208. Levine, E.M. and E.S. Green, *Cell-intrinsic regulators of proliferation in vertebrate retinal progenitors.* Semin Cell Dev Biol, 2004. **15**(1): p. 63-74.

209. Bar-Yosef, U., et al., *CHX10 mutations cause non-syndromic microphthalmia/ anophthalmia in Arab and Jewish kindreds.* Hum Genet, 2004. **115**(4): p. 302-9.

210. Wong, G., S.B. Conger, and M. Burmeister, *Mapping of genetic modifiers affecting the eye phenotype of ocular retardation (Chx10or-J) mice.* Mamm Genome, 2006. **17**(6): p. 518-25.

211. Burmeister, M., et al., *Ocular retardation mouse caused by Chx10 homeobox null allele: impaired retinal progenitor proliferation and bipolar cell differentiation.* Nat Genet, 1996. **12**(4): p. 376-84.

212. Quiring, R., et al., *Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans.* Science, 1994. **265**(5173): p. 785-9.

213. Terzic, J. and M. Saraga-Babic, *Expression pattern of PAX3 and PAX6 genes during human embryogenesis.* Int J Dev Biol, 1999. **43**(6): p. 501-8.

214. Hatakeyama, J. and R. Kageyama, *Retinal cell fate determination and bHLH factors.* Semin Cell Dev Biol, 2004. **15**(1): p. 83-9.

215. Baulmann, D.C., et al., *Pax6 heterozygous eyes show defects in chamber angle differentiation that are associated with a wide spectrum of other anterior eye segment abnormalities.* Mech Dev, 2002. **118**(1-2): p. 3-17.

216. Glaser, T., et al., *PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects.* Nat Genet, 1994. **7**(4): p. 463-71.

217. Ferda Percin, E., et al., *Human microphthalmia associated with mutations in the retinal homeobox gene CHX10.* Nat Genet, 2000. **25**(4): p. 397-401.

218. Wang, P., et al., *Evaluation of MFRP as a candidate gene for high hyperopia.* Mol Vis, 2009. **15**: p. 181-6.

219. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

220. VassarStats, R.L. *VassarStats: Website for Statistical Computation*. Vassarstats.

221. Barrett, J.C., *Haploview: Visualization and analysis of SNP genotype data.* Cold Spring Harb Protoc, 2009. **2009**(10): p. pdb ip71.

222. Guo, Y., et al., *Gains in power for exhaustive analyses of haplotypes using variable-sized sliding window strategy: a comparison of association-mapping strategies.* Eur J Hum Genet, 2009. **17**(6): p. 785-92.

223. Tregouet, D.A. and V. Garelle, *A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies.* Bioinformatics, 2007. **23**(8): p. 1038-9.

224. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.

225. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

226. Yuan, H.Y., et al., *FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W635-41.

227. Cheng, J., A. Randall, and P. Baldi, *Prediction of protein stability changes for single-site mutations using support vector machines.* Proteins, 2006. **62**(4): p. 1125-32.

228. Puigbo, P., I.G. Bravo, and S. Garcia-Vallve, *CAIcal: a combined set of tools to assess codon usage adaptation.* Biol Direct, 2008. **3**: p. 38.

229. Cartegni, L., et al., *ESEfinder: A web resource to identify exonic splicing enhancers.* Nucleic Acids Res, 2003. **31**(13): p. 3568-71.

230. Mutti, D.O., et al., *Axial growth and changes in lenticular and corneal power during emmetropization in infants.* Invest Ophthalmol Vis Sci, 2005. **46**(9): p. 3074-80.

231. Flitcroft, D.I., *Is myopia a failure of homeostasis?* Exp Eye Res, 2013. **114**: p. 16-24.

232. Wallman, J. and J. Winawer, *Homeostasis of eye growth and the question of myopia.* Neuron, 2004. **43**(4): p. 447-68.

233. Siegwart, J.T., Jr. and T.T. Norton, *Perspective: how might emmetropization and genetic factors produce myopia in normal eyes?* Optom Vis Sci, 2011. **88**(3): p. E365-72.

234. Ayala-Ramirez, R., et al., *A new autosomal recessive syndrome consisting of posterior microphthalmos, retinitis pigmentosa, foveoschisis, and optic disc drusen is caused by a MFRP gene mutation.* Mol Vis, 2006. **12**: p. 1483-9.

235. Penha, A.M., F. Schaeffel, and M. Feldkaemper, *Insulin, insulin-like growth factor-1, insulin receptor, and insulin-like growth factor-1 receptor expression in the chick eye and their regulation with imposed myopic or hyperopic defocus.* Mol Vis, 2011. **17**: p. 1436-48.

236. Jiang, B., et al., *PAX6 haplotypes are associated with high myopia in Han chinese.* PLoS One, 2011. **6**(5): p. e19587.

237. Kannabiran, C., et al., *Mutations in TULP1, NR2E3, and MFRP genes in Indian families with autosomal recessive retinitis pigmentosa.* Mol Vis, 2012. **18**: p. 1165-74.

238. Zakharov S, W.T., Aung T, *Combined genotype and haplotype tests for region-based association studies.* BMC Genomics, 2013. **14**: p. 569.

239. Fingert, J.H., et al., *Analysis of ASB10 variants in open angle glaucoma.* Hum Mol Genet, 2012.

240. Sauna, Z.E. and C. Kimchi-Sarfaty, *Understanding the contribution of synonymous mutations to human disease.* Nat Rev Genet, 2011. **12**(10): p. 683-91.

241. Reis, L.M., et al., *VSX2 mutations in autosomal recessive microphthalmia.* Mol Vis, 2011. **17**: p. 2527-32.

242. Tsonis, P.A. and E.J. Fuentes, *Focus on molecules: Pax-6, the eye master.* Exp Eye Res, 2006. **83**(2): p. 233-4.

243. Chen, K.C., et al., *MicroRNA-328 may influence myopia development by mediating the PAX6 gene.* Invest Ophthalmol Vis Sci, 2012. **53**(6): p. 2732-9.

244. Tkatchenko, A.V., et al., *Form deprivation modulates retinal neurogenesis in primate experimental myopia.* Proc Natl Acad Sci U S A, 2006. **103**(12): p. 4681-6.

245. Gentle, A. and N.A. McBrien, *Modulation of scleral DNA synthesis in development of and recovery from induced axial myopia in the tree shrew.* Exp Eye Res, 1999. **68**(2): p. 155-63.

246. Hammond, C.J., et al., *A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: a genomewide scan of dizygotic twins.* Am J Hum Genet, 2004. **75**(2): p. 294-304.

247. Mutti, D.O., et al., *Candidate gene and locus analysis of myopia.* Mol Vis, 2007. **13**: p. 1012-9.

248. Simpson, C.L., et al., *The Roles of PAX6 and SOX2 in Myopia: lessons from the 1958 British Birth Cohort.* Invest Ophthalmol Vis Sci, 2007. **48**(10): p. 4421-5.

249. Tsai, Y.Y., et al., *A PAX6 gene polymorphism is associated with genetic predisposition to extreme myopia.* Eye (Lond), 2008. **22**(4): p. 576-81.

250. Liang, C.L., et al., *A functional polymorphism at 3'UTR of the PAX6 gene may confer risk for extreme myopia in the Chinese.* Invest Ophthalmol Vis Sci, 2011. **52**(6): p. 3500-5.

251. Zayats, T., et al., *Comment on 'A PAX6 gene polymorphism is associated with genetic predisposition to extreme myopia'.* Eye (Lond), 2008. **22**(4): p. 598-9; author reply 599.

252. Hysi, P.G., et al., *A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25.* Nat Genet, 2010. **42**(10): p. 902-5.

253. Solouki, A.M., et al., *A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14.* Nat Genet, 2010. **42**(10): p. 897-901.

254. Li, Y.J., et al., *Genome-wide association studies reveal genetic variants in CTNND2 for high myopia in Singapore Chinese.* Ophthalmology, 2011. **118**(2): p. 368-75.

255. Li, Z., et al., *A genome-wide association study reveals association between common variants in an intergenic region of 4q25 and high-grade myopia in the Chinese Han population.* Hum Mol Genet, 2011. **20**(14): p. 2861-8.

256. Shi, Y., et al., *Genetic variants at 13q12.12 are associated with high myopia in the Han Chinese population.* Am J Hum Genet, 2011. **88**(6): p. 805-13.

257. Yang, Z., et al., *Clinical and linkage study on a consanguineous Chinese family with autosomal recessive high myopia.* Mol Vis, 2009. **15**: p. 312-8.

258. Stojic, J., H. Stohr, and B.H. Weber, *Three novel ABCC5 splice variants in human retina and their role as regulators of ABCC5 gene expression.* BMC Mol Biol, 2007. **8**: p. 42.

259. Nongpiur, M.E., et al., *ABCC5, a gene that influences the anterior chamber depth, is associated with primary angle closure glaucoma.* PLoS Genet, 2014. **10**(3): p. e1004089.

260. Paluru, P.C., et al., *Identification of a novel locus on 2q for autosomal dominant high-grade myopia.* Invest Ophthalmol Vis Sci, 2005. **46**(7): p. 2300-7.

261. Nagel, M., S. Nagorka, and O. Gross, *Novel COL4A5, COL4A4, and COL4A3 mutations in Alport syndrome.* Hum Mutat, 2005. **26**(1): p. 60.

## Appendices

### 1. Patient recruitment and clinical examination

All the studies were approved by the institutional ethics board adhered to the guidelines in the Declaration of Helsinki and was conducted at the SNONGC department of genetics and molecular biology, Vision Research Foundation, Sankara Nethralaya, India.

***Chapter 2.1*:**

Large consanguineous families from South India were identified through the resources available at Sankara Nethralaya eye hospital. Pedigrees for these families were drawn using information obtained from the proband. The criteria for selecting the families included i) a minimum of atleast 20 members ii) presence of more than one consanguineous marriage, and iii) a minimum of atleast 6 individuals per sibship. After identifying the pedigrees all the available family members were ascertained for complete phenotypic characterization, and blood collection.

The exclusion criteria included:

1. Members are not affected by any single-gene Mendelian disorder eg: congenital glaucoma or retinitis pigmentosa.
2. Using family history to exclude families which show extreme phenotypes so that the clinical measurements are not skewed. Eg: a family affected by congenital glaucoma would not be useful for QT mapping of IOP because they would have abnormally high IOP which would overwhelm the values of the other families.

Four large normal consanguineous families were identified which fulfilled the above criteria from the Chennai Glaucoma study [132] and they were recruited. The study participants were above the age of 18 years and without any ophthalmological abnormality. A total of 83 individuals were recruited from these 4 families after complete ophthalmic examination.

***Phenotype characterization*** The members of the selected pedigrees underwent complete ocular examination including visual acuity, measurement of refractive error, IOP measurement, slit lamp examination and funduscopic evaluation. It was taken into point to exclude individuals below 16 yrs of age as some of the parameters measured may be affected by growth of the eye.

***Quantitative trait measurements*** The QTs included for the current study were selected based on their significance as risk factors for glaucoma. These Qts measured are: CCT, optic cup size, optic disc size and VCDR. CCT was measured using an ultrasonic pachymeter. Three measurements were made consecutively on each eye for each individual with each device. The average value was recorded in the database. Measurements of the optic nerve diameter and the optic cup diameter was obtained from from Heidelberg Retinal Tomograph III.

***Chapter 2.2:*** **POAG cases and controls**

Ninety seven patients diagnosed with POAG (both high tension (HTG=66) and normal tension glaucoma (NTG=31)) were recruited from the glaucoma clinic of Sankara Nethralaya eye hospital and the Chennai glaucoma study, an epidemiology study on the prevalence of glaucoma in South India. The patients were all >40 years of age and underwent complete ophthalmic examination which included measurement of best corrected visual acuity, slit lamp biomicroscopy, applanation tonometry, gonioscopy, tachymetry, dilated fundus examination including stereobiomicroscopic evaluation of the optic disc and macula with a 78 D lens and examination of the retina using the indirect ophthalmoscope, Humphrey visual fields and optic disc documentation were included in the study. The standardized inclusion criteria were presence of glaucomatous optic neuropathy with compatible visual field loss, open angle on gonioscopy and in case of HTG a mean IOP consistently higher than 21mmHg without treatment. Unrelated age matched healthy controls (n=371) were recruited after complete ophthalmic examination and informed consent.

***Chapter 2.3*** **XFS/XFG cases and controls**

All subjects and controls underwent a comprehensive ocular examination including vision and refraction, slit lamp assessment, applanation tonometry, gonioscopy and dilated stereobiomicroscopic optic disc and retinal examination. Presence of pseudoexfoliation was diagnosed if the pseudoexfoliative material was detected on pupillary ruff or anterior lens capsule. The corneal endothelium and inferior angle were also examined for additional deposits. XFG was diagnosed in the presence of glaucomatous optic disc damage with associated visual field damage (Humphrey visual field 24-2). Cases with an elevated IOP (22 mm Hg or higher) with a healthy optic disc and visual fields were classified as ocular hypertension (OHT). Only bilaterally phakic subjects were enrolled as controls.

***Chapter 3.1and Chapter 3.2- Myopia***:

Cases were defined with a refractive error worse than -6.00 D and controls with +0.50 to -0.50D in the least myopic eye. Subjects with other ocular diseases that predispose / associated with myopia were excluded from the study. Informed consent was obtained from the patients and unrelated healthy controls for the research use of peripheral blood samples after comprehensive ocular examination which included the standard retinoscopic technique to evaluate the refractive status and the corneal curvature values were obtained from Bausch and Lomb keratometer. Measurement ACD and AL were performed with an optical biometer (IOLMaster; Carl Zeiss, Germany).

2. **Blood collection:** Venous blood was collected from cases and controls / families after obtaining their consent. The blood was collected using 10ml vacutainers with sodium heparin as the anticoagulant.

**3. DNA Extraction:**

**Preparation of reagents**

*Proteinase K*: 126 mg of proteinase K [lyophilized form] is dissolved in 5.75 ml of proteinase buffer. It is stored at 4 °C.

*Wash Buffer:* To 50 ml of wash buffer [BQ2] provided in the kit, 200 ml of absolute alcohol (ethanol) is added.

**Procedure**

Lysing of Blood samples

- 500 ml of proteinase K is taken in a 50 ml collection tube.
- Blood sample is added along the sides of the tube containing proteinase K.
- If the volume of the blood is <10 ml, the volume is made up to 10 ml with PBS.
- If the volume of the blood is <5 ml, the volume is made up to 5 ml with PBS and the volume of proteinase K, lysis buffer and ethanol is also reduced to half the volume mentioned.
- 10 ml of lysis buffer is added to the blood samples.
- The tube is shaken vigorously for 4 min.
- The tube is placed in 56 °C water bath in shaking for 15 min.

Adjusting DNA binding conditions

- 10 ml of chilled absolute alcohol is added to the tube after the lysate had cooled to the room temperature and shaken vigorously for 4 min.

Binding of DNA

- Half of the lysate is transferred to the Nucleospin® Blood XL column placed in a 50 ml collection tube and centrifuged at 4500 rpm for 3 min.
- The flow through is discarded into 3% sodium hypochloride solution after removing the column.
- The sides and the rim of the collection tubes are wiped clean and the remaining lysate is transferred and centrifuged at 4500 rpm for 4 min.
- The flow through is discarded and the sides of the tubes are wiped clean.

Washing and drying of the silica membrane

- The column is washed with 7.5 ml of wash buffer and centrifuged at 4500 rpm for 2 min.
- The above step is repeated with the centrifugation step being increased to 20 min to ensure complete drying of the column.
- The column is transferred to a fresh 50 ml elution tube.

Elution of DNA

- 750 µl of elution buffer pre-heated to 70 °C is added to the column and left at room temperature, overnight.
- The DNA is eluted completely by centrifuging at 5000 rpm for 6 min.
- This is transferred to a 1.5 ml vial labeled as first elute. It is placed in a rotor overnight at room temperature and stored at 4 °C if it is going to be used immediately or at -20 °C for long term storage.
- 1000 µl of elution buffer pre-heated to 70 °C is added to the column and left at room temperature, overnight.
- The next day the tube is centrifuged at 5000 rpm for 6 min.
- The flow through is transferred onto the column and this process is repeated everyday till fifth day.
- At the end of fifth day, the final elute is transferred in to a 1.5 ml vial labeled as 2-5 elute and stored at 4 °C or -20 °C.

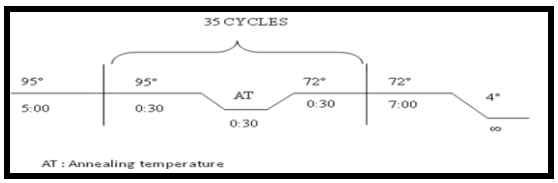
## 4. Quantification of extracted DNA

Quantification of nucleic acid is done in Nanodrop which is based on the principle of a spectrophotometer, at a wavelength of 260nm and 280nm according to the principle of Beer Lambert's law. The ratio between the optical density (OD) reading at 260nm and 280nm provides an estimate of the purity of DNA and should have a value of 1.8. If there was contamination with protein or phenol the OD 260/OD 280 will be significantly lesser

than the value mentioned above and accurate quantification of nucleic acid will not be possible.

## 5. PCR Reaction protocol

| Reagents | Concentration | Volume(µl) |
|---|---|---|
| DNA | 50ng | 1 |
| Froward primer | 2pmol | 0.2 |
| Reverse primer | 2pmol | 0.2 |
| Taq | 0.6units | 0.2 |
| DNTPs | 0.5µmol each | 2 |
| Taq buffer | 1.5mM final con | 2 |
| Water | | 14.4 |




## 6. Agarose Gel Electrophoresis:

The amplified products were then loaded into the agarose gel (2%) to check for amplification.

**Requirements**

1. Gene Ruler 100Bp: Molecular weight marker (Thermoscientific)
2. Agarose(SRL)
3. 10X TBE buffer

Tris - 500gms

Boric acid - 27.5 gms

EDTA - 3.72gms

Distilled water - 500ml

4. Ethidium bromide (2mg/ml concentration)
5. Tracking dye- Bromophenol blue (BPB)

Bromophenol blue - 0.1gm

1X TBE buffer - 100ml

Sucrose - 4 gms in 100 ml water

Mix equal volumes of 0.1% BPB and sucrose.

**Preparation of agarose gel**

Clean the gel trough with ethanol and seal the ends with the cellophane tape. Place the Combs in the respective positions to form wells. Weigh 0.5gm of agarose and dissolve in 25ml of 1X TBE Buffer (2% agarose gel). Melt the agarose in microwave oven and add 8µl of ethidium bromide to the molten agarose. Mix it and pour on to the sealed trough and allow it to set in dark.

**Agarose gel electrophoresis**

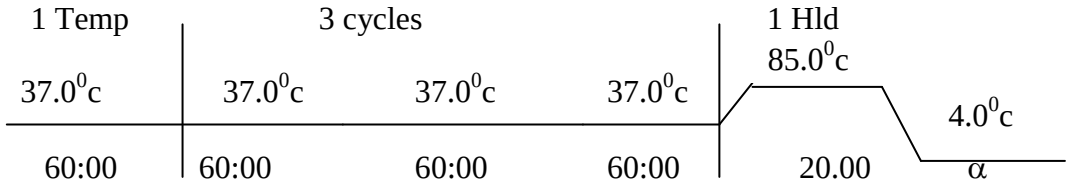
After the gel gets solidified , remove the cellophane tapes and combs & place the trough in electrophoresis tank containing 250ml of 1X TBE Buffer.Mix 5µl of amplified DNA product with 5µl of 0.1% bromo phenol blue and load on to the wells. Dilute 1µl of Molecular weight marker with 4µl of 1X TBE Buffer and mix with 5µl of bromo phenol blue and load onto the wells. Run electrophoresis at 120v for 20 to 30 min. Capture the gel image by gel documentation system from Biorad and analysis using Quantity one 1-D software.

**7. RFLP**

**RFLP reaction protocol**

| Reagents | Volume(µl) |
|---|---|
| Enzyme (2units) | 0.1 |
| Buffer | 2.0 |
| Product | 1.0 |
| Water | 5.0 |

**Machine protocol for RFLP**:




**8. Cycle Sequencing Protocol**

| Reagents | 10µl |
|---|---|
| Amplified Product (µl) | 1.0 |
| Primer(diluted) (µl) | 1.0 |
| Sequencing buffer(µl) | 1.0 |
| RR mix (µl) | 0.1-0.6 |
| Milli QWater (µl) | 6.9-6.4 |

Depending on the length of the amplified product the RR mix volumes were changed and the corresponding volume was adjusted with water

**Reaction conditions of Cycle Sequencing**

| | | |
|---|---|---|
| Initial denaturation | 96°C | 60 seconds |
| Denaturation | 96°C | 10 seconds |
| Annealing | 50°C | 5 seconds |
| Extension | 60°C | 4 minutes |

The reaction was carried out for 25 cycles.

*Purification of Extension Products*

Purify the extension products helps to remove the unincorporated dye terminators before the samples are analysed by electrophoresis. Excess dye terminators in sequencing reactions obscure data in the early part of the sequence and can interfere with base calling.

**Reagents Required**:

I. 50mM EDTA.
II. 3M Sodium acetate(pH 4.8).
III. Absolute Ethanol.
IV. 70% Ethanol.

**Procedure:**

- Do a one in four dilution of 50mM EDTA with milliQwater.
- Add in a 0.5ml vial 10µl of autoclaved milliQ followed by 2µl of diluted EDTA.
- Add to it 50µl of absolute ethanol and 2µl of 3M sodium acetate.
- To the above mixture add 10µl of the cycle sequencing product and mix.
- Keep it in room temperature for 15 minutes. Microfuge at 8,000rpm for 20 minutes. Discard the supernatant and to this add 250µl of 70% ethanol.
- Vortex and microfuge at 8000rpm for 10 minutes.
- Discard the supernatant and cover the vial with paraffin film , perforated and dried at $37^0$C.
- Before loading 10-15uL of Hi-di formamide / template suppressor reagent is added, vortexed, denatured at $95^0$C for 5 minutes and then loaded on to the ABI PRISM genetic analyzer.

9. **Sequencing in *ABI Prism 3100 AVANT* genetic analyzer**

The sequence of the PCR amplified DNA is deduced with the help of the ABI Prism 3100 AVANT genetic analyzer that works based on the principle of Sanger dideoxy sequencing. The enzymatic PCR reaction involves amplifiation of the target gene sequence in multiple copies with the help of Taq polymerase enzyme that helps in the addtiion of the nucleotides to the 3'end thus resulting in chain elongation through the formation of phosphodiester bond between the 3' hydroxyl end at the growing end of the primer and the 5' phosphate group of the incoming deoxynucleotide. Thus a block in this phospho di ester bond results in abrupt chain termination that is aided by the addition of dideoxy nucleotide tri phosphate. This method was devoloped by Sanger *et al* in 1981. Thus at the end of the reaction, a pool of products with the base pair length ranging from 1 to the total product size were separated by electrophoretic methods.

The fluroscent based detection by automated sequencer adopts the Sangers method and incorporates the fluorescent dyes into DNA extension products using 5'-dye labeled primers or 3'-dye labeled ddNTPs (dye terminators called commercially as RRMIX). Each dye emits light at a different wavelength when extited by an argon ion laser. All four colours and therefore all four bases can be detected and distinguished in a single gel lane or capillary injection.

The amplified products with the dye at the terminated 3'end were subjected to capillary electrophoresis by an automated sample injection. The emitted flurorescence from the dye labels on crossing the laser area were collected in the rate of once per second by cooled, charge-coupled device (CCD) camera at particular wavelength bands (virtual filters) and stored as digital signals on the computer for processing that are analysed by software called the sequence analysis software (seqscape manager in ABI 31000 AVANT machine). The ABI AVANT genetic analyser can be upgraded from 4 capillary to 16 capillary to that facilitates the electrophoresis of 16 samples at a given time.

The four nucleotide bases with the respective acceptor dyes and colour emission.

| Terminator | Acceptor dye | Colour of raw data on electrophoretogram |
|---|---|---|
| A | dR6G | Green |
| C | dROX | Red |
| G | dR110 | Blue |
| T | dTAMRA | Black |

10. **Details of Bioinformatic tools**

**i) Fischer Exact test**: Is a statistical significance test used in the analysis of contingency tables. This method is usually employed for small sample sizes. It is one of the class of 'exact test' wherein the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, as with many statistical tests. The test is considered to be significant if the p value is equal to or less than 0.05.

ii)**. Haploview**: It is commonly used bioinformatics software which is designed to analyze and visualize patterns of linkage disequilibrium (LD) in genetic data. Haploview can also perform association studies, choosing tagSNPs and estimating haplotype frequencies. Haploview is developed and maintained by Dr. Mark Daly's lab at the MIT/Harvard Broad Institute.

In the present work Haploview program was used for:

- LD & haplotype block analysis
- Haplotype population frequency estimation
- Single SNP and haplotype association tests
- Permutation testing for association significance

The data input was given in linkage format (pedigree file (.ped) and data file (.dat)). The linkage disequilibrium (LD) pattern was created using solid spine analysis. This internally developed method searches for a spine of strongest LD running from one marker to the other along the legs of the triangle in the LD chart, meaning that the first and last marker of the block are in strong LD with each other while the intermediate markers are not necessarily be in LD.

**iii) PLINK:** It is being developed by Shaun Purcell at the Centre for Human Genetic Research (CHGR), Massachusetts General Hospital (MGH), and the Broad Institute of Harvard & MIT, with the support of others. In the present study logistic regression was performed for haplotype association using sliding window technique in this software. This method of assessing the haplotype association provides more power in identification of association rather than a single marker analysis because here for a given window size

within a gene under consideration, the test was performed for all window sizes shifting one SNP at a time toward the 3' end of the gene .

**iv) THESIAS- Testing Haplotype Effects In Association studies**: The objective of the program is to perform haplotype based association analysis in unrelated individuals. It is based on maximum likelihood model and linked to SEM algorithm. It allows the simultaneous estimation of haplotype frequencies and their associated effects on the phenotype of interest. The main output includes the allele frequency of all the studied polymorphisms, estimated haplotype frequencies, estimated haplotype and covariate effects, LD matrix, $R^2$ information etc...

**v) Polyphen:** Polymorphism Phenotyping is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.

**vi) SIFT:** Single nucleotide polymorphism (SNP) studies and random mutagenesis projects identify amino acid substitutions in protein-coding regions. Each substitution has the potential to affect protein function. SIFT (Sorting Intolerant From Tolerant) is a program that predicts whether an amino acid substitution affects protein function so that users can prioritize substitutions for further study.

**vii) SNAP (SNP annotation and Proxy search)**: this method is used to evaluate the effects of SNPs on protein function. The SNPs are either ascertained as neutral or non neutral. Neutral SNPs are those that do not change the functionality of the mutated from the wild type whereas the SNPs are quoted as non neutral if the mutant and wild type differ in function. The ability to identify non-neutral substitutions in an ocean of SNPs could significantly aid targeting disease causing detrimental mutations, as well as SNPs that increase the fitness of particular phenotypes. It was developed by Yana Bromberg in Rost Lab, at Columbia University, New York.

SNAP is a neural-network based method that uses in silico derived protein information (e.g. secondary structure, conservation, solvent accessibility, etc.) in order to make predictions regarding functionality of mutated proteins. The network takes protein sequences and lists of mutants as input, returning a score for each substitution. These scores can then be translated into binary predictions of effect (present/absent) and reliability indices (RI).

**viii) FastSNP (Function Analysis and Selection Tool for Single Nucleotide Polymorphisms**): is a web server that allows users to efficiently identify the SNPs most

likely to have functional effects. It prioritizes SNPs according to twelve phenotypic risks and putative functional effects, such as changes to the transcriptional level, pre-mRNA splicing, protein structure, etc. A unique feature of FASTSNP is that the prediction of functional effects is always based on the most up-to-date information, which FASTSNP extracts from eleven external Web servers at query time using a team of re-configurable Web wrapper agents. These Web wrapper agents automate Web browsing and data extraction and can be easily configured and maintained with a tool that uses a machine learning algorithm. This allows users to configure/repair a Web wrapper agent without programming. Another benefit of using Web wrapper agents is that FASTSNP is extendable, so we can include new functions by simply deploying more Web wrapper agents. One of the new functionalities included in is the ESE finder which is used to detect whether the SNP is inside any Exonic Splicing Enhancer (ESE) and how it reacts to various SR proteins (seriene arginine rich proteins) which are required for alternative and constitutive splicing.

ix) **MUpro software:** It is a set of machine learning programs which studies the protein sequence given and predicts how SNPs affect the protein stability. The two machine learning methods contained in this suite is Support Vector Machine (predicts the value of energy change ($\Delta\Delta G$)) and Neural Networks. The result is depicted in the form of confidence score of values -1 to 1. Values less than zero means stability decreases with the mutation, conversely score greater than 0 indicates increased protein stability.

**x) CAIcal (Codon adaptation index) software:** This was developed to measure the synonymous codon usage bias for a DNA and RNA sequence. It calculates the synonymous codon usage of the gene with that of the reference set. The index value ranges from 0 to 1, being 1 if the gene always uses the most frequently used synonymous codons in the reference set.

**xi) Target scan software:** This software predicts biological targets of miRNAs by searching for the presence of conserved 8mer and 7mer sites that match the seed region of each miRNA. In mammals the predictions are ranked based on the predicted efficiency of targeting as calculated by the context +scores of the sites.

11. **Affymetrix work flow**:

i) **Genomic DNA preparation:** The concentrations of the genomic DNA were determined and dilutions were made accordingly to obtain a working stock of 50ng/µl

concentration using reduced EDTA TE buffer (pH 8.0). 5.0 µl of each DNA aliquot was taken for the next step.

ii) **Restriction Enzyme Digestion:** The digestion master mix was prepared on ice according to table 3.2.1

**Reaction protocol for restriction enzyme digestion**

| **Reagent stock** | **Per sample (in µl)for 10K** | **Per sample (in µl) for 250K** |
|---|---|---|
| Water | 10.5 | 11.55 |
| NE buffer 2(10X) | 2 | 2 |
| BSA | 2(10X(1mg/mL)) | 0.2(100X(10mg/mL)) |
| Restriction enzyme | 0.5 XbaI (20U/ µl) | 1 µl Nsp I(10U/ µl) |
| Total | 15 | 14.75 |

5 µl of genomic DNA was added to the digestion reaction mixture and vortexed and the reaction was performed in thermocycler for 37˚C (120min) followed by 70˚C (20min) for 10K and 65˚C (20min) for 250K and kept at 4˚C hold.

iii) **Ligation:** The ligation mastermix was prepared on ice according to the table 3.2.2

**Reaction protocol for preparation of ligation mixture**

| **Reagent** | **Per sample (in µl) for 10K** | **Per sample (in µl) for 250K** |
|---|---|---|
| Adaptor | 1.25 Xba (5µM) | 0.75 Nsp I (50 µM) |
| T4 DNA Ligase buffer (10X) | 2.5 | 2.5 |
| T4 DNA Ligase (400U/ µl) | 1.25 | 2 |
| Total | 3.75 | 5.25 |

The digested genomic DNA was added to the ligation mixture. This mixture was vortexed and kept in thermocycler for 16˚C (120min, 10K; 180min, 250K) followed by 70˚C (20min) and 4˚C hold. This ligated product was diluted with 75 µL molecular biology grade water giving a total volume of 100 µL.

iv) **PCR:** PCR master mix was prepared on ice according to table 3.2.3

**PCR reaction protocol**

| **Reagent** | **Per sample (in µl) for 10K** | **Per sample (in µl) for 250K** |
|---|---|---|
| PCR buffer (10X) | 10 | 10 (Titanium *Taq*) |
| dNTP (2.5mM each) | 10 | 14 |
| MgCl2 (25mM) | 10 | - |
| GC-Melt (5M) | - | 20 |
| PCR Primer | 7.5 Xba (10µM) | 4.5 primer 002(100 µM) |

| | | |
|---|---|---|
| *Taq* (5U/µL) | 2 Amplitaq Gold | 2 (Titanium *Taq*) |
| Water | 50.5 | 39.5 |
| Total | 90 | 90 |

A total four reactions were set for each sample in case of 10K while 3 reactions were set for 250K. 90 µL of PCR master mix was added to 10 µL of diluted ligated DNA making a total of 100 µL. This reaction mixture was then placed in thermocycler and programmed as follows

| **Temperature** ˚C | **Time (sec)** | **Cycles** |
|---|---|---|
| 95 | 180 | 1X |
| 95<br>59<br>72 | 30<br>30(10K),45(250K)<br>30(10K),15(250K) | 35X(10K)/30X(250K) |
| 72 | 420 | 1X |
| 4 | Hold | |

3 µL of the PCR product was mixed with 3 µL of 2X gel loading dye on 2% TBE gel at 120V for 1hr.

v) **Purification, elution and quantification of PCR product:** Purification was performed using 96 UF PCR purification plate and vacuum manifold. All four PCR products were consolidated into a single well of the purification plate and vacuum of ~ 800mbar (10K)/ ~ 600mbar (250K) was applied. In case 250K the PCR products were diluted with 0.1M EDTA (8 µL) before loading into the purification plate. It takes about 90mins-2hrs to dry PCR sample. The PCR products were washed three times with 50µL of biology grade water for approximately 20 mins. Following this 40 µL elution buffer (10K)/ 45µL of RB buffer (250K) was added to elute the product. A 40fold (10K)/ 100 fold (250K) dilution was performed for each sample and then quantitated using NanoDrop. Dilutions were made according to the readings obtained to make a final concentration of 20 µg/45 µL (10K)/ 90 µg/45µL, this was further taken for fragmentation.

vi) **Fragmentation:** 5 µL of fragmentation buffer (10X) was added to the purified PCR product making a total volume of 50µL. The fragmentation reagent was prepared separately according the manufacturer's instructions. 5 µL of this reagent was then added to the fragmentation mix and kept in thermocycler at 37˚C (30min,10K/35min, 250K), 95˚C (15mins) and 4˚C hold. 4 µL of fragmented PCR product was diluted with 4 µL gel loading dye and run on 4% TBE buffer at 120V for 30 min-1hr.

vii) **Labelling:** The labelling master mix was prepared on ice according to the following table

| Reagent | Per sample (in µl) for 10K | Per sample (in µl) for 250K |
|---|---|---|
| 5 TdT buffer | 14 | 14 |
| Gene chip DNA labelling reagent | 2 | 2 |
| TdT (30U/ µL) | 3.4 | 3.5 |
| Total | 19.4 | 19.5 |

To labelling mixture, fragmented DNA was added. After brief vortexing and spinning the mixture was kept in thermocycler at 37˚C (2hrs, 10K/4hrs, 250K), 95˚C (15min) and 4˚C hold.

viii) **Hybridization :** The hybridization cocktail was prepared according to the table 3.2.5

**Preparation of hybridisation reaction protocol**

| Reagent | Per sample (in µl) | Final concentration |
|---|---|---|
| MES (12;1.22M) | 12 | 0.056M |
| DMSO (100%) | 13 | 5% |
| Denhardt's solution (50X) | 13 | 2.5 |
| EDTA (0.5M) | 3 | 5.77mM |
| HSDNA (10mg/mL) | 3 | 0.115mg/mL |
| Oligonucleotide control | 2 | 1X |
| Human Cot-1 (1mg/mL) | 3 | 11.5µg/mL |
| Tween-20 (3%) | 1 | 0.0115% |
| TMACL (5M) | 140 | 2.69M |
| Total | 190 | |

70µL of each sample was transferred to the hybridization mix. This was then subjected to denaturation at 95˚C in a heat block for 10min and cooled in crushed ice for 10secs. Following this 80 µL of hybridisation mix was injected into the array chip and placed in hybridisation chamber at 48˚C (10K)/ 49˚C (250) for 16-18hrs at 60rpm.

ix) **Washing & Staining:** After hybridisation the cocktail was removed from the probe array and in turn filled with 80 µL (10K)/ 270 µL (250K) of array holding buffer. Following this step staining of the probe was done, this involves 3 major steps stains, Streptavidin Phycoerythin (SAPE), followed by antibody amplification step and final stain with SAPE. All these procedures are done using automated protocol in the fluidics station.

**Preparation wash buffers and stains for Affymetrix chip**:

**i) Stain buffer:**

| Components | 1x (in µl) | Final concentration |
|---|---|---|
| Water | 800.04 | |
| SSPE (20X) | 360 | 6x |
| Tween-20 (3%) | 3.96 | 0.01% |

| | | |
|---|---|---|
| Denhardt's (50X) | 24 | 1x |
| Subtotal | 1188 | |

**ii) SAPE solution:**

| **Components** | **1x (in µl)** | **Final concentration** |
|---|---|---|
| Stain buffer | 594 | 1x |
| 1mg/ml SAPE | 6 | 10µg/mL |
| total | 600 | |

The solution must be prepared immediately before use and stored away from light.

**iii) Antibody stain solution**

| **Components** | **1x (in µl)** | **Final concentration** |
|---|---|---|
| Stain buffer | 594 | 1x |
| 0.5 mg/ml SAPE | 6 | 5µg/mL |
| total | 600 | |

**iv) Array holding buffer**

| **Components** | **Volume (ml)** |
|---|---|
| MES stock buffer (12X) | 8.3 |
| 5M NaCl | 18.5 |
| Tween-20 (10%) | 0.1 |
| Water | 73.1 |
| Total | 100 |

x) **Scanning:** Scanning was performed in GeneChip Scanner which is operated using Affymetrix GeneChip Operating Software 1.4 (Patch 5 or higher). After staining the chip was taken for scanning. As the scanning proceeds the image of the array image appears on the screen.

**12. File formats for EasyLinkage software:**

***.pro file*:**

| **Individual id** | **Family id** | **Father** | **Mother** | **Gender** | **Affected status** |
|---|---|---|---|---|---|
| **1** | **1** | **0** | **0** | **1** | **0** |
| **1** | **2** | **0** | **0** | **2** | **0** |
| **1** | **3** | **0** | **0** | **1** | **0** |

***.snp file***

| SNP_ID | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP_A-1517367 | AB | AA | AA | AA | AA | AB | AB | AB | AA | AA | AA | AB | AA | |
| SNP_A-1517259 | AA | 00 | AB | AA | AA | AA | AB | AB | AB | AB | AA | AA | BB | |
| SNP_A-1508602 | AB | AB | AB | AB | AB | AB | AA | BB | AB | AA | BB | AA | AB | |
| SNP_A-1508076 | AB | 00 | AB | AB | AA | AB | AB | AB | AB | AB | AB | AA | BB | |
| SNP_A-1510443 | AB | 00 | AB | AB | AA | AB | AB | AB | AA | BB | AA | AB | AB | |
| SNP_A-1511588 | AA | AA | AA | AA | AA | AA | AA | AA | AA | AA | AA | AA | AA | |
| SNP_A-1509634 | AA | 00 | AA | AA | AB | AA | AB | AB | AA | AA | AB | AB | AA | |
| SNP_A-1510330 | AB | AA | AB | BB | AB | BB | AB | AB | AB | BB | BB | AB | BB | |
| SNP_A-1512761 | AB | 00 | AB | AA | AA | AA | AA | AA | AB | AB | AA | AA | AA | |
| SNP_A-1509046 | AB | 00 | AB | AB | AB | AA | AA | AA | AA | AB | AA | AA | AA | |
| SNP_A-1517931 | AA | 00 | AA | AA | AB | AB | AA | AA | AA | AA | AA | AB | AA | |
| SNP_A-1518938 | AA | AA | AA | AA | AB | AA | AA | AB | AA | AA | AB | AB | AA | |

## 13. Reference Sequences

**Note:** Bold font refer to forward and reverse primers

**rs1900004**

**ACCAAATGCTCACCCAGTGT**TAACTCCAATTCTTATAACCCTCTTTGAGTGAGGACAGTT
GTAATCCCATTTTGTGGCCAAGAAAACCAAGGCCCTAGTTAAACTGAGCAACACAGCAGC
CAGGTTTGAACCCAGCTCTCTTTCTACACAGAGCTCATGCTCTGTGCTGCACTGGCCCTTGT
CTGTGAGAGCTTTGATGGCAGGAACA**C/T**GGCCTTATTCTTCACGGTGTACCCAGTTCCTAG
AAAAGGAGCTGGGACCCGGTTGCACTCATTAAATATGCTGGGATGTGTAAGCAGTCACTGT
GTCAGGCCCAGCTGGCCCCCATTCTGCTAAACTATAGGCACA**GGATCTAGCAATCTGCAG**
**GC**

**rs3858145**

**TGCAAAGCAAAGTGTTTACCT**AAGTTACCAGTGCTAAATCTAAATATCCATGCAATTATA
AATTTAGATGGAAACTAAAATTGTGTTTCCAATTAAGGAAAGAAACCCCTTAGACTTTTTC
AGTTCAGCCTAACAAGCAAGTTACTAATTGTTTCAGTTTTGGTTTTGCTACAAATCATTTAG
AAGGAAAGAGGATCTGAGGTGCAAAGAGGAAGGCA**R**TTGTAATAATGGGAAATTGGACT
AAATTAAAATGATCCTAAGTCACAGCACAACTAGGCCTCTTTGCCACCAACTTTTGACATT
CTAATTAGGGCAAATGATTATGCAAATCCAGTCAGTCGCTCTTTTCCCCCCAAGCTGAAGT
GCTTGTGATTGTTC**AAGTGAGGACTGTTGGTGGC**

**rs17231602**

**GCACAGCTGTCCTTTTGTCC**TGCCTGCCACCAGGACGCCTCGAGTTTCCCCATCAAGTAT
AATAGAATGTGACACTTGAGGTGTTCTCAAGCCTTGGTCCTCATGAGTGGAGTGGAGGACA
GGACAGCAGGAGTCCAGCACACCTAGGTCACCTCCTTCTGGGCATGTTAGCTGAGGAAAC
AGGGTTGGAGGGTRAACCCAAACATGGAGCACAGAGCACCTGAGTTTGGGGGCACCTATG
AAATAAAATTGATCATGGGTTGATAAGCATTGAAACTGCATGACATGTACTTGGTTGGTTG
GTGGCGGGGTAGGGTAAGTATAGTAATTTGCTTTGCTTTTGTATATGTTTAAAATATTAATT
TTTTAAAATTAATGC**ATGCAATGCCAAAAATAGCA**

**rs4746741**

**CATGTTGGTCAGGCTGTTC**TCAAACTTCTAACCTCAGGTGATCCACCTGCCTCGGCCTCC
CAAAGTACTGGGATTACAGGCATGAGCCACTGCGCCTGGCCCACCTGTGATTTTTTTAAGT
GTAAATGTATGTATGGGCTGGGGAAGACACTTTCCCTCATTACAMACCAAACGGATTTTGA
TAGGGATAACTAAGAAAATAAAAATTGTAATAAAGTCTGTAGATAAACTGTGAAGAGGCA
GCTGTGTCTGAGGTTAGGCTTGGGAATCCAAAATAAAAGAGTTTTGGGAGTCATTGGTGTT
TTTGTTGACTTGTGACAAGACAAATTCCAAATG**CATGGATAACTTAGAGTCCTGC**

**rs77330683**

**GGCTCCGTCTTTAAGCACAA**ATCCCCACCGCCCCSCCGGCCACCTCCCTGTCATTTTCGA
AGACATCATTTTCTTTCCCCAYGTGCCCCCTTATCTCCCTCCTGTTAAAGGATCTCGAATAG
GCACCCCGATAA**GACAGGACGAACACCATTCC**

**rs10483727**

**ATGGGAGCCTACCTCTTGCTTC**GTTATGAATCTCTAAACAAAAAATGGACACTTACTYAT
ATGGCATGAGGCCATTATCATATTTAACAAAATTAGCAATAATTTTTAGTGGCATATAACA
AGAGTTTATGAATAAATTTCCAAAATTGTTTCCAAAAACTCTTTTTATAGTGTGTTTGAATT
TAGTTGTTATGTCTCCTAAATCTCTT**TCTCGTTTTTAACCCCCATG**

**rs1063192**

**AGTTGCATTATACTGGGTCATGA**AAAATTATCCCTTGAAATAGATATGAAACATGTTACT
TCATTTCTGGTTTAAATAACTTGTGGAATCTTTCCTAATGAC**A**ACYTGATATTAAGGGAAA
CTAAAGAAAATGTTATTGTGGATCCCACAGTACTATATTACACTGTTTTTTTTGTTTGTTTT
GTTAGTTTTTTTTATTTAAAGCAAACCTCAAACATTATTGGGTATCAATTACCACCTGGTTG
TATTGAAATAGTAACTTATCAATGCCATGTAAAAATTAATTCCATTTTCGAAGCCACCTGG
CAGACAGGTTTAGCTGTTTCATCAGCAGCCTAATATATACTGTTAAATTTGTTAAGGATTTC
ACTTTGAAGGATACATGCAAAACATATAGTTACTATTTTCATGAGTCCTGCTTCTA**GCTCC**
**ATTGTGGAATACAGAAA**

**rs9607469**

**GAGAACAGCCTCCTCCTGTG**CTTTGGGAGAGACAGAGGATTGACAGASAGGATGGGGGA
AAGGTYGGCAGAGAGACCTTGAGGTTGCGTMTTTATTCAGCCTATCAAAGC**R**CCTTATTTG
GCAGGTATCACTTTCTGAGCCCCGAGAGCTATTTCCTCTCTATCTGACATTTTCCTGGTGCT
TGCTTTGCTGCACCAGGAAAATGTCAGCAACCTGTKW**GTTGCCTACTTCCATCCCAA**

**rs1192415**

**GCCTTAAGATCTTCACAAACAAACAA**ATCTTTCTGAAATAGAAAAAAAAAAAAAACAGAC
CCAAAAATCTAACACTATAGTGACCTCCTACTAGGCTTTAAGTTTAAACGGATTAGGGCAG
GAAGTTTTAC**R**TTCCTGTCGTATTTGCTGAAGGTGTATGGCAGGGTGCGGCGCCTGTTGCA
GGATTCCAGCAGGGCCTGCCCTGGCTCCCCAGGGCTTTGCAAGGAAGCTG**GAGAAGTCCG**
**CCACCTCG**

***LOXL1***

**ATTCGGCTTTGGCCAGGT**GCCCGACAACTGGCGCGAGGTGGCCGTCGGGGACAGCACGG
GCATGGCCCGGGCCCGCACCTCCGTCTCCCAGCAACGGCACGGGGGCTCCGCCTCCTCGGT
CTCGGCTTCGGCCTTCGCCAGCACCTACCGCCAGCAGCCC**TCCTACCCGCAGCAGTTC**

***VSX2***

**EXON 1**

**GGCACCTGGGACCAACTTCGC**GAAGCGGGAAGCCCGGCGGGGGGGTGGGGGGAGCTAA
AGACCTGCGGCCTCAGCCCCTCCAAAGAACAGGGAGATGACGGGGAAAGCAGGGGAAGC
GCTGAGCAAGCCCAAATCCGAGACAGTGGCCAAGAGTACCTCGGGGGGCGCCCCGGCCAG
GTGCACTGGGTTCGGCATCCAGGAGATCCTGGGCTTGAACAAGGAGCCCCCGAGCTCCCA
CCCGCGGGCAGCGCTCGACGGCCTGGCCCCCGGGCACTTGCTGGCGGCGCGCTCAGTGCTC
AGCCCCGCGGGGGTGGGCGGCATGGGGCTTCTGGGGCCCGGGGGGCTCCCTGGCTTCTAC
ACGCAGCCCACCTTCCTGGAAGTGCTGTCCGACCCGCAGAGCGTCCACTTGCAGCCATTGG
GCAGAGCATCGGGGCCGCTGGACACCAGCCAGACGGCCAGCTCGGGTAGGTGAGGAGAG
GTT**GCGCTGCTGCCTGACTGGGG**

**EXON2**

**GTTTCGGGCACAGCGGAGCGC**GTCCCCCACTCTGCCGCATGTCCCTACGCCCGCTTTTCA
GATTCTGAAGATGTTTCCTCCAGCGATCGAAAAATGTCCAAATCTGCTTTAAACCAGACCA
AGAAACGGAAGAAGCGGCGACACAGGTGAGGCCCCCGCTTTCTGTTACCTCTCCTGCCCG
CGCACCCCGCTGCCGTCCCCCGTTCCGGGATCGGGTCTCTCGGACCCTATTTT**CGCCGACA**
**ACGGATCTGAGGTC**

**EXON 3**

**GTTCTGTCTTGTCTGAGACAGGC**TCTTTTAGTTTCTGGGGTCCTGAGTGTTCGGTCTTGCT
CCCCTCAGGACAATCTTTACCTCCTACCAGCTAGAGGAGCTGGAGAAGGCATTCAACGAA
GCCCACTACCCAGACGTCTATGCCCGGGAGATGCTGGCCATGAAAACGGAGCTGCCGGAA
GACAGGATACAGGTAACAGCCCTGAGCCCCTCTCCCCTCCACTCTCCCCTCTCTCG**GTGAC**
**ATCTACCACTCCAGGGTT**

**EXON 4**

**GAAATCTTCACTCCAAGCCTACAAG**GGGTAAGGGCCTTGGGCCCAGCACCTCTCAGAGC
AAGCCTCTGACCTGTTCTGTGCACCTGCAGGTCTGGTTCCAGAACCGTCGAGCCAAGTGGA
GGAAGCGGGAGAAGTGCTGGGGCCGGAGCAGTGTCATGGCGGAGTATGGGCTCTACGGGG
CCATGGTGCGGCACTCCATCCCCCTGCCCGAGTCCATCCTCAAGTCAGCCAAGGATGGCAT

CATGGACTCCTGTGCCCCGTGGCTACTGGGTAAGAGCCCGCACCCTCCTTGGGGTCCTGCC
**CTGCGGTGAGGAGAGCGGGCT**

**EXON 5**

**GGGAGTAAGGCTTTCTGCTCGTC**CTTAATTCTGGCCTCTCTCTATCTTTGCCGTTTTCAGT
TCAAGATGGCTTTCCCAGGCGCTTTTCTAAACCCGAATACCAACAATTCTTTCTAGGGATG
CACAAAAAGTCGCTGGAGGCAGCAGCCGAGTCGGGGAGGAAGCCCGAGGGGGAACGCCA
GGCCCTG/ACCCAAGCTCGACAAGATGGAGCAGGACGAGCGGGGCCCCGACGCTCAGGCG
GCCATCTCCCAGGAGGAACTGAGGGAGAACAGCATTGCGGTGCTCCGGGCCAAAGCTCAG
GAGCACAGCACCAAAGTGCTGGGGACTGTGTCTGGGCCGGACAGCCTGGCCCGGAGTACC
GAGAAGCCAGAGGAGGAGGAGGCCATGGATGAAGACAGGCCGGCGGAGAGGCTCAGTCC
ACCGCAGCTGGAGGACATGGCTTAGGTCAAGGCGCGCT**CAGATGCCGGAGCCCCAAGAC**
**TCT**

## ***MFRP***

**EXON 1-2:**

**CTTTTCCCTCGGGTAGTAGA**GGCTGCCCCTCTGACAGGGAGTTTGTGGTCAGCGAGTGGG
CCCTTGGGGGTCTTGGGCTGTCACAGGGACTCCAGAAAGCCATGCAGAGCCTTGTGCCATG
AAGGACTTCTCAGATGTCATCCTCTGCATGGAGGCAACAGAATCGAGCAAGGTAAGGACC
ACATAGAATCGTGCTCCTAAGACCCAGCACCAGTGGTCACTGCTGCCCTCAGCCCTCAAGT
ATCTCCCCCTTCCAGACCGAGTTCTGCAATCCTGCCTTCGAGCCTGAGTCTGGGCCACCCTG
CCCTCCCCCAGTTTTCCCAGAGGATGCCAGCTACAGCGTCCCAGCTCCCTGGCATGGTAAC
CATCCCAGGGGGTGACCTGGGGGAGGACAGAGAGGCCCAGGAGCCAAGGACAGCATGAG
GAATACCCAACAGAAGGGGCT**TTCAAAACCCCACCTGCCAA**

**EXON 3-4**

**AGGACAGCATGAGGAATA**CCCAACAGAAGGGGCTTTCAAAACCCCACCTGCCAACCCGC
AGGTCGGCGTCCTCGAGGGCTACGGCCAGACTGCCGCTTCTCCTGGCTCTGTGTCCTCCTG
CTCTCCAGCCTGCTCCTCCTGCTGCTTGGGCTGCTGGTGGCCATCATCCTGGCCCGTAAGTA
CCCGGGGACAGCTTGAAACGAGGGCTTTGGAATGAAACTCCATGAGCCGCTCTCCCAGCCT
CCCAGGACCCAGCTCCTCTGAACGCCACCCTCCATCTTCTCGCCTCCAGAGCTGCAGGCTG
CACCCCCATCTGGGGCGTCCCATAGCCCACTGCCTGCCGGAGGCCTTACCACGACCACCAC
CACCCCCACCATCACCACCTCTCAGGCAGCTGGGACCCCTAAAGGGCAGCAGGAGTCAGG
CGTGAGCCCCTCCCCACAGTCCAGTGAGTACTGGGGGCAGATCTTCCGGGAGAAGGGCC**C**
**AAGATGACGGGGTGGGGTT**

**EXON 5**

**GAGAAGGGCCCAAGATGA**CGGGGTGGGGTTGAGGGGAGGGTGAGGAAGAGGCACCTTG
AGAATAAAGGACCTCACTCCTTTTGACCTCTGAACTCTAACCTCCATCTCTGCAGCCTGTGG
AGGCCTCCTCTCTGGCCCAAGGGGCTTCTTCAGCAGCCCTAACTACCCAGACCCTTACCCC
CCCAACACCCACTGCGTGTGGCATATCCAGGTGGCCACAGACCATGCAATACAGCTCAAG
ATCGAAGCCCTCAGCATAGAGAGTGTGGCCTCTTGCCTTTTTGATCGCTTGGAACTCTCCCC
TGAGCCTGAAGGCCCCCTCCTCAGGTAGGTCCCTCTGTCCCCAGGACATCCCACCGTGTCC
TGAACCACTATCTAAAGAGCA**GAATCAGGGAGTGGCAGGGA**

**EXON 6-7**

**TTGGGGGTTGAGAAAATA**GGACTGCAAGGCCCAGGCAAGCTCCTCTCTGGCTGACCCTG
CTCTTGGAGCCTCCCTCACCCCCAACCTGGCTTCTTGCAGGGTTTGTGGAAGGGTGCCTCCC
CCCACGCTCAACACCAATGCCAGCCACCTCCTGGTGGTCTTCGTCTCTGACAGCAGTGTGG
AAGGATTTGGTTTCCATGCCTGGTACCAGGCTATGGCCCCTGGGCGCGGTGAGTGTGTTCC
CACCTGCCTGTTGAGTGTCCACTTTCTTCCCCTCCACCCACTCCCAGTCCTGTCAAAGGGGT
GGAGACTCCGGGCCTCCCTGACTCCTGGTGCCCACATCCAGGGAGCTGTGCCCATGATGAG
TTCCGCTGTGACCAGCTCATCTGCCTGCTACCTGACTCAGTGTGTGATGGTTTTGCCAACTG
TGCTGACGGCAGTGATGAGACCAATTGCAGTGCCAAGTTCTCGGGTACGGGCCAGGCATG
GGGGTCCTCTCTTCAGGCGTCTGATCTGGGTTCAAGCTGGGCTCCCCAGCTCTGGTCCAGC
CTCAAACCGGGCCCTCCGTCAGCATTTCCTGC**TCAGTCATTCTTTGGCCCAG**

**EXON 8**

**ATCACCGCCAGCCCTATT**GGATGGAGGCACAGATCCTAGCAAACTCATGAGCCCTTCCAC
CTCTCTACAGGGTGTGGGGGGAATCTGACTGGCCTCCAGGGCACTTTCTCTACTCCCAGCT

ACCTGCAGCAGTACCCTCACCAACTGGTAAGCACAGTGCTCCATGCAGAACGGGAGGGGC
ACACACGGAAGCACCTGATCCCCAAGTTAGTGTAATGGGAATGGCATGGTAGACACTACC
TTCTACTTGGGAAGGTCAGAG**ATCAAGCAGACGGGGGATG**

**Exon 9:**
**GGTGCCCGGGATGAGACA**GGAGAAGGGGCCATGAATTGCTCCACCTATGCCCTCCCCCA
GCTCTGCACCTGGCATATCTCGGTGCCTGCCGGACACAGCATAGAACTACAGTTCCACAAC
TTCAGCCTGGAGGCTCAGGACGAGTGCAAGTTTGACTACGTGGAGGTGTATGAGACCAGC
AGCTCAGGGGCCTTCAGCCTCCTGGGCAGGTACAGGAGCCAGGGAAGAGCCATGGATAAC
TCCATCCTC**ACTGTCTGCCACCCCCGG**

**Exon 10-11:**
**GCAGTGCCCCCTCAG**TCAGCCAGGGCTGGTGCCCAGAACAGCTGTCTGCTTTGGGCAGGT
TCTGTGGAGCAGAGCCACCCCCCCACCTCGTCTCCTCGCACCATGAGCTGGCTGTGCTGTT
TAGGACAGATCATGGCATCAGCAGTGGAGGCTTCTCAGCCACCTACCTGGCCTTCAATGCC
ACGGAGAGTAGGTGCCCCTGGGCAAGTGGGTGGGGGCAGTAGAGGCACCTCCAGGGGGA
CAGACAAGGGCTCTGGACCCTGGTCAGGACACTCTGGGGGTACTGGGGGTGTAGGTATGT
CCCCGGGCACTGCCATGCCCCTTGTGCCTGCAGACCCCTGTGGGCCCAGTGAGCTCTCCTG
CCAGGCAGGAGGGTGTAAGGGTGTGCAGTGGATGTGTGACATGTGGAGAGACTGCACCGA
TGGCAGCGATGACAACTGCAGCGGCCCCTTGTTCCCACCCCCAGGTGAGAAGCCTGCCCCT
GGGGACCCT**GAGCAGGCTGGGTGCCCAC**

**Exon 12-13:**
**CAGGCCACAGAGCCA**GTGAGCAGTCCCCACAGTGGCCCTCCCCTCCACCCCTGCAGAGCT
GGCCTGTGAGCCTGTCCAGGTGGAGATGTGCCTCGGTCTGAGCTACAACACCACAGCCTTC
CCTAACATCTGGGTGGGCATGATCACCCAGGAGGAGGTGGTAGAGGTCCTCAGCGGTTAC
AAGGTCTTCTGGGTGGAGGGAAGGAGGGCGGGCCTGGGAGGGGAGGAGCCTGGGAGCAG
GTGCCTGAAGGCCACTCTCCCTCCGCAGAGCCTGACAAGCCTGCCCTGCTACCAGCATTTC
CGGAGGCTCCTGTGTGGGCTGCTTGTGCCCCGTTGCACCCCACTAGGCAGTGTTCTGCCCC
CTTGCCGCTCTGTCTGCCAGGAAGCGGAGCACCAGTGCCAGTCTGGCCTGGCACTACTGGG
CACCCCCTGGCCCTTCAACTGCAACAGGCTGCCAGAGGCAGCTGACCTGGAAGCTTGTGCC
CAGCCCTGACCCTGAAGCCGGCCCCTGCCCTCTTCCTGCCCGTC**CTCTTTTGCCGGTCAG
GGCT**

## ***PAX6 - 5'*** **Upstream**

**ACTGGGGAAGGAATGGACTT**GAAACAAGGACCTTTG[TATACAGAAGGCACGATATCAG
TTGGAA
CAAATCTTCATTTTGGTATCCAAACTTTTATTCATTTTGGTGTATTATTTGTAAATGGGCATT
TGTATGTTATAATGAAAAAAAGAACAATGTAGACTGGATGGATGTTTGATCTGTGTTGGTC
ATGAAGTTGTTTTTTTTTTTTTTAAAAAGAAAACCATGATCAACAAGCTTTGCCACGAATTT
AAGAGTTTTATCAAGATAT***ATCGAATACTTCTACCCATCTGTT***CATAGTTTATGGACTGATG
TTCCAAGTTTGTATCATTCCTTTGCATATAATTAA**ACCTGGAACAACATGCACTAGA**TTTA
TGTCAGAAATATCTGTTGGTTTTCCAAAGGTTGTTAACAGATGAAGTTTATGTGCAAAAAA
GGGTAAGATATAAATTCAAGGAAGAAAAAAAGTTGATAGCTAAAAGGTAGAGTGTGTCTT
CGATATAATCCAATTTGTTTTATGTCAAAATGTAAGTATT**TGTCTTCCCTAGAAATCCTCA
GA**ATGATTTCTATAATAAAGTTAATTTCATTTATATTTGACAAGAATATAGATGTTTTATAC
ACATTTTCATGCAATCATACGTTTCTTTTTTGGCCAGCAAAAGTTAATTGTTCTTAGATATA
GTTGTA***TTACTGTTCACGGTCCAATCA***TTTTGTGCATCTAGAGTTCATTCCTAATCAATTAAA
AGTGCTTGCAAGAGTTTTAAACTTAAGTGTTTTGAAGTTGTTCACAACTACATATCAAAATT
AACCATTGTTGATTGTAAAAAACCATGCCAAAGCCTTTGTATTTCCTTTATTATACAGTTTT
CTTTTTAACCTTATAGTGTGGTGTTACAAATTTTATTTCCATGTTAGATCAACATTCTAAAC
CAATGGTTACTTTCA**CACACACTCTGTTTTACATCCTGA**

## List of Publications and Presentations

### Publications

1. **Ferdinamarie Sharmila**, Abinayapriya, Karthikeyan Ramprabhu, Govindasamy Kumaramanickavel, RR Sudhir, Sarangapani Sripriya. Genetic analysis of axial length genes in high grade myopia from Indian population. Meta gene 2 (2014) 164-175.

2. Vedam Lakshmi Ramprasad, Ronnie George, Nagasamy Soumittra, **Ferdinamarie Sharmila**, Lingam Vijaya, Govindasamy Kumaramanickavel. Association of non-synonymous single nucleotide polymorphisms in the LOXL1 gene with pseudoexfoliation syndrome in India. Molecular Vision 2008; 14:318-322.

3. **Ferdinamarie Sharmila**, Rashima Asokan, Vishwananthan N, Ronnie George, Lingam Vijaya, Sarangapani Sripriya. Genetic Association of SNPs near *ATOH7*, *CARD10*, CDKN2B, *CDC7* and *SIX1/SIX6* with the Endophenotypes of Primary Open Angle Glaucoma in Indian Population. PLOS ONE (under review).

4. Ramakrishnan Gayatri, Karunakaran Coral, **Ferdinamarie Sharmila**, Sarangapani Sripriya, Vetrivel Umashankar, Sivashanmugam Muthukumaran, Sripriya AV, Pandy Manish, Shantha B, Ronnie George, Lingam Vijaya, Narayanasamy Angayarkanni. Regulation of Lysyl oxidase activity in aqueous humor of Pseudoexfoliation patients-role of TGF beta, Homocysteine and *LOXL1* SNP. Current Eye Research (under review).

5. **Ferdinamarie Sharmila**, Rashima Asokan, Vishwananthan N, Ronnie George, Lingam Vijaya, Sarangapani Sripriya. Genetic Association of SNPs in exon 1 of *LOXL1* with the Endophenotypes of Primary Open Angle Glaucoma in Indian Population (submitted).

6. **Ferdinamarie Sharmila**, Rashima Asokan, Vishwananthan N, Ronnie George, Lingam Vijaya, Sarangapani Sripriya. Effect modification of SNPs near *ATOH7*, *CARD10*, CDKN2B, *CDC7* and *SIX1/SIX6* with environmental factors in POAG cases and controls (submitted)

**Abstracts**

B. J. Fan, **P. Ferdina Marie Sharmila**, N. Soumittra, S. Sripriya, J. Madhavan, D. S. Friedman3, L. Vijaya, J. L. Haines, R. J. George, J. L. Wiggs. Genome-wide linkage analysis using large consanguineous pedigrees from South India identifies new loci for ocular quantitative traits. **ARVO 2014**

B. J. Fan, **P. Ferdina Marie Sharmila**, N. Soumittra, S. Sripriya, J. Madhavan, D. S. Friedman3, L. Vijaya, J. L. Haines, R. J. George, J. L. Wiggs. Linkage and association analyses identify a VCDR locus on chromosome 14q32 in consanguineous pedigrees from South India. **ARVO 2014**

B. J. Fan, **P. Ferdina Marie Sharmila**, N. Soumittra, S. Sripriya, J. Madhavan, D. S. Friedman3, L. Vijaya, J. L. Haines, R. J. George, J. L. Wiggs. Family based genome-wide association study in South Indian consanguineous pedigrees identifies an association between *WNT7B* and central corneal thickness. **ARVO 2014**

B. J. Fan, **P. Ferdina Marie Sharmila**, N. Soumittra, S. Sripriya, J. Madhavan, D. S. Friedman3, L. Vijaya, J. L. Haines, R. J. George, J. L. Wiggs. Family-based genome-wide association study in South Indian consanguineous pedigrees confirms association between ZNF469 and central corneal thickness. **ARVO 2014**

**Ferdinamarie Sharmila**, Sathypriya, Abinayapriya, R.R.Sudhir, and Sripriya Sarangapani. Genetic association study of MFRP and VSX2 genes with myopia and correlation of MFRP CGTCGG haplotype with decreased AXL in Indian population. **ISHG, Punjab 2012.**

**P. Ferdina Marie Sharmila**, B. J. Fan, N. Soumittra, S. Sripriya, J. Madhavan, D. S. Friedman3, L. Vijaya, J. L. Haines, R. J. George, J. L. Wiggs. Power of consanguineous pedigrees in mapping ocular quantitative traits. Ranbaxy Young scientist award 2012, Gurgoan, Delhi

**Ferdina Marie Sharmila**, S. Sacikala, S. Aparna, N. Soumittra, S. Sripriya, J. Madhavan, G. Ronnie, L. Vijaya, D.S. Friedman, J.L. Wiggs. Association Analysis of ATOH7 and Quantitative Traits in Indian Families. **Asia ARVO, Singapore 2011.**

**Ferdinamarie Sharmila**, Ramprabhu, Anuradha Elamparudhi, Abinaya Priya, RR Sudhir, Madhavan Jagadeesan, Sarangapani Sripriya. Genotyping rs1800470 and rs1800471 in Myopia Patients from South India for disease association. **IERG, Hyderabad 2010.**

**Ferdinamarie Sharmila**, Salomi Sathya Sangeetha, Sacikala Srinivasan , Kavitha Kalaivani Natarajan, Madhavan Jagadeesan, Sarangapani Sripriya. Candidate gene analysis in Small Eye Syndrome. **ISHG, Lucknow 2010.**

## Brief Biography of the Candidate

Mrs. P. Ferdinamarie Sharmila completed her under graduation in B.Sc (MLT) (JIPMER) from Pondicherry University in the year 2005. She joined Sankara Nethralaya for MS (MLT) BITS program and graduated in the year 2008. Following this she joined as research fellow in SNONGC Dept of Genetics and Molecular biology under the INDO_US project for a period of two years and ICMR funded project for a span of 3 years. During this fellowship she has attended several conferences, symposiums including IERG, ISHG. She was selected for "Hands on training in molecular biological techniques - Microarray workshop" conducted in Cancer institute, Adayar. She also received travel fellowship grant award at the IERG conference held in CCMB, Hyderabad and for "Ranbaxy young scientist" held in Ranbaxy centre- Gurgaon. She has 2 papers published, 2 under review and 2 submitted. She has good knowledge and hands on experience in various techniques including DNA sequencing, linkage and association analysis, whole genome scans using high density Microarray chips, and analysis of data in command based softwares. Her area of research include genomics especially QTL mapping in ocular disorders, identification of mutations, and genetic counselling.

## Brief Biography of the Supervisor

Dr. Saragapani Sripriya is, Lecturer at SN ONGC Department of Genetics & Molecular Biology, Vision Research Foundation, Sankara Nethralaya, Chennai, India. She obtained her Doctorate degree from Birla Institute of Technology and Science, Pilani, in 2006 on Title: "Genetic association studies, genome wide linkage analysis and mutation screening in primary open and closed angle glaucomas". She did MS (Medical Laboratory Technology) Birla Institute of Technology and Science, Pilani, in 1999 with course work at Medical Research Foundation, Sankara Nethralaya, Chennai. She has received the following awards: DST Fast Track fellowship (2012), ICMR fellowship for exchange visit under Indo French collaborative project (2010), Young Scientist at 31st Indian Society of Human Genetics, New Delhi, 2006, Best Poster at 29th Indian Society of Human Genetics, Bangalore, 2004, Chennai Willingdon Corporate Foundation PhD Fellowship, 2001, Young Investigator, 1st Singapore Eye Research Institute International Meeting, Singapore, 2001, Travel fellowship, IX Indian Eye Research Group Meeting, Hyderabad, 1999, Bangalore Genie Best Outgoing Student Award for Clinical genetics, 1999. She works on molecular genetics of inherited ophthalmic diseases and involved in genetic diagnostics of the same. She received funding from ICMR, DST etc for various projects. She has 18 publications in peer reviewed journals including Nature Genetics, BMJ Open, Ophthalmic genetics, Molecular Vision, Clinical Genetics, Molecular diagnostics and therapy, Journal of current glaucoma practice, Ophthalmic Epidemiology, Journal Diabetes and Complications, Diabetes Research and Clinical Practice, In press 3: Clinical Genetics, Meta gene, Current eye research. She has reviewed articles on ocular genetic disorders for journals molecular vision, British Journal of Ophthalmology. She has trained 15 masters' students who worked on specific genetics dissertation topic. She guides 3 PhD students registered under BITS, Pilani and SASTRA, Tanjore, Tamil Nadu. She is also involved in genetic counseling for patients with ocular genetic diseases. Areas of Interest: Homozygosity mapping candidate gene analysis and genetic association studies for oligogenic disorders Bardet Biedl syndrome, ocular developmental anomalies, type 2 diabetic retinopathy, open angle and angle closure glaucoma.