THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

This thesis in electronic version is provided to the Library by the author. In the case where its contents is different from the printed version, the printed version shall prevail.

The Hong Kong Polytechnic University

School of Optometry

Mapping of Myopia Susceptibility Genes Using Population-based Association Studies (Case-control Studies)

LO Ka Kin

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

December 2009

CERTIFICATE OF ORIGINALITY

I, Lo Ka Kin, declare that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Except where reference is made in the text of this dissertation, this dissertation contains no material published elsewhere or extracted in whole or in part from a dissertation presented by me for another degree or diploma.

This dissertation has not been submitted for the award of any other degree or diploma in any other tertiary institution.

____

Lo Ka Kin

Dated: 31/09/2010

ABSTRACT

Myopia is the most common eye disorder in the world. The prevalence of myopia is up to 30% in western countries, but may be as high as 80% in some Chinese populations. A refractive error in excess of –6.00 diopters (D) is defined as high myopia, also called pathologic myopia because of its associated potential eye complications, which can lead to blindness. The prevalence of high myopia can reach as high as 24% in Chinese populations while it is only up to 5% in western countries.

The present case-control study started with recruitment of Chinese subjects who were highly myopic (≤ -6.0 D or worse, n=300) or emmetropic (within ±0.75 D, n=300). The first genetic association study was a replication study for the myocilin (*MYOC*) gene and the second study was a separate study exploring candidate genes in the *MYP2* region by a DNA pooling approach, followed by confirmation using individual genotyping.

Several studies have reported the association of several *MYOC* polymorphisms with high myopia, including one family-based study from our group. However, other studies reported negative findings. We attempted to replicate previous studies in a case-control study involving 300 cases and 300 controls. *MYOC* polymorphisms previously found associated with high myopia were genotyped together with other SNPs in strong linkage disequilibrium with the positive markers: two microsatellites and eight single nucleotide polymorphisms (SNPs) in total. Five correlated SNPs at the 3' end of the gene showed significant differences between high myopes and controls under three genetic models tested (genotypic, additive and allelic):

rs12076134, rs1602244, rs6425356, rs10737323 and rs743994. The results remained significant after correction for multiple comparisons by false discovery rate at 0.05 levels. The two most significant associations with rs64252356 and rs743994 were further confirmed in our original families. These SNPs have not been investigated by any other groups up to date. Three polymorphisms previously associated with high myopia failed to be replicated, suggesting that the original positive results were probably chance findings. One previously positive SNP failed to be replicated because of local variation in the linkage disequilibrium patterns in the case-control subjects.

Previous studies using linkage analysis of families with highly myopic members identified a myopia locus at chromosome 18p11 – the *MYP2* locus. We selected seven candidate genes (*CLUL1*, *EMLIN-2*, *LPIN2*, *MYOM1*, *MYL12A*, *MYL12B* and *ZFP161*) from the *MYP2* region and examined 62 tag SNPs with a two-stage DNA pooling approach. In the first stage, 6 case pools and 6 control pools were constructed using DNA samples from 300 high myopes and 300 controls. Each DNA pool was prepared by mixing equal amounts of DNA from 50 distinct subjects of the same affection status. Allele frequencies of SNPs were estimated by analysis of primer-extended products in a denaturing high performance liquid chromatography system, and compared across 3 replicates of each pool and across two sets of pools by means of nested analysis of variance. In the second stage, nine promising SNPs ($P \leq 0.10$) were further evaluated by individual genotyping of samples included in the pools. One SNP (rs589318) within the *LPIN2* gene was found to be associated with high myopia under three genetic models (genotypic, additive and allelic), and the significance survived correction by false discovery rate at 0.05 level. This is a novel

finding not previously reported.

In conclusion, the results of the present study implicated the involvement of 3’ polymorphisms of the *MYOC* gene in the predisposition to high myopia. Moreover, we successfully applied a DNA pooling strategy to screen candidate genes in the *MYP2* region and identified the *LPIN2* gene to be associated with high myopia. These are novel findings and should be replicated using independent sample sets.

ACKNOWLEDGEMENTS

This work was carried out in the School of Optometry (subject recruitment) and the Department of Health Technology and Informatics (laboratory work) at The Hong Kong Polytechnic University, Hong Kong SAR, China, during the years 2004-2009.

I wish to express my deepest gratitude to my supervisor, Professor Shea-ping Yip, for his continuing encouragement and optimism throughout these years. It has been a privilege to have the opportunity to work in such a great research group under his guidance. I am most grateful to my other supervisor Professor Maurice Yap for introducing me to the world of science and for his patience and encouragement. During these years, they have always been available when I needed help with my research. Thank them for teaching me the basics of lab work at the start of this research project. I wish to express my respect for Dr. Leung KH for providing excellent research facilities and an inspiring scientific environment at the department. Prof. To CH is acknowledged for their scientific expertise and valuable comments on the thesis. Thanks also go to Professor Pak Sham, Hong Kong University, for his advice on the use of nested analysis of variance for analysing pooled DNA data.

Colleagues at the Department of Health Technology and Informatics, The Hong Kong Polytechnic University, deserve my warmest thanks. Pik Shan Choi, Sau Yin Lee and Wai Yan Fung are acknowledged for all their patient help with the practical matters and their excellent technical assistance. I warmly thank my dear friends and colleagues Percy Po Wah Ng for his friendship over the years during and after optometry school.

I wish to express my warmest gratitude to my parents, for always believing in me no matter what I was planning to do. The loving support from my whole family, my brother and his dear spouses has carried me over the difficulties.

This work was supported by grants from Research Grant Count (Ref. No. PolyU 5411/06M; account code: B-Q04A) and from The Hong Kong Polytechnic University (RGM7, 87MS and J-BB7P).

PUBLICATIONS ARISING FROM THE THESIS

Lo KK, Yip SP, Ng PW, Fung WY, Maurice KH Yap. Myocilin is associated with high myopia: a replication study. The 12th International Myopia Conference, Australia, July 2008

Lo KK, Yip SP, Ng PW, Fung WY, Maurice KH Yap. A DNA pooling approach to indentifying myopia susceptibility gene in the *MYP2* locus. The 9th Asia-Pacific Conference on Human Genetics, Hong Kong, Nov – Dec 2010

Lo KK, Yip SP, Ng PW, Fung WY, Maurice KH Yap (2010). Polymorphisms at the 3' flanking region of the myocilin (*MYOC*) gene are associated with high myopia. *Molecular Vision* (submitted).

TABLE OF CONTENTS

	Page
CERTIFICATE OF ORIGINALITY	**ii**
ABSTRACT	**iii**
ACKNOWLEDGEMENTS	**vi**
PUBLICATIONS ARISING FROM THE THESIS	**viii**
TABLE OF CONTENTS	**ix**
LIST OF FIGURES	**xv**
LIST OF TABLES	**xvii**
LIST OF ABBREVIATIONS	**xix**
Chapter 1 Introduction	**1**
1.1 Overview	1
1.1.1 Global public health concern of myopia	1
1.1.1.1 Description and classification of myopia	2
1.1.1.2 Prevalence of myopia	8
1.1.1.3 Public health concern	10
1.1.2 Epidemiology	11
1.1.2.1 Ethnicity and race	11
1.1.2.2 Education, intelligence and IQ	12
1.1.3 Aetiology and pathogenesis	13
1.1.3.1 Theories	14
1.1.3.2 Ocular Morbidity	17
1.2 The genetics of myopia	19
1.2.1 Familial aggregation studies & win studies	20
1.2.1.1. Family Studies	20
1.2.1.2. Modes of inheritance	21
1.2.1.3. Twin Studies	22
1.2.2 Genetics of ocular refractive components	23
1.2.3 Animal models - Gene expression and functional Studies	25
1.2.3.1 Active emmetropization mechanism	25

1.2.3.2 Knockout animal models 26
1.2.3.3 Eye growth 27
1.2.4 Molecular level - Myopia loci 27
1.2.4.1 The loci of high myopia 28
1.2.4.2 The loci of common myopia 33
1.2.5 Molecular studies on myopia – Candidate gene studies 39
1.2.5.1 Positional candidate gene studies for non-syndromic myopia loci 40
1.2.5.2. Functional Candidate Gene Studies 43
1.2.5.3 Scleral gene studies 44
1.3 Genetic approach to studying myopia 46
1.3.1 Introduction of SNP genotyping methods 47
1.3.1.1 Single nucleotide polymorphisms (SNPs) : the unit of variation 47
1.3.1.2 Types of SNPs 48
1.3.1.3 Databases 49
1.3.1.4 Use and importance of SNPs 50
1.3.1.5 SNPs and common diseases 50
1.3.1.6 SNPs–haplotype blocks and linkage disequilibrium 51
1.3.1.7 Genotyping of SNPs 53
1.3.2 Genetic mapping of common complex disease genes 78
1.3.2.1 The rationale of the population-based genetic association studies 78
1.3.2.1.1 Introduction to Genetic Association 79
1.3.2.2 Genetic dissection of common complex traits 80
1.3.2.2.1 Linkage studies 81
1.3.2.2.2 Association studies 82
1.3.2.3 Challenges of candidate gene approach 97
1.3.2.3.1 Replication 97
1.3.2.3.2 Novel gene exploration 98
1.3.2.3.3 Population stratification 98
1.3.2.4 Statistical programmes of association analysis 98
1.3.2.5 Power of studies 103
1.3.3 Pooling Approach 106
1.3.3.1 Quantitative genotyping and platforms 106
1.3.3.2 Primer extension (PE) reaction coupled with DHPLC Analysis for DNA Pools 108
1.3.3.3 Quantitative genotyping - Allele frequency estimation in DNA pools 111
1.3.3.4 Comparison between real allele frequencies and allele frequencies estimated in the pools 113
1.3.3.5 Marker selection for pooling studies 113

1.3.3.6 Statistical analysis of pooling data 114
1.3.3.6.1 Statistical analysis of estimated allele frequencies in DNA pools 114
1.3.3.6.2 Statistical analysis of allele frequencies in individual genotyping 114
1.3.3.6.3 Sources of error in DNA pooling approach 115
1.4 Study Objectives 116

Chapter 2 Materials & methods 118

2.1 Research design 118
2.2 Materials 119
2.2.1 Materials for laboratory experiments 119
2.2.1.1 Chemicals and reagents 119
2.2.1.2 Solutions 121
2.2.1.2.1. DNA extraction 121
2.2.1.2.2 Electrophoresis 122
2.2.1.2.3 DHPLC buffer 123
2.2.1.3 Equipment 123
2.3 Methods 123
2.3.1 Sample collection and diagnosis 123
2.3.1.1 Subject recruitment 123
2.3.1.2 Phenotypic data 125
2.3.2 Experimental background and protocols 126
2.3.2.1 DNA extraction 126
2.3.2.2 DNA quantification 127
2.3.3 Sample purity 128
2.3.4 Polymerase chain reaction 129
2.3.4.1 Primer design 129
2.3.4.2 Thermal cycles 130
2.3.5 Variations of the basic PCR technique 131
2.3.5.1 Allele-specific PCR 131
2.3.5.2 Nested PCR 131
2.3.5.3 Touchdown PCR 132
2.3.6 Gel electrophoresis 132
2.3.7 Cycle sequencing 133
2.4 Marker Selection for the *MYOC* gene 134
2.4.1 Genotyping of microsatellite (MS) markers of the MYOC locus by Genescan 136

2.4.2 Genotyping of SNPs of the MYOC locus 141
2.4.2.1 Allele-specific (AS) PCR 141
2.4.2.2 Restriction Fragment Length Polymorphism (RFLP) 145
2.4.2.3 Primer Extension with DHPLC 152
2.5 Statistical analysis of MYOC data 155
2.5.1 Case-control data 155
2.5.1.1 Hardy-Weinberg equilibrium (HWE) 155
2.5.1.2 Tests of association: single markers 155
2.5.1.3 Tests of main effects by stepwise logistic regression 156
2.5.1.4 Tests of association: multiple SNPs by building *MYOC* haplotype blocks 157
2.5.1.5 Linkage disequilibrium (LD) patterns 158
2.5.2 Tests of association: Family samples with two positive SNPs of the *MYOC* locus 159
2.6 Testing *MYP2* candidate genes by a DNA pooling approach 160
2.6.1 Pool construction 161
2.6.2 Tag SNP of Candidate genes 162
2.6.3 Touchdown PCR amplification of DNA samples 171
2.6.4 Characterisation of the screened SNPs 171
2.6.5 SNP confirmation and validation by screening of heterozygotes 172
2.6.6 Genotyping by PE-DHPLC analysis 173
2.6.6.1 Post-PCR purification 173
2.6.6.2 Primer extension (PE) reactions 173
2.6.6.3 DHPLC analysis 175
2.7 Statistical analysis of estimated allele frequencies of DNA pools 176
2.7.1 Nested analysis of variance 177
2.8 Individual genotyping for pooling screening 180
2.9 Statistical analysis for individual genotyping 185
2.10 Computer programs 185

Chapter 3 Replication study for the *MYOC* gene 187

3.1 Overview 187
3.2 Background of myocilin (MYOC) 188
3.3 Results 189
3.3.1 Analysis of ocular data in unrelated subjects 189
3.3.2 Tests of genetic association: single locus 192
3.3.2.1 Tests of association under three genetic models 192

3.3.2.2 Detection of main effects by stepwise logistic regression 199
3.3.3 Tests of association: multiple SNPs 202
3.3.3.1 Linkage disequilibrium patterns 202
3.3.3.2 Haplotype analysis 205
3.3.4 Replication by family-based association study 207
3.3.4.1 Analysis of ocular data in myopic offspring 207
3.3.4.2 Analysis of genotype data in nuclear families 207
3.3.4.3 Linkage disequilibrium patterns in nuclear families 208
3.4 Discussion 212
3.4.1 Existing *MYOC* studies 212
3.4.2 MYOC polymorphisms associated with high myopia 215
3.4.3 Summary 219
3.4.4 Future studies 220
3.4.5 Conclusion 221

Chapter 4 A DNA pooling approach to identify new myopia susceptibility genes in the *MYP2* locus **223**

4.1 Overview 223
4.2 Background of Myopia-2 (*MYP2*) 224
4.3 Results 226
4.3.1 Elution times of extension products 226
4.3.2 k correction factors 227
4.3.3 Estimating allele frequencies in case and control pools 234
4.3.4 Tests of genetic association for individual genotyping 234
4.3.5 LD patterns of 9 initial positive SNP markers of *MYP2* 240
4.3.6 Comparison of allele frequency differences estimated from pools with the frequency differences determined from individual genotypes 240
4.4 Discussion 243
4.4.1 Myopia-2 locus (*MYP2*) 243
4.4.2 LPIN2 244
4.4.3 MYP2 candidate genes examined in this study 249
4.4.4 Other candidate genes in the *MYP2* locus 253
4.4.5 DNA pooling strategy 254
4.4.6 The validity of the primer extension reaction coupled with DHPLC analysis in DNA pooling study 256
4.4.7 Potential sources of bias or imprecision in pooled DNA methods 257
4.4.7.1 The k correction factor 257
4.4.7.2 Pool-specific measurement error 259

4.4.7.3 Standard sampling error 260
4.4.7.4 Haplotype frequency estimates 260
4.5 Conclusion 261

Chapter 5 Overall discussion 262

5.1 Mapping susceptibility genes for myopia 263
5.1.1 Phenotype definition and related phenotypes 263
5.1.2 Disease gene hunting approaches 267
5.1.2.1 Linkage analysis 267
5.1.2.2 Association studies 268
5.1.2.3 Study design: Candidate Gene vs Genome-Wide association study 270
5.2 Detection of polymorphisms 274
5.2.1 SNP genotyping methods 274
5.2.2 SNP screening with primer extension DHPLC platform in DNA pooling studies 279
5.3 Future studies 283
5.3.1 Future studies for the *MYOC* gene 283
5.3.2 Future studies for the *MYP2* locus 287
5.4 Other considerations for future studies 289
5.4.1 DNA pooling approach as the first stage of GWAS 289
5.4.2 Copy number variation (CNV) 290
5.4.3 Gene-environment interactions 291
5.4.4 Gene-gene interactions 294
5.5 Limitations 295
5.5.1 Limitations of DNA pooling methods 295

Chapter 6 Summary and conclusion 297

References 298

APPENDICES 336

Information sheets Informed consent 337
Informed consent 339

LIST OF FIGURES

Figure no.	Title	Page
1.1	The principle of PCR-based SSCP analysis.	55
1.2	The principle of dHPLC with heteroduplex formation	56
1.3	Identifying heterozygosity by dHPLC	56
1.4	The principle of allele specific oligonucleotide hybridization	58
1.5	Principle of OLA-based SNP genotyping.	60
1.6	Microarray-based platform for genetic screening	62
1.7	Restriction fragment length polymorphism-based platform for genetic screening.	64
1.8	Allele-Specific Primer Extensions	66
1.9	The TaqMan Allelic Discrimination assay	68
1.10	The principle of the Invader assay	70
1.11	The genomic localization of selected candidate genes within the MYP2 interval	88
1.12	A schematic diagram of the Cluster-in like 1 (CLUL1) gene	92
1.13	A schematic diagram of the EMILIN2 gene	92
1.14	A schematic diagram of the LPIN2 gene	95
1.15	Structure of the MYOM1 gene.	95
1.16	Structure of the Myosin regulatory light chain (MLCB) gene	96
1.17	Structure of the ZFP161 gene.	96
1.18	The principle of the DHPLC	110
1.19	Correction for SNP Markers in estimating the allele frequency	112
2.1	Electrophoretograms for microsatellite markers *NGA19* and *NGA17*.	138
2.2	Tm-shift genotyping of SNPs	143
2.3	SNP genotyping by comparative thresholds (d*C*t) method	144
2.4	Genotyping of SNP rs2421853 by restriction fragment length polymorphism (RFLP) analysis with the use of the restriction endonuclease BanII.	147
2.5	Genotyping of SNP rs6425356 by restriction fragment length	149

	polymorphism (RFLP) analysis with the use of the restriction endonucleae HphI	
2.6	Genotyping of SNP rs743994 by restriction fragment length polymorphism (RFLP) analysis with the use of the restriction endonuclease HinfI.	150
2.7	RFLP with three SNPs for MYOC gene	151
2.8	Genotyping of SNP rs6425356 and rs1602244 by primer extension coupled with denaturing high performance liquid chromatography (PE-DHPLC).	154
2.9	DHPLC elution profiles for a heterozygous individual	174
2.10	The nested design of the DNA pooling study	179
2.11	RFLP with nine SNPs for individual genotyping	182
3.1	Linkage disequilibrium (LD) patterns of *MYOC* single nucleotide polymorphisms in the Han Chinese subjects of the current case-control study and the HapMap database	203
3.2	Linkage disequilibrium (LD) patterns of *MYOC* single nucleotide polymorphisms in the parents of the current family study	211
3.3	The *MYOC* gene is bound by two hotspots of recombination.	222
4.1	Derivation of the k correction factors from heterozygous individuals	231
4.2	Linkage disequilibrium (LD) patterns of 9 initial positive SNP markers of *MYP2*.	241

LIST OF TABLES

Table no.	Caption	Page
1.1	Classification Systems for myopia.	4
1.2	Classification of myopia based on aetiology	6
1.3	Summary of Myopia Loci	37
1.4	Positional candidate genes within myopia loci	41
2.1	*MYOC* polymorphisms genotyped in the present study	135
2.2	MYOC polymorphisms: genotyping methods and PCR information	139
2.3	Construction of DNA pools	163
2.4	Characteristics of 62 tag SNPs selected from 7candidate genes in the *MYP2* region	164
2.5	Primers (PCR and PE) and PCR product size for tag SNPs from MYP2 locus	166
2.6	PCR primers, reaction component, amplicon sizes and RFLP digestion enzymes for nine top-ranking SNPs in the *MYP2* region	181
2.7	Software used in study and their applications and sources.	186
3.1	Clinical Characteristics of the highly myopic patients and the control subjects	191
3.2	Comparison of *MYOC* genotype frequencies between males and females within the control and the case groups	194
3.3	MYOC microsatellite markers: frequency distribution in cases and controls, and association analysis	195
3.4	*MYOC* single nucleotide polymorphisms: frequency distribution in cases and controls, and association analysis	196
3.5	Correction for multiple comparisons by false discovery rate (FDR) at a level of 0.05 for association analysis of *MYOC* markers	198
3.6	Detection of main effects by forward stepwise logistic regression analysis of positive *MYOC* single nucleotide polymorphisms	200
3.7	Detection of main effects by backward stepwise logistic regression analysis of positive *MYOC* single nucleotide polymorphisms	201
3.8	Linkage disequilibrium (LD) measures (r^2) between *MYOC* microsatellite and SNP markers	204
3.9	*MYOC* haplotype analysis across rs6425356, rs10737323 and rs743994 by Plink	206

3.10 Summary of ocular data of 233 myopic siblings from nuclear families 209

3.11 *MYOC* single nucleotide polymorphisms: summary of genetic data for parents and association analysis by family-based association testing under additive model in 162 nuclear families 210

3.12 Comparison of studies investigating the relationship between high myopia and *MYOC* polymorphisms 214

4.1 DHPLC analysis of SNPs in the *MYP2* locus: elution times of extension products 229

4.2 Pooled DNA analysis of SNPs in the *MYP2* locus: k correction factor and estimated allele frequencies and their differences 232

4.3 Comparison of *MYP2* genotype frequencies between males and females within the control and the case groups 235

4.4 *MYP2* SNPs determined by individual genotyping: frequency distribution in cases and controls, and association analysis 237

4.5 Correction for multiple comparisons by false discovery rate (FDR) at a level of 0.05 for association analysis of *MYP2* markers 239

4.6 Comparison of allele frequency differences estimated from pools with the frequency differences determined from individual genotypes. 242

4.7 SNPs tagged by rs589318 in Han Chinese (HapMap data) 246

5.1 Comparison of case-control and families studies 266

LIST OF ABBREVIATIONS

°C	Degrees Celsius
A	Adenine
ACD	Anterior chamber depth
ACN	Acetonitrile
AD	Autosomal dominant
ASO	Allele-specific oligonucleotide hybridization
ASPCR	Allele-specific polymerase chain reaction
bp	Base pair
C	Cytosine
CC	Corneal curvature
CC	Case and control
CDCV	Common disease common variant
cDNA	complementary DNA
CI	Confidence interval
cM	CentiMorgan
CNP	Copy-number polymorphism
CNV	Copy number variation
CP	Crossing point
Ct	Cycle threshold
dbSNP	SNP database from National Center for Biotechnology Information
ddNTP	Dideoxynucleotide
df	Degrees of freedom
DHPLC	Denaturing high-performance liquid chromatography
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
dsDNA	Double stranded DNA
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
ExoI	Exonuclease I

FDR	False discovery rate
FEN	Flap Endonuclease
FLU	Fluorescein
FRET	Fluorescence energy resonance transfer
G	Guanine
GEM	Genes in Myopia
HGF	Hepatocyte growth factor gene
HLA	Haploid genotype
HWE	Hardy-Weinberg equilibrium
IOP	Intraocular pressure
IQ	Intelligence quotient
L	Litre
LASIK	Laser-Assisted in Situ Keratomileusis
LD	Lens thickness
LNA	Locked nucleic acid
LOD	likelihood of the odds
MDR	Multifactor Dimensionality Reduction
Mg^{2+}	Magnesium
MIM	Mendelian Inheritance in Man
MS	Mass Spectrometry
MS	Microsatellite
MYOC	Myocilin
n	Sample size
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	Nanograms
nm	Nanometres
OD	Optical density
OLA	Oligonucleotide ligation assay

OMIM	Mendelian Inheritance in Man
OR	Odds ratio
p	Shorter segment of the chromosome
PAGE	Polyacrylamide gel
PCR	Polymerase chain reaction
PE	Primer Extension
POAG	Primary open angle glaucoma
POP-4	Performance optimized polymer 4
q	Longer segment of the chromosome
QTL	Quantitative trait locus
RD	Retinal detachment
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RPE	Retinal pigment epithelium
RPM	Revolutions per minute
RR	Relative risk
RT-PCR	Reverse transcription polymerase chain reaction
s	Seconds
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SER	Spherical equivalent refraction
SLRP	Small leucine-rich repeat proteoglycan
SNPs	Single-nucleotide polymorphisms
SSCP	Single-strand conformation polymorphism
T	Thymine
Taq	Taq polymerase
TBE	Tris/Borate/EDTA
TEAA	Triethylamine acetate buffer
TIGR	Trabecular Meshwork Inducible Glucocorticoid Response gene
TL	Lens thickness
Tm	Melting temperature
Tris-HCl	Tris-hydrochloric acid

TSR	Template suppression reagent
u	Unit(s)
UK	United Kingdom of Great Britain
UV	Ultraviolet light
VCD	Vitreous chamber depth
VCDR	Vertical cup-disc ratio
α	Significant level
ε	Effect size
λs	Sibling recurrence risk ratio
μ	Micro
$\chi 2$	Chi square test
Ψ	Statistical power

Chapter 1
Introduction

1.1 Overview

Myopia is a significant public health problem in the world. Globally, more than 2 billion people have myopia, and the overall prevalence of myopia is projected to be one third of the world's population by the year 2020 (Pechmann & Czepita, 2000). Sustainable costs are spent on the correction of the blurred vision caused through spectacles, contact lenses or refractive surgery. Recent advances in molecular biology research have led to a better understanding of the genetics and pathophysiology of myopia. With the updated knowledge, new therapies could hopefully be designed to prevent and help treat it. This chapter reviews the clinical and genetic features of myopia.

1.1.1 Global public health concern of myopia

1.1.1.1 Description and classification of myopia

1.1.1.1.1 Description

Myopia is the scientific term for nearsightedness. It is a common refractive defect in which parallel rays of light entering the eye at rest are brought to focus in front of the retina (Curtin, 1988; Zadnik, et al., 1994). The major sign and symptom of myopia can generally be corrected by the use of negative power optical aids, such as glasses or contact lens, and also by refractive surgery. Usually, myopia has its onset during the school years and progresses with age until almost 20 years old.

Concept of myopia

The first concept of myopia from the ancient Greek was the loss of distance vision (Bambridge, 2002). In the mid of the 18th century, a more reasonable observation on the cause of myopia was related to the poor visual hygiene and prolonged tension of eyes during close work. This vision problem was still not clearly understood in the mid 19th century. The use of glasses was thought to promote the progression of myopia, and even eye care professionals believed that ciliary muscle was weakened and could no longer focus the lens. A more precise understanding of myopia was not revealed until the equipment of ultrasound scans was available. The machine measures the size of eyeballs *in vivo* and objectively shows that myopia is mainly due to the elongation of the eyeball (Francois & Goes, 1977).

1.1.1.1.2 Classification of Myopia

Various classification systems have been described for myopia (Grosvenor, 1987) and are described below.

- **By cause**

Duke-Elder classified myopia by cause (Duke-Elder, 1968). *Axial myopia* is attributed to an increase in the eye's axial length while *refractive myopia* is attributed to the condition of the refractive elements of the eye (Cline, et al., 1989). Refractive myopia can further be subclassified as follows (Borish & Indiana University. Division of Optometry, 1970). *Curvature myopia* is attributed to excessive, or increased, curvature of one or more of the refractive surfaces of the eye, especially the cornea (Cline et al., 1989). In those with Cohen syndrome, myopia appears to

result from high corneal and lenticular power (Summanen, et al., 2002). Finally, *index myopia* is attributed to variation in the index of refraction of one or more of the ocular media (Cline et al., 1989).

- **Clinical entity**

Myopia can be classified based on clinical concept: simple myopia, nocturnal myopia, pseudomyopia, degenerative myopia, and induced (acquired) myopia (Grosvenor, 1987) **(Tables 1.1 and 1.2)**.

Simple myopia is a physiological form of myopia because the only deviation from normal structure and function is the need for minus power lenses for normal distance visual acuity. Both genetic and environmental factors contribute to the development.

Nocturnal myopia, also known as *night myopia*, occurs only in dim illumination. It is a condition in which the eye has a greater difficulty of seeing in low illumination areas, even though its daytime vision is normal. Night myopia is believed to be caused by increased accommodative response associated with low levels of light (Hope & Rubin, 1984; Owens & Leibowitz, 1976). Because there is insufficient contrast for an adequate accommodative stimulus, the eye assumes the intermediate dark focus accommodative position rather than focusing for infinity. A stronger prescription for myopic night drivers is often needed. Younger people are more likely to be affected by night myopia than the elderly (Brabyn, et al., 2005; Chen, et al., 2003).

Table 1.1 Classification systems for myopia (Grosvenor, 1987)

Type of Classification	Classes of Myopia
Clinical	Simple myopia
	Nocturnal myopia
	Pseudomyopia
	Degenerative myopia
	Induced myopia
Degree	Low myopia
	Medium myopia (from -3.00D to -6.00D)
	High myopia (<-6.00 D)
Age of Onset	Congenital myopia (present at birth and persisting through infancy)
	Youth-onset myopia (<20 years of age)
	Early adult-onset myopia (20-40 years of age)
	Late adult-onset myopia (>40 years of age)

Pseudomyopia is generally encountered in younger subjects performing excessive close work. Sustained or excessive near-work demands result in hypertonicity of the ciliary body so that an emmetropic or slightly hyperopic subject clinically appears to be myopic or a myopic subject appears to be more so. In psychogenic accommodative spasm, psychological influences can produce spasm of the near reflex (Koyama, 1970). *Degenerative myopia*, also called pathological or progressive myopia, is characterised by structural defects in the posterior segment of the eye such as posterior staphyloma, and associated with a high refractive error and subnormal visual acuity after correction. Severe congenital myopia during infancy typically becomes degenerative myopia.

Induced myopia may be viewed as a secondary form of myopia that is pathological in nature, and depends upon the initiating condition or agent, which is often temporary. A refractive shift towards myopia in the elderly is usually associated with the development of nuclear sclerosis of the crystalline lens (Sperduto & Hiller, 1984).

Table 1.2 Classification of myopia based on aetiology (Grosvenor, 1987)

Type of Myopia	Aetiologies
Simple Myopia	Inheritance Significant amounts of near work
Nocturnal Myopia	Significant levels of dark focus of accommodation
Pseudomyopia	Accommodative disorder High exophoria Cholinergic agonist agents
Degenerative Myopia	Inheritance Retinopathy of prematurity Interruption of light passing through ocular media
Induced Myopia	Age-related nuclear cataracts Exposure to sulfonamides and other pharmaceutical agents Significant variability in blood sugar level

- **Degree**

Myopia, which is measured in dioptres by the strength or optical power of a corrective lens that focuses distant images on the retina, can also be classified by degree or severity (i.e., low, medium, or high) (Grosvenor, 1987).

Low myopia usually describes myopia of –3.00 dioptres or less (closer to 0.00) while ***medium myopia*** usually describes myopia between –3.00 and –6.00 dioptres (Cline et al., 1989). ***High myopia*** usually describes myopia of –6.00 or worse (towards -10.00) (Cline et al., 1989). People with high myopia are more likely to have retinal detachments (Gozum, et al., 1997) and primary open angle glaucoma (Podos, et al., 1966). They are also more likely to experience floaters, shadow-like shapes which appear singly or in clusters in the field of vision (Messmer, 1992).

- **Age of onset**

Congenital myopia is the least common form of myopia and is present at birth but with a high severity.

Youth-onset myopia has its onset in childhood between the age of five and mid-to late teens. ***School myopia*** appears during childhood, particularly the school-age years (Morgan & Rose, 2005). This form of myopia is attributed to the use of the eyes for close work during the school years (Cline et al., 1989). ***Early adult onset myopia*** occurs between ages 20 and 40 while ***late adult onset myopia*** occurs after age 40 (Grosvenor & Scott, 1991).

1.1.1.2 Prevalence of myopia

The global prevalence of refractive errors has been estimated to be in the range of 800 million and 2.3 billion, which is rapidly increasing throughout the world (Pechmann & Czepita, 2000). By 2020, it is estimated the number of people with myopia will grow to one third of the world's population. It is not unexpected that the prevalence of myopia within a sampled population often varies with age, country, sex, race, ethnicity, occupation, environment, and other factors (Fredrick, 2002). Variability in testing and data collection methods makes comparisons of prevalence and progression difficult (Matsumura & Hirai, 1999).

Myopia is the most common eye disorder especially in Asia. In Singapore, a series of studies have shown an increase in myopia in males aged 15-25 from 26% of this group in the late 1970s to 83% in the late 1990s (Wu, et al., 2001). In India, the prevalence of myopia in the general population has been reported to be only 6.9% (Mohan, et al., 1988). In Hong Kong, the prevalence of myopia among Hong Kong Chinese adults was up to 70 to 90% (Edwards & Lam, 2004).The early onset and fast progression of myopia among Chinese children is also well documented (Edwards, 1999; Fan, et al., 2004b; Lam, et al., 2004).

A recent study involving first-year undergraduate students in the United Kingdom found that 50% of British whites and 53.4% of British Asians were myopic (Logan, et al., 2005). In Greece, the prevalence of myopia among 15 to 18 year old students was found to be 36.8% (Mavracanas, et al., 2000). A study of Jordanian adults aged 17 to 40 found that over half (53.7%) were myopic (Mallen, et al., 2005).

In Australia, the overall prevalence of myopia (< –0.50D) has been estimated to be 17% (Wensor, et al., 1999). In one recent study, less than 1 in 10 (8.4%) Australian children between the age of 4 and 12 were found to have myopia worse than –0.50D (Junghans & Crewther, 2005). A recent review found that 16.4% of Australians aged 40 or over have at least –1.00D of myopia and 2.5% have at least –5.00D (Kempen, et al., 2004).

In the United States, the prevalence of myopia has been estimated at 20% (Kempen et al., 2004). Nearly 1 in 10 (9.2%) American children between the age of 5 and 17 have myopia (Kleinstein, et al., 2003). Approximately 25% of Americans between the age of 12 and 54 have the condition (Sperduto, et al., 1983). A recent review found that 25.4% of Americans aged 40 or over have at least –1.00D of myopia and 4.5% have at least –5.00D (Kempen et al., 2004).

In Brazil, a 2005 study estimated that 6.4% of Brazilians between the age of 12 and 59 had –1.00D of myopia or more, compared with 2.7% of the indigenous people in northwestern Brazil (Thorn, et al., 2005). Another study found nearly 1 in 8 (13.3%) of the students in the city of Natal were myopic (Garcia, et al., 2005).

High myopia (refractive spherical dioptric power of –6.00 or worse) affects 27% to 33% of all myopic eyes, up to 2% of the general population in the United States (Angle & Wissmann, 1980). It is especially common in Asian populations. In Japan, high myopia reportedly affects 6% to 18% of the myopic population and almost 2% of the general population (Matsumura & Hirai, 1999). High myopia most commonly appears as a complex disease caused by a combination of genetic and environmental

factors working together (Klein, et al., 2005).

However, it sometimes presents as one of the features in a wide variety of genetic disorders, including Stickler syndrome, Marfan syndrome, and chromosome abnormalities such as Down syndrome. High myopia is a major cause of legal blindness because of its association with an increased risk for premature cataracts, glaucoma, retinal detachment, and macular degeneration.

1.1.1.3 Public health concern

Myopia is a significant public health problem as it is associated with increased risk for visual loss (Curtin, 1979; Curtin, 1982; Goss & Winkler, 1983; Pararajasegaram, 1999). Myopic chorioretinal degeneration is the fourth most frequent cause of blindness leading to visual services and disability registration, and accounts for 8.8% of all causes of blindness. An estimated 5.6% of blindness among schoolchildren in the U.S. is attributable to myopia (Nallasamy, et al., 2007).

Remarkable resources are required for optical correction of myopia with spectacles, contact lenses, and, more recently, surgical procedures such as Laser-Assisted in Situ Keratomileusis (LASIK) (Javitt & Chiang, 1994; Pruett, 1995). The market for optical aids in the U.S. was estimated to exceed $8 billion in annual sales in 1990; most dollars were spent for the correction of myopia (Javitt & Chiang, 1994). The development of methods for preventing or slowing the onset of myopia, or for limiting its progression is highly significant (Pararajasegaram, 1999).

1.1.2 Epidemiology

1.1.2.1 Ethnicity and race

Prevalence of myopia varies in different parts of the world. The prevalence of myopia has been reported as high as 70–90% in some Asian countries, 30–40% in Europe and the United States, and 10–20% in Africa (Fredrick, 2002). In the younger Taiwanese population (Chen, et al., 1985) reported a prevalence of 84% in people by 16 years of age. Among Asian children of similar age, it was 30 to 50% (Lam, et al., 1994). Several recent studies among white children have shown contrasting findings that the prevalence of myopia was usually around the range 9 to 20% in European and North American countries (Kleinstein et al., 2003; Laatikainen & Erkkila, 1980).

Ethnic difference in myopia prevalence has been observed in young school children. Myopia starts at a very young age in countries such as Singapore and Japan and among the Chinese population in Taiwan and Hong Kong. In general, once the children start school, myopia starts to appear. There has also been a trend of increasing prevalence of myopia and the severity of myopia in the last decade. Females are reported to have an earlier onset and a slightly higher prevalence than males (Goss & Winkler, 1983; Wang, et al., 1994b).

Myopia usually starts after 6 years of age, progresses between the age of 7 and 16 years, and remains stable after the age of 20. Cross-sectional refractive data for individuals beyond age 16 suggest that the onset and increase of myopia after that age occurs although it is smaller in degree and appears limited to a subgroup. Few

myopia cases have onset among young adults and their progression is smaller than juvenile myopia.

1.1.2.2 Education, intelligence, and IQ

A number of studies have shown that the prevalence of myopia increases with level of education (Mavracanas et al., 2000; Sperduto et al., 1983) and many studies have shown a relationship between myopia and intelligence quotient (IQ). Myopic high school students aged 17 or 18 years performed better on IQ tests than their non-myopic classmates (Karlsson, 1976).

Comparison with test results obtained 10 years earlier before development of myopia suggested that the influence of the gene on the brain was of fundamental importance. According to Arthur Jensen, myopes average 7–8 IQ points higher than non-myopes (Jacobsen, et al., 2007). The relationship also holds within families: siblings with a higher degree of refraction error average higher IQs than siblings with less refraction error. Jensen believes that this indicates myopia and IQ are pleiotropically related as they are caused or influenced by the same genes. No specific mechanism that could cause a relationship between myopia and IQ has yet been identified.

In addition, other personal characteristics, such as value systems, school achievements, time spent in reading for pleasure, language abilities and time spent in sport activities, also correlated to the occurrence of myopia in studies (Mutti, et al., 2002; Saw, et al., 2001).

1.1.3 Aetiology and pathogenesis

The eye length is too long for most cases of myopia. Therefore, any etiologic explanation must account for such axial elongation. To date, no single theory is able to solely explain this elongation.

In the mid-1900s, investigators believed myopia to be primarily hereditary; the influence of near work in its development seemed "incidental" and the increased prevalence of the condition with increasing age was viewed as a "statistical curiosity" (Borish & Indiana University. Division of Optometry, 1970; Duke-Elder, 1935). While a number of theories regarding the aetiology of myopia exist, it is now widely believed that myopia results from a combination of multiple genes and environmental factors. (Chen et al., 1985; Goldschmidt, 2003; Hammond, et al., 2001; Morgan & Rose, 2005)

Currently two basic mechanisms are believed to cause myopia: form deprivation (also known as pattern deprivation) (Howlett & McFadden, 2006) and optical defocus (Saw, et al., 2002a). Form deprivation occurs when the image quality on the retina is reduced while optical defocus occurs when light focuses in front of or behind the retina. Numerous experiments with animals have shown that myopia can be artificially generated by inducing either of these conditions. In animals wearing negative spectacle lenses, axial myopia has been shown to occur as the eye elongates to compensate for optical defocus (Saw et al., 2002a). The exact mechanism of this image-controlled elongation of the eye is still unknown (Feldkamper & Schaeffel, 2003). It has been suggested that accommodative lag leads to blur (i.e. optical defocus) which in turn stimulates axial elongation and myopia (Schor, 1999). In

humans, it has been found that children with ptosis become myopic in the closed eye (O'Leary & Millodot, 1979). Children with corneal opacities tend to be myopic as do those with mild retrolental fibroplasia which distorts vision (Alfano, 1958).

1.1.3.1 Theories

- **Combination of genetic and environmental factors** — In China, myopia is more common in those with higher education background (Xu, et al., 2005). Some studies suggest that near work may exacerbate a genetic predisposition to develop myopia (Wolffsohn, et al., 2003). Other studies have shown that near work (reading, computer games, etc) may not be associated with myopic progression (Saw, et al., 2005b). However, "genetic susceptibility" to environmental factors has been postulated as one explanation for the varying degrees of myopia among individuals or populations (Hammond, et al., 2004) although there exists some difference of opinion as to whether it exists (Morgan & Megaw, 2004; Morgan & Rose, 2005). High heritability simply means that the variation in a particular population at a particular time is mainly due to genetic differences. If the environment changes — as, for example, it has by the introduction of televisions and computers — the incidence of myopia can change accordingly together with change in heritability. From a slightly different point of view, it could be concluded that — determined by heritage — some people are at a higher risk to develop myopia when exposed to modern environmental conditions with a lot of extensive near work like reading. In other words, it is often not the myopia itself which is inherited, but the reaction to specific environmental conditions and this reaction can be the onset and the progression of myopia.

- **Genetic factors** — The wide variability of the prevalence of myopia in different ethnic groups has been reported as additional evidence supporting the role of genetics in the development of myopia (Saw, et al., 1996). Measures of the heritability of myopia have yielded figures as high as 89%. Recent research has suggested genes that may be responsible: defective versions of the *PAX6* gene seem to be associated with myopia in twin studies (Hammond et al., 2004). Under this theory, the eye is slightly elongated front to back as a result of faults during development, causing images to be focused in front of the retina rather than directly on it. It is usually discovered during the pre-teen years between eight and twelve years of age. It most often worsens gradually as the eye grows during adolescence and then levels off as a person reaches adulthood. Genetic factors can work in various biochemical ways to cause myopia, and a weak connective tissue is a very essential one. Genetic factors include an inherited, increased susceptibility to environmental influences like excessive near work, and the fact that some people do not develop myopia in spite of very adverse conditions is a clear indication that heredity is involved somehow in such cases.

- **Environmental factors** — It has been suggested that genetic susceptibility to myopia does not exist (Morgan & Rose, 2005). A high heritability of myopia (as for any other condition) does not mean that environmental factors and lifestyle have no effect on the development of the condition. Some recommend a variety of eye exercises to strengthen the ciliary muscle. Other theories suggest that the eyes become strained by the constant extra work

involved in "near-work" and get stuck in the near position, and eye exercises can help loosen the muscles up thereby freeing it for far vision. These primarily mechanical models appear to be in contrast to research results, which show that the myopic elongation of the eye can be caused by the image quality with biochemical processes as the actuator. Common to both views is, however, that extensive near work and corresponding accommodation can be essential for the onset and the progression of myopia.

One recent Austrian study confirmed that the axial length of the eye does mildly increase while reading, but attributed this elongation to contraction of the ciliary muscle during accommodation (the process by which the eye increases optical power to maintain a clear image focus), not "squeezing" of the extraocular muscles (Drexler, et al., 1998). Near work and nightlight exposure in childhood have been hypothesised as environmental risk factors for myopia (Saw, et al., 2002b). Although one initial study indicated a strong association between myopia and nightlight exposure (Quinn, et al., 1999), recent research has found none (Guggenheim, et al., 2003; Gwiazda, et al., 2000; Zadnik, et al., 2000).

- Near work. Near work has been implicated as a contributing factor to myopia in some studies, but refuted in others (Saw et al., 2005b). One recent study suggested that students exposed to extensive "near work" may be at a higher risk of developing myopia, whereas extended breaks from near work during summer or winter vacations may retard myopic progression (Jiang, et al., 2005). Near work in certain cultures

(e.g. Vanuatu) does not result in greater myopia (Garner, et al., 1985; Grosvenor, 1988). It has been hypothesised that this outcome may be a result of genetics or environmental factors such as diet or over-illumination, changes which seem to occur in Asian, Vanuatu and Inuit cultures acclimating to intensive early studies (Mann & Ward, 2004).

- Time spent outdoors. A number of studies have shown that children who spend more time outdoors have lower rates of myopia. It is theorized that the higher brightness or the larger distances outdoors play a role (Dirani, et al., 2009; Rose, et al., 2008).

- Diet and nutrition. it was recently suggested that myopia may be caused by over-consumption of bread in childhood (Cordain, et al., 2002), or in general by diets too rich in carbohydrates, which can lead to chronic hyperinsulinemia. Various other components of the diet, however, were made responsible for contributing to myopia as well, as summarised in the article.

1.1.3.2 Ocular morbidity

Several eye diseases were shown to associate with high myopia, e.g. cataract, glaucoma, retinal detachment (RD) and posterior staphyloma with retinal degenerative changes (Curtin, 1982; Curtin, 1985; Grossniklaus & Green, 1992; Hotchkiss & Fine, 1981; Noble & Carr, 1982; Perkins, 1960; Perkins, 1979; Rabb, et al., 1981). These eye diseases are the leading causes of irreversible blindness. The

reason underlying the disease association is the excessive elongation of the eyeball in high myopia, which may be accompanied by degenerative changes in the fundus, e.g. the sclera, choroid, Bruch's membrane, retinal pigment epithelium (RPE), and neural retina. These changes include geographic areas of atrophy of the RPE and choroid, lacquer cracks in Bruch's membrane, sub-retinal hemorrhage, and choroidal neovascularisation. Among these fundus lesions, macular neovascularisation is the most common vision-threatening complication of high myopia (Avila, et al., 1984; Hayasaka, et al., 1990; Noble & Carr, 1982; Rabb et al., 1981). Clinical and histopathological studies have documented choroidal neovascularisation in 4 to 11% of highly myopic eyes (Avila et al., 1984; Burton, 1989; Vongphanit, et al., 2002). Low vision can result from myopic choroidal neovascularisation, often affecting relatively young patients. RD risk is 3 to 7 times greater for persons with myopia of at least -5.00D, relative to myopia of less than -5.00D (Burton, 1989). Myopia of -5.00 to -10.00D is associated with a higher likelihood of RD. The lifetime risk for RD is estimated to be 1.6% for patients with refractive error less than -3.00D and 9.3% for those with myopia greater than -5.00D (Wang, et al., 1994a). A sub-group with lattice degeneration and greater than -5.00D of myopia has a lifetime risk of 35.9% (Yura, 1998). The prevalence of lattice degeneration increases with higher myopia using axial length (AL) as a biometric measure of refractive error. Glaucoma was observed in 3% of patients with myopia who had ALs of >26.5 mm, in 11% with ALs between 26.5 and 33.5 mm, and in 28% of those with longer lengths (Vidic & Lerchner, 1990).

In conclusion, with the increasing prevalence of myopia worldwide, it is expected that the advanced forms of myopia will have an increasing effect on the quality of

ocular health and impose an increasing economic burden on society to treat the related complications. Several models of pathogenesis have been explored by characterisation of tissue changes in myopic eyes and the use of epidemiological models. However, genetic factors play a major role in the risk of developing myopia. In next section, current molecular genetic approaches to identifying genetic components that are contributory to myopia will be discussed.

1.2 The Genetics of myopia

In recent years, molecular genetic studies based on family linkage analysis and case-control association studies have transformed our understanding of myopia. The potential role of family history as a risk factor has been recognized for more than 30 years. However, there has been considerable debate as to the extent that genetic factors are critical to myopia pathogenesis.

While a number of theories regarding the aetiology of myopia exist, it is now widely believed that myopia is caused by a combination of multiple genes and multiple environmental factors that work together. Increased near work such as reading or higher educational levels may play a significant role in the abnormal eye growth and development of myopia. Additionally, animal models have demonstrated that myopia can be induced by artificially altering the visual experience early in life. However, these environmental factors cannot fully explain the whole story. In fact, it only explains a small part of the development of myopia.

1.2.1 Familial aggregation studies & twin studies

1.2.1.1. Family studies

Many family studies report the correlation relationship between parental myopia and myopia in their children, indicating a hereditary factor in myopia susceptibility (Goldschmidt, 1968; Pacella, et al., 1999; Patten & Howland, 1996; Yap, et al., 1993; Zadnik, 1997; Zadnik et al., 1994)

In Gwiazda's study, if both parents were myopic, nearly half the children became myopic by age 18; if only one parent was myopic, a quarter of their children were nearsighted too; but if neither parent was myopic, only 8 percent of the children were myopic. Yap et al. (1993) also reported that the prevalence of myopia in 7-year-old children was 7.3% when neither parent was myopic, 26.2% when one parent was myopic, and 45% when both parents were myopic. Pacella et al. (1999) also found that children with a family history of myopia had on average less hyperopia, deeper anterior chambers, and longer vitreous chambers even before becoming myopic, suggesting that they had inherited a myopic tendency and genes played a large part in the development of the initial shape and subsequent growth of the eye.

Adding a lot of near work to a parental history of myopia increased the likelihood of myopia in their children. The investigators explained that the familial patterns in the children may result from some combination of heredity and environment (Zadnik et al., 1994). A particular eye size and shape may be inherited from myopic parents along with a certain level of intelligence and a tendency for different patterns of near-work activity. Another group examined the influences of ethnicity, parental

myopia, and 'near work' on spherical equivalent refraction (SER) and axial length (AL) in a population-based sample of 2,353 Australian children (mean age, 12 years) (Ip, et al., 2008; Ip, et al., 2007). In multivariate analyses, odds of childhood myopia did not change with higher levels of near work. Interactions between parental myopia and ethnicity were significant for SER and AL, reflecting greater decreases in SER and greater increases in AL with the number of myopic parents in the children of East Asian ethnicity (15% of the sample) than in the children of European Caucasian ethnicity (60% of the sample). In the nonmyopic children, there was no association between parental myopia and AL. Thus, in this study, parental myopia was associated with more myopic SER and longer AL with significant ethnic interactions.

1.2.1.2. Modes of inheritance

In the past, several modes of inheritance for myopia have been proposed (Goss et al, 1988) although the majority of myopia is multifactorial in causation. An autosomal dominant (AD) mode of inheritance was suggested by a segregation analysis of French multiplex families with high myopia (Naiglin, et al., 1999). The sibling recurrence risk ratio (λs) for myopia (the increase in risk to siblings of a person with a disease compared with the population prevalence) has been estimated to be approximately 4.9 to 19.8 for high myopia (<–6.00D) and approximately 1.5 to 3.0 for low or common myopia (approximately –1.00 to –3.00D) (Guggenheim, et al., 2000), suggesting a definite genetic basis for high myopia and a strong genetic basis for low myopia. Other modes of inheritance have also been suggested, like autosomal recessive (Der Kaloustian & Baghdassarian, 1972; Edwards & Lewis, 1991), or X-linked recessive trait (Schwartz, et al., 1990). A high degree of familial

aggregation of refraction, particularly myopia, was recently reported in the Beaver Dam Eye Study population after accounting for the effects of age, sex, and education (Klein et al., 2005). Segregation analysis suggested the involvement of multiple genes rather than a single major gene effect.

1.2.1.3. Twin studies

Twin studies provide the most convincing evidence to support a strong genetic basis. Studies of twins are particularly useful for understanding myopia because they allow researchers to look at the interaction between heredity and environmental influences. Identical twins share 100% of their genes, as well as much of their early childhood environment. Non-identical twins also share a similar childhood environment, but are not genetically alike. Comparing these two groups allows researchers to assess the relative influence and interaction of genes and environment for complex diseases.

Multiple studies reported an increased concordance of myopia (Chen et al., 1985; Karlsson, 1974; Lin & Chen, 1987) as well as refractive biometric parameters (axial length, corneal curvature, lens power) (Rosner, et al., 1988; Toh, et al., 2005) in monozygotic twins compared to dizygotic twins .

The data from British twin study have found that an identical twin has an 80% chance of developing myopia if the other twin has myopia, versus a 40% chance if the other is a non-identical twin (McBrien, et al., 2008). The British twin study with 506 pairs of twins (both identical and non-identical) confirmed genetic influence as the most important factor for myopia with a heritability of 89% while environmental factors only accounted for 11%. This breakthrough contributes to a growing

scientific debate about whether the causes of a recent worldwide increase in myopia are genetic or due to changes in lifestyle such as children spending more time playing computer games and surfing the internet.

Another large myopia twin study with 627 twin pairs performed in Australian adults was named as the Genes in Myopia (GEM) (Dirani, et al., 2008a; Dirani, et al., 2008b). The study has provided an extensive twin database for genetic analysis. Heritability estimates for refraction and axial length reached 88% and 90% respectively. The axial length and refraction were suggested to share common genes in myopia aetiology.

In conclusion, multiple familial aggregation analyses and twin studies reinforced the importance of genetic risk factors for myopia. However, the epidemiological association between educational period (near work) and myopia, the evidence of increasing myopia prevalence within a few generations, and the theory of gene-environment interaction may imply that some individuals might be genetically liable to develop myopia if exposed to certain environmental factors. These studies could not establish how many or which genes were potentially involved in the disease. In the next part, the molecular approach to underpinning the myopia will be discussed.

1.2.2 Genetics of ocular refractive components

Refraction is determined by coordinated contribution of ocular biometric components such as axial length (AL), anterior chamber depth (ACD), corneal

curvature (CC), and lens thickness (LD) (Paget, et al., 2008). Separately, these components may be assessed as quantitative traits intimately related to the clinical phenotype of myopia. Multiple reports have examined familial aggregation and heritability of these ocular components (Goldschmidt, 1968; Teikari, et al., 1991; von Noorden & Lewis, 1987; Yap et al., 1993). These may be intimately related to the phenotype of myopia, and the genetic effects of these traits should be examined.

The inverse relationship of AL to refraction has been fully examined (Ashton, 1985; Farbrother, et al., 2004a; Goss, et al., 1988; Guggenheim et al., 2000; Naiglin et al., 1999). AL is the largest contributor to the determination of refractive error (Klein et al., 2005; Sorsby, et al., 1966). Estimates of heritability for AL range from 40% to 94% (Goldschmidt, 1968; Sorsby & Fraser, 1964; Wallman, et al., 1978; Yap et al., 1993). A study of three large Sardinian families found modest evidence for linkage on chromosome 2p24 with a logarithm of the odds (LOD) score of 2 (Biino, et al., 2005). Chromosome 5q together with 6, 10, and 14 loci were identified for the implication of ocular axial length (Zhu, et al., 2008). AL includes anterior chamber depth (ACD), and studies have shown that increased ACD has an inverse relationship as well to refractive error (Farbrother et al., 2004a). The estimates of heritability reported for ACD range from 70% to 94% (Schwartz et al., 1990; Wallman et al., 1978; Yap et al., 1993), and the same Sardinian study found modest linkage evidence to chromosome 1p32.2 with a LOD score of 2.32 (Biino, et al., 2005).

Corneal curvature steepness is more likely to result in myopia as hyperopic eyes are more likely to have flatter corneal curvatures (Ashton, 1985; Schwartz et al., 1990;

Young, et al., 2004). Heritability estimates for corneal curvature range from 60% to 92% (Goldschmidt, 1968; Klein et al., 2005; Wallman et al., 1978; Yap et al., 1993). The same Sardinian study also noted modest linkage of corneal curvature to chromosomes 2p25, 3p26, and 7q22 with LOD scores ranging from 2.34 to 2.50 (Biino, et al., 2005). Increased lens thickness correlates with increased myopia (Farbrother et al., 2004a). A di- and mono-zygotic twin study showed 90% to 93% heritability for lens thickness (Alsbirk, 1977; Lyhne, et al., 2001; Yap et al., 1993).

1.2.3 Animal models - Gene expression and functional Studies

Animal models provide clues to genes and protein systems important in relevant behaviours, but it is unlikely that animal behaviours will fully parallel the physiological myopia in humans.

1.2.3.1 Active emmetropisation mechanism

Animal models have revealed an active emmetropisation mechanism: eyeball grows to the correct size for the image to be in focus on the retina (Norton, 1999; Wallman & McFadden, 1995; Wallman, et al., 1995; Wildsoet, 1997). The active process is most easily studied in newborn animals because their eyes show substantial growth over the first few weeks of life. This mechanism generates signals initially from the retina (Fischer, et al., 1999; Mertz & Wallman, 2000). Constant hyperopic defocus produces retinal signals to the growing wall of the eyeball and cause axial elongation (Mertz & Wallman, 2000; Siegwart & Norton, 1999). This, in turn, reduces the

hyperopic defocus so that this feedback system is self-limiting until the axial length matches to optical power. Therefore, the local control of the eyeball size is within the retina from bipolar or amacrine cells; and no signals from ganglion cells are transmitted to the visual cortex.

Genes expressed in retina, RPE, choroid, or sclera that are involved in the normal emmetropisation process could be involved in myopia development with irregular expression so that the emmetropisation mechanism is disrupted, causing the eye to elongate and become myopic (Gentle, et al., 2003; Guggenheim & McBrien, 1996; Mertz & Wallman, 2000; Norton & Rada, 1995; Siegwart & Norton, 2002). The retina is suspected of having the highest likelihood of harbouring expressed candidate genes that influence scleral growth, based on multiple studies including optic nerve sectioning in animal models (Wallman, et al., 1987). Human retinal gene expression has been well studied (Bortoluzzi, et al., 2000; Bowes Rickman, et al., 2006; den Hollander, et al., 1999; Sohocki, et al., 1999).

1.2.3.2 Knockout animal models

Chickens and knockout mice have been used to evaluate the early growth response (*Egr-1*) gene that was shown to be related to myopia via axial eye growth (Schippert, et al., 2007). This model involved the immediate early gene transcription factor ZENK (egr-1). ZENK expression is suppressed by minus lenses and form deprivation, which lead to ocular elongation, and is enhanced by plus lenses and termination of form deprivation, which suppress ocular elongation (Bitzer & Schaeffel, 2002). Researchers revealed that ZENK might be involved the regulation of the axial eye growth. ZENK knockout mice had longer eyes and a myopic shift

relative to heterozygous and wild-type mice with identical genetic background (Schippert et al., 2007).

Another double-knockout mice model also suggested lumican and fibromodulin as potential candidate genes for susceptibility to high myopia (Chakravarti, et al., 2003). When compared with the wild-type mice, the lumican and fibromodulin double-knockout mice had increased ocular axial length and thinner sclera with altered collagen architecture. These are some of the key features of high myopia. Mutations or altered expression of these proteoglycans may contribute to myopia in humans.

1.2.3.3 Eye growth

A mouse model was developed to study the genetic and environmental factors for eye growth and axial components during emmetropisation in C57BL/6 mice (Zhou, et al., 2008). One major advantage of the mouse model is the substantial heritability for mapping genes that specifically modulate growth of different parts of the eye.

1.2.4 Molecular level - Myopia loci

Molecular genetic studies of myopia by family-based linkage analysis have yielded valuable insights into the risks of developing this condition and potential disease-causing mechanisms. With the present genetic studies, one finds a variety of myopia definitions, including discrete categories of affected and unaffected states, and grading systems that allow for quantitative trait analyses.

Twenty-four chromosomal loci have been identified for different types of myopia. Among them, ***MYP1*** to ***MYP5***, ***MYP11*** to ***MYP13***, ***MYP16***, and ***MYP18*** are linked to high myopia, and ***MYP2***, ***11***, ***13***, ***16***, and ***18*** are found in the Chinese population. The fact that some loci have not been replicated by other studies is partly explained by the different definitions of the phenotype (**Table 1.3**). Genetic heterogeneity can also explain the same clinical findings in separate families related to mutations in different genes. On the other hand, variable expressivity can account for the same mutation of a gene leading to variable clinical findings among multiple individuals (Taylor, et al., 2004).

1.2.4.1 The loci of high myopia

MYOPIA 1 (*MYP1*)

The first myopia locus was mapped on chromosome Xq28 (named ***MYP1***) from a family with X-linked recessive form of myopia (Schwartz et al., 1990). The syndrome was later renamed as the Bornholm eye disease, and includes myopia, lazy eyes and red-green color blindness as the clinical features. A Danish family with X-linked myopia was reported with astigmatism, impaired vision and moderate hypoplasia of the optic nerve heads (Haim, et al., 1988). Deuteranopia was present in all affected males. The syndrome was traced in 5 generations of the family which had its origin on the island of Bornholm. Schwartz et al. (1990) found linkage to the factor VIII gene with a maximum LOD score of 4.8. Thus, the locus appears to be located on Xq28. However, a recent Chinese family linkage study with X-linked recessive high myopia did not identify any potential causative mutation within the loci and its adjacent intronic regions (Guo, et al., 2010).

MYOPIA 2 (*MYP2*)

Following the first myopia locus, linkage studies of a region on chromosome 18p11.31 were conducted with multiple high myopia families (Farbrother et al., 2004a; Lam, et al., 2003a; Young, et al., 2001; Young, et al., 1998b). The area implicated is broad, proximal to marker D18S52, with a likely interval of 0.8 cM between markers D18S63 and D18S52 (Young et al., 2001). Research in this region demonstrates the intersections of linkage with chromosomal and association methods. In an initial study of the candidate gene *TGIF*, six SNPs showed a significant difference ($P < 0.05$) between patients and control subjects in univariate analysis, and therefore *TGIF* was suggested as a myopia candidate gene (Lam et al., 2003a). However, subsequent association studies failed to replicate the positive signal (Hasumi, et al., 2006; Pertile, et al., 2008; Wang, et al., 2009b).

MYOPIA 3 (*MYP3*)

Another myopia locus at 12q21-q23 was identified in a large German/Italian family segregating autosomal dominant high-grade myopia (refractive error greater than or equal to 6 D) (Nurnberg, et al., 2008; Young et al., 1998b). The average age at diagnosis of myopia was 5.9 years. The average spherical component refractive error for the affected individuals was -9.47 D. The maximum LOD score with 2-point linkage analysis was 3.85 at a recombination fraction of 0.0010 with markers D12S1706 and D12S327. Recombination events defined a 30.1-cM interval on 12q21-q23 for this second autosomal myopia locus.

Within this region, Young et al. (1998) identified decorin (*DCN*) at 12q23 and lumican (*LUM*) at 12q21.3-q22 as potential candidate genes. These are members of

the small interstitial proteoglycan family of proteins that are expressed in the extracellular matrix (ECM) of various tissues. Both proteins interact with collagen and limit the growth of fibril diameter. Dermatan sulfate proteoglycan-3 (*DSPG3*), which maps to 12q21, is another small interstitial proteoglycan that is expressed in cartilage, ligaments and placental tissues. Young et al. (1998) suggested that fibrillogenesis of the sclera may be affected by mutations in these candidate proteins, as has been demonstrated in connective tissue disorders such as Stickler syndrome and Marfan syndrome.

MYOPIA 4 (*MYP4*)

Twenty one French and two Algerian families with autosomal dominant high-grade myopia (<-6 D) were reported and the families varied from small nuclear families to extended multi-generational pedigrees (Naiglin et al., 1999). They excluded previously identified myopia loci and found suggestive evidence of linkage to chromosome 7q36, with a maximum multipoint LOD score of 2.81. No locus heterogeneity was detected.

In a combined analysis of two genomewide scans of high-grade myopia for the families reported by Naiglin et al. (2002) and a subset of 9 newly collected families, Paget et al. (2008) did not found linkage evidence in 7q36. However, a non-parametric model demonstrated significant linkage to chromosome 7p15 in all of the families (Z-NPL = 4.07, P = 0.00002). The interval was 7.81 cM between markers D7S2458 and D7S2515.

A quantitative trait locus (QTL) linkage analysis was performed for ocular refraction

in 96 African American families (Ciner et al., 2008). Evidence of linkage was found for a region on 7p15 (maximum LOD score of 5.87) in a 17.5Mb region between markers D7S1802 and D7S2846. The authors noted that a previous study in European-derived families by Klein et al (2007) had found evidence of linkage to chromosome 7p21.

MYOPIA 5 (*MYP5*)

Paluru et al (2003) reported a novel locus for autosomal dominant high-grade myopia in a multigenerational English/Canadian family, which showed linkage to 17q21-q22. Fine mapping and haplotype analysis refined the critical interval to a 7.71-cM region on 17q21-q22 between markers D17S787 and D17S1811. Previously identified myopia loci were excluded in this family.

MYOPIA 11 (*MYP11*)

A large Chinese family with an autosomal dominant inheritance pattern for myopia was reported to show linkage to 4q22-q27 (Zhang, et al., 2005). The family included 12 affected individuals. The linkage region on chromosome 4 was between D4S1578 and D4S1612 with a maximum LOD score of 3.11 at marker D4S1564. The investigators prioritized the rhodopsin homolog *RRH* gene within the linked region as a candidate gene for mutation screening, but no causative mutations were found.

MYOPIA 12 (*MYP12*)

Paluru et al. (2005) presented linkage evidence for 2q37.1. They found that high-grade myopia in a large U.S. family of northern European extraction was tightly linked (maximum multipoint LOD score = 4.75) to the microsatellite (MS) marker D2S2344. This is the same site as two candidate genes, S-antigen (*SAG*; MIM ID#181031) and diacylglycerol kinase-delta (*DGKD*; MIM ID#601826), but no causative mutations were found.

MYOPIA 13 (*MYP13*)

Zhang et al (2006) conducted an X-chromosome linkage analysis of a 4-generation Chinese family in which 6 males had high myopia inherited in an X-linked recessive pattern. They mapped the high myopia locus to a 25-cM region on Xq23-q25 between DXS1210 and DXS8057with maximum 2-point LOD scores of 2.75 and 2.29 for DXS1001 and DXS8059, respectively. X-chromosome linkage analysis in another Chinese family further identified a candidate locus for high myopia on Xq25-q27.2, which overlapped *MYP13* by approximately 4.80 cM (Zhang, et al., 2007).

MYOPIA 15 (*MYP15*)

A genomewide scan with 382 MS markers was recently conducted for a large Hutterite family from South Dakota and this family included 7 patients with non-syndromic high-grade myopia (Nallasamy et al., 2007). They detected linkage of high myopia to 10q21.1 with a maximum multipoint LOD score of 3.22 under an autosomal dominant model. Haplotype analysis demonstrated 2 distinct haplotypes segregating with the disorder, indicative of 2 distinct mutations in the same gene.

MYOPIA 16 (*MYP16*)

A Hong Kong research group conducted a genomewide linkage search in 3 Chinese pedigrees from Hong Kong with autosomal dominant high myopia (Lam, et al., 2008). The MS markers spanning the whole genome with an average spacing of 10 cM were adopted. Significant evidence for linkage with maximal LOD scores of 4.81 was observed on chromosome 15.33-p15.2 with a 17.45-cM interval. Five positional genes (*IRX2, IRX1, POLS, CCT5*, and *CTNND2*) were screened, but no segregation of polymorphisms with high myopia was found.

1.2.4.2 The loci of common myopia

Early linkage studies showed that the genes for high myopia (on chromosome 18 and 12) did not play any role in causing the milder form of myopia. Evidence for a new region on chromosome 22 was found when extended families of the Ashkenazi Jews and the Amish community in Lancaster, Pennsylvania, were studied (Ibay, et al., 2004). These analyses demonstrated the complex genetic basis of ocular refraction, and the discovery of disease-causing variants may be used to aid identification of other susceptibility loci for myopia of different severity.

MYOPIA 6 (*MYP6*)

A recent linkage study concluded that the localization to chromosome 22q12 is important for susceptibility to mild/moderate myopia rather than high myopia (Stambolian, et al., 2004). The mild/moderate form of myopia is more prevalent than high myopia, especially in Orthodox Jewish males. They investigated the contribution of this locus to the inheritance of common myopia in 44 large American

families of Ashkenazi Jewish descent, each with at least 2 affected sibs. They refined the localization to a region near D16S409, obtaining a maximum LOD score of 3.54 at marker D22S685.

Another study also suggested the possible involvement of a locus on 22q12 (Klein, et al., 2007). They performed non-parametric sib-pair and genomewide linkage study of ocular refraction, adjusting for age, education, and nuclear sclerosis, in 834 sib pairs in 486 extended pedigrees in the Beaver Dam Eye Study (Klein et al., 2007).

MYOPIA 7-10 (*MYP7 to MYP10*)

Hammond and his associates used 737 MS markers in a genomewide linkage screen for the quantitative measurement of refraction in 221 of the dizygotic twin pairs of the Twins UK Registry (Hammond et al., 2004). They observed evidence for linkage at 11p13 (*MYP7*), with a LOD score of 6.1. Other linkage peaks were observed at chromosomes 3q26 (*MYP8*), 4q12 (*MYP9*), and 8p23 (*MYP10*) (Stambolian, et al., 2006), with LOD scores of 3.7, 3.3, and 4.1, respectively. These analyses demonstrated the complex genetic basis of ocular refraction, and that the discovery of disease-causing variants may help in identifying additional susceptibility loci for myopia.

The *PAX6* gene (homeobox gene), at the precise location of chromosome 11p13, was directly beneath the highest linkage peak. *PAX6* is known as a master-control for eye development and other neural functions, so there is much promise of finding a gene or genes in this region (Hammond et al., 2004; Mutti, et al., 2007).

However, the British investigators failed to find a phenotypic association with common SNPs in the *PAX6* gene. They suggested that further mapping is required to confirm whether linkage to *PAX6* is due to regulatory loci or to linkage to an unrecognized nearby gene (Hammond et al., 2004).

MYOPIA 14 (*MYP14*)

Wojciechowski and co-workers undertook a systematic screening of the entire genome with 387 MS markers for identifying susceptibility genes for ocular refraction in 186 affected sib pairs from 49 multigenerational Ashkenazi Jewish families (Wojciechowski, et al., 2006). They provided strong evidence for the presence of susceptibility loci for common myopia on chromosomes 1p36 between markers D1S552 and D1S1622. Multipoint regression-based QTL linkage analysis yielded a LOD score of 9.5 for refractive error.

Previous genetic linkage studies of myopia have recently been reviewed and summarised (Hornbeak & Young, 2009; Jacobi, et al., 2005; Morgan & Rose, 2005; Tang, et al., 2008; Young, 2004). Some of these loci have been repeatedly detected even with different disease models and analytical methods. In contrast, there are several loci that have been reported in one or two studies only. Some may represent false-positive linkage signals while others may be specific to the manner in which the myopia was defined and/or analysed. The replication of linkage signals requires a considerably larger sample size than the original cohort used for the initial discovery. Nearly all of the present linkage studies are insufficiently powered to test all of the reported loci, some of which may have only a limited contribution to the risk of developing myopia. Given the large genetic intervals that are implicated by these

linkage studies, subsequent follow-up by association studies is essential to test potential candidate genes in the implicated regions. The first dramatic success of association studies for myopia was the discovery of a variant in the paired box gene 6 (*PAX6*) gene by genomewide and focused chromosome 11 genotyping of single nucleotide polymorphisms (SNPs) (Hammond et al., 2004; Mutti et al., 2007). It is valuable to test whether reported associations within these genetic intervals actually account for the linkage evidence that has been reported. For this fundamental ocular gene, the power of the associations and the consistency of the results among multiple studies provide convincing evidence that they play major roles in the pathogenesis of myopia (Han, et al., 2009; Ng, et al., 2009; Tsai, et al., 2008; Zayats, et al., 2008). These associations and the reported odds ratios (ORs) are summarised by Mutti et al (2007). A few of these reports have been replicated. Given the possibility for false-positive association signals, one must adjust the level of significance to account for the number of variants that are tested and use a highly stringent criterion for accepting the result. One should generally view more modest association results as no more than suggestive even when the association is with a variant that is plausible from a biological perspective (such as the case for the *LIPIN2* gene) (Scavello, et al., 2005; Zhou & Young, 2005).

Table 1.3 Summary of Myopia Loci (based on Tang et al, 2008)

Myopia locus	Authors (years)	Inheritance / QTL †	Chr. No.	Location	Ethnicity of subjects	Types of families	Linkage analysis ‡	Affected status§	Max LOD¶
MYP1	Schwartz et al. (1990)	XR	X	**Xq28**	Danish	Large pedigrees	PL	–	4.8
MYP2	Young et al. (1998)	AD	18	**18p11.31**	American and Chinese	Moderate to large multigenerational families	PL	≤ -6.00 D SE	9.59
MYP2	Lam et al. (2002)	AD	18	**18p11.31**	Hong Kong Chinese	Moderate pedigrees	PL	≤ -6.00 D	2.1
MYP3	Young et al. (1998)	AD	12	**12q21-23**	German/Italian	A large pedigree	PL	≤ -6.00 D SE	3.85
MYP3	Farbrother et al. (2004)	AD	17	**12q21-23**	UK population	Nuclear families	PL, NPL	≤ -6.00 D in the least negative meridian of both eyes	2.54
MYP4	Naiglin et al. (2002)	AD	7	**7q36**	French and Algerian	Large to moderate pedigrees	PL, NPL	≤ -6.00 D both eyes	2.81
MYP5	Paluru et al. (2003)	AD	17	**17q21-22**	English/Canadian	A large pedigree	PL	≤ -6.00 D SE	3.17
MYP6	Stambolian et al. (2004)	AD	22	**22q12**	American families of Ashkenazi Jewish descent	Large pedigrees	PL, NPL	≤ -1.00 D in each meridian for both eyes	3.54
MYP6	Stambolian et al. (2006)	AD	22	**22q12**	Additional Jewish descent	pedigrees	PL, NPL	≤ -1.00 D in each meridian for both eyes	4.73
MYP6	Klein et al. (2007)	QTL	22	**22q12**	Americans of Northern European and/or German ancestry	Sib-pairs	NPL	Mean +0.44 D SE; range: -12.12 to +8.38D	P value = 0.00330
MYP7	Hammond et al. (2004)	QTL	11	**11p13**	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	6.1
MYP8	Hammond et al. (2004)	QTL	3	**3q26**	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	3.7
MYP9	Hammond et al. (2004)	QTL	4	**4q12**	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	3.3
MYP10	Hammond et al. (2004)	QTL	8	**8p23**	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	4.1

(Continued on next page)

Table 1.3 (continued)

Myopia locus	Authors (years)	Inheritance / QTL †	Chr. No.	Location	Ethnicity of subjects	Types of families	Linkage analysis ‡	Affected status§	Max LOD¶
MYP10	Stambolian et al. (2005)	AD	8	**8p23**	Old Order Amish	Families with affected sibs	PL, NPL	≤ -1.00 D in each meridian for both eyes	2.03
MYP11	Zhang et al. (2005)	AD	4	**4q22-27**	Han Chinese in a small village of central China	A large pedigree	PL	Range: -5.00 to -20.00 D	3.11
MYP12	Paluru et al. (2005)	AD	2	**2q37.1**	US family of northern European	A large pedigree	PL	≤ -6.00 D SE; range: -7.25 to -27.00 D	4.75
MYP13	Zhang et al. (2006)	XR	X	**Xq23-25**	Chinese	A large pedigree	PL	≤ -6.00 D SE; range: -6.00 to -20.00 D	2.75
MYP13	Zhang et al. (2007)	XR	X	**Xq23-27.2**	Chinese	A large pedigree	PL	≤ -6.00 D SE; range: -7.00 to -16.00 D	2.79
MYP14	Wojciechowski et al. (2006)	QTL	1	**1q36**	Ashkenazi Jewish	Moderate to large multigenerational families	PL	Mean -3.46 D SE; ≤ -1.00 D in each meridian for both eyes	9.54
MYP15	Klein et al. (2007)	QTL	1	**1q41**	Americans of Northern European and/or German ancestry	Sib-pairs	NPL	Mean +0.44 D SE; range: -12.12 to +8.38D	P value = 0.00019
MYP16	Klein et al. (2007)	QTL	7	**7p21**	Americans of Northern European and/or German ancestry	Sib-pairs	NPL	Mean +0.44 D SE; range: -12.12 to +8.38D	P value = 0.0023
MYP17	Nallasamy et al. (2007)	AD	10	**10q21.2**	Hutterite population from South Dakota	A large pedigree	PL	Mean -7.04 D; range -3.75 to -13.25 D	3.22

† Inheritance is indicated as X-linked recessive (XR), autosomal dominant (AD). QTL represents quantitative trait locus.

‡ PL represents parametric linkage analyses whereas NPL represents non-parametric linkage analyses, usually affected sib pair methods.

§ SE represents spherical equivalent in diopters (D)

¶ Max. LOD stands for maximum “logarithm of the odds” score.

1.2.5 Molecular studies on myopia – Candidate gene studies

The next steps of verifying the above results involve twin studies and segregation analyses, performing analyses on more markers within a smaller region of the genome, and using single nucleotide polymorphisms (SNPs) to identify associations with the disease. Most of these genes have also been selected on the basis of findings in linkage studies or animal models.

Myopia is only one example of the challenge for the genetics of complex diseases. But the technologies are rapidly evolving and enable investigators to test association at a number of candidate genes in smaller regions. Therefore, finding the genes for myopia is truly promising.

In this chapter, we will introduce the recent discoveries regarding the genes that have been postulated for myopia, such as *PAX6* (Hammond et al., 2004), *TGIFB1* (Lam et al., 2003a), *HGF* (Han, et al., 2006), *MMP3* (Liang, et al., 2006), *MMP9* (Liang et al., 2006), *MFN1* (Andrew, et al., 2008), *PSAR* (Andrew et al., 2008), *COL2A1* (Mutti et al., 2007), *COL1A1* (Inamori, et al., 2007), *TGFB1* (Lin, et al., 2006; Zha, et al., 2009), *LUM* (Chen, et al., 2009; Lin, et al., 2010; Majava, et al., 2007; Paluru, et al., 2004; Solomon, et al., 2009), and *TEX28* (Metlapally, et al., 2009b).

1.2.5.1 Positional candidate gene studies for non-syndromic myopia loci

Investigators have also tested association for a number of candidate genes in myopia loci, as shown in **Table 1.4**.

PAX6

The most compelling candidate gene, *PAX6* on chromosome 11p13, was discovered in dizygotic twins from a genomewide scan for the quantitative measurement of refraction. *PAX6* is expressed in the human eye (Nishina, et al., 1999) and plays an evolutionarily conserved role in ocular development.

Although initial studies did not implicate the coding region of *PAX6* in Caucasian populations (Favor, et al., 2008; Hammond et al., 2004; Mutti et al., 2007; Simpson, et al., 2007), further study with independent Chinese population samples was warranted after association to dinucleotide repeats within the P1 promoter as reported (Ng et al., 2009). The upstream promoter or regulatory variants in *PAX6* have been reported to associate with high myopia in Han Chinese nuclear families (Han et al., 2009) and with extreme myopia in a Taiwan Chinese population (Tsai et al., 2008). Interestingly, SNPs or haplotypes of SNPs at the 3′ untranslated region of the *PAX6* gene were recently found to be associated with high myopia in two separate Chinese studies, one in Hong Kong and another in Taiwan (Jiang, et al., 2011; Liang, et al., 2011).

Table 1.4 Positional candidate genes within myopia loci

Category	Chromosomal Localization	Subject (family)	Candidate genes/proteins
MYP1	Xq27.3-28	Danish	*TEX28* (Metlapally et al., 2009b)
MYP2	18p11.31	American/Chinese	*CLUL1,TGIF,EMILIN-2* (Scavello et al., 2005)
MYP3	12q21-q23	Italian/German	*LUM,DCN* (Paluru et al., 2004)
MYP4	7q36	French/Algerian	*VIP receptor 2* (Vessey, et al., 2005)
MYP5	17q21-q23	English/Canadian	*COL1A1,CHAD* (Inamori et al., 2007)
MYP6	22q12	Ashkenazi Jews	*EMO2* (Li, et al., 2009)
MYP7	11p13	English (dizygotic twins)	*MFRP* (Schache, et al., 2007) *PAX6* (Favor et al., 2008; Hammond et al., 2004; Mutti et al., 2007; Simpson et al., 2007)
MYP8	3q26	English (dizygotic twins)	*RP1L1* (Conte, et al., 2003)
MYP11	4q22-q27	Chinese	*RRH,* (Zhang et al., 2005)

TGIFβ

Another positional candidate gene within the *MYP2* locus (Young et al., 1998b) on 18p11.3 region is transforming growth factor β–induced factor (*TGIFβ*) (Lam et al., 2003a). The gene was responsible for the control of retinoid-responsive transcription (Bertolino, et al., 1995). Mutations lead to holoprosencephaly, a prevalent congenital disorder of brain and craniofacial malformation (Overhauser, et al., 1995). A modest association was reported for SNPs of the gene in Hong Kong Chinese patients with high myopia, but this could not be replicated by other groups (Hasumi et al., 2006; Pertile et al., 2008; Scavello, et al., 2004).

COL1A1

The collagen type 1 alpha 1 gene (*COL1A1*) is expressed as a component of the ECM in the scleral wall, and maps within the *MYP5* locus for high myopia on chromosome 17q22-q23.3. A case-control study of a Japanese cohort showed association between high myopia and 2 SNPs of this gene (Inamori et al., 2007), but the subsequent studies failed to replicate with other populations (Liang, et al., 2007; Nakanishi, et al., 2009a).

UMODL1

The uromodulin-like 1 (*UMODL1*) gene, on chromosome 21q22.3, was identified as a positional candidate gene during a case-control association analysis in Japanese high-myopia patients (Nishizaki, et al., 2009). One associated SNP was found within a frequent recombinant region, confirming the gene’s candidacy as a disease susceptibility gene.

MFN1 and PSAR

The *MFN1* gene is located on chromosome 3q26 (*MYP8*), upstream from alternate-splicing *SOX20T* and *PSARL* (Andrew et al., 2008). The *MFN1* and *PSARL* gene products have mitochondrial regulatory function in the retina, and thus mitochondrial-related pathogenetic pathway might be involved in common myopia development.

1.2.5.2. Functional Candidate Gene Studies

Here we report the hypothesised candidate genes for myopia based on the current understanding of the pathophysiology.

HGF

Han et al (2006) concluded from a family-based association study using SNPs that high myopia was associated with the hepatocyte growth factor gene (*HGF*). *HGF* is an important multifunctional cytokine, is expressed in the eye, and maps to 7q21.1 interval, which is homologous to the mouse *Eye1* locus (Grierson, et al., 2000; Veerappan, et al., ; Yanovitch, et al., 2009).

TGFB1

Lin et al concluded from the *Eye2* locus using SNP polymorphisms that high myopia was associated with the transforming growth factor β1 gene (*TGFB1*) in a Chinese Taiwanese adult population (Lin et al., 2006). The positive finding was recently replicated by our group (Zha et al, 2009). TGFB1 is a transcription factor and modulates the production of ECM (Rohrer & Stell, 1994). It is a growth regulator of

scleral chondrocytes and scleral fibroblasts that in turn affects scleral shape such as AL.

MFRP

The membrane-type frizzled-related protein (*MFRP*) gene is hypothesised to play a role in axial length regulation (Sundin, et al., 2005). Nanophthalmos – an autosomal recessive form of extreme hyperopia – is caused by null mutations in the gene encoding the membrane-type frizzled-related protein (*MFRP*). Intriguingly, a recent study reported no association between 16 *MFRP* SNPs and non-syndromic high myopia (Metlapally et al., 2008).

TEX28

A recent molecular approach for complex disease emphasized the significance of copy number variations (CNVs) of the *TEX28* (also named CXorf2) on the *MYP1* X-linked myopia phenotypes (Metlapally et al., 2009b). The gene is a nested gene within this cone pigment gene array. This gene was more highly expressed in SAGE libraries of the human macular retina relative to peripheral retina

1.2.5.3. Scleral Gene Studies

Genes that are important for constituent organization and maintenance of connective tissue function may be physiologically important in myopic ocular shape change as it is related to scleral wall expansion. The human and mouse ocular tissue expression studies reported candidate genes that may be relevant to an ECM-associated function in the sclera.

COL2A1

The myopia-associated collagen-related genes that have been reported from case-control association studies include the collagen 2 alpha 1 (*COL2A1*) gene, which maps to chromosome 12q13.11. Owing to their ECM expression in the sclera, *COL2A1* has been suggested as a candidate gene in the aetiology of familial high myopia. The association of structural gene mutations with high myopia is well established from Stickler disease (Zechi-Ceide, et al., 2008). The association has been remarkably consistent among the multiple studies (Metlapally, et al., 2009a; Mutti et al., 2007), thus confirming that myopia can result from defects in type II collagen.

LUM

The lumican *(LUM)* gene falls within the *MYP3* locus on chromosome 12q21-q23 (Paluru et al., 2004). It encodes a member of the family of structural proteins called small leucine-rich repeat proteoglycans (SLRPs) (Majava et al., 2007). Owing to their ECM expression in the sclera, *LUM* has been suggested as a candidate gene for myopia development in the scleral theory.

Lumican-deficient mice were shown to have a disrupted collagen fibril formation in the sclera, which was associated with larger eyes on the basis of volumetric estimations (Austin, et al., 2002; Chakravarti et al., 2003). However, no mutation was found in the *lumican* gene in the family with 12q21-22-linked high myopia (Paluru et al., 2004). Two other studies in Taiwan found that common polymorphisms in the regulatory domains of the *LUM* gene were associated with

high myopia (Chen et al., 2009; Lin, et al., 2009).

Matrix Metalloproteinases

It was hypothesised that the three genes coding for matrix metalloproteinases (*MMP-1*, *MMP-3*, *MMP-9*) represented good functional candidate genes for common myopia (Andrew et al., 2008). The genes are related to the degradation of matrix proteins and might thus modulate scleral extensibility. In a study of white English patients, a strong association was found with the dose of polymorphic alleles in these three genes (Hall, et al., 2009). However, associations of refraction to polymorphisms near *MMP1* and within *MMP2* were identified in the Amish but not among the Ashkenazi families (Wojciechowski, et al., 2010). Genetic heterogeneity was suggested as a reason for the inconsistent association results between different population sample sets.

1.3. Genetic approach to studying myopia

In recent years, there has been a greater interest in studying the genetic basis of common disorders, in which multiple genes of small effect are involved. Current studies take advantage of SNPs which are easy to type, highly abundant and stable. This opens up unprecedented opportunities for understanding the complex genetic traits (Altshuler, et al., 2000). In these studies, a large number of individuals must be genotyped for a large number of SNPs (Risch & Merikangas, 1996). Thus, fulfilling this promise requires technologies that can be used to genotype SNPs efficiently and inexpensively.

Here, the current status of SNP genotyping is discussed in terms of the principles of

allelic discrimination, the reaction formats, the detection modalities and the running cost. A number of genotyping methods currently in use are described to illustrate the approaches being taken. Although no single genotyping method is ideally suited for all applications, a number of good genotyping methods are available to meet the needs of many study designs. The challenges for SNP genotyping in the near future include increasing the speed of assay development, reducing the cost of the assays, and performing multiple assays in parallel.

1.3.1 Introduction of SNP genotyping methods

The increase of interest in SNPs is reflected from the rapid development of a diverse range of SNP genotyping methods. This section provides an overview of the major strategies for interrogating SNPs.

1.3.1.1 Single nucleotide polymorphisms (SNPs): the unit of variation

The human genome comprises 3.2×10^9 basepairs (bp), of which approximately 99.9% are identical among us all. In the remaining 0.1% lies the variability that makes us unique and is largely responsible for our phenotype, including our appearance, susceptibility to diseases and response to drugs (Collins, 1997).

Although the variations can be in the form of insertions, deletions and copy number of repeated sequences, most are in the form of SNPs. A SNP usually has two alleles, where the rare allele frequency is $\geq$ 1%. It is a single base pair mutation or variation at a specific position. The genetic code is specified by the four nucleotide "letters": A

(adenine), C (cytosine), G (guanine) and T (thymine). A SNP occurs when a single nucleotide, such as an A, replaces one of the other three nucleotide letters (C, G, or T). Deleterious SNPs are often found to be the aetiology of many human diseases (Yue & Moult, 2006).

On average, SNPs occur in at least 1% of the population, and make up about 90% of all human genetic variation. Therefore, the 3-billion-base human genome probably harbours about 10 million SNPs (Sachidanandam, et al., 2001; Yue & Moult, 2006).

1.3.1.2 Types of SNPs

SNPs may fall within coding sequences of genes, non-coding regions of genes, or in the intergenic regions between genes (Vignal, et al., 2002).

- **Coding SNPs**

SNPs within a coding sequence may not necessarily change the amino acid sequence of the protein that is produced, due to degeneracy of the genetic code. A SNP in which both forms lead to the same polypeptide sequence is termed *synonymous*. If a different polypeptide sequence is produced, they are *non-synonymous*. A non-synonymous change may either be missense or nonsense, where a missense change results in a different amino acid while a nonsense change results in a premature stop codon. SNPs found within a coding sequence are of particular interest to researchers because they are more likely to alter the biological function of a protein. SNP discovery and detection are the current trend because of the unique characteristics of these genetic variations in facilitating gene identification.

Non-coding SNPs

Most SNPs are found outside "coding sequences" because only about 3 to 5 percent of a person's DNA sequence codes for proteins, SNPs that are not in protein-coding regions may still have consequences for gene splicing, transcription factor binding, or the sequence of non-coding RNA.

1.3.1.3 Databases

As there are for genes, there are also bioinformatics databases for SNPs. *dbSNP* is a SNP database in the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). The *OMIM* (http://www.ncbi.nlm.nih.gov/omim/) database describes the association between polymorphisms and diseases in text form while HGVbaseG2P (http://www.hgvbaseg2p.org/index) allows users to visually interrogate the actual summary-level association data. The International HapMap Project concentrated on making genetic maps and estimated the frequencies of variants across the human genome in different human populations. In the latest phase - HapMap 3 - the researchers looked for variants across the genome in 1184 samples from 11 populations. They chose the large sample set and the wide range of populations to maximize the variation they could capture. The project includes both SNPs as well as copy number polymorphisms. The HapMap data set is available at (http://www.hapmap.org).

1.3.1.4 Use and importance of SNPs

Over the past 2 decades, more than 4 million SNPs have been identified. Therefore, SNPs are the most abundant markers (with an approximate frequency of one in every kilobase of sequence). Because SNPs are evolutionarily conserved, they are stable genetic markers well suited for use in QTL analysis and in association studies in place of MSs (Kruglyak, 1997). Technologies for genotyping these biallelic variants are relatively easier to be automated because SNPs can be screened in a digital format by analysing the presence or absence of a sequence when compared with other genetic markers like minisatellites and MSs. To take advantage of this resource, typing of identified SNPs with rapidly emerging novel genotyping strategies is the key to minimizing the cost and to increasing throughput of DNA genotyping.

1.3.1.5 SNPs and common diseases

Many common diseases in humans are not caused by a genetic variation within a single gene but are influenced by complex interactions among multiple genes (the effects of any one gene might be small) as well as environmental and lifestyle factors. Although both environmental and lifestyle factors add tremendously to the uncertainty of developing a disease, it is currently difficult to measure and evaluate their overall effect on a disease process. Therefore, we refer here mainly to a person's genetic predisposition, or the potential of an individual to develop a disease based on genes and hereditary factors.

According to the hypothesis of “Common Disease, Common Variant (CDCV)”, genetic variations with frequency greater than 1% in the population but low penetrance are the main contributors to genetic susceptibility in common diseases

(Schork, et al., 2009). Genetic factors may also confer susceptibility or resistance to a disease and determine the severity or progression of disease. Researchers have found it difficult to develop screening tests for most diseases such as diabetes because we do not yet know all of the factors involved in these intricate pathways. Standard linkage analysis using large pedigrees has only limited power to detect such small effects in these disorders (Risch & Merikangas, 1996). Association studies using unrelated cases and controls, or using smaller family groups such as sibling pairs or 'two parents and affected child' trios have been proposed to be more likely to detect these small effects. Quantitative analysis and mathematical modeling have suggested that genome-wide association studies using SNPs are more effective than linkage analysis for identifying complex disease genes (Kruglyak, 1997).

By studying stretches of DNA that have been found to harbour a SNP associated with a disease trait, researchers may begin to reveal relevant genes associated with a disease. Defining and understanding the role of genetic factors in disease will also allow researchers to better evaluate the role of non-genetic factors (such as behaviour, diet, lifestyle and physical activity) in disease.

1.3.1.6 SNPs - Haplotype blocks and linkage disequilibrium

Each person's genetic material contains a unique SNP pattern that is made up of many different genetic variations. Researchers have found that about 10 million SNPs common in human populations are not inherited independently. Rather, sets of adjacent SNPs are present on the same chromosome with a specific combination of alleles in a block pattern, so called haplotype. The term 'haplotype', a contraction of haploid genotype, was coined by the HLA geneticist Ruggero Ceppellini in 1967 to

apply to haploid (single chromosome) combinations of human leukocyte antigen coding locus alleles. Many haplotype blocks in humans have been transmitted through many generations without recombination. This means that although a block may contain many SNPs, it takes a few SNPs to identify or tag each haplotype in the block. By studying SNP profiles or haplotypes associated with a disease trait, researchers may begin to reveal relevant genes associated with a disease.

Researchers have shown that most SNPs are not responsible for a disease state because of lack of direct effect on the function of the gene in which they are located. Instead, they serve as biological markers for locating a disease gene on the human genome map because they are usually near a gene found to be associated with a certain disease. A SNP may be in linkage disequilibrium (LD) with the 'true' functional variant. LD is generally defined as a measure of the degree of association between two genetic markers and can be used to identify regions of the genome that are associated with the disease. LD, also known as allelic association, is a population phenomenon and refers to the non-random association of alleles at different loci in the gametes of the population. Occasionally, a SNP may actually cause a disease, result in an amino acid change or alter exon–intron splicing, thereby directly modifying the relevant protein. It may also exist in a regulatory region, altering the level of expression or the stability of the mRNA.

To date, most studies have used a linkage approach with the more informative and polymorphic MS markers for the genome-wide screen and followed this up with an LD mapping approach in the smaller linked region or around specific candidate genes. This approach substantially reduces the number of SNPs that need to be

analysed by focusing on those SNPs in and around selected candidate genes.

Therefore, SNPs can be used to search for and isolate the disease-causing gene. Then, it will only be a matter of time before physicians can screen individuals for susceptibility to a disease just by analysing their DNA samples for specific SNP patterns.

1.3.1.7 Genotyping of SNPs

Association studies with SNPs typically use genomic DNA from hundreds of individuals and numerous SNPs. The development of high-throughput technologies has been vital to the widespread use of SNPs in research and industry. The spectra of methods currently available for genotyping SNPs in individual samples can be divided into four classes (Shi, et al., 1999; Syvanen, 2001).

1) Conformation-based mutation scanning methods based on physical properties of DNA after amplification

- ***Single-strand conformation polymorphism (SSCP) analysis*** is one of the most widely used methods for mutation detection. In SSCP analysis, DNA regions with potential polymorphisms are first amplified by polymerase chain reaction (PCR). Single-stranded DNAs are then generated by denaturation of the PCR products and separated on a nondenaturing polyacrylamide gel (Costabile, et al., 2006). A fragment with a single base change generally forms a different conformer and migrates differently when compared with wild-type DNA **(Figure 1.1)** (Orita, et al., 1989). The sensitivity can be increased to nearly 100% by dideoxy-sequencing

fingerprinting (Sarkar, et al., 1992).

The advantages of this method are its simplicity and relatively low costs. The disadvantages are low throughput, difficulty of optimization and restriction on the fragment length (Myers et al, 1987). The sensitivity drops when sequences longer than 400 bp are used.

- ***Denaturing high performance liquid chromatography*** (***DHPLC)*** relies on the melting characteristics of double-stranded DNA, which are largely defined by the DNA sequence (O'Donovan, et al., 1998). Therefore, a single-base mismatch can produce conformational changes in the double helix that cause the differential migration of homoduplexes and heteroduplexes containing base mismatches during passage through a separation matrix packed in a column maintained at high temperature **(Figure 1.2)**. This method has been shown to be highly sensitive for identifying mutations in areas of highly repetitive and GC-rich sequences (Korkko, et al., 1998). The DHPLC method is relatively fast, highly specific, and is easily automated because no labeling of the DNA fragments is needed. Reactions from different samples can be pooled into a single reaction for variant detection **(Figure 1.3)**. Therefore, it can substantially increase the throughput when compared with DNA sequencing. One major drawback of DHPLC is that the column temperature must be optimized for each target in order to achieve the right degree of denaturation for multiple samples.

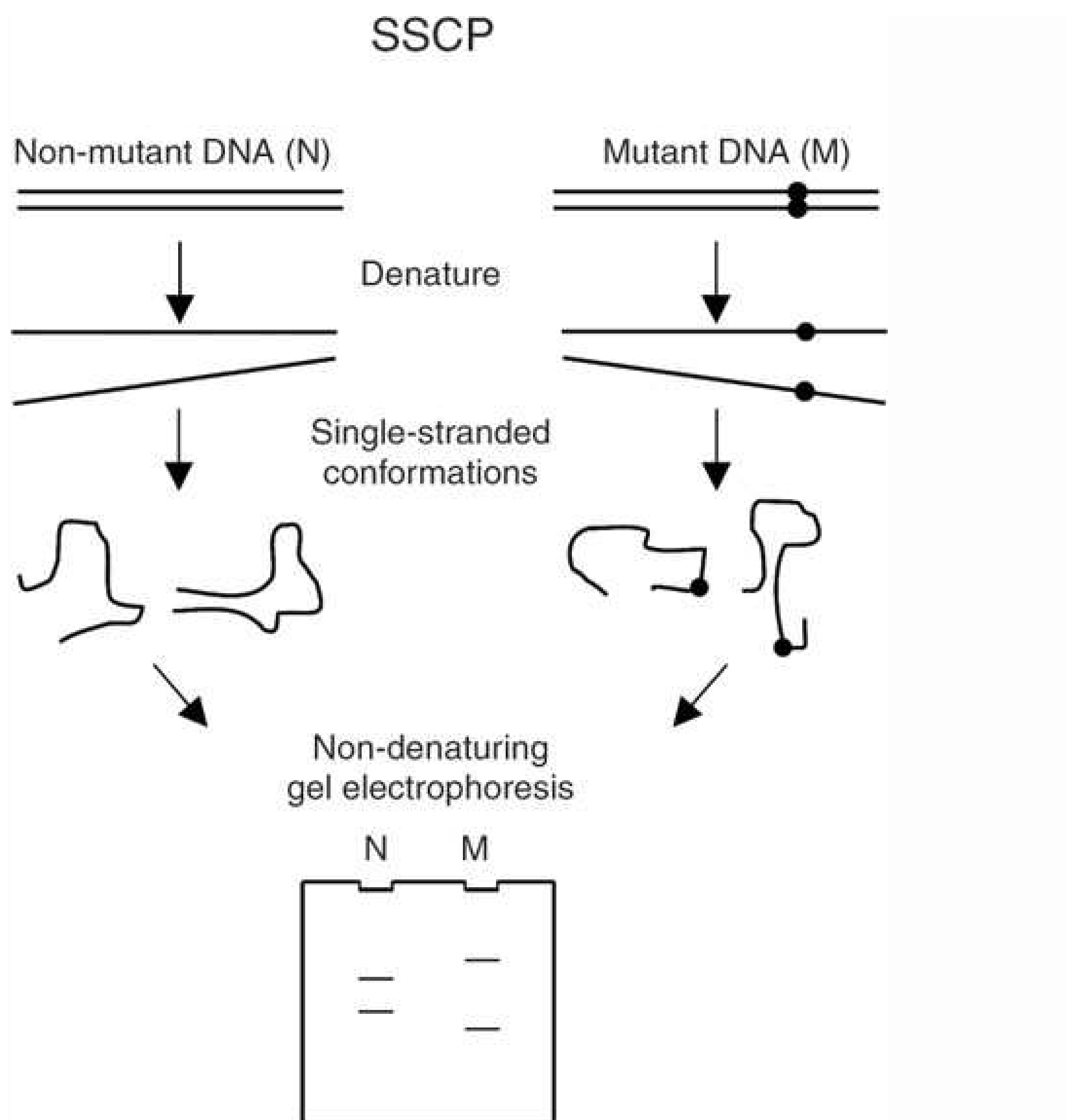


Figure 1.1
The principle of PCR-based SSCP analysis (Gasser & Zhu, 1999)

A point mutation (represented by a dot on a DNA strand) leads to the formation of different single-strand conformations of the mutant DNA (M) when compared with the non-mutant molecule (N), resulting in differential mobility in a non-denaturing gel matrix.

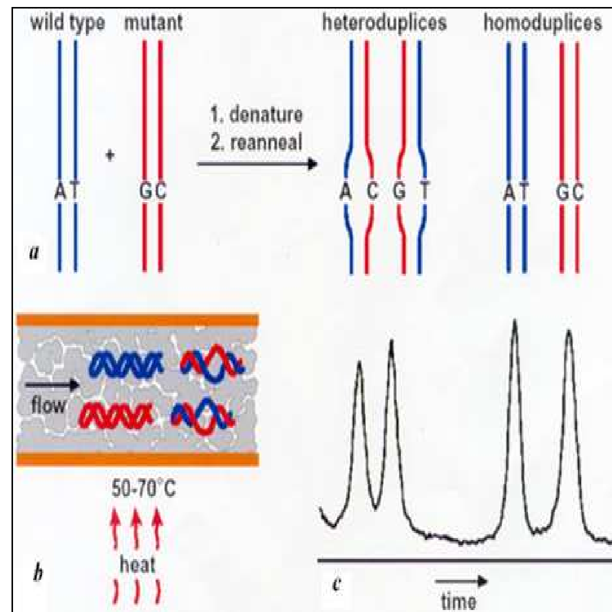


Figure 1.2
The principle of DHPLC with heteroduplex formation (Huber, et al., 2001)

Wild-type and mutant PCR products are heated to denature DNA duplexes and then allowed to cool slowly. The result is a mixed population of the original homoduplexes plus heteroduplexes containing the mismatched bases.

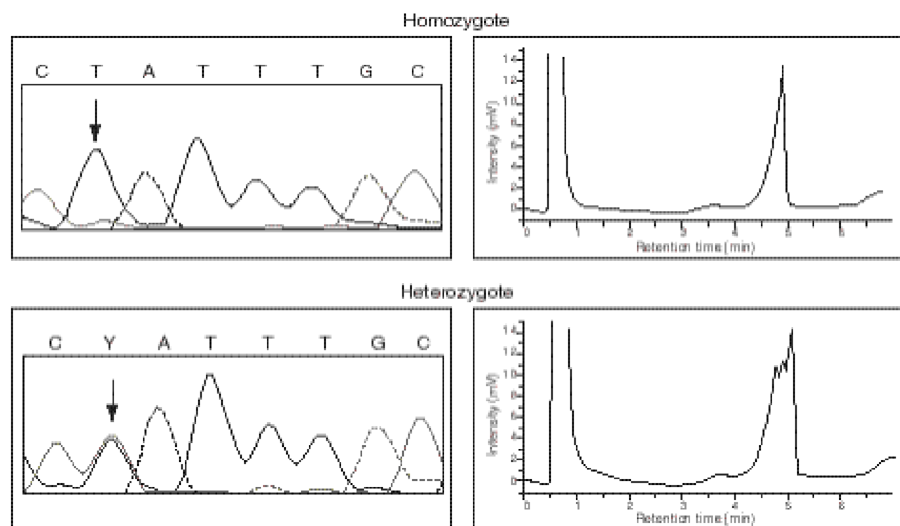


Figure 1.3
Identifying heterozygosity by DHPLC (Bonner & Ballard, 1999).

The PCR product from a heterozygote naturally forms heteroduplexes because of the sequence variation of each allele. This heterozygosity is easily identified by DHPLC when the PCR product is analysed under partially denaturing temperatures.

2) Hybridization genotyping methods

Several methods that interrogate known polymorphisms are based on hybridization, amplification or ligation in an allele-specific manner. The challenge of this approach is reducing cross-hybridization between the allele-specific probes. This challenge is generally overcome by manipulating the hybridization stringency conditions (Rapley & Harbron, 2004).

➢ ***Allele specific hybridization***

Also known as allele specific oligonucleotide hybridization (ASO), this protocol relies on distinguishing between two DNA molecules differing by one base by hybridization (You, et al., 2008). Fluorescence-labelled PCR fragments are applied to immobilized oligonucleotides representing SNP sequences. After stringent hybridization and washing conditions, fluorescence intensity is measured for each SNP oligonucleotide. **(Figure 1.4)** Because ASO genotyping is measuring a quantifiable change in Tm, it is capable of measuring all types of mutations, not just SNPs. Other benefits of ASO include its ability to work with label-free probes and its simple design and performance conditions.

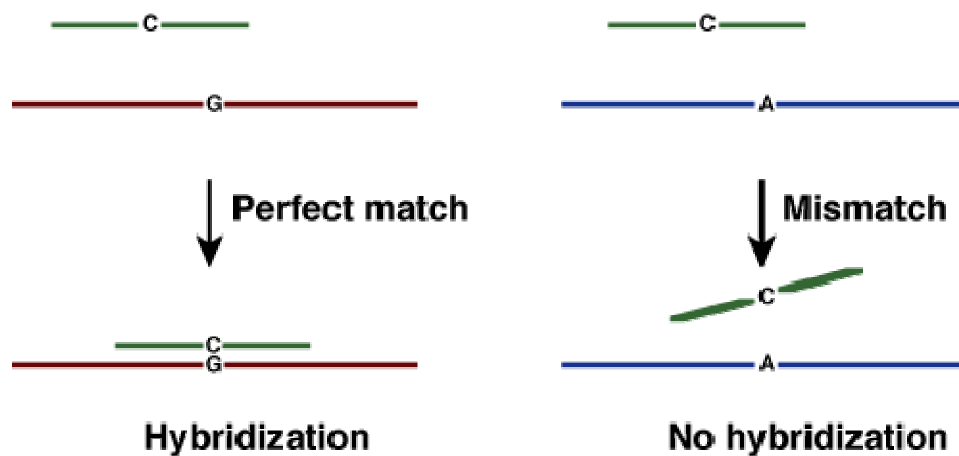


Figure 1.4
The principle of allele specific oligonucleotide hybridization (You et al., 2008)

The target DNA is generally obtained using the polymerase chain reaction and specific primers. Allele-specific oligonucleotides are then used to detect single base changes in the DNA samples. Typically, target DNA is immobilized on a solid support and denatured. Labelled (radioactive or fluorescent) oligonucleotides are then allowed to anneal. Complementary sequences bind while non-complementary sequences do not. Sequences that match the oligonucleotide are detected by fluorescence or by exposure to X-ray film when the oligonucleotide is radiolabelled.

- ***Oligonucleotide ligation assay (OLA) genotyping***

The OLA relies on hybridization with specific oligonucleotide probes that can effectively discriminate between the wild-type and variant sequences. Three oligonucleotides are used in OLA: two allele-specific oligonucleotide probes (one specific for the wild-type allele and the other specific for the mutant allele) plus a fluorescent common probe. The 3' ends of the allele-specific probes are immediately adjacent to the 5' end of the common probe. The gene fragment containing the polymorphic site is amplified by PCR and incubated with the probes. In the presence of thermally stable DNA ligase, ligation of the fluorescence-labelled probe to the allele-specific probe(s) occurs only when there is a perfect match between the variant or the wild-type probe and the PCR product template. **(Figure 1.5)** The ligation products are then separated by electrophoresis, which permits the recognition of the wild-type genotypes, the variants, the heterozygotes, and the unligated probes. By varying the combination of color dyes and probe lengths, multiple mutations can be detected in a single reaction (Baron, et al., 1996).

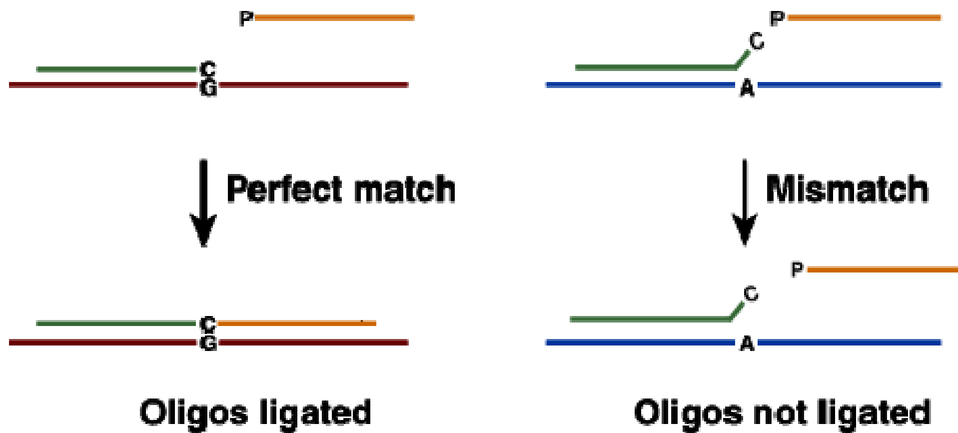


Figure 1.5
Principle of OLA-based SNP genotyping (Tang, et al., 2005).

By designing oligonucleotides complementary to the target sequence, with the allele-specific base at its 3'-end or 5-'end, one can determine the genotype of the PCR amplified target sequence by determining whether an oligonucleotide complementary to the DNA sequencing adjoining the polymorphic site is ligated to the allele-specific oligonucleotide or not.

➢ ***Microarray genotyping***

The DNA microarray is a hybridization-based genotyping platform that offers simultaneous analysis of many polymorphisms. High-density microarrays are created by attaching hundreds of thousands of oligonucleotides to a solid silicon surface in an ordered array **(Figure 1.6)**. The DNA sample of interest is amplified to incorporate fluorescently labelled nucleotides and then hybridized to the array. Each oligonucleotide in the high-density array acts as an allele-specific probe. Perfectly matched sequences hybridize more efficiently to their corresponding oligomers on the array and, therefore, give stronger fluorescent signals than mismatched probe-target combinations (Pollak, et al., 2001). The hybridization signals are quantified by high-resolution fluorescence scanning and analysed by computer software. DNA alterations such as heterozygous base-pair polymorphisms or mutations, insertions, and deletions can be identified (Chee, et al., 1996). Although oligonucleotide microarrays have a comparatively lower specificity and sensitivity, the large number of SNPs that can be interrogated in a single experiment is a major benefit. Affymetrix and Illumina make quartz chips for DNA microarrays, which used to determine genotypes of DNA samples (or gene expression levels in RNA samples). Running thousands of screens in a few seconds for each chip allows high speed analysis although this can still be too costly for some small laboratories. In recent years, the price decreases substantially as more competitive platforms appear in the market (Kaller, et al., 2007).

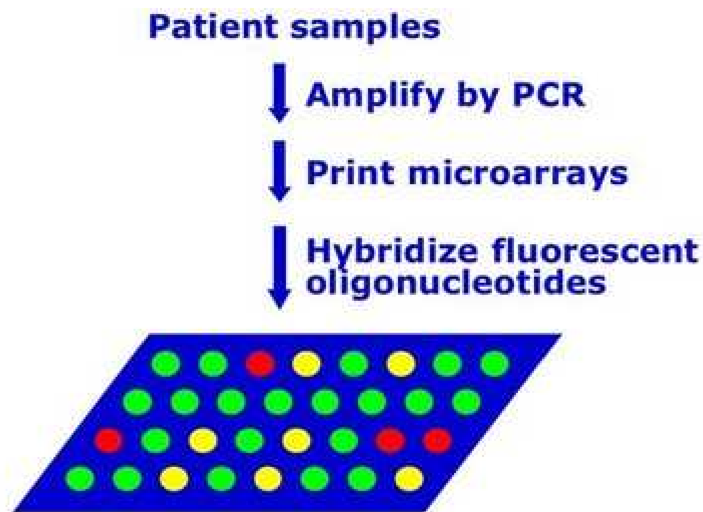


Figure 1.6
Microarray-based platform for genetic screening (Khan, 2004).

Specific chromosomal loci are amplified by use of the polymerase chain reaction, printed into microarrays, and hybridized with fluorescent oligonucleotides. The fluorescent microarrays are then scanned for specific fluorescence emission and signal strengths provide genotyping information.

3) Enzyme-based methods

A broad range of enzymes including DNA ligase, DNA polymerase and nucleases have been employed to generate high-fidelity SNP genotyping methods

➢ ***Restriction fragment length polymorphism (RFLP)***

It is considered to be the earliest and simplest method to detect SNPs. SNP-RFLP makes use of the many different restriction endonucleases and their high affinity to unique and specific restriction sites **(Figure 1.7)**. By performing a digestion on a genomic or PCR-amplified sample and determining fragment lengths through a gel assay, it is possible to ascertain whether or not the enzymes cut the expected restriction sites (Lange & Boehnke, 1983). A failure to cut the genomic sample results in an identifiably larger than expected fragment, implying that there is a mutation at the point of the restriction site, which protects it from nuclease activity.

Unfortunately, the combined factors of the high complexity of most eukaryotic genomes, the requirement for specific endonucleases, the fact that the exact mutation cannot be necessarily be resolved in a single experiment, and the slow nature of gel assays make RFLP a poor choice for high throughput analysis.

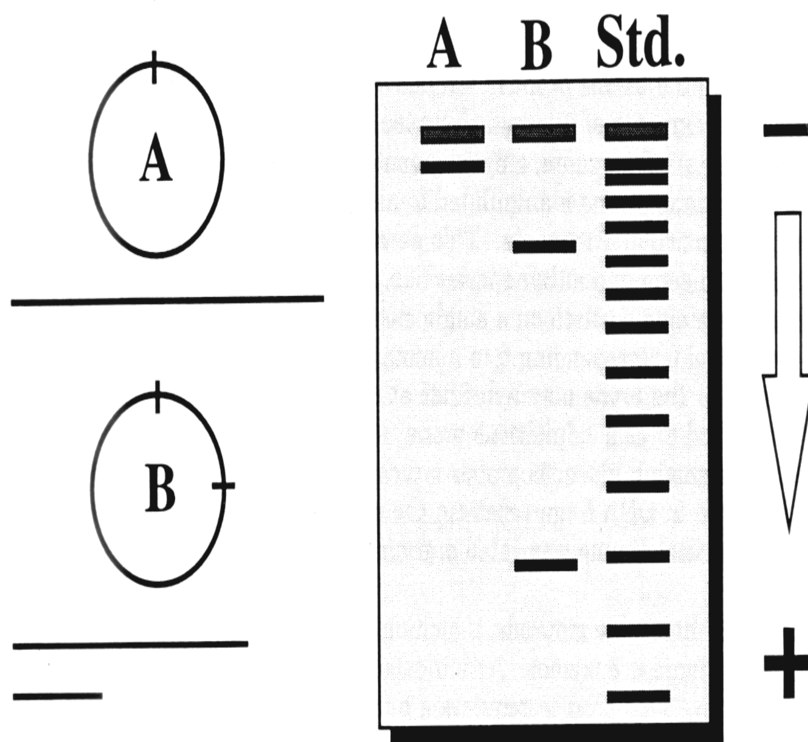


Figure 1.7
Restriction fragment length polymorphism-based platform for genetic screening (Lange & Boehnke, 1983).

Fragments "**A**" & "**B**" represent the genomes of two individuals. The perpendicular tick marks on the circles indicate restriction enzyme recognition sites. A base substitution in individual **A** has resulted in the loss of one recognition site relative to **B**, producing a single larger fragment rather than two smaller fragments. The restriction fragments are shown below the circles; electrophoretic separation of the restriction fragments is illustrated on the right. The number of bands on the gel indicates the number of cleavage sites. The slots at the top of the gel are wells into which the DNA samples are loaded. The distance the fragments have migrated from the well is a function of their relative size. In the third lane of the gel is a DNA size standard, which allows estimation of the fragment sizes.

- ***Primer extension (Minisequencing)***

In the single-base extension approach, the target region is amplified by PCR and then a single-base sequencing reaction performed using a primer that anneals one base next to the polymorphic site (Pastinen, et al., 1996; Syvanen, 2001). Several detection methods have been described. One can label the primer and apply the extension products to gel electrophoresis. Alternatively, the single base extension product can be measured by mass spectrometry or DHPLC. The most popular detection method involves fluorescence-labelled dideoxynucleotide terminators that stop the chain extension **(Figure 1.8)**.

It is an efficient way to detect SNPs through the addition of specific nucleotides to a single primer (Syvanen, 2001). To increase the throughput, flexibility and specificity, primer extension probes can be arrayed on slides so that many SNPs can be genotyped at once.

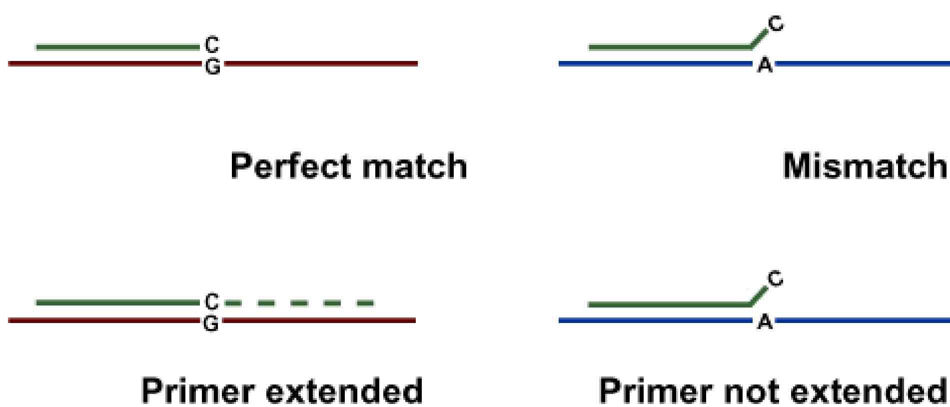


Figure 1.8
Allele-Specific Primer Extensions (Ericsson, et al., 2003).

In the single base extension approach, the target region is amplified by PCR followed by a single base sequencing reaction using a primer that anneals one base next to the polymorphic site. Several detection methods have been described. One can label the primer and apply the extension products to gel electrophoresis. The single base extension product can also be measured by mass spectrometry or DHPLC. The most popular detection method involves fluorescence-labelled dideoxynucleotide terminators that stop the chain extension.

- ***5'- nuclease***

The TaqMan allelic discrimination assay uses the 5' nuclease activity of *Taq* DNA polymerase to detect a fluorescent reporter signal generated during PCR (Livak, et al., 1995). For SNP genotyping, one pair of TaqMan probes and one pair of PCR primers are used. The assay uses two TaqMan probes that differ at the polymorphic site with one probe complementary to the wild-type allele and the other to the variant allele. A 5' reporter dye and a 3' quencher dye are covalently linked to the wild-type or variant allele probes. When the probes are intact, fluorescence is quenched because of the physical proximity of the reporter and quencher dyes. This is one example of fluorescence energy resonance transfer (FRET) and the energy transferred to the acceptor is quenched (i.e. dissipated as heat rather than emitted as photons). During the PCR annealing step, the TaqMan probes hybridize to the targeted polymorphic site. During the PCR extension phase, the 5' reporter dye is cleaved by the 5' nuclease activity of *Taq* DNA polymerase, leading to an increase in the characteristic fluorescence of the reporter dye. Specific genotyping is determined by measuring the signal intensity of the two different reporter dyes after the PCR. **(Figure 1.9)** In addition to detecting SNPs, small gene deletions and insertions can also be identified by this method (Shi et al., 1999). This assay can be multiplexed by combining the detection of up to seven SNPs in one reaction. However, since each SNP requires a distinct probe, the TaqMan assay is limited by how close the SNPs can be situated. The scale of the assay can be drastically increased by performing many simultaneous reactions in microtitre plates. Generally, it is limited to applications that involve interrogating a small number of SNPs since optimal probes and reaction conditions must be designed for each SNP (Syvanen, 2001).

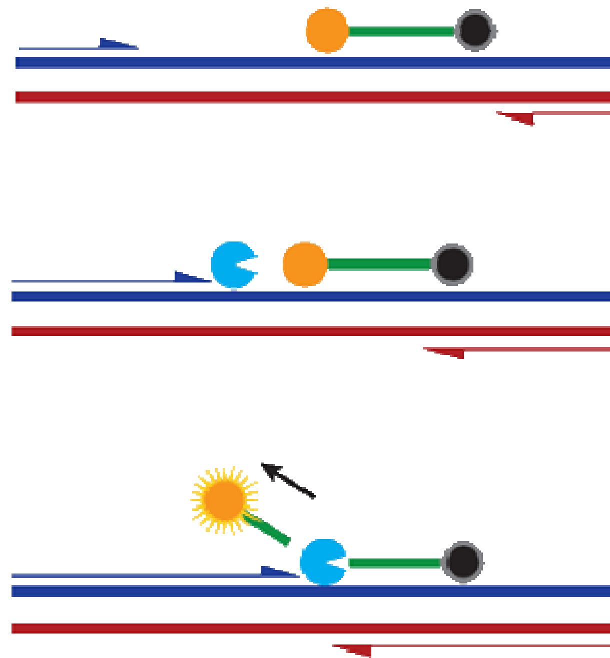


Figure 1.9
The TaqMan allelic discrimination assay (Ranade, et al., 2001)

Two TaqMan probes targeted at the polymorphic site are labelled with reporter dyes and a common quencher dye. The 5' nuclease activity of thermostable polymerases used in the PCR cleaves hydrolysis probes during the amplicon extension step. Only the perfectly hybridized probe will be cleaved. A mismatched probe will not be recognized by the Taq polymerase. Cleavage separates the detectable reporter fluorophore from a quencher. Fluorescence emitted when excited by an external light source at each PCR cycle is proportional to the amount of product formed.

- ***Flap endonuclease (FEN)***

The endonuclease catalyzes structure-specific cleavage. This cleavage is highly sensitive to mismatches and can be used to interrogate SNPs with a high degree of specificity (Olivier, 2005). In the basic Invader assay, a FEN called cleavase is used with two specific oligonucleotide probes that, together with the target DNA, can form a tripartite structure recognized by the cleavase (Kaiser et al., 1999). Invader assays are conducted isothermally, and a linear increase in signal over time will be produced (**Figure 1.10**). Each cleavage product then serves as an Invader oligonucleotide in a secondary reaction, in which it directs the cleavage of a combined labelled probe-template construct. This secondary oligonucleotide probe is 5' end-labelled with the donor fluorophore, which is quenched by an internal acceptor dye. When the DNA is cleaved, the donor and the acceptor dyes are no longer in close proximity, the quenching is abolished and fluorescence generated. Assays are read with a fluorescence plate reader, and genotypes are assigned after determination of the net wild-type/variant signal ratio for each sample (Olivier, 2005).

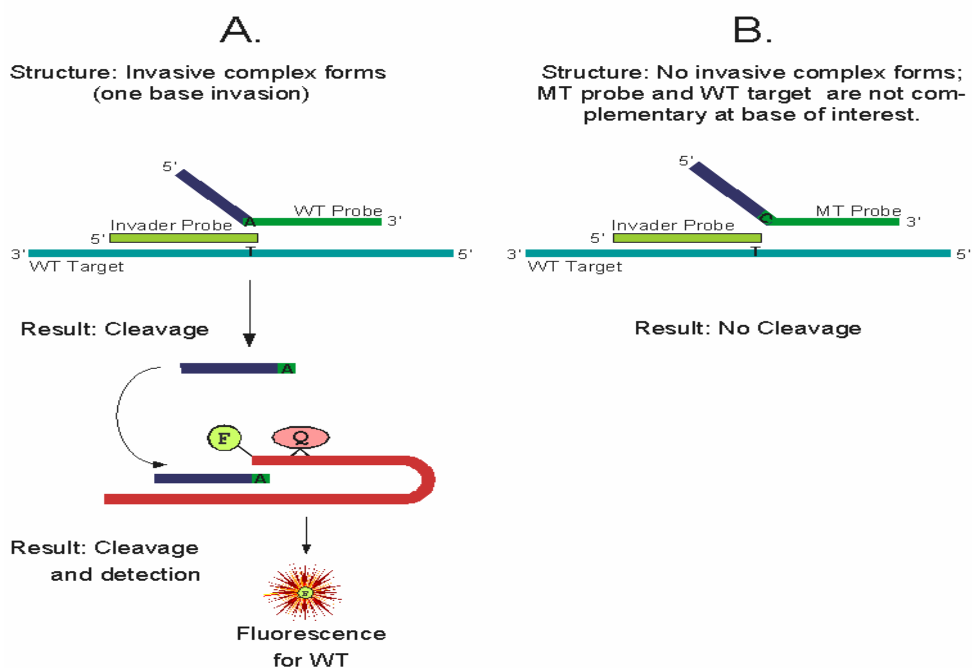


Figure 1.10
The principle of the Invader assay (Olivier, 2005)

F = fluorescent molecule; Q = quencher molecule.

- ***Mass spectrometry (MS)***

Large molecules can be identified by MS through electrospray or matrix-assisted laser desorption ionization, and ion-trap or time-of-flight detectors. MS yields precise information on the molecular mass of the DNA fragments. The procedure can be fully automated, and both DNA strands can be analysed in parallel (Ross, et al., 1998). Unlike the fluorescent genotyping methods described above, MS offers specificity and accuracy without requiring special labelled probes or primers. One disadvantage of this technique is that it requires purified samples free of ions and other impurities, thus increasing the technical time and sample-processing costs. A chip-based genotyping method using MS has been described (Tang, et al., 1999). PCR can be performed in 1-μL reactions directly in the chip wells in parallel. The PCR product can be detected *in situ* using matrix-assisted laser desorption ionization MS. This miniaturization technique has the potential for high-throughput, low-cost genotyping.

4) Sequencing

Sequencing is the procedure of choice for SNP discovery. Once the potential regions have been confirmed to contain putative polymorphisms, these regions can be sequenced to locate the final polymorphic sites. The most common forms of sequencing are based on primer extension using either a) dye-primers and unlabelled terminators or b) unlabelled primers and dye-terminators. The products of the reaction are then separated using electrophoresis in either capillaries or slab gels. With the improvement of computer software and detection systems, fluorescent DNA sequencing has become fully automated. The SOLiD (Applied Biosystems), Illumina (Solexa) and 454 Pyrosequencing GS FLX sequencer (Roche) systems are fully

automated non-capillary-based sequencers that are emerging as high-throughput sequencing technology platforms for large-scale DNA sequencing projects. High-throughput sequencing technologies lower the cost of DNA sequencing when compared with standard dye-terminator methods (Schuster, 2008). Compared to other SNP genotyping methods, sequencing is suited to identifying multiple SNPs in a small region, but not for genotyping known SNPs in a large number of samples.

Considerations for the ease of use:

The success of large-scale genotyping studies depends on user-friendly technologies that can detect polymorphisms rapidly and accurately on a large scale. Each of the methods discussed above has been demonstrated to work in a variety of settings, but each method has its limitations. These methods mainly rely on a PCR step to increase the concentration of a segment of DNA sequence carrying the SNP. Since assays based on TaqMan and molecular beacons incorporate allele-specific probes in the PCR, they combine the amplification and detection steps and require no post-PCR processing for determining genotypes for each reaction. In fact, fluorescence is merely measured after PCR and genotypes are inferred based on these values. The other methods, in contrast, require significant post-PCR processing. On the other hand, in the microarray method used by Wang et al. (1998), amplified products are purified to remove nucleotides, enzymes, primers, etc. These purified samples are then hybridized for 15 hours to oligonucleotides arrayed on chips and the genotypes are determined after several additional washing and developing steps. In some assays, such as the Invader assay, separate reactions are performed after amplification to distinguish the two alleles (Pielberg, et al., 2003). These separate reactions could potentially lead to errors because if one reaction fails and the other

succeeds, then a heterozygote could be misinterpreted as a homozygote. Thus, accurate and high-throughput genotyping method should not require processing of amplified products.

Of all the options available, FRET-based and chip-based genotyping technologies will most likely evolve as the ultra-high-throughput detection systems to meet the requirements of large-scale SNP genotyping. For DNA microarrays, a better approach is to fabricate a generic array containing tag sequences close to the polymorphic sites. Genotyping can be conducted by use of fluorescently labelled single-base extension reactions.

Methods based on hybridization or on physical–chemical properties are likely to be ruled out as each assay must be optimized. The method with minisequencing from a primer adjacent to the site of the SNP such as SNaPshot™ (Applied Biosystems) is robust and requires little optimization (Syvanen, 1990). However, multiplexing of SNP assays is less straightforward, as the signal strength varies between assays. SNP genotyping using primer extension followed by DHPLC requires extensive optimization of the primer extension reaction and does not allow easy multiplexing. Optimization of the gradient that is best suited to the elution of each product is also required and multiplex reactions are difficult to perform.

Cost consideration

The cost of genotyping is highly dependent on the ability to multiplex reactions and minimize reaction volume. For these reasons, primer extension followed by DHPLC is not as cost effective as the SNaPshot™ method. Methods such as Pyrosequencing,

TaqMan or bioluminometric assay coupled with modified primer extension reactions (BAMPER) that use modified primers are relatively expensive.

Determining criteria

Choosing a method for genotyping, particularly if this implies the purchase of expensive equipment, is difficult and no golden rule can be applied. The deciding factors include the number of genes/SNPs to be typed, the sample size, the need for single genotyping versus pool genotyping, the level of throughput required and whether there is a need for SNP discovery as well as genotyping.

Some critical factors for selecting protocols for SNP typing (Kwok, 2002) are discussed below.

1. Scope of genotyping

This would include the number of SNPs to be screened, and the number of samples to be tested, as well as how many genotyping projects are conducted at the same time. This will help determine if one needs a higher throughput format with relative ease to switch from one project to another.

2. Level of molecular biology expertise in the laboratory

Some assays are relatively simple to set up and perform, while others require considerable amount of experience in assay optimization and software support.

3. Funding for capital instrument

Capital investment for commercial SNP platforms can range from US$30K to over

US$500K. Typically, expensive sample processing automation is included in the high capital investment.

4. Consumable cost

Cost for consumables (reagents, plates, and pipette tips, etc) range from US $0.20 to US$5.00 per genotype. This may not sound like a lot. However, when the project involves screening 1000 SNPs in a population of 1000, it can cost up to US$5 million for a total of 1,000,000 genotypes for a single project.

Automation

Due to the vast numbers of genotypes that are being studied, automation of genotyping will play a key role in this field. Recently, the application of automation in genomics was extensively described in two reviews (Meldrum, 2000a; Meldrum, 2000b). Several SNP genotyping technologies have reached maturity in the last few years and are being integrated into large scale genotyping operations supported by automation. SNP genotyping methods are very diverse, as are automation solutions for them. Appropriate automation entirely depends on the method. On the other hand, the choice of method depends on the scale and the scientific question a project is trying to answer. A project might require genotyping of a limited number of SNPs in a large population or the analysis of a large number of SNP markers in one individual. Currently, there are few one-stop-shops for a high-throughput SNP genotyping process. All systems are combinations of products from different suppliers and are the results of alliances. Still most systems are custom products that are built around commercial elements to specification of individual projects.

Control of errors

Genotyping errors can be divided into two broad categories: operational errors (e.g., sample swaps, pipetting errors or DNA template contamination) and genotype scoring errors. Because of an increased use of robotic workstations, stringent quality control procedures, and optimized experimental conditions, the occurrence of operational errors has been greatly reduced for high-throughput genotyping technologies developed in recent years. In contrast, genotype scoring errors remain a significant challenge for automated scoring programmes. In circumstances when genotype clusters are not sufficiently separated, which can be caused by (1) wide variations in fluorescence signals for different subjects and (2) unbalanced amplifications of the two alternative alleles for heterozygotes, genotype scoring is typically performed manually. However, this is very time-consuming and error-prone (humans are likely to make errors due to fatigue or oversight when manual scoring becomes routine). Moreover, manual scoring rules are difficult to standardize, and different readers can have different subjective views (van den Oord, et al., 2003).

Conclusion

In conclusion, most common diseases are thought to result from a mixture of genetic and environmental risk factors. Therefore, the contribution of each gene is likely to be relatively small. Allelic association methods are more powerful in the detection of these genetic risk factors than conventional linkage approaches. However, allelic association methods require genetic markers to be very closely spaced because they rely on LD between the marker and the disease allele.

Generating human SNPs is no longer a bottleneck step for genetic studies of disease. Recent advances in molecular genetic technology including the identification of over a million of SNPs in the human genome and high throughput genotyping methods now make incredible opportunities for identifying association between genetic polymorphisms and disease-related phenotypes.

SNP genotypes are by nature biallelic, which makes them easy to call because they are relatively free of mistyping errors in the laboratory. They also occur at the high density essential for association mapping, which relies on LD among markers. High-throughput genotyping makes this a realistic and affordable mapping strategy. Choosing a method for genotyping, particularly if this implies the purchase of expensive equipment, is difficult and no golden rule can be applied. The deciding factors include the number of genes/SNPs to be typed, the need for single genotyping versus pool genotyping, the level of throughput required and whether there is a need for SNP detection as well as genotyping. New genotyping methods that are high-throughput, accurate and cheap are urgently needed for gaining full access to the abundant genetic variation of organisms.

Ultimately, we hope that in the coming future physicians can screen patients for susceptibility to a disease by analysing their DNA for specific SNP profiles.

1.3.2 Genetic mapping of common complex disease genes

Genetic epidemiology is the study of the role of genetic factors in determining disease in families and in populations, and the interplay of such genetic factors with environmental factors. Traditionally, the study of the role of genetics in disease progresses through the following study designs (Brennan, 1999; Rudan, et al., 1999):

- Familial aggregation studies
- Segregation studies
- Linkage studies
- Association studies

This traditional approach has proved highly successful in identifying monogenic disorders and locating the causal genes. More recently, the scope of genetic epidemiology has expanded to include common diseases for which many genes each play a smaller contribution (polygenic, multifactorial or multigenic disorders). This has developed rapidly in the first decade of the 21st century following the completion of the Human Genome Project, as advances in genotyping technology and associated reduction in cost have made it feasible to conduct large-scale genome-wide association studies that genotype many thousands of SNPs in thousands of individuals. These have led to the discovery of many genetic polymorphisms that influence the risk of developing common diseases.

1.3.2.1 The rationale of the population-based genetic association studies

The genetic models of common disorders tend to fall into 2 major models: either "common variants for a common disease" or "rare variants for a common disease"

(Iyengar & Elston, 2007). In the first model, there are a few relatively common variants that each confers some detectable degree of risk on the carriers. Association studies are a powerful tool for identifying such variants.

In contrast, the rare variant model is based on the concept that many independent mutations have arisen in one or more genes and all these variants contribute to the risk of developing disease within the population. In this situation, association study is not an appropriate approach because each variant contributes only a relatively small percentage of the cases. Without large numbers of cases and controls, one would not observe statistically significant differences.

However, if different variants within the same gene have cumulatively a major effect on disease risk, then family linkage studies may show positive signals, even in the absence of an allele-specific association. This is because linkage studies ask how often the same genetic material is shared by affected family members, regardless of which variant within that DNA is responsible in a given family (Polychronakos, 2008).

1.3.2.1.1 Introduction to genetic association

Genetic association studies are performed to determine whether a genetic variant is associated with a disease or trait. If association is present, a particular allele, genotype or haplotype of a polymorphism or polymorphism(s) will be seen more often than expected by chance in individuals carrying the trait. Thus, a person carrying one or two copies of a high-risk variant is at increased risk of developing the associated disease or having the associated trait.

A statistical association between genotypes at the marker locus and the phenotype is usually thought to imply physical linkage between the marker locus and a disease locus. They occur together in a population more often than can be readily explained by chance. This can be between phenotypes, e.g. visible characteristics such as flower color or height, between a phenotype and a genetic polymorphism such as a SNP, or between two genetic polymorphisms.

Association between genetic polymorphisms occurs when there is non-random association of their alleles as a result of their proximity on the same chromosome. This is known as genetic linkage.

1.3.2.2 Genetic dissection of common complex traits

Most common diseases that can be studied in human populations are complex genetic traits. In contrast to simple Mendelian traits that result from mutations in a single gene, sometimes with locus and allelic heterogeneity, several genes and variants within genes are expected for complex diseases (Buchanan, et al., 2009). Many genetic and non-genetic factors interact to determine the final phenotype.

Identifying the genes that underlie the population variation in these phenotypes has been challenging. Recently, databases of common genetic variants, recognition of the patterns of genetic variation, and rapid genotyping methodologies have emerged. The combination of these tools and resources will greatly facilitate genetic association studies, a potentially powerful method to map the genes for complex traits (Pawitan,

et al., 2009).

Two major approaches have been used to map genetic variants that influence disease risk: linkage analysis and association studies.

1.3.2.2.1 Linkage studies

In linkage analysis, a genome-wide set of markers spaced millions of bases apart is typed in families with multiple affected relatives (Menotti-Raymond, et al., 1999). Markers that segregate with diseases (or the traits) in relatives more often than expected are used to localise the disease genes. This approach has the advantage of being an unbiased, comprehensive search across the genome for susceptibility alleles, and has been successfully in identifying genes for simple (Mendelian) diseases such as Huntington disease and cystic fibrosis.

However, linkage analysis has been used less successfully for searching genes that are implicated in polygenic diseases or traits (Altmuller, et al., 2001; Hirschhorn & Daly, 2005), in part because of a limited power to detect the effect of common alleles with modest effects on disease, limiting the potential for early application to determine individual disease risk (Altmuller et al., 2001; Risch & Merikangas, 1996). This failure is the result of four main features of complex diseases:

- The variety in severity of symptoms and age of onsets, which results in difficulty in defining an appropriate phenotype and selecting the best population to study;
- The variety of possible etiologies responsible for disease, which might involve various biological pathways;

- Multiple causative genes contribute in small portions and cause small relative risk. The complex interplay of several physiological systems that regulate the final clinical manifestation of disease; and
- The interaction between environment and disease.

Because of these features, attention has turned to association studies in the expectation that they might be more effective in identifying genes that are involved in complex diseases. Association studies look for a particular marker to be statistically correlated with disease (or trait values) across a population. The basic approach of association studies is to compare genotype (or allele) frequencies between cases and controls (Risch & Merikangas, 1996). These studies have much greater power than linkage analysis in detecting several genes of small effect (Risch, 1990; Risch & Merikangas, 1996). However, association studies require more markers and samples than linkage analysis (Carlson, et al., 2004). Thus, association studies are usually limited to candidate genes or regions because of the expense of a genome-wide approach.

1.3.2.2.2 Association studies

The candidate gene approach is defined as the study of the genetic influences on a complex trait. To identify candidate genes that might have a role in the aetiology of the disease, the following steps are involved: identifying variants in or near those genes that might either cause a change in the protein or its expression, or be in LD with functional changes; genotyping the variants in a population; and by using statistical methods to determine whether there is a correlation between those variants and the phenotype. A prominent example of the 'candidate gene' approach was the

search for genes causing severely high or low blood pressure (Botstein & Risch, 2003).

- **Case-control designs**

The design is a widely used and powerful approach for genetic association studies (Risch, 2000). Genotype frequencies are compared between case and control samples to identify candidate genes or nearby markers that are associated with the susceptibility to a disease. Although association studies may be subject to the possibility of population stratification, it has been recognized that this effect is small in magnitude in well designed studies that sample controls and cases from a homogeneous population, or that match cases and controls for the major confounding variables such as age, gender, and race-ethnicity (Risch, 2000).

- **Family-based designs**

To avoid the false-positive results produced by population stratification, family-based methods, e.g., the transmission-disequilibrium test, discordant sib pairs and affected family-based controls, were suggested to evaluate association between genetic markers and disease status (Laird & Lange, 2006). Unlike traditional case-control studies in which all individuals are unrelated, cases from the same family are often correlated because these individuals share genetic and environmental conditions. The parents are used as controls for the case, which is their affected offspring. Consequently, the frequency of risk alleles at a marker locus is usually increased among related cases relative to unrelated cases. Using related cases sampled from families or ascertained from family linkage studies and related controls may have fewer false-positive results produced by population stratification.

However, the power of such methods may be lower than the traditional case-control design based on independent samples (Tang et al., 2008).

- **Selecting candidate genes for myopia**

Although candidate gene lists will rarely include all relevant genes, it is also true that some parts of the genome are better places to start searching than others. Although candidate genes for disease susceptibility are less obvious, there are many useful pointers:

1) The first group of candidate genes appears from multiple genetic syndromes with systemic findings that have myopia as a consistent clinical feature.

For example, Stickler syndrome is an autosomal dominant connective tissue disorder characterised by ocular, orofacial, and skeletal abnormalities. Associated ocular manifestations include high myopia, glaucoma, cataract, vitreoretinal degeneration, and retinal detachment (Richards, et al., 1996). Marfan syndrome is an autosomal dominant disorder with clinical features of myopia, lens dislocation, tall body stature, and increased aortic wall distensibility (Nijbroek, et al., 1995). Knobloch syndrome has an autosomal recessive high myopia presentation with vitreous degeneration and encephalocele (Sertie, et al., 2000). It is reasonable to assume that genes that have severe mutations causing Mendelian forms might also harbour less severe sequence changes that predispose to disease. These Mendelian genes constitute high priority candidate genes for the myopic conditions.

2) A second group of candidate genes often originates from ideas concerning the

biology of the conditions, based on animal or cellular models, or the pathophysiology of the condition.

For example, the sclera, the tough outer coat of the eye, is a typical connective tissue that provides the structural framework for the eye. As the sclera defines the shape of the eye, it is also likely to determine the axial length of the eyeball. The ECM of the sclera has been shown to contain collagen fibrils in close association with proteoglycans and glycoproteins (Chakravarti et al., 2003). Alterations in any of these ECM components are likely to lead to changes in eye shape. Recent studies have shown that the scleral ECM undergoes significant changes during growth and aging (Rada, et al., 2000) and is dramatically altered during the development of myopia (Norton & Rada, 1995). The sclera of highly myopic human eyes differs considerably from normal sclera in both its physical dimensions and its biomechanical properties (Avetisov, et al., 1983). Many of the pathological changes seen in highly myopic human eyes are a consequence of gross scleral thinning, particularly at the posterior pole of the eye.

Genes responsible for several syndromic genetic disorders with myopia as a consistent clinical finding have been identified: *COL2A1* and *COL11A1* for Stickler syndromes type 1 and 2, respectively (Richards et al., 1996), lysyl-protocollagen hydroxylase for type 4 Ehler-Danlos syndrome (Heikkinen, et al., 1999; Pousi, et al., 1994), collagen 18A1 for Knobloch syndrome (Sertie et al., 2000), and fibrillin for Marfan syndrome (Dietz, et al., 1991). Each of these genes is expressed in the sclera, demonstrating how knowledge of gene expression in the scleral wall is critical to of the developmentof the eyeball elongation and myopia.

Therefore, we might make a hypothesis that the non-syndromic myopias result from developmental defects of scleral wall growth control, and that their causative genes may be functionally or structurally related to one another and have parallel functions in the development of the eyeball elongation. Moreover, genes expressed by the human sclera and examined using both complementary DNA (cDNA) library and microarray techniques have been identified to aid in the selection of candidate genes for high myopia.

Eye1 and ***Eye2*** are the loci known to control normal variation in eye size in mice (Zhou & Williams, 1999). One strong candidate gene for ***Eye1*** is the hepatic growth factor gene (Hgf), a potent mitogen expressed in the retina, pigment epithelium, and choroid. The human homolog of ***Eye2*** should map to chromosome 6p, 16q13.3, or 19q13, whereas that of ***Eye1*** should map to 7q.

3) A third group of candidates emerges from linkage studies.

The linkage analysis might provide information about genomic regions that can be explored further. All genes that map within myopia loci region might be candidate disease genes based on position. Therefore, genes likely to influence gene expression or function can then be prioritized. For the Myopia-2 locus *MYP2* mapped by linkage analysis, multiple candidate genes have been identified within this critical region (Young et al., 2004).

Molecular genetic studies of *MYP2*

Recent family linkage analysis and transmission disequilibrium test have identified a genomic region on chromosome 18p11. The core region extending from markers D18S481 to D18S52 **(Figure 1.11)** showed linkage with high myopia in seven high-grade myopia families. Across the 7.6 cM genomic region on the short arm of the chromosome 18, the first genetic linkage study of non-syndromic autosomal dominant high myopia showed linkage at 18p11.31 by haplotype analysis with a maximum cumulative logarithm of the odds (LOD) score of 9.59 (Young et al., 1998b). Refinement of the region by transmission disequilibrium testing suggests that candidate genes for this locus named myopia 2 (*MYP2*) (OMIM No. 160700) is likely to be in an interval between markers D18S63 and D18S52 (Young et al., 2001).

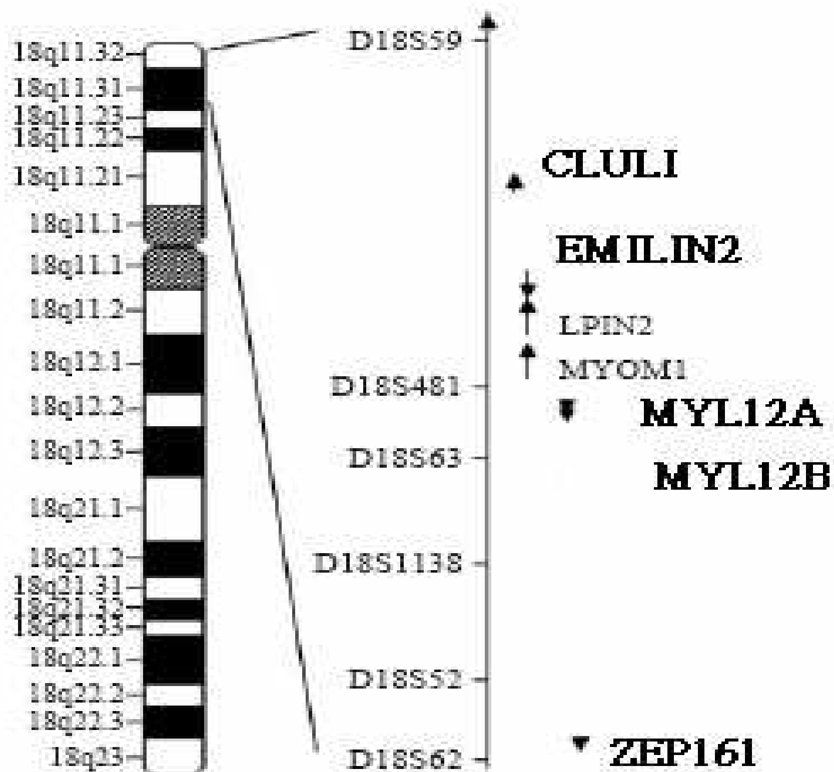


Figure 1.11
The genomic localization of selected candidate genes within the *MYP2* interval on chromosome 18p11.31 based on Homo sapiens genome view build 34 versions 3 (Scavello et al., 2005).

Flanking markers of the *MYP2* region are D18S59 and D18S62. Markers D18S63 and D18S52 flank the contracted *MYP2* interval defined by transmission disequilibrium testing. Arrows point in the directions of gene transcription.

This locus has been repeatedly detected in different population samples even with different disease models and analytical methods (Heath, et al., 2001; Lam, et al., 2003b; Naiglin et al., 1999). One replication study used a Hong Kong cohort of 6 myopic families, finding linkage between markers D18S476 and D18S62 (Lam et al., 2003b). Another replication study was a Sardinian Italian cohort study of 15 myopic individuals with a genetically homogeneous background, showing strongest linkage at marker D18S63 (Heath et al., 2001). However, there are a few other studies reporting negative results (Chang, et al., 2008; Farbrother, et al., 2004b; Ibay et al., 2004; Li et al., 2009; Naiglin et al., 1999; Yamane, et al., 2007). Some may represent false-positive association signals, while others may be specific to the manner in which the myopia was defined and/or analysed. Underpowered studies are one of the main reasons for failure to replicate an initial linkage study. The replication of linkage signals requires a considerably larger sample size than the original cohort used for the initial discovery, and nearly all of the present linkage studies are insufficiently powered to test all of the reported loci, some of which may have only a limited contribution to the risk of developing myopia.

Candidate genes within the *MYP2* critical region

All genes that map within the *MYP2* critical region are candidate disease genes based on position. There are 9 known and 20 hypothetical genes that map within the 2.2 cM interval.

The ECM genes should be prioritized for myopia study because the high myopia phenotype uniformly involves the axially elongated eyeball and thinning scleral wall. Two candidate genes, the *LAMA1* (alpha subunit of laminin) gene and the

transforming growth beta-induced factor (*TGIF*) gene, have been shown with no association with high myopia in subsequent association studies (Pertile et al., 2008; Scavello et al., 2004).

No coding mutation sequences were found with *TGIF*, *LOC91941* and *PRPH* gene in 180 Han Chinese patients and 60 controls by using exon-by-exon PCR-heteroduplex-SSCP analysis and sequencing (Li, et al., 2003).

However the remaining seven positional biologically plausible candidate genes should be prioritized: clusterin-like 1 (*CLUL1*) (Zhang, et al., 2003), elastin-microfibril located interface protein (*EMILIN-2*) (Doliana, et al., 2001), lipin 2 (*LPIN2*) (Scavello et al., 2005; Zhou & Young, 2005), myomesin 1 (*MYOM1*) (Wiesen, et al., 2007), myosin, light chain 12A, regulatory (*MYL12A),* myosin, light chain 12B, regulatory (*MYL12B)* (Redowicz, 2002; Satpathy, et al., 2004), and zinc finger protein 161 homolog (*ZFP161*) (Wang, et al., 2004). It is valuable to test whether reported associations within these genetic intervals actually account for the linkage evidence that has been reported.

We hypothesised that the *MYP2* locus also contain common genetic variants predisposing to high myopia, and conducted a case-control study using 62 SNPs located primarily in 7 positional candidate genes within the locus region.

Selection of candidate genes within *MYP2* in pooling

The aim of studying positional candidate genes within the *MYP2* locus is to provide insight into the molecular basis of eye growth and myopia development.

Clusterin-like 1 (*CLUL1*)

CLUL1 produces clusterin family glycoprotein that is predominantly expressed in the cone photoreceptors retina (Zhang et al, 2003). The ~53 kb human *CLUL1* gene is encoded by 10 exons **(Figure 1.12)**. *CLUL1* expression is down-regulated in some forms of retinal disease and thought to be involved in a defense response against neuronal damage (Sturgill, et al., 2006).

Elastin microfibril interface located Protein *(EMILIN-2)*

An elastic fiber-associated glycoprotein gene is essential for constituent organization and maintenance of connective tissue. The glycoprotein was found at the interface between amorphous elastin and microfibrils and regulates elastic fiber formation (Doliana et al., 2001). It is expressed in eye tissues based on the RT-PCR studies (Young, 2004). The *EMILIN-2* gene is located between the markers D18S476 and D18S481 right upstream of the *KIAA0249* gene, and close to the *LAMA1* gene coding for the laminin α1 chain. The genomic structure for *EMILIN-2* is shown in **Figure 1.13.**

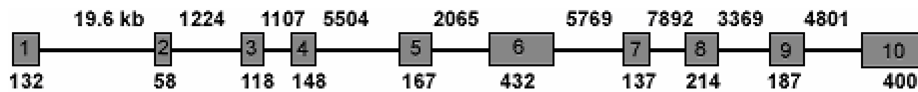


Figure1.12
A schematic diagram of the Clusterin-like 1 (*CLUL1*) gene (Scavello et al., 2005).

Exons, represented as squares/rectangles, are numbered from 1 to 10, with respective base pair size listed directly below. Intronic size in base pairs, unless stated otherwise, is listed between exons.

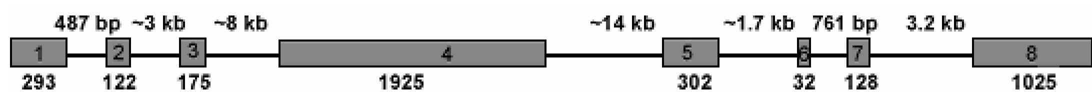


Figure 1.13
A schematic diagram of the *EMILIN2* gene (Scavello et al., 2005).

Exons, represented as boxes, are numbered from 1 to 8. Numbers below the various exons and above introns refer to their length in base pairs, unless stated otherwise.

Lipin 2 (*LPIN2*)

It is located in proximity to the *MYP2* contracted interval near marker D18S481 which is located from base pair 2,916,991 to base pair 3,011,944 on chromosome 18 (Young, et al., 1998a). *LPIN2* belongs to a family of nuclear proteins. Three closely related members of the lipin family, Lipin-1, -2, and -3, have been identified in both mouse and human. The lipin 1 (*LPIN1*) gene was originally characterised as a candidate gene for mouse lipodystrophy and plays an important role in lipid metabolism (Peterfy, et al., 2001). The *LPIN2* gene was identified based on sequence similarity to *LPIN1* **(Figure 1.14),** and also plays a role in the lipid metabolism. It may be involved in controlling inflammation and in cell division. With the expression profile by RT-PCR, the *LPIN2* gene is found to be expressed in various eye tissues including cornea, retina, optic nerve, and sclera in humans (Young, 2004; Zhou & Young, 2005).

Myomesin (*MYOM1*)

The *MYP2* critical region is on the p arm, now almost a single contig. It contains one gap just centromeric to D18S481, which is spanned by the myomesin (MYOM1) gene. By the human scleral microarray analysis using six Affymetrix U95A chips, Young identified *MYOM1* as a structural gene expressed in the ECM of the sclera (Young, 2004). The gene important for constituent organization and maintenance of connective tissue function should be given priority for testing the myopia association **(Figure 1.15)**.

Myosin regulatory light chain genes: Myosin, light chain 12A, regulatory *MYCL12A* (*MRCL2),* Myosin, light chain 12B, regulatory *MYCL12B* (*MRCL3)*

The *MRCL2* and *MRCL3* genes were shown to be involved in scleral formation or regulation by human scleral microarray analysis (Young, 2004) **(Figure 1.16)**.

Zinc finger protein 161 homolog (*ZFP161*)

The gene is near the boundary of the *MYP2* interval. The relation of *ZFP161* to myopia might be due to the transfection of *ZFP161* induced dopamine transporter mRNA expression (Scavello et al., 2005). Animal model of experimental myopia indicated neonatal deprivation of form vision alters retinal dopamine metabolism at the same time as the eye enlarges (Stones. et al., 1989). The findings suggest local growth control in the retina and the participation of *ZFP161* and retinal dopamine in the the regulation of axial growth of the eye **(Figure 1.17).**

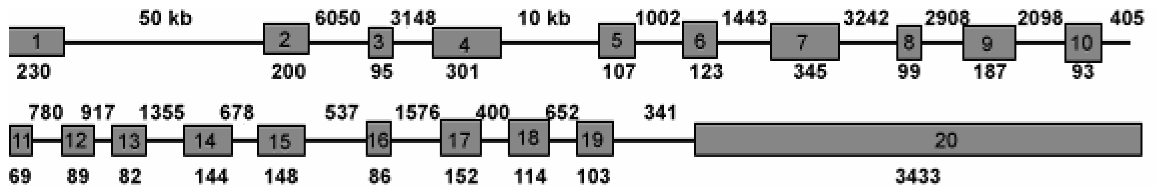


Figure 1.14
A schematic diagram of the *LPIN2* gene (Scavello et al., 2005).

Exons, represented as squares/rectangles, are numbered from 1 to 20, with respective base pair size listed directly below. Intronic size in base pairs, unless stated otherwise, is listed between exons.

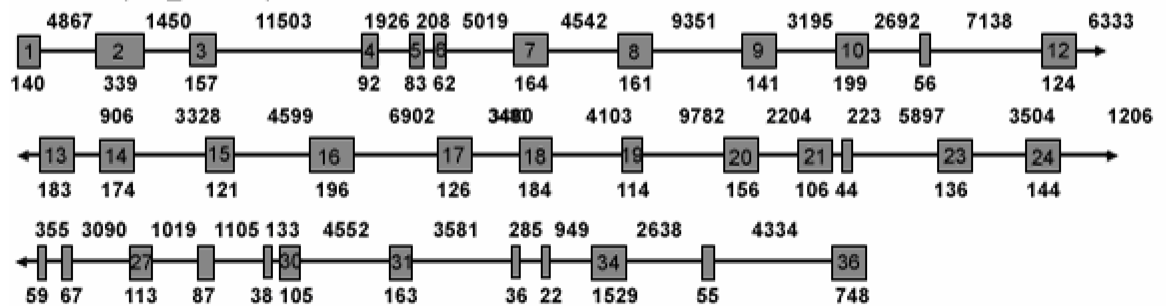


Figure 1.15
Structure of the *MYOM1* gene (Scavello et al., 2005).

Exons, represented as squares/rectangles, are numbered from 1 to 36, with respective base pair size listed directly below. Intronic size in base pairs, unless stated otherwise, is listed between exons.

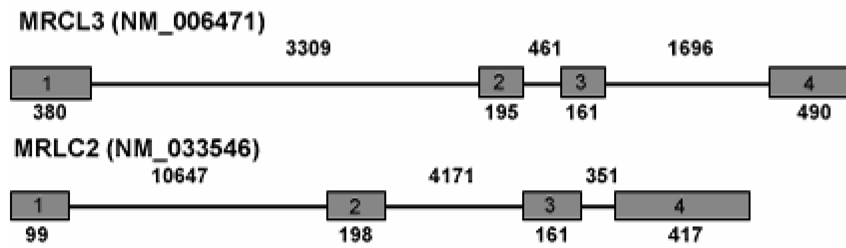


Figure 1.16
Structure of the Myosin regulatory light chain (*MLCB*) gene *MYCL12A* (*MRCL2),* Myosin, light chain 12B, regulatory *MYCL12B*(*MRCL3)* (Scavello et al., 2005).

Exons are represented as squares/rectangles with respective base pair size listed directly below. Intronic sizes in base pairs are listed between exons.

Figure 1.17
Structure of the *ZFP161* gene (Scavello et al., 2005).

Exons are represented as squares/rectangles with respective base pair size listed directly below. Intronic size in base pairs is listed between exons.

1.3.2.3 Challenges of candidate gene approach

1.3.2.3.1 Replication

The candidate gene association studies can detect genomic variants underlying susceptibility to complex diseases. It is vital to replicate previous studies to prove that associations that are identified reflect interesting biological processes rather than methodological issues. However, obtaining replication of initial association findings has proved difficult (Ioannidis, 2007). In a comprehensive review of over 600 positive association studies, only 166 were studied more than three times and only six were consistently replicated (Hirschhorn, et al., 2002). The inconsistency can be attributed to the inadequacies in study design, implementation, and interpretation, which resulted in a significant loss of statistical power to detect the genetic effects (Bacanu, et al., 2000).

In order to improve study reproducibility, a series of follow-up studies should be considered during the initial study design whenever possible. Data pooling from multiple investigations can also be used to improve study reproducibility (Ioannidis, et al., 2006).

Another issue is publication bias, which results from the preferential publication of positive associations and the reduced likelihood that negative findings will be reported (Colhoun, et al., 2003). The publication of negative findings resulting from an adequately powered genetic association study is as important as publishing statistically significant findings. To overcome this problem, alternative approaches

are being developed for reporting all genetic association studies on the online database.

1.3.2.3.2 Novel gene exploration

It is not proven to be that successful with respect to complex phenotypes because of our biological knowledge predicted bioinformatically for phenotypes is insufficient or even incorrect, which greatly diminishes our ability to select relevant genes a priori (Devlin & Roeder, 1999).

1.3.2.3.3 Population stratification

Population stratification can be thought of as confounding by ethnicity. If the case and control populations are not well matched for ethnicity or geographic origin, false positive, false-negative association effects and deviations from Hardy-Weinberg equilibrium might be obtained, which cause inconsistent association results across studies (Deng, 2001). Family-based data can be used to overcome this confounding problem, and the effect of population stratification can easily be controlled using methods such as the transmission disequilibrium test (Pritchard & Rosenberg, 1999). Software packages such as Structure and Strat can be used with genomic controls to correct for effects due to population stratification (Devlin, 1999; Pritchard, 1999).

1.3.2.4 Statistics of association analysis

Association analysis is to determine the contribution of genes to disease susceptibility. It can be family-based or population-based.

1.3.2.4.1 Univariate statistical tests

Analysis typically starts with univariate statistical tests (e.g. chi-squared tests, logistic regression, etc) of each marker individually.

Chi-squared tests (χ^2)

The χ^2 test can be used for the comparison of two independent proportions (p_1, p_2). To calculate the χ^2 statistic, a 2×2 table is first constructed, actual numbers of occurrences are placed, and the expected frequencies in each cell are calculated. The expected frequency in a cell is the product of the relevant row and column totals divided by the sample size (grand total; N = a+b+c+d). For the cell with observed frequency 'a', for example, the expected value is (a+b)(a+c)/N. The difference between observed (O) and expected (E) values (residual) is the same for each cell but with different signs (- or +). The difference between O and E for each cell (O-E) is then calculated. The χ^2 is calculated as the sum of $(O-E)^2/E$ for all four cells:

$$\chi^2 = \Sigma[(O-E)^2/E]$$

The χ^2 relies on a normal approximation to the distribution of the cell counts, and this approximation may be poor for small sample sizes. The *P* value for the χ^2 test is obtained from tables.

For a case control study, most people perform chi-squared tests in the initial association scan. This can be an allele-based test or a genotype-based test. For a genotype-based test, one can use a 2 degrees of freedom (df) test on the 2×3 table (3 genotypes), a 1 df trend test, or a 1 df test that combines the heterozygote class with the rarer homozygote class. A contingency table showing the distribution of alleles or genotypes can be used for significance testing, odds ratio (OR) estimation and

confidence interval (CI) calculation. To compare the power of these different statistical procedures, genotype-based trend test is more meaningful than allele-based test (Li, et al., 2008b). For a multiallelic locus like the MS markers, however, the number of genotype categories is large. The allele-based test analysis does not take into account the underlying genetic model (recessive, dominant, codominant, additive, multiplicative) (Lewis, 2002). All genetic models were suggested to explore if the genetic model for susceptibility is unknown (Sellers, 2004).

Armitage trend test

The test is the best if the model is additive. For the significance level to give a p-value with the usual probabilistic interpretation, the weights must be specified before examining the data, and only one set of weights may be used (Lewis, 2002). The gradual increase in significance level is best assessed by the trend test by collapsing all other alleles into one category so that there will be three genotypes to compare in cases and controls. For SNP data, usually there are three genotypes (AA, AB, BB).

Confidence interval estimation

When a value is the estimate of an unknown "true" value, CIs can be applied to them. CI is more informative than the simple results of hypothesis tests, where a null hypothesis is rejected or accepted. In the case of comparing two groups, a CI enables the researcher to see how large the difference between two proportions may be, not simply whether it is different from zero. CIs can be calculated for different confidence levels. If a CI is calculated at a 95% level (as usually done), 5% of the time the true population parameter will not be contained within the interval calculated from the

sample statistics. More technically, it means that 95% of all samples drawn from the same population will have the population parameter within this interval. The width of the CI also gives us an idea about how uncertain we are about the unknown population parameter. The most common CIs are calculated for a mean, for the difference between two means and for a relative risk (RR) or odds ratio. A very wide interval may indicate that the sample size should be increased to be more confident about the parameter.

Logistic regression

Instead of chi-squared tests, logistic regression is used to incorporate covariates into the initial scan (Amos, et al., 1997). It is used when the outcome variable is binomial or categorical; the outcome is a logistic regression equation in the following format:

$$\text{logarithm of odds} = a + b_1x_1 + b_2x_2 + ... + b_nx_n$$

Here, logarithm of odds is the natural logarithm of the overall OR for all variables included in the model. By using different values of the explanatory variables in the formula, different ORs can be calculated for any combination of the values the variables can take (for example, gender, parental history, educational level and age, etc). Each coefficient (b_i) provides a measure of the degree of association between each variable and the outcome. This coefficient is the logarithm of the OR for that variable ($OR = e^b$) controlled (adjusted or corrected) for the other variables in the model. It is also possible to calculate confidence intervals for the estimated OR as well as the statistical significance of each coefficient.

1.3.2.4.2 Haplotype tests

To increase statistical power in an association study, the use of haplotypes rather than

alleles for analysis is a more reliable method to determine the associated disease risks (Collins, et al., 1997). However, current genotyping technologies are unable to determine the phases of several different markers or a haplotype. Current experimental approaches by means of mouse cell line hybrid, cloning into a plasmid and allele specific PCR is not efficient in routine high-throughput analysis. The alternative common way is to computationally determine haplotypes (Botstein & Risch, 2003). In less favorable situations, historic crossover points may be less obvious. For these cases, new computer algorithms based on haplotype analysis have been developed to estimate statistically the likely locations of the disease mutation (Zhu et al, 2003).

1.3.2.4.3 Resampling statistics

Resampling procedures (bootstrap, permutation and other simulations) have recently become the method of choice for hypothesis testing and estimation of confidence intervals. With resampling, the data are repeatedly resampled, if needed according to a model suggested by the data, to assess the variability of a statistic or estimate calculated from the observed data. The software Haploview can be used to do permutation test.

Haploview is a commonly used bioinformatics software tool which is designed to analyse and visualize patterns of LD in genetic data (Barrett, et al., 2005). Haploview also provides functionality for choosing tagSNPs, performing association studies and estimating haplotype frequencies. Haploview is developed and maintained by Dr. Mark Daly's lab at the MIT/Harvard Broad Institute.

There are many computer packages for analysing genetic association, such as UNPHASED (Dudbridge, 2008), WHAP (Purcell, et al., 2007a), weighted variance FBAT (Lu & Cantor, 2007), Merlin (Abecasis, et al., 2002) and PLINK (Purcell, et al., 2007b). However, simple genotypic or allelic association with a dichotomous trait can be examined using the chi-squared test for significance. Meta-analyses of multiple studies can provide clues but not definitive results. The imputation-based approach for association mapping allows genetic variants that are not actually typed in an association study to be tested for association. Testing imputed variants can increase the power to detect associations, particularly when it is used to combine data from multiple studies. The imputation programme predicts unmeasured genotypes that are correlated with measured genotypes, relaying on the correlation patterns from a database consisting of a panel of densely genotyped individuals.

1.3.2.5 Power of studies

Power calculation is important in genetic association studies of complex diseases (Montana & Pritchard, 2004; Schliekelman, 2008). Genetic susceptibility to such diseases involves a large number of alleles, each conferring only a small genotypic risk (like OR = 1.2 to 2.0), working together with environmental factors to produce a phenotype. Apart from increasing the sample sizes to give greater power and validity, healthy controls or genetically severe cases (with family history of disease) can be used. Other strategies to increase statistical power in an association study are the use of haplotypes rather than alleles in analysis.

The statistical power (Ψ) of a genetic association study is a variable parameter and can be determined as a function of three parameters:

1) significant level (α),

2) sample size (n),

3) effect size (ε)

1) The significant level (α) influences the likelihood of statistical significance. As the stated α level increases, statistical significance is more readily achieved, but the chance of committing a Type I statistical error increases proportionally.

2) Sample size (n)

The sample size of a study (n) and its relationship to Ψ provides a mean of protection from committing a Type I statistical error. The power of a study will increase in proportion to the increase in n. If α is set to a more stringent statistical threshold (0.05 to 0.01), then an increase in n is needed to retain the same Ψ.

3) Effect size (ε)

This parameter accounts for the difference in genetic variation frequency that truly exists between a case group and a control group. For example, an effect size of 3 indicates a three-fold difference in the frequency of a tested genetic variation in a case group compared to a matched control group. As ε decreases, n needs to increase in order to retain Ψ. Smaller values of n create greater risks of failure unless ε is large.

The most updated: complex diseases

Genetic association studies offer a potentially powerful approach to mapping causal genes with modest effects, but are limited because only a small number of genes can be studied at a time. Genome-wide association studies, without a priori assumptions concerning the potential importance of genes or chromosomal regions, recently discovered loci associated with complex diseases, which then become the foci of more intensive follow-up analyses. High-throughput next-generation sequencing (NGS) platforms produce tens to hundreds of millions of short reads (25 - 50 bp) in a single run, and can be used for the resequencing of targeted regions to identify causal disease-susceptibility alleles, such as the 1000 Genomes Project and the Personal Genome Project (http://www.personalgenomes.org) (Pfeifer & Hainaut, 2011).

Summary

There is a move in human genetics from the mapping of Mendelian conditions by linkage analysis to the identification of genes underlying complex diseases by association studies. Myopia has a significant genetic basis, and identifying genes that confer susceptibility to myopia could aid in preventing and effectively treating the disease in the long run. Several genome studies have been conducted to identify major susceptibility loci that are linked with myopia, and candidate genes that relates to the components of the sclera.

The candidate gene approach is commonly used to identify genetic risk factors for complex disorders such as myopia, and directly tests the effects of genetic variants of a potentially contributing gene in an association study. These studies can be

performed relatively quickly and inexpensively, and may allow identification of genes with small effects. However, the candidate gene approach is limited by how much is known of the biology of the disease being investigated. As researchers identify potential candidate genes using animal studies or linkage analysis with further development in construction of LD and haplotype block, the candidate gene approach will continue to be used and will provide new insights into the identification of novel genes and pathways related to myopia.

1.3.3 Pooling approach

By testing DNA pools rather than individual samples, the number of tests for a case-control association study can be decreased to only two for each marker: one for the case pool and one for the control pool. The use of DNA pools saves a lot of time, DNA quantity, and reagent costs. This method has been regarded as an efficient tool for high-throughput association screening with a certain level of experimental error. In this section, the pooling approach will be discussed for the estimation of the allele frequencies of SNPs in DNA pools by primer extension reaction followed by DHPLC analysis.

1.3.3.1 Quantitative genotyping and platforms

To identify common disease genes by means of case-control association studies requires a large number of markers. A recent report on the scale of markers for a whole genome screening (Collins, et al., 1999) suggested a mean distance of about 100 kb per marker, projecting a required number of 10000 markers. High-throughput

genotyping methods are, therefore, essential to allow the analysis of markers, making the most advantageous and immediate use of the SNP data. The need of testing a large number of genetic markers in a large number of DNA samples entails an unacceptable working load. To address this challenge in mapping common disease genes, testing case and control DNA pools rather than individual samples was initially proposed by the Barcellos group (Barcellos, et al., 1997). This method involves the preparation of a pool of affected individuals and a pool of unaffected individuals; the pools are then genotyped for each marker and allele frequencies are compared. Thus, theoretically, only two amplifications are needed for each marker.

Current genetic markers, microsatellites and single nucleotide polymorphisms, are abundant for the identification of genes underlying common diseases. We only concentrate on SNP-based methodologies because of the difficulty in the estimation of MS allele frequencies in pooled DNAs confounded by the stutter bands around marker peaks (Barcellos et al., 1997). PCR amplification of MSs produces a stutter artifact, which generates additional (usually shorter) DNA fragments corresponding to 1, 2, 3, etc. units shorter than the authentic fragments. When DNA pools are genotyped, these stutter bands overlap and increase the height of shorter allele peaks. This can lead to a distortion in the estimated differences in allele frequencies between two groups analysed by the DNA pooling approach. Another artifact of PCR is differential amplification, which refers to the less efficient amplification of longer alleles. This can also distort the estimated differences.

1.3.3.2 Primer extension (PE) reaction coupled with DHPLC Analysis for DNA Pools

Primer extension is a two-step process that first involves the hybridization of a primer to the bases immediately upstream of the SNP nucleotide followed by a 'mini-sequencing' reaction, in which DNA polymerase extends the hybridized primer by adding a base that is complementary to the SNP nucleotide. This incorporated base is detected and determines the SNP allele (Syvanen, 2001). Since primer extension is based on the highly accurate DNA polymerase enzyme, the method is generally very reliable. Primer extension is able to genotype most SNPs under very similar reaction conditions, making it also highly flexible.

The primer extension method is used in a number of assay formats. These formats use a wide range of detection techniques that include MALDI-TOF Mass spectrometry (Storm, et al., 2003) and DHPLC-based assays (Hua, et al., 2004; Yip, et al., 2006).

A primer hybridizes to the target DNA immediately upstream of SNP nucleotide, and a single dideoxynucleotides (ddNTP) complementary to the SNP allele is added to the 3′ end of the primer (the missing 3'-hydroxyl in dideoxynucleotide prevents additional nucleotides from being added). If each ddNTP is labelled with a different fluorescent dye, this allows the detection of all four alleles in the same reaction. If unlabelled ddNTPs are used, the extended products each with a different ddNTP incorporated can still be distinguished from each other by means of MALDI-TOF mass spectrometry or DHPLC. In this study, the two extended alleles are discriminated by DHPLC, which also allows a precise quantification of the allelic

ratio in the tested DNA sample. The elution profile corresponds to SNP is obtained (**Figure 1.18**).

Pooling problems for SNP marker

The drawback of the PE essay is that partial self-complementarily at the 3′ end of primers might result in self-annealing, which allows extension to occur independently of the target template. However, this problem can be identified readily by carrying out a control extension reaction in the absence of a template.

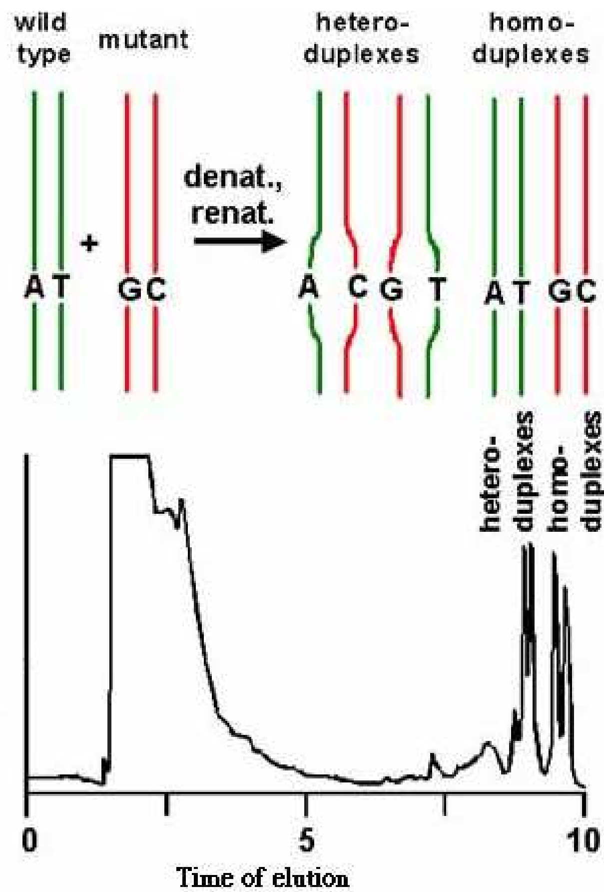


Figure 1.18
The principle of the DHPLC (Xiao & Oefner, 2001)

Heteroduplex forms through hybridization after heating and cooling the PCR products. Separation of heteroduplexes from homoduplexes is accomplished under partially denaturing conditions.

1.3.3.3 Quantitative genotyping - Allele frequency estimation in DNA pools

The genotyping approach based on primer extension followed by denaturing high performance liquid chromatography (PE-DHPLC) can be applied to the quantification of SNP allele frequencies in DNA pools. The DHPLC elution profile of the primer extension reaction gives reproducible chromatograms in which the two alleles are resolved as two separate peaks. However, the ratio of the two peaks does not always directly correspond to the allele frequency as demonstrated by the unequal peak heights in heterozygous samples for many SNPs. The reason is that the extension enzyme (e.g. Therminator) incorporates different ddNTPs with different efficiencies in the primer extension reaction. Therefore, direct estimation of allele frequency is inaccurate.

Correction for SNP markers

To improve unbiased estimates of pooled allele frequencies, a solution is to use known individuals heterozygous for a SNP to calibrate an observed peak height (Le Hellard, et al., 2002). The principle of measuring heterozygous signals is a base reference to pooled allele amplification. The k-correction method is a mathematical correction for differential primer extension reaction efficiency (Le Hellard et al., 2002) (**Figure 1.19**). In order to reduce the standard error of the estimate of this calibration factor, data from several individuals heterozygous for a marker are required (Le Hellard et al., 2002). This is a significant problem for SNPs with low minor allele frequencies as it means hundreds of individuals must be genotyped to be sure of having sufficient heterozygotes, negating the benefits of DNA pooling.

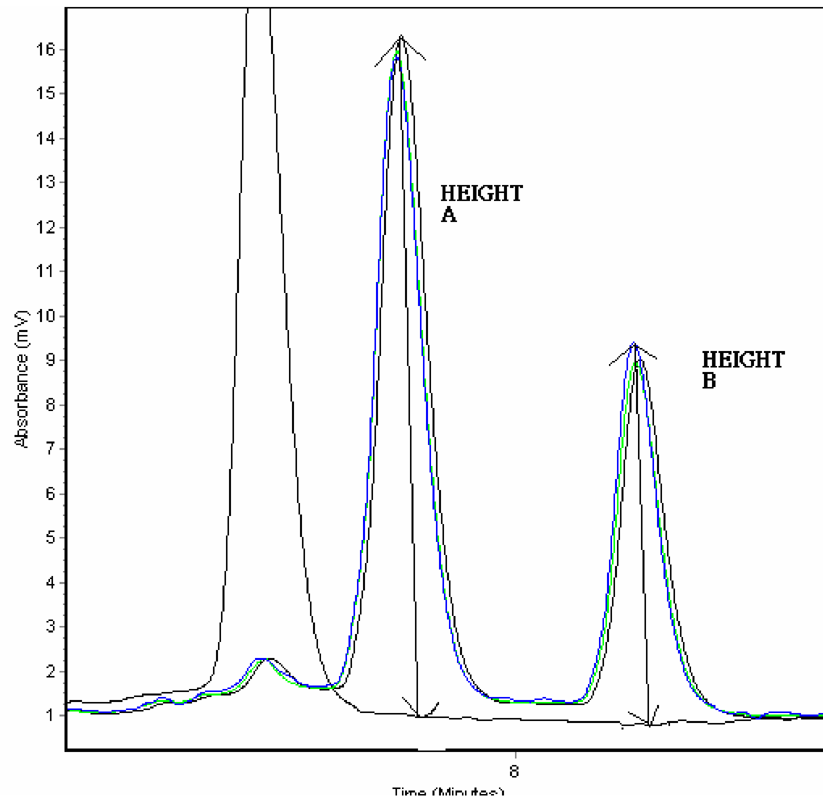


Figure 1.19
Correction for SNP Markers in estimating the allele frequency

Unequal representation of alleles within the DNA pool is corrected by applying the k correction factor, which is defined as the ratio of the peak heights in a heterozygous DNA sample. A and B are the peak heights of alleles A and B, Frequency of allele A=A/ (A+kB), where k=A/B in a heterozygote.

1.3.3.4 Comparison between real allele frequencies and allele frequencies estimated in the pools

The composition of the fragments amplified from the pool should reflect the true frequency of the two alleles in the included samples (Giordano, et al., 2001). This requires that each DNA sample is equally represented in the pool and that all DNA are equally amplifiable. This depends on the DNA purification procedure, the accuracy of DNA quantification, DNA degradation, etc.

If initial comparison of DNA pools shows statistically significant difference in allele frequencies, confirmatory study is carried out by genotyping individual samples. If initial comparison does not show any significant difference, then the SNP will not be further investigated. Therefore, this approach allows more time and effort to be spent on sequence variations that are worthy of further investigation, and unpromising sequence variations are abandoned after initial testing.

1.3.3.5 Marker selection for pooling studies

A **tag SNP** is a representative SNP in a region of the genome with high LD (the non-random association of alleles at two or more loci). It is possible to identify genetic variation without genotyping every SNP in a chromosomal region. Tag SNPs are useful in whole-genome SNP association studies in which hundreds of thousands of SNPs across the entire genome are genotyped. For this reason, the International HapMap Project aims to provide tag SNPs for discovering genes responsible for certain disorders.

1.3.3.6 Statistical analysis of pooling data

1.3.3.6.1 Statistical analysis of estimated allele frequencies in DNA pools

Allele frequencies of SNPs were determined based on the peak heights generated from DHPLC for each SNP and for each of the pools. Frequencies were averaged across three measurements and corrected for unequal detection of the two alleles with relative amplification efficiencies of both alleles (k correction factor) based on the ratios in individual heterozygote (Shifman, et al., 2002). Frequency estimates for high myopia and emmetropia controls were obtained from the average frequency in the six pools belonging to each category. Significance of allele frequency differences between case and control pools were estimated by nested ANOVA, which handles the error variance introduced by the primer extension DHPLC-specific DNA pooling procedure. We applied a lenient level of significance ($P < 0.1$) for selecting SNPs for follow-up individual genotyping because this might underestimate some allele frequency differences.

1.3.3.6.2 Statistical analysis of allele frequencies in individual genotyping

Only those SNPs meeting the lenient level of significance ($P < 0.1$) were considered for follow-up. For individual genotyping, significance of case–control association and deviation from Hardy-Weinberg equilibrium (HWE) were tested by chi-square tests. Empirical corrected p values were calculated using the rank of the chi-square test value divided by the successful genotyping markers this study.

1.3.3.6.3 Sources of error in DNA pooling approach

Three (**E1, E2 and E3**) potential sources of experimental errors in allele frequency measurement were estimated across DNA pools (Le Hellard et al., 2002; Visscher & Le Hellard, 2003):

- The standard sampling error (**E1**). DNA pool constructs can be underpowered if a finite number of individuals from a population cohort are randomly sampled (i.e. small sample size). E1 can therefore be minimized by increasing the sample size (n) of individual DNA samples in pool constructs.
- A second potential source of DNA pooling error (**E2**) originates in the initial correction factor (k) estimate. Spurious estimates of k (**E2**) can be adjusted by genotyping an appropriate number of individual heterozygotes or an appropriate number of replicated single heterozygotes.
- Third, a pool specific measurement error (**E3**) is evident if insufficient numbers of pool genotype replicates are used to accommodate for varying degrees of variation in allele frequency estimates between DNA pools.

Conclusion

This chapter has highlighted the potential increase in efficiency that DNA pooling can offer for association studies. However, it has also drawn attention to several methodological issues that must be attended to for DNA pooling to work effectively. Clearly, laboratory procedures must be optimized to obtain reliable results with minimal biases and errors. Further reduction of errors will depend on experimental design and statistical analysis. The second stage of follow-up individual genotyping should be adopted for markers that show putative association in the initial pooling

screening. The frequency of the two SNP alleles in a DNA pool can be determined by a sensitive and reproducible approach by coupling allele discrimination by primer extension and quantitative detection by DHPLC (Hoogendoorn, et al., 2000). A sophisticated selection of SNP markers might promote the effective screening of human genomes for disease association.

1.4 Study Objectives

Myopia or short-sightedness is very common in Orientals like Chinese and as many as 80% of young Chinese adults are myopic. Myopia is a multifactorial condition caused by multiple genes, environmental factors and possibly their interactions.

To identify genes predisposing humans to myopia, DNA samples from highly myopic adults and emmetropic control subjects were recruited and tested. Two approaches to studying the candidate genes were adopted in this study: 1) replication study of the *MYOC* gene, 2) investigation of candidate genes in the *MYP2* interval by DNA pooling method.

1) Replication study of the *MYOC* gene

Myocilin (*MYOC*) is one of the candidate genes more likely to be associated with myopia. The purpose of this study was to determine whether the genetic polymorphisms of *MYOC* were associated with high myopia in the Chinese population. A total of 300 unrelated Chinese Han subjects with high myopia and 300 unrelated emmetropic control subjects were recruited and genotyped for eight SNPs

and two MSs in the *MYOC* gene. Their genotypes, allele frequencies, and ORs were analysed.

2) Identification of myopia susceptibility gene in the *MYP2* interval by DNA pooling.

The inheritance of autosomal dominant high-grade myopia (refractive error greater than or equal to 6 dioptres) has been confirmed in multiple genome-wide studies on a region of chromosome *MYP2* at 18p11.31. In this region, Young et al. (1998) identified a list of positional candidate genes which are highly expressed in the ocular tissue, and might thus be related to myopic development. The strategy of selective DNA pooling was used as a primary screen to identify implicated genes affecting refraction. DNA pools representing the cases and controls were compared for differential allelic enrichment using widely dispersed SNP variants. To evaluate putative susceptibility alleles within the *MYP2* locus, subsequent individual genotyping was employed to estimate the gene effects and assess statistical significance.

Chapter 2
Materials & methods

2.1 Research design

Han Chinese subjects with high myopia and unrelated control subjects were recruited at The Hong Kong Polytechnic University as a result of poster promotion inviting volunteers to participate in the study. In addition, an additional set of 162 Chinese families with high myopia (557 subjects) recruited for a previous study (Tang, et al., 2007) in our group was also used for the study focusing on the *MYOC* gene. Of the recruited families, 95 families had one myopic offspring, 63 families two myopic offspring, and 4 families three myopic offspring. Ophthalmological examination and body check (e.g. laboratory tests for blood counts and blood groups) were offered free to the subjects. Ophthalmological clinical diagnosis of each individual was performed by optometrists, including visual acuity, refraction, slit lamp and dilated fundus examination, at The Hong Kong Polytechnic University Optometry Eye Clinic. The body check and venous blood collection were carried out in the private laboratory PHC Medical Diagnostic Centre. DNA was extracted from blood samples within a few days after collection. Ethical approval for this project was obtained from the Human Subjects Ethics Subcommittee of The Hong Kong Polytechnic University and was in accordance with the tenets of the Declaration of Helsinki. Written informed consent was obtained from every study subject **(Appendix A)**. Both subject groups were analysed statistically to determine whether any significant association exists between the gene and myopia susceptibility.

2.2 Materials

2.2.1 Materials for laboratory experiments

2.2.1.1 Chemicals and reagents

DNA extraction

Absolute ethanol (AnalaR grade) and isopropanol (AnalaR grade) were obtained from Riedel-de Haën (Seelze, Germany).

Polymerase Chain Reactions (PCR)

Reagents used : Two thermostable DNA polymerases with their corresponding reaction buffers being employed for all PCR applications, AmpliTaq Gold DNA Polymerase (Applied Biosystems) and HotStarTaq Plus Polymerase (Qiagen); deoxyribonucleoside triphosphates (dNTPs) from GE Healthcare (formerly Amersham Pharmacia; Piscataway, USA).

Primers used for PCR and SNP genotyping were ordered from either Integrated DNA Technologies (Coralville, IA, USA) or Tech Dragon Limited (Hong Kong, China).

Gel electrophoresis

SeaKem LE agarose (Lonza, USA) with ethidium bromide (Sigma, USA) or polyacrylamide gel (PAGE) (Bio-Rad, USA) with SYBR Green I (Invitrogen, USA) were used for gel electrophoresis whereas 1 Kb plus DNA Ladder was purchased from Invitrogen (formerly Life Technologies; Carlsbad, CA, USA) while 6× loading dye contained 30% glycerol (AJAX Chemicals, Australia).

Denaturing high performance liquid chromatography (DHPLC)

Triethylammonium acetate (TEAA) with pH 7.0 from Transgenomic (San Jose, CA, USA) and acetonitrile (ACN) from Fisher Scientific (NJ, USA) were used for the preparation of different types of buffers.

Cycle Sequencing

PCR products were purified by exonuclease I (ExoI) (New England Biolabs, Beverly, USA) and shrimp alkaline phosphatase (SAP) (GE Healthcare, Piscataway, USA). BigDye Terminator Cycle Sequencing Kit version 1.1, template suppression reagent (TSR), performance optimized polymer 4 (POP-4) and HiDi Formamide for resuspending cycle sequencing products were obtained from Applied Biosystems.

Enzymes used for SNP genotyping

- *PCR-RFLP analysis*

The restriction enzyme BseNI was from New England Biolabs while BanII, Eco130I, MvaI, Hpy166II, BseNI, HpyF3I, RsaI, Alw26I, HinfI and HphI were obtained from Fermantas (CHINAGEN, Hong Kong).

- *Primer extension DHPLC analysis*

Therminator DNA Polymerase used in primer extension reaction from New England Biolabs

Microsatellite Genotyping

POP-4, HiDi Formamide and GeneScan 500 [TAMRA] size standard were all from Applied Biosystems.

DNA concentration measurement for pooling experiments

PicoGreen dsDNA Quantitation Reagent Kit (Molecular Probes, Eugene, OR, USA) was used for measuring DNA concentration before DNA pooling. The kit contained PicoGreen solution in DMSO, 20× TE buffer and Lambda DNA standard in TE.

2.2.1.2 Solutions

All solutions were prepared using water purified by reverse osmosis (Millipore, Bedford, USA). Water used in PCR work and DNA preparation was further irradiated by ultra-violet light (MilliQ) and autoclaved.

2.2.1.2.1 DNA extraction

- *Red cell lysis (RBC) lysis solution*

It contained 155 mM ammonium chloride (NH_4Cl) [BDH, Poole, UK], 10 mM potassium bicarbonate ($KHCO_3$) [BDH, Poole, UK] and 1 mM disodium ethylenediamine tetra-acetic acid (Na_2EDTA) [BDH, Poole, UK]. The solution was adjusted to pH 7.4 with concentrated sodium hydroxide (NaOH) [Riedel Haen, Seelze, Germany].

- *White cell (WBC) lysis solution*

It consisted of 25 mM Na_2EDTA (BDH, Poole, UK) and 2% sodium dodecyl sulphate (SDS) [Sigma, St. Louis, USA] with pH 7.4.

- *Protein precipitation solution*

It contained 10 M ammonium acetate (BDH, Poole, UK). It was sterilized by membrane filtration and then stored at room temperature.

- *TE buffer*

It contained 10 mM Tris-hydrochloric acid (Tris-HCl) [Sigma, St. Louis, USA], 1 mM Na_2EDTA [BDH, Poole, UK], and was adjusted to pH 8.0. All solutions were sterilized by autoclaving before use except protein precipitation solution.

2.2.1.2.2 Electrophoresis

- *10× Tris-Borate-EDTA (TBE) buffer*

It contained 890 mM Tris (hydroxymethy)-aminoethane (Tris) [Life Technologies, Rockville, USA], 890 mM boric acid (Riedel-de Haën, Seelzw, Germany), 25 mM Na_2EDTA (BDH, Poole, UK), pH 8.3. It was then diluted to 0.5× solution for agarose gel electrophoresis.

- *6x loading buffer.*

It contained 30% glycerol (AJAX Chemicals, Auburn, Australia), 1× TE buffer and 0.05% bromophenol blue (Sigma, St. Louis, USA).

2.2.1.2.3 DHPLC buffer

The components of DHPLC buffers are shown below:

Buffer A: 0.1 M TEAA (pH 7.0), 0.25% ACN

Buffer B: 0.1 M TEAA (pH 7.0), 25% ACN

Buffer C: 8% ACN

Buffer D: 75% ACN

2.2.1.3 Equipment

DNA concentration was measured by spectrophotometry with MBA 2000 (Perkin Elmer) or Victor3™ V 1420 Mutilabel Counter (Perkin Elmer, Waltham, MA, USA). PCR was performed in the 96-well GeneAmp PCR system 9700 (Applied Biosystems, Foster City, USA). Allelic specific PCR was carried out in ABI 7500 Real-time PCR system (Applied Biosystems, Foster City, USA) or LightCycler 480 Real-time PCR System (Roche Applied Science, Indianapolis, IN, USA). DHPLC analysis was performed on the WAVE DNA Fragment Analysis System containing DNASep Cartridge (Transgenomic, Omaha, NE, USA). ABI PRISM™ 310 Genetic Analyzer (Applied Biosystems, Foster City, USA) was used for DNA sequencing and for DNA fragment length determination in MS genotyping.

2.3 Methods

This part gives the methodology of 2 main projects ("Replication Study for the *MYOC* gene" & "A DNA pooling approach to identifying myopia susceptibility genes in the *MYP2* locus"). The details are described in the following sections.

2.3.1 Sample collection and diagnosis

2.3.1.1 Subject recruitment

Mapping 'myopia genes' using molecular genetics techniques is a lengthy, meticulous exercise (Guggenheim et al., 2000). Essentially, the first step is to recruit myopic subjects – both "**Individual Case/Control**" and "**Family"** sample sets.

2.3.1.1.1 Individual Case/Control sample set

In total, 600 DNA samples from unrelated Han Chinese individuals were recruited in The Hong Kong Polytechnic University (300 subjects with high myopia and 300 control subjects). Clinically, the high myopia subjects had an average age of onset of ~12 years. The unaffected control group was recruited from the same geographical location as the affected group (Hong Kong, China), and was carefully matched to the high myopic group for age (within 3 years) and ethnicity (Southern Han Chinese). This would thus reduce the possibility of spurious results due to underlying population stratification. The entry criteria of refractive error of both eyes were at least -8.00 dioptres (D) for high myopes (cases), and refractive error of both eyes within 1 D for emmetropes (controls). To promote the statistical power in current genetic association study of myopia candidate genes, the entry criterion for highly myopic subjects was more stringent than previous myopia studies (Leung et al, 2000; Zayats et al, 2009) as we recruited case subjects with an extreme phenotype with a spherical power worse than - 8.00 D for both eyes. The entry age of subjects was from 18 to 45 as we thought that they were unlikely to develop high myopia due to senile cataract. Subjects were excluded from the study if they showed obvious signs of ocular disease or other inherited disease associated with myopia, such as Stickler or Marfan syndromes.

2.3.1.1.2 Family sample set

In addition, 162 unrelated Chinese nuclear families with high myopia (557 subjects in total) were recruited in a previous family study in our group (Tang et al., 2007). Each family with at least one highly myopic offspring and their parents were

recruited in The Hong Kong Polytechnic University. The entry criteria of refractive error of both eyes were at least -5.00 dioptres (D).

The key to unlocking the secrets of myopia genetics lies in the process of recruiting individual subjects and families. By means of placing posters in the campus and Optometry Clinic of The Hong Kong Polytechnic University (PolyU), visual screening activities in local society and referral of myopic subjects from optometrists across Hong Kong, I recruited over 200 families for the Family Study of Myopia during my post of being a research assistant, and then recruited about 500 unrelated case-control subjects in the first two years of my PhD study. Another optometrist Mr. Percy Ng Po Wah in our team then followed up the task of recruiting more subjects willing to take part in the study.

2.3.1.2 Phenotypic data

Objective refraction was taken using open field autorefractor (Shin-Nippon SRW-5000, Tokyo, Japan) after the subject was given one to two drops of 1% tropicamide per eye. Spherical equivalent (SE) was calculated as the sum of sphere dioptres and half cylinder dioptres. Intra-ocular pressure (IOP) by non-contact tonometer (NIDEK NT-2000 Ver1.03, Japan) was measured before any eye-drop instillation. Corrected distance visual acuity was measured by LogMAR chart at 3 metres. Central corneal curvature was measured using autokeratometry (Canon RK-5 Auto Ref-keratometer, Canon, Inc., Tokyo, Japan). Then, axial length, vitreous chamber depth (VCD) and lens thickness (TL) were measured using A-Scan ultrasound (Advent A/B System; Mentor, Santa Barbara, CA) after one drop of 0.4%

benoxinate hydrochloride was instilled in each eye to produce anesthesia. Fundus examination was reviewed to assess the ocular health of each subject. Subjects were excluded from the study if they showed obvious signs of ocular diseases (such as glaucoma and keratoconus) other than retinal changes typically associated with myopia, known genetic diseases (such as Marfan syndrome and Stickler syndrome) with myopia as one of the presenting features, or history of ocular trauma. Age of onset and family history were obtained from verbal interview.

2.3.2 Experimental background and protocols

DNA extraction and spectrophotometric quantification at 260 nm (Sections 2.3.2.1 to 2.3.2.2) were carried out by another member of the myopia genetics research team.

2.3.2.1 DNA extraction

Venous blood (9-12 ml) was collected from the subjects after eye examination, and DNA extracted from the leukocytes by a modified salt precipitation method (http://leedsdna.info/HUGO/2000/manual/manual.htm).

There were three basic steps in DNA extraction: cell lysis to expose the DNA within, removing membrane lipids by adding a detergent, removing proteins by adding a protease, and precipitating the DNA with alcohol. Since DNA is insoluble in alcohol, it would aggregate together, giving a pellet upon centrifugation. This step also removes alcohol-soluble salt. Cellular and histone proteins bound to the DNA could be removed either by adding a protease or by precipitating the proteins with sodium

or ammonium acetate, or extracting them with a phenol-chloroform mixture prior to the DNA-precipitation.

Three ml EDTA-anticoagulated blood samples were placed into polypropylene centrifuge tube. Then, 9 ml of red cell lysis buffer were added to each tube. The tubes were centrifuged for 3300 RPM (2000 *g*) for 10 minutes. The supernatant was discarded and 3 ml of white cell lysis solution were added to resuspend. Then the mixture was incubated for 30 minutes at 37°C, followed by cooling to below 21 ^{o}C for 5 minutes on ice. The mixture was mixed well with 1 ml protein precipitation solution and centrifuged for 10 minutes at 3300 RPM (2000 *g*). The supernatant was transferred to a clean tube, leaving behind the precipitated protein pellet. Three ml of isopropanol were added and the tube inverted several times until the DNA precipitate was visible. The mixture was spun at 2000 g for 10 minutes, the supernatant poured off and the tube blotted on a paper towel. The DNA pellet was washed twice with 70% ethanol with centrifugation at 2000 *g* for 10 minutes to pack the pellet. The DNA was resuspended in 110 μl TE buffer. The dissolution of DNA was performed at room temperature for overnight before quantification.

2.3.2.2 DNA quantification

Two methods were adopted to measure the concentration of a DNA solution: spectrophotometric quantification (for individual genotyping) and ultraviolet light (UV) fluorescence in presence of a DNA-binding dye (for pooling screening).

2.3.2.2.1 Spectrophotometric quantification

Because DNA absorbs UV light with an absorption peak at 260nm wavelength, spectrophotometers are commonly used to determine the concentration of DNA in a solution. Inside a spectrophotometer, a sample is exposed to ultraviolet light at 260 nm, and a photo-detector measures the light that passes through the sample. The more light absorbed by the sample, the higher the DNA concentration in the sample. Stock dsDNA solution concentration was measured by spectrophotometer with MBA 2000 or Victor3™ V 1420 Mutilabel Counter (only for accurate quantification by the PicGreen method). A working DNA solution at a concentration of 10 ng/μl was prepared by diluting with 1× TE buffer from the stock DNA. The stock DNA was stored at –70°C whereas the working DNA solution for PCR was placed at –20°C.

2.3.3 Sample purity

DNA extraction quality was monitored by UV spectrophotometry (MBA 2000 [Perkin Elmer] and Victor3™ V 1420 Mutilabel Counter [Perkin Elmer, Waltham, MA, USA]). It is common for DNA samples to be contaminated with other molecules (e.g. protein, phenol, and other organic compounds). To assess DNA sample purity, the absorbance at wavelength 280 nm was compared to that at 260 nm. DNA absorbs UV light at 260 and 280 nm, and aromatic proteins absorb UV light at 280 nm. A pure sample of DNA with the 260/280 ratio at 1.8 is relatively free from protein contamination.

2.3.4 Polymerase chain reaction

The polymerase chain reaction (PCR), developed in 1984 by Kary Mullis (Deepak, et al., 2007), is a routine technique to amplify copies of specific fragments of DNA. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA fragments) containing sequences complementary to the target region along with a DNA polymerase are key components to enable selective and repeated amplification.

2.3.4.1 Primer design

Based on information of the databases of GeneWindow and National Centre of Biotechnology Information (NCBI), oligonucleotide primer pairs specific for the target were designed with the software Oligo (Version 6.57; Molecular Biology Insights, Cascade, USA). Primers are short oligonucleotides, i.e., chemically synthesized, single-stranded DNA fragments, usually only 18 to 25 bases long containing nucleotides that are complementary to the nucleotides at both ends of the DNA fragment to be amplified.

The melting temperature (Tm) of the primers was calculated using the equation:

$$69.3°C + (0.41 \times GC\ \%) + (650 / \text{primer length in bp})$$

The Tm was used as a reference for the annealing temperature during PCR optimization.

Primer specificity was further checked with the online tool Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), which blasts the primer sequences against the human genome sequences available in NCBI. In the presence of suitably thermostable DNA polymerase and DNA precursors (the four deoxynucleoside triphosphates: dATP, dCTP, dGTP and dTTP), they initiate the synthesis of new DNA strands which are complementary to the individual DNA strands of the target DNA segment.

2.3.4.2 Thermal cycles

Two thermostable DNA polymerases were employed in the study: AmpliTaq Gold DNA polymerase in the initial stage and HotStar Taq Plus DNA polymerase in a later stage. The whole myopia genetics research team switched to use HotStarTaq Plus DNA polymerase because of its higher efficiency and yield. For the standard protocol, typical 10-μl reactions contained 10× reaction buffer, 0.1-0.3 μM of each of forward and reverse primers, 0.2 unit of Taq DNA polymerase, 0.2 mM each of dNTPs and 10-20 ng of genomic DNA template. For purposes of DNA sequencing and genotyping, PCRs were routinely carried out in volumes of 10-15 μl.

PCR involved an initial denaturing step at 95°C for 5 minutes, followed by 30-40 cycles of denaturation (95°C for 30 seconds), primer annealing (variable, depending on the annealing temperature of the primers) and extension (72°C for 30 seconds) followed by a final extension of 7 minutes at 72°C.

2.3.5 Variations of the basic PCR technique

PCR was modified to perform a wide variety of genetic manipulations.

2.3.5.1 Allele-specific PCR

It is used to genotype SNPs. It requires prior knowledge of a DNA sequence, including differences between alleles, and uses primers whose 3' ends encompass the SNP. PCR amplification under stringent conditions is much less efficient in the presence of a mismatch between template and primer, so successful amplification with a SNP-specific primer signals the presence of a specific allele in a sequence (Newton, et al., 1989).

2.3.5.2 Nested PCR

To increase the specificity of DNA amplification, nested PCR was performed to reduce the background due to non-specific amplification of DNA. Two sets of primers were used in two successive PCRs. In the first reaction, one pair of primers was used to generate DNA products, which might still include non-specifically amplified DNA fragments in addition to the intended target. The product(s) were then used in a second PCR with a set of primers whose binding sites are completely or partially different from and located 3' of each of the primers used in the first reaction. Nested PCR is often more successful in specifically amplifying long DNA fragments than conventional PCR, but it requires more detailed knowledge of the target sequences Nested PCR suffers from some disadvantages. One is the obvious necessity to add the second set of primers. It might also cause allelic drop-out or allelic bias.

2.3.5.3 Touchdown PCR

This variation of PCR aimed to reduce non-specific background by gradually lowering the annealing temperature as PCR cycling progresses (Don, et al., 1991). The annealing temperature in the initial cycles was usually a few degrees (5-10°C) above the Tm of the primers used, while it is a few degrees (5-10°C) below the primer Tm in the later cycles. The higher temperatures gave greater specificity for primer binding, and the lower temperatures permitted more efficient amplification from the specific products formed during the initial cycles.

2.3.6 Gel electrophoresis

Agarose and polyacrylamide gel electrophoresis was used for analysing DNA fragment size and quality. This was achieved by moving negatively charged DNA through a gel matrix with an electric field (Breen, et al., 1977). The agarose gel matrix was made by melting agarose powder in a microwave oven or polyacrylamide gels in TBE buffer. Gels were made routinely between 1 and 3% w/vol. The gels were cast with the addition of 50ng/mL ethidium bromide EtBr (Sigma) or SYBR Green I (Invitrogen) and using plastic combs for wells into which samples could be loaded. Once set, gels were submerged in 0.5× TBE buffer (pH 8.0) in the electrophoresis tank. Samples were pre-mixed with 6× loading dye (AJAX Chemicals, Auburn, Australia) and loaded into the wells of the gel. DNA samples were subjected to electrophoresis for approximately 30 minutes to 1 hour at 100 to160 mV depending on the size and percentage of the gel. Lower concentration of a gel speeds up the migration and enables separation of longer DNA molecules. The higher the voltage, the faster the DNA moves. But voltage is limited by the fact that

it heats up the gel and the buffer. High voltages also decrease the resolution. The estimation of the band size is by comparison with 1 kb DNA Plus Ladder (Invitrogen). The DNA was visualized under UV transillumination.

2.3.7 Cycle sequencing

PCR products were purified with 0.5 unit of ExoI and 1 unit of SAP at 37°C for 30 minutes, after which the enzymes were inactivated at 80°C for 15 minutes. The conditions for cycle sequencing were as follows: 96°C for 1 minute, then 30 cycles of 95°C for 10 seconds, 50-55°C for 10 seconds and 60°C for 4 minutes. A 10-µl reaction mix was prepared as follows: 5µl of purified PCR products, 2µl of ABI terminator ready reaction mix (Big Dye), 0.17 µl of one sequencing primer (10µM) and 3µl of H_2O. Cycle sequencing programme consisted of 96°C for 1 minute, then 30 cycles of 95°C for 10 seconds, 55°C for 10 seconds and 60°C for 4 minutes.

The final product was purified by ethanol precipitation and resuspended in a solution of HiDi formamide (15 µl) before analysis by capillary electrophoresis. Automated DNA sequencing was performed in ABI PRISM 310 (Applied Biosystems). The DNA sequences obtained from samples were compared to consensus sequences obtained from GenBank sequences.

2.4 Marker selection for the *MYOC* gene

Based on the previous report from Tang et al (2007), the 3' flanking region of *MYOC* may contribute to the genetic susceptibility to high myopia in Han Chinese. We attempted to replicate these genetic markers with an independent set of Chinese samples (cases and controls). The SNPs genotyped in this study, their positions and genotyping methods are shown in **Table 2.1**. The four *MYOC* polymorphisms (NGA17, NGA19, rs2421853 and rs235858) that were found associated with high myopia in our previous family study were included in the present study. In addition, six other flanking SNPs (rs12076134, rs1602244, rs171000, rs6425356, rs10737323 and rs743994) were also investigated because they were tagged by rs2421853 or rs235858 in the HapMap Han Chinese subjects with the criteria of r2 >0.8 and minor allele frequency (MAF) of at least 0.1, as determined by the pairwise algorithm of the software Tagger (de Bakker, et al., 2005) implemented online in the HapMap website. In total, two MSs and eight SNPs were studied, and designated as S1 to S10 in sequential order from the 5' end of the gene for the sake of easy reference (**Table 2.1**).

Table 2.1 *MYOC* polymorphisms genotyped in the present study*

Reference No	Markers	Position (bp)	Location	Genotyping method(s)
S1	NGA17	-339	5’ flanking region	Genescan
S2	rs12076134	15697	Intron 2	Comparative Thresholds
S3	NGA19	17307	3’ flanking region	Genescan
S4	rs2421853	21798	3’ flanking region	RFLP
S5	rs1602244	22811	3’ flanking region	Tm-Shift Method & Primer Extension with DHPLC
S6	rs171000	23669	3’ flanking region	Tm-Shift Method
S7	rs6425356	23755	3’ flanking region	Comparative thresholds, Primer extension with DHPLC & RFLP (Family sample)
S8	rs10737323	25085	3’ flanking region	Comparative Thresholds
S9	rs235858	25249	3’ flanking region	Comparative Thresholds
S10	rs743994	25534	3’ flanking region	Comparative Thresholds & RFLP (Family sample)

* The table shows the positions of the polymorphisms relative to the exon structure of the *MYOC* gene and its genotyping methods in the present study. The positions of the polymorphisms are numbered with reference to the first base (A, numbered as +1) of the start codon ATG of the gene.

2.4.1 Genotyping of microsatellite (MS) markers of the *MYOC* locus by GeneScan

Two polymorphic GT microsatellites on the *MYOC* locus, NGA17 at the promoter and NGA19 at the 3' flanking region (**Table 2.1**), were found associated with myopia (Tang et al., 2007). To validate the relationship between the *MYOC* MSs and high myopia, the two MSs were genotyped for the present independent set of Chinese case-control samples. *Homo sapiens* chromosome 1 genomic contig NT_004487 was used as the reference genomic DNA sequences.

Primers previously designed by Tang et al (2007) were used to amplify the sequences flanking the two MSs by PCR (**Table 2.2**). The forward primers (Myocpm-F and Myoc3pm-F) for both markers were labelled with fluorescein at the 5' end. The reaction mixture (15 µl) contained 1× Gold Buffer (15 mM Tris-HCl, 50 mM KCl, pH 8.0), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 0.3 µM of each primer for NGA17 (or 0.5 µM for NGA19), 1 U AmpliTaq Gold DNA polymerase and 20 ng genomic DNA. Touchdown PCR was used to avoid excessive optimization of the reaction conditions (Roux & Hecker, 1997). PCR amplification consisted of initial denaturation for 5 min at 95°C, 8 touchdown cycles, 30 main cycles, and final extension for 7 min at 72°C. Both touchdown and main cycles consisted of 30 seconds at 95°C, 45 seconds at the annealing temperature (**Table 2.2**) and 45 seconds at 72°C. The annealing temperature of the main cycles was 53°C for NGA17 and 57°C for NGA19, and the initial annealing temperature for the touchdown cycles was 7°C above this with 1°C reduction for each successive touch-down cycle. As the two PCR fragments were sufficiently different in size (about 330 bp for NGA17 and

about 145 bp for NGA19), the two PCR products separately amplified from the same individual were mixed and genotyped in the same injection. The PCR products were separated on the ABI PRISM 310 Genetic Analyzer using GeneScan software together with GeneScan-500 (TAMRA) size standard according to the manufacturer's instructions, and the genotypes were called manually **(Figure 2.1)**. For both NGA17 and NGA19, the alleles were designated as 3 to 8 if the alleles had 13 to 18 GT repeats, respectively.

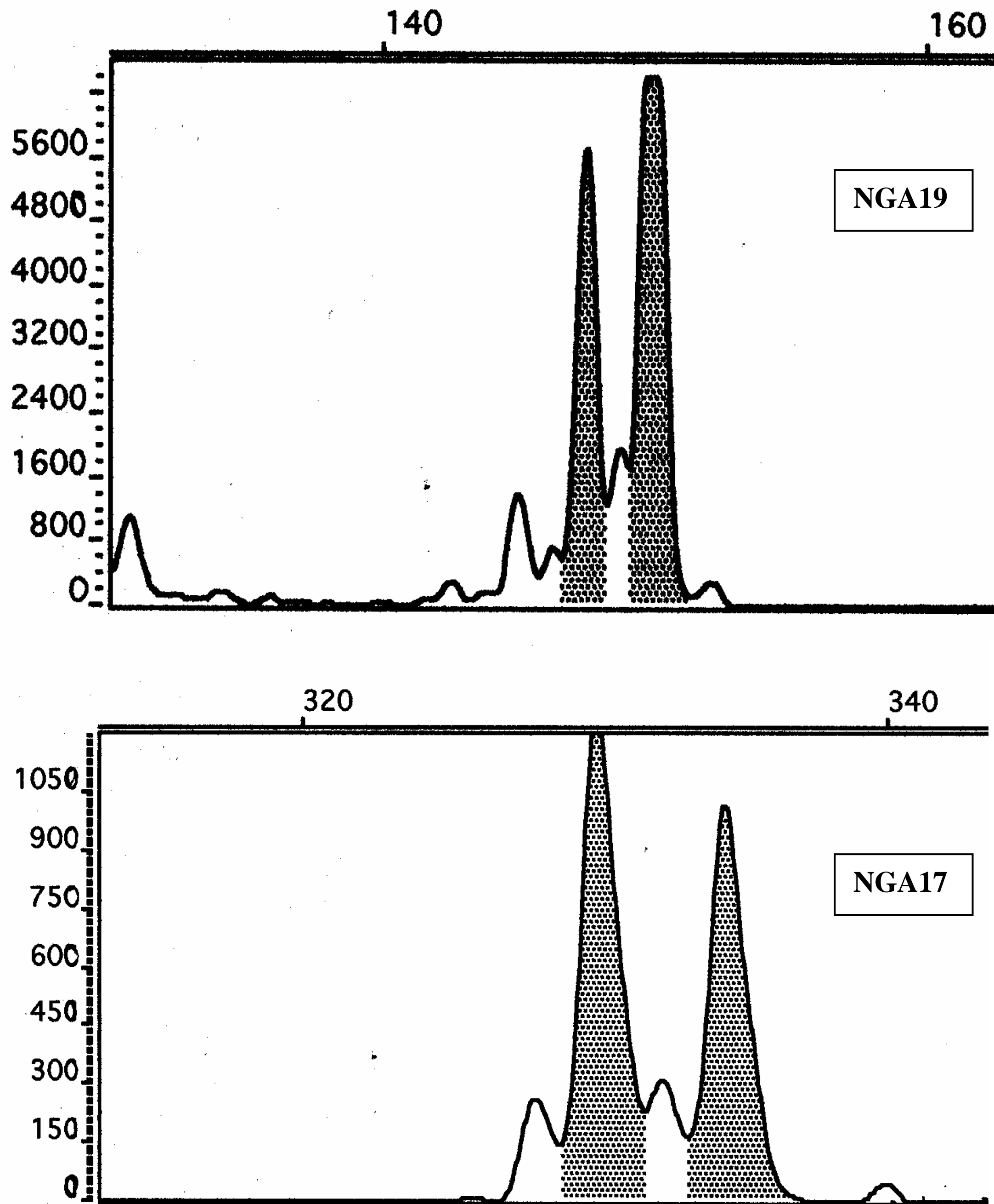


Figure 2.1
Electrophoretograms for microsatellite markers *NGA19* and *NGA17*

Microsatellite markers are analysed by capillary electrophoresis. Results show graphical outputs from the ABI PRISM 310 Genetic Analyzer Peaks (about 145 bp for NGA19 and about 330 bp for NGA17).

Table 2.2 *MYOC* polymorphisms: genotyping methods and PCR information (sequences, product size, cycle number and key reaction components)

Genotyping method*	Marker	Ref. No	Alleles	Primer	Sequence (5'>3')**	Tm (°C)	Product size (bp)	PCR conditions		
								Cycle no.	$MgCl_2$ (mM)	Primer (µM)
GeneScan	NGA17	S1	3 - 7	Myocpm-F	FLU-GGCTGTTATTTTTCTCTGT	53	~330	38 cycles	1.5	0.3
				Myocpm-R	TGCCAGCAAGATTCTTAGAA			Touch Down		
GeneScan	NGA19	S3	3 - 8	Myoc3pm-F	FLU-GTTGGGAGATGTGATTGCAG	57	~145	38 cycles	1.5	0.3
				Myoc3pm-R	AGATGGAGGTGGGAAAGTGT			Touch Down		
Comparative Thresholds	rs12076134	S2	G	MY12076134ASLG	GCG GGC CGG GCG GC TAC GCC TCA GTT ATC GCA T**G**	58	66	40 cycles	2.0	0.2
(ASPCR with LNA)			T	MY12076134ASLT	GCG GGC CTA CGC CTC AGT TAT CGC AT**T**		67	Touch Down		
				MY12076134RC	TAC AAC AGC CCT ACT ACC CAA AAC					
Comparative Thresholds	rs6425356	S7	C	MY6425356ASLT	ACA TCT TTT CCA TTA TAC TCA TC**G**	58	81	40 cycles	1.5	0.1
(ASPCR with LNA)			T	MY6425356ASLC	ACA CAT CTT TTC CAT TAT ACT CAT C**A**		80	Touch Down		
				MY6425356FC	GAC CCT CTC TCA AAA ACA AAA C					
Comparative Thresholds	rs10737323	S8	A	MY10737323ASLG	GCG GGC CGG GCG GC TGA TCC CAT GCA TTT AAT AAA ACC A**A**	60	120	45 cycles	2.5	0.3
(ASPCR with LNA)			G	MY10737323ASLT	GCG GGC ATC CCA TGC ATT TAA TAA AAC CA**G**		119	Touch Down		
				MY10737323RC	AAC TAC TAT GGG GGT AGA AGA ACC G					
Comparative Thresholds	rs235858	S9	C	MY235858ASLC	GCG GGC CGG GCG GC GTT CAG GAG TAA TGA CTA GG**C**	58	95	40 cycles	2.0	0.2
(ASPCR with LNA)			T	MY235858ASLT	GCG GGC AGT TCA GGA GTA ATG ACT AGG **T**		95	Touch Down		
				MY235858RC	CAG GGC TTT AAA TTC CAA CTC					
Comparative Thresholds	rs743994	S10	G	MY743994ASLG	GCG GGC CGG GCG GC TGG GGC TAA GGT GGT T**G**	60	167	40 cycles	2.5	0.2
(ASPRCR with LNA)			A	MY743994ASLA	GCG GGC CTG GGG CTA AGG TGG TT**A**		168	Touch Down		
				MY743994RC	TGG AGA GGA CTC AAG CAA TGG AC					
Tm-Shift	rs1602244	S5	T	MY1602244ASLC	[GCG GGC CGG GCG GC] TGG TGC ATT CAA GGA GAG **C**	58	59	35 cycles	2.0	0.2
(ASPCR with LNA)			C	MY1602244ASLT	[GCG GGC] CTG GTG CAT TCA AGG AGA G**T**		60	Conventional		
				MY1602244RC	TCC CAC AAA TGT CCC TGA TCC AC					
Tm-Shift	rs171000	S6	G	MY171000ASLG	[GCG GGC CGG GCG GC] GTT GGT TGT AGT GTT AAA CAG **G**	58	59	35 cycles	2.0	0.2
(ASPCR with LNA)			T	MY171000ASLT	[GCG GGC] GGT TGG TTG TAG TGT TAA ACA G**T**		60	Conventional		
				MY171000RC	CAG TCT CGG GTA TGC CTT TAT C					
PE with DHPLC	rs1602244	S5	T	MYpe1602244AST	CTG GTG CAT TCA AGG AGA GTA TC	58	186	38 cycles	1.5	0.3
			C	MYpe1602244ASC	TGG TGC ATT CAA GGA GAG CCT GC			Conventional		

(Continued on next page)

Table 2.2 (continued)

Genotyping method*	Marker	Ref. No	Alleles	Primer	Sequence (5’>3’)**	Tm (°C)	Product size (bp)	PCR conditions Cycle no.	MgCl$_2$ (mM)	Primer (µM)
				MYrs1602244PER	CAC TCT ACT CAA GCT TCC TC		21			
PE with DHPLC	rs6425356	S7	T	MYpe6425356F	GGA GGT TGC CAT AAG CTG AGA TT	58	264	38 cycles	1.5	0.3
			C	MYpe6425356R	TCT CGG GTA TGC CTT TAT CAG TG			Conventional		
				MYrs6425356-PER	TTA ACA CAT CTT TTC CAT TAT ACT CAT C		29			
RFLP by BanII	rs2421853	S4	G	MYOCrs2421853F6	CAG ATT TAC CTG GGT GCA ATG G	58	1038	30 cycles nested	1.5	0.3
(CC sample)			A	MYOCrs2421853R6	TCT CTC CAT AAC CTG CTG CAT C			(1st PCR)		
				MYOCrs2421853R3	GCT AGT CTT GAA CTC CTG ACC	58	605	35 cycles nested	1.5	0.3
				MYOCrs2421853F3	GGC AGG AGG ATT GTT TGA GG			(2nd PCR)		
RFLP by HphI	rs6425356	S7	T	MYOCrs6425356F3	GGA GGT TGC CAT AAG CTG AGA TTA	60	244	40 cycles	2.0	0.2
(Family sample)			C	MYOCrs6425356R3	GTG GCA TTA AAA TGG AGT CAT ACA CC			Touch Down		
RFLP by HinfI	rs743994	S10	G	MYOCrs743994F3	TTG CTG TAG TCA GTT GTA TTT AAA TTA ATA AAC	60	265	40 cycles	2.0	0.2
(Family sample)			A	MYOCrs743994R3	$(A)_{13}$ ACT TTA AAA ACA CTG TCA ATA AAT GTG AGA CT			Touch Down		
For confirming the genotypes of representing samples										
Direct cycle sequencing	rs12076134	S2	G	MY12076134Fseq	CAT CCA CAC ACC ATA CTT GCC	58	334	38 cycles	2.0	0.2
			T	MY12076134RC	TAC AAC AGC CCT ACT ACC CAA AAC			Conventional		
Direct cycle sequencing	rs1602244	S5	T	MY1602244Fseq	CCC TTC CAC CTA TGA GTC TGT AAA ATC	58	385	38 cycles	2.0	0.2
			C	MY1602244RC	TCC CAC AAA TGT CCC TGA TCC AC			Conventional		
Direct cycle sequencing	rs6425356	S7	T	MY6425356Fseq	GGA GGT TGC CAT AAG CTG AGA TT	58	225	38 cycles	1.5	0.2
			C	MY6425356RC	TCT CGG GTA TGC CTT TAT CAG TG			Conventional		
Direct cycle sequencing	rs171000	S6	G	MY171000FSeq	TGG AGG TTG CCA TAA GCT GAG A	58	153	38 cycles	2.0	0.3
			T	MY171000RC	CAG TCT CGG GTA TGC CTT TAT C			Conventional		
Direct cycle sequencing	rs10737323	S8	A	MY10737323Fseq	GGC TCT TCT GAA TTG TTC GAA GGC A	58	209	38 cycles	2.0	0.2
			G	MY10737323RC	AAC TAC TAT GGG GGT AGA AGA ACC G			Conventional		
Direct cycle sequencing	rs235858	S9	C	MYrs235858Fseq	CAG GAG TTT GAG GCT ACA GTG AG	58	297	40 cycles	2.0	0.3
			T	MY235858RC	CAG GGC TTT AAA TTC CAA CTC			Conventional		
Direct cycle sequencing	rs743994	S10	G	MY743994Fseq	CAA AAC GCA ACC TGA GAC ACT ACA	58	236	38 cycles	1.5	0.2
			A	MY743994RC	TGG AGA GGA CTC AAG CAA TGG AC			Conventional		

* FLU = fluorescein;, ASPCR = allele-specific polymerase chain reaction; LNA = locked nucleic acid; PE with DHPLC = primer extension coupled with denaturing high performance liquid chromatography;, CC samples = case-control samples; Tm = annealing temperature for PCR

** The 3’ base of the allele-specific primer is a locked nucleic acid (boldface and underlined). For Tm-shift method, the 14-base and 6-base GC-tails are shown within square brackets [ ].

2.4.2 Genotyping of SNPs of the *MYOC* locus

2.4.2.1 Allele-specific (AS) PCR

For ASPCR, we first tried the *one-tube* melting temperature (Tm)-shift method with one common primer and two allele-specific primers that were each tagged by a GC-rich tail (one 14 bases and one 6 bases long) at the 5' end (Germer & Higuchi, 1999; Wang, et al., 2005). Briefly, PCR primers were designed for products ranging from 100 to 200 bp around the polymorphism. PCRs were performed with 10 ng genomic DNA in 10-μl volumes for the data shown here. PCRs were set up as follows: 0.2 μM each of the three primers, 0.2 U AmpliTaq Gold DNA polymerase, 1.5-2.5 mM $MgCl_2$, 5% dimethyl sulfoxide (DMSO), 2.5% glycerol in the presence of SYBR Green I in a LightCycler 480 Real-time PCR System according to a published protocol with minor modifications (**Table 2.2**).

2.4.2.1.1 Tm-Shift method

The genotypes of rs1602244 (S5) and rs171000 (S6) were determined based on this method (**Table 2.2**). Genomic DNA was amplified in a *single-tube* multiplex reaction with two forward allele-specific primers and a common reverse primer. Samples homozygous for allele A would be amplified with the short GC-tailed (6 bp) primer and only gave a product with lower temperature peak in a melting curve. Samples homozygous for allele B would be amplified with the long GC-tailed (14 bp) primer and only give a higher temperature peak. The heterozygous samples will be amplified with both GC-tailed primers, and the melting curves will have both peaks (**Figure 2.2**). The sequences of the primers are shown in **Table 2.2.**

2.4.2.1.2 Comparative thresholds (d*C*t) method

If the Tm difference between the two alleles was not large enough for easy genotype call or the specificity of the allele-specific primers did not allow reliable end-point readout, we used the same primers to perform a *two-tube* real-time ASPCR and called the genotypes on the basis of the difference in the threshold cycles (Ct) of the two allele-specific reactions (the so-called delta Ct or dCt method) (Wu, et al., 2003). For each sample, touchdown thermocycling of the two reactions was carried out in parallel using AmpliTaq Gold DNA polymerase in the presence of SYBR Green I in a LightCycler 480 Real-time PCR System according to a reported protocol with minor modifications (Wu, et al., 2005). In general, higher C_t values reflected lower efficiency of ASPCRs. For each SNP and each sample, two C_t values were obtained from the two ASPCRs (**Figure 2.3**). We defined the dC_t by subtracting the C_t value between the two ASPCRs of each sample (d$C_{tA\text{-}B}$ = C_{tA} - C_{tB}). Theoretically, the d$C_{tA\text{-}B}$ value for a given DNA sample may be high positive for the homozygous allele B, low negative for the homozygous allele A, and close to zero for the heterozygous allele AB.

Five SNPs were genotyped by this method: rs12076134 (S2), rs6425356 (S7), rs10737323 (S8), rs235858 (S9) and rs743994 (S10). Locked nucleic acid (LNA) was used instead of the conventional nucleotide at the 3' end of the allele-specific primer to enhance the affinity and specificity of the allele-specific PCR primers (Reynisson, et al., 2006).

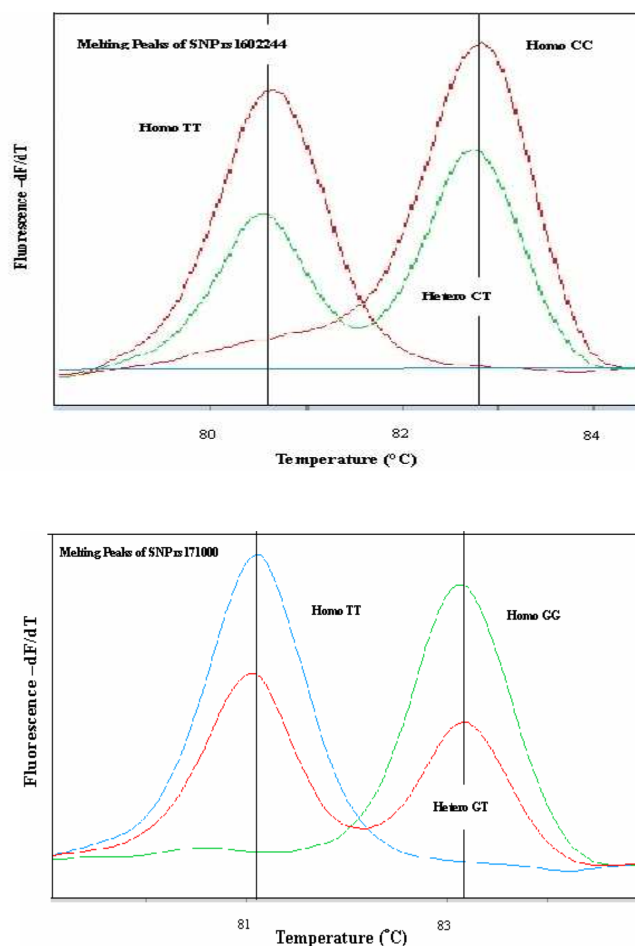
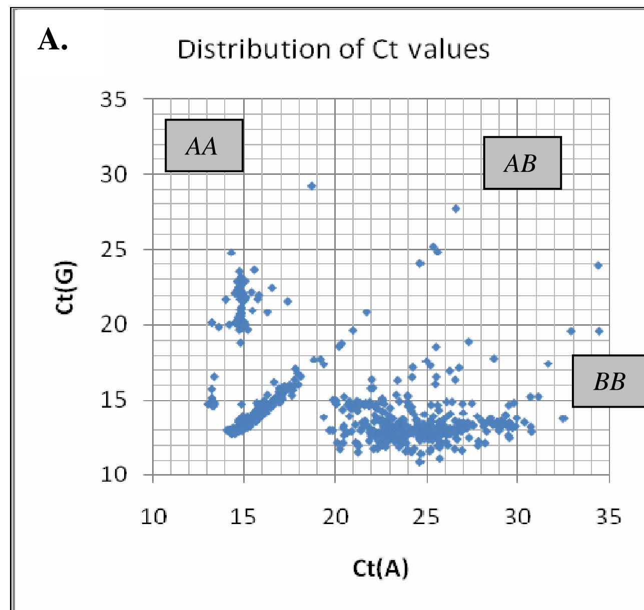


Figure 2.2
T_m-shift genotyping of SNPs

Two SNPs rs1602244 (alleles C and T) and rs171000 (alleles G and T) were genotyped with this method, and three groups of melting curves corresponding to three genotypes are seen. The y-axis is –dF/dT (negative slope of the rate of change of fluorescence signal against temperature) and the x-axis is temperature (°C). For these two SNPs, samples homozygous for allele T will be amplified with the short GC-tailed primer and only give a product with lower temperature peak (~81°C) in a melting curve. Samples homozygous for allele C (rs1602244) or G (rs171000) will be amplified with the long GC-tailed primer and only give a higher temperature peak (~83°C). The heterozygous samples will be amplified with both GC-tailed primers, and the melting curves will have both peaks.



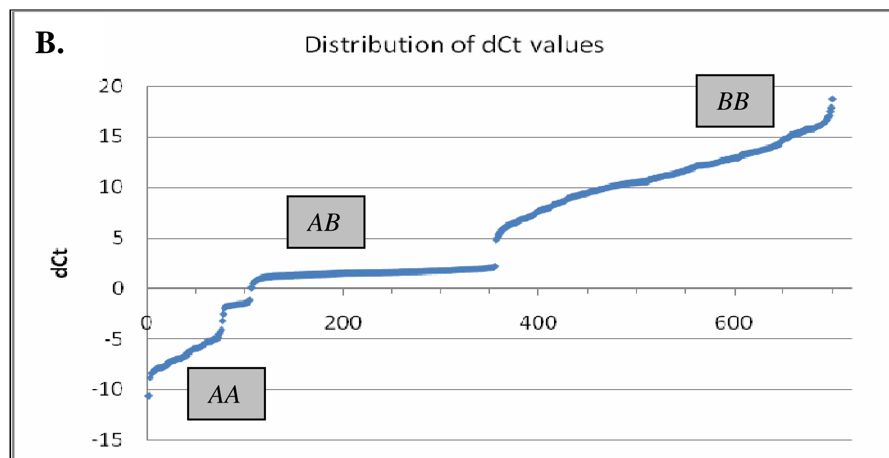


Figure 2.3
SNP genotyping by comparative thresholds (d*C*t) method

(A) Cluster analysis. The two threshold cycle (*C*t) numbers are obtained from a two-tube allele-specific PCR, and plotted against each other for individual samples. (B) The genotype is determined by the d*C*t value, which is obtained by subtracting the *C*t values between the two ASPCRs of each sample ($dC_{tA\text{-}B}=C_{tA}\text{-}C_{tB}$). High positive dC_t indicates homozygosity for allele *B* or low negative for the homozygous allele *A*, while close to zero indicates heterozygosity *AB*. These data are those of rs743994: the A allele here is the A allele and the B allele here is the G allele of this SNP.

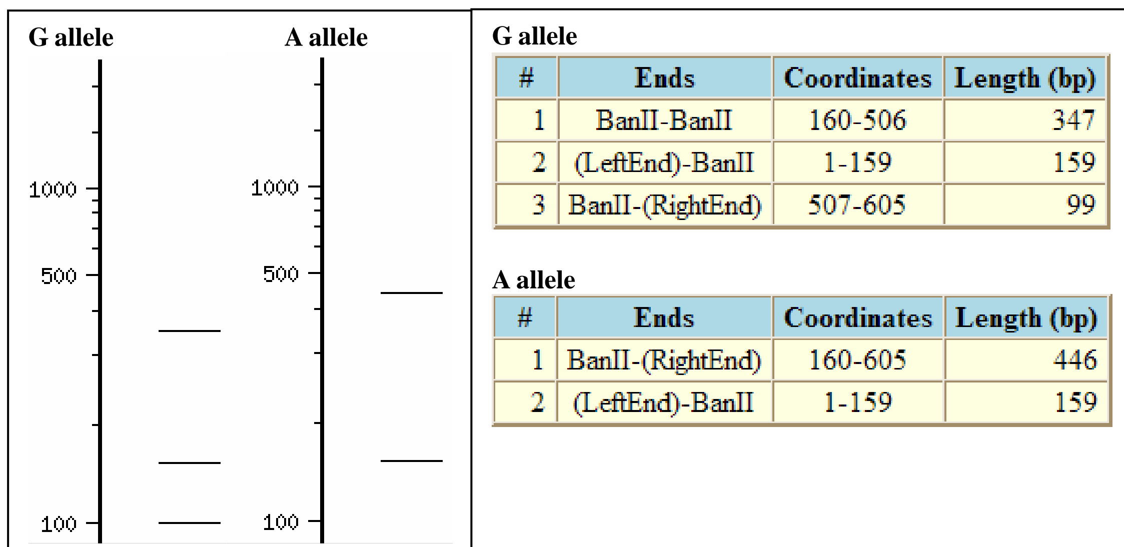
2.4.2.2 Restriction fragment length polymorphism (RFLP)

Polymorphisms of genes may be at the position of a restriction site, leading to a restriction fragment length polymorphism (RFLP) that can be detected by electrophoresis analysis.

2.4.2.2.1 RFLP analysis of rs2421853

We developed a nested PCR RFLP method for typing rs2421853 (S4) (**Table 2.2**). This SNP failed to be genotyped with traditional PCR assays because of non-specific amplification. PCR was performed in a 10-μL reaction mixture containing 10 ng of genomic DNA. Amplification was performed in 96-well plates with a GeneAmp system 9700 PCR system. The first round PCR was performed using outer primers (MYOCrs2421853F6 and MYOCrs2421853R6) specific for the rs2421853, and a 1038-bp DNA fragment was amplified. The reaction mixture contained 1.5 mM $MgCl_2$, 0.1 μM each of the forward and reverse primers, 0.2 mM each dNTP and 1× buffer and 0.1 U HotStarTaq Plus DNA Polymerase. The cycling conditions consisted of 1 cycle of initial denaturation for 5 minutes at 95°C, 30 cycles of 30 seconds at 95°C, 30 seconds at 58°C, and 30 seconds at 72°C, plus 1 cycle of final extension for 7 minutes at 72°C. In the second round PCR, a shorter 605-bp DNA fragment was amplified separately from the diluted first PCR product using 0.3 μM nested primers (MYOCrs2421853F3 and MYOCrs2421853R3), 1.5 mM $MgCl_2$, 0.2 mM each dNTP and 1× buffer, and 0.2 U HotStarTaq Plus DNA Polymerase. The thermal profile comprised 1 cycle of initial denaturation for 5 minutes at 95°C, 35 cycles of 30 seconds at 95°C, 30 seconds at 58°C, and 30 seconds at 72°C, plus 1 cycle of final extension for 7 minutes at 72°C.

Three RFLP patterns were obtained by digestion of the shorter PCR product with the restriction endonuclease BanII, and digested products were resolved in 8% polyacrylamide gels by electrophoresis. DNA fragments were electrophoresed at 130 V. The gel was stained with SYBR Green I and the bands visualized with gel doc. This enzyme was used to distinguish the A and G alleles of SNP rs2421853 by cleaving the G allele fragment into 347-bp and 159-bp fragments. All alleles discriminated based on combinations of these RFLP patterns are shown in **Figure 2.4** and **2.7**). Although the nested primers were not totally specific for the rs2421853, clear RFLP banding patterns were obtained because specificity was guaranteed by the use of the outer primers MYOCrs2421853F4 and MYOCrs2421853R4 in the first round PCR.



G allele

#	Ends	Coordinates	Length (bp)
1	BanII-BanII	160-506	347
2	(LeftEnd)-BanII	1-159	159
3	BanII-(RightEnd)	507-605	99

A allele

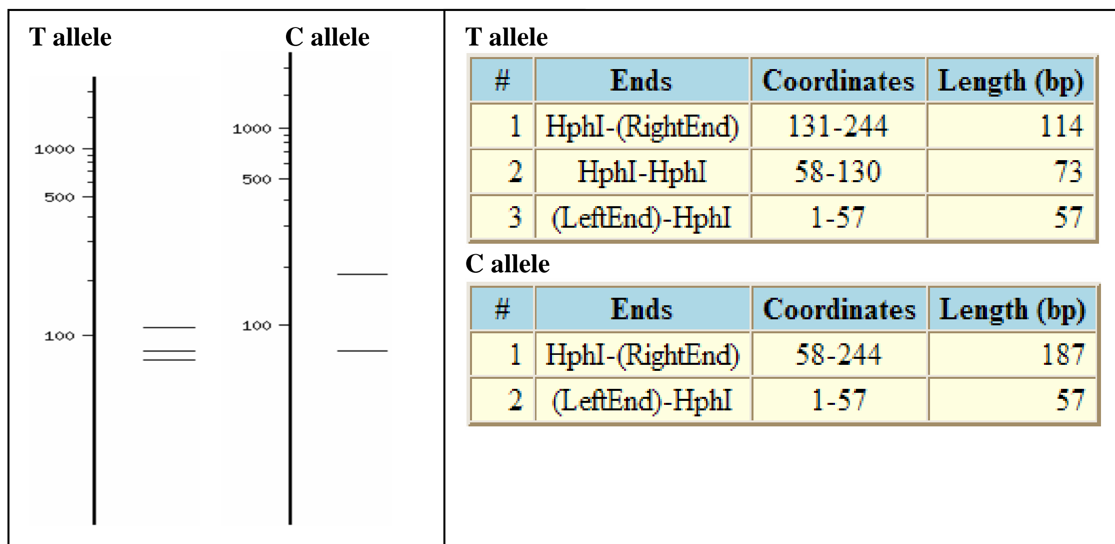
#	Ends	Coordinates	Length (bp)
1	BanII-(RightEnd)	160-605	446
2	(LeftEnd)-BanII	1-159	159

Figure 2.4
Genotyping of SNP rs2421853 by restriction fragment length polymorphism (RFLP) analysis with the use of the restriction endonuclease BanII.

The presence of the 446-bp and 159-bp fragments indicates allele A while the presence of the 347-bp, 159-bp, 99-bp fragments indicates allele G. The vertical bars on the left panel show the mobility of DNA size markers in multiples of 100 bp.

2.4.2.2.2 RFLP analysis of rs6425356 and rs743994

A confirmatory nuclear family dataset was used to test two positive SNPs, rs6425356 and rs743994, found in the case-control dataset. The genotyping method was RFLP with PAGE in 163 nuclear families **(Figures 2.5, 2.6** and **2.7)**. PCR was performed in a 10-μL reaction mixture containing 10 ng of genomic DNA, 2.0 mM $MgCl_2$, 0.2 μM each of the forward and reverse primers, 0.2 mM each dNTP and 1× buffer, and 0.2 U HotStarTaq Plus DNA Polymerase. Amplification was performed in 96-well plates with a GeneAmp system 9700 PCR system, consisting of 1 cycle of initial denaturation for 5 minutes at 95°C, 40 cycles of 30 seconds at 95°C, 30 seconds at the annealing temperature (**Table 2.2**), and 30 seconds at 72°C, plus 1 cycle of final extension for 7 minutes at 72°C. Restriction digestion was carried out overnight (at least 20 hours) according to the manufacturer's instructions. Digested products were separated by electrophoresis in horizontal 102-well polyacrylamide gels.



T allele

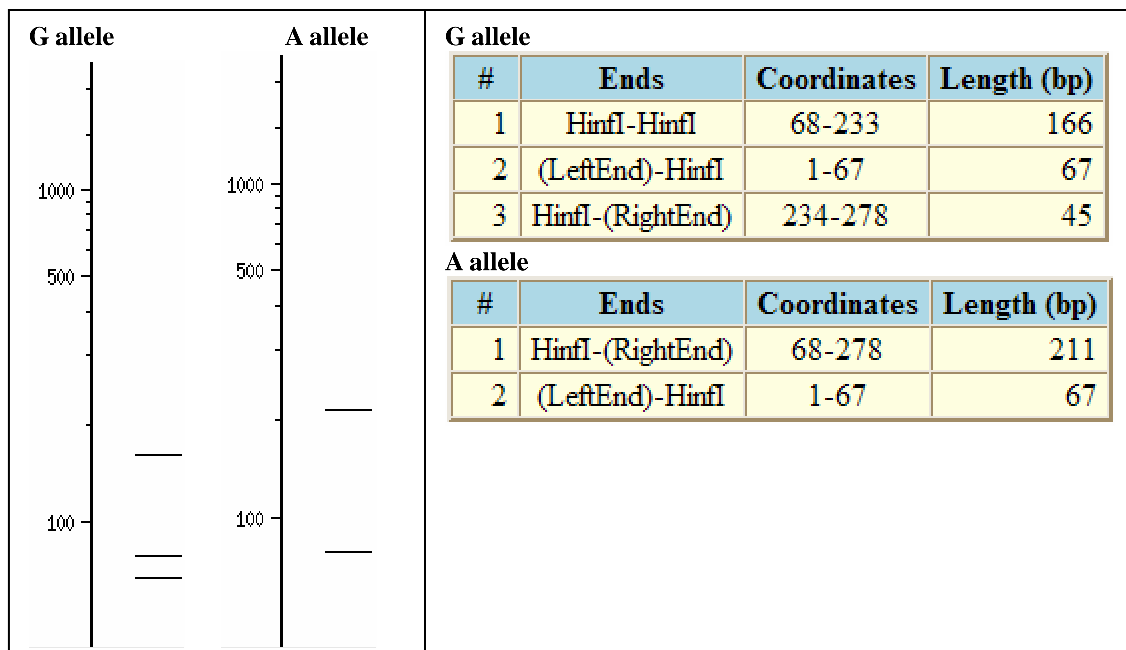
#	Ends	Coordinates	Length (bp)
1	HphI-(RightEnd)	131-244	114
2	HphI-HphI	58-130	73
3	(LeftEnd)-HphI	1-57	57

C allele

#	Ends	Coordinates	Length (bp)
1	HphI-(RightEnd)	58-244	187
2	(LeftEnd)-HphI	1-57	57

Figure 2.5
Genotyping of SNP rs6425356 by restriction fragment length polymorphism (RFLP) analysis with the use of the restriction endonuclease HphI.

The presence of the 187-bp and 57-bp fragments indicates allele C while the presence of the 114-bp, 73-bp, 57-bp fragments indicates allele T. The vertical bars on the left panel show the mobility of DNA size markers in multiples of 100 bp.



G allele

#	Ends	Coordinates	Length (bp)
1	HinfI-HinfI	68-233	166
2	(LeftEnd)-HinfI	1-67	67
3	HinfI-(RightEnd)	234-278	45

A allele

#	Ends	Coordinates	Length (bp)
1	HinfI-(RightEnd)	68-278	211
2	(LeftEnd)-HinfI	1-67	67

Figure 2.6
Genotyping of SNP rs743994 by restriction fragment length polymorphism (RFLP) analysis with the use of the restriction endonuclease HinfI.

The presence of the 211-bp and 67-bp fragments indicates allele A while the presence of the 166-bp, 67-bp, 45-bp fragments indicates allele G. The vertical bars on the left panel show the mobility of DNA size markers in multiples of 100 bp.

Figure 2.7 RFLP with three SNPs for *MYOC* gene

Sample	SNP	Enzymes	PAGE	Polymorphic band in bp*	Pattern
Case & Control	rs2421853	BanII	5%	**GT**: 446, 347, 159, (99) **GG**: 347, 159, (99) **TT**: 446,159	GG GT TT LADDER
Family	rs6425356	HphI	10%	**CT**: 187,114,73,(57) **CC**: 187, (57) **TT**: 114, 73, (57)	LADDER CT CC TT
Family	rs743994	HinfI	8%	**GT**: 211,166, 67, (45) **GG**: 166, 67, (45) **TT**: 211,67	LADDER GG GA AA

* Bands with size shown in brackets ( ) are not obvious on the gel

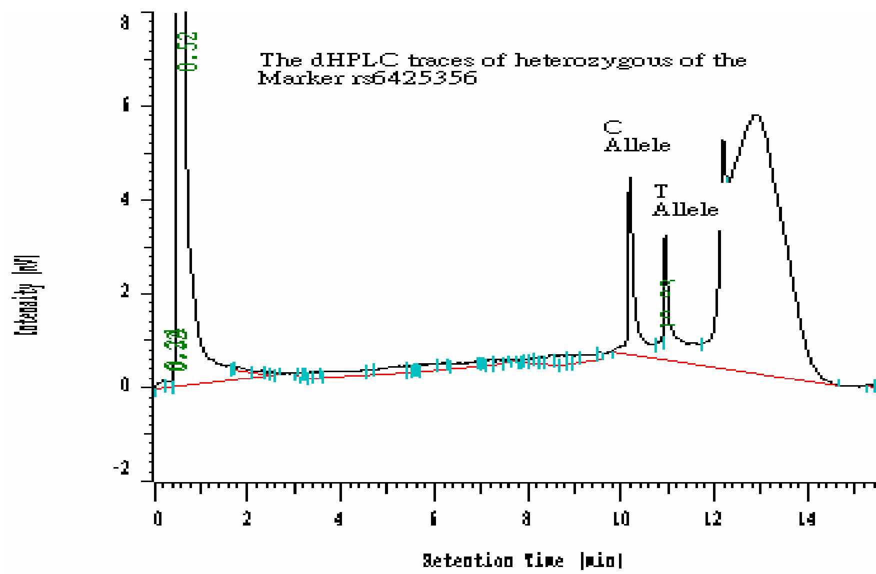
2.4.2.3 Primer extension with DHPLC

The genotypes of two positive SNPs, rs1602244 (S5) and rs6425356 (S7), were further validated by an independent genotyping platform – primer extension coupled with denaturing high-performance liquid chromatography (PE-DHPLC) (Yip, et al., 2003). PCR was performed in a 10-μL reaction mixture containing 20 ng of genomic DNA, 1.5 mM $MgCl_2$, 0.3 μM each of the forward and reverse primers, 0.2 mM each dNTP and 1× Gold buffer, and 0.2 U of AmpliTaq Gold DNA polymerase. PCR was performed in 96-well plates with a GeneAmp system 9700 PCR system, including 1 cycle of initial denaturation for 5 minutes at 95°C, 40 cycles of 30 seconds at 95°C, 30 seconds at the annealing temperature (**Table 2.2**) and 30 seconds at 72°C, plus 1 cycle of final extension for 7 minutes at 72°C. The thermal cycling condition for PE was as follows: an initial denaturation of 2 minutes at 95°C, followed by 50 cycles of 5 seconds at 95°C, 5 seconds at 43°C and 5 seconds at 60°C. The WAVE DNA Fragment Analysis System was used to analyse the extension products. (**Figure 2.8**) PE products were denatured by heating at 96°C for 1 minute and then cooling in ice over a period of 10 minutes. Then, 10 μl of PE products were automatically loaded into the DNASep column and eluted on a linear acetonitrile gradient in a 0.1 M TEAA (pH 7) with a constant flow rate of 0.9 ml/min. The gradient was combined by buffer A (0.1 M TEAA) and buffer B (25% acetonitrile in 0.1 M TEAA). Different DNA segments were separated by changing the concentration of buffer A relative to that of buffer B. The gradient profile of proportion in buffer B was adjusted according to the size of the PCR amplicons for the 2 markers.

The DHPLC system automatically processes the samples at a rate of 6 (S5) and 12 (S7) minutes per sample. UV detection was set at 260 nm. After each run, we washed

the column with 25% acetonitrile for 1 minute and equilibrated it for 1 minute before the next sample injection. Optimized peak separation between the 2 markers can be achieved with the selection of temperature and acetonitrile gradient. **(Figure 2.8)** A DHPLC analysis was performed at optimal 70°C and 20% buffer B starting concentration. DHPLC data analysis was based on chromatograms profile for samples with retention time and height of peaks. The DNA segments of each SNP marker were eluted under essentially the same conditions of temperature and solvent gradient. Each SNP has a corresponding elution profile, or signature, at a given set of elution conditions of temperature and gradient.

(A)



(B)

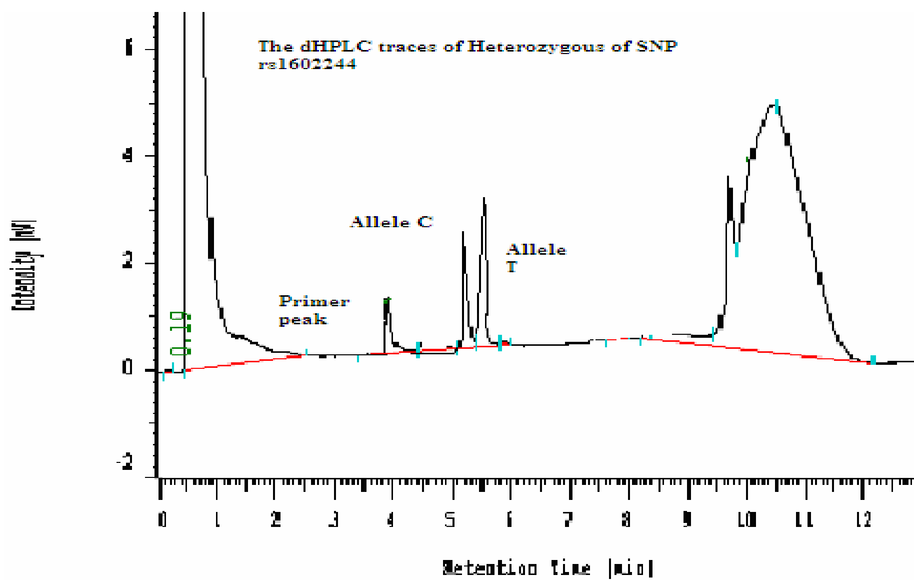


Figure 2.8
Genotyping of SNP rs6425356 and rs1602244 by primer extension coupled with denaturing high performance liquid chromatography (PE-DHPLC).

The diagram shows the elution profiles of DHPLC. The elution times of PE products are longer for rs6425356 (10-12 min; A) than rs1602244 (5-6 min; B).

2.5 Statistical analysis of *MYOC* data

2.5.1 Case-control data

2.5.1.1 Hardy-Weinberg equilibrium (HWE)

Consistency of the genotype frequencies with the Hardy-Weinberg equilibrium was tested using the GenAssoc package (Cordell & Clayton, 2002) executed within the software STATA (version 8.2). Genotypes were tested for controls and cases separately. Deviations from HWE can indicate inbreeding, genetic drift, founder effect, population stratification. However, genotyping error is an important cause of deviation from HWE (Gomes, et al., 1999). In samples of affected individuals, these deviations may indicate association. Tests of HWE are commonly performed using a simple χ^2 goodness-of-fit test (Wigginton, et al., 2005). Moderately high values for HWE χ^2 tests are indicative of genetic association to a susceptibility association, while extremely high values may indicate genotyping errors. Marker loci close to or inside susceptibility genes will tend to show larger values (Gordon, et al., 2001).

2.5.1.2 Tests of association: single markers

Three genetic models were tested for each polymorphism: genotypic, additive and allelic models. Chi-squared analysis was used to compare the allele frequencies and genotype frequencies of *MYOC* gene polymorphisms between the case and the control groups. Armitage trend test was used to test the additive effect of the risk allele (the minor allele) across the genotypes in cases and controls. ORs were calculated to estimate the effect size with reference to the more frequent homozygote or allele, together with the corresponding 95% confidence intervals. Association analysis was performed with the GenAssoc package (Cordell & Clayton, 2002)

executed within the STATA software.

Multiple comparisons between cases and controls were corrected by means of false discovery rate (FDR) at a level of 0.05. FDR is defined as the proportion of true null hypothesis that is rejected out of the total number of null hypotheses rejected (Benjamini & Yekutieli, 2005; Hochberg, 1995). Correction for multiple comparisons by controlling FDR is a *post hoc* maximizing procedure, and is more powerful than the conventional Bonferroni procedure. There were a total of 30 comparisons: 10 polymorphisms each tested for 3 genetic models. The observed P values were sorted from smallest to largest: $P_1 \leq P_2 \ldots \leq P_n$, where n=30. Starting from the largest P value (P_n), comparison was made between P_j and the value of $(0.05 \times j/n)$, and continued as long as $P_j > (0.05 \times j/n)$. If k was the *first* time when P_k was *less than* or *equal to* $(0.05 \times k/n)$, then the hypotheses corresponding to the smallest k observed P values were declared as significant (Benjamini, et al., 2001; Benjamini & Hochberg, 1995).

2.5.1.3 Tests of main effects by stepwise logistic regression

For polymorphisms found to be significantly associated with high myopia after adjustment for multiple testing, stepwise logistic regression was carried out using STATA (version 8.2) to investigate whether any of these polymorphisms could account for the positive effects of the other polymorphisms (Cordell & Clayton, 2002). Both forward and backward procedures were used. In this way, a final model could be generated in which all polymorphisms in the model could be regarded to have a significant effect on the development of high myopia. Each of these polymorphisms in the final model was therefore likely to be genuinely functional or

to be in LD with another functional polymorphism that had not been typed.

In the forward procedure, for each positive polymorphism in turn, the effect of adding another positive polymorphism to the model was considered. If at least one of the positive polymorphisms gave a significant improvement in fit over the null model, this most significant polymorphism was added to the current model. Then, the effect of adding additional polymorphisms was each considered again, and the one giving the greatest improvement in fit was added to the current model. This procedure was repeated until no more polymorphisms gave a significant improvement in fit.

In the backward procedure, a model including all positive polymorphisms was first considered. Each polymorphism was then deleted in turn from this full model, and the one giving the least significant worsening in fit was removed from the current model (full model in the first round). In the second stage, each polymorphism still in the model was again deleted in turn, and the one giving the least significant drop in fit was again removed from the current model. This backward procedure was continued until no more polymorphisms could be removed without significantly weakening the fit.

2.5.1.4 Tests of association: multiple SNPs by building *MYOC* haplotype blocks

SNPs are relatively uninformative individually, i.e. more than one is required to obtain an amount of information equivalent to more informative markers, such as MSs. One way to increase the information from SNPs is to use haplotypes

constructed from multiple SNPs, which is more powerful for detecting an association than using all SNPs individually. Instead of analysing haplotypes for SNPs within an LD block or haplotypes for sliding windows of varying size, haplotypes were constructed and analysed only the SNPs found by stepwise logistic regression to contribute main effects to the association with high myopia. Haplotype analysis was executed with the software PLINK (Purcell et al., 2007b) with correction of multiple comparisons by permutations. Haplotype frequencies were estimated by PLINK using expectation-maximization algorithm.

2.5.1.5 Linkage disequilibrium (LD) patterns

Assume that one locus has two alleles *A* and *a*, and a second locus has two alleles *B* and *b*. Two-locus Lewontin's LD parameter (D) is defined for two alleles at different loci as $D_{AB} = P_{AB} - (P_A \times P_B)$, where P_{AB} is the observed frequency of haplotype consisting of allele *A* at one locus and allele *B* at another locus, and P_A and P_B are the frequencies of alleles *A* and *B* at the respective loci (Ardlie, et al., 2002). The correlation coefficient r^2 is defined as $D^2/(P_A \times P_a \times P_B \times P_b)$. The biallelic loci, all four possible allele-pair LD statistics are equivalent although the signs may be different. For multiallelic loci, the locus-pair LD statistics are defined as a weighted mean of the absolute allele-pair LD statistics: $LD = \Sigma\Sigma(P_j\ P_k\ |LD|)$, where LD can be D or r^2, j is one allele of one locus and k is one allele of another locus, and P_j and P_k are their corresponding allele frequencies. In this study, only r^2 was presented because it is the measure used to select tag SNPs by Tagger (de Bakker, 2009).

LD statistics were calculated and plotted by the software Haploview for biallelic polymorphisms (SNPs) (Barrett et al., 2005). LD blocks were constructed using

Gabriel's definition (Gabriel, et al., 2002). On the other hand, multiallelic LD statistics were calculated using the software PowerMarker (Liu & Muse, 2005) for multiallelic polymorphisms, i.e., MSs.

2.5.2 Tests of association: Family samples with two positive SNPs of the *MYOC* locus

The four *MYOC* polymorphisms (NGA17, NGA19, rs2421853 and rs235858) that were found associated with high myopia in our previous family study (Tang, et al, 2007) gave negative results in the present case-control study (see Chapter 3 for details). Two positive SNPs rs6425356 (S7) and rs743994 (S10) that gave the most significant results in the present study were further investigated with the family samples. The genotypes of the parents were tested for HWE. Family-based association analysis was conducted using Family-based Association Test (FBAT) package (Horvath, et al., 2001; Rabinowitz & Laird, 2000). This is a generalized algorithm derived from the original transmission disequilibrium test (TDT). A generalized score is used to perform a variety of tests similar in spirit to TDT. FBAT first defines a test statistic that reflects association between the trait under study (high myopia in this study) and the polymorphism. It then computes the distribution of the test statistic under the null hypothesis by treating the offspring genotypes as random, and conditioning on the observed trait and the parental genotype distribution. FBAT is not affected by biases resulting from population stratification, mis-specification of the trait distribution, the presence of multiple siblings in a family, and/or selection based on trait. The tool FBAT was also used in the original study by Tang et al (2007).

2.6 Testing *MYP2* candidate genes by a DNA pooling approach

To investigate the genetic polymorphisms underlying common multifactorial myopia trait, whole genome association studies using SNPs requires genotyping of hundreds of thousands to a million markers in a large number of cases and controls. A DNA pooling approach has been used to reduce the amount of genotyping work in very large population sample sets and then select putative genes for subsequent study (Darvasi & Soller, 1994). The approach has been demonstrated to be a reasonably accurate and powerful screening tool as an alternative to individual genotyping to detect the allele frequency differences between two groups (Daniels, et al., 1998). Primer extension coupled with DHPLC is one of the efficient platforms for estimating allele frequencies in this kind of study (Hoogendoorn et al., 2000). In order to make an efficient scan of a large candidate region with greatly reduced genotyping cost and time for suggestive evidence of genetic association, the use of DNA pools consisting of equal amounts of genomic DNA from individual samples was proposed in this study.

To evaluate the genetic contribution of the *MYP2* locus to high myopia, 62 SNPs from 7 positional candidate genes within the locus were tested for association by using 2 sets of pools containing DNA prepared respectively from individuals with high myopia and from emmetropic controls. The genes (or markers) showing a significant difference in allele frequencies between the case pools and the control pools were further investigated by sample-by-sample genotyping for their

susceptibility to high myopia.

A complete DNA pooling experiment consists primarily of three stages. The **first stage** is a pilot study in which heterozygous individuals are collected to estimate the coefficient of preferential amplification (i.e., the k correction factor). The coefficient is subsequently used to correct the estimates of allele frequencies in the second stage. In the **second stage**, DNA pooling experiments are conducted for a large number of SNPs, and pooling association tests are carried out to screen for potential genetic markers. Only a small proportion of markers selected from the second stage experiments are included in the **third stage** in which all individuals are genotyped to confirm the validity of the markers selected from the second stage. As a consequence of the preliminary screen in the second stage, the number of SNPs in the third stage is drastically reduced, thereby lowering genotyping costs.

2.6.1 Pool construction

Genomic DNA extracted from whole blood was quantified again by using spectrophotometry at 260 nm with Victor3™ V 1420 Mutilabel Counter before construction of DNA pools. The DNA samples were then diluted to a standard concentration (10 ng/μl).

DNA samples were further accurately quantified again using the Quant-iT PicoGreen dsDNA Assay Kit according to the manufacturer's instructions. The fluorescence readings were measured using Victor3™ V 1420 Mutilabel Counter with a 480 nm excitation filter and a 520 nm emission filter. A standard curve was constructed from

a series of dilutions of the Lambda DNA standard provided in the kit, and used to determine the concentration of the DNA samples. The accurately quantified DNA samples were then further diluted to a final target concentration of 5 ± 0.5 ng/μl, and re-quantified with the PicoGreen assay. Any DNA samples outside of the target concentration range were re-adjusted and re-quantified. This was followed by pooling equal volumes (30 μl) of DNA from 50 subjects of the same affection status (case or control) to construct a DNA pool. Finally, 6 distinct DNA pools were prepared from 300 cases (H1 to H6), and 6 distinct DNA pools from 300 controls (E1 to E6) for subsequent analysis (**Table 2.3**). The pool size has been suggested to influence the accuracy of the estimates. Smaller pool size with 50 individuals may control the error to a negligible level (Barratt et al., 2002). In addition, replication measurement of the pools was suggested to control for measurement error.

2.6.2 Tag SNPs of candidate genes

In total, 62 tag SNPs (**see Table 2.4**) were selected from HapMap Chinese data (corresponding to samples with Han Chinese ancestry) (Frazer, et al., 2007) using Haploview software (Barrett et al., 2005) from seven potential candidate genes based on location in the *MYP2* locus. The coverage of the set was limited to SNPs with a pairwise $r^2 \geq 0.8$ and minor allele frequencies over 0.1. SNP annotation was based on the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov/) dbSNP database, ref-SNP, Build 118. Genomic annotation was based on NCBI Genome Build 34. Gene annotation was based on Entrez Gene entries for which NCBI provided positions on the Mapview FTP site. The locations of these SNPs are shown in **Table 2.4.**

Table 2.3 Construction of DNA pools*

Code of DNA pools	Samples
H1	H001-H050
H2	H051-H100
H3	H101-H150
H4	H151-H200
H5	H201-H250
H6	H251-H300
E1	E001-E050
E2	E051-E100
E3	E101-E150
E4	E151-E200
E5	E201-E250
E6	E251-E300

* Equal volumes of 50 DNA samples (5±0.5 ng/μl) are mixed to form each pool.

Table 2.4 Characteristics of 62 tag SNPs selected from 7candidate genes in the *MYP2* region

No	Candidate gene	SNP	Position	Alleles
1	*CLUL1*	rs546696	609050	T/G
2	*CLUL1*	rs1004961	610709	A/C
3	*CLUL1*	rs505140	614631	T/G
4	*CLUL1*	rs485562	619035	A/G
5	*CLUL1*	rs11662827	621398	C/G
6	*CLUL1*	rs11661043	621526	C/T
7	*CLUL1*	rs11660005	625288	A/G
8	*CLUL1*	rs11663153	631024	A/C
9	*CLUL1*	rs2342700	634527	C/G
10	*CLUL1*	rs9966612	639311	A/G
11	*EMILIN2*	rs4797088	2863848	C/G
12	*EMILIN2*	rs680173	2864364	G/C
13	*EMILIN2*	rs7226712	2864552	C/T
14	*EMILIN2*	rs637647	2871839	T/C
15	*EMILIN2*	rs604050	2883413	T/A
16	*EMILIN2*	rs6506038	2888535	A/G
17	*EMILIN2*	rs16944003	2889324	C/T
18	*EMILIN2*	rs1790994	2903431	C/T
19	*EMILIN2*	rs1059281	2903896	C/G
20	*LPIN2*	rs1985	2907223	T/A
21	*LPIN2*	rs16944051	2912716	C/T
22	*LPIN2*	rs3819090	2916402	G/T
23	*LPIN2*	rs10460009	2938029	C/T
24	*LPIN2*	rs1628891	2938632	A/G
25	*LPIN2*	rs2298786	2957456	T/C
26	*LPIN2*	rs589318	2973942	T/C
27	*LPIN2*	rs16944193	2994756	C/T
28	*MYOM1*	rs1042731	3056944	T/C
29	*MYOM1*	rs4613146	3067464	A/C
30	*MYOM1*	rs7233983	3067569	A/C
31	*MYOM1*	rs4413045	3078903	C/T

(Continued on next page)

Table 2.4 (continued)

No	Candidate gene	SNP	Position	Alleles
32	*MYOM1*	rs11081004	3086165	A/G
33	*MYOM1*	rs6506057	3086717	A/G
34	*MYOM1*	rs1071600	3116811	G/A
35	*MYOM1*	rs9948582	3128199	A/G
36	*MYOM1*	rs6506074	3146244	C/T
37	*MYOM1*	rs9952207	3152390	A/G
38	*MYOM1*	rs4798069	3153771	C/T
39	*MYOM1*	rs7235847	3161224	A/G
40	*MYOM1*	rs9951849	3161473	G/T
41	*MYOM1*	rs7238703	3170709	C/T
42	*MYOM1*	rs8091916	3182682	A/C
43	*MYOM1*	rs4340411	3199224	T/C
44	*MYOM1*	rs8090956	3200634	C/T
45	*MYOM1*	rs1662315	3205230	T/C
46	*MYOM1*	rs9947162	3207184	A/G
47	*MYOM1*	rs4441365	3207622	T/C
48	*MYOM1*	rs4507002	3209591	G/A
49	*MYOM1*	rs12605942	3209851	A/G
50	*MYCL12B*	rs1662347	3240326	C/A
51	*MYCL12B*	rs717183	3242924	G/A
52	*MYCL12B*	rs3786458	3244460	T/A
53	*MYCL12B*	rs1662342	3245301	G/A
54	*MYCL12A*	rs1662336	3253391	G/A
55	*MYCL12A*	rs949303	3254349	A/G
56	*MYCL12A*	rs1630702	3257808	C/T
57	*MYCL12A*	rs1791067	3258550	C/T
58	*MYCL12A*	rs7239576	3259893	C/T
59	*MYCL12A*	rs6506094	3261359	A/C
60	*ZFP161*	rs2789	5279887	C/G
61	*ZFP161*	rs990072	5280198	C/T
62	*ZFP161*	rs620652	5282030	G/A

Table 2.5 Primers (PCR and PE) and PCR product size for tag SNPs from *MYP2* locus

Gene	Tag SNP	Primer name*	Primer sequence (5'>3')	Product size (bp)
CLUL1				
1	rs546696	CLU5466F	GGT CTC GTG CCT CCT TAT TAG	
	G/T	CLU5466R	TCA GGG TGC TCA TTA GAT TGG TG	558
		CLU5466PE	TGG TCC GTG TGG CAG CTG	
2	rs1004961	CLU1004F	ATA ATA AAA TCC ATC AGT CTA CCT	
	A/C	CLU1004R	TTA AAC TGC AAG CTA TTC AAT GT	392
		CLU1004PE	GCT ATA GCT TGA ATT ACA TAT TTT ATC	
3	rs505140	CLU5051F	GGC ATC TTC ATT GTC AGG TCA C	
	T/G	CLU5051R	GCA TAG CCT TTC TTC TTC CTC C	366
		CLU5051PER	ATC TTA TTC ATC TGT TCA TTC ATT TGT TCA	
4	rs485562	CLU4855F	GGA TTA CAG GTG CCC GCC	
	A/G	CLU4855R	CAC TTA ATA CCA CCC AGC CTG	547
		CLU4855PE	CTG AAT CAG ATT CTC AAA ATC GCC	
5	rs11662827	CLU11662F	TGG AAG GAA CAG ATG AGA TTG AGT	
	G/C	CLU11662R	CAA TCC ATC ACC TCG CTT AGT C	230
		CLU11662PE	CGG GAG ACC TGG TTC TAA T	
6	rs11661043	CLU11661F	GAG AAA CAG ACA GGC AGG T	
	T/C	CLU11661R	GAG GCT GTG TCA TAA TCA TCA	237
		CLU11661PE	GAA GGG TTC TGA GGT TCT G	
7	rs11660005	CLU11660F	AAA GAT GGT GGG GCA GGG G	
	G/A	CLU11660R	ATG ACT GCC TGA GCC CCG	167
		CLU11660PER	GCA AAG GGT CTA GGT TGA G	
8	rs11663153	CLU11663F	TAT GTG GTT CTT TTC CTC TTT TCC C	
	C/A	CLU11663R	CCT TTT GCC TGT GGT AGT TTT TAG T	328
		CLU11663PE	GAA GGG GCA TAG TAG GGA	
9	rs2342700	CLU2342F	GCC AAA ATC ACA CCT ACA ACC ATA AAT	
	C/G	CLU2342R	GGC TGG GTT GTG CTG AGG	457
		CLU2342PE	CCC CTT GTG GTT AAA CGT TG	
10	rs9966612	CLU9966F	TCA ATG CAT TCA GGG TCC CAA AC	
	A/G	CLU9966R	CAG CTA TGA AAC TTG GGG ATG G	298
		CLU9966PE	TCA CTG AAA CAT GAA TTC CAA TTT TAT A	
EMILIN2				
11	rs4797088	EMI4797F	AGG CAG GAG AAT GGC GTG AA	
	C/G	EMI4797R	CAG AGT GGT TCA GAG AGC TAG	317
		EMI4797PE	TGT AGC AGC CCC AAA TCT TC	
12	rs680173	EMI6801F	CTC CTC TTT CCT TCT TAC CCC C	
	C/G	EMI6801R	GCC AGA GCC CAA CCC ACC	442
		EMI6801PE	GCA CCC AGC TGT AGT TGA TTA	

(Continued on next page)

Table 2.5 (continued)

Gene	Tag SNP	Primer name*	Primer sequence (5'>3')	Product size (bp)
13	rs7226712	EMI7226F	CAT CAG TTC TTC TCT TTA TCT CGC T	
	T/C	EMI7226R	GGA AGG AAA AGG ACC CAA ACA TTA T	259
		EMI7226PE	GTT GGG CTC TGG CAC TTT	
14	rs637647	EMI6376F	CCG ATG AAC AAA ACC CGA GAG	
	T/C	EMI6376R	TAA ACC CAA GCC TCC TCC CT	388
		EMI6376PE	TGA TGC TTT CTT CCC TTT TTA CAT TT	
15	rs604050	EMI6040F	GCT TGG GGC AGT GGG TGT T	
	A/T	EMI6040R	CGC CCA GAC TCC GAC TAG A	441
		EMI6040PE	AGA GCA GTT TAG GGA CAG G	
16	rs6506038	EMI6506F	ACC CTG CCT TCC CAC TGT TC	
	A/G	EMI6506R	TGA GCA GAA AGG GGT GGG AG	211
		EMI6506PE	CCA GAT GTC ACG CTT CCA	
17	rs16944003	EMI694F	GGA TCT ACC CAC CCC AGG	
	C/T	EMI694R	AGG AGA ATT GCT TGA ACC CAG GA	380
		EMI694PE	GAA CCA CTG ACT GAC ACT G	
18	rs1790994	EMI1790F	TCG TGG TGA CTG GGG GCA A	
	T/C	EMI1790R	CTT CCA GTT GTC CAG TTA GAG TG	338
		EMI1790PE	GGA GAT GTC AGG GGA AAG A	
19	rs1059281	EMI1059F	TCT GAC TGT GGG CTG GGA G	
	C/G	EMI1059R	GCT GCT GTG GGA TCT GAG TG	635
		EMI1059PE	CCG ACT TTA GTT TGG GCT GTT	
LPIN2				
20	rs1985	LIPIN1985F	CAC ACA GCC CTT CCA CAG TT	
	T/A	LIPIN1985R	ATT CAG TTT ATG TTA TGT TCG TTT ATT GTT G	483
		LIPIN1985PE	GAA AGT AAG AAA GGG AGG GG	
21	rs16944051	LIPIN169440F	AGG AAA CTG CTT AGG ATT ATA GAG G	
	C/T	LIPIN169440R	CCC CTA TAT TTT TTC TTC ACA GTG G	203
		LIPIN169440PE	TGG GAA AAA AAT GGA GGT ACA GG	
22	rs3819090	LIPIN3819F	GGA CAG GCA GAA GCA GGA AC	
	T/G	LIPIN3819R	CAG CGT ATG GGA AGG GGC	225
		LIPIN3819PER	GGA GCC AAA GGG AGA CAT A	
23	rs10460009	LIPIN1245F	GCC ACG TAA TTC CTA GTC CTC	
	C/T	LIPIN1245R	TGC TGC CAC CTG GAG GAG	417
		LIPIN1245PE	ATG GGC CCA GTC TGT CTA	
24	rs1628891	LIPIN1628F	AGC TCC TCC AGG TGG CAG	
	A/G	LIPIN1628R	GTG TAG GTG GGC AAG AAA CGT ATT	600
		LIPIN1628PE	CTC TTT TTG GTA GAT CTT GGA G	
25	rs2298786	LIPIN2298F	CTA TAC TCA GGG TCC CAA CTT TC	
	C/T	LIPIN2298R	ACC TCC TTT CGT TCT TTG TAA TCC C	408
		LIPIN2298PE	GCA AAG GTA CCC ACG CAG	
26	rs589318	LIPIN5893F	GTA TTA AGC ATC ACC AGA CAC CTA TT	

(Continued on next page)

Table 2.5 (continued)

Gene	Tag SNP	Primer name*	Primer sequence (5'>3')	Product size (bp)
	A/G	LIPIN5893R	5' CTA AAA ACC GCA AAC ACA CAA CTC	661
		LIPIN5893PE	5' GTC CTC AGT TCT TCC TTA CCT	
27	rs16944193	LIPIN169441F	GCC AAC ACT TAG GGA AAA CAG AAA	
	C/T	LIPIN169441R	GGA AGA TGG GGA GGT GAT GT	514
		LIPIN169441PE	GCC CGA GAG CAA TAT TTA GG	
MYOM1				
28	rs1042731	MYOM1042F	CCT GTG TGA CTT GGG TGT GA	
	A/G	MYOM1042R	AAG CCA CCC AAC CTA GCA G	476
		MYOM1042PER	ACA GCT GCA GCA AAT CCC AA	
29	rs4613146	MYOM4613F	ATG TCT GGC AGG TCG CTT CA	
	A/C	MYOM4613R	CCA TTC CCA CCT CTC TGT CA	278
		MYOM4613PER	GAG AAA AAA TCA GGA CTA GCA ACT	
30	rs7233983	MYOM7233F	AGG GTA GAC CAG TTA GGC GG	
	A/C	MYOM7233R	GCA GTC AGT GGG TGG AAA ATG	244
		MYOM7233PE	TCC CCG CAC CAC TCT GAC	
31	rs4413045	MYOM4413F	ACT CAG CTA CAC ACC ACA TGC	
	T/C	MYOM4413R	TGA AAT GCA AGG TAA GAG GTA GAT AAA T	573
		MYOM4413PE	CTT AAA ACA AAA ACG TGA GAG TCT T	
32	rs11081004	MYOM1108F	CTA TGT GAG TGA GGA GGT TAC C	
	A/G	MYOM1108R	GTG GCG ATT TGA TTT TAG AGG GG	324
		MYOM1108PER	CTG CTC CCC TGA CTG TGA	
33	rs6506057	MYOM6506057F	AGA TTC TTG TAC CTT AGC CTC CA	
	A/G	MYOM6506057R	TAT TTT GAG GGG GTG GGT GG	275
		MYOM6506057PE	CCT CCC CAC ATA TAG TGC AT	
34	rs1071600	MYOM1071F	GCT ACA GCA CCA CCA TCT CC	
	T/C	MYOM1071R	GCA AGC ATA GCG TCA TAC ATA GG	259
		MYOM1071PER	CAG TAA TTT CTG CCC CTC CA	
35	rs9948582	MYOM9948F	GAA GGA CGT GGT GTA TAT TTG AG	
	A/G	MYOM9948R	CTG TGT GGT GGA CTG ATA AAC T	268
		MYOM9948PE	GCC CAT GGT TGG TAA GAG	
36	rs6506074	MYOM6506074F	GGA GGG AGA GAA GGG GG	
	T/C	MYOM6506074R	TTG CCT AGA TGC TTT CCT GCT G	399
		MYOM6506074PE	CGG AGT TAG TAG AAG AGG C	
37	rs9952207	MYOM9952F	CCA GGC AAG TAG GGA CCA C	
	A/G	MYOM9952R	GTC AGA AGG GGC AGG AGG A	157
		MYOM9952PE	CCT ACT GAA ATT ATT TGA ACT AGC C	
38	rs4798069	MYOM4798F	AGC CAA CTT GAA CTT CTG CCT G	
	T/C	MYOM4798R	GAT TAG CTG GGC ATG GCG G	500
		MYOM4798PE	CTA GGT CTA TTT TAT TGG AAA CGA A	
39	rs7235847	MYOM7235F	GGA AGG CAA AAT AAC TGA CCA C	

(Continued on next page)

Table 2.5 (continued)

Gene	Tag SNP	Primer name*	Primer sequence (5'>3')	Product size (bp)
	A/G	MYOM7235R	GGG GAC CTA CTT GCC TTT C	332
		MYOM7235PER	GAG ATG TGC TTC CAA AGA GGT TA	
40	rs9951849	MYOM9951F	TGG GAG GCT GGG CTG GTT	
		MYOM9951R	GCC AAC CCC TGC CCT AAA C	291
		MYOM9951PER	CCA AAG TTA TCG CAG ACA GTG	
41	rs7238703	MYOM7238F	CCC AGT GAG GCA TTT AGG TTT G	
	T/C	MYOM7238R	CTC GGT GTG CAT TAT TTG TGG G	477
		MYOM7238PER	AAA CTT ATA GTA TAC TGA AAG CTC ATT A	
42	rs8091916	MYOM8091F	GCT GCC AAC TCT TTT CCA CAT G	
	A/C	MYOM8091R	AAA TAA AAT AGA ATG TAG AGA AGA GGA AGC	419
		MYOM8091PE	CTT AAA ACA AAA ACG TGA GAG TCT T	
43	rs4340411	MYOM4340F	CAG TAA GAG CGG ATG TGC GA	
	A/G	MYOM4340R	GCC CAG ACT ATC AGC AGC TT	440
		MYOM4340PE	TTG AGG CTG TGA AAT GGT CT	
44	rs8090956	MYOM8090F	GTG GGA GTG GGG GTG GTT G	
	T/C	MYOM8090R	CAC CCA ATG CCA GTT ACA GTC A	407
		MYOM8090PE	TGC TGT TCC CTG CTG TAA ATC T	
45	rs1662315	MYOM1662F	TGG GTG TAG ACG GCG GAG	
		MYOM1662R	GGG CTC AAG CAA TCC TCC TG	426
		MYOM1662PER	TGG CCC GGT TCC TTC AAG	
46	rs9947162	MYOM9947F	GGA GAC ACA GGG AGA AGA CA	
	T/C	MYOM9947R	GGG AGG TCT AAC AGG CAT C	462
		MYOM9947PER	CCG GGC TAT CAT AGC TAA G	
47	rs4441365	MYOM4441F	GTC TCA AAC TCC TGG GCT CA	
	A/G	MYOM4441R	GAA GCT GGA TAC ATG ACT CTG G	262
		MYOM4441PE	CGT GTA CAA TGG TTC TCA CCT	
48	rs4507002	MYOM4507F	AGA ATA ACA GAG AAA TAG GAA AGT GAA G	
	T/C	MYOM4507R	GCC TCC TGT CTC TCA TCC A	234
		MYOM4507PE	CAC AGC TCT TCT CTC ACA TCA	
49	rs12605942	MYOM1260F	GGC TGT TCT GGT TTC CCT CCC	
	A/G	MYOM1260R	ATG CCC CTC TCC TCC TCC TG	437
		MYOM1260PER	AGT CTT CAT TCT CTC ATC CAG	
MYCL12B (MRCL3)				
50	rs1662347	MRCL1662347F	GGG GAG GGA TAG CAT TAG GAG	
	A/C	MRCL1662347R	GCG ATA GGG AGT TTG GCT GC	458
		MRCL1662347PE	TTA TTA CAT GAA GAG TGA CAG TCT	
51	rs717183	MRCL7171F	TCT TAC TAT TCT TCA CCT TTC TGT CC	
	A/G	MRCL7171R	GGT CTT GGT CTT TGT TCT TTT GCT	445
		MRCL7171PER	TAT AGG GGT GCT GAG GGA	
52	rs3786458	MRCL3786F	GAG TTC TCT TGG CAT GTT TGT CAC	
	T/A	MRCL3786R	GTT TTC TCT TAC ATA CAC TTC TGC C	594

(Continued on next page)

Table 2.5 (continued)

Gene	Tag SNP	Primer name*	Primer sequence (5'>3')	Product size (bp)
		MRCL3786PE	AAA TTG TTT AAA TCT GAC ATC TAA CCT T	
53	rs1662342	MRCL1662342F	ACA TCC CCA CTG CCC AGC	
	G/A	MRCL1662342R	TGC TCC GCT CTA TCA TTT GCT C	459
		MRCL1662342PE	ACT GGG CCT CAA GCT TCC	
MYCL12A (MRCL2)				
54	rs1662336	MRLC1662F	GAA CAT CGA GTC AGC CAT CAT AAA AG	
	T/C	MRLC1662R	GGA TGT TTA GAG GAT TCG TTT GGC	274
		MRLC1662PER	CTT TGC GTT TTA TAA CTG AGG AAA	
55	rs949303	MRLC9493F	GTT GTC CCT GGT TGC TAC TGC	
	A/G	MRLC9493R	GGG TGG CAG GGT AGA GAC A	490
		MRLC9493PE	GGA TTA TTT GAG CAC CTG CC	
56	rs1630702	MRLC1630F	TAC AGT AGT CCT ATA TGA TGC TTC TTT C	
	T/C	MRLC1630R	ACC CAT ACA GTT AGC CCT CCA T	572
		MRLC1630PE	ATC TGT ATC TGC AGA GGG	
57	rs1791067	MRLC1791F	GCA GAG GAA GGA AGA AGT GTT TTT AC	
	C/T	MRLC1791R	CAA CAC TGA TTT CTG AGC ACC TAC AA	396
		MRLC1791PER	ACA GAA TAG AGT TGC AGT GAG TAA	
58	rs7239576	MRLC7239F	CAA GGC TCC AAG GAA AGG TCA	
	T/C	MRLC7239R	GCT TCC TCC ACC TTC TCA CT	529
		MRLC7239PER	GAA CTT ACG CTC TAA TCA GGA	
59	rs6506094	MRLC6506F	GGG GAG GGT AAG GAG TTG G	
	A/C	MRLC6506R	AGG ACT TCT GGT ATT GGG ATT TTT TAT	166
		MRLC6506PER	GCT TGG CTT AGT GGT GTT GT	
ZFP161				
60	rs2789	ZEP2789F	TCT GAT GTT TGT TGC TGA TGG AAG	
	C/G	ZEP2789R	GTG GGG GGT CTT GGA GC	459
		ZEP2789PE	TTT ATG GAT AAA TCA TGT GCC CCA	
61	rs990072	ZEP9900F	GTA CTG CCC CAA ATA AGA GGA AC	
	C/T	ZEP9900R	GAC GCA GAA AAT TGG TGA GTG G	292
		ZEP9900PE	GGT TTG CTT GGC CTT TAG GA	
62	rs620652	ZEP6206F	CTG AAA ACA CTA AAT GAA CAA CGC CT	
	C/T	ZEP6206R	TTT TCA TCG GGA CTG GAC ACA TC	341
		ZEP6206PER	CTT CTT GAA AAG CTT TTT AAA GTA	

* The three primers listed for each SNP are in the following order down to the column: forward PCR primer, reverse PCR primer and primer for primer extension reaction.

2.6.3 Touchdown PCR amplification of DNA samples

Rather than optimizing each primer pair, all assays were processed with the same touchdown PCR conditions in this study (Roux & Hecker, 1997). A 25-µl reaction mixture was prepared, and contained the following: genomic DNA (25 ng, either pooled or individual), 1 unit of HotStar Taq Plus DNA Polymerase, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, and 0.2 µM of each SNP-specific PCR primer. The thermal profiles were as follows: initial denaturation at 95°C for 5 minutes followed by 8 touchdown cycles, 30 main cycles, and final extension for 7 minutes at 72°C. Both touchdown and main cycles consisted of 30 seconds at 95°C, 45 seconds at the annealing temperature and 45 seconds at 72°C. The annealing temperature of the main cycles was 58°C, and the initial annealing temperature for the touchdown cycles was 7°C above this with 1°C reduction for each successive touch-down cycle. In brief, 62 markers of the *MYP2* locus were successfully amplified. Individual DNA samples were used in initial trial runs and screening of heterozygous samples for subsequent estimation of k correction factors (see below). Pooled DNA samples were *each* amplified in triplicates and hence each SNP had 36 PCR products.

2.6.4 Characterisation of the screened SNPs

The DNA fragments that contained the expected SNPs were characterised by direct sequencing. The aim of characterisation was to confirm the presence of SNPs and identify samples with known genotypes for subsequent investigation. In total, all 62 SNPs from 7 candidate genes were successfully characterised by the sequencing **(Table 2.5).**

2.6.5 SNP confirmation and validation by screening of heterozygotes

In total, 62 selected SNPs were confirmed to be polymorphic. For each DNA pool and each SNP, three allele frequency estimates *each* derived from separate PCR amplifications were calculated.

To improve the accuracy of allele frequency estimates from pooled DNA samples, corrections have to be made to account for biases in allelic representation (Barcellos et al., 1997). This allelic representation bias can be caused by allele specific preferential amplification of the genomic DNA, differences in hybridization properties for the different probe sequences, or differential incorporation of ddNTPs in primer extension in the present study.

The output from the DHPLC analysis was the heights of two peaks (1st peak was “A” and 2nd peak was “B”) corresponding to two polymorphic alleles at the SNP locus. For a heterozygote, the heights of A and B were not necessarily the same (**Figure 2.9**). Inference about the allele frequency was made from the ratio of the peak heights. The most common correction method was the k correction which used a correction factor k empirically derived from the signal intensity pattern of heterozygote individuals. The mathematical definition of k correction was

$$k = A/ B,$$

where A and B are the peak intensities of alleles A and B (Hoogendoorn et al., 2000; Hoogendoorn, et al., 1999). The parameters were estimated from heterozygous individuals who provide a standard for a 1:1 ratio for a pair of peak intensities. For k

> 1, the allele A was amplified more than allele B; for $k < 1$, allele B was amplified more than the allele A; for $k = 1$, amplification was equal. One could expect to type 10 individuals to find one to two heterozygotes for an SNP with a minor allele frequency of 0.1.

2.6.6. Genotyping by PE-DHPLC analysis

2.6.6.1. Post-PCR purification

The residual single-stranded primers and unincorporated dNTPs were then removed from 10 µl of PCR amplicon solution (see Section 2.6.3) by incubating with 5 U of shrimp alkaline phosphatase and 25 U of exonuclease I at 37°C for 30 minutes, followed by 20 minutes at 80°C for enzyme inactivation.

2.6.6.2. Primer extension (PE) reactions

Primer extension reactions were performed by using 10 µl purified DNA pool amplicon, 50 µM each of two ddNTPs appropriate for the SNP-specific termination mixes (ddATP, ddCTP, ddGTP and ddTTP), 1.5 µM of specific extension primer and 1 U of Therminator in 1× supplied reaction buffer. Thermocycling conditions for the PE reactions were initial denaturation step at 96°C for 1 minute, followed by 55 cycles of 10 seconds at 96°C, 20 seconds at 43°C and 1 minute at 60°C.

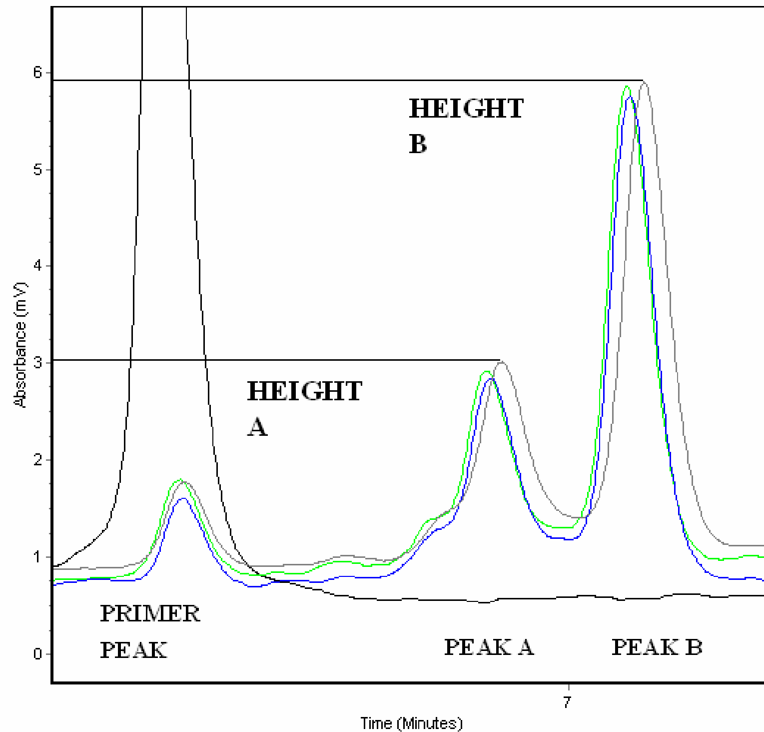


Figure 2.9
DHPLC elution profiles for a heterozygous individual.

The elution profile of primer extension products with ‘A’ and ‘B’ peaks represent allele C and allele A, respectively. In this example, the major allele is allele C and the minor allele was allele A. The peak height is equal to the signal strengths (y-axis) of the two alleles. The k correction factor is used for correction of differential amplification, and derived from the ratio of three independent readings for C and A alleles.

2.6.6.3. DHPLC analysis

PE amplicons from pooled DNA were analysed by DHPLC in the Transgenomic WAVE system. Prior to DHPLC analysis, PE products were denatured by heating at 96°C for 1 minute and then cooled in ice over a period of 10 minutes. Then, 10 µl of PE product were automatically loaded into the DNASep column and eluted on a linear acetonitrile gradient in a 0.1 M TEAA (pH 7) with a constant flow rate of 0.9 ml/min. The gradient was combined by buffer A (0.1 M TEAA) and buffer B (25% acetonitrile in 0.1 M TEAA). Different DNA fragments were separated by changing the concentration of buffer A relative to that of buffer B. The gradient profile of proportion in buffer B was adjusted according to the size of the PE amplicon.

The DHPLC system automatically processed the samples at a rate of 5 to 10 minutes per sample. UV detection was set at 260 nm. After each run, we washed the column with 25% acetonitrile for 1 minute and equilibrated it for 1 minute before the next sample injection. Optimized peak separation can be achieved with the selection of temperature and acetonitrile gradient. A DHPLC analysis was performed at 70°C and 20% to 26% buffer B starting concentration. DHPLC data analysis was based on chromatographic profile for samples with retention time and height of peaks. The PE amplicons of each SNP were eluted under essentially the same conditions of temperature and solvent gradient. Each SNP had a specific elution profile, or signature, at a given set of elution conditions of temperature and gradient.

2.7 Statistical analysis of estimated allele frequencies of DNA pools

Allele frequencies of 62 SNPs were determined based on the peak heights generated from DHPLC for each SNP and for each of the 12 pools. Since each DNA pool was measured in triplicates, each SNP had a total of 36 estimated relative allele frequencies. For each SNP, an average k correction factor was estimated from 3 separate measurements, and then used to calculate the relative allele frequencies of DNA pools. The relative allele frequency of allele A was estimated as

$$f(A) = A / (A + kB),$$

where A and B are the peak heights of the respective allele-specific primer extension products of each DNA pool, and k is the average correction factor for the SNP concerned (Hoogendoorn, et al, 2000). Significance of allele frequency differences between case and control pools were tested by nested one-way analysis of variance (ANOVA) (Qian & Shen, 2007). Since each DNA pool was constructed from 50 *distinct* subjects and its relative allele frequencies were estimated in triplicates, variance due to sampling error and variance due to technical procedures (pooling, PCR amplification, ddNTP incorporation and estimation of peak heights by DHPLC) were all handled by nested ANOVA. Since the DNA pooling approach was a screening procedure, a lenient level of significance ($P < 0.1$) was used for selecting SNPs for subsequent follow-up by individual genotyping to avoid missing genuine association.

2.7.1 Nested analysis of variance

If two or more groups are being compared, then use of non-independent observations within groups leads to inflation of degrees of freedom for statistical tests and increased probability of falsely rejecting the null hypothesis (type I error). The multiple observations for a single pool essentially represent subsamples; use of such subsampling can yield a more precise estimate of mean values for each individual pool, but the individual pool means should be used for comparisons among groups. Besides leading to appropriate statistical tests of differences among groups, nested ANOVA provide a basis for estimating components of variance among and within individuals, which is an important foundation for studying individual variation (Hayes & Jenkins, 1997; Lessells & Boag, 1987)

Nested design is an extension of a one-way ANOVA in that each group is divided into subgroups. The simple use of this would be an equivalent multivariate, paired-sample experimental design. All assumptions of ANOVA hold, e.g. normality of residuals, constant variance, etc.

Mathematically, if factor B (pool) is nested within factor A (group), then a level of factor B can only occur within one level of factor A and there can be no interaction. This gives the following statistical model (**Figure 2.10**):

$$Y_{ijk} = \mu + \alpha_i + \beta_{(i)j} + \gamma_{(ij)k}$$

where

i indexes A (called the “major factor” – case group or control group),

(i)j indexes B within A (B is called the “minor factor” – pools within group), and

(ij)k indexes replication (replicates per pool);

$i = 1, 2$ (groups)

$j = 1, 2, 3, 4, 5, 6$ (pools within group), and

$k = 1, 2, 3$ (replicates per pool); and

Y_{ijk} = individual response variable

μ = overall mean

α_i = effect for ith group

$\beta_{(i)j}$ = effect for jth pool within ith group

$\gamma_{(ij)k}$ = random error (technical replicates)

This equation indicates that each data value is the sum of a common value (mean), the level effect for factor A (group), the level effect of factor B (pool) nested within factor A, and the residual (random error). For a nested design we typically use variance component (i.e., ANOVA) methods to perform the analysis.

Case group (H)						Control group (E)					
P1(H)	P2(H)	P3(H)	P4(H)	P5(H)	P6(H)	P7(E)	P8(E)	P9(E)	P10(E)	P11(E)	P12(E)
X	X	X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	X

Figure 2.10
The nested design of the DNA pooling study

There are two groups (case and control), six pools per group (P1 – P6 for case [**H**igh myopia] group, and P7 – P12 for control [**E**mmetropia] group), and three replicates (X) per pool. Note that there is no link from any pools of the case group to any pools of the control group. In other words, the level of the control group is *not* cross-classified with the case group, but is nested within the group – the pools are nested within the group. The nested design can test two things. **First**, it tests the difference between the case (high myopia) group and the control (emmetropia) group, which is the main purpose of the pooling experiments – to test for allele frequency difference between these two groups. **Second**, it also tests for the variability of the pools within groups, which is expected to be large because each pool is prepared from 50 distinct samples.

2.8 Individual genotyping for pooling screening

Only those SNPs that met the designated level of significance (P <0.1) were considered for selection. Individual genotyping was performed using the RFLP method.

PCR amplification was carried out in a total volume of 10 μL containing 1× reaction buffer, 0.2-0.3μM of each of forward and reverse primers, 0.2 units of HotStarTaq Plus DNA polymerase, 0.2 mM dNTPs, 1.5-2.5 mM $MgCl_2$ and 10 ng of genomic DNA template. The sequences of primers used are listed in **Table 2.6**. All PCRs were carried out in a Gene-Amp PCR System 9700. In a total volume of 15 μL, amplified DNA (10 μL) was digested overnight with 2 U of restriction endonuclease using the buffers and temperatures recommended by the manufacturers. Digested PCR products were separated by electrophoresis in polyacrylamide gels (8-12%, depending on the size of restriction fragments) and stained with SYBR Green I (**Figure 2.11**). To verify the RFLP results, at least 3 samples with different genotypes were sequenced in an automated ABI Prism 310 Genetic Analyzer for quality control. An internal control site was always introduced into the PCR fragment to guard against incorrect genotype calls due to faulty restriction enzymes or incomplete digestion. The PCR product contains a restriction recognition site common to both alleles of a SNP, and cleavage at this site serves as an internal positive control for the restriction digestion.

Table 2.6 PCR primers, reaction component, amplicon sizes and RFLP digestion enzymes for nine top-ranking SNPs in the *MYP2* region

Gene	SNP	Primer name	Primer sequence (5'>3')	Size (bp)	PCR cycles	$MgCl_2$ (mM)	Primer (μM)	RFLP enzymes
CLUL1	rs546696	CLU546696F1	$(A)_{10}$G GTA TGG TCC GTA TGG CAC CTG	183	38	1.5	0.3	MvaI
		CLU546696R1	$(A)_{26}$ CAG ATT ACA AAA CAC CAG GTT CAC CAA CC					
CLUL1	rs1004961	CLU1004961F1	CTC TTA CAT GGC AAC TCA TTA CAA TT	321	38	1.5	0.3	Eco130I
		CLU1004961R1	TGA ACT TTA TAC TGT GGC ATA TGA AC					
EMILIN2	rs637647	EM637647F1	CTG AGA TTA CAC ATT CTT CCA GCA T	192	38	2	0.2	Eco130I
		EM637647R2	$(A)_{26}$TC CCA CCC TTG GCC AGG					
LPIN2	rs589318	LP589318F1	$(A)_{24}$ GGG GAA AAG GTC ATT CAG GCA AGG TA	180	38	1.5	0.2	RsaI
		LP589318R1	$(A)_{23}$ T GTG AAG GTA CAA TGG ATG TGG ATC AG					
MYOM1	rs9948582	MY9948582F1	$(T)_{19}$ GC CCA AGT CTT AGA ATA ATT AAT GG	296	38	1.5	0.2	Alw26I
		MY9948582R1	$(T)_{35}$ ATC AGG CAA GAG TTT CTA AAT GTC T					
MYOM1	rs7235847	MY7235847F1	CAT AGG AAG GCA AAA TAA CTG ACC AC	197	38	2.5	0.2	RsaI
		MY7235847R1	$(T)_{37}$ GAG ATG TGC TTC CAA AGA GGG TA					
MYOM1	rs4340411	MYOM1rs4340411F1	GAT TCA TTG TAT GCG CAA AAT GTC CA	208	38	1.5	0.2	Hpy166II
		MYOM1rs4340411R1	AGG GCC AGA TGT GCA CTA AGA GC					
MYOM1	rs12605942	MY12605942F1	CAT AAG CTA TCA AGT GCA TTC TTC TG	198	38	2	0.2	BseNI
		MY12605942R1	$(T)_{29}$ GGT CCT GTC TTT AGC ACT GCC					
MYCL12A	rs1791067	MRL1791067F1	$(A)_{15}$ TGG TTA AAC ATT AAT AAA TTG GAC AGT CTT AGA TC	200	38	2.5	0.1	HpyF3I
		MRL1791067R1	$(A)_{19}$ CC AAG RTA CAG AAT AGA GTT GCA GTG ACT AA					

Figure 2.11 RFLP with nine SNPs for individual genotyping

Gene	SNP	Enzymes	PAGE	Polymorphic band in bp*	Pattern
CLUL1	rs546696	MvaI	10%	**TT**: 138,45 **GT**: 138, 110, 45, 28 **GG**: 110, 45, 28	
CLUL1	rs1004961	Eco130I	8%	**CC**: 151, 94, (76) **AA**: 277,94 **CA**: 277, 151, 94, (76)	
EMILIN2	rs637647	Eco130I	12%	**CC**: 153, 38 **TT**: 89, 64, (38) **CT**: 153, 89, 64, 38	

(Continued on next page)

Figure 2.11 (continued)

Gene	SNP	Enzymes	PAGE	Polymorphic band (bp)*	Pattern
LPIN2	rs589318	RsaI	12%	**TT**: 148, (32) **CT**: 148, 99,(49),(32) **CC**: 99,(49),(32)	TT CT CC LADDER
MYOM1	rs9948582	Alw26I	8%	**AA**: 240,56 **AG**: 240, 174, 66, 56 **GG**: 174, 66, 56	AA AG GG
MYOM1	rs7235847	RsaI	12%	**GG**: 89, 59, 49 **AA**: 108, 89 **GG**: 108, 89, 59, 49	GG AA AG LADDER

(Continued on next page)

Figure 2.11 (continued) * Bands with size shown in brackets ( ) are not obvious on the gel.

Gene	SNP	Enzymes	PAGE	Polymorphic band in bp*	Pattern
MYOM1	rs4340411	Hpy166II	10%	**CC**: 120, (30), (25) **CT**: 150, 120, (30), (25) **TT**: 150, (25)	CC CT TT LADDER
MYOM1	rs12605942	BseNI	12%	**AA**: 115, 58, 44 **GG**: 115, 44 **AG**: 115, 58, 56, 44	AA GG LADDER AG
MYCL12A (*MRCL2*)	rs1791067	HpyF3I	10%	**TT**: 157, (43) **CC**: 107, 50, (43) **CT**: 157, 107, 50, (43)	TT CC CT LADDER

2.9 Statistical analysis for individual genotyping

Genotypes were tested for HWE by chi-square test. Allele- and genotype-wise analysis with chi-square tests were carried out to derive *P* values for association.

Finally, all 600 DNA samples were genotyped individually for the 9 SNPs found to be associated with high myopia by the DNA pooling approach. Multiple testing was accounted for by false discovery rate at a level of 0.05. See Section 2.5.1.2 for details of the procedure. For this part, the total number of comparison (n) was 89 as shown below: 62 comparisons of estimated allele frequencies between case pools and control pools, and 27 comparisons of 9 SNPs individually genotyped and analysed under 3 genetic models.

2.10 Computer programmes

The computer programmes for experimental applications and statistical analysis are listed in **Table 2.7.**

Table 2.7 Software used in study and their applications and sources.

Software	Application	Sources / URL
FBAT	Family-based association test program	http://www.biostat.harvard.edu/~fbat/default.html
GeneScan Analysis Software (ver 3.1)	Fragment length analysis for microsatellite genotyping	Applied Biosystems, Foster City, USA
Haploview	Haplotype analysis	http://www.broadinstitute.org/haploview/haploview
NEBcutter (ver 2.0)	restriction enzyme cutting sites for RFLP	http://tools.neb.com/NEBcutter2/index.php
OLIGO (ver 6.57)	Primer design program Molecular Biology Insights, Cascade, US	Molecular Biology Insights, Cascade, US
Plink	Genetic Analysis family data	http://pngu.mgh.harvard.edu/~purcell/plink/
PowerMarker	Calculation of LD measures for multiallelic markers	http://statgen.ncsu.edu/powermarker/
Primer-BLAST	similar sequences for primer design	http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome
SPSS (ver 17.0)	Correlation between Spherical Equvalent and other ocular parameters	SPSS Inc., Chicago, IL
STATA (ver 8.2) installed with GenAssoc	HWE testing, association analysis, stepwise logistic regression	Stata Corporation, College Station, TX, USA
TM Utility (ver 1.3)	Estimates the melting temperature of primers with or without base mismatch	http://www.idahotech.com/downloads_up/index.html
WAVEMAKER (ver 4.1.4)	DHPLC melting curve analysis for estimating melting temperature for SNP	Transgenomic, Omaha, NE, USA

Chapter 3
Replication study for the *MYOC* gene

3.1 Overview

Objectives. Polymorphisms of the myocilin (*MYOC*) gene on chromosome 1q24.3 have been reported to be associated with high myopia. However, subsequent studies showed inconsistent results. We attempted to replicate these genetic markers found associated in the recent reports in a case-control study of 300 unrelated Chinese subjects with high myopia (≤-8.0 D; n=300) and 300 emmetropic Chinese control subjects (±0.75 D; n=300).

Methods. *MYOC* polymorphisms previously found associated with high myopia were genotyped together with other SNPs in strong LD with the positive SNPs. Their frequencies were compared between high myopes and controls by statistical tests.

Results. Five correlated SNPs at the 3' end of the *MYOC* gene showed significant differences in allele and genotype frequencies between high myopes and controls: rs12076134 (S2), rs1602244 (S5), rs6425356 (S7), rs10737323 (S8) and rs743994 (S10). The results remained significant after correction for multiple comparisons. Four *MYOC* polymorphisms previously found associated with high myopia failed to be replicated. Two most significant associations with rs6425356 (S7) and rs743994 (S10) were further confirmed in the original families.

Conclusions. *MYOC* polymorphisms at the 3' flanking region may predispose to development of high myopia.

3.2 Background of myocilin (*MYOC*)

The myocilin (*MYOC*; OMIM 601652) gene is also known by the name trabecular meshwork inducible glucocorticoid response (*TIGR*) gene. In 1997, myocilin was identified as the causal gene for primary open angle glaucoma – the glaucoma locus *GLC1A* (Stone, et al., 1997).

In the eye, the trabecular meshwork cells help regulate eye pressure by controlling the drainage of fluid from the eye as new fluid is produced. When treated with steroids, the cultured cells secreted the same protein, which Nguyen *et al*. called TIGR (for trabecular meshwork inducible glucocorticoid response protein) (Nguyen, et al., 1998). The gene (*MYOC/TIGR*), located on chromosome 1q24.3, spans about 17 kb in the human genome and contains three exons. Fingert *et al*. showed by northern blot analysis that *MYOC* was expressed as a 2.3-kb transcript and the translated product was predicted to contain 504 amino acids (57 kDa) (Fingert, et al., 1998). The protein is normally present in a variety of ocular and non-ocular tissues (Fingert et al., 1998; Tomarev, et al., 1998). In the eye, MYOC may be produced in greater amounts in times of stress (Johnson, 2000).

Myopia has been also linked to glaucoma (Tripathi, et al., 1999; Wu, et al., 2000), which in turn is associated with the *MYOC* gene. The importance of genetic effects of *MYOC* for high myopia was first suggested in an abstract published by Wu et al. in Singapore. The study found in the 5' proximal promoter region of the *MYOC* gene genetic variants that were associated with high myopia (≤-8 D) (Wu H, 1999). Later on, the same group reported the association of high myopia with a MS marker at the 3' flanking region, but the not the original MS in the 5' proximal promoter region

(Wu H, 2000). However, another study failed to replicate the association with a small samples size (n=69) in Hong Kong Chinese (Leung, et al., 2000). Recently, a family-based association study in Hong Kong confirmed a high degree of association between high myopia and polymorphism in the 3' flanking region of *MYOC* (Tang et al., 2007). Intriguingly, another recent study did not find association between high myopia and *MYOC* polymorphisms in subjects of European origin (Zayats, et al., 2009). To clarify the conflicting relationship between high myopia and *MYOC* polymorphisms, we attempted to further investigate the putative association between polymorphisms in *MYOC* and high myopia with an independent Chinese sample set with a larger sample size.

3.3 Results

3.3.1 Analysis of ocular data in unrelated subjects

The present study recruited 300 control subjects and 300 case subjects with high myopia. The demographic and clinical characteristics of the controls and of cases are shown in **Table 3.1**. The case subjects with high myopia were slightly older and had longer eyeball than control subjects. For control subjects, the mean age was 24.9 years, mean SE +0.03 D, and mean axial length (AXL) 23.85 mm. For subjects with high myopia, the mean age was 27.7 years, mean SE -10.53 D, and mean AXL 27.76 mm. The ocular data presented were for right eyes. The correlation between right and left eyes was the best for both SE (0.97) and AXL (0.96) when all subjects (n = 600) were included in the analysis. The partial correlation of SE with other ocular

components was the best for AXL irrespective of whether the controls and the cases were analysed separately or combined.

Among ocular components, axial length is considered to be the most important determinant of refraction (Kinge, et al., 1999), which has been associated with chromosome 2p24 in the isolated Sardinian population (Biino et al., 2005) and chromosome 5q (Zhu et al., 2008).

- **Case and control data**

There were more females in highly myopic subjects than in control subjects (72.5% vs 56.3%; $p < 0.001$, χ^2 test).

Table 3.1 Clinical Characteristics of the highly myopic patients and the control subjects *

Characteristics	Controls (n=300)	Cases (n=300)	All subjects (n=600)
Age, mean±SD (yr)	24.9 ±6.1	27.7 ±6.9	26.3 ±6.6
Females, No. (%)	169 (56.3%)	215 (72.4%)	384 (64.3%)
SE, mean±SD (D)	+0.03 ±0.43	-10.53 ±2.48	-5.22 ±5.58
AXL, mean±SD (mm)	23.85 ±0.83	27.76 ±1.13	25.79 ±2.19
CP, mean±SD (D)	43.50 (1.50	44.40 ±2.71	43.95 ±2.24
ACD, mean±SD (mm)	3.62 ±0.32	3.72 ±0.32	3.67 ±0.33
LT, mean±SD (mm)	3.99 ±0.64	4.07 ±0.67	4.03 ±0.66
Partial correlation (r) of SE with ocular components			
AXL, r (*P* value)	-0.34 (<0.001)	-0.67 (<0.001)	-0.94 (<0.001)
CP, r (*P* value)	-0.25 (<0.001)	-0.25 (<0.001)	-0.47 (<0.001)
ACD, r (*P* value)	0.04 (<0.46)	0.27 (<0.001)	0.24 (<0.001)
LT, r (*P* value)	0.16 (0.008)	0.10 (0.100)	0.10 (0.020)

* SE = spherical equivalent; AXL = axial length; CP = corneal power; ACD = anterior chamber depth; LT = lens thickness. The ocular data are for right eyes only.

3.3.2 Tests of genetic association: single locus

3.3.2.1 Tests of association under three genetic models

Although there were more females in the case group than in the control group, chi-square tests indicated that there were no significant difference (all P values > 0.05) in the distribution of genotypes between the two genders for all 10 polymorphisms (**Table 3.2**). This justified the direct comparison of genetic data between cases and controls *without* stratification into males and females. For the sake of easy reference and discussion, the polymorphisms were also designated S1 to S10 in sequential order from the 5' to the 3' end of the *MYOC* gene

The genotypes of all polymorphisms were in Hardy-Weinberg equilibrium ($P > 0.05$) for the control group and the case group, except NGA17 in the case group (P = 0.0226) (**Tables 3.3** and **3.4**). Therefore, all polymorphisms were tested for association with high myopia. At the nominal significance level of 0.05, five SNPs showed significant differences in distribution of genotypes and alleles between cases and controls under all three genetic models (genotypic, additive and allelic): rs12076134 (S2), rs1602244 (S5), rs6425356 (S7), rs10737323 (S8) and rs743994 (S10). The MS NGA19 gave marginal significance under the additive model (P = 0.043).

With correction for 30 comparisons by false discovery rate (FDR) at a level of 0.05, the FDR-adjusted significance level was set at 0.0250 (**Table 3.5**). Therefore, only the above-mentioned SNPs remained significantly associated with high myopia under all three genetic models even after correction for multiple comparisons; and the corresponding P values are shown in **boldface** in **Table 3.4**. The most significant

results were given by rs6425356 (S7) and rs743994 (S10) with the *P* values in the range between 10^{-5} and 10^{-6}. In particular, rs6425356 (S7) gave a *P* value of 1.98×10^{-6} under the additive model, and rs743994 (S10) showed the most impressive *P* value of 1.41×10^{-6} under the allelic model.

It is interesting to examine the ORs for the five SNPs significantly associated with high myopia (**Table 3.4**). In the **genotypic model** with the more frequent homozygote as the reference genotype, the OR varied from 1.503 for rs1602244 (S5) to 2.031 for rs743994 (S10) for the heterozygote, and from 1.340 for rs10737323 (S8) to 3.218 for rs6425356 (S7) for the less frequent homozygote. In the **additive model** with the more frequent homozygote as the reference genotype, the common OR varied from 1.420 for rs10737323 (S8) to 1.842 for rs6425356 (S7) for an increase of one copy of the risk allele (i.e. the less frequent allele or the minor allele). In the **allelic model** with the more frequent allele (i.e. the major allele) as the reference allele, the OR varied from 1.414 for rs10737323 (S8) to 1.876 for rs743994 (S10) for the risk allele. It is also intriguing to note that the ORs were very similar for both rs6425356 (S7) and rs743994 (S10), which also demonstrated the most convincing evidence (i.e. impressively low *P* values) of association with high myopia.

Of particular interest and importance was the finding that the four polymorphisms found positive by our group in a previous family-based study were all negative in the present case-control study. These four polymorphisms were NGA17 (S1), NGA19 (S3), rs2421853 (S4) and rs235858 (S9) (**Tables 3.3** and **3.4**).

Table 3.2 Comparison of *MYOC* genotype frequencies between males and females within the control and the case groups

Polymorphism	Ref. no. (this study)	*P* value (χ^2 test)	
		Controls	Cases
NGA17	S1	0.3625	0.6978
rs12076134	S2	0.1846	0.9715
NGA19	S3	0.0666	0.4274
rs2421853	S4	0.3174	0.6654
rs1602244	S5	0.2270	0.7365
rs171000	S6	0.9364	0.2182
rs6425356	S7	0.0901	0.9904
rs10737323	S8	0.4254	0.7269
rs235858	S9	0.7433	0.8086
rs743994	S10	0.2035	0.9773

Table 3.3 *MYOC* microsatellite markers: frequency distribution in cases and controls, and association analysis*

				HWE test		Genotypic test		Trend test					Allelic test	
		Controls	Cases	Controls	Cases		χ^2	Common OR	χ^2		Controls	Cases		χ^2
MS**	Gt	No. (Freq)	No. (Freq)	*P* value	*P* value	OR (945% CI)	*P* value	(945% CI)	*P* value	Allele	No. (Freq)	No. (Freq)	OR (95% CI)	*P* value
NGA17	3/3	81 (0.270)	90 (0.300)	0.3930	0.0226	(reference)	0.0898	1.016 (0.950 - 1.088)	0.6407	3	318 (0.530)	317 (0.528)	(reference)	0.6516
(S1)	3/4	42 (0.140)	59 (0.197)			1.264 (0.769 - 2.078)				4	75 (0.125)	87 (0.145)	1.163 (0.823 - 1.645)	
	3/5	114 (0.380)	77 (0.257)			0.608 (0.401 - 0.922)				5	204 (0.340)	191 (0.318)	0.939 (0.730 - 6.033)	
	3/6	0 (0.000)	1 (0.003)			-				6	3 (0.005)	4 (0.007)	1.337 (0.297 - 6.032)	
	4/4	4 (0.013)	3 (0.010)			0.675 (0.147 - 3.107)				7	0 (0.000)	1 (0.002)	-	
	4/5	24 (0.080)	21 (0.070)			0.788 (0.408 - 1.521)								
	4/6	1 (0.003)	1 (0.003)			0.900 (0.055 - 14.62)								
	5/5	32 (0.107)	45 (0.150)			1.266 (0.735 - 2.180)								
	5/6	2 (0.007)	2 (0.007)			0.900 (0.124 - 6.537)								
	5/7	0 (0.000)	1 (0.003)			-								
NGA19	3/3	16 (0.053)	15 (0.050)	0.2289	0.8792	(reference)	0.1400	1.063 (1.002 - 1.129)	0.0432	3	160 (0.267)	133 (0.222)	0.780 (0.598 - 1.019)	0.3367
(S3)	3/4	5 (0.017)	0 (0.000)			0				4	5 (0.008)	4 (0.007)	0.751 (0.200 - 2.819)	
	3/5	118 (0.393)	100 (0.333)			0.904 (0.425 - 1.923)				5	415 ((0.692)	442 (0.737)	(reference)	
	3/6	4 (0.013)	3 (0.010)			0.800 (0.149 - 4.285)				6	19 (0.032)	0 (0.000)	1.038 (0.550 - 1.959)	
	3/8	1 (0.003)	0 (0.000)			0				7	0 (0.000)	21 (0.035)	-	
	4/4	0 (0.000)	1 (0.003)			-				8	1 (0.002)	0 (0.000)	0	
	4/5	0 (0.000)	2 (0.007)			-								
	5/5	141 (0.470)	161 (0.537)			1.218 (0.580 - 2.556)								
	5/6	15 (0.050)	18 (0.060)			1.280 (0.475 - 3.453)								

* MS = microsatellite marker; Gt = genotype; HWE = Hardy-Weinberg equilibrium; OR = odds ratio; CI = confidence interval

** For both NGA17 and NGA19, alleles 3 to 8 have 13 to 18 GT repeats, respectively.

Table 3.4 *MYOC* single nucleotide polymorphisms: frequency distribution in cases and controls, and association analysis*

				HWE test		Genotypic test**		Trend test**					Allelic test**	
		Controls	Cases	Controls	Cases		χ^2	Common OR	χ^2		Controls	Cases		χ^2
SNP	Gt	No. (Freq)	No. (Freq)	*P* value	*P* value	OR (945% CI)	*P* value	(95% CI)	*P* value	Allele	No. (Freq)	No. (Freq)	OR (95% CI)	*P* value
rs12076134	T/T	184 (0.613)	138 (0.460)	0.6185	0.4590	(reference)	**0.0007**	1.622 (1.257 - 2.094)	**0.0002**	T	468 (0.780)	411 (0.685)	(reference)	**0.0002**
(S2)	T/G	100 (0.333)	135 (0.450)			1.800 (1.281 - 2.529)				G	132 (0.220)	189 (0.315)	1.630 (1.257 - 2.115)	
	G/G	16 (0.053)	27 (0.090)			2.250 (1.167 - 4.338)								
rs2421853	G/G	159 (0.530)	147 (0.490)	0.2308	0.3463	(reference)	0.6033	1.128 (0.887 - 1.435)	0.3262	G	431 (0.718)	415 (0.692)	(reference)	0.3114
(S4)	G/A	113 (0.377)	121 (0.403)			1.158 (0.824 - 1.628)				A	169 (0.282)	185 (0.308)	1.137 (0.886 - 1.458)	
	A/A	28 (0.093)	32 (0.107)			1.236 (0.710 - 2.152)								
rs1602244	C/C	182 (0.607)	147 (0.490)	0.9309	0.8480	(reference)	**0.0075**	1.502 (1.164 - 1.938)	**0.0018**	C	467 (0.778)	419 (0.698)	(reference)	**0.0016**
(S5)	C/T	103 (0.343)	125 (0.417)			1.503 (1.070 - 2.110)				T	133 (0.222)	181 (0.302)	1.517 (1.168 - 1.969)	
	T/T	15 (0.050)	28 (0.093)			2.311 (1.190 - 4.488)								
rs171000	T/T	90 (0.300)	105 (0.350)	0.9356	0.1697	(reference)	0.4099	0.896 (0.717 - 1.120)	0.3332	T	328 (0.547)	345 (0.575)	(reference)	0.3229
(S6)	T/G	148 (0.493)	135 (0.450)			0.782 (0.542 - 1.127)				G	272 (0.453)	255 (0.425)	0.891 (0.709 - 1.120)	
	G/G	62 (0.207)	60 (0.200)			0.829 (0.527 - 1.305)								

Table 3.4 (continued)

				HWE test		Genotypic test**		Trend test**					Allelic test**	
		Controls	Cases	Controls	Cases		χ^2	Common OR	χ^2		Controls	Cases		χ^2
SNP	Gt	No. (Freq)	No. (Freq)	*P* value	*P* value	OR (945% CI)	*P* value	(95% CI)	*P* value	Allele	No. (Freq)	No. (Freq)	OR (95% CI)	*P* value
rs6425356	C/C	177 (0.590)	121 (0.403)	0.7767	0.2605	(reference)	**1.03×10^{-5}**	1.842 (1.432 - 2.370)	**1.98×10^{-6}**	C	462 (0.770)	388 (0.647)	(reference)	**2.63×10^{-6}**
(S7)	C/T	108 (0.360)	146 (0.487)			1.978 (1.408 - 2.777)				T	138 (0.230)	212 (0.353)	1.829 (1.416 - 2.362)	
	T/T	15 (0.050)	33 (0.110)			3.218 (1.676 - 6.181)								
rs10737[illegible]	[illegible]	[illegible]	143 (0.477)	0.2710	0.0900	(reference)	**0.0045**	1.420 (1.096 - 1.841)	**0.0081**	A	463 (0.772)	423 (0.705)	(reference)	**0.0086**
(S8)	A/G	99 (0.330)	137 (0.457)			1.761 (1.255 - 2.472)				G	137 (0.228)	177 (0.295)	1.414 (1.090 - 1.834)	
	[illegible]	[illegible]063)	20 (0.067)			1.340 (0.689 - 2.605)								
rs235858	T/T	88 (0.293)	94 (0.313)	0.7965	0.7776	(reference)	0.8127	0.928 (0.738 - 1.166)	0.5201	T	327 (0.545)	338 (0.563)	(reference)	0.5231
(S9)	T/C	151 (0.503)	150 (0.500)			0.930 (0.644 - 1.344)				C	273 (0.455)	262 (0.437)	0.922 (0.739 - 1.166)	
	C/C	61 (0.203)	56 (0.187)			0.859 (0.540 - 1.368)								
rs743994	G/G	188 (0.627)	131 (0.437)	0.1827	0.8016	(reference)	**1.07×10^{-5}**	1.795 (1.405 - 2.293)	**2.81×10^{-6}**	G	470 (0.783)	395 (0.658)	(reference)	**1.41×10^{-6}**
(S10)	G/A	94 (0.313)	133 (0.443)			2.031 (1.437 - 2.869)				A	130 (0.217)	205 (0.342)	1.876 (1.447 - 2.433)	
	A/A	18 (0.060)	36 (0.120)			2.870 (1.562 - 5.273)								

(Continued on next page)

* SNP = single nucleotide polymorphism; Gt = genotype; HWE = Hardy-Weinberg equilibrium; OR = odds ratio; CI = confidence interval

** The *P* values for association analyses that remain significant after correction for multiple comparisons at a false discovery rate of 0.05 are shown in **boldface**. For details, please refer to **Table 3.5**.

Table 3.5 Correction for multiple comparisons by false discovery rate (FDR) at a level of 0.05 for association analysis of *MYOC* markers*

Observed *P* values (P_j)	Rank (j)	FDR thresholds (0.05 × j/30)	Observed *P* values (P_j)	Rank (j)	FDR thresholds (0.05 × j/30)
1.41×10^{-6}	1	0.0017	0.0432	16	0.0267
1.98×10^{-6}	2	0.0033	0.0898	17	0.0283
2.63×10^{-6}	3	0.0050	0.14	18	0.0300
2.81×10^{-6}	4	0.0067	0.3114	19	0.0317
1.03×10^{-5}	5	0.0083	0.3229	20	0.0333
1.07×10^{-5}	6	0.0100	0.3262	21	0.0350
0.0002	7	0.0117	0.3332	22	0.0367
0.0002	8	0.0133	0.3367	23	0.0383
0.0007	9	0.0150	0.4099	24	0.0400
0.0016	10	0.0167	0.5201	25	0.0417
0.0018	11	0.0183	0.5231	26	0.0433
0.0045	12	0.0200	0.6033	27	0.0450
0.0075	13	0.0217	0.6407	28	0.0467
0.0081	14	0.0233	0.6516	29	0.0483
0.0086	**15**	**0.0250**	0.8127	30	0.0500

* Association analysis was performed for 10 *MYOC* markers (2 microsatellites and 8 SNPs; **Tables 3.3** and **3.4**) each with three genetic models (genotypic, additive and allelic). Thus, there were 30 comparisons for correction. The list of 30 observed *P* values is shown in two columns above from left to right, and sorted from the smallest (P_1) to the largest (P_{30}). The list of FDR threshold *P* values (0.05 × j/n, where j is the rank; and n is equal to 30 comparisons) is also arranged from the smallest to the largest (0.05). Starting from the largest observed *P* value (P_n or 0.8127 in the table above), compare P_n with 0.05 × j/n. Continue as long as $P_j > 0.05 \times j/n$. Let k be the *first* time when $P_k \leq 0.05 \times k/n$, and declare the comparisons corresponding to the smallest k observed *P* values as statistically significant. In the table above, k = 15 and the FDR threshold *P* value is 0.0250. Therefore, the smallest 15 observed *P* values (0.0086 to 1.41×10^{-6} shown in **boldface** in the table and in **Table 3.4**) are significant.

3.3.2.2 Detection of main effects by stepwise logistic regression

To investigate which of the five significant SNPs contributed the main effects to the association, a stepwise logistic regression procedure was carried out. In the **forward procedure**, the genotypes of each SNP were tested in turn together with another SNP under an additive model for the copy number of the risk allele (**Table 3.6**). Given the main effect of S7, only S10 remained significant, but not S2, S5 or S8. Given the main effect of S2, both S7 and S10 still contributed significantly to the association. Given the main effect of S8, only S5 did not contribute any significant effect to the association. In other words, S7 and S10 consistently showed significant main effects even in the presence of the other three SNPs. When both S7 and S10 were in the model, only S8 remained significant. When S7, S8 and S10 were included in the model, the remaining two SNPs (S2 and S5) did not contribute significant effects anymore. In conclusion, the forward stepwise procedure found that rs6425356 (S7), rs10737323 (S8) and rs743994 (S10) contributed main effects to the association with high myopia and could also account for the positive association results of rs12076134 (S2) and rs6425356 (S5).

In the **backward procedure**, S5 was first removed from the full model because its deletion deteriorated the fit of the model least (**Table 3.7**). This was followed by the removal of S2. With the remaining three SNPs (S7, S8 and S10) in the model, deletion of any one worsened the fit significantly. In conclusion, the backward procedure also identified S7, S8 and S9 as the contributors of main effects to the association while the same conclusion as in the forward procedure. Selection procedures testing the genotypes as categorical variables also gave the same conclusion.

Table 3.6 Detection of main effects by forward stepwise logistic regression analysis of positive *MYOC* single nucleotide polymorphisms*

Null hypothesis	Alternative hypothesis	Wald test	
		$\chi^2(1)$	*P* value
S7	S7 + **S10**	5.31	**0.0212**
S7	S7 + S2	1.17	0.2785
S7	S7 + S5	0.31	0.5784
S7	S7 + S8	1.92	0.1659
S10	**S10** + **S7**	5.99	**0.0144**
S10	S10 + S2	0.15	0.7008
S10	S10 + S5	0.09	0.7620
S10	S10 + S8	1.31	0.2528
S2	S2 + **S7**	10.01	**0.0016**
S2	S2 + **S10**	8.25	**0.0041**
S2	S2 + S5	1.02	0.3127
S2	S2 + S8	0.01	0.9179
S5	S5 + **S7**	12.72	**0.0004**
S5	S5 + **S10**	12.04	**0.0005**
S5	S5 + **S2**	5.05	**0.0247**
S5	S5 + S8	0.08	0.7782
S8	S8 + **S7**	16.33	**0.0001**
S8	S8 + **S10**	14.93	**0.0001**
S8	S8 + **S2**	6.67	**0.0098**
S8	S8 + S5	2.83	0.0923
S7 + S10	S7 + S10 + S2	0.11	0.7397
S7 + S10	S7 + S10 + S5	1.76	0.1849
S7 + **S10**	**S7** + **S10** + **S8**	8.58	**0.0034**
S7 + S8 + S10	S7 + S8 + S10 + S2	0.50	0.4811
S7 + S8 + S10	S7 + S8 + S10 + S5	0.00	0.9470

* The single nucleotide polymorphisms (SNPs) tested are as follows: rs12076134 (S2), rs1602244 (S5), rs6425356 (S7), rs10737323 (S8), rs743994 (S10). The genotypes of each SNP are treated as a metric variable, and tested as an additive model for the copy number of the minor allele. $\chi^2(1)$ indicates chi-square statistic with one degree of freedom.

Table 3.7 Detection of main effects by backward stepwise logistic regression analysis of positive *MYOC* single nucleotide polymorphisms*

		Wald test	
Null hypothesis	Alternative hypothesis	$\chi^2(1)$	*P* value
S5 + S7 + S8 + S10	S2 + S5 + S7 + S8 + S10	0.49	0.4830
S2 + S7 + S8 + S10	S2 + **S5** + S7 + S8 + S10	**0.00**	0.9971
S2 + S5 + S8 + S10	S2 + S5 + S7 + S8 + S10	11.22	0.0008
S2 + S5 + S7 + S10	S2 + S5 + S7 + S8 + S10	7.50	0.0062
S2 + S5 + S7 + S8	S2 + S5 + S7 + S8 + S10	8.07	0.0045
S7 + S8 + S10	**S2** + S7 + S8 + S10	**0.50**	0.4811
S2 + S8 + S10	S2 + S7 + S8 + S10	12.63	0.0004
S2 + S7 + S10	S2 + S7 + S8 + S10	8.90	0.0028
S2 + S7 + S8	S2 + S7 + S8 + S10	8.07	0.0045
S8 + S10	S7 + S8 + S10	13.00	0.0003
S7 + S10	S7 + S8 + S10	8.58	0.0034
S7 + S8	S7 + S8 + S10	11.57	0.0007

* The single nucleotide polymorphisms (SNPs) tested are as follows: rs12076134 (S2), rs1602244 (S5), rs6425356 (S7), rs10737323 (S8), rs743994 (S10). The genotypes of each SNP are treated as a metric variable, and tested as an additive model for the copy number of the minor allele. $\chi^2(1)$ indicates chi-square statistic with one degree of freedom.

3.3.3 Tests of association: multiple SNPs

3.3.3.1 Linkage disequilibrium patterns

The distribution of r^2 values in controls and in cases was similar although the values were generally higher in the controls than in the cases (**Figure 3.1**). The r^2 values in the combined group of cases and controls were intermediate between the two constituent groups with a similar pattern of distribution. In the combined group, the strongest LD was observed between rs1602244 (S5) and rs10737323 (S8) with an r^2 value of 0.65. No LD block could be constructed based on Gabriel's definition. This was in sharp contrast to strong LD expected among most of the SNPs under study for Han Chinese as documented in the HapMap database (**Figure 3.1D**) because rs2421853 (S4) or rs235858 (S9) was supposed to tag efficiently (i.e. r^2 close to 1.0) the other SNPs under study － the original algorithm that other SNPs under study were selected in the first place.

A similar distribution pattern of r^2 values was also observed between MSs (NGA17 and NGA19) and SNPs for the control group, the case group and the combined group (**Table 3.8**). In general, the r^2 values were very low: 0.15 or less.

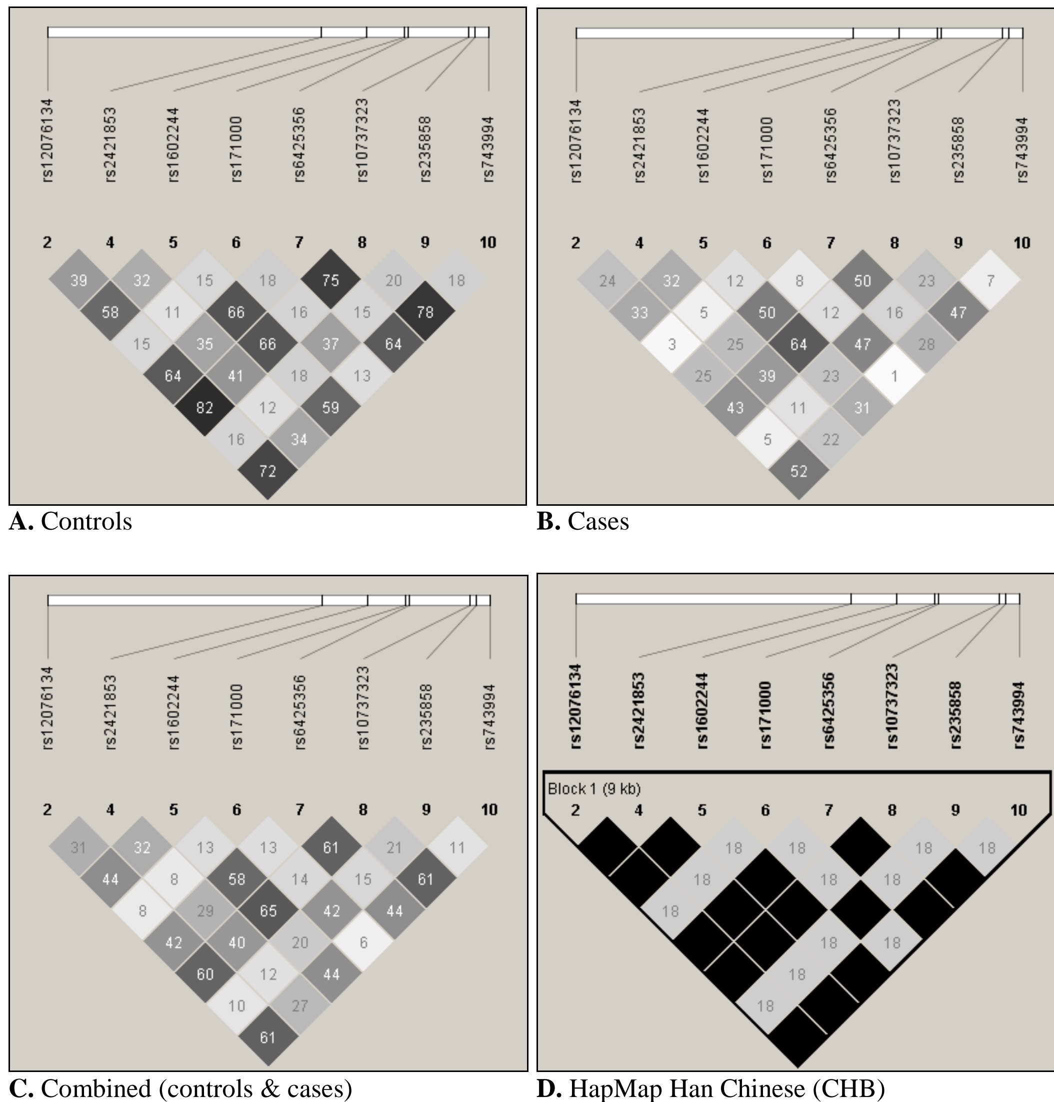


Figure 3.1
Linkage disequilibrium (LD) patterns of *MYOC* single nucleotide polymorphisms in the Han Chinese subjects of the current case-control study and the HapMap database

The LD measures (r^2) are calculated and plotted by Haploview for the 8 SNPs under study for (A) controls, (B) cases (high myopia), (C) controls and cases combined, and (D) HapMap Han Chinese subjects. The r^2 values shown above are the actual values multiplied by 100 (e.g., 44 means 0.44, and 8 means 0.08). The shades of grey indicate the magnitude of the measures with black equal to 100% (i.e. 1.00), which is omitted in the diagram to avoid cluttering the display. Haplotype blocks are constructed based on Gabriel's definition and indicated by triangles encompassing the SNPs involved and their corresponding LD measures. As Haploview cannot handle multiallelic markers, the LD measures for MSs NGA17 and NGA19 are shown separately.

Table 3.8 Linkage disequilibrium (LD) measures (r^2) between *MYOC* microsatellite and SNP markers*

Markers	Controls		Cases		Combined		Parents	
	NGA17 (S1)	NGA19 (S3)	NGA17 (S1)	NGA19 (S3)	NGA17 (S1)	NGA19 (S3)	NGA17 (S1)	NGA19 (S3)
rs12076134 (S2)	0.05	0.10	0.01	0.02	0.03	0.05	**	**
NGA19 (S3)	0.01	-	0.00	-	0.01	-	0.02	-
rs2421853 (S4)	0.03	0.05	0.02	0.01	0.02	0.03	0.07	0.12
rs1602244 (S5)	0.05	0.05	0.04	0.03	0.04	0.04	**	**
rs171000 (S6)	0.09	0.11	0.05	0.10	0.07	0.11	**	**
rs6425356 (S7)	0.04	0.08	0.04	0.03	0.04	0.05	0.09	0.14
rs10737323 (S8)	0.05	0.09	0.05	0.05	0.05	0.07	**	**
rs235858 (S9)	0.13	0.15	0.15	0.12	0.14	0.13	0.21	0.19
rs743994 (S10)	0.04	0.05	0.02	0.01	0.03	0.03	0.08	0.14

* The values of r^2 are shown here for controls, cases (high myopia), combined (controls & cases) and the parents of the nuclear samples. Parental data for rs6425356 (S7) and rs743994 (S10) are parts of the current study while the remaining parental data are taken from Tang et al (2007). Genotype data and hence LD measures are not available for NGA17 and NGA19 microsatellites from the HapMap database.

** Data not available from either the current study or Tang et al (2007).

3.3.3.2 Haplotype analysis

Since rs6425356 (S7), rs10737323 (S8) and rs743994 (S10) contributed main effects to the association, haplotypes were constructed from these three SNPs and compared between cases and controls (**Table 3.9**). The omnibus test was highly significant ($P = 6.03\times10^{-13}$). The OR of each haplotype was estimated with the remaining haplotypes as the reference.

The common haplotype CAG, consisting of all major alleles of these 3 SNPs, were expectedly protective in nature (OR = 0.434) and gave an impressive empirical P value of 1.00×10^{-6}. The estimated frequency of the haplotype CAG was 0.5545 in cases, and 0.7410 in controls. Three other haplotypes consisting of one or two risk alleles of the constituent SNPs were high-risk in nature with ORs ranging from 2.930 to 13.30. Therefore, these high-risk haplotypes were much less common in controls (total frequency = 0.0343) than in cases (total frequency = 0.1552). Note that these risk haplotypes had a much larger effect size than the risk alleles of the constituent SNPs in terms of the ORs: 2.930 – 13.30 (**Table 3.9**) vs 1.414 – 1.876 (**Tables 3.4**).

Table 3.9 *MYOC* haplotype analysis across rs6425356, rs10737323 and rs743994 by Plink*

S7-S8-S10 Haplotype	Haplotype frequency		OR	P	Empirical P
	Cases	Controls			
CAG (111)	0.5545	0.7410	**0.434**	1.55×10^{-11}	$\mathbf{1.00\times10^{-6}}$
CAA (112)	0.0640	0.0106	**5.280**	1.12×10^{-6}	**0.0003**
TA G (211)	0.0670	0.0208	**2.930**	9.81×10^{-5}	**0.0021**
CGA (122)	0.0258	0.0174	1.380	0.3150	0.9452
TAA (212)	0.0242	0.0029	**13.30**	0.0014	**0.0189**
TGG (221)	0.0345	0.0205	1.610	0.1397	0.6785
TGA (222)	0.2300	0.1869	1.300	0.0670	0.3387

* The haplotypes consisting of rs6425356 (S7), rs10737323 (S8) and rs743994 (S10) are analysed by Plink, and the omnibus is highly significant ($P = 6.03\times10^{-13}$). For the sake of easy interpretation, the major allele is designated as 1 and the minor/risk allele as 2 for each SNP; and the haplotypes are also indicated in this 1-2 format. The empirical P values were generated for correction by 1,000,000 permutations.

3.3.4 Replication by family-based association study

Four *MYOC* polymorphisms were first found associated with high myopia in our published family study: NGA17 (S1), NGA19 (S3), rs2421853 (S4) and rs235858 (S9) (Tang et al., 2007). However, they were all negative in the present case-control study (**Tables 3.3** and **3.4**). On the other hand, a few SNPs that were found significantly associated with high myopia in the case-control study were not genotyped in the original nuclear families. Two most top-ranking SNPs were genotyped again for the family samples: rs6425356 (S7) and rs743994 (S10).

3.3.4.1 Analysis of ocular data in myopic offspring

For the sake of clarity and completeness, the demographic and ocular data of the family study are reproduced below (Tang et al., 2007). The family study included 162 nuclear families with a total of 557 subjects (324 parents and 233 highly myopic offspring). Of the recruited families, 95 families had one myopic offspring, 63 families two myopic offspring, and 4 families three myopic offspring. The mean age of the myopic offspring was 24.9 years (SD, 7.5 years) and females were more frequent than males among the offspring. The ocular data of the right eyes of the myopic offspring are summarised in **Table 3.10**. The mean was -8.38 D (SD, -1.90 D) for spherical power and -9.06 D (SD, -2.10 D) for spherical equivalent power.

3.3.4.2 Analysis of genotype data in nuclear families

The FBAT package was used to analyse the family data. The null hypothesis was "no linkage and no association between the polymorphism and any myopia-influencing locus" while the alternate hypothesis was "both linkage and association between the polymorphism and high myopia". One nuclear family with two affected offspring

was removed from analysis because of problematic genotyping. Both rs6425356 (S7) and rs743994 (S10) showed significant association with high myopia under the additive model (**Table 3.11**) – a finding consistent with the results of the case-control study (**Table 3.4**). The two alleles of each SNP showed opposite preferential transmission/non-transmission under the additive model: the Z scores were -2.837 or +2.837 for S7 (P = 0.0046), and -2.817 or +2.817 for S10 (P = 0.0049). The minor alleles of both SNPs showed increased transmission (i.e. positive Z scores) and hence were high-risk alleles – a finding also consistent with the case-control study (**Table 3.4**).

3.3.4.3 Linkage disequilibrium patterns in nuclear families

Four SNPs had been genotyped for both the case-control samples and the family samples with two sets of genotypes done in our previous family study: rs2421853 (S4), rs6425356 (S7), rs235858 (S9) and rs743994 (S10). The LD patterns for the four SNPs were examined for the parents of the nuclear families (**Figure 3.2**). The r^2 values and their distribution were extremely similar to those seen in the Han Chinese of the HapMap database (**Figure 3.1D**). These r^2 values were thus much higher than the corresponding values seen in the case-control subjects (**Figure 3.1A** to **C**).

Table 3.10 Summary of ocular data of 233 myopic siblings from nuclear families (Tang, et al, 2007)

Ocular parameter (unit)	Right eye	
Spherical power, mean±SD (D)	-8.38	± 1.90
SE, mean±SD (D)	-9.06	± 2.01
AXL, mean±SD (mm)	26.88	± 1.05
CP, mean±SD (D)	42.84	± 1.62
ACD, mean±SD (mm)	3.57	± 0.34
LT, mean±SD (mm)	3.72	± 0.43

* SE = spherical equivalent; AXL = axial length; CP = corneal power; ACD = anterior chamber depth; LT = lens thickness.

Table 3.11 *MYOC* single nucleotide polymorphisms: summary of genetic data for parents and association analysis by family-based association testing under additive model in 162 nuclear families*

SNP	rs6425356 (S7)		rs743994 (S10)	
Allele	C	T	G	A
Freq in parents	0.7388	0.2612	0.7368	0.2632
No. of families	101	101	99	99
Z score	-2.837	2.837	-2.817	2.817
P value	0.0046	0.0046	0.0049	0.0049
Global statistic	$\chi^2(1) = 8.048$; $P = 0.0046$**		$\chi^2(1) = 7.934$; $P = 0.0049$**	

* The additive model is being test. The null hypothesis is "no linkage and no association between the polymorphism and high myopia" while the alternate hypothesis is "both linkage and association between the polymorphism and high myopia. "No. of families" refers to the number of informative families in which there was at least one heterozygous parent. $\chi^2(1)$ indicates chi-square statistic with one degree of freedom.

** Both comparisons remained significant after correction by either false discovery rate at a level of 0.05 (cut-off threshold = 0.05) or Bonferroni adjustment (cut-off threshold = 0.025).

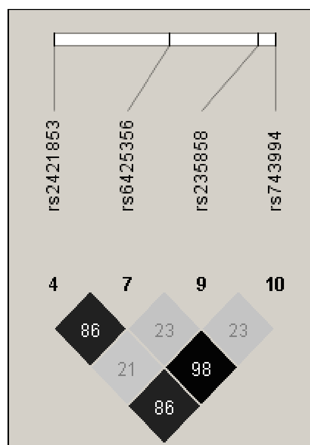


Figure 3.2
Linkage disequilibrium (LD) patterns of *MYOC* single nucleotide polymorphisms in the parents of the current family study

The LD measures (r^2) are calculated and plotted by Haploview for four SNPs genotyped in the parents of the family study. Genotype data for rs2421853 (S4) and rs235858 (S9) are taken from Tang et al (2007) and those for rs6425356 (S7) and rs743994 (S10) are from the current study. The r^2 values shown above are the actual values multiplied by 100 (e.g., 86 means 0.86 and 21 means 0.21). The shades of grey indicate the magnitude of the measures with black equal to 100% (i.e. 1.00), which is omitted in the diagram to avoid cluttering the display.

3.4 Discussion

Myocilin was originally known as the trabecular meshwork-induced glucocorticoid response protein (TIGR), which has been associated with cytoskeletal function. In humans, the gene maps to chromosome 1q23-q24. It is highly expressed in many ocular tissues, most notably in the ciliary body, iris, trabecular meshwork, sclera, choroid, and retina (Tamm, et al., 1999).

In addition to glaucomatous involvement (Stone et al., 1997), genetic variants in *MYOC* have also been implicated in causing susceptibility to high myopia (**Table 3.12**). This potential involvement would be consistent with the increased frequency of myopia in patients with open-angle glaucoma (Nomura, et al., 2004; Wong, et al., 2003), and the observation that intraocular pressure (IOP) is higher in myopes than in emmetropes (Abdalla & Hamdi, 1970). Recently, the Beaver Dam Eye Study also identified a suggestive linkage peak on 1q24 (D1S1589, exact P value = 0.00229) in a non-parametric linkage study investigating refractive error as a quantitative trait and involving 486 extended families with 486 sibling pairs (Klein et al., 2007). It is also noteworthy that some factors that stimulate myocilin expression in trabecular meshwork (Tamm et al., 1999) have also been implicated in the regulation of eye growth (Rohrer & Stell, 1994) and myopia (Lin et al., 2006; Zha et al., 2009) .

3.4.1 Existing *MYOC* studies

Since the first report of the probable association of *MYOC* polymorphisms with high myopia, there are now a total of seven studies investigating this relationship, including the present study. These studies are summarised in **Table 3.12**. They vary

in the ethnicity of the study subjects, study design (population-based or family-based or both), threshold refractive error defining high myopia, sample size, the polymorphisms genotyped and the results. In fact, the results are conflicting and the information currently available in the literature makes meta-analysis across studies very difficult.

Five studies were conducted in Chinese populations in Singapore or Hong Kong, and two studies in Caucasian populations in the UK, the USA or Croatia (**Table 3.12**). In terms of study designs, three studies adopted a case-control approach, two studies employed a family-based association approach, and two studies (including the present one) had both unrelated case-control samples and family samples. With regard to case definition, four studies defined high myopia as -6.0 D or worse (spherical or spherical equivalent), one study used a threshold of -8.0 D, the current study adopted a cut-off of -8.0 D for unrelated cases and -6.0 D for family study, and one study did not specify the threshold. As for sample size, small studies each had <100 cases and <100 controls, and the present study had 300 cases and 300 controls. One study had an extremely biased case: control ratio of 1:53 (Vatavuk, et al., 2009). The largest family study combined 142 British families with 86 American families with a total number of 909 subjects, plus additional number of unrelated cases and controls (Leung et al., 2000; Zayats et al., 2009). Our own family study had 162 families with 324 parents and 233 myopic offspring (Tang et al., 2007).

Table 3.12 Comparison of studies investigating the relationship between high myopia and *MYOC* polymorphisms*

Study	Population	Subjects	Markers tested & results
Wu, et al, 1999	Chinese (Singapore)	97 high myopes (<= -8.0D) 92 matched emmetropes	6 MSs, 1 SNP: Including NGA17 (S1) & NGA19 (S3) NGA17 (S1): relative risk = 0.7 - 4.3 with increasing repeat length
Wu, et al, 2000b	Chinese (Singapore)	104 families each with at least myopic child (threshold not clear)	2 MSs, 1 SNP: Including NGA17 (S1) & NGA19 (S3) NGA19 (S3) : *P* = 0.0014
Leung, et al, 2000	Chinese (Hong Kong)	70 high myopes (<=-6.0D) 69 non-myopic subjects	NGA17 (S1) : *P* = 0.84
Tang, et al, 2007	Chinese (Hong Kong)	162 families each with 1-3 myopic children (Spherical <=-6.0D): 324 parents & 233 myopic offspring	2 MSs, 5 SNPs: Including NGA17 (S1), NGA19 (S3), rs2421853 (S4), rs235858 (S9) NGA17 (S1) : *P* = 0.0084 NGA19 (S3) : *P* = 0.0172 rs2421853 (S4) : *P* = 0.0009 rs235858 (S9) : *P* = 4.0e-6
Zayats, et al, 2009	Europeans (Cardiff, UK)	142 families, total no. of subjects=551 cases=121, controls=116	2 MSs, 15 SNPs: Including NGA17 (S1), NGA19 (S3), rs2421853 (S4), rs235858 (S9)
	Europeans (Durham, USA)	86 families, total no. of subjects=358 cases=56 (High myopia: SE<=-6.0D)	No association for all markers
Vatavuk, et al, 2009	Europeans (Croatia)	16 high myopes (<=-6.0D) 844 controls without high myopia	rs2421853 (S4): *P* = 0.006 (bivariate analysis)** *P* = 0.049 (age- & sex-adjusted analysis)
Current study	Chinese (Hong Kong)	300 high myopes (SE<=-8.0D) 300 emmetropes	2 MSs, 8 SNPs (S1 - S10) rs12076134 (S2) : *P* = 0.0002 rs1602244 (S5) : *P* = 0.0018 rs6425356 (S7) : *P* = 1.98e-6 rs10737323 (S8) : *P* = 0.0081 rs743994 (S10): *P* = 2.81e-6
		162 families each with 1-3 myopic children (Spherical <=-6.0D): 324 parents & 233 myopic offspring	rs6425356 (S7) : *P* = 0.0045 rs743994 (S10): *P* = 0.0049

* MS = microsatellite; SNP = single nucleotide polymorphism; SE = spherical equivalent. The reference numbers (S1 – S10) used in the present study is also indicated for the sake of easy discussion. Polymorphisms whose results are not particularly specified above are negative.

** Based on the genotype counts given in the paper, manual calculation gives *P* = 0.049.

3.4.2 *MYOC* polymorphisms associated with high myopia

Six of the seven studies examined the role of the MS markers in susceptibility to high myopia (**Table 3.12**). High myopia was found associated with the MS NGA17 at the promoter of the *MYOC* gene, but not the MS NGA19 at the 3' flanking region, in Chinese in one case-control study in Singapore (Wu H, 1999). However, the results were just opposite in another family-based Chinese study conducted by the same group in Singapore (Wu H, 2000). Both MS markers were associated with high myopia in Hong Kong Chinese in a family-based study conducted by our group (Tang et al., 2007), but the results could not be replicated in the present case-control study involving a larger size and adopting a more stringent threshold for defining high myopia (**Table 3.3**). We note that the LD between these MSs and other SNPs under study was very weak in the parents of the nuclear families and in the unrelated cases and controls (**Table 3.8**). This means that they are not efficient at all in representing other SNPs. In fact, the positive association of these two MSs with high myopia could be accounted for by two other SNPs investigated in the same family-based study (Tang et al., 2007). Association between these MSs and high myopia was not demonstrated either in the British family cohort (Zayats et al., 2009), which had a very similar sample size as our own family study (**Table 3.12**). Another small case-control study also failed to demonstrate the association between NGA17 and high myopia in Chinese (Leung et al., 2000). Inadequate power could be one of the reasons for negative finding in small studies, as had been suggested by Zayats et al (2009). However, the cumulative evidences tend to suggest that these MSs do not influence the susceptibility to high myopia, and the reported positive findings could probably be chance findings.

The SNP rs235858 (S9) was first found to be associated with high myopia in our family-based study and could apparently account for the other positive markers tested in the same study (Tang et al., 2007). Nevertheless, this positive result failed to be replicated by a very large study, which examined 142 British families (551 subjects), unrelated British cases (121 subjects) and controls (116 subject), 86 American families (358 subjects) and unrelated American cases (56 subjects) (**Table 3.12**; Zayats et al., 2009). This collaborative study collectively investigated 1202 subjects of European origin although the failure rate of genotype calling was 13% for this particular SNP (one of the highest failure rates among all SNPs examined). The results were the same whether only families or all groups of subjects (families, cases and controls) were analysed. Intriguingly, the present case-control study of Chinese subjects also refuted the positive association claimed in our original family-based study, and instead supported the results of the large British-American collaborative study. We then asked the question "What about other SNPs tagged by rs235858 (S9)?" We followed up this by genotyping rs171000 (S6), which was supposed to be in perfect LD with ($r^2 = 1.0$) and efficiently tagged rs235858 (S9) as documented by the HapMap Han Chinese data (**Figure 3.1D**). No association was demonstrated between rs171000 (S6) and high myopia either (**Table 3.4**). Overall, the evidence indicates that rs235858 (S9) was not associated with high myopia and our initial positive association was more likely a chance finding. It is also worth noting that the LD between rs235858 (S9) and rs171000 (S6) was not very strong as observed in our group of 600 case-control subjects ($r^2 = 0.46$; **Figure 3.1C**).

The SNP rs2421853 (S4) was also first found to be associated with high myopia in our family-based study (Tang et al., 2007). In a way similar to rs235858 (S9)

discussed above, the positive results failed to be replicated by both the large British-American collaborative study of Caucasian subjects (Zayats et al., 2009) and the present case-control study of Chinese subjects. We looked up the HapMap Han Chinese database and followed up other SNPs tagged by rs2421853 (S4): $r^2 = 0.96$ for rs12076134 (S2) and $r^2 = 1.00$ for rs1602244 (S5), rs6425356 (S7), rs10737323 (S8) and rs7439994 (S10) (**Figure 3.1D**). Surprisingly, all five SNPs were found to be associated with high myopia under the three genetic models tested (**Table 3.4**) and the significance survived the multiple testing correction by FDR at a level of 0.05 (**Table 3.5**). S7 gave the most significant result under the genotypic and the additive models ($P = 1.03 \times 10^{-5}$ and 1.98×10^{-6}, respectively) while S10 showed the most significant association under the allelic model ($P = 1.41 \times 10^{-6}$). Even more surprising was the weak LD observed between each other for these six SNPs: the combined group of cases and controls showed the highest r^2 value of 0.65 between S5 and S8, and most other r^2 values were much smaller (**Figure 3.1C**). In particular, S4 gave an r^2 value of 0.31 with S2, 0.32 with S5, 0.29 with S7, 0.40 with S8, and 0.27 with S10. In other words, S4 could not tag other SNPs efficiently and, similarly, could not be tagged efficiently by other SNPs either. Now that S7 and S10 gave the most significant results, the same significant results could be achieved by testing S4 (the original positive marker) only if the sample size was increased by a factor of $1/r^2$ (i.e. 1/0.27 or 3.7) to 2220 subjects ($3.7 \times 600 = 2220$ cases and controls combined) – a property of the r^2 value (Pritchard & Przeworski, 2001).

The positive signals now fell upon the other SNPs rather than the original rs2421853 (S4). We wondered whether these new signals could be replicated in our original families. We genotyped the family samples for the two SNPs demonstrating the most

significant association in the case-control samples: rs6425356 (S7) and rs7439994 (S10). Indeed, these two SNPs were also found to be associated with high myopia in the family study: $P = 0.0046$ for S7 and $P = 0.0049$ for S10 (**Table 3.11**). The results were reassuring. S4 was positive in the family study because it *could* tag S7 and S10 very efficiently ($r^2 = 0.86$ for both SNPs; **Figure 3.2**). S4 was negative in the case-control study because it *could not* tag S7 and S10 efficiently ($r^2 = 0.29$ and 0.27; **Figure 3.1C**).

Many factors influence LD patterns: population growth, admixture or migration, population structure, natural selection, variable recombination rates, variable mutation rate, and gene conversion (Ardlie et al., 2002). It is obvious from **Figure 3.1** that the *distribution patterns* of LD in terms of r^2 values were similar for the control group, the case group, the combined group and the HapMap Han Chinese except that the r^2 values were dramatically reduced for subjects of the present study. The reduction in the r^2 values was most extreme for the case group and less extreme for the control group, with the combined group in between. The control group consisted of emmetropic subjects (within ±1.0 D) and the case group high myopes (< -8.0 D or worse). Therefore, both groups were not representative of the general Chinese population in Hong Kong. The parents of the nuclear families were unrelated subjects and, as a group belonging to the "last" generation, had a lower prevalence of myopia. Nevertheless, they were comparatively more representative of the general Chinese population in Hong Kong than the control group or the case group, and had the corresponding r^2 values (**Figure 3.2**) closer to those of the Chinese subjects of the HapMap Project. The subjects in our study were southern Han Chinese. The Chinese subjects genotyped in the HapMap Project were northern

Han Chinese from Beijing. These differences in the Chinese subjects of the groups considered might contribute to the observed differences in the r^2 values although other factors cannot be ruled out.

3.4.3 Summary

In summary, the present study strongly suggests that certain polymorphisms at the 3' end of the *MYOC* gene contribute to the susceptibility to high myopia in Chinese. **First**, a few correlated SNPs (S2, S5, S7, S8 and S10) were found to be associated with high myopia at varying levels of significance in the case-control subjects. **Second**, the most significant association for S7 and S10 were replicated in our original families. It should be noted that our family-based study had a smaller sample size and hence was less powerful when compared to our own case-control study, and that the threshold for defining high myopia was less stringent (-6.0 D vs -8.0 D). **Third**, the discrepancy arising from the initial positive SNP rs2421853 (S4) could be explained by the local variation in the LD patterns *actually observed* in our case-control subjects, in the parents of our nuclear families and in Han Chinese subjects (n = 45) included in the International HapMap Project (**Figures 3.1** and **3.2**). **Fourth**, we reduced genotyping errors, if any, to an extremely low level. As a standard quality control procedure, 5% of the samples were randomly chosen and re-genotyped with very high concordance rates. Occasional discrepancies were resolved by direct DNA sequencing. Two positive SNPs (S5 and S7) were genotyped twice with two different methods (**Table 2.1**) when the first genotyping methods were found to give many ambiguous results. The concordance rates between the methods were very high, and occasional discrepancies were also resolved by direct

DNA sequencing. It was also reassuring that the allele frequencies of S7 and S10 were very similar for the *controls* of the case-control study and the *parents* of the family study (P = 0.3349 for S7 and P = 0.1241 for S10, χ^2 test; **Tables 3.4** and **3.11**).

3.4.4 Future studies

Both forward and backward stepwise logistic regression procedures identified rs6425356 (S7), rs10737323 (S8) and rs7439994 (S10) as the contributors of main effects to the association with high myopia. We used the online tool of SNP Function Prediction (http://manticore.niehs.nih.gov/snpfunc.htm) (Xu & Taylor, 2009) to explore the potential functions of these three SNPs (and the other two positive SNPs). However, these non-coding SNPs were predicted *not* to show such functional changes as transcriptional regulation by altering the activity of transcription factor binding sites and changing of splicing pattern or efficiency by disrupting splice site. Therefore, they might not be the causal SNPs directly associated with the susceptibility to high myopia, but might be in strong LD with another genuine casual SNP in the nearby region. Intriguingly, S2, S4, S5, S7, S8 and S10 (**Figure 3.1D**) capture or represent each other very well, but not other SNPs as now *documented* in HapMap database for Chinese in the 48kb region examined in this study. However, there are many neighboring SNPs that are not genotyped in the HapMap Project. Thus, such neighboring SNPs are worth exploring in this regard. Moreover, in view of the weak LD among SNPs in this region as actually observed in our own case-control samples, we suggest sequence this region (3' flanking region of the *MYOC* gene) for a small number of Chinese subjects (e.g. 40 subjects) to discover

any SNPs not yet identified and to establish the fine-scale LD structures for this region in Chinese. In this regard, data from the 1000 Genomes Project should be very useful (Via, et al., 2010) since the Project includes 100 samples from southern Han Chinese (http://www.1000genomes.org/page.php). From there, we can zoom in to identify the potential casual SNP through additional genetic association study and relevant functional studies. It should also be noted that the *MYOC* gene is bound by two recombination hotspots (**Figure 3.3**). The search for causal variants should be restricted to the region bound by these hotspots since the positive SNPs found in the present study will not tag SNPs outside this region with any appreciable efficiency.

The SNPs found positive in the present study have not been investigated by any other groups. We suggest that the positive SNPs be replicated by independent sample sets, preferably from different populations. When the genotype data are to be analysed, attention should also be paid to the LD pattern actually observed in the study population.

3.4.5 Conclusion

In conclusion, our results highlight the reproducible association of the 3' *MYOC* SNPs with high myopia. Replication studies in other populations are recommended. Additional studies to explore the functional effects of all associated variants and to determine how it modifies the genetic risk will greatly help in understanding the pathophysiology of high myopia.

Figure 3.3
The *MYOC* gene is bound by two hotspots of recombination.

The figure shows the LD patterns for Han Chinese subjects in the HapMap database for the region running from rs235858 to rs12082573 on human chromosome 1 (position 142819774 to 142844986 of Genome Build 36.3 of the NCBI Human Reference sequence). Note the hotspots of recombination on either of the *MYOC* gene. The green lines link the SNPs examined in the present study to the LD pattern on the lower panel.

Chapter 4

A DNA pooling approach to identifying myopia susceptibility genes in the *MYP2* locus

4.1 Overview

Objectives. Association analysis of seven positional candidate genes in the *MYP2* locus was carried out to assess the differences of SNP allelic frequencies between subjects with high myopia and emmetropic controls by the use of pooled DNA samples and prioritizing them for further confirmation in individual samples.

Methods. The study employed two sets of DNA pools (6 case pools and 6 control pools) consisting of DNA samples from 300 high myopia cases and 300 controls. Each DNA pool was constructed by mixing equal amounts of DNA from 50 *distinct* subjects of the same affection status. To screen 62 tag SNPs from 7 candidate genes in the *MYP2* interval, estimated allele frequencies of SNPs were obtained by analysis of primer-extended products in a denaturing high performance liquid chromatography system, and then compared across replicates of each pool and across two sets of pools by means of nested ANOVA. The most promising SNPs ($P \leq 0.10$) were further evaluated by individual genotyping of samples included in the pools.

Results. Nine SNPs exhibiting significant allele frequency differences between case pools and controls pools were genotyped in individual samples. One SNP within the *LPIN2* gene was found to be associated with high myopia and survived correction

for multiple comparisons by FDR at a level of 0.05: rs589318 (P = 0.0015 under the genotypic model, 0.0006 under the additive model, and 0.0005 under the allelic model). For rs589318, the C allele (minor allele) was less common in the cases than in the controls, and hence was protective against high myopia with OR significantly less than 1.000.

Conclusions. We investigated potential myopia susceptibility genes in the *MYP2* locus and found the *LPIN2* gene to be associated with high myopia. Replication studies should be performed using independent sample sets.

4.2 Background of Myopia-2 (*MYP2*)

Identification of the *MYP2* Locus

Recent family linkage analysis identified a genomic region on chromosome 18p11 (*MYP2*, OMIM. 160700) closely linked with high myopia with a maximum cumulative logarithm of the odds (LOD) score of 9.59 (Young et al., 1998b). The core region identified extends 7.6 cM from marker D18S481 to D18S52 **(Figure 1.11)**. Refinement of the region by transmission disequilibrium test suggests that candidate genes for this locus are likely in an interval between markers D18S63 and D18S52 (Young et al., 2001).

This locus has been confirmed through replication studies in independent family studies with different population samples (Heath et al., 2001; Lam et al., 2003b; Naiglin et al., 1999). One successful replication study used a Hong Kong Chinese cohort of 6 myopic families, finding linkage between markers D18S476 and D18S62

(Lam et al., 2003b). Another replication study was a Sardinian Italian cohort study of 15 myopic individuals with a genetically homogeneous background, showing strongest linkage at marker D18S63 (Heath et al., 2001). In contrast, there were a few other studies reporting negative results (Chang et al., 2008; Farbrother et al., 2004b; Ibay et al., 2004; Yamane et al., 2007). Nearly all of the present linkage studies are insufficiently powered to test the locus for high myopia. The replication of linkage signals requires a considerably larger sample size than the original cohort used for the initial discovery.

Candidate genes within the *MYP2* critical region

All genes that map within the *MYP2* critical region are candidate disease genes based on position. There are 9 known and 20 hypothetical genes within the 2.2 cM interval. Seven positional biologically plausible candidate genes were prioritized: clusterin-like 1 (*CLUL1*) (Zhang et al., 2003), elastin-microfibril located interface protein (*EMILIN-2*) (Doliana et al., 2001), lipin 2 (*LPIN2*) (Zhou & Young, 2005), myomesin 1 (*MYOM1*) (Wiesen et al., 2007), myosin, light chain 12A, regulatory (*MYL12A),* myosin, light chain 12B, regulatory (*MYL12B)* (Redowicz, 2002; Satpathy et al., 2004), and zinc finger protein 161 homolog (*ZFP161*) (Wang et al., 2004). It is valuable to test whether reported associations within these genetic intervals actually account for the linkage evidence that has been reported.

We hypothesised that the *MYP2* locus also contained common genetic variants associated with high myopia, and conducted a case-control study using 62 SNPs located primarily in 7 positional candidate genes within the *MYP2* region. We first used a screening approach based on DNA pooling, and followed up the top-ranking

SNPs with individual genotyping.

4.3 Results

In total, 12 DNA pools were constructed with 6 case pools and 6 control pools, each of which was prepared by mixing equal amounts of DNA from 50 *distinct* subjects of the same affection status. Each DNA pool was separately amplified, primer-extended and analysed by DHPLC in triplicates.

4.3.1 Elution times of extension products

The shortest mean elution time was 3.27 minutes for the extension product of allele C of rs43404011 of the *MYOM1* gene while the longest mean elution time was 10.33 minutes for the extension product of allele T of rs7238703 of the *MYOM1* gene (**Table 4.1**). The elution times were very reproducible with most of the coefficient of variation (CV) below 2% except for 3 SNPs (rs485562 of *CLUL1*, rs7226712 of *EMLIN2* and rs3819090 of *LPIN2*). The elution times of these 3 SNPs were quite varied with CV ranging from 8.07% to 16.38%. Overall, the mean CV was 1.27% for the elution times of both the first and the second extension products across all SNPs examined.

We observed that extension products were eluted in the following order with only occasional exceptions: (short elution time) C allele < G allele < T allele < A allele (long elution time) (**Table 4.1**). The only exception was rs3786458 of *MYCL12B*, where the extension product of A allele was eluted more quickly than that of T allele.

This general trend of elution order reflects the differential interaction between the incorporated ddNTPs and the stationary phase of the DNASep column in the presence of increasing concentration of acetonitrile.

4.3.2 k correction factors

The correction factors were estimated from heterozygous individuals who provide a standard for a 1:1 ratio for a pair of peak intensities of two heterozygous alleles. The k correction factor was defined as the height of the first extension product divided by the height of the second extension product, irrespective of whether the first eluted allele was the major or the minor allele. Error was reduced by using more heterozygotes to estimate k and/or more replicates (x3) from a single heterozygote, and by using replicate (x3) samples of the pools. In practice, this has implications for resource allocation because a balance needs to be struck between the number of SNPs to be tested and the number of replicate pools per SNP. It is time-consuming to search heterozygous for the k factors in the pilot study ($p = 1/10$, MAF>0.1). In the present study, the estimated k correction factor was based on the mean of three replicate readings from a sample heterozygous for the SNP concerned. The three replicates were from separate PCR and dHPLC runs. Simulation results (Visscher & Le Hellard 2003) indicated that the precision of estimation does not need to be high. For example, a scenario where the standard error of the estimate of k is less than 30% of the mean value, the impact on type I error was negligible. However, failing to estimate k, by implicitly assuming that the peak ratio is unity, gives a systematic bias in the test unless the true value is close to unity. **Figure 4.1** shows the elution profiles of heterozygous samples for three SNPs with different values of the k correction factors. For the 62 SNPs studied, the mean was 1.21 and the median 1.09

(**Table 4.2**). The minimum was 0.30 and the maximum 4.37 while the 5th percentile was 0.79 and 95th percentile 2.11.

Table 4.1 DHPLC analysis of SNPs in the *MYP2* locus: elution times of extension products*

Gene	SNP	1st peak			2nd peak		
		Allele	Mean ± SD (min)	CV (%)	Allele	Mean ± SD (min)	CV (%)
CLUL1	rs546696	G	5.79 ± 0.01	0.23	T	6.56 ± 0.01	0.22
CLUL1	rs1004961	C	3.64 ± 0.04	1.03	A	3.96 ± 0.04	0.98
CLUL1	rs505140	G	9.07 ± 0.04	0.49	T	9.34 ± 0.04	0.47
CLUL1	rs485562	G	5.30 ± 0.43	**8.16**	A	5.46 ± 0.44	**8.07**
CLUL1	rs11662827	C	4.27 ± 0.09	2.01	G	4.49 ± 0.06	1.27
CLUL1	rs11661043	C	6.54 ± 0.01	0.16	T	7.12 ± 0.01	0.17
CLUL1	rs11660005	G	7.84 ± 0.02	0.25	A	8.23 ± 0.02	0.21
CLUL1	rs11663153	C	5.72 ± 0.03	0.46	A	6.42 ± 0.03	0.48
CLUL1	rs2342700	C	4.28 ± 0.04	0.86	G	4.56 ± 0.02	0.45
CLUL1	rs9966612	G	6.74 ± 0.05	0.67	A	7.10 ± 0.01	0.16
EMILIN2	rs4797088	C	4.63 ± 0.03	0.64	G	5.07 ± 0.07	1.32
EMILIN2	rs680173	C	5.14 ± 0.01	0.28	G	5.38 ± 0.01	0.26
EMILIN2	rs7226712	C	6.59 ± 0.78	**11.80**	T	7.24 ± 0.85	**11.70**
EMILIN2	rs637647	C	3.59 ± 0.06	1.63	T	3.75 ± 0.07	1.99
EMILIN2	rs604050	T	7.93 ± 0.02	0.29	A	8.09 ± 0.02	0.30
EMILIN2	rs6506038	G	4.54 ± 0.02	0.39	A	4.77 ± 0.02	0.36
EMILIN2	rs16944003	C	3.69 ± 0.03	0.77	T	4.05 ± 0.02	0.60
EMILIN2	rs1790994	C	5.68 ± 0.02	0.32	T	6.30 ± 0.02	0.32
EMILIN2	rs1059281	C	5.22 ± 0.04	0.69	G	5.48 ± 0.02	0.30
LPIN2	rs1985	T	3.91 ± 0.02	0.60	A	4.33 ± 0.02	0.49
LPIN2	rs16944051	C	3.99 ± 0.03	0.70	T	4.18 ± 0.02	0.58
LPIN2	rs3819090	G	5.04 ± 0.83	**16.38**	T	5.70 ± 0.92	**16.20**
LPIN2	rs10460009	C	3.87 ± 0.02	0.61	T	4.37 ± 0.02	0.52
LPIN2	rs1628891	G	5.11 ± 0.01	0.26	A	5.43 ± 0.01	0.22
LPIN2	rs2298786	C	3.96 ± 0.03	0.77	T	4.40 ± 0.03	0.57
LPIN2	rs589318	C	5.03 ± 0.03	0.54	T	5.29 ± 0.03	0.50
LPIN2	rs16944193	C	3.83 ± 0.02	0.47	T	4.00 ± 0.05	1.24
MYOM1	rs1042731	C	6.49 ± 0.01	0.21	T	7.10 ± 0.01	0.20
MYOM1	rs4613146	C	7.19 ± 0.02	0.27	A	7.94 ± 0.02	0.25
MYOM1	rs7233983	C	4.33 ± 0.11	2.61	A	4.91 ± 0.13	2.55
MYOM1	rs4413045	C	3.76 ± 0.06	1.71	T	4.02 ± 0.05	1.17
MYOM1	rs11081004	G	3.34 ± 0.05	1.47	A	3.58 ± 0.03	0.88

(Continued on next page)

Table 4.1 (continued)

Gene	SNP	1st peak			2nd peak		
		Allele	Mean ± SD (min)	CV (%)	Allele	Mean ± SD (min)	CV (%)
MYOM1	rs6506057	G	3.49 ± 0.04	1.02	A	3.63 ± 0.05	1.30
MYOM1	rs1071600	G	4.51 ± 0.02	0.38	A	4.73 ± 0.02	0.35
MYOM1	rs9948582	G	3.87 ± 0.03	0.68	A	4.36 ± 0.02	0.47
MYOM1	rs6506074	C	3.86 ± 0.03	0.70	T	4.23 ± 0.02	0.52
MYOM1	rs9952207	G	4.16 ± 0.07	1.70	A	4.34 ± 0.07	1.66
MYOM1	rs4798069	C	3.84 ± 0.03	0.71	T	4.01 ± 0.02	0.60
MYOM1	rs7235847	G	4.51 ± 0.16	3.61	A	4.80 ± 0.16	3.35
MYOM1	rs9951849	G	7.38 ± 0.01	0.14	T	7.68 ± 0.01	0.13
MYOM1	rs7238703	C	9.99 ± 0.02	0.18	T	10.33 ± 0.02	0.18
MYOM1	rs8091916	C	3.85 ± 0.03	0.78	A	4.01 ± 0.03	0.69
MYOM1	rs4340411	C	3.27 ± 0.02	0.67	T	3.50 ± 0.02	0.57
MYOM1	rs8090956	C	3.78 ± 0.03	0.71	T	4.09 ± 0.02	0.55
MYOM1	rs1662315	C	4.63 ± 0.02	0.33	T	4.82 ± 0.01	0.28
MYOM1	rs9947162	G	3.86 ± 0.02	0.60	A	4.29 ± 0.02	0.47
MYOM1	rs4441365	C	4.53 ± 0.02	0.44	T	4.76 ± 0.02	0.37
MYOM1	rs4507002	G	3.88 ± 0.03	0.72	A	4.25 ± 0.02	0.58
MYOM1	rs12605942	G	4.03 ± 0.03	0.79	A	4.36 ± 0.04	0.90
MYCL12B	rs1662347	C	4.71 ± 0.04	0.81	A	4.80 ± 0.04	0.81
MYCL12B	rs717183	G	3.85 ± 0.08	2.09	A	4.27 ± 0.03	0.70
MYCL12B	rs3786458	A	4.23 ± 0.05	1.20	T	4.31 ± 0.05	1.14
MYCL12B	rs1662342	G	3.79 ± 0.04	0.96	A	3.92 ± 0.03	0.66
MYCL12A	rs1662336	G	9.16 ± 0.01	0.13	A	9.49 ± 0.01	0.12
MYCL12A	rs949303	G	3.80 ± 0.03	0.85	A	4.10 ± 0.03	0.69
MYCL12A	rs1630702	C	4.95 ± 0.04	0.83	T	5.74 ± 0.05	0.86
MYCL12A	rs1791067	C	4.33 ± 0.02	0.37	T	4.56 ± 0.01	0.30
MYCL12A	rs7239576	C	4.67 ± 0.02	0.46	T	4.93 ± 0.02	0.39
MYCL12A	rs6506094	C	4.67 ± 0.01	0.25	A	4.92 ± 0.01	0.28
ZFP161	rs2789	C	4.73 ± 0.02	0.41	G	4.99 ± 0.02	0.40
ZFP161	rs990072	C	4.99 ± 0.02	0.30	T	5.52 ± 0.01	0.22
ZFP161	rs620652	G	6.10 ± 0.06	1.01	A	6.72 ± 0.06	0.90

* SNPs are arranged down the column in ascending chromosomal positions. The elution times indicated above are each the mean of 39 measurements (12 pools each in triplicates plus 3 replicates for a heterozygous sample).

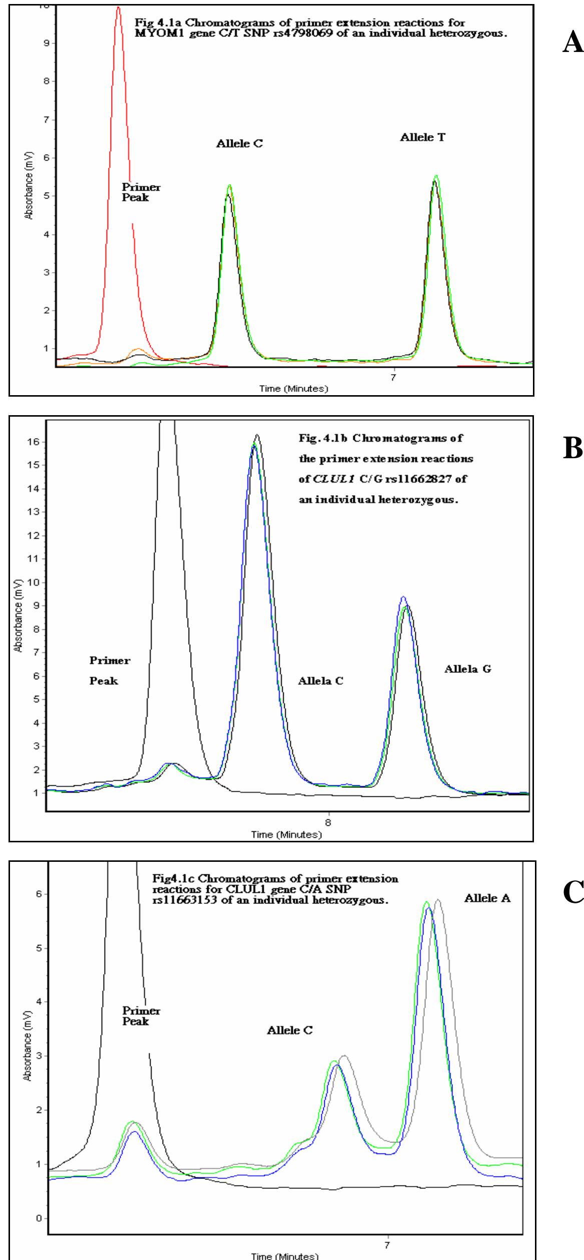


Figure 4.1
Derivation of the k correction factors from heterozygous individuals

The DHPLC chromatograms show the elution profiles of the PCR primer and the extension products for SNPs with different values of the k correction factors. The mean k correction factor is (A) 0.97 for rs4798069 (*MYOM1*), (B) 1.30 for rs11662827 (*CLUL1*), and (C) 0.42 for rs11663153 (*CLUL1*). The peaks are identified in the order of increasing elution time as follows: *the first* unextended primer, *the second and third are* the two extension products of each SNP. The peak heights are equal to the signal strengths (absorbance) of the two alleles. The k correction factor is the height of the second peak (i.e. first extension product) divided by the height of the third peak (i.e. second extension product). Note that the scales of the x- and y-axes are different for the chromatograms shown.

Table 4.2 Pooled DNA analysis of SNPs in the *MYP2* locus: k correction factor and estimated allele frequencies and their differences*

		k correction	Estimated allele frequencies			nANOVA
Gene	SNP	factor	Controls	Cases	Difference	*P* value
CLUL1	**rs546696**	**1.11**	**0.6900**	**0.6696**	**0.0204**	**0.0429**
CLUL1	**rs1004961**	**1.10**	**0.4974**	**0.4847**	**0.0127**	**0.0031**
CLUL1	rs505140	0.96	0.5861	0.5687	0.0174	0.2414
CLUL1	rs485562	0.95	0.4060	0.4017	0.0043	0.7469
CLUL1	rs11662827	1.30	0.5065	0.4744	0.0321	0.1782
CLUL1	rs11661043	0.42	0.5089	0.4899	0.0190	0.5469
CLUL1	rs11660005	1.01	0.7834	0.7717	0.0117	0.6925
CLUL1	rs11663153	2.92	0.5307	0.5393	-0.0086	0.1670
CLUL1	rs2342700	1.17	0.3037	0.2821	0.0216	0.2317
CLUL1	rs9966612	1.12	0.7753	0.7614	0.0139	0.4657
EMILIN2	rs4797088	0.98	0.4934	0.4943	-0.0009	0.9704
EMILIN2	rs680173	1.12	0.6071	0.6017	0.0054	0.8180
EMILIN2	rs7226712	1.10	0.7406	0.7203	0.0203	0.2681
EMILIN2	**rs637647**	**1.22**	**0.3672**	**0.4208**	**-0.0536**	**0.0204**
EMILIN2	rs604050	1.08	0.6267	0.5951	0.0316	0.2069
EMILIN2	rs6506038	1.12	0.3590	0.3676	-0.0086	0.3282
EMILIN2	rs16944003	1.13	0.5156	0.5162	-0.0006	0.9776
EMILIN2	rs1790994	1.19	0.4482	0.4527	-0.0045	0.7640
EMILIN2	rs1059281	1.10	0.5902	0.5764	0.0138	0.1490
LPIN2	rs1985	0.99	0.6842	0.6889	-0.0047	0.8229
LPIN2	rs16944051	0.93	0.3519	0.3902	-0.0383	0.1233
LPIN2	rs3819090	1.12	0.4291	0.4025	0.0266	0.2758
LPIN2	rs10460009	0.88	0.4423	0.4658	-0.0235	0.4179
LPIN2	rs1628891	0.93	0.3839	0.4161	-0.0322	0.1081
LPIN2	rs2298786	0.92	0.5143	0.5332	-0.0189	0.4243
LPIN2	**rs589318**	**1.33**	**0.4218**	**0.4581**	**-0.0363**	**0.0407**
LPIN2	rs16944193	0.93	0.3655	0.4009	-0.0354	0.2897
MYOM1	rs1042731	1.10	0.5426	0.5761	-0.0335	0.1054
MYOM1	rs4613146	1.05	0.404	0.4086	-0.0046	0.8511
MYOM1	rs7233983	1.00	0.5165	0.4939	0.0226	0.4837
MYOM1	rs4413045	1.14	0.4557	0.4482	0.0075	0.6791
MYOM1	rs11081004	2.13	0.4174	0.4123	0.0051	0.7845
MYOM1	rs6506057	1.07	0.7562	0.741	0.0152	0.4277

(Continued on next page)

Table 4.2 (continued)

Gene	SNP	k correction factor	Estimated allele frequencies			nANOVA
			Controls	Cases	Difference	*P* value
MYOM1	rs1071600	4.37	0.3372	0.3148	0.0224	0.1717
MYOM1	**rs9948582**	**1.08**	**0.4132**	**0.4646**	**-0.0514**	**0.0512**
MYOM1	rs6506074	1.41	0.5602	0.5659	-0.0057	0.7518
MYOM1	rs9952207	1.02	0.4004	0.4276	-0.0272	0.2978
MYOM1	rs4798069	0.30	0.5365	0.5557	-0.0192	0.3810
MYOM1	**rs7235847**	**0.97**	**0.1989**	**0.2251**	**-0.0262**	**0.0694**
MYOM1	rs9951849	0.91	0.6458	0.6028	0.0430	0.2059
MYOM1	rs7238703	0.79	0.4158	0.3688	0.0470	0.1797
MYOM1	rs8091916	1.73	0.4961	0.501	-0.0049	0.3514
MYOM1	**rs4340411**	**1.08**	**0.5568**	**0.5891**	**-0.0323**	**0.0035**
MYOM1	rs8090956	1.58	0.5342	0.5658	-0.0316	0.1453
MYOM1	rs1662315	1.28	0.4006	0.3762	0.0244	0.3984
MYOM1	rs9947162	0.83	0.5704	0.5311	0.0393	0.1439
MYOM1	rs4441365	1.27	0.3544	0.3658	-0.0114	0.3851
MYOM1	rs4507002	1.13	0.5282	0.5328	-0.0046	0.5491
MYOM1	**rs12605942**	**1.05**	**0.6475**	**0.6042**	**0.0433**	**0.0756**
MYCL12B	rs1662347	0.66	0.3911	0.3845	0.0066	0.5544
MYCL12B	rs717183	1.21	0.5577	0.5475	0.0102	0.6613
MYCL12B	rs3786458	1.13	0.2466	0.2315	0.0151	0.4035
MYCL12B	rs1662342	1.17	0.4934	0.4917	0.0017	0.3571
MYCL12A	rs1662336	1.00	0.1658	0.1730	-0.0072	0.4795
MYCL12A	rs949303	1.06	0.3477	0.3633	-0.0156	0.3790
MYCL12A	rs1630702	1.06	0.5118	0.4958	0.0160	0.6468
MYCL12A	**rs1791067**	**1.20**	**0.3516**	**0.3943**	**-0.0427**	**0.0879**
MYCL12A	rs7239576	1.06	0.3257	0.3558	-0.0301	0.1662
MYCL12A	rs6506094	1.31	0.8077	0.8175	-0.0098	0.2420
ZFP161	rs2789	1.03	0.3247	0.3492	-0.0245	0.3077
ZFP161	rs990072	1.06	0.4443	0.4361	0.0082	0.8142
ZFP161	rs620652	3.55	0.7602	0.7667	-0.0065	0.8032

* SNPs are arranged down the column in ascending chromosomal positions. The estimated allele frequencies refer to those of the first eluted allele (shorter elution time), disregarding whether it is a major or minor allele. The difference (controls – cases) in estimated allele frequencies is tested by nested analysis of variance (nANOVA). If the *P* value is <0.10, the SNP and its corresponding data are highlighted in **boldface** for the sake of easy recognition.

4.3.3 Estimating allele frequencies in case and control pools

For the sake of easy manipulation and calculation, the allele frequencies estimated refer to those of the first eluted allele (shorter elution time), irrespective of whether it was a major or minor allele (**Table 4.2**). For the control pools, the estimated frequencies of the first eluted alleles ranged from 0.1658 to 0.8077. For the case pools, they ranged from 0.1730 to 0.8175. The estimated differences in allele frequencies between two sets of DNA pools (controls – cases) ranged from -0.0536 to 0.0470 and the *P* values from nested ANOVA ranged from 0.9776 to 0.0031 (**Table 4.2**). Nine SNPs gave putative significant difference in the estimated allele frequencies ($P \leq 0.10$) and were thus genotyped for individual samples.

4.3.4 Tests of genetic association for individual genotyping

There were more females in the case group than in the control group. Chi-square tests indicated that there was no significant difference ($P > 0.05$) in the distribution of genotypes between the two genders for putative significant polymorphisms except for rs12605942 of the *MYOM1* gene (**Table 4.3**). For rs12605942, the difference in genotype distribution was significant for both the control group ($P = 0.0106$) and the case group ($P = 0.0050$). This justified the direct comparison of genetic data between cases and controls *without* stratification into males and females for 8 SNPs while Mantel-Haenszel test controlling for the gender was used to test the association between of rs12605942 and high myopia. It is interesting to note that the allele frequencies estimated by DNA pooling method were in general higher than the corresponding ones estimated by individual genotyping (**Tables 4.2** and **4.4**) except for rs7235847 of *MYOM1*.

Table 4.3 Comparison of *MYP2* genotype frequencies between males and females within the control and the case groups

Gene	SNP	*P* value (χ^2 test)	
		Controls	Cases
CLUL1	rs546696	0.5646	0.4353
CLUL1	rs1004961	0.8976	0.8824
EMILIN2	rs637647	0.5579	0.0463
LPIN2	rs589318	0.1699	0.5321
MYOM1	rs9948582	0.5614	0.4428
MYOM1	rs7235847	0.9933	0.9984
MYOM1	rs4340411	0.0943	0.1991
MYOM1	**rs12605942**	**0.0106**	**0.0050**
MYCL12B	rs1791067	0.5079	0.6788

The genotypes of all polymorphisms were in Hardy-Weinberg equilibrium ($P > 0.05$) for the control group and the case group, except rs9948582 ($P = 0.0378$) and rs7235847 ($P = 2.36\times10^{-8}$) of *MYOM1* in the case group (**Tables 4.4**). Therefore, all polymorphisms were tested for association with high myopia. At the nominal significance level of 0.05, only one SNP showed significant differences in distribution of genotypes and alleles between cases and controls under all three genetic models: rs589318 of the *LPIN2* gene ($P = 0.0015$ for genotypic model, 0.0006 for additive model, and 0.0005 for allelic model).

In total, there were 89 comparisons for correction: 62 comparisons of estimated allele frequencies between case pools and control pools for 62 SNPs in the *MYP2* locus (**Table 4.2**), and 27 comparisons for 9 SNPs individually genotyping and each analysed under three genetic models (**Table 4.4**). At a FDR at a level of 0.05, the FDR-adjusted significance level was set at 0.0017 (**Table 4.5**). The same SNP (rs589318 of *LPIN2*) remained significant under all three genetic models.

For rs589318 of the *LPIN2* gene, we examined the ORs with reference to the major allele (T allele) or the common homozygote (TT). The ORs were all less than 1.000 under the three genetic models tested, indicating that the C allele or C-containing genotypes were protective in nature. The genotypic OR was 0.775 (95% CI: 0.544 – 1.104) for C/T heterozygote, and 0.409 (95% CI, 0.250 – 0.670) for the C/C homozygote under the genotypic model (**Table 4.4**). The common OR was 0.671 (95% CI: 0.534 – 0.843) for an increase of one copy of the C allele under the additive model. The allelic OR was 0.662 (95% CI: 0.524 – 0.836) for the C allele under the allelic model.

Table 4.4 *MYP2* SNPs determined by individual genotyping: frequency distribution in cases and controls, and association analysis*

SNP (Gene)	Gt	Controls No. (Freq)	Cases No. (Freq)	HWE test Controls *P* value	HWE test Cases *P* value	Genotypic test** OR (945% CI)	Genotypic test** χ^2 *P* value	Trend test** Common OR (945% CI)	Trend test** χ^2 *P* value	Allele	Controls No. (Freq)	Cases No. (Freq)	Allelic test** OR (95% CI)	Allelic test** χ^2 *P* value
rs546696	G/G	83 (0.277)	83 (0.277)	0.7532	0.9606	(reference)	0.9531	0.980 (0.782 - 1.228)	0.8630	G	313 (0.522)	316 (0.527)	(reference)	0.8623
(*CLUL1*)	G/T	147 (0.490)	150 (0.500)			1.020 (0.698 - 1.492)				T	287 (0.478)	284 (0.473)	0.980 (0.781 - 1.230)	
	T/T	70 (0.233)	67 (0.223)			0.957 (0.609 - 1.505)								
rs1004961	C/C	142 (0.473)	132 (0.440)	0.4892	0.9993	(reference)	0.7073	1.085 (0.854 - 1.377)	0.5032	C	409 (0.682)	398 (0.663)	(reference)	0.4986
(*CLUL1*)	A/C	125 (0.417)	134 (0.447)			1.153 (0.821 - 1.620)				A	191 (0.318)	202 (0.337)	1.087 (0.854 - 1.383)	
	A/A	33 (0.110)	34 (0.113)			1.108 (0.650 - 1.891)								
rs637647	T/T	136 (0.453)	122 (0.407)	0.2295	0.3789	(reference)	0.4876	1.147 (0.912 - 1.442)	0.2419	T	397 (0.662)	377 (0.628)	(reference)	0.2276
(*EMILIN2*)	C/T	125 (0.417)	133 (0.443)			1.186 (0.840 - 1.676)				C	203 (0.338)	223 (0.372)	1.157 (0.913 - 1.466)	
	C/C	39 (0.130)	45 (0.150)			1.286 (0.785 - 2.107)								
rs589318	T/T	97 (0.323)	126 (0.420)	0.2830	0.6007	(reference)	**0.0015**	0.671 (0.534 - 0.843)	**0.0006**	T	333 (0.555)	392 (0.653)	(reference)	**0.0005**
(*LPIN2*)	C/T	139 (0.463)	140 (0.467)			0.775 (0.544 - 1.104)				C	267 (0.445)	208 (0.347)	0.662 (0.524 - 0.836)	
	C/C	64 (0.213)	34 (0.113)			0.409 (0.250 - 0.670)								
rs9948582	A/A	89 (0.297)	78 (0.260)	0.3302	0.0378	(reference)	0.0944	0.973 (0.773 - 1.224)	0.8142	A	319 (0.532)	323 (0.538)	(reference)	0.8169
(*MYOM1*)	A/G	141 (0.470)	167 (0.557)			1.351 (0.926 - 1.972)				G	281 (0.468)	277 (0.462)	0.974 (0.776 - 1.222)	
	G/G	70 (0.233)	55 (0.183)			0.897 (0.563 - 1.429)								

(Continued on next page)

Table 4.4 (continued)

SNP		Controls	Cases	HWE test Controls	HWE test Cases	Genotypic test**		Trend test**			Controls	Cases	Allelic test**	
(Gene)	Gt	No. (Freq)	No. (Freq)	*P* value	*P* value	OR (945% CI)	χ^2 *P* value	Common OR (945% CI)	χ^2 *P* value	Allele	No. (Freq)	No. (Freq)	OR (95% CI)	χ^2 *P* value
rs7235847	A/A	102 (0.340)	72 (0.240)	0.0593	2.36×10^{-8}	(reference)	0.0943	1.233 (0.952 - 1.598)	0.1128	A	363 (0.605)	339 (0.565)	(reference)	0.1597
(*MYOM1*)	A/G	159 (0.530)	195 (0.650)			1.737 (1.203 - 2.508)				G	237 (0.395)	261 (0.435)	1.179 (0.937 - 1.484)	
	G/G	39 (0.130)	33 (0.110)			1.199 (0.689 - 2.084)								
rs4340411	C/C	133 (0.443)	125 (0.417)	0.7943	0.2638	(reference)	0.7579	1.056 (0.827 - 1.358)	0.6636	C	401 (0.668)	394 (0.657)	(reference)	0.6691
(*MYOM1*)	C/T	135 (0.450)	144 (0.480)			1.135 (0.809 - 1.593)				T	199 (0.332)	206 (0.343)	1.054 (0.829 - 1.339)	
	T/T	32 (0.107)	31 (0.103)			1.031 (0.594 - 1.788)								
rs12605942[@]	G/G	146 (0.487)	141 (0.470)	0.2882	0.2085	(reference)	0.9189	1.054 (0.814 - 1.367)	0.6862	G	424 (0.707)	418 (0.697)	(reference)	0.6961
(*MYOM1*)	A/G	132 (0.440)	136 (0.453)			1.075 (0.772 - 1.498)				A	176 (0.293)	206 (0.303)	1.051 (0.817 - 1.353)	
	A/A	22 (0.073)	23 (0.077)			1.075 (0.572 - 2.022)								
rs1791067	T/T	115 (0.383)	136 (0.453)	0.0702	0.1711	(reference)	0.0583	0.908 (0.715 - 1.153)	0.4289	T	383 (0.638)	396 (0.660)	(reference)	0.4316
(*MYCL12B*)	C/T	153 (0.510)	124 (0.413)			0.685 (0.486 - 0.966)				C	217 (0.362)	204 (0.340)	0.909 (0.717 - 1.153)	
	C/C	32 (0.107)	40 (0.133)			1.057 (0.624 - 1.790)								

* SNP = single nucleotide polymorphism; Gt = genotype; HWE = Hardy-Weinberg equilibrium; OR = odds ratio; CI = confidence interval

** The *P* values for association analyses that remain significant after correction for multiple comparisons at a false discovery rate of 0.05 are shown in **boldface**. For details, please refer to **Table 4.3.**

@ Association between the SNP rs12605942 and high myopia was examined using Mantel-Haenszel tests controlling for gender

Table 4.5 Correction for multiple comparisons by false discovery rate (FDR) at a level of 0.05 for association analysis of *MYP2* markers*

Observed *P* values (P_j)	Rank (j)	FDR thresholds (0.05 × j/89)	Observed *P* values (P_j)	Rank (j)	FDR thresholds (0.05 × j/89)	Observed *P* values (P_j)	Rank (j)	FDR thresholds (0.05 × j/89)
0.0005	1	0.0006	0.2276	31	0.0174	0.5491	61	0.0343
0.0006	2	0.0011	0.2317	32	0.0180	0.5544	62	0.0348
0.0015	**3**	**0.0017**	0.2414	33	0.0185	0.6468	63	0.0354
0.0031	4	0.0022	0.2419	34	0.0191	0.6613	64	0.0360
0.0035	5	0.0028	0.242	35	0.0197	0.6636	65	0.0365
0.0204	6	0.0034	0.2681	36	0.0202	0.6691	66	0.0371
0.0407	7	0.0039	0.2758	37	0.0208	0.6791	67	0.0376
0.0429	8	0.0045	0.2897	38	0.0213	0.6862	68	0.0382
0.0512	9	0.0051	0.2978	39	0.0219	0.6925	69	0.0388
0.0583	10	0.0056	0.3077	40	0.0225	0.6961	70	0.0393
0.0694	11	0.0062	0.3282	41	0.0230	0.7073	71	0.0399
0.0756	12	0.0067	0.3514	42	0.0236	0.7469	72	0.0404
0.0879	13	0.0073	0.3571	43	0.0242	0.7518	73	0.0410
0.0943	14	0.0079	0.3790	44	0.0247	0.7579	74	0.0416
0.0944	15	0.0084	0.3810	45	0.0253	0.7640	75	0.0421
0.1054	16	0.0090	0.3851	46	0.0258	0.7845	76	0.0427
0.1081	17	0.0096	0.3984	47	0.0264	0.8032	77	0.0433
0.1128	18	0.0101	0.4035	48	0.0270	0.8142	78	0.0438
0.1233	19	0.0107	0.4179	49	0.0275	0.8142	79	0.0444
0.1439	20	0.0112	0.4243	50	0.0281	0.8169	80	0.0449
0.1453	21	0.0118	0.4277	51	0.0287	0.818	81	0.0455
0.149	22	0.0124	0.4289	52	0.0292	0.8229	82	0.0461
0.1597	23	0.0129	0.4316	53	0.0298	0.8511	83	0.0466
0.1662	24	0.0135	0.4657	54	0.0303	0.8623	84	0.0472
0.1670	25	0.0140	0.4795	55	0.0309	0.863	85	0.0478
0.1717	26	0.0146	0.4837	56	0.0315	0.9189	86	0.0483
0.1782	27	0.0152	0.4876	57	0.0320	0.9531	87	0.0489
0.1797	28	0.0157	0.4986	58	0.0326	0.9704	88	0.0494
0.2059	29	0.0163	0.5032	59	0.0331	0.9776	89	0.0500
0.2069	30	0.0169	0.5469	60	0.0337			

* Estimated allele frequencies were compared between case pools and control pools for 62 SNPs in the *MYP2* locus (**Table 4.2**), and 9 SNPs were followed up by individual genotyping and each analysed under three genetic models (27 comparisons, **Table 4.4**). Thus, there were 89 comparisons for correction. The list of 89 observed *P* values is shown in three columns above from left to right, and sorted from the smallest (P_1) to the largest (P_{89}). The list of FDR threshold *P* values (0.05 × j/n, where j is the rank; and n is equal to 89 comparisons) is also arranged from the smallest to the largest (0.05). Starting from the largest observed *P* value (P_n or 0.9776 in the table above), compare P_n with 0.05 × j/n. Continue as long as $P_j > 0.05 \times j/n$. Let k be the *first* time when $P_k \leq 0.05 \times k/n$, and declare the comparisons corresponding to the smallest k observed *P* values as statistically significant. In the table above, k = 3 and the FDR threshold *P* value is 0.0017. Therefore, the smallest 3 observed *P* values (0.0015 to 0.0005 shown in **boldface** in the table) are significant.

4.3.5 LD patterns of 9 initial positive SNP markers of *MYP2*

The nine SNPs that showed initial association evidence with high myopia in pooling screening (rs546696 and rs1004961 of *CLUL1*, rs637647 of *EMILIN2*, rs589318 of *LPIN2*, rs9948582, rs7235847, rs4340411 and rs12605942 of *MYOM1*, rs1791067 in *MYCL12B*) show weaker evidence of LD with each other (**Figure 4.2**). However, information from the other functional SNPs may not have been captured as efficiently by the SNPs selected in the present study. Therefore, the future study should focus on the markers in LD with the positive markers of *LIPIN2* gene.

4.3.6 Comparison of allele frequency differences estimated from pools to the frequency differences determined from individual genotypes

It is interesting to note that the allele frequencies estimated by DNA pooling method were in general higher than the corresponding ones estimated by individual genotyping except for rs7235847 of *MYOM1*. For the 9 SNPs, we compared the estimated allele frequency differences based on case and control pools with the frequency differences estimated from genotyping individuals comprising the pools (**Table 4.36**). The mean absolute error in estimating the allele frequency difference between pools calculated from 9 SNP-pool comparisons was 0.0354, and the maximum and minimum absolute errors were 0.0536 and 0.0127.

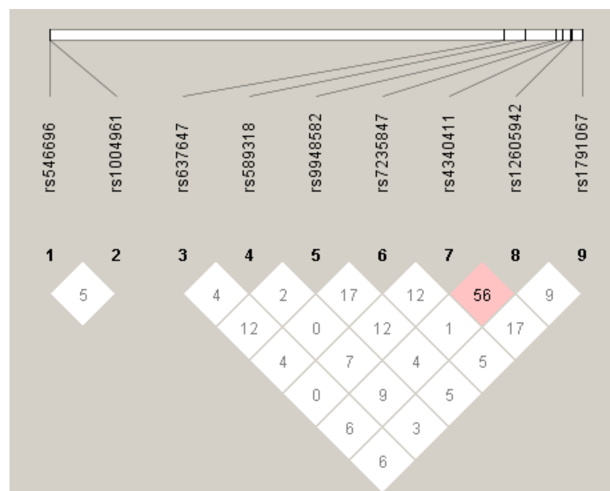


Figure 4.2
Linkage disequilibrium (LD) patterns of 9 initial positive SNP markers of *MYP2*.

The LD measures (r^2) are calculated and plotted by Haploview for the 9 initial positive SNPs genotyped individually. The r^2 values shown above are the actual values multiplied by 100 (e.g., 56 means 0.56, and 5 means 0.05). The shades of grey indicate the magnitude of the measures.

Table 4.6 Comparison of allele frequency differences estimated from pools with the frequency differences determined from individual genotypes*

Gene	SNP	k correction factor	**Estimated allele frequencies:** Absolute Difference	**Actual allele frequencies:** Absolute Difference
CLUL1	rs546696	1.11	0.0204	0.005
CLUL1	rs1004961	1.10	0.0127	0.019
EMILIN2	rs637647	1.22	0.0536	0.034
LPIN2	rs589318	1.33	0.0363	0.098
MYOM1	rs9948582	1.08	0.0514	0.006
MYOM1	rs7235847	0.97	0.0262	0.040
MYOM1	rs4340411	1.08	0.0323	0.011
MYOM1	rs12605942	1.05	0.0433	0.010
MYCL12A	rs1791067	1.20	0.0427	0.022

* The correlation between estimates of allele frequencies from pooled DNA and individual genotyping is 0.084.

4.4 Discussion

In spite of consistent replication for the *MYP2* locus in different linkage studies (Heath et al., 2001; Lam et al., 2003b; Naiglin et al., 1999; Young et al., 2001; Young et al., 1998b), specific candidate genes or disease-causing alleles have not been identified at this locus. Such consistency might emerge as these genes are subjected to more detailed and systematic analysis in sufficiently large samples by the genetic association approach (Pertile et al., 2008; Sasaki, et al., 2007; Scavello et al., 2005; Zhou & Young, 2005). To identify the genetic influences of complex diseases, much hope is now hinged on the results of genome-wide association studies (GWAS) (Amos, 2007). Such studies analyse the genome with hundreds of thousand markers and >1000 pairs of case and control subjects to obtain a reasonable power after correction for multiple testing (Ohashi & Tokunaga, 2001; Sale, et al., 2009). However, the studies are expensive and require such a large sample size that makes it unfeasible for this study (Li, et al., 2008a). DNA pooling studies on defined candidate loci can dramatically cut the cost and provide an adequate study power to investigate the genetic association of high myopia with the *MYP2* locus in the Hong Kong Han Chinese population (Elston, et al., 2007).

4.4.1 Myopia-2 locus (*MYP2*)

In 1998, Young et al (1988b) performed a linkage study of eight families with high myopia in two or three successive generations, containing a total of 82 individuals to obtain a maximum LOD score of 9.59 for the MS marker D18S481. This region, which mapped at 7.6 cM on the short arm of chromosome 18, was indicated to be a

susceptibility genetic locus for high myopia. In this report, we present the results of a pooling technique as a first screen in order to identify markers over large stretches of genomic DNA for a subsequent targeted genotyping on an individual level. From the data obtained for the entire 2.2-cM interval region, we selected a panel of 7 positional candidate genes in the *MYP2* locus, and screened 62 SNPs tagging these candidate genes. Nine SNPs exceeded a lenient significance threshold (P < 0.1). Upon follow-up by individual genotyping, one SNP remained significantly associated with high myopia even after correcting for multiple tests: rs589318 (*LPIN2*). This SNP has not been previously investigated for association with high myopia. No association was found with any of the other candidate genes *(CLUL1, EMILIN2, MYOM1, MYCL12B, MYCL12A, and ZFP161*) and high myopia.

4.4.2 *LPIN2*

We screened 8 SNPs within the *LIPIN2* gene for 300 cases and 300 controls in the initial DNA pooling tests (**Table 4.2**). Five SNPs localized in introns and the other three in the 3' UTR of the *LIPIN2* gene. One of these 8 SNPs was followed up by individual genotyping. We first reported a common polymorphism (rs589318) in *LIPIN2* associated with high myopia for the *MYP2* locus investigated by the pooled DNA approach. Allele and genotype frequencies of rs589318 in controls and cases are listed in **Table 4.4**. Statistically significant result (P < 0.05) was observed in both allele and genotype frequencies between the controls and the cases (**Table 4.4**). This SNP is located in intron 1 in the potential binding elements for homeodomain transcription factors, where it is unlikely to regulate *LPIN2* expression and stability. This significant SNP tags 11 other SNPs documented in the HapMap database, but

none leads to an amino acid change (**Table 4.7**). It is probable that there are many unidentified SNPs in this LD cluster, which may include the SNP responsible for this association. The C allele of rs589318 was protective in nature (**Table 4.4**) and is less common in Chinese than in Caucasians (**Table 4.7**). This incidental observation is intriguing in that myopia is much more common in Chinese than in Caucasians in general. This inverse relationship of allele frequencies in Chinese and Caucasians is also obvious with other SNPs tagged by rs589318 (**Table 4.7**).

Expression studies of this gene by reverse transcription-polymerase chain reaction (RT-PCR) revealed that it was expressed in various tissues including brain, kidney, lung, heart and skeletal muscles (Donkor, et al., 2007). The gene was also expressed in various eye tissues including cornea, retina, optic nerve, and sclera (Zhou & Young, 2005).

High myopia with the clinical state of ocular axial elongation and pathological changes in the retina and the sclera might be due to visual deprivation from blurred image although the exact mechanism underlying the elongation has yet to be defined (Morgan, 2003; Zejmo, et al., 2009). It is presumed that the retina transfers the blurred image signal to the sclera, thus leading to scleral remodeling. Therefore, genetic polymorphisms in the *LPIN2* gene may potentially cause individual differential expression of the gene in the retinal and scleral components for different severity of scleral extension and remodeling.

Table 4.7 SNPs tagged by rs589318 in Han Chinese (HapMap data)*

		Position on	Location	Allele frequencies (HapMap)			r^2 with
SNP	Allele	Chr 18	within *LPIN2*	Allele	CEU	CHB	rs589318 (CHB)
rs681670	A/G	2941876	Intron 1	G	0.312	0.625	0.952
rs643015	T/C	2942988	Intron 1	T	0.692	0.356	0.952
rs603080	C/T	2943798	Intron 1	C	0.289	0.644	0.952
rs584853	T/G	2946812	Intron 1	G	0.704	0.371	0.940
rs8096026	A/G	2947166	Intron 1	G	0.673	0.389	0.888
rs751375	A/G	2949758	Intron 1	G	0.692	0.356	0.952
rs652587	T/A	2951033	Intron 1	A	0.692	0.356	0.952
rs660716	G/A	2956840	Intron 1	G	0.321	0.631	0.903
rs625527	A/C	2958752	Intron 1	C	0.321	0.625	0.952
rs635836	T/C	2959355	Intron 1	C	0.321	0.631	0.903
rs638331	A/G	2959374	Intron 1	G	0.325	0.656	0.903
rs589318	**T/C**	**2973942**	**Intron 1**	**T**	**0.321**	**0.631****	**1.000**

* When tagging SNPs are selected from HapMap Han Chinese (CHB) database by the Tagger software with the criteria of $r^2 \geq 0.8$ and minor allele frequency ≥0.10, the SNPs in the table above form one group and are tagged by rs589318 (shown in **boldface**) with the corresponding r^2 values in the rightmost column. Also shown are the allele frequencies for American residents with northern and western European ancestry (CEU). Note that the SNPs are listed according to their position on chromosome 18.

** For rs589318, the frequency of the T allele is 0.631 for Han Chinese from HapMap database, but is 0.555 in the Han Chinese control group of the present study (**Table 4.4**).

The *LPIN2* gene contains 20 exons spanning 95 kb (**Figure 1.14**), and encodes a nuclear protein (known as lipin 2) essential to normal adipose tissue development and the processing of fats (lipid metabolism) (Reue, 2009). This gene represents a candidate gene for human lipodystrophy characterised by loss of body fat, fatty liver, hypertriglyceridemia, and insulin resistance (Reue, 2009). The protein lipin 2 acts as a magnesium-dependent phosphatidate phosphatase enzyme which catalyzes the conversion of phosphatidic acid to diacylglycerol during triglyceride, phosphatidylcholine and phosphatidylethanolamine biosynthesis in the reticulum endoplasmic membrane (Carman & Han, 2006). It also acts as a nuclear transcriptional coactivator for PPARGC1A to modulate lipid metabolism (Finck, et al., 2006). Mouse studies suggest that the gene may also involve in controlling inflammation and in cell division (Ferguson, et al., 2005).

The role of *LPIN2* mutations in myopia is uncertain. One rare *LPIN2* mutation has been described in three Middle Eastern families (Majeed syndrome) with bone disease, anemia, and inflammation of the skin (Al-Mosawi, et al., 2007; Ferguson et al., 2005). The mutations in the *LPIN2* gene alter the structure and function of lipin-2 and lead to chronic disease, and cause abnormal inflammation in some tissues.

It has been hypothesised that the development and progression of myopia may be attributed to carbohydrate diets with high glycaemic load, possibly as a result of altering the sensitivity to insulin or increasing levels of circulating insulin-like growth factor 1 (IGF1) (Cordain et al., 2002). Recent findings suggested that IGF1 was required for normal insulin sensitivity and impaired synthesis of IGF1 would worsen the state of insulin resistance (Clemmons, 2004). A Danish study found that

myopia was more prevalent in diabetic patients than in the general population, and suggested that poor metabolic control of glucose was a potential risk factor for myopia (Jacobsen, et al., 2008). *IGF1* polymorphisms were indeed recently shown to be associated with high myopia in an international Caucasian family cohort (Metlapally, et al., 2010) and in our own case-control study (Mak JY, 2010). Interestingly, the *IGF1* gene is located within the *MYP3* locus at chromosome 12q22-q23. On the other hand, it was recently reported that a SNP of the *LPIN2* gene was associated with type 2 diabetes, glucose metabolism (as measured by composite insulin sensitivity index) and body composition (Aulchenko, et al., 2007). This circumstantial link is intriguing, particularly in light of the positive association between a *LPIN2* polymorphism and high myopia in the present study. More future work along this direction is warranted.

4.4.2.1 Previous work studying *LPIN2* and myopia

Mutation screens of the *LPIN2* gene by direct sequencing were performed in two studies (Scavello et al., 2005; Zhou & Young, 2005), but no particular sequence alterations were identified in high myopia. Coding regions, intron-exon boundaries and untranslated exons of the gene were sequenced using samples from 10 patients with high myopia cases (mean spherical refractive error of myopia was -9.48 D) and 6 healthy control subjects in Scavell's study and 8 patients with high myopia cases (mean spherical refractive error of myopia was −10.50 D) and 6 healthy control subjects in Zhou & Young's study. The discrepancy in results between the previous *LPIN2* study (Scavello et al., 2005; Zhou & Young, 2005) and the present study might be due to the differences in genetic method to elucidate high myopia genes. The present study is the first large-scale association analysis of the *LPIN2* gene in

subjects with high myopia by genotyping *common* SNPs, which are expected to have *small effect size*, if any. In addition, our larger sample size (n=600) offered a stronger study power to detect the *LPIN2* association with high myopia. The past *LPIN2* studies were mutation screening studies that searched for *rare* disease-causing mutations of *strong effect size*. The previous studies did not account for the aging effect on myopia phenotype, as the older subjects over the age of 50 years might have myopic shifts associated with cataract development (Pararajasegaram, 1999).

4.4.3 *MYP2* candidate genes examined in this study

A genomic interval of 2.2 centiMorgans (cM) was defined on chromosome band 18p11.31 using 7 families diagnosed with autosomal dominant high myopia and was designated the *MYP2* locus (Young et al., 2001). To characterise this region, we analysed 7 positional candidate genes localized within the 2.2 cM interval by an initial screen based on the estimation of relative allele frequencies in DNA pools. Apart from *LIPIN2*, there was no convincing evidence to prove a connection between nucleotide sequence variations in *CLUL1*, *EMILIN2*, *MYOM1*, *MYCL12B*, *MYCL12A*, and *ZFP161* with high myopia.

4.4.3.1 Clusterin-like 1(*CLUL1*)

Clusterin-like 1 is a secreted clusterin family glycoprotein expressed predominantly in the retina, specifically in cone photoreceptor cells, and shares 25% identity to clusterin at the protein level (Zhang et al., 2003). Its major function is to protect stressed proteins (Sturgill et al., 2006). It is upregulated in many different forms of neurodegeneration and is thought to represent a defense response against neuronal

damage (Park, et al., 2007). It has been a candidate gene for retinopathies and age-related macular degeneration (Sturgill et al., 2006). As retinal pathological change is very common in high myopia, we therefore investigated *CLUL1* as a candidate gene for high myopia. A mutation screen of the coding region of the *CLUL1* gene in 10 high myopia and 6 healthy control subjects was performed in Scavell's study, but no sequence alterations were found to segregate with myopia (Scavello et al., 2005). The present study is the first association analysis of the *CLUL1* gene for high myopia by screening common variants. We selected a total of 10 tag SNPs of the *CLUL1* gene, ensuring a wide distribution over the region ranging from exons 1 to 9 and occupying 35 kb within the genome. Given the low LD in the *CLUL1* gene, the results of 10 SNPs in this study did not necessarily reflect the relations between high myopia and all polymorphisms in the *CLUL1* gene. Thus, more detailed SNP analysis may be necessary for the complete screen of the entire *CLUL1* gene.

4.4.3.2 Extracellular matrix associated function, or structural genes such as Elastin microfibril interfacer 2 (*EMILIN2)* and Myomesin-1 (*MYOM1*)

Structural candidate genes that map to the *MYP2* region were given high priority in our initial screen. *EMILIN2* and *MYOM1*genes were found to be important for constituent organization and maintenance of connective tissue function in the eye tissue (Bressan, et al., 1993). Emilin and myomesin are strong candidates of ECM proteins for the maintenance of scleral elasticity, strength, and thickness (Rada, et al., 2006). The knockdown of *EMILIN2* increases transformed cell survival, and over-expression impairs clonogenicity in soft agar and three-dimensional growth in

natural matrices due to massive apoptosis (Mongiat, et al., 2007). *MYOM1* is a giant protein containing structural modules with strong homology to either fibronectin type III (motif I) or immunoglobulin C2 (motif II) domains (Reddy, et al., 2008). Both genes are expressed in the sclera, cornea, optic nerve and retina of human eye based on RT-PCR results (Young, 2004). Thus, we speculate that aberrations of *EMLIN2 and MYOM1* genes in cases of high myopia might be implicated in scleral thinning during myopia development.

A total of 9 *EMLIN2* and 22 *MYOM1* tag SNPs were screened by pooling method in this study. One candidate SNP (rs637647) from *EMLIN2* and four candidate SNPs (rs9948582, rs7235847, rs4340411 and rs12605942) from *MYOM1* were identified from the pooling screening for follow-up study, using pre-specified criteria ($P \leq 0.10$). Evaluation of those SNPs by individual genotyping to reveal additional evidence of association with high myopia did not produce fruitful results. The two genes might not play a critical role in disease susceptibility to high myopia.

4.4.3.3 Myosin regulatory light chains (*MYCL12A* & *MYCL12B*)

Based on the study of gene expression in human donor sclera with cDNA library (Young, et al., 2003), the myosin regulatory light chain genes mapped to *MYP2* locus were selected as positional candidate genes for high myopia in our *MYP2* study. A total of 6 *MYCL12A* and 4 *MYCL12B* tag SNPs were screened by the pooling method in this study. Only one candidate SNP (rs1791067) for *MYCL12A* was identified from the pooling screening based on a pre-specified threshold ($P \leq 0.10$). The SNP was genotyped in the individual sample set using a RFLP assay to reveal additional evidence of association with high myopia, but no significant association

was found between the SNP and high myopia. The results indicate that the myosin regulatory light chain genes may not play a role in high myopia.

4.4.3.4 Zinc finger protein 161 (*ZFP161*)

Transcription factors and regulatory genes expressed in the retina may play a role in regulating eye growth (Schippert, et al., 2008). Dopamine was shown to be important in the control pathways for eye ball elongation during myopia development in a mouse model (Kroger, et al., 1999). *ZFP161* is a transcriptional activator of the dopamine transporter (Lee, et al., 2004). Therefore, *ZFP161* was identified as a candidate gene for high myopia because of its role in ocular growth in animal studies.

A recent family-based study of probands from 211 families with high myopia (<=-6.0D) and 116 unrelated emmetropic controls analysed the sequence variations of *ZFP161* in relation to high myopia in Chinese (Wang et al., 2004). The coding regions of *ZFP161* were analysed by using PCR-based heteroduplex-SSCP analysis and sequencing, where a synonymous mutation (Codon56, GCC-->GCA, Ala56Ala) was also present in 5 patients and 3 controls. No other coding variation was found for high myopia.

In this study, 3 tag SNPs were screened by the pooling method to investigate the association of *ZFP161* with high myopia. None was identified on the basis of a pre-specified cut-off ($P \leq 0.10$). The lack of significant association with tag SNPs suggests that *ZFP161* was an unlikely candidate gene for myopia.

4.4.4 Other candidate genes in the *MYP2* locus

Other genes located in or near the *MYP2* region with implication in eye growth may provide alternative candidate genes for further exploration in the analysis of high myopia or related biometric phenotypes.

4.4.4.1 Transforming growth beta-induced factor (*TGIF*)

Transforming growth beta-induced factor (*TGIF*) was initially identified as a candidate gene for high myopia because of its location within the *MYP2* region. *TGIF* is expressed in the sclera, retina, cornea, and optic nerve (Young, 2004). Animal studies using form-deprivation myopia also demonstrated that transforming growth factor (TGF)-beta, which is induced by *TGIF*, mediates the retinal control of ocular growth (Jobling, et al., 2004). Genetic evidence supporting a role for *TGIF* in myopia came from the analysis of a Hong Kong Chinese cohort (71 subjects with high myopia versus 105 control subjects) where SNP rs2229336 (c.657T/G, synonymous) was significantly associated with high myopia (<=-6.0D) (Lam et al., 2003a). However, association of this locus was not supported by subsequent studies. This significant association could not be replicated in a second Chinese case/control study of high myopia individuals (Li et al., 2003; Wang et al., 2009b). A Japanese case/control study of high myopia (<=-9.25D) also failed to replicate this gene by using 13 SNPs across the *TGIF* gene (Hasumi et al., 2006). In the only Caucasian study to date, coding regions and intron–exon boundaries of *TGIF* were sequenced in 10 cases (<=-6.0D) from European high-myopia families and 10 unrelated emmetropic controls, but no disease-causing sequence variants were detected in the high-myopia individuals when compared with control subjects (Scavello et al., 2004). Currently, no replication association of *TGIF* with myopia has so far been conclusive,

a further study on *TGIF* was suggested with appropriate criteria and careful selection of cases to minimize false positive and enhance the possibilities of identifying genes predisposing to myopia (Colhoun et al., 2003; Dahlman, et al., 2002; Newton-Cheh & Hirschhorn, 2005).

4.4.4.2 Alpha subunit of laminin (*LAMA1*)

LAMA1 was analysed as a candidate gene for myopia because of its location near the *MYP2* region and its role in controlling sclera collagen fibrillogenesis (Dietlein, et al., 1997). This gene was excluded as a candidate gene in a study of 330 unrelated Japanese patients with high myopia (<-9.25 D) and 330 ethnically and sex matched controls (Sasaki et al., 2007). Two of the SNPs were monomorphic and none of the 11 SNPs showed statistically significant association with high myopia in the Japanese population.

4.4.5 DNA pooling strategy

Our study raises a number of issues for pooled DNA association studies in general and the genetic basis of high myopia in particular. Evaluating the susceptibility genes of complex disease requires genotyping of large numbers of SNPs. The major prohibition to this approach is the huge cost of genotyping and the considerable expense involved. Recent progress in high throughout genotyping techniques with decreasing genotyping cost has alleviated this problem to some extent, but the creation of polymorphism maps at specific loci to study specific diseases may still be a challenge. DNA pooling allows substantial savings in genotyping costs (Mohlke, et al., 2002; Sham, et al., 2002) and has been reliably applied in the analysis of

complex disorders (Hinds, et al., 2004; Sawcer, et al., 2002; Shifman et al., 2002). It reduces the number of required genotyping reactions by a factor of 100 - 1000 (Shaw, et al., 1998). The statistical power was shown to be high and the false-positive rate was well controlled (Kirov, et al., 2000).

For our small sample size (n=600), it is too risky to undertake a genome-wide association study (GWAS) because of the necessity of correction for multiple testing and the reduction of study power by DNA pooling. Instead of GWAS, utilizing the DNA pooling method to screen for candidate genes within a specific myopia locus is a more feasible approach.

4.4.5.1 Genotyping platforms for allele frequency estimation in pooling project

Several platforms are generally considered suitable for DNA pooling strategies, including SNaPshot (Le Hellard et al., 2002), SNaPIT (Curran, et al., 2002), pyrosequencing (Gruber, et al., 2002), TaqMan and DHPLC (Le Hellard et al., 2002; Norton, et al., 2002).

Recently, an outstanding technique combined the Affymetrix platform and DNA pooling study for GWAS had been applied to search high-risk genes of complex diseases (Macgregor, et al., 2006). In the present study, if we apply a genetic fine mapping design to candidate genome regions potentially associated with diseases, it seems too wasteful to use the Affymetrix platform. Therefore, to search high-risk genetic variants of diseases in specific genome regions by using DNA pooling design, other platforms could be more adaptable to this study. The primer extension reaction

coupled with DHPLC analysis has the properties of easy-to-use and low cost in DNA pooling study (Norton, et al., 2004). The exact costs of various genotyping methods vary among countries and even among laboratories depending upon purchase agreements. In general, the primer extension reaction coupled with DHPLC is an inexpensive and accurate method for the estimation of allele frequencies in pooled DNA samples.

4.4.6 The validity of the primer extension reaction coupled with DHPLC analysis in DNA pooling study

The foundation of a successful DNA-pooling association test is a precise and accurate estimation of allele frequencies. In comparison to other SNP genotyping methods for screening DNA pools for association, the primer extension reaction coupled with DHPLC analysis is reasonably precise when compared to other pooling DNA methods. The average allele frequency SD of 0.005–0.029 we reported is similar to the 0.014 reported for other primer extension-denaturing high-performance liquid chromatography (Giordano, et al., 2001), the 0.009–0.017 reported for fluorescent nucleotide primer extension-capillary electrophoresis (Mátyás et al., 2002; Norton, et al., 2002), and the 0.011 reported for pyrosequencing (Wasson, et al., 2002), and less than the 0.038 reported for bioluminometric-primer extension (Zhou, et al., 2001). The validity of the current pooling method was shown by the correlation of 0.084 between estimates of allele frequency from pooled DNA and individual genotyping (**Table 4.6**). With this level of validity, this pooling approach DNA estimates are not compared well with individual data, but the intent is to use

this method not as a definitive analysis, but rather as a rapid initial screen to reduce the number of candidate SNPs to be submitted to individual genotyping for the large samples needed to detect SNP of small effect size (Sham et al., 2002). With 62 chosen SNPs for the *MYP2* locus, a large number of false positive results are expected by chance and so a multi-stage replication design was adopted in order to balance false positives and false negatives in the selection of SNPs for individual genotyping.

4.4.7 Potential sources of bias or imprecision in pooled DNA methods

A number of experimental complications may arise from the allele frequencies estimated from a pool of DNA. Biased or unreliable estimation of allele frequencies can lead to spurious results in association studies. Variation in the data from a DNA-pooling study may arise from several different sources, such as pool formation, polymerase chain reaction (PCR) amplification, allele frequency measurement and other uncontrollable experimental errors (Barratt, et al., 2002; Visscher & Le Hellard, 2003).

4.4.7.1 The k correction factor

Inaccurate estimates of allele frequencies in DNA pools are introduced by SNPs showing unequal degrees of amplification during the process of primer extension, since one allele is more efficiently amplified than the other (Le Hellard et al., 2002). It arises both from heterogeneous nucleotide incorporation during primer extension and differential efficiency of nucleotide detection during DNA quantification (Sham

et al., 2002). In DNA pooling study, the k correction method is often carried out to adjust the unequal amplification of heterozygous alleles. But the practice of k adjustment is under the assumption that the degree of unequal amplification holds constant across different allele frequencies. Not all SNPs exhibit tight correlations between actual allelic frequency and the amplified results. Thus, using k correction may result in biased estimates of allele frequencies in DNA pooling study with certain polymorphic types of SNPs, such as G/C and G/T polymorphisms. It has been reported that significantly inaccurate estimates were obtained by using k correction at least for certain types of SNPs when actual minor allele frequencies were in the range of 20% to 40% (Gruber et al., 2002). Moreover, it had been reported that the variations in k factor is large enough to result in unacceptable error rates if association studies are conducted without regard to the variation of the k factor (Moskvina, et al., 2005).

Use of standard curves to estimate allele frequencies has been proven to be an alternative to the k correction method (Chen, et al., 2008). The principle of polynomial standard curve is conceptually simple and does not required complex statistics. To obtain accurate estimates of allele frequencies and unbiased results of genetic association studies, combination of k correction with second-degree polynomial standard curves had been applied to DNA pooling study with the Affymetrix platform (Brohede, et al., 2005). The difficulty in this method is accurately mixing DNA samples to make standard DNA solution with the heterozygous and homozygous samples. Moreover, many more individual samples (n = 16-20) are required for genotyping in pilot studies to reveal marker's allelic type for their future use in constructing standard curve compared to the k correction

method (n = 5-10).

The range of values for the k correction factor in current pooling study was from 0.30 to 4.37 with a mean of 1.21, which is not too large to result in unacceptable error rates (Moskvina et al., 2005). We therefore modified the k correction method in this study by measuring the allele peak height ratio from three repeat pooled DNA samples. Using the three peak height proportions, three estimates of allele frequency were obtained from the k correction method, could lead to more accurate and precise estimates of the allele frequencies.

4.4.7.2. Pool-specific measurement error

The estimation of the sample allele frequency can be imprecise due to unequal amounts of DNA per individual in the pool and due to experimental errors. To address the challenges of allele frequency estimation, our sample of 300 cases and 300 controls was split into 6 case pools and 6 control pools each containing equal amounts of DNA from 50 distinct individuals. The pool size has been suggested to influence the accuracy of the estimates. Smaller pool size with 50 individuals may control the error to a negligible level (Barratt et al., 2002). In addition, repeat measurement of the pools was suggested to control for measurement error.

In practice, this has been prevented by financial limitations. A balance needs to be struck between the number of SNPs to be tested and the number of replicate pools per SNP. We used three replicates per pool and this appeared to be adequate.

4.4.7.3. Standard sampling error

Sampling a finite number of individuals from a population can inflate the sample error in the pooling analysis. One weakness of the current study is that only 300 cases with extreme myopic phenotypes and 300 controls are available for screening 62 SNP markers within the *MYP2* locus. Such a relatively small sample size could lead to a false positive finding (type 1 error) and loss in power when compared with individual genotyping (Jawaid & Sham, 2009). However, the possibility of type 1 error is minimized because our pooling analysis was confirmed by the second step of individual typing. In addition, the association findings were based on an analysis of a locus targeted by our hypothesis, a promising *MYP2* region previously linked to high myopia by several population groups.

4.4.7.4. Haplotype frequency estimates

Haplotype and single-marker approaches are equally important in the association tests for disease gene mapping. However, the pooling approach in current study did lose information in LD (Pirinen, et al., 2008). A modified version of PHASE for estimating population haplotype frequencies has been demonstrated for pooled DNA data. The algorithm is compared with (i) a maximum likelihood estimation under the multinomial model and (ii) a deterministic greedy algorithm, on both simulated and real data sets (HapMap data). The software takes into account correlated genealogical histories of the haplotypes by modeling mutations and recombination. Although the accuracy of the haplotype frequency estimates decreases as the level of genotyping error increases, this decrease is small and, even in the presence of genotyping error, the estimates of the haplotype frequencies are accurate (Quade, et al., 2005). Ideally, it would be most beneficial to design studies with a large number

of individuals per pool to minimize the genotyping costs. Under all genotyping error levels, 10 individuals per pool were suggested to obtain accurate estimates (Quade et al., 2005). Despite the proven utility of pooling, it does provide less information, hence less power, than individual genotyping (Zou & Zhao, 2003).

Although errors are present when estimating the allele frequencies from pooled samples, pooling provided reasonably accurate estimates. In a comprehensive review of DNA pooling, Sham (2002) concludes that pooling can be considered both cost- and time-effective.

4.5 Conclusion

In conclusion, our findings are important for the understanding of the genetic basis of high myopia, and prove that association analysis can be accelerated when pools of DNA are used. This approach successfully identified a novel positive signal in the *LPIN2* gene within the *MYP2* locus region. Replication investigation using an independent sample set is required to confirm the association and, if confirmed, follow-up studies to identify the causal variant and relevant functional studies are required to investigate the mechanism that *LPIN2* is linked to the pathogenesis of high myopia. In addition, the estimation of allele frequencies in DNA pools with subsequent individual genotyping of selected SNPs demonstrated the feasibility of screening other myopia loci undergoing preliminary investigation in this medium-throughput manner.

Chapter 5
Overall discussion

Genetic predisposition plays an important role in the pathogenesis of myopia development. Unraveling the genetics of myopia has so far been proven difficult. Association studies are powerful enough to detect small gene effects in complex diseases such as myopia. The *myocilin* gene was recently identified as a novel myopia susceptibility gene. The present case-control study together with a family dataset served to confirm or dispute this recent finding applied to the Hong Kong Chinese population. Another objective of this study was to apply a two-stage DNA pooling method to investigate any role for positional candidate genes within a high myopia locus located on chromosome 18p11. Discussions and suggestions for the future studies on these genes were separately discussed in the following sections.

We were able to replicate the association of the *MYOC* polymorphisms with high myopia. Confirmation of the *MYOC* pathway for myopia genesis requires identification of the causal variant(s) within this region and subsequent additional functional studies. Suggestions for the future studies will be summarised again in this chapter, but the main focus of future work should be on the 3' flanking region of *MYOC* and its immediate neighboring regions.

Another novel finding is the association of the *LPIN2* gene in the *MYP2* region with high myopia by using the pooling approach and its impact on future research will be also be discussed.

5.1 Mapping susceptibility genes for myopia

The recent completion of the Human Genome Project allows rapid analysis and assessment of our genome to enhance our understanding of the human genome (Little, et al., 2002). Based on the understanding of the molecular biology of complex diseases, genetic diagnosis, preventive medicine and the development of new therapies will promote the management of such diseases (Helgason & Stefansson, 2010). Therefore, the development of molecular medicine or gene therapy will be the ultimate goal to mapping genes for complex diseases (Butcher, et al., 2004) like myopia.

Several approaches are available to map susceptibility genes of complex traits like myopia and these have already been reviewed in Chapter 1. In this discussion part, emphasis will be on the approaches used in this study such as **case-control association test** and **candidate gene approach**. Several SNP genotyping methods were used in the present study. These approaches and methods will be discussed in terms of their advantages and disadvantages. Issues related to collection and analysis of data and improvement of the study will also be discussed.

5.1.1 Phenotype definition and related phenotypes

The phenotypes of refractive errors used in myopia genetic research are often ill-defined in the literature. A poor case definition for high myopia can pose challenges when investigators seek to link that phenotype to other related phenotypes (in order to better define a multivariate phenotype) or to genotypes (in order to define

a genetic association). Poorly defined phenotypes can lead to negative results and failures to replicate findings, as is frequently seen in myopia genetic studies. For instance, the discrepancy between the results of the two recent hepatocyte growth factor (*HGF*) studies may have occurred because the studies used different grades of high myopia as inclusion criteria (Han et al., 2006; Yanovitch et al., 2009). The inclusion criteria adopted by the American group was myopia ≤-5.00 D, which resulted in a smaller extreme high myopia sample size, while Han et al. targeted subjects with myopia ≤-10.00 D. Due to significant differences in phenotype definition, it is inappropriate to perform meta-analyses and difficult to make cross comparisons. Because of the high prevalence of low and medium myopia, the common type of myopia is expected to be heterogeneous in a population (Morgan & Rose, 2005). Poorly defined phenotypes are the most challenging issues for the ocular biometric measurements, including axial length (Nakanishi et al., 2009a; Nakanishi, et al., 2009b; Paget et al., 2008), lens thickness, and corneal power (Young, et al., 2007). The axial length of the eyeball in adults is approximately 24 mm, and its elongation by 1 mm without other compensatory changes results in a myopic shift of −2.5 to −3.0 D (Saw, et al., 2005a). In the study of the pathological myopia at locus 11q24.1, Nakanishi *et al.* (2009b) took axial length (greater than 28.0 mm) into account in the phenotype definition for recruiting myopic subject in order to reduce the complexity of the genetic background contributing to the myopia trait of the subjects. The definition of high myopia have been redefined as axial length being greater than 25.5 mm or 26 mm (Morita, et al., 1995; Pierro, et al., 1992).

The benefit of good phenotype definition can be seen in the recent successes of genome-wide association studies (GWAS) outside the field of myopia. For instance,

in studies of type 1 diabetes, recent advances in technology and knowledge have allowed researchers to determine four new genes from GWAS, by examining a clear (endo)phenotype in immune cell histocompatibility antigen molecules (Barratt, et al., 2004; Bottini, et al., 2006; Lowe, et al., 2007; Qu, et al., 2007; Vella, et al., 2005). Thus even with a polygenic disease, a coherent phenotype for diabetes has allowed the genetic contributions to be fully realized in these association studies.

An over-stringent entry criterion increases the difficulty in subject recruitment. In current study, high myopia was defined as refractive error ≤-8.0D of spherical equivalent for both eyes of case subjects because the ease of subject recruitment could give a larger sample size within a reasonable period. Such refraction criteria may be strict enough to enhance the contribution of the genetic component to myopia trait in the subjects (Rebbeck, 1999). High myopia accounted for nearly 30% of all myopic eyes, corresponding to a prevalence of 4% in the general Chinese population in Hong Kong (Fan, et al., 2004c). Hong Kong is one of the high-risk areas in the world for myopia, and high myopia is much more common in South Asian city (Asano, et al., 2004; Fotouhi, et al., 2007; Xu et al., 2005). The prevalence of myopia was 85% in 13~15 year-old students in Hong Kong in 2004 (Fan et al., 2004c). In contrast, recruitment of families with highly myopic siblings is difficult owing to its lower prevalence and the requirement of involving the parents. In order to give sufficient sample size, we assigned criteria for myopic siblings from nuclear families with less stringent definition (≤-6.0 D). This strategy might be one of the reasons why family dataset gave less impressive *P* values for the two positive SNPs rs6425356 (S7) and rs743994 (S10) when compared to the case-control dataset (**Table 5.1**).

Table 5.1 Comparison of case-control and families studies*

Study	Cutoff refraction for recruitment (D)	Mean ± SD for dataset		*P* values (additive model)	
		SE (D)	AXL (mm)	rs6425356 (S7)	rs743994 (S10)
Case-control	-8.0	-10.53 ± 2.48	27.76 ± 1.13	1.98×10^{-6}	2.81×10^{-6}
Family	-6.0	-9.06 ± 2.01	26.88 ± 1.04	0.0046	0.0049

* D = dioptres; SE = spherical equivalent; AXL = axial length. The data are taken from **Tables 3.1**, **3.4**, **3.10** and **3.11**.

With standardized phenotype definitions for myopia genetic studies in the future, we can construct multivariate phenotypic models, universal criteria about spherical refraction equivalent and correlated ocular biometric data. High myopia is usually defined as ≤ -6.00D for both eyes in most of the myopia genetics studies (Lam et al., 2003a; Lin et al., 2006; Scavello et al., 2005; Wang, et al., 2009a). An axial length of greater than 26.0 mm should also be used as the cut-off threshold for high myopia in future replication study for myopia genes. This will facilitate comparison across different studies.

5.1.2 Disease gene hunting approaches

There were two main approaches for mapping genes for complex traits like myopia: linkage analysis and association studies.

5.1.2.1 Linkage analysis

This approach is used to identify chromosomal regions that co-segregate with the trait within families, and thus utilizes every family member's phenotypic and genotypic information. It investigates the relationship between loci rather than between alleles because the disease alleles and the markers linked in a given family may be different from those in another family. However, parental genotypes of an affected child are sometimes difficult to find, particularly when the disease is late-onset. Genetic linkage studies have been used successfully to map simple monogenic Mendelian diseases. However, this parametric method requires the assumption of a single locus inheritance, specification of disease gene frequency and penetrance, and is poor at providing a precise location of the disease gene (Greenberg, 1990). The

linkage approach is used to map myopia loci with large effects to relatively large chromosomal regions (Tang et al., 2008), but has not yet yielded consistent evidence for mapping complex disease genes, like those for myopia (Foroud, 1997; Hornbeak & Young, 2009).

5.1.2.2 Association studies

Association studies are used to identify loci that are associated with the trait at the population level. The method is widely used for genetic analysis when attention is focused on a relatively small chromosomal segment, and investigates the relationship between alleles rather than between loci (Johnson & Todd, 2000; Lander & Schork, 1994; Rannala, 2001). No clear pattern of Mendelian inheritance is required. Therefore, association studies are usually carried out following the identification of potential candidate genes in that region.

Theoretically, linkage methods are less powerful than association studies in genetically mapping a complex disease (Chen & Deng, 2001). Practical studies also confirmed this (Jones, 1998). For example, evidence from conventional linkage studies for the search of myopia gene loci lagged behind that from association studies (Jacobi & Pusch, 2010). Population-based case-control association studies are a popular alternative to the linkage studies (Tang et al., 2008). The candidate genes of myopia are expected to be expressed in the eye tissues and can be selected based on hypotheses of biological functions related to myopia development. However, population stratification and admixture may lead to spurious associations, and hence biased findings for the gene-disease association (McKeigue, 1997). This can occur when both disease risk and genetic mutation frequencies vary among

different ethnic groups (Cardon & Palmer, 2003; Wacholder, et al., 2002). To avoid the problem of population stratification bias, matching cases to controls on ethnic background is required and family-based association studies can be used (Cheng & Lin, 2007).

Family-based association designs offer a compromise between traditional linkage studies and case-control association studies (Jacobi & Pusch, 2010). Moreover, compared to the linkage analysis, family-based association analysis is more powerful in detecting small to moderate genetic effects (Laird & Lange, 2006). TDT compares the frequency in heterozygous parents of specific alleles that are transmitted to affected offspring with the frequency (in the parents) of the alleles not transmitted (Ewens, et al., 2008). Alleles that are *not* transmitted to affected offspring act as internal controls and avoid the issues of population stratification that may lead to false-positive results in case-control association studies (Nicodemus, et al., 2007). Although it is more difficult to recruit family samples than case-control ones, it is still worth the efforts because of this advantage.

Pros and cons of the current case-control myopia genetics studies

To detect genetic risk factors for high myopia which occurs in only 4% of the general population, case-control association studies may be the only way to dissect the genetic basis of such complex eye disorder. The approach can study multiple aetiological factors simultaneously and less time-consuming in subject recruitment. However, multiple biases may give spurious evidence of association between candidate genes and myopia. Another independent family set for replication is essential to avoid the biases. In the present *MYOC* studies (Jacobi & Pusch, 2010),

the positive association of two SNPs in the 3' untranslated region of the gene was also successfully replicated and confirmed with an independent family data set (**Tables 3.4** and **3.11**), and hence the confounding due to population heterogeneity was very unlikely. The consistently positive results for the *MYOC* gene strongly suggest that this gene does play a significant role in the aetiology of high myopia. For the family-based association designs, TDT utilizes transmission information from heterozygous parents to the affected offspring from nuclear families. In this study, both parents were available for all nuclear families since the onset of myopia is not too late for our study. On the other hand, parents may be missing for late-onset diseases and the statistical power will be reduced. Some studies suggested the use of unaffected siblings from the same families to increase the power of TDT for detecting small genetic effects in complex diseases (Curtis, 1997; Guo, et al., 2007; Moroldo, et al., 1998; Rieger, et al., 2001).

5.1.2.3 Study designs: Candidate-gene vs Genome-wide association study (GWAS)

Successful gene mapping strategies for common disease continue to require careful consideration of basic study design. During the initial planning of the current project, most genetic association studies examined a single polymorphism or a set of polymorphisms near a single gene or focused on a candidate region defined by a linkage peak from family-based studies. With the ever-improving genotyping technology, the decrease in the genotyping cost, genome-wide association studies with hundreds of thousands or even millions of polymorphisms genotyped have become feasible and popular in recent years (Bosse, et al., 2009). The GWAS has been suggested as an unbiased approach to investigating complex diseases since it

does not require prior hypotheses of candidate genes or polymorphisms related to the phenotypes of interest (Kitsios & Zintzaras, 2009). In other words, it is hypothesis-generating rather than hypothesis-testing. Consider a typical association study of 1000 case and 1000 control samples using 10,000 SNPs. Individual genotyping would require a minimal of 2×10^7 typings, which is usually beyond the capability of most research laboratories as an in-house project (Peters, 2009). Moreover, this method often ignores the use of biological information available, such as disease-specific biochemical pathways, known functional properties of SNPs, comparative genomics, prior evidence of genetic linkage, and LD (Moore, et al., 2010). Another issue is the reduction in power of GWAS study because of testing small sample size with a large number of markers compounded by multiple comparisons (Duggal, et al., 2008).

The genome-wide approach interrogates all variations throughout the genome and has two main strategies in use. Family linkage study assumes ~10Mb regions of LD with 400 MS markers while unrelated case-control study assumes ~10kb regions of LD with 300-500K SNPs, Multiple technologies now allow a GWAS design to be implemented with high fidelity and low cost (per genotype), such as Affymetrix SNP Array 6.0 and Illumina HumanHap550 BeadChip. The alleles, genotypes, or haplotypes of these SNPs are tested directly for association with disease. The same genotyping platforms also capture CNV information. Thus, GWAS is by far the most detailed and complete method of whole genome interrogation currently available.

During the past few years, there have been three articles focusing on myopia using the GWAS approach (Hysi, et al., 2010; Nakanishi et al., 2009b; Solouki, et al.,

2010). They are discussed below.

A GWAS for refractive error examined 5,328 individuals from a Dutch population with replication in four independent cohorts (combined 10,280 individuals in the replication stage), and demonstrated the association of three genetic loci associated with optic disc area, and another six with vertical cup-disc ratio (VCDR) (Solouki et al., 2010). The optic disc region is an important measure with myopia. The most interesting candidate genes for optic disc area were *TGFBR3* on chromosome 1p22, *ATOH7* on chromosome 10q21.3-22.1 (also for VCDR) and *SALL1* on chromosome 16q12. Regions of interest for VCDR were *CDKN2B* on chromosome 9p21, *SIX1* on chromosome 14q22-23, *SCYL1* on chromosome 11q13, *CHEK2* on chromosome 22q12.1, *DCLK1* on chromosome 13q13, and *BCAS3* on chromosome 17q23. There are several pathways implicated, but the most interesting is the TGFβ signaling pathway that appears to play a key role.

GWAS has also been used to identify myopia susceptibility genes in Japanese, analysing 411,777 SNPs in 830 cases and 1,911 general population controls with a two-stage design (Nakanishi et al., 2009b). Two genes, *BLID* and LOC399959, were identified within a 200-kb DNA region encompassing rs577948. RT-PCR analysis demonstrated that both genes were expressed in human retinal tissue. Another finding was a novel susceptibility locus for pathological myopia at 11q24.1.

The most recent GWAS from the Twins UK cohort examined 4,270 individuals and identified SNPs on 15q25 to be associated with refractive error (rs8027411, $P = 7.91 \times 10^{-8}$) (Hysi et al., 2010). This association was replicated in six adult cohorts of

European ancestry with a combined 13,414 individuals (combined P = 2.07×10^{-9}). This locus overlaps the transcription initiation site of *RASGRF1*, which is highly expressed in neurons and the retina, and has previously been implicated in retinal function and memory consolidation. Rasgrf1 (-/-) mice show a heavier average crystalline lens ($P = 0.001$).

Our candidate-gene approach, individual or pooled sample testing, was limited to a small subset of the genome and focused on a set of candidate genes potentially associated with myopia development. However, even in the context of a candidate gene study, we might end up testing hundreds of SNPs. Tag SNPs helped to minimize the number of SNPs tested. In fact, a causal variant is unlikely to be typed in the study because it may not be documented in the relevant databases or it may not be a SNP (it might be an insertion or deletion, inversion, or copy-number polymorphism). Nevertheless, our study is well-designed and will have a good chance of including one or more SNPs that are in strong LD with a common potential causal variant, e.g. positive SNP markers found in ***MYOC*** and ***LPIN2*** genes. The present association studies are powerful and robust enough to detect small genetic effects. The phenotype definition of severe myopia ≤ -8.00D is likely to enhance the statistical power. It tends to increase the probability of detecting the gene effects with severe myopia as the cases. However, it was very challenging in practice to recruit a large number of high myopia cases, especially nuclear families, with stringent refraction criteria. Collaborations of multiple myopia research centres should facilitate the collection of a huge number of subjects for association studies and increase the chance of identifying the myopia genes.

The two associations might detect the indirect association between marker locus and disease phenotype. Replication with an independent population sample set and functional analyses are required to evaluate the reported associations between candidate genes and myopia.

5.2 Detection of polymorphisms

Since SNPs are sufficiently abundant in the human genome and are less mutable than MSs, the SNPs are used as useful markers for mapping genes that contribute to myopia in the current project.

5.2.1. SNP genotyping methods

Several genotyping methods were used in the present study, such as allele-specific PCR (comparative thresholds and Tm-Shift methods) with real-time PCR System**,** RFLP and PE-DHPLC platform. Some positive SNPs (rs1602244 and rs6425356) were genotyped twice with two different methods to rule out the possibility of false positive results because of genotyping errors. Genotypes of both case-control samples and the family samples were successfully determined by these methods. The details of their principle are described in Chapter 2. The pros and cons of these genotyping methods are discussed below.

5.2.1.1. RFLP

SNP can be genotyped by restriction enzymes which only cut certain specific DNA sequences of the selected allele. The restriction fragments are then separated

according to length by gel electrophoresis.

This method is reliable and cost-effective, but rather time-consuming and laborious. The obvious drawback of this method is that some SNPs do not carry restriction enzyme cutting sites and thus cannot be genotyped by RFLP. However, this problem can be circumvented by introducing a base change into a PCR primer to create a new restriction enzyme recognition site for a given allele of a SNP (Kimura, et al., 2000). In this study, SNPs rs637647, rs4340411, rs12605942, rs1791067, rs589318, rs9948582 and rs7235847 of the candidate genes in the *MYP2* locus were genotyped by this build-in mismatch primer method (**Table 2.6**). One drawback of the mismatch primer method is that the size difference between cut and uncut fragments might be too small (about the size of a PCR primer) to be determined by agarose gel electrophoresis (Haliassos, et al., 1989). To deal with this possible problem, polyacrylamide gel was used for separating the small fragments.

An internal control site was *always* introduced into the PCR fragment to guard against incorrect genotype calls due to faulty restriction enzymes or incomplete digestion. The PCR product contains a restriction recognition site common to both alleles of a SNP, and cleavage at this site serves as an internal positive control for the restriction digestion. After electrophoresis is complete, the PCR restriction fragments in the gel can be stained to make them visible. Ethidium bromide and SYBR Green I, which were used for staining gels, are toxic and carcinogenic. There is a risk of exposure to and contamination by this chemical during the genotyping process. Allele-specific PCR assay is an alternative method to overcome this limitation.

5.2.1.2 Allele-specific (AS) PCR

Allele-specific PCR, also known as amplification refractory mutation system (ARMS), is a well-established method for discriminating between different alleles at specific loci resulting from single base mutations (Newton et al., 1989). We used this principle to establish two assays: comparative thresholds (d*C*t) and melting temperature (T_m)-shift SNP methods (Germer & Higuchi, 1999). For both methods, the most important step was the design of allele-specific PCR primers that were able to specifically discriminate the two alleles of a SNP. The specificity of PCR was controlled by both the annealing temperature and the inability of *Taq* DNA polymerase to extend a primer mismatched at the 3' end with the template.

For the d*C*t method, locked nucleic acid (LNA) was used instead of the conventional nucleotide at the 3' end of the allele-specific primer to enhance the affinity and specificity of the allele-specific PCR primers (Reynisson et al., 2006). For the T_m-shift method, in addition to an allele-specific LNA at the 3' end, GC-rich tails of different lengths were added to the 5' end of allele-specific PCR primers such that the two SNP alleles of DNA samples can be discriminated by the T_m's of the PCR products (Wang et al., 2005). For the d*C*t method, two separate allele-specific PCRs were performed for each sample, one for each allele, and hence consumed twice the amount of DNA and reagents when compared with all other methods. For the T_m-shift method, all reactions and measurements were performed in a single closed tube.

In the d*C*t method, each PCR mixture contains a common primer and the primer specific for a particular SNP allele. The modified allele-specific primers used in this study were unlabelled since the real-time fluorescence signal was generated by a

non-specific DNA binding dye, SYBR Green I. Ideally, only completely matched primers are extended and only the matching allele is amplified. In practice, however, there will be amplification of the mismatched allele, which will occur much less efficiently such that many more amplification cycles are needed to generate detectable levels of the product. “Mismatched” amplification is frequently delayed by 6 - 10 cycles when amplification is monitored on a cycle-by-cycle basis (Higuchi, et al., 1993). A delay of around six cycles is usually adequate for the discrimination of the two alleles of a given SNP. In general, the greater the delay for the *C*t of the mismatched allele, the more reliable and reproducible is the method.

Allele-specific PCR is relatively inexpensive. By using 96- or 384-well real-time PCR thermal cycles, it could be further scaled up to genotyping many samples in parallel. It is fast and does not require post-PCR analysis (e.g. gel electrophoresis). However, this assay does require expertise in primer design and extensive PCR optimization. The method of primer extension coupled with DHPLC (PE-DHPLC) is an alternative method to minimize difficulties in optimization.

5.2.1.3 Primer extension DHPLC platform

In this method, a primer is annealed to a PCR product immediately adjacent to the SNP position. One or a few bases are added to the extension primer, depending on the sequence at the SNP site in the presence of appropriate dNTPs and unlabelled ddNTPs. In the present study, DHPLC was used for the extension product analysis and genotype determination. The single-stranded primer extension products were analysed by DHPLC under completely denaturing conditions (DNASep column kept at 70°C). The method has been demonstrated to be superior for SNP detection in

terms of sensitivity, efficiency and economy since it is not based on expensive fluorescently labelled primers or probes, and operates under uniform conditions without the need for marker-specific assay optimization. The main disadvantage is the requirement of post-PCR processing: treatment with Exo I and SAP, primer extension reactions and fragment analysis by DHPLC. The cost of SNP genotyping per sample using primer extension is relatively higher than those based on AS-PCR and RFLP methods. Although analysis of primer extension products by DHPLC is automatic and requires minimal optimization, it requires sequential sample injection and longer running time (8-12 mins per sample). This becomes a rate-limiting step when large numbers of individual samples are genotyped.

5.2.1.4 Direct cycle-sequencing

Besides the above methods, direct cycle sequencing is a reliable method for both mutation screening and genotyping, and fluorescence-based sequencing is widely used because the genotyping process is automated. The high cost for genotyping a large number of samples is prohibitive. This method also requires several post-PCR procedures such as cycle sequencing reaction after PCR and removal of unused primers and dNTPs. In this study, direct sequencing was used for confirming the specificity of PCR and the genotypes of representative samples before large-scale genotyping, and for resolving problematic genotypes obtained with other methods.

5.2.2 SNP screening with PE-DHPLC platform in DNA pooling studies

Genotyping large numbers of SNPs individually is performed routinely but is cost-prohibitive for large-scale genetic studies. DNA pooling is a reliable and cost-saving alternative genotyping method. DNA samples are pooled before PCR to reduce the cost and increase the efficiency of SNP screening by DHPLC (Han, et al., 2005; Wolford, et al., 2000).

5.2.2.1 SNP selection for screening

We examined 62 tag SNPs from the seven positional candidate genes in the *MYP2* locus. These tag SNPs were chosen initially for availability of assays, and spacing across the gene. For this initial screen of candidates, we did not attempt to perform a comprehensive analysis of all polymorphisms in each gene. LD results from HapMap data confirm the emerging idea of haplotype blocks, with the SNPs in several genes (*LPIN2, MYCL12B, MYCL12A* and *ZFP161*) in strong LD across the entire gene. These data suggest that we effectively reduced the number of independent SNPs.

5.2.2.2 SNP frequencies in DNA pools

For screening candidate genes for high myopia in the *MYP2* locus, estimated allele frequencies of SNP markers were derived by averaging relative allele signals obtained by analysis of primer-extended products in a denaturing high performance liquid chromatography system. Under completely denaturing condition with the column temperature kept at 70°C, DHPLC is capable of separating single-based extended products that have the same length, but differ from each other by one base

(the single base incorporated in the primer extension reaction) (Xiao & Oefner, 2001).

In this study, two sets of pools (6 case pools and 6 control pools) consisting of DNA samples from 300 high myopia cases and 300 controls were used. The platform has been demonstrated as an useful tool to estimate allele frequency for pooling studies (Hoogendoorn et al., 2000), but the speed of analysis is not fast enough to handle massive screening of hundreds of thousands of markers at one time for a larger number of loci or GWAS.

The PE-DHPLC analysis of DNA pools was suggested to be of a lower resolution towards rare alleles when compared with other pooling techniques (Blazej, et al., 2003; Chen, et al., 2002; Lavebratt & Sengul, 2006). In addition, the method is quite limited in the multiplex design as used in SNaPshot assays (Applied Biosystems), which separate primer extension products only on the basis of size and hence can be multiplexed by using primers of different lengths without laborious optimization. This strategy is nevertheless a comprehensive, rapid, efficient and economical method for accurately estimating allele frequencies in DNA pools.

5.2.2.3 Two-stage approach

It is important that positive results generated from the frequency estimation in the pooled cases and controls be validated by individual genotyping (Risch & Teng, 1998). The first reason for the validation is to rule out false positive results from the analysis of pooled DNA by obtaining individual genotyping results (Shaw et al., 1998). Testing of differences between cases and controls can be based merely on the

allele frequencies in the initial screen of a dense set of markers in pooled samples. Hence, the second reason for individual genotyping is that, if deviations from Hardy-Weinberg equilibrium are seen in either group, a test on genotype frequencies should be performed.

Currently, the most effective use of DNA pooling consists of a two-stage design in which markers showing putative association are followed up by individual genotyping (Sham et al., 2002). However, several statistical methods are available to prioritize SNPs for the second stage. By comparing many statistical tests, testing difference in allelic frequency with the silhouette statistic scores, not by *P* value calculation, was one common method to rank SNPs if less than 3,000 SNPs are considered (Pearson, et al., 2007). In the current study, we used the nested ANOVA method to identify the most promising markers. The superiority of the nested ANOVA method decreases as the number of replicate measurements increases (Colliver, et al., 2000). In the *MYP2* locus, the statistical analysis of the pooled data identified the SNP rs1004961 in the *CLUL1* gene as the top candidate marker (P = 0.0031, nested ANOVA), but it was refuted upon individual genotyping (P = 0.4986, χ^2 test; **Table 4.2).** This suggests that strong genetic effects are likely to be found irrespective of the ranking based on the *P* values from nested ANOVA. In future *MYP2* pooling studies, if more markers for many more loci are considered for validation, combining the methods of ranking should be suggested to increase the number of true genetic markers compared to the nested ANOVA alone. Emphasis should be placed on the top-ranking SNPs which overlap between different ranking methods.

5.2.2.4 LD within the*MYP2* locus

The number of markers required for case-control studies depends on the extent of LD in the *MYP2* locus since all LD blocks should be tagged in this region. Although the expected LD of the studied region 18p11.31 is small, there are some recombination hotspots in 7.6 cM recombinant interval (Scavello et al., 2005). The genes being studied have 2 (*ZFP161*) to 36 exons (*MYOM1*), and show LD extending about 385 kb. The coverage of the current study was limited to SNPs with a pairwise $r^2 \geq 0.8$ and minor allele frequencies over 0.1.

By using the tag SNPs from the HapMap Project across the positional candidate genes in the *MYP2* interval, in which SNP information for the Han Chinese population is available (The international HapMap Consortium, phase 1 and 2), it saved a lot of genotyping cost and time for this current association study.

5.3 Future studies

The SNP-based strategies for complex diseases have had recent success in the myopia genetic studies. The positive association findings from the studies should routinely be followed up by:

- Generalisability studies to assess the full scope of replicated associations across different races, different endpoints, different interactions (Thomas, et al., 2009);
- Fine-mapping or resequencing to try to identify the causal variant (Wiltshire, et al., 2008); and
- Experimental studies of the biological function of the association genes (Lusa, et al., 2007).

5.3.1 Future studies for the *MYOC* gene

5.3.1.1 Meta-analysis

In view of population heterogeneity and potential bias in case-control studies, the association between *MYOC* and high myopia needs further confirmation across different studies in different populations. To perform a meta-analysis of high myopia on combined *MYOC* results from all available data sets (Leung et al., 2000; Tang et al., 2007; Vatavuk et al., 2009; Zayats et al., 2009), the basic methodology method would be to summarise each set of *MYOC* genotypes, then analyse refractive error separately in each dataset, and then use a meta-analysis technique to combine the *P* values across all data sets (Look, et al., 2002; Nakanishi et al., 2009b). To enlarge the sample pool for *MYOC* study, ethnically different groups should be invited for separate and combined meta-analysis as well. It is very important to replicate

association studies with other sets of independent samples, preferably from different ethnic groups (Colhoun et al., 2003).

5.3.1.2 Requirements for *MYOC* replication studies

It would be extremely helpful if future studies to replicate our *MYOC* results can be unified or as similar as possible for the following aspects. This would certainly facilitate meta-analysis across studies.

With respect to subject recruitment criteria, it would be ideal to have consensus on the refraction cut-off threshold for high myopia definition (as discussed in Chapter 5.1.1). Many association studies cannot be replicated because of differences in phenotype criteria (Gorroochurn, et al., 2007). High myopia has usually been defined as ≤ -6.00D for both eyes in most myopia genetics studies (Hornbeak & Young, 2009). To enhance the contribution of the genetic component to the myopia trait, a few myopia studies have recruited extremely myopic subjects (≤-10.00D) (Han et al., 2006; Wang, et al., 2006b). Subject recruitment is a bottleneck for most studies and it is very often difficult to change the recruitment criteria once started. We suggest that the data be reported for several levels of threshold refraction (≤-6.0 D, ≤-8.0 D and ≤-10.D) to allow meta-analysis to be performed at these different levels.

5.3.1.3 3′ flanking region of *MYOC*

Only polymorphisms at the 3′ end of the *MYOC* gene demonstrated association with high myopia. Thus, this region should be the main focus for further studies. According to the HapMap data, two recombination hotspots (**Figures 3.3)** surround the *MYOC* gene within a region of 60 kb. The potential causal *MYOC* SNPs

associated with high myopia would be very likely located within this region. Tag SNPs that are located in the 3′ flanking region and within the recombination hotspot region should be prioritized for further association tests.

Recent evidence has suggested that binding sites in the 3′ flanking region of a gene are important for regulation of gene expression via microRNAs, which are small RNA molecules that regulate gene expression post-transcriptionally (Bergauer, et al., 2009; Majewski & Ott, 2002). As a consequence, the amount of transcribed protein would be altered if the alleles of a SNP alter the binding sites for microRNAs.

5.3.1.4 Functional studies of *MYOC*

If several replications of all positive SNPs are confirmed across studies from different population samples, then the markers can be prioritized for functional studies. The problem of cost is compounded by the fact that each replicated SNP may be in LD with many other potential causal variants. Statistical evidence of association alone cannot be used in the prioritization scheme because these LD proxies will have similar association results. Studies will often use biological data to guide the prioritization process. For example, genetic variants in biochemical pathways related to the disease can be given greater weights. Failure to do this systematically may artificially inflate the biological importance of a gene (Chanock, et al., 2007). There are now many public databases that provide data on biochemical pathways and metabolic networks (Altman, 2007; Arakawa, et al., 2005; Vastrik, et al., 2007; von Mering, et al., 2007). There are many other forms of genomic annotation and biological data that could be included for prioritizing SNPs (Jegga, et al., 2007; Lee & Shatkay, 2008; Wang, et al., 2006a; Yuan, et al., 2006). However, it

is unclear how to go from one parameter system to the other, and therefore difficult to compare methodologies.

As predicted using online bioinformatics tool, the positive *MYOC* SNPs identified in this study do not seem to have any functional roles in regulating the expression of the *MYOC* gene and hence the amount of the wildtype myocilin in the target tissue (see **Section 3.4.4**). The functional roles of these SNPs and other potential causal variants identified in future studies should ultimately tested by experimental methods. In vitro functional assays can be the first step. Gel mobility shift assay, DNA footprinting and reporter gene analysis can be used to address this important question (Knight, 2003). We selected *MYOC* as one of the first candidate genes tested for allelic association with high myopia because of the close relationship between glaucoma and myopia on the one hand (Nomura et al., 2004; Wong et al., 2003), and the causation of primary open angle glaucoma by *MYOC* mutations on the other hand (Stone et al., 1997). Intriguingly, there is now ample evidence that myocilin-associated glaucoma is caused by misfolded mutant myocilin (Joe, et al., 2003; Liu & Vollrath, 2004), but not by a reduced amount (Fan, et al., 2004a) or an increased expression (Gould, et al., 2004) of wildtype myocilin.

Functional studies which can help clarify mechanism can counterbalance deficiencies in some genetic association studies. Such data, however, should not be required in all genetic association studies since SNPs found to be important in genetic association studies may produce no change in protein expressed (synonymous), may be located in non-coding regions (introns or 5' or 3' untranslated

regions), and/or may require an intact organism/animal/human in order to demonstrate the physiological relevance. Many genetic animal models are available for functional studies of ocular diseases, e.g. cataract, retinal degeneration, glaucoma and retinal degeneration. Mouse models have become dominant and are used more often than other mammalian species (Budzynski, et al., 2006) in studying heritable ocular disease or in test interventions. Therefore, mouse models are easier to perform for *MYOC* function study (Allen, et al., 2008; Fenoy, et al., 2001; Lindsey & Weinreb, 2005). Inherited and/or induced models exist for various diseases and conditions, but not all of these models are well characterised in a uniform fashion. Although experimentally induced models do not allow us to study the initial causes of inherited disease, they can be useful for assessing genetic susceptibility factors (Pastino, et al., 2000) and testing treatments (Kim, et al., 2009).

5.3.2 Future studies for the *MYP2* locus

With a pooling approach, the likelihood of false-positives is high when performing an initial screen on the *MYP2* locus. Accordingly, validation is essential to confirm the findings observed in the first stage of the study. Within the expanded 7.6-cM interval of the *MYP2* locus at chromosome 18p11.31, individual genotyping of the 9 putative markers from 5 candidate genes was performed to ensure that the construction of the genomic DNA pools did not inflate the true allelic frequency difference between cases and controls. To follow up the positive association results for the current *MYP2* study, independent replication in a second population and functional studies will still be required. Although this pooling screening was based

on pooled DNA samples, it is likely to have missed or overrated some genetic markers implicated in high myopia. It is unclear whether the results obtained in the Hong Kong Chinese population can be directly applied to other populations. Further empirical studies will be required to confirm these results.

Other researchers screening for myopia candidate genes in the *MYP2* interval may wish to avoid repeated screening of those genes that have been excluded. Other than the seven candidate genes evaluated in the current study, the *YES1*, *C18ORF2*, *METTL4*, *KNTC2*, *ENOSF1*, *ADCYAP1*, *TYMS*, *CR627458*, *COLEC12*, *THOC1* and *CETN1* genes within the *MYP2* locus should then be screened for the association for myopia although such positional candidate genes have been found to be less related to eye function.

In summary, the pooling approach is useful as a screening tool for the myopia locus, and more sophisticated statistical methods should be developed. The limitation of current DNA pooling analyses would then be overcome and DNA pooling may therefore be further applied to the study of other myopia loci.

5.4 Other considerations for future studies

5.4.1 DNA pooling approach as the first stage of GWAS

Because of the high genotyping cost of GWAS, many investigators have used a two-stage design for GWAS. This involves using pooling strategies on high-density genotyping arrays for an initial screening of "promising" SNPs at a less stringent significance level, and is followed by the independent "exact replication" study in a similar population of the same promising SNPs that capture allelic differences *between* cases and controls. Substantial cost savings (approximately 20-fold reduction in the total cost) were demonstrated with satisfactory levels of overall type I error when compared with individual genotyping with the optimal design depending primarily upon the ratio of costs per genotype for stages I and II (Bosse et al., 2009). By using a sufficient number of replications, array-based pooling could minimize errors from pool construction (Macgregor, 2007). Whole blood pooling has also been shown to reduce GWAS cost and speed up initial screening stage I in identifying the associated variants for eye color, age-related macular degeneration, and pseudoexfoliation syndrome (Craig, et al., 2009). This involves mixing equal volume of whole blood aliquots from case and control cohorts *prior to* DNA extraction, and the whole blood samples have *not* been measured for the white cell counts. The simplicity and robustness of "pooling of blood" (POB) method will promote association gene discovery for the complex disease by the ease of performing GWAS from smaller research groups. The drawback of POB is that the lack of stored DNA from each individual included in the blood pools does not permit individual genotyping in the second stage. In addition, white cells and hence the

DNA in the whole blood samples may have degraded too much before enough numbers of blood samples are collected for pooling.

5.4.2 Copy number variation (CNV)

SNPs have occupied the predominant position among many human variations (Bae, et al., 2008). However, the concept of CNV offers a new tool for understanding human genomic variation. CNVs influence gene expression (Henrichsen, et al., 2009), phenotypic variation (Hasin-Brumshtein, et al., 2009), and adaptation by altering gene dosage (Lee & Jeon, 2008). Therefore, it may affect disease susceptibility to various complex traits and diseases (Ionita-Laza, et al., 2008; Plagnol, 2009). Recently, several studies have reported a relationship between CNV and complex diseases including autism (Sebat, et al., 2007), bipolar disorder (Grozeva, et al.), autoimmunity (Schaschl, et al., 2009), and rheumatoid arthritis (Lee & Jeon, 2008). The accurate and reliable genotyping methods and statistical analysis for CNV are increasingly important to the dramatically increased use of CNV in association studies. HumanHap300 BeadChip (Illumina, San Diego) (Alonso, et al., 2010) and BeadStudio 3.0 software (Colella, et al., 2007) have been proven to be able to type multi-allelic CNV markers for studies (Seo, et al., 2007). The copy number variation of CXorf2/TEX28 and opsin genes has been used for myopia studies and *TEX28* gene CNV was demonstrated to be associated with the *MYP1* X-linked myopia phenotypes in five pedigrees (with high myopia and either protanopia or deuteranopia) (Metlapally et al., 2009b). CXorf2/TEX28 is a nested, intercalated gene within the red-green opsin cone pigment gene tandem array on Xq28. To examine CNVs, ultra-high resolution array-comparative genomic

hybridization (array-CGH) assays were performed to compare the patient genomic DNA with control samples (two pairs from two pedigrees). Opsin or TEX28 gene-targeted quantitative real-time gene expression assays (comparative Ct method) were performed to validate the array-CGH findings. The Array-CGH findings revealed predicted duplications in affected patient samples.

CNV data offers a challenge to investigators as there are a number of analytic steps necessary to achieve reliable and accurate copy number measurements. First, the high limits of detection in DNA concentration between conditions, thus the absolute concentration of DNA to be analysed must exceed certain thresholds (Ohashi, 2009). The normalization process might be required to reduce variation between chips and to minimize false negatives. Second, higher frequency differences of some CNVs were found between geographic regions and among populations, but very few methods have been developed to control for population structure in CNV association studies. Therefore, most of the methods developed for SNP data to correct for population structure have been applied to CNV data. To address the above limitations, the inclusion of CNV-targeting probes on high-density SNP arrays should be prepared for joint analysis of SNPs and CNVs in genome-wide association studies (McCarroll, 2008).

5.4.3 Gene-environment interactions

Future studies of myopia should examine both genetic and environmental variables and examine for gene-environment interactions. There is strong evidence for rapid, environmentally induced change in the prevalence of myopia from one generation to

the next in South East Asia, associated with industrialized settings (Lewallen, et al., 1995), increased education (Lam et al., 2004) and urbanization (Dirani et al., 2009). Myopia is hypothesised to be highly complex, with interactions among genes and environmental risk factors playing a major role in the process (Klein et al., 2005). Consideration should be given to the role of environmental impacts on genetic influences, such as interactions of early-age near-work and genotype. The results of genetic studies have been varied even in the same ethnic populations (Lakshmi, et al., 2002). There may be other genetic and environmental factors playing a role in the different groups studied. Knowledge of genetic mechanisms involved in myopia refractive error susceptibility may allow treatment to prevent progression or to further examine gene-environment interactions. Early genetic predisposition detection for developing severe refractive errors may be useful for efficient and cost-effective screening programme design. Consideration also needs to be given to the identification of phenotypes indicating aetiologically homogeneous subgroups, e.g., early age-of-onset, with/without retinal degenerative changes, or classification by individual response to treatments that reduce accommodation to near objects, such as progressive addition lens use.

In recent years, a number of statistical models and bioinformatics approaches have been designed and developed to tackle issues of gene-environment interaction in many complex diseases such as asthma, Alzheimer disease and brain cancer (Ritchie, 2005; Sengler, et al., 2002). Eventually, an integrated platform bringing together all of the above will probably be necessary to secure relevant information specific to a particular combination of conditions and settings (age, geo-ethnicity and exposure).

Multifactor Dimensionality Reduction (MDR) method is a powerful statistical approach used to detect gene-gene or gene-environment interactions in the presence or absence of statistically detectable main effects (Hahn, et al., 2003). This computational approach is a non-parametric and model-free approach that has been shown to have reasonable power to detect epistasis in both theoretical and empirical studies (Motsinger & Ritchie, 2006). Polymorphisms in candidate genes and a variety of environmental risk factors consisting of a medical history, familial history, gender, education, nearwork tasks, outdoor activities, refraction and biometric measurements can be collected to characterise the development of myopia development. Questionnaires can be designed to collect data on probable environmental conditions and behaviour of the subjects in order to enhance the contribution of the environmental factors and gene-environment interaction to the myopia trait of the subjects. Examples include ocular development and health measures collected from Sydney school children for the Sydney Myopia Study design (Ojaimi, et al., 2005), the development of refractive errors and ocular biometrics in an Australia twin population (Dirani, et al., 2006), and familial influences on school-aged siblings in Singapore Cohort Study of the Risk factors for Myopia (SCORM) (Guggenheim, et al., 2007). Standardized questionnaire protocols and methodology were recommended for comparison with international population-based data.

Another interesting gene-environment interaction that predisposes patients to myopia is abnormal lighting conditions or visual blur in the developing eye (Ciuffreda & Wallis, 1998). Degrading the retinal image provides signals to the underlying ocular

tissues to promote or restrict axial elongation of the globe with refractive error in the developing eyes demonstrated with several animal models (chicken, tree shrew, and primate animal models) (Wallman et al., 1978). There are uncertainties about the applicability of these experimental plasticity of eye growth and refractive error development to physiological human myopia (Zadnik & Mutti, 1995). For the daily environment faced by school-aged children, there is no severe form deprivation vision that is used to induce myopia in animals; the sensitive period for myopia development in children (5 to12 years of age) is much later than that in deprivation myopia of animals (15 days to 4 weeks) (Norton, 1990).

5.4.4 Gene-gene interactions

Candidate gene studies indicate that myopia is very complex. Disease expression of myopia involves more than 1 gene (epistasis), of which some may display incomplete penetrance or variable expressivity (Feldkamper & Schaeffel, 2003). There is a need for a genome-wide approach, incorporating candidate genes but not restricted to the study of candidate genes, to explore the relative contributions and interactions between known candidate genes and possibly novel genes in increased myopia susceptibility (Ziegler, et al., 2008). Interaction analyses of multiple candidate genes could apply a logistic regression model or multifactor dimensionality reduction to determine polygenic aetiology of myopia (Duell, et al., 2008). From myopia biochemical pathways, combinations of polymorphisms from several candidate genes can be included in these analyses to identify potentially new risk factor combinations. Variations in three genes coding for matrix metalloproteinases (*MMP1*, *MMP3* and *MMP9* genes) were assessed for common myopia in United Kingdom (Hall et al., 2009); MMP enzymes degrade matrix

proteins and modulate scleral extensibility, and have differential expression in experimental myopia. By using logistic regression, the strength of association with different *MMP* polymorphisms was assessed. Risk of myopia increased progressively with the dose of these three alleles, showing a greater than 10-fold difference across the range. The population distribution of refractive error is not normally distributed, and myopia is more suited to analysis as a categorical variable.

5.5 Limitations

Results from our study are based on individuals of Southern Chinese descent only and this study should be replicated in other ethnicities. Moreover, our study was not powerful enough to detect less common variants (say, MAF < 0.1) in association with high myopia. Finally, the functional significance of the positive SNPs identified in the *MYOC* gene and the *MYP2* locus study is unknown, and true causal variants remain to be determined. The proportion of refractive errors explained by the SNPs identified here is extreme myopia, the genetic component of common myopia or hyperopia might be different from that of high myopia (Klein et al., 2005). However, the genes discovered in our study provide a basis for future studies of myopia.

5.5.1 Limitations of DNA pooling methods

DNA pooling methods clearly have limitations when compared to individual genotyping. First, DNA pooling adds extra experimental error (e.g. pipetting for pool construction) to the allele frequency measurement that directly influences the power to detect small effect sizes (Barratt et al., 2002; Jawaid & Sham, 2009). In addition,

pooling results in a loss in the ability to study haplotypes, or specific genetic models as well as to undertake gene-gene and gene-environment interaction studies (Zeng & Lin, 2005). Lastly, DNA pooling does not allow detection of and adjustment for population stratification (Sham et al., 2002). It is thus important to examine population stratification using a limited number of individual genotypes before pooling DNA. Nevertheless, despite these limitations, the pooled genome-wide screen technique provides an attractive alternative to the currently expensive GWAS (Craig et al., 2009).

Chapter 6

Summary and conclusion

Population-based and family-based association approaches were employed to test for association between high myopia and one candidate gene and one myopia locus in Hong Kong Chinese. The *MYOC* gene was selected based on the results of our previous family study on high myopia and the striking consistent associations across our case-control samples and family samples confirmed the role of *MYOC* in myopia susceptibility.

From the pooling results in the *MYP2* locus screening, *LPIN2* gene showed a definitive association with high myopia. According to the results at this stage, the *LPIN2* gene may contain potential polymorphisms affecting myopia susceptibility in the Hong Kong Chinese. First, replication with an independent set of samples, preferably from populations of different ethnic origins (Colhoun et al., 2003), is essential to confirm association between these genetic variations and high myopia. Second, further replication studies or meta-analysis with sufficient power and comprehensive coverage will be needed to clarify the role of *LPIN2* sequence variants in myopia development.

References

Abdalla, M.I., & Hamdi, M. (1970). Applanation ocular tension in myopia and emmetropia. *Br J Ophthalmol, 54* (2), 122-125.

Abecasis, G.R., Cherny, S.S., Cookson, W.O., & Cardon, L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet, 30* (1), 97-101.

Al-Mosawi, Z.S., Al-Saad, K.K., Ijadi-Maghsoodi, R., El-Shanti, H.I., & Ferguson, P.J. (2007). A splice site mutation confirms the role of LPIN2 in Majeed syndrome. *Arthritis Rheum, 56* (3), 960-964.

Alfano, J.E. (1958). Myopia of prematurity. *Am J Ophthalmol, 46* (1, Part 1), 45-49.

Allen, J.E., Adjei, O., Bain, O., Hoerauf, A., Hoffmann, W.H., Makepeace, B.L., Schulz-Key, H., Tanya, V.N., Trees, A.J., Wanji, S., & Taylor, D.W. (2008). Of mice, cattle, and humans: the immunology and treatment of river blindness. *PLoS Negl Trop Dis, 2* (4), e217.

Alonso, A., Julia, A., Tortosa, R., Canaleta, C., Canete, J.D., Ballina, J., Balsa, A., Tornero, J., & Marsal, S. (2010). CNstream: a method for the identification and genotyping of copy number polymorphisms using Illumina microarrays. *BMC Bioinformatics, 11*, 264.

Alsbirk, P.H. (1977). Variation and heritability of ocular dimensions. A population study among adult Greenland Eskimos. *Acta Ophthalmol (Copenh), 55* (3), 443-456.

Altman, R.B. (2007). PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. *Nat Genet, 39* (4), 426.

Altmuller, J., Palmer, L.J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet, 69* (5), 936-950.

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., & Lander, E.S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature, 407* (6803), 513-516.

Amos, C.I. (2007). Successful design and conduct of genome-wide association studies. *Hum Mol Genet, 16 Spec No. 2*, R220-225.

Amos, C.I., Wan, Y., Siminovitch, K.A., & Rubin, L.A. (1997). Estimating the strength of genetic effects: a comparison of maximum likelihood and transmission disequilibrium methods in the study of ankylosing spondylitis. *Hum Immunol, 57* (1), 44-50.

Andrew, T., Maniatis, N., Carbonaro, F., Liew, S.H., Lau, W., Spector, T.D., & Hammond, C.J. (2008). Identification and replication of three novel myopia common susceptibility gene loci on chromosome 3q26 using linkage and linkage disequilibrium mapping. *PLoS Genet, 4* (10), e1000220.

Angle, J., & Wissmann, D.A. (1980). The epidemiology of myopia. *Am J Epidemiol, 111* (2), 220-228.

Arakawa, K., Kono, N., Yamada, Y., Mori, H., & Tomita, M. (2005). KEGG-based pathway visualization tool for complex omics data. *In Silico Biol, 5* (4), 419-423.

Ardlie, K.G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the

human genome. *Nat Rev Genet, 3* (4), 299-309.

Asano, T., Kondo, M., Kondo, N., Ueno, S., Terasaki, H., & Miyake, Y. (2004). High prevalence of myopia in Japanese patients with multiple evanescent white dot syndrome. *Jpn J Ophthalmol, 48* (5), 486-489.

Ashton, G.C. (1985). Segregation analysis of ocular refraction and myopia. *Hum Hered, 35* (4), 232-239.

Aulchenko, Y.S., Pullen, J., Kloosterman, W.P., Yazdanpanah, M., Hofman, A., Vaessen, N., Snijders, P.J., Zubakov, D., Mackay, I., Olavesen, M., Sidhu, B., Smith, V.E., Carey, A., Berezikov, E., Uitterlinden, A.G., Plasterk, R.H., Oostra, B.A., & van Duijn, C.M. (2007). LPIN2 is associated with type 2 diabetes, glucose metabolism, and body composition. *Diabetes, 56* (12), 3020-3026.

Austin, B.A., Coulon, C., Liu, C.Y., Kao, W.W., & Rada, J.A. (2002). Altered collagen fibril formation in the sclera of lumican-deficient mice. *Invest Ophthalmol Vis Sci, 43* (6), 1695-1701.

Avetisov, E.S., Savitskaya, N.F., Vinetskaya, M.I., & Iomdina, E.N. (1983). A study of biochemical and biomechanical qualities of normal and myopic eye sclera in humans of different age groups. *Metab Pediatr Syst Ophthalmol, 7* (4), 183-188.

Avila, M.P., Weiter, J.J., Jalkh, A.E., Trempe, C.L., Pruett, R.C., & Schepens, C.L. (1984). Natural history of choroidal neovascularization in degenerative myopia. *Ophthalmology, 91* (12), 1573-1581.

Bacanu, S.A., Devlin, B., & Roeder, K. (2000). The power of genomic control. *Am J Hum Genet, 66* (6), 1933-1944.

Bae, J.S., Cheong, H.S., Kim, J.O., Lee, S.O., Kim, E.M., Lee, H.W., Kim, S., Kim, J.W., Cui, T., Inoue, I., & Shin, H.D. (2008). Identification of SNP markers for common CNV regions and association analysis of risk of subarachnoid aneurysmal hemorrhage in Japanese population. *Biochem Biophys Res Commun, 373* (4), 593-596.

Bambridge, A. (2002). Approaching myopia holistically: a case study and theoretical exploration. *J Altern Complement Med, 8* (3), 371-377.

Barcellos, L.F., Klitz, W., Field, L.L., Tobias, R., Bowcock, A.M., Wilson, R., Nelson, M.P., Nagatomi, J., & Thomson, G. (1997). Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am J Hum Genet, 61* (3), 734-747.

Baron, J.M., Rubben, A., & Grussendorf-Conen, E.I. (1996). Evaluation of a new general primer pair for rapid detection and differentiation of HSV-1, HSV-2, and VZV by polymerase chain reaction. *J Med Virol, 49* (4), 279-282.

Barratt, B.J., Payne, F., Lowe, C.E., Hermann, R., Healy, B.C., Harold, D., Concannon, P., Gharani, N., McCarthy, M.I., Olavesen, M.G., McCormack, R., Guja, C., Ionescu-Tirgoviste, C., Undlien, D.E., Ronningen, K.S., Gillespie, K.M., Tuomilehto-Wolf, E., Tuomilehto, J., Bennett, S.T., Clayton, D.G., Cordell, H.J., & Todd, J.A. (2004). Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes, 53* (7), 1884-1889.

Barratt, B.J., Payne, F., Rance, H.E., Nutland, S., Todd, J.A., & Clayton, D.G. (2002). Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet, 66* (Pt 5-6), 393-405.

Barrett, J.C., Fry, B., Maller, J., & Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics, 21* (2), 263-265.

Ben-Shlomo, G., & Ofri, R. (2006). Development of inner retinal function, evidenced by the pattern electroretinogram, in the rat. *Exp Eye Res, 83* (2), 417-423.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav Brain Res, 125* (1-2), 279-284.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. 57 (pp. 289-300): JSTOR.

Benjamini, Y., & Yekutieli, D. (2005). Quantitative trait loci analysis using the false discovery rate. 171 (p. 783): Genetics Soc America.

Bergauer, T., Krueger, U., Lader, E., Pilk, S., Wolter, I., & Bielke, W. (2009). Analysis of putative miRNA binding sites and mRNA 3' ends as targets for siRNA-mediated gene knockdown. *Oligonucleotides, 19* (1), 41-52.

Bertolino, E., Reimund, B., Wildt-Perinic, D., & Clerc, R.G. (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem, 270* (52), 31178-31188.

Biino, G., Palmas, M.A., Corona, C., Prodi, D., Fanciulli, M., Sulis, R., Serra, A., Fossarello, M., & Pirastu, M. (2005). Ocular refraction: heritability and genome-wide search for eye morphometry traits in an isolated Sardinian population. *Hum Genet, 116* (3), 152-159.

Bitzer, M., & Schaeffel, F. (2002). Defocus-induced changes in ZENK expression in the chicken retina. *Invest Ophthalmol Vis Sci, 43* (1), 246-252.

Blazej, R.G., Paegel, B.M., & Mathies, R.A. (2003). Polymorphism ratio sequencing: a new approach for single nucleotide polymorphism discovery and genotyping. *Genome Res, 13* (2), 287-293.

Bonner, M., & Ballard, L. (1999). Considerations in adding mutation detection services to a sequencing core facility. *J Biomol Tech, 10* (4), 177-186.

Borish, I.M., & Indiana University. Division of Optometry. (1970). Clinical refraction. (pp. xxi, 1381 p.). Chicago, Ill.]: Professional Press.

Bortoluzzi, S., d'Alessi, F., & Danieli, G.A. (2000). A novel resource for the study of genes expressed in the adult human retina. *Invest Ophthalmol Vis Sci, 41* (11), 3305-3308.

Bosse, Y., Bacot, F., Montpetit, A., Rung, J., Qu, H.Q., Engert, J.C., Polychronakos, C., Hudson, T.J., Froguel, P., Sladek, R., & Desrosiers, M. (2009). Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. *Hum Genet, 125* (3), 305-318.

Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet, 33 Suppl*, 228-237.

Bottini, N., Vang, T., Cucca, F., & Mustelin, T. (2006). Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol, 18* (4), 207-213.

Bowes Rickman, C., Ebright, J.N., Zavodni, Z.J., Yu, L., Wang, T., Daiger, S.P., Wistow, G.,

Boon, K., & Hauser, M.A. (2006). Defining the human macula transcriptome and candidate retinal disease genes using EyeSAGE. *Invest Ophthalmol Vis Sci, 47* (6), 2305-2316.

Brabyn, J.A., Schneck, M.E., Lott, L.A., & Haegerstrom-Portnoy, G. (2005). Night driving self-restriction: vision function and gender differences. *Optom Vis Sci, 82* (8), 755-764.

Breen, G.A., Lusis, A.J., & Paigen, K. (1977). Linkage of genetic determinants for mouse beta-galactosidase electrophoresis and activity. *Genetics, 85* (1), 73-84.

Brennan, P. (1999). Design and analysis issues in case-control studies addressing genetic susceptibility. *IARC Sci Publ,* (148), 123-132.

Bressan, G.M., Daga-Gordini, D., Colombatti, A., Castellani, I., Marigo, V., & Volpin, D. (1993). Emilin, a component of elastic fibers preferentially located at the elastin-microfibrils interface. *J Cell Biol, 121* (1), 201-212.

Brohede, J., Dunne, R., McKay, J.D., & Hannan, G.N. (2005). PPC: an algorithm for accurate estimation of SNP allele frequencies in small equimolar pools of DNA using data from high density microarrays. *Nucleic Acids Res, 33* (17), e142.

Buchanan, A.V., Sholtis, S., Richtsmeier, J., & Weiss, K.M. (2009). What are genes "for" or where are traits "from"? What is the question? *Bioessays, 31* (2), 198-208.

Budzynski, E., Lee, Y., Sakamoto, K., Naggert, J.K., & Nishina, P.M. (2006). From vivarium to bedside: lessons learned from animal models. *Ophthalmic Genet, 27* (4), 123-137.

Burton, T.C. (1989). The influence of refractive error and lattice degeneration on the incidence of retinal detachment. *Trans Am Ophthalmol Soc, 87*, 143-155; discussion 155-147.

Butcher, E.C., Berg, E.L., & Kunkel, E.J. (2004). Systems biology in drug discovery. *Nat Biotechnol, 22* (10), 1253-1259.

Cardon, L.R., & Palmer, L.J. (2003). Population stratification and spurious allelic association. *Lancet, 361* (9357), 598-604.

Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L., & Nickerson, D.A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet, 74* (1), 106-120.

Carman, G.M., & Han, G.S. (2006). Roles of phosphatidate phosphatase enzymes in lipid metabolism. *Trends Biochem Sci, 31* (12), 694-699.

Chakravarti, S., Paul, J., Roberts, L., Chervoneva, I., Oldberg, A., & Birk, D.E. (2003). Ocular and scleral alterations in gene-targeted lumican-fibromodulin double-null mice. *Invest Ophthalmol Vis Sci, 44* (6), 2422-2432.

Chang, J.X., Zhang, D.D., Lin, Y., Liu, B., Lu, F., Tang, X., Yang, Y., Chen, B., Zhang, C.J., & Yang, Z.L. (2008). [Linkage analysis of one family with autosomal dominant high myopia]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 25* (4), 424-426.

Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E., Brooks, L.D., Cardon, L.R., Daly, M., Donnelly, P., Fraumeni, J.F., Jr., Freimer, N.B., Gerhard, D.S., Gunter, C., Guttmacher, A.E., Guyer, M.S., Harris, E.L., Hoh, J., Hoover, R., Kong, C.A., Merikangas,

K.R., Morton, C.C., Palmer, L.J., Phimister, E.G., Rice, J.P., Roberts, J., Rotimi, C., Tucker, M.A., Vogan, K.J., Wacholder, S., Wijsman, E.M., Winn, D.M., & Collins, F.S. (2007). Replicating genotype-phenotype associations. *Nature, 447* (7145), 655-660.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., & Fodor, S.P. (1996). Accessing genetic information with high-density DNA arrays. *Science, 274* (5287), 610-614.

Chen, C.J., Cohen, B.H., & Diamond, E.L. (1985). Genetic and environmental effects on the development of myopia in Chinese twin children. *Ophthalmic Paediatr Genet, 6* (1-2), 353-359.

Chen, H.H., Jou, Y.S., Lee, W.J., & Pan, W.H. (2008). Applying polynomial standard curve method to correct bias encountered in estimating allele frequencies using DNA pooling strategy. *Genomics, 92* (6), 429-435.

Chen, J., Germer, S., Higuchi, R., Berkowitz, G., Godbold, J., & Wetmur, J.G. (2002). Kinetic polymerase chain reaction on pooled DNA: a high-throughput, high-efficiency alternative in genetic epidemiological studies. *Cancer Epidemiol Biomarkers Prev, 11* (1), 131-136.

Chen, J.C., Schmid, K.L., & Brown, B. (2003). The autonomic control of accommodation and implications for human myopia development: a review. *Ophthalmic Physiol Opt, 23* (5), 401-422.

Chen, W.M., & Deng, H.W. (2001). A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genet Epidemiol, 21* (1), 53-67.

Chen, Z.T., Wang, I.J., Shih, Y.F., & Lin, L.L. (2009). The association of haplotype at the lumican gene with high myopia susceptibility in Taiwanese patients. *Ophthalmology, 116* (10), 1920-1927.

Cheng, K.F., & Lin, W.J. (2007). Simultaneously correcting for population stratification and for genotyping error in case-control association studies. *Am J Hum Genet, 81* (4), 726-743.

Ciner, E., Wojciechowski, R., Ibay, G., Bailey-Wilson, J.E., Stambolian, D. (2008) Genomewide scan of ocular refraction in African-American families shows significant linkage to chromosome 7p15. *Genet Epidemiol, 32*(5)*,* 454–463.

Ciuffreda, K.J., & Wallis, D.M. (1998). Myopes show increased susceptibility to nearwork aftereffects. *Invest Ophthalmol Vis Sci, 39* (10), 1797-1803.

Clemmons, D.R. (2004). Role of insulin-like growth factor iin maintaining normal glucose homeostasis. *Horm Res, 62 Suppl 1*, 77-82.

Cline, D., Hofstetter, H.W., & Griffin, J.R. (1989). Dictionary of visual science. (pp. xxi, 820 p.). Radnor, Pa.: Chilton.

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., & Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res, 35* (6), 2013-2025.

Colhoun, H.M., McKeigue, P.M., & Davey Smith, G. (2003). Problems of reporting genetic

associations with complex outcomes. *Lancet, 361* (9360), 865-872.

Collins, A., Lonjou, C., & Morton, N.E. (1999). Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci U S A, 96* (26), 15173-15177.

Collins, F. (1997). Human genetics: where do we stand? *Origins, 26* (28), 464-468.

Collins, F.S., Guyer, M.S., & Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science, 278* (5343), 1580-1581.

Colliver, T.L., Hess, E.J., Pothos, E.N., Sulzer, D., & Ewing, A.G. (2000). Quantitative and statistical analysis of the shape of amperometric spikes recorded from two populations of cells. *J Neurochem, 74* (3), 1086-1097.

Conte, I., Lestingi, M., den Hollander, A., Alfano, G., Ziviello, C., Pugliese, M., Circolo, D., Caccioppoli, C., Ciccodicola, A., & Banfi, S. (2003). Identification and characterisation of the retinitis pigmentosa 1-like1 gene (RP1L1): a novel candidate for retinal degenerations. *Eur J Hum Genet, 11* (2), 155-162.

Cordain, L., Eaton, S.B., Brand Miller, J., Lindeberg, S., & Jensen, C. (2002). An evolutionary analysis of the aetiology and pathogenesis of juvenile-onset myopia. *Acta Ophthalmol Scand, 80* (2), 125-135.

Cordell, H.J., & Clayton, D.G. (2002). A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet, 70* (1), 124-141.

Costabile, M., Quach, A., & Ferrante, A. (2006). Molecular approaches in the diagnosis of primary immunodeficiency diseases. *Hum Mutat, 27* (12), 1163-1173.

Craig, J.E., Hewitt, A.W., McMellon, A.E., Henders, A.K., Ma, L., Wallace, L., Sharma, S., Burdon, K.P., Visscher, P.M., Montgomery, G.W., & MacGregor, S. (2009). Rapid inexpensive genome-wide association using pooled whole blood. *Genome Res, 19* (11), 2075-2080.

Curran, S., Hill, L., O'Grady, G., Turic, D., Asherson, P., Taylor, E., Sham, P., Craig, I., & Vaughan, P. (2002). Validation of single nucleotide polymorphism quantification in pooled DNA samples with SNaPIT. A glycosylase-mediated methods for polymorphism detection method. *Mol Biotechnol, 22* (3), 253-262.

Curtin, B.J. (1979). Physiologic vs pathologic myopia: genetics vs environment. *Ophthalmology, 86* (5), 681-691.

Curtin, B.J. (1982). Posterior staphyloma development in pathologic myopia. *Ann Ophthalmol, 14* (7), 655-658.

Curtin, B.J. (1985). The myopias : basic science and clinical management. (pp. xv, 495 p.). Philadelphia: Harper & Row.

Curtin, B.J. (1988). Pathologic myopia. *Acta Ophthalmol Suppl, 185*, 105-106.

Curtis, D. (1997). Use of siblings as controls in case-control association studies. *Ann Hum Genet, 61* (Pt 4), 319-333.

Dahlman, I., Eaves, I.A., Kosoy, R., Morrison, V.A., Heward, J., Gough, S.C., Allahabadia, A., Franklyn, J.A., Tuomilehto, J., Tuomilehto-Wolf, E., Cucca, F., Guja, C.,

Ionescu-Tirgoviste, C., Stevens, H., Carr, P., Nutland, S., McKinney, P., Shield, J.P., Wang, W., Cordell, H.J., Walker, N., Todd, J.A., & Concannon, P. (2002). Parameters for reliable results in genetic association studies in common disease. *Nat Genet, 30* (2), 149-150.

Daniels, J., Holmans, P., Williams, N., Turic, D., McGuffin, P., Plomin, R., & Owen, M.J. (1998). A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. *Am J Hum Genet, 62* (5), 1189-1197.

Darvasi, A., & Soller, M. (1994). Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics, 138* (4), 1365-1373.

de Bakker, P.I. (2009). Selection and evaluation of Tag-SNPs using Tagger and HapMap. *Cold Spring Harb Protoc, 2009* (6), pdb ip67.

de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nat Genet, 37* (11), 1217-1223.

Deepak, S., Kottapalli, K., Rakwal, R., Oros, G., Rangappa, K., Iwahashi, H., Masuo, Y., & Agrawal, G. (2007). Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes. *Curr Genomics, 8* (4), 234-251.

den Hollander, A.I., van Driel, M.A., de Kok, Y.J., van de Pol, D.J., Hoyng, C.B., Brunner, H.G., Deutman, A.F., & Cremers, F.P. (1999). Isolation and mapping of novel candidate genes for retinal disorders using suppression subtractive hybridization. *Genomics, 58* (3), 240-249.

Deng, H.W. (2001). Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. *Genetics, 159* (3), 1319-1323.

Der Kaloustian, V.M., & Baghdassarian, S.A. (1972). The autosomal recessive variety of congenital stationary night-blindness with myopia. *J Med Genet, 9* (1), 67-69.

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics, 55* (4), 997-1004.

Dietlein, T.S., Jacobi, P.C., Paulsson, M., Smyth, N., & Krieglstein, G.K. (1997). [Variability of laminin isoforms in the limbus region of the eye]. *Klin Monbl Augenheilkd, 211* (3), 188-191.

Dietz, H.C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M., & et al. (1991). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature, 352* (6333), 337-339.

Dirani, M., Chamberlain, M., Shekar, S.N., Islam, A.F., Garoufalis, P., Chen, C.Y., Guymer, R.H., & Baird, P.N. (2006). Heritability of refractive error and ocular biometrics: the Genes in Myopia (GEM) twin study. *Invest Ophthalmol Vis Sci, 47* (11), 4756-4761.

Dirani, M., Islam, A., & Baird, P.N. (2008a). Body stature and myopia-The Genes in Myopia (GEM) twin study. *Ophthalmic Epidemiol, 15* (3), 135-139.

Dirani, M., Shekar, S.N., & Baird, P.N. (2008b). Evidence of shared genes in refraction and axial length: the Genes in Myopia (GEM) twin study. *Invest Ophthalmol Vis Sci, 49* (10), 4336-4339.

Dirani, M., Tong, L., Gazzard, G., Zhang, X., Chia, A., Young, T.L., Rose, K.A., Mitchell, P., & Saw, S.M. (2009). Outdoor activity and myopia in Singapore teenage children. *Br J Ophthalmol, 93* (8), 997-1000.

Doliana, R., Bot, S., Mungiguerra, G., Canton, A., Cilli, S.P., & Colombatti, A. (2001). Isolation and characterization of EMILIN-2, a new component of the growing EMILINs family and a member of the EMI domain-containing superfamily. *J Biol Chem, 276* (15), 12003-12011.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., & Mattick, J.S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res, 19* (14), 4008.

Donkor, J., Sariahmetoglu, M., Dewald, J., Brindley, D.N., & Reue, K. (2007). Three mammalian lipins act as phosphatidate phosphatases with distinct tissue expression patterns. *J Biol Chem, 282* (6), 3450-3457.

Drexler, W., Findl, O., Schmetterer, L., Hitzenberger, C.K., & Fercher, A.F. (1998). Eye elongation during accommodation in humans: differences between emmetropes and myopes. *Invest Ophthalmol Vis Sci, 39* (11), 2140-2147.

Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered, 66* (2), 87-98.

Duell, E.J., Bracci, P.M., Moore, J.H., Burk, R.D., Kelsey, K.T., & Holly, E.A. (2008). Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer. *Cancer Epidemiol Biomarkers Prev, 17* (6), 1470-1479.

Duggal, P., Gillanders, E.M., Holmes, T.N., & Bailey-Wilson, J.E. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics, 9*, 516.

Duke-Elder, S. (1935). The practice of refraction. (Philadelphia: P. Blakiston's Son & Company).

Duke-Elder, S. (1968). Marc Amsler, 1891-1968. *Br J Ophthalmol, 52* (7), 575.

Edwards, M., & Lewis, W.H. (1991). Autosomal recessive inheritance of myopia in Hong Kong Chinese infants. *Ophthalmic Physiol Opt, 11* (3), 227-231.

Edwards, M.H. (1999). The development of myopia in Hong Kong children between the ages of 7 and 12 years: a five-year longitudinal study. *Ophthalmic Physiol Opt, 19* (4), 286-294.

Edwards, M.H., & Lam, C.S. (2004). The epidemiology of myopia in Hong Kong. *Ann Acad Med Singapore, 33* (1), 34-38.

Elston, R.C., Lin, D., & Zheng, G. (2007). Multistage sampling for genetic studies. *Annu Rev Genomics Hum Genet, 8*, 327-342.

Ericsson, O., Sivertsson, A., Lundeberg, J., & Ahmadian, A. (2003). Microarray-based resequencing by apyrase-mediated allele-specific extension. *Electrophoresis, 24* (19-20), 3330-3338.

Ewens, W.J., Li, M., & Spielman, R.S. (2008). A review of family-based tests for linkage disequilibrium between a quantitative trait and a genetic marker. *PLoS Genet, 4* (9),

e1000180.

Fan, B.J., Leung, Y.F., Pang, C.P., Fan, D.S., Wang, D.Y., Tong, W.C., Tam, P.O., Chua, J.K., Lau, T.C., & Lam, D.S. (2004a). Polymorphisms in the myocilin promoter unrelated to the risk and severity of primary open-angle glaucoma. *J Glaucoma, 13* (5), 377-384.

Fan, D.S., Cheung, E.Y., Lai, R.Y., Kwok, A.K., & Lam, D.S. (2004b). Myopia progression among preschool Chinese children in Hong Kong. *Ann Acad Med Singapore, 33* (1), 39-43.

Fan, D.S., Lam, D.S., Lam, R.F., Lau, J.T., Chong, K.S., Cheung, E.Y., Lai, R.Y., & Chew, S.J. (2004c). Prevalence, incidence, and progression of myopia of school children in Hong Kong. *Invest Ophthalmol Vis Sci, 45* (4), 1071-1075.

Farbrother, J.E., Kirov, G., Owen, M.J., & Guggenheim, J.A. (2004a). Family aggregation of high myopia: estimation of the sibling recurrence risk ratio. *Invest Ophthalmol Vis Sci, 45* (9), 2873-2878.

Farbrother, J.E., Kirov, G., Owen, M.J., Pong-Wong, R., Haley, C.S., & Guggenheim, J.A. (2004b). Linkage analysis of the genetic loci for high myopia on 18p, 12q, and 17q in 51 U.K. families. *Invest Ophthalmol Vis Sci, 45* (9), 2879-2885.

Favor, J., Gloeckner, C.J., Neuhauser-Klaus, A., Pretsch, W., Sandulache, R., Saule, S., & Zaus, I. (2008). Relationship of Pax6 activity levels to the extent of eye development in the mouse, Mus musculus. *Genetics, 179* (3), 1345-1355.

Feldkamper, M., & Schaeffel, F. (2003). Interactions of genes and environment in myopia. *Dev Ophthalmol, 37*, 34-49.

Fenoy, S., Ollero, M.D., Guillen, J.L., & del Aguila, C. (2001). Animal models in ocular toxocariasis. *J Helminthol, 75* (2), 119-124.

Ferguson, P.J., Chen, S., Tayeh, M.K., Ochoa, L., Leal, S.M., Pelet, A., Munnich, A., Lyonnet, S., Majeed, H.A., & El-Shanti, H. (2005). Homozygous mutations in LPIN2 are responsible for the syndrome of chronic recurrent multifocal osteomyelitis and congenital dyserythropoietic anaemia (Majeed syndrome). *J Med Genet, 42* (7), 551-557.

Finck, B.N., Gropler, M.C., Chen, Z., Leone, T.C., Croce, M.A., Harris, T.E., Lawrence, J.C., Jr., & Kelly, D.P. (2006). Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. *Cell Metab, 4* (3), 199-210.

Fingert, J.H., Ying, L., Swiderski, R.E., Nystuen, A.M., Arbour, N.C., Alward, W.L., Sheffield, V.C., & Stone, E.M. (1998). Characterization and comparison of the human and mouse GLC1A glaucoma genes. *Genome Res, 8* (4), 377-384.

Fischer, A.J., McGuire, J.J., Schaeffel, F., & Stell, W.K. (1999). Light- and focus-dependent expression of the transcription factor ZENK in the chick retina. *Nat Neurosci, 2* (8), 706-712.

Foroud, T. (1997). Introduction to genetic linkage analysis. *Cancer Invest, 15* (6), 548-552.

Fotouhi, A., Hashemi, H., Khabazkhoob, M., & Mohammad, K. (2007). The prevalence of refractive errors among schoolchildren in Dezful, Iran. *Br J Ophthalmol, 91* (3), 287-292.

Francois, J., & Goes, F. (1977). Ultrasonographic study of 100 emmetropic eyes. *Ophthalmologica, 175* (6), 321-327.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., & Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature, 449* (7164), 851-861.

Fredrick, D.R. (2002). Myopia. *BMJ, 324* (7347), 1195-1199.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science, 296* (5576), 2225-2229.

Garcia, C.A., Orefice, F., Nobre, G.F., Souza Dde, B., Rocha, M.L., & Vianna, R.N. (2005). [Prevalence of refractive errors in students in Northeastern Brazil.]. *Arq Bras Oftalmol, 68* (3), 321-325.

Garner, L.F., Kinnear, R.F., Klinger, J.D., & McKellar, M.J. (1985). Prevalence of myopia in school children in Vanuatu. *Acta Ophthalmol (Copenh), 63* (3), 323-326.

Gasser, R.B., & Zhu, X.Q. (1999). Sequence-based analysis of enzymatically amplified DNA fragments by mutation detection techniques. *Parasitol Today, 15* (11), 462-465.

Gentle, A., Liu, Y., Martin, J.E., Conti, G.L., & McBrien, N.A. (2003). Collagen gene

expression and the altered accumulation of scleral collagen during the development of high myopia. *J Biol Chem, 278* (19), 16587-16594.

Germer, S., & Higuchi, R. (1999). Single-tube genotyping without oligonucleotide probes. *Genome Res, 9* (1), 72-78.

Giordano, M., Mellai, M., Hoogendoorn, B., & Momigliano-Richiardi, P. (2001). Determination of SNP allele frequencies in pooled DNAs by primer extension genotyping and denaturing high-performance liquid chromatography. *J Biochem Biophys Methods, 47* (1-2), 101-110.

Goldschmidt, E. (1968). [On the etiology of myopia. An epidemiological study]. *Acta Ophthalmol (Copenh)*, Suppl 98:91+.

Goldschmidt, E. (2003). The mystery of myopia. *Acta Ophthalmol Scand, 81* (5), 431-436.

Gomes, I., Collins, A., Lonjou, C., Thomas, N.S., Wilkinson, J., Watson, M., & Morton, N. (1999). Hardy-Weinberg quality control. *Ann Hum Genet, 63* (Pt 6), 535-538.

Gordon, D., Hoh, J., Finch, S.J., Levenstien, M.A., Edington, J., Li, W., Majewski, J., & Ott, J. (2001). Two approaches for consolidating results from genome scans of complex traits: selection methods and scan statistics. *Genet Epidemiol, 21 Suppl 1*, S396-402.

Gorroochurn, P., Hodge, S.E., Heiman, G.A., Durner, M., & Greenberg, D.A. (2007). Non-replication of association studies: "pseudo-failures" to replicate? *Genet Med, 9* (6), 325-331.

Goss, D.A., Hampton, M.J., & Wickham, M.G. (1988). Selected review on genetic factors in myopia. *J Am Optom Assoc, 59* (11), 875-884.

Goss, D.A., & Winkler, R.L. (1983). Progression of myopia in youth: age of cessation. *Am J Optom Physiol Opt, 60* (8), 651-658.

Gould, D.B., Miceli-Libby, L., Savinova, O.V., Torrado, M., Tomarev, S.I., Smith, R.S., & John, S.W. (2004). Genetically increasing Myoc expression supports a necessary pathologic role of abnormal proteins in glaucoma. *Mol Cell Biol, 24* (20), 9019-9025.

Gozum, N., Cakir, M., Gucukoglu, A., & Sezen, F. (1997). Relationship between retinal lesions and axial length, age and sex in high myopia. *Eur J Ophthalmol, 7* (3), 277-282.

Greenberg, D.A. (1990). Linkage analysis assuming a single-locus mode of inheritance for traits determined by two loci: inferring mode of inheritance and estimating penetrance. *Genet Epidemiol, 7* (6), 467-479.

Grierson, I., Heathcote, L., Hiscott, P., Hogg, P., Briggs, M., & Hagan, S. (2000). Hepatocyte growth factor/scatter factor in the eye. *Prog Retin Eye Res, 19* (6), 779-802.

Grossniklaus, H.E., & Green, W.R. (1992). Pathologic findings in pathologic myopia. *Retina, 12* (2), 127-133.

Grosvenor, T. (1987). A review and a suggested classification system for myopia on the basis of age-related prevalence and age of onset. *Am J Optom Physiol Opt, 64* (7), 545-554.

Grosvenor, T. (1988). Myopia in Melanesian school children in Vanuatu. *Acta Ophthalmol Suppl, 185*, 24-28.

Grosvenor, T., & Scott, R. (1991). Comparison of refractive components in youth-onset and early adult-onset myopia. *Optom Vis Sci, 68* (3), 204-209.

Grozeva, D., Kirov, G., Ivanov, D., Jones, I.R., Jones, L., Green, E.K., St Clair, D.M., Young, A.H., Ferrier, N., Farmer, A.E., McGuffin, P., Holmans, P.A., Owen, M.J., O'Donovan, M.C., & Craddock, N. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch Gen Psychiatry, 67* (4), 318-327.

Gruber, J.D., Colligan, P.B., & Wolford, J.K. (2002). Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing. *Hum Genet, 110* (5), 395-401.

Guggenheim, J.A., Hill, C., & Yam, T.F. (2003). Myopia, genetics, and ambient lighting at night in a UK sample. *Br J Ophthalmol, 87* (5), 580-582.

Guggenheim, J.A., Kirov, G., & Hodson, S.A. (2000). The heritability of high myopia: a reanalysis of Goldschmidt's data. *J Med Genet, 37* (3), 227-231.

Guggenheim, J.A., & McBrien, N.A. (1996). Form-deprivation myopia induces activation of scleral matrix metalloproteinase-2 in tree shrew. *Invest Ophthalmol Vis Sci, 37* (7), 1380-1395.

Guggenheim, J.A., Pong-Wong, R., Haley, C.S., Gazzard, G., & Saw, S.M. (2007). Correlations in refractive errors between siblings in the Singapore Cohort Study of Risk factors for Myopia. *Br J Ophthalmol, 91* (6), 781-784.

Guo, C.Y., Lunetta, K.L., DeStefano, A.L., Ordovas, J.M., & Cupples, L.A. (2007). Informative-transmission disequilibrium test (i-TDT): combined linkage and association mapping that includes unaffected offspring as well as affected offspring. *Genet Epidemiol, 31* (2), 115-133.

Guo, X., Xiao, X., Li, S., Wang, P., Jia, X., & Zhang, Q. (2010). Nonsyndromic high myopia in a Chinese family mapped to MYP1: linkage confirmation and phenotypic characterization. *Arch Ophthalmol, 128* (11), 1473-1479.

Gwiazda, J., Ong, E., Held, R., & Thorn, F. (2000). Myopia and ambient night-time lighting. *Nature, 404* (6774), 144.

Hahn, L.W., Ritchie, M.D., & Moore, J.H. (2003). Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics, 19* (3), 376-382.

Haim, M., Fledelius, H.C., & Skarsholm (1988). X-linked myopia in Danish family. *Acta Ophthalmol (Copenh), 66* (4), 450-456.

Haliassos, A., Chomel, J.C., Grandjouan, S., Kruh, J., Kaplan, J.C., & Kitzis, A. (1989). Detection of minority point mutations by modified PCR technique: a new approach for a sensitive diagnosis of tumor-progression markers. *Nucleic Acids Res, 17* (20), 8093-8099.

Hall, N.F., Gale, C.R., Ye, S., & Martyn, C.N. (2009). Myopia and polymorphisms in genes for matrix metalloproteinases. *Invest Ophthalmol Vis Sci, 50* (6), 2632-2636.

Hammond, C.J., Andrew, T., Mak, Y.T., & Spector, T.D. (2004). 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, 75* (2), 294-304.

Hammond, C.J., Snieder, H., Gilbert, C.E., & Spector, T.D. (2001). Genes and environment in refractive error: the twin eye study. *Invest Ophthalmol Vis Sci, 42* (6), 1232-1236.

Han, W., Leung, K.H., Fung, W.Y., Mak, J.Y., Li, Y.M., Yap, M.K., & Yip, S.P. (2009). Association of PAX6 polymorphisms with high myopia in Han Chinese nuclear families. *Invest Ophthalmol Vis Sci, 50* (1), 47-56.

Han, W., Lou, D.H., Wang, J., Yip, S., & Yap, M. (2005). [Application of DHPLC in human genomic SNP screening and genotyping for all-trans-retinol dehydrogenase]. *Zhejiang Da Xue Xue Bao Yi Xue Ban, 34* (3), 248-254.

Han, W., Yap, M.K., Wang, J., & Yip, S.P. (2006). Family-based association analysis of hepatocyte growth factor (HGF) gene polymorphisms in high myopia. *Invest Ophthalmol Vis Sci, 47* (6), 2291-2299.

Hasin-Brumshtein, Y., Lancet, D., & Olender, T. (2009). Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet, 25* (4), 178-184.

Hasumi, Y., Inoko, H., Mano, S., Ota, M., Okada, E., Kulski, J.K., Nishizaki, R., Mok, J., Oka, A., Kumagai, N., Nishida, T., Ohno, S., & Mizuki, N. (2006). Analysis of single nucleotide polymorphisms at 13 loci within the transforming growth factor-induced factor gene shows no association with high myopia in Japanese subjects. *Immunogenetics, 58* (12), 947-953.

Hayasaka, S., Uchida, M., & Setogawa, T. (1990). Subretinal hemorrhages with or without choroidal neovascularization in the maculas of patients with pathologic myopia. *Graefes Arch Clin Exp Ophthalmol, 228* (4), 277-280.

Hayes, J.P., & Jenkins, S.H. (1997). Individual variation in mammals. 78 (pp. 274-293): JSTOR.

Heath, S., Robledo, R., Beggs, W., Feola, G., Parodo, C., Rinaldi, A., Contu, L., Dana, D., Stambolian, D., & Siniscalco, M. (2001). A novel approach to search for identity by descent in small samples of patients and controls from the same mendelian breeding unit: a pilot study on myopia. *Hum Hered, 52* (4), 183-190.

Heikkinen, J., Pousi, B., Pope, M., & Myllyla, R. (1999). A null-mutated lysyl hydroxylase gene in a compound heterozygote British patient with Ehlers-Danlos syndrome type VI. *Hum Mutat, 14* (4), 351.

Helgason, A., & Stefansson, K. (2010). The past, present, and future of direct-to-consumer genetic tests. *Dialogues Clin Neurosci, 12* (1), 61-68.

Henrichsen, C.N., Chaignat, E., & Reymond, A. (2009). Copy number variants, diseases and gene expression. *Hum Mol Genet, 18* (R1), R1-8.

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y), 11* (9), 1026-1030.

Hinds, D.A., Seymour, A.B., Durham, L.K., Banerjee, P., Ballinger, D.G., Milos, P.M., Cox, D.R., Thompson, J.F., & Frazer, K.A. (2004). Application of pooled genotyping to scan candidate regions for association with HDL cholesterol levels. *Hum Genomics, 1* (6), 421-434.

Hirschhorn, J.N., & Daly, M.J. (2005). Genome-wide association studies for common

diseases and complex traits. *Nat Rev Genet, 6* (2), 95-108.

Hirschhorn, J.N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet Med, 4* (2), 45-61.

Hochberg, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B, 57*, 289-300.

Hoogendoorn, B., Norton, N., Kirov, G., Williams, N., Hamshere, M.L., Spurlock, G., Austin, J., Stephens, M.K., Buckland, P.R., Owen, M.J., & O'Donovan, M.C. (2000). Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. *Hum Genet, 107* (5), 488-493.

Hoogendoorn, B., Owen, M.J., Oefner, P.J., Williams, N., Austin, J., & O'Donovan, M.C. (1999). Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. *Hum Genet, 104* (1), 89-93.

Hope, G.M., & Rubin, M.L. (1984). Night myopia. *Surv Ophthalmol, 29* (2), 129-136.

Hornbeak, D.M., & Young, T.L. (2009). Myopia genetics: a review of current research and emerging trends. *Curr Opin Ophthalmol, 20* (5), 356-362.

Horvath, S., Xu, X., & Laird, N.M. (2001). The family based association test method: strategies for studying general genotype--phenotype associations. *Eur J Hum Genet, 9* (4), 301-306.

Hotchkiss, M.L., & Fine, S.L. (1981). Pathologic myopia and choroidal neovascularization. *Am J Ophthalmol, 91* (2), 177-183.

Howlett, M.H., & McFadden, S.A. (2006). Form-deprivation myopia in the guinea pig (Cavia porcellus). *Vision Res, 46* (1-2), 267-283.

Hua, L., Zhu, H., Li, X.R., Li, J., Mo, Q.H., Liao, C., Hou, Y.X., Zhong, M., & Xu, X.M. (2004). [Accurate and rapid prenatal diagnosis of beta-thalassemia by a multiplex primer extension and denaturing high-performance liquid chromatography technique.]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 21* (6), 600-603.

Huber, C.G., Premstaller, A., Xiao, W., Oberacher, H., Bonn, G.K., & Oefner, P.J. (2001). Mutation detection by capillary denaturing high-performance liquid chromatography using monolithic columns. *J Biochem Biophys Methods, 47* (1-2), 5-19.

Hysi, P.G., Young, T.L., Mackey, D.A., Andrew, T., Fernandez-Medarde, A., Solouki, A.M., Hewitt, A.W., Macgregor, S., Vingerling, J.R., Li, Y.J., Ikram, M.K., Fai, L.Y., Sham, P.C., Manyes, L., Porteros, A., Lopes, M.C., Carbonaro, F., Fahy, S.J., Martin, N.G., van Duijn, C.M., Spector, T.D., Rahi, J.S., Santos, E., Klaver, C.C., & Hammond, C.J. (2010). A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. *Nat Genet, 42* (10), 902-905.

Ibay, G., Doan, B., Reider, L., Dana, D., Schlifka, M., Hu, H., Holmes, T., O'Neill, J., Owens, R., Ciner, E., Bailey-Wilson, J.E., & Stambolian, D. (2004). Candidate high myopia loci on chromosomes 18p and 12q do not play a major role in susceptibility to common myopia. *BMC Med Genet, 5*, 20.

Inamori, Y., Ota, M., Inoko, H., Okada, E., Nishizaki, R., Shiota, T., Mok, J., Oka, A., Ohno,

S., & Mizuki, N. (2007). The COL1A1 gene and high myopia susceptibility in Japanese. *Hum Genet, 122* (2), 151-157.

Ioannidis, J.P. (2007). Why most published research findings are false: author's reply to Goodman and Greenland. *PLoS Med, 4* (6), e215.

Ioannidis, J.P., Gwinn, M., Little, J., Higgins, J.P., Bernstein, J.L., Boffetta, P., Bondy, M., Bray, M.S., Brenchley, P.E., Buffler, P.A., Casas, J.P., Chokkalingam, A., Danesh, J., Smith, G.D., Dolan, S., Duncan, R., Gruis, N.A., Hartge, P., Hashibe, M., Hunter, D.J., Jarvelin, M.R., Malmer, B., Maraganore, D.M., Newton-Bishop, J.A., O'Brien, T.R., Petersen, G., Riboli, E., Salanti, G., Seminara, D., Smeeth, L., Taioli, E., Timpson, N., Uitterlinden, A.G., Vineis, P., Wareham, N., Winn, D.M., Zimmern, R., & Khoury, M.J. (2006). A road map for efficient and reliable human genome epidemiology. *Nat Genet, 38* (1), 3-5.

Ionita-Laza, I., Perry, G.H., Raby, B.A., Klanderman, B., Lee, C., Laird, N.M., Weiss, S.T., & Lange, C. (2008). On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test. *Genet Epidemiol, 32* (3), 273-284.

Ip, J.M., Huynh, S.C., Robaei, D., Kifley, A., Rose, K.A., Morgan, I.G., Wang, J.J., & Mitchell, P. (2008). Ethnic differences in refraction and ocular biometry in a population-based sample of 11-15-year-old Australian children. *Eye (Lond), 22* (5), 649-656.

Ip, J.M., Huynh, S.C., Robaei, D., Rose, K.A., Morgan, I.G., Smith, W., Kifley, A., & Mitchell, P. (2007). Ethnic differences in the impact of parental myopia: findings from a population-based study of 12-year-old Australian children. *Invest Ophthalmol Vis Sci, 48* (6), 2520-2528.

Iyengar, S.K., & Elston, R.C. (2007). The genetic basis of complex traits: rare variants or "common gene, common disease"? *Methods Mol Biol, 376*, 71-84.

Jacobi, F.K., & Pusch, C.M. (2010). A decade in search of myopia genes. *Front Biosci, 15*, 359-372.

Jacobi, F.K., Zrenner, E., Broghammer, M., & Pusch, C.M. (2005). A genetic perspective on myopia. *Cell Mol Life Sci, 62* (7-8), 800-808.

Jacobsen, N., Jensen, H., & Goldschmidt, E. (2007). Prevalence of myopia in Danish conscripts. *Acta Ophthalmol Scand, 85* (2), 165-170.

Jacobsen, N., Jensen, H., Lund-Andersen, H., & Goldschmidt, E. (2008). Is poor glycaemic control in diabetic patients a risk factor of myopia? *Acta Ophthalmol, 86* (5), 510-514.

Javitt, J.C., & Chiang, Y.P. (1994). The socioeconomic aspects of laser refractive surgery. *Arch Ophthalmol, 112* (12), 1526-1530.

Jawaid, A., & Sham, P. (2009). Impact and quantification of the sources of error in DNA pooling designs. *Ann Hum Genet, 73* (1), 118-124.

Jegga, A.G., Gowrisankar, S., Chen, J., & Aronow, B.J. (2007). PolyDoms: a whole genome database for the identification of non-synonymous coding SNPs with the potential to impact disease. *Nucleic Acids Res, 35* (Database issue), D700-706.

Jiang, B.C., Schatz, S., & Seger, K. (2005). Myopic progression and dark focus variation in optometric students during the first academic year. *Clin Exp Optom, 88* (3), 153-159.

Jiang, B., Yap, M.K., Leung, K.H., Ng, P.W., Fung, W.Y., Lam, W.W., Gu, Y.S., & Yip, S.P. (2011) PAX6 haplotypes are associated with high myopia in Han Chinese. *PLoS One, 12* (5), e19587.

Jobling, A.I., Nguyen, M., Gentle, A., & McBrien, N.A. (2004). Isoform-specific changes in scleral transforming growth factor-beta expression and the regulation of collagen synthesis during myopia progression. *J Biol Chem, 279* (18), 18121-18126.

Joe, M.K., Sohn, S., Hur, W., Moon, Y., Choi, Y.R., & Kee, C. (2003). Accumulation of mutant myocilins in ER leads to ER stress and potential cytotoxicity in human trabecular meshwork cells. *Biochem Biophys Res Commun, 312* (3), 592-600.

Johnson, D.H. (2000). Myocilin and glaucoma: A TIGR by the tail? *Arch Ophthalmol, 118* (7), 974-978.

Johnson, G.C., & Todd, J.A. (2000). Strategies in complex disease mapping. *Curr Opin Genet Dev, 10* (3), 330-334.

Jones, H.B. (1998). The relative power of linkage and association studies for the detection of genes involved in hypertension. *Kidney Int, 53* (6), 1446-1448.

Junghans, B.M., & Crewther, S.G. (2005). Little evidence for an epidemic of myopia in Australian primary school children over the last 30 years. *BMC Ophthalmol, 5*, 1.

Kaller, M., Lundeberg, J., & Ahmadian, A. (2007). Arrayed identification of DNA signatures. *Expert Rev Mol Diagn, 7* (1), 65-76.

Karlsson, J.I. (1974). Concordance rates for myopia in twins. *Clin Genet, 6* (2), 142-146.

Karlsson, J.L. (1976). Genetic factors in myopia. *Acta Genet Med Gemellol (Roma), 25*, 292-294.

Kempen, J.H., Mitchell, P., Lee, K.E., Tielsch, J.M., Broman, A.T., Taylor, H.R., Ikram, M.K., Congdon, N.G., & O'Colmain, B.J. (2004). The prevalence of refractive errors among adults in the United States, Western Europe, and Australia. *Arch Ophthalmol, 122* (4), 495-505.

Khan, A.S. (2004). Genomics and microarray for detection and diagnostics. *Acta Microbiol Immunol Hung, 51* (4), 463-467.

Kim, H., Pawlikowska, L., Chen, Y., Su, H., Yang, G.Y., & Young, W.L. (2009). Brain arteriovenous malformation biology relevant to hemorrhage and implication for therapeutic development. *Stroke, 40* (3 Suppl), S95-97.

Kimura, R., Mandrell, R.E., Galland, J.C., Hyatt, D., & Riley, L.W. (2000). Restriction-site-specific PCR as a rapid test to detect enterohemorrhagic Escherichia coli O157:H7 strains in environmental samples. *Appl Environ Microbiol, 66* (6), 2513-2519.

Kinge, B., Midelfart, A., Jacobsen, G., & Rystad, J. (1999). Biometric changes in the eyes of Norwegian university students--a three-year longitudinal study. *Acta Ophthalmol Scand, 77* (6), 648-652.

Kirov, G., Williams, N., Sham, P., Craddock, N., & Owen, M.J. (2000). Pooled genotyping of microsatellite markers in parent-offspring trios. *Genome Res, 10* (1), 105-115.

Kitsios, G.D., & Zintzaras, E. (2009). Genomic convergence of genome-wide investigations

for complex traits. *Ann Hum Genet, 73* (Pt 5), 514-519.

Klein, A.P., Duggal, P., Lee, K.E., Klein, R., Bailey-Wilson, J.E., & Klein, B.E. (2005). Support for polygenic influences on ocular refractive error. *Invest Ophthalmol Vis Sci, 46* (2), 442-446.

Klein, A.P., Duggal, P., Lee, K.E., Klein, R., Bailey-Wilson, J.E., & Klein, B.E. (2007). Confirmation of linkage to ocular refraction on chromosome 22q and identification of a novel linkage region on 1q. *Arch Ophthalmol, 125* (1), 80-85.

Kleinstein, R.N., Jones, L.A., Hullett, S., Kwon, S., Lee, R.J., Friedman, N.E., Manny, R.E., Mutti, D.O., Yu, J.A., & Zadnik, K. (2003). Refractive error and ethnicity in children. *Arch Ophthalmol, 121* (8), 1141-1147.

Knight, J.C. (2003). Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci (Lond), 104* (5), 493-501.

Korkko, J., Annunen, S., Pihlajamaa, T., Prockop, D.J., & Ala-Kokko, L. (1998). Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. *Proc Natl Acad Sci U S A, 95* (4), 1681-1685.

Koyama, A. (1970). [Transitional mechanism of pseudomyopia into myopia]. *Ganka, 12* (3), 212-219.

Kroger, R.H., Hirt, B., & Wagner, H.J. (1999). Effects of retinal dopamine depletion on the growth of the fish eye. *J Comp Physiol A, 184* (4), 403-412.

Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nat Genet, 17* (1), 21-24.

Kwok, P.Y. (2002). SNP genotyping with fluorescence polarization detection. *Hum Mutat, 19* (4), 315-323.

Laatikainen, L., & Erkkila, H. (1980). Refractive errors and other ocular findings in school children. *Acta Ophthalmol (Copenh), 58* (1), 129-136.

Laird, N.M., & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet, 7* (5), 385-394.

Lakshmi, N., Demarchi, D.A., Veerraju, P., & Rao, T.V. (2002). Population structure and genetic differentiation among the substructured Vysya caste population in comparison to the other populations of Andhra Pradesh, India. *Ann Hum Biol, 29* (5), 538-549.

Lam, C.S., Goh, W.S., Tang, Y.K., Tsui, K.K., Wong, W.C., & Man, T.C. (1994). Changes in refractive trends and optical components of Hong Kong Chinese aged over 40 years. *Ophthalmic Physiol Opt, 14* (4), 383-388.

Lam, C.S., Goldschmidt, E., & Edwards, M.H. (2004). Prevalence of myopia in local and international schools in Hong Kong. *Optom Vis Sci, 81* (5), 317-322.

Lam, C.Y., Tam, P.O., Fan, D.S., Fan, B.J., Wang, D.Y., Lee, C.W., Pang, C.P., & Lam, D.S. (2008). A genome-wide scan maps a novel high myopia locus to 5p15. *Invest Ophthalmol Vis Sci, 49* (9), 3768-3778.

Lam, D.S., Lee, W.S., Leung, Y.F., Tam, P.O., Fan, D.S., Fan, B.J., & Pang, C.P. (2003a).

TGFbeta-induced factor: a candidate gene for high myopia. *Invest Ophthalmol Vis Sci, 44* (3), 1012-1015.

Lam, D.S., Tam, P.O., Fan, D.S., Baum, L., Leung, Y.F., & Pang, C.P. (2003b). Familial high myopia linkage to chromosome 18p. *Ophthalmologica, 217* (2), 115-118.

Lander, E.S., & Schork, N.J. (1994). Genetic dissection of complex traits. *Science, 265* (5181), 2037-2048.

Lange, K., & Boehnke, M. (1983). Some combinatorial problems of DNA restriction fragment length polymorphisms. *Am J Hum Genet, 35* (2), 177-192.

Lavebratt, C., & Sengul, S. (2006). Single nucleotide polymorphism (SNP) allele frequency estimation in DNA pools using Pyrosequencing. *Nat Protoc, 1* (6), 2573-2582.

Le Hellard, S., Ballereau, S.J., Visscher, P.M., Torrance, H.S., Pinson, J., Morris, S.W., Thomson, M.L., Semple, C.A., Muir, W.J., Blackwood, D.H., Porteous, D.J., & Evans, K.L. (2002). SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res, 30* (15), e74.

Lee, J.H., & Jeon, J.T. (2008). Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels. *Cytogenet Genome Res, 123* (1-4), 333-342.

Lee, K.H., Kwak, Y.D., Kim, D.H., Chang, M.Y., Lee, Y.S., & Lee, Y.S. (2004). Human zinc finger protein 161, a novel transcriptional activator of the dopamine transporter. *Biochem Biophys Res Commun, 313* (4), 969-976.

Lee, P.H., & Shatkay, H. (2008). F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res, 36* (Database issue), D820-824.

Lessells, C.M., & Boag, P.T. (1987). Unrepeatable repeatabilities: a common mistake. *Auk,* 104(1), 116-121

Leung, Y.F., Tam, P.O., Baum, L., Lam, D.S., & Pang, C.C. (2000). TIGR/MYOC proximal promoter GT-repeat polymorphism is not associated with myopia. *Hum Mutat, 16* (6), 533.

Lewallen, S., Lowdon, R., Courtright, P., & Mehl, G.L. (1995). A population-based survey of the prevalence of refractive error in Malawi. *Ophthalmic Epidemiol, 2* (3), 145-149.

Lewis, C.M. (2002). Genetic association studies: design, analysis and interpretation. *Brief Bioinform, 3* (2), 146-153.

Li, C., Li, M., Long, J.R., Cai, Q., & Zheng, W. (2008a). Evaluating cost efficiency of SNP chips in genome-wide association studies. *Genet Epidemiol, 32* (5), 387-395.

Li, J., Zhang, Q.J., Xiao, X.S., Li, J.Z., Zhang, F.S., Li, S.Q., Li, W., Li, T., Jia, X.Y., Guo, L., & Guo, X.M. (2003). [The SNPs analysis of encoding sequence of interacting factor gene in Chinese population]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 20* (5), 454-456.

Li, W., Suh, Y.J., & Yang, Y. (2008b). Exploring case-control genetic association tests using phase diagrams. *Comput Biol Chem, 32* (6), 391-399.

Li, Y.J., Guggenheim, J.A., Bulusu, A., Metlapally, R., Abbott, D., Malecaze, F., Calvas, P., Rosenberg, T., Paget, S., Creer, R.C., Kirov, G., Owen, M.J., Zhao, B., White, T., Mackey, D.A., & Young, T.L. (2009). An international collaborative family-based whole-genome linkage scan for high-grade myopia. *Invest Ophthalmol Vis Sci, 50* (7), 3116-3127.

Liang, C.L., His, E., Chen, K.C., Pan, Y.R., Wang, Y.S., & Juo, S.H. (2011) A functional polymorphism at 3'UTR of the PAX6 gene may confer risk for extreme myopia in the Chinese. *Invest Ophthalmol Vis Sci, 52* (6), 3500-3505.

Liang, C.L., Hung, K.S., Tsai, Y.Y., Chang, W., Wang, H.S., & Juo, S.H. (2007). Systematic assessment of the tagging polymorphisms of the COL1A1 gene for high myopia. *J Hum Genet, 52* (4), 374-377.

Liang, C.L., Wang, H.S., Hung, K.S., Hsi, E., Sun, A., Kuo, Y.H., & Juo, S.H. (2006). Evaluation of MMP3 and TIMP1 as candidate genes for high myopia in young Taiwanese men. *Am J Ophthalmol, 142* (3), 518-520.

Lin, H.J., Kung, Y.J., Lin, Y.J., Sheu, J.J., Chen, B.H., Lan, Y.C., Lai, C.H., Hsu, Y.A., Wan, L., & Tsai, F.J. (2010). Association of the lumican gene functional 3'-UTR polymorphism with high myopia. *Invest Ophthalmol Vis Sci, 51* (1), 96-102.

Lin, H.J., Kung, Y.J., Lin, Y.J., Sheu, J.J., Chen, B.H., Lan, Y.C., Lai, C.H., Shu, Y.A., Wan, L., & Tsai, F.J. (2009). Association of the Lumican gene functional 3' UTR polymorphism with high myopia. *Invest Ophthalmol Vis Sci,*

Lin, H.J., Wan, L., Tsai, Y., Tsai, Y.Y., Fan, S.S., Tsai, C.H., & Tsai, F.J. (2006). The TGFbeta1 gene codon 10 polymorphism contributes to the genetic predisposition to high myopia. *Mol Vis, 12*, 698-703.

Lin, L.L., & Chen, C.J. (1987). Twin study on myopia. *Acta Genet Med Gemellol (Roma), 36* (4), 535-540.

Lindsey, J.D., & Weinreb, R.N. (2005). Elevated intraocular pressure and transgenic applications in the mouse. *J Glaucoma, 14* (4), 318-320.

Little, J., Bradley, L., Bray, M.S., Clyne, M., Dorman, J., Ellsworth, D.L., Hanson, J., Khoury, M., Lau, J., O'Brien, T.R., Rothman, N., Stroup, D., Taioli, E., Thomas, D., Vainio, H., Wacholder, S., & Weinberg, C. (2002). Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol, 156* (4), 300-310.

Liu, K., & Muse, S.V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics, 21* (9), 2128-2129.

Liu, Y., & Vollrath, D. (2004). Reversal of mutant myocilin non-secretion and cell killing: implications for glaucoma. *Hum Mol Genet, 13* (11), 1193-1204.

Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl, 4* (6), 357-362.

Logan, N.S., Davies, L.N., Mallen, E.A., & Gilmartin, B. (2005). Ametropia and ocular biometry in a U.K. university student population. *Optom Vis Sci, 82* (4), 261-266.

Look, M.P., van Putten, W.L., Duffy, M.J., Harbeck, N., Christensen, I.J., Thomssen, C., Kates, R., Spyratos, F., Ferno, M., Eppenberger-Castori, S., Sweep, C.G., Ulm, K., Peyrat, J.P., Martin, P.M., Magdelenat, H., Brunner, N., Duggan, C., Lisboa, B.W., Bendahl, P.O., Quillien, V., Daver, A., Ricolleau, G., Meijer-van Gelder, M.E., Manders, P., Fiets, W.E., Blankenstein, M.A., Broet, P., Romain, S., Daxenbichler, G., Windbichler, G., Cufer, T., Borstnar, S., Kueng, W., Beex, L.V., Klijn, J.G., O'Higgins, N., Eppenberger, U., Janicke, F., Schmitt, M., & Foekens, J.A. (2002). Pooled analysis of prognostic impact of urokinase-type

plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst, 94* (2), 116-128.

Lowe, C.E., Cooper, J.D., Brusko, T., Walker, N.M., Smyth, D.J., Bailey, R., Bourget, K., Plagnol, V., Field, S., Atkinson, M., Clayton, D.G., Wicker, L.S., & Todd, J.A. (2007). Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet, 39* (9), 1074-1082.

Lu, A.T., & Cantor, R.M. (2007). Weighted variance FBAT: a powerful method for including covariates in FBAT analyses. *Genet Epidemiol, 31* (4), 327-337.

Lusa, L., McShane, L.M., Reid, J.F., De Cecco, L., Ambrogi, F., Biganzoli, E., Gariboldi, M., & Pierotti, M.A. (2007). Challenges in projecting clustering results across gene expression-profiling datasets. *J Natl Cancer Inst, 99* (22), 1715-1723.

Lyhne, N., Sjolie, A.K., Kyvik, K.O., & Green, A. (2001). 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, 85* (12), 1470-1476.

Macgregor, S. (2007). Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *Eur J Hum Genet, 15* (4), 501-504.

Macgregor, S., Visscher, P.M., & Montgomery, G. (2006). Analysis of pooled DNA samples on high density arrays without prior knowledge of differential hybridization rates. *Nucleic Acids Res, 34* (7), e55.

Majava, M., Bishop, P.N., Hagg, P., Scott, P.G., Rice, A., Inglehearn, C., Hammond, C.J., Spector, T.D., Ala-Kokko, L., & Mannikko, M. (2007). Novel mutations in the small leucine-rich repeat protein/proteoglycan (SLRP) genes in high myopia. *Hum Mutat, 28* (4), 336-344.

Majewski, J., & Ott, J. (2002). Distribution and characterization of regulatory elements in the human genome. *Genome Res, 12* (12), 1827-1836.

Mak JY, Y.M.K., Fung WY , Ng PW, Lo KK , Yip SP. (2011). Insulin-like growth factor 1 (IGF1) gene haplotypes are associated with high myopia in a Chinese population. *Arch Opthalmol (in press).*

Mallen, E.A., Gammoh, Y., Al-Bdour, M., & Sayegh, F.N. (2005). Refractive error and ocular biometry in Jordanian adults. *Ophthalmic Physiol Opt, 25* (4), 302-309.

Mann, T., & Ward, A. (2004). To eat or not to eat: implications of the attentional myopia model for restrained eaters. *J Abnorm Psychol, 113* (1), 90-98.

Matsumura, H., & Hirai, H. (1999). Prevalence of myopia and refractive changes in students from 3 to 17 years of age. *Surv Ophthalmol, 44 Suppl 1*, S109-115.

Mátyás, G., Giunta, C., Steinmann, B., Hossle, J.P., Hellwig, R. (2002). Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. *Hum Mutat. 9*(1),58-68.

Mavracanas, T.A., Mandalos, A., Peios, D., Golias, V., Megalou, K., Gregoriadou, A., Delidou, K., & Katsougiannopoulos, B. (2000). Prevalence of myopia in a sample of Greek students. *Acta Ophthalmol Scand, 78* (6), 656-659.

McBrien, N.A., Young, T.L., Pang, C.P., Hammond, C., Baird, P., Saw, S.M., Morgan, I.G., Mutti, D.O., Rose, K.A., Wallman, J., Gentle, A., Wildsoet, C.F., Gwiazda, J., Schmid, K.L., Smith, E., 3rd, Troilo, D., Summers-Rada, J., Norton, T.T., Schaeffel, F., Megaw, P., Beuerman, R.W., & McFadden, S.A. (2008). Myopia: Recent Advances in Molecular Studies; Prevalence, Progression and Risk Factors; Emmetropization; Therapies; Optical Links; Peripheral Refraction; Sclera and Ocular Growth; Signalling Cascades; and Animal Models. *Optom Vis Sci, 86* (1), 45-46.

McCarroll, S.A. (2008). Extending genome-wide association studies to copy-number variation. *Hum Mol Genet, 17* (R2), R135-142.

McKeigue, P.M. (1997). Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am J Hum Genet, 60* (1), 188-196.

Meldrum, D. (2000a). Automation for genomics, part one: preparation for sequencing. *Genome Res, 10* (8), 1081-1092.

Meldrum, D. (2000b). Automation for genomics, part two: sequencers, microarrays, and future trends. *Genome Res, 10* (9), 1288-1303.

Menotti-Raymond, M., David, V.A., Lyons, L.A., Schaffer, A.A., Tomlin, J.F., Hutton, M.K., & O'Brien, S.J. (1999). A genetic linkage map of microsatellites in the domestic cat (Felis catus). *Genomics, 57* (1), 9-23.

Mertz, J.R., & Wallman, J. (2000). Choroidal retinoic acid synthesis: a possible mediator between refractive error and compensatory eye growth. *Exp Eye Res, 70* (4), 519-527.

Messmer, D.E. (1992). [Retinal detachment]. *Schweiz Rundsch Med Prax, 81* (19), 622-625.

Metlapally, R., Ki, C.S., Li, Y.J., Tran-Viet, K.N., Abbott, D., Malecaze, F., Calvas, P., Mackey, D.A., Rosenberg, T., Paget, S., Guggenheim, J.A., & Young, T.L. (2010). Genetic association of insulin-like growth factor-1 polymorphisms with high-grade myopia in an international family cohort. *Invest Ophthalmol Vis Sci, 51* (9), 4476-4479.

Metlapally, R., Li, Y.J., Tran-Viet, K.N., Abbott, D., Czaja, G.R., Malecaze, F., Calvas, P., Mackey, D., Rosenberg, T., Paget, S., Zayats, T., Owen, M.J., Guggenheim, J.A., & Young, T.L. (2009a). COL1A1 and COL2A1 genes and myopia susceptibility: evidence of association and suggestive linkage to the COL2A1 locus. *Invest Ophthalmol Vis Sci, 50* (9), 4080-4086.

Metlapally, R., Li, Y.J., Tran-Viet, K.N., Bulusu, A., White, T.R., Ellis, J., Kao, D., & Young, T.L. (2008). Common MFRP sequence variants are not associated with moderate to high hyperopia, isolated microphthalmia, and high myopia. *Mol Vis, 14*, 387-393.

Metlapally, R., Michaelides, M., Bulusu, A., Li, Y.J., Schwartz, M., Rosenberg, T., Hunt, D.M., Moore, A.T., Zuchner, S., Rickman, C.B., & Young, T.L. (2009b). Evaluation of the X-linked high-grade myopia locus (MYP1) with cone dysfunction and color vision deficiencies. *Invest Ophthalmol Vis Sci, 50* (4), 1552-1558.

Mohan, M., Pakrasi, S., & Zutshi, R. (1988). Myopia in India. *Acta Ophthalmol Suppl, 185*, 19-23.

Mohlke, K.L., Erdos, M.R., Scott, L.J., Fingerlin, T.E., Jackson, A.U., Silander, K., Hollstein, P., Boehnke, M., & Collins, F.S. (2002). High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. *Proc Natl Acad Sci U S*

A, 99 (26), 16928-16933.

Mongiat, M., Ligresti, G., Marastoni, S., Lorenzon, E., Doliana, R., & Colombatti, A. (2007). Regulation of the extrinsic apoptotic pathway by the extracellular matrix glycoprotein EMILIN2. *Mol Cell Biol, 27* (20), 7176-7187.

Montana, G., & Pritchard, J.K. (2004). Statistical tests for admixture mapping with case-control and cases-only data. *Am J Hum Genet, 75* (5), 771-789.

Moore, J.H., Asselbergs, F.W., & Williams, S.M. (2010). Bioinformatics challenges for genome-wide association studies. *Bioinformatics, 26* (4), 445-455.

Morgan, I., & Megaw, P. (2004). Using natural STOP growth signals to prevent excessive axial elongation and the development of myopia. *Ann Acad Med Singapore, 33* (1), 16-20.

Morgan, I., & Rose, K. (2005). How genetic is school myopia? *Prog Retin Eye Res, 24* (1), 1-38.

Morgan, I.G. (2003). The biological basis of myopic refractive error. *Clin Exp Optom, 86* (5), 276-288.

Morita, H., Funata, M., & Tokoro, T. (1995). A clinical study of the development of posterior vitreous detachment in high myopia. *Retina, 15* (2), 117-124.

Moroldo, M.B., Donnelly, P., Saunders, J., Glass, D.N., & Giannini, E.H. (1998). Transmission disequilibrium as a test of linkage and association between HLA alleles and pauciarticular-onset juvenile rheumatoid arthritis. *Arthritis Rheum, 41* (9), 1620-1624.

Moskvina, V., Norton, N., Williams, N., Holmans, P., Owen, M., & O'Donovan, M. (2005). Streamlined analysis of pooled genotype data in SNP-based association studies. *Genet Epidemiol, 28* (3), 273-282.

Motsinger, A.A., & Ritchie, M.D. (2006). Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Hum Genomics, 2* (5), 318-328.

Mutti, D.O., Cooper, M.E., O'Brien, S., Jones, L.A., Marazita, M.L., Murray, J.C., & Zadnik, K. (2007). Candidate gene and locus analysis of myopia. *Mol Vis, 13*, 1012-1019.

Mutti, D.O., Mitchell, G.L., Moeschberger, M.L., Jones, L.A., & Zadnik, K. (2002). Parental myopia, near work, school achievement, and children's refractive error. *Invest Ophthalmol Vis Sci, 43* (12), 3633-3640.

Naiglin, L., Clayton, J., Gazagne, C., Dallongeville, F., Malecaze, F., & Calvas, P. (1999). Familial high myopia: evidence of an autosomal dominant mode of inheritance and genetic heterogeneity. *Ann Genet, 42* (3), 140-146.

Nakanishi, H., Yamada, R., Gotoh, N., Hayashi, H., Otani, A., Tsujikawa, A., Yamashiro, K., Shimada, N., Ohno-Matsui, K., Mochizuki, M., Saito, M., Saito, K., Iida, T., Matsuda, F., & Yoshimura, N. (2009a). Absence of association between COL1A1 polymorphisms and high myopia in the Japanese population. *Invest Ophthalmol Vis Sci, 50* (2), 544-550.

Nakanishi, H., Yamada, R., Gotoh, N., Hayashi, H., Yamashiro, K., Shimada, N., Ohno-Matsui, K., Mochizuki, M., Saito, M., Iida, T., Matsuo, K., Tajima, K., Yoshimura, N., & Matsuda, F. (2009b). A genome-wide association analysis identified a novel susceptible

locus for pathological myopia at 11q24.1. *PLoS Genet, 5* (9), e1000660.

Nallasamy, S., Paluru, P.C., Devoto, M., Wasserman, N.F., Zhou, J., & Young, T.L. (2007). Genetic linkage study of high-grade myopia in a Hutterite population from South Dakota. *Mol Vis, 13*, 229-236.

Newton-Cheh, C., & Hirschhorn, J.N. (2005). Genetic association studies of complex traits: design and analysis issues. *Mutat Res, 573* (1-2), 54-69.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., & Markham, A.F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res, 17* (7), 2503-2516.

Ng, T.K., Lam, C.Y., Lam, D.S., Chiang, S.W., Tam, P.O., Wang, D.Y., Fan, B.J., Yam, G.H., Fan, D.S., & Pang, C.P. (2009). AC and AG dinucleotide repeats in the PAX6 P1 promoter are associated with high myopia. *Mol Vis, 15*, 2239-2248.

Nguyen, T.D., Chen, P., Huang, W.D., Chen, H., Johnson, D., & Polansky, J.R. (1998). Gene structure and properties of TIGR, an olfactomedin-related glycoprotein cloned from glucocorticoid-induced trabecular meshwork cells. *J Biol Chem, 273* (11), 6341-6350.

Nicodemus, K.K., Luna, A., & Shugart, Y.Y. (2007). An evaluation of power and type I error of single-nucleotide polymorphism transmission/disequilibrium-based statistical methods under different family structures, missing parental data, and population stratification. *Am J Hum Genet, 80* (1), 178-185.

Nijbroek, G., Sood, S., McIntosh, I., Francomano, C.A., Bull, E., Pereira, L., Ramirez, F., Pyeritz, R.E., & Dietz, H.C. (1995). Fifteen novel FBN1 mutations causing Marfan syndrome detected by heteroduplex analysis of genomic amplicons. *Am J Hum Genet, 57* (1), 8-21.

Nishina, S., Kohsaka, S., Yamaguchi, Y., Handa, H., Kawakami, A., Fujisawa, H., & Azuma, N. (1999). PAX6 expression in the developing human eye. *Br J Ophthalmol, 83* (6), 723-727.

Nishizaki, R., Ota, M., Inoko, H., Meguro, A., Shiota, T., Okada, E., Mok, J., Oka, A., Ohno, S., & Mizuki, N. (2009). New susceptibility locus for high myopia is linked to the uromodulin-like 1 (UMODL1) gene region on chromosome 21q22.3. *Eye (Lond), 23* (1), 222-229.

Noble, K.G., & Carr, R.E. (1982). Pathologic myopia. *Ophthalmology, 89* (9), 1099-1100.

Nomura, H., Ando, F., Niino, N., Shimokata, H., & Miyake, Y. (2004). The relationship between intraocular pressure and refractive error adjusting for age and central corneal thickness. *Ophthalmic Physiol Opt, 24* (1), 41-45.

Norton, N., Williams, N.M., O'Donovan, M.C., & Owen, M.J. (2004). DNA pooling as a tool for large-scale association studies in complex traits. *Ann Med, 36* (2), 146-152.

Norton, N., Williams, N.M., Williams, H.J., Spurlock, G., Kirov, G., Morris, D.W., Hoogendoorn, B., Owen, M.J., & O'Donovan, M.C. (2002). Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet, 110* (5), 471-478.

Norton, T.T. (1990). Experimental myopia in tree shrews. *Ciba Found Symp, 155*, 178-194;

discussion 194-179.

Norton, T.T. (1999). Animal Models of Myopia: Learning How Vision Controls the Size of the Eye. *Ilar J, 40* (2), 59-77.

Norton, T.T., & Rada, J.A. (1995). Reduced extracellular matrix in mammalian sclera with induced myopia. *Vision Res, 35* (9), 1271-1281.

Nurnberg, G., Jacobi, F.K., Broghammer, M., Becker, C., Blin, N., Nurnberg, P., Stephani, U., & Pusch, C.M. (2008). Refinement of the MYP3 locus on human chromosome 12 in a German family with Mendelian autosomal dominant high-grade myopia by SNP array mapping. *Int J Mol Med, 21* (4), 429-438.

O'Donovan, M.C., Oefner, P.J., Roberts, S.C., Austin, J., Hoogendoorn, B., Guy, C., Speight, G., Upadhyaya, M., Sommer, S.S., & McGuffin, P. (1998). Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics, 52* (1), 44-49.

O'Leary, D.J., & Millodot, M. (1979). Eyelid closure causes myopia in humans. *Experientia, 35* (11), 1478-1479.

Ohashi, J. (2009). A practical case-control association test for detecting a susceptibility allele at a copy number variation locus. *J Hum Genet, 54* (3), 169-173.

Ohashi, J., & Tokunaga, K. (2001). The power of genome-wide association studies of complex disease genes: statistical limitations of indirect approaches using SNP markers. *J Hum Genet, 46* (8), 478-482.

Ojaimi, E., Rose, K.A., Smith, W., Morgan, I.G., Martin, F.J., & Mitchell, P. (2005). Methods for a population-based study of myopia and other eye conditions in school children: the Sydney Myopia Study. *Ophthalmic Epidemiol, 12* (1), 59-69.

Olivier, M. (2005). The Invader assay for SNP genotyping. *Mutat Res, 573* (1-2), 103-110.

Orita, M., Suzuki, Y., Sekiya, T., & Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics, 5* (4), 874-879.

Overhauser, J., Mitchell, H.F., Zackai, E.H., Tick, D.B., Rojas, K., & Muenke, M. (1995). Physical mapping of the holoprosencephaly critical region in 18p11.3. *Am J Hum Genet, 57* (5), 1080-1085.

Owens, D.A., & Leibowitz, H.W. (1976). Night myopia: cause and a possible basis for amelioration. *Am J Optom Physiol Opt, 53* (11), 709-717.

Pacella, R., McLellan, J., Grice, K., Del Bono, E.A., Wiggs, J.L., & Gwiazda, J.E. (1999). Role of genetic factors in the etiology of juvenile-onset myopia based on a longitudinal study of refractive error. *Optom Vis Sci, 76* (6), 381-386.

Paget, S., Vitezica, Z.G., Malecaze, F., & Calvas, P. (2008). Heritability of refractive value and ocular biometrics. *Exp Eye Res, 86* (2), 290-295.

Paluru, P.C., Scavello, G.S., Ganter, W.R., & Young, T.L. (2004). Exclusion of lumican and fibromodulin as candidate genes in MYP3 linked high grade myopia. *Mol Vis, 10*, 917-922.

Pararajasegaram, R. (1999). VISION 2020-the right to sight: from strategies to action. *Am J*

Ophthalmol, 128 (3), 359-360.

Park, C.H., Kim, Y.S., Lee, H.K., Kim, Y.H., Choi, M.Y., Jung, D.E., Yoo, J.M., Kang, S.S., Choi, W.S., & Cho, G.J. (2007). Citicoline reduces upregulated clusterin following kainic acid injection in the rat retina. *Curr Eye Res, 32* (12), 1055-1063.

Pastinen, T., Partanen, J., & Syvanen, A.C. (1996). Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. *Clin Chem, 42* (9), 1391-1397.

Pastino, G.M., Flynn, E.J., & Sultatos, L.G. (2000). Genetic polymorphisms in ethanol metabolism: issues and goals for physiologically based pharmacokinetic modeling. *Drug Chem Toxicol, 23* (1), 179-201.

Patten, B., & Howland, H.C. (1996). Clinical findings before the onset of myopia in youth. I. Ocular optical components. *Surv Ophthalmol, 41* (2), 184-185.

Pawitan, Y., Seng, K.C., & Magnusson, P.K. (2009). How many genetic variants remain to be discovered? *PLoS One, 4* (12), e7969.

Pearson, J.V., Huentelman, M.J., Halperin, R.F., Tembe, W.D., Melquist, S., Homer, N., Brun, M., Szelinger, S., Coon, K.D., Zismann, V.L., Webster, J.A., Beach, T., Sando, S.B., Aasly, J.O., Heun, R., Jessen, F., Kolsch, H., Tsolaki, M., Daniilidou, M., Reiman, E.M., Papassotiropoulos, A., Hutton, M.L., Stephan, D.A., & Craig, D.W. (2007). Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *Am J Hum Genet, 80* (1), 126-139.

Pechmann, A., & Czepita, D. (2000). [Epidemiology of myopia]. *Klin Oczna, 102* (4), 297-300.

Perkins, E.S. (1960). Glaucoma in the younger age groups. *Arch Ophthalmol, 64*, 882-891.

Perkins, E.S. (1979). Morbidity from myopia. *Sight Sav Rev, 49* (1), 11-19.

Pertile, K.K., Schache, M., Islam, F.M., Chen, C.Y., Dirani, M., Mitchell, P., & Baird, P.N. (2008). Assessment of TGIF as a candidate gene for myopia. *Invest Ophthalmol Vis Sci, 49* (1), 49-54.

Peterfy, M., Phan, J., Xu, P., & Reue, K. (2001). Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat Genet, 27* (1), 121-124.

Peters, S.P. (2009). Reporting and evaluating genetic association studies. *Respir Res, 10*, 109.

Pfeifer, G.P., & Hainaut, P. (2011). Next-generation sequencing: emerging lessons on the origins of human cancer. *Curr Opin Oncol, 23* (1), 62-68.

Pielberg, G., Day, A.E., Plastow, G.S., & Andersson, L. (2003). A sensitive method for detecting variation in copy numbers of duplicated genes. *Genome Res, 13* (9), 2171-2177.

Pierro, L., Camesasca, F.I., Mischi, M., & Brancato, R. (1992). Peripheral retinal changes and axial myopia. *Retina, 12* (1), 12-17.

Pirinen, M., Kulathinal, S., Gasbarra, D., & Sillanpaa, M.J. (2008). Estimating population haplotype frequencies from pooled DNA samples using PHASE algorithm. *Genet Res, 90* (6), 509-524.

Plagnol, V. (2009). Association tests and software for copy number variant data. *Hum Genomics, 3* (2), 191-194.

Podos, S.M., Becker, B., & Morton, W.R. (1966). High myopia and primary open-angle glaucoma. *Am J Ophthalmol, 62* (6), 1038-1043.

Pollak, E.S., Feng, L., Ahadian, H., & Fortina, P. (2001). Microarray-based genetic analyses for studying susceptibility to arterial and venous thrombotic disorders. *Ital Heart J, 2* (8), 568-572.

Polychronakos, C. (2008). Common and rare alleles as causes of complex phenotypes. *Curr Atheroscler Rep, 10* (3), 194-200.

Pousi, B., Hautala, T., Heikkinen, J., Pajunen, L., Kivirikko, K.I., & Myllyla, R. (1994). Alu-Alu recombination results in a duplication of seven exons in the lysyl hydroxylase gene in a patient with the type VI variant of Ehlers-Danlos syndrome. *Am J Hum Genet, 55* (5), 899-906.

Pritchard, J.K., & Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *Am J Hum Genet, 69* (1), 1-14.

Pritchard, J.K., & Rosenberg, N.A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet, 65* (1), 220-228.

Pruett, R.C. (1995). Socioeconomic aspects of laser refractive surgery. *Arch Ophthalmol, 113* (9), 1093.

Purcell, S., Daly, M.J., & Sham, P.C. (2007a). WHAP: haplotype-based association analysis. *Bioinformatics, 23* (2), 255-256.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., & Sham, P.C. (2007b). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81* (3), 559-575.

Qian, S.S., & Shen, Z. (2007). Ecological applications of multilevel analysis of variance. *Ecology, 88* (10), 2489-2495.

Qu, H.Q., Montpetit, A., Ge, B., Hudson, T.J., & Polychronakos, C. (2007). Toward further mapping of the association between the IL2RA locus and type 1 diabetes. *Diabetes, 56* (4), 1174-1176.

Quade, S.R., Elston, R.C., & Goddard, K.A. (2005). Estimating haplotype frequencies in pooled DNA samples when there is genotyping error. *BMC Genet, 6*, 25.

Quinn, G.E., Shin, C.H., Maguire, M.G., & Stone, R.A. (1999). Myopia and ambient lighting at night. *Nature, 399* (6732), 113-114.

Rabb, M.F., Garoon, I., & LaFranco, F.P. (1981). Myopic macular degeneration. *Int Ophthalmol Clin, 21* (3), 51-69.

Rabinowitz, D., & Laird, N. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered, 50* (4), 211-223.

Rada, J.A., Achen, V.R., Penugonda, S., Schmidt, R.W., & Mount, B.A. (2000).

Proteoglycan composition in the human sclera during growth and aging. *Invest Ophthalmol Vis Sci, 41* (7), 1639-1648.

Rada, J.A., Shelton, S., & Norton, T.T. (2006). The sclera and myopia. *Exp Eye Res, 82* (2), 185-200.

Ranade, K., Chang, M.S., Ting, C.T., Pei, D., Hsiao, C.F., Olivier, M., Pesich, R., Hebert, J., Chen, Y.D., Dzau, V.J., Curb, D., Olshen, R., Risch, N., Cox, D.R., & Botstein, D. (2001). High-throughput genotyping with single nucleotide polymorphisms. *Genome Res, 11* (7), 1262-1268.

Rannala, B. (2001). Finding genes influencing susceptibility to complex diseases in the post-genome era. *Am J Pharmacogenomics, 1* (3), 203-221.

Rebbeck, T.R. (1999). Inherited genetic predisposition in breast cancer. A population-based perspective. *Cancer, 86* (11 Suppl), 2493-2501.

Reddy, K.B., Fox, J.E., Price, M.G., Kulkarni, S., Gupta, S., Das, B., & Smith, D.M. (2008). Nuclear localization of Myomesin-1: possible functions. *J Muscle Res Cell Motil, 29* (1), 1-8.

Redowicz, M.J. (2002). Myosins and pathology: genetics and biology. *Acta Biochim Pol, 49* (4), 789-804.

Reue, K. (2009). The lipin family: mutations and metabolism. *Curr Opin Lipidol, 20* (3), 165-170.

Reynisson, E., Josefsen, M.H., Krause, M., & Hoorfar, J. (2006). Evaluation of probe chemistries and platforms to improve the detection limit of real-time PCR. *J Microbiol Methods, 66* (2), 206-216.

Richards, A.J., Yates, J.R., Williams, R., Payne, S.J., Pope, F.M., Scott, J.D., & Snead, M.P. (1996). A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in alpha 1 (XI) collagen. *Hum Mol Genet, 5* (9), 1339-1343.

Rieger, R.H., Kaplan, N.L., & Weinberg, C.R. (2001). Efficient use of siblings in testing for linkage and association. *Genet Epidemiol, 20* (2), 175-191.

Risch, N. (1990). Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol, 7* (1), 3-16; discussion 17-45.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273* (5281), 1516-1517.

Risch, N., & Teng, J. (1998). The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res, 8* (12), 1273-1288.

Risch, N.J. (2000). Searching for genetic determinants in the new millennium. *Nature, 405* (6788), 847-856.

Ritchie, M.D. (2005). Bioinformatics approaches for detecting gene-gene and gene-environment interactions in studies of human disease. *Neurosurg Focus, 19* (4), E2.

Rohrer, B., & Stell, W.K. (1994). Basic fibroblast growth factor (bFGF) and transforming

growth factor beta (TGF-beta) act as stop and go signals to modulate postnatal ocular growth in the chick. *Exp Eye Res, 58* (5), 553-561.

Rose, K.A., Morgan, I.G., Ip, J., Kifley, A., Huynh, S., Smith, W., & Mitchell, P. (2008). Outdoor activity reduces the prevalence of myopia in children. *Ophthalmology, 115* (8), 1279-1285.

Rosner, J., Aker, J.S., & Rosner, J. (1988). Similarities and differences in a set of 7-year-old triplets, two of whom are monozygotic. *Am J Optom Physiol Opt, 65* (11), 890-892.

Ross, P., Hall, L., Smirnov, I., & Haff, L. (1998). High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol, 16* (13), 1347-1351.

Roux, K.H., & Hecker, K.H. (1997). One-step optimization using touchdown and stepdown PCR. *Methods Mol Biol, 67*, 39-45.

Rudan, I., Campbell, H., & Rudan, P. (1999). Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies. *Coll Antropol, 23* (2), 531-546.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., & Altshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409* (6822), 928-933.

Sale, M.M., Mychaleckyj, J.C., & Chen, W.M. (2009). Planning and executing a genome wide association study (GWAS). *Methods Mol Biol, 590*, 403-418.

Sarkar, G., Yoon, H.S., & Sommer, S.S. (1992). Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. *Genomics, 13* (2), 441-443.

Sasaki, S., Ota, M., Meguro, A., Nishizaki, R., Okada, E., Mok, J., Kimura, T., Oka, A., Katsuyama, Y., Ohno, S., Inoko, H., & Mizuki, N. (2007). A single nucleotide polymorphism analysis of the LAMA1 gene in Japanese patients with high myopia. *Clin Ophthalmol, 1* (3), 289-295.

Satpathy, M., Gallagher, P., Lizotte-Waniewski, M., & Srinivas, S.P. (2004). Thrombin-induced phosphorylation of the regulatory light chain of myosin II in cultured bovine corneal endothelial cells. *Exp Eye Res, 79* (4), 477-486.

Saw, S.M., Gazzard, G., Au Eong, K.G., & Tan, D.T. (2002a). Myopia: attempts to arrest progression. *Br J Ophthalmol, 86* (11), 1306-1311.

Saw, S.M., Gazzard, G., Shih-Yen, E.C., & Chua, W.H. (2005a). Myopia and associated pathological complications. *Ophthalmic Physiol Opt, 25* (5), 381-391.

Saw, S.M., Katz, J., Schein, O.D., Chew, S.J., & Chan, T.K. (1996). Epidemiology of myopia. *Epidemiol Rev, 18* (2), 175-187.

Saw, S.M., Tong, L., Chua, W.H., Chia, K.S., Koh, D., Tan, D.T., & Katz, J. (2005b). Incidence and progression of myopia in Singaporean school children. *Invest Ophthalmol Vis Sci, 46* (1), 51-57.

Saw, S.M., Wu, H.M., Seet, B., Wong, T.Y., Yap, E., Chia, K.S., Stone, R.A., & Lee, L. (2001). Academic achievement, close up work parameters, and myopia in Singapore military conscripts. *Br J Ophthalmol, 85* (7), 855-860.

Saw, S.M., Zhang, M.Z., Hong, R.Z., Fu, Z.F., Pang, M.H., & Tan, D.T. (2002b). Near-work activity, night-lights, and myopia in the Singapore-China study. *Arch Ophthalmol, 120* (5), 620-627.

Sawcer, S., Maranian, M., Setakis, E., Curwen, V., Akesson, E., Hensiek, A., Coraddu, F., Roxburgh, R., Sawcer, D., Gray, J., Deans, J., Goodfellow, P.N., Walker, N., Clayton, D., & Compston, A. (2002). A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. *Brain, 125* (Pt 6), 1337-1347.

Scavello, G.S., Jr., Paluru, P.C., Zhou, J., White, P.S., Rappaport, E.F., & Young, T.L. (2005). Genomic structure and organization of the high grade Myopia-2 locus (MYP2) critical region: mutation screening of 9 positional candidate genes. *Mol Vis, 11*, 97-110.

Scavello, G.S., Paluru, P.C., Ganter, W.R., & Young, T.L. (2004). Sequence variants in the transforming growth beta-induced factor (TGIF) gene are not associated with high myopia. *Invest Ophthalmol Vis Sci, 45* (7), 2091-2097.

Schache, M., Richardson, A.J., Pertile, K.K., Dirani, M., Scurrah, K., & Baird, P.N. (2007). Genetic mapping of myopia susceptibility loci. *Invest Ophthalmol Vis Sci, 48* (11), 4924-4929.

Schaschl, H., Aitman, T.J., & Vyse, T.J. (2009). Copy number variation in the human genome and its implication in autoimmunity. *Clin Exp Immunol, 156* (1), 12-16.

Schippert, R., Burkhardt, E., Feldkaemper, M., & Schaeffel, F. (2007). Relative axial myopia in Egr-1 (ZENK) knockout mice. *Invest Ophthalmol Vis Sci, 48* (1), 11-17.

Schippert, R., Schaeffel, F., & Feldkaemper, M.P. (2008). Microarray analysis of retinal gene expression in chicks during imposed myopic defocus. *Mol Vis, 14*, 1589-1599.

Schliekelman, P. (2008). Statistical power of expression quantitative trait loci for mapping of complex trait loci in natural populations. *Genetics, 178* (4), 2201-2216.

Schor, C. (1999). The influence of interactions between accommodation and convergence on the lag of accommodation. *Ophthalmic Physiol Opt, 19* (2), 134-150.

Schork, N.J., Murray, S.S., Frazer, K.A., & Topol, E.J. (2009). Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev, 19* (3), 212-219.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. *Nat Methods, 5* (1), 16-18.

Schwartz, M., Haim, M., & Skarsholm, D. (1990). X-linked myopia: Bornholm eye disease. Linkage to DNA markers on the distal part of Xq. *Clin Genet, 38* (4), 281-286.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimaki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., & Wigler, M. (2007). Strong association of de novo copy

number mutations with autism. *Science, 316* (5823), 445-449.

Sellers, T.A. (2004). Genetic ancestry and molecular epidemiology. *Cancer Epidemiol Biomarkers Prev, 13* (4), 499-500.

Sengler, C., Lau, S., Wahn, U., & Nickel, R. (2002). Interactions between genes and environmental factors in asthma and atopy: new developments. *Respir Res, 3*, 7.

Seo, B.Y., Park, E.W., Ahn, S.J., Lee, S.H., Kim, J.H., Im, H.T., Lee, J.H., Cho, I.C., Kong, I.K., & Jeon, J.T. (2007). An accurate method for quantifying and analyzing copy number variation in porcine KIT by an oligonucleotide ligation assay. *BMC Genet, 8*, 81.

Sertie, A.L., Sossi, V., Camargo, A.A., Zatz, M., Brahe, C., & Passos-Bueno, M.R. (2000). Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum Mol Genet, 9* (13), 2051-2058.

Sham, P., Bader, J.S., Craig, I., O'Donovan, M., & Owen, M. (2002). DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet, 3* (11), 862-871.

Shaw, S.H., Carrasquillo, M.M., Kashuk, C., Puffenberger, E.G., & Chakravarti, A. (1998). Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome Res, 8* (2), 111-123.

Shi, M.M., Bleavins, M.R., & de la iglesia, F.A. (1999). Technologies for detecting genetic polymorphisms in pharmacogenomics. *Mol Diagn, 4* (4), 343-351.

Shifman, S., Pisante-Shalom, A., Yakir, B., & Darvasi, A. (2002). Quantitative technologies for allele frequency estimation of SNPs in DNA pools. *Mol Cell Probes, 16* (6), 429-434.

Siegwart, J.T., Jr., & Norton, T.T. (1999). Regulation of the mechanical properties of tree shrew sclera by the visual environment. *Vision Res, 39* (2), 387-407.

Siegwart, J.T., Jr., & Norton, T.T. (2002). The time course of changes in mRNA levels in tree shrew sclera during induced myopia and recovery. *Invest Ophthalmol Vis Sci, 43* (7), 2067-2075.

Simpson, C.L., Hysi, P., Bhattacharya, S.S., Hammond, C.J., Webster, A., Peckham, C.S., Sham, P.C., & Rahi, J.S. (2007). The Roles of PAX6 and SOX2 in Myopia: lessons from the 1958 British Birth Cohort. *Invest Ophthalmol Vis Sci, 48* (10), 4421-4425.

Sohocki, M.M., Malone, K.A., Sullivan, L.S., & Daiger, S.P. (1999). Localization of retina/pineal-expressed sequences: identification of novel candidate genes for inherited retinal disorders. *Genomics, 58* (1), 29-33.

Solomon, K.D., Fernandez de Castro, L.E., Sandoval, H.P., Biber, J.M., Groat, B., Neff, K.D., Ying, M.S., French, J.W., Donnenfeld, E.D., & Lindstrom, R.L. (2009). LASIK world literature review: quality of life and patient satisfaction. *Ophthalmology, 116* (4), 691-701.

Solouki, A.M., Verhoeven, V.J., van Duijn, C.M., Verkerk, A.J., Ikram, M.K., Hysi, P.G., Despriet, D.D., van Koolwijk, L.M., Ho, L., Ramdas, W.D., Czudowska, M., Kuijpers, R.W., Amin, N., Struchalin, M., Aulchenko, Y.S., van Rij, G., Riemslag, F.C., Young, T.L., Mackey, D.A., Spector, T.D., Gorgels, T.G., Willemse-Assink, J.J., Isaacs, A., Kramer, R., Swagemakers, S.M., Bergen, A.A., van Oosterhout, A.A., Oostra, B.A., Rivadeneira, F., Uitterlinden, A.G., Hofman, A., de Jong, P.T., Hammond, C.J., Vingerling, J.R., & Klaver,

C.C. (2010). A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14. *Nat Genet, 42* (10), 897-901.

Sorsby, A., & Fraser, G.R. (1964). Statistical Note on the Components of Ocular Refraction in Twins. *J Med Genet, 55*, 47-49.

Sorsby, A., Leary, G.A., & Fraser, G.R. (1966). Family studies on ocular refraction and its components. *J Med Genet, 3* (4), 269-273.

Sperduto, R.D., & Hiller, R. (1984). The prevalence of nuclear, cortical, and posterior subcapsular lens opacities in a general population sample. *Ophthalmology, 91* (7), 815-818.

Sperduto, R.D., Seigel, D., Roberts, J., & Rowland, M. (1983). Prevalence of myopia in the United States. *Arch Ophthalmol, 101* (3), 405-407.

Stambolian, D., Ibay, G., Reider, L., Dana, D., Moy, C., Schlifka, M., Holmes, T., Ciner, E., & Bailey-Wilson, J.E. (2004). Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. *Am J Hum Genet, 75* (3), 448-459.

Stambolian, D., Ibay, G., Reider, L., Dana, D., Moy, C., Schlifka, M., Holmes, T.N., Ciner, E., & Bailey-Wilson, J.E. (2006). Genome-wide scan of additional Jewish families confirms linkage of a myopia susceptibility locus to chromosome 22q12. *Mol Vis, 12*, 1499-1505.

Stone, E.M., Fingert, J.H., Alward, W.L., Nguyen, T.D., Polansky, J.R., Sunden, S.L., Nishimura, D., Clark, A.F., Nystuen, A., Nichols, B.E., Mackey, D.A., Ritch, R., Kalenak, J.W., Craven, E.R., & Sheffield, V.C. (1997). Identification of a gene that causes primary open angle glaucoma. *Science, 275* (5300), 668-670.

Stone, R.A., Lin, T., Laties, A.M., Iuvone, P.M. (1989). Retinal dopamine and form-deprivation myopia. *Proc Natl Acad Sci U S A, 86*(2):704-706.

Storm, N., Darnhofer-Patel, B., van den Boom, D., & Rodi, C.P. (2003). MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol Biol, 212*, 241-262.

Sturgill, G.M., Pauer, G.J., Bala, E., Simpson, E., Yaniglos, S.S., Crabb, J.W., Hollyfield, J.G., Lewis, H., Peachey, N.S., & Hagstrom, S.A. (2006). Mutation screen of the cone-specific gene, CLUL1, in 376 patients with age-related macular degeneration. *Ophthalmic Genet, 27* (4), 151-155.

Summanen, P., Kivitie-Kallio, S., Norio, R., Raitta, C., & Kivela, T. (2002). Mechanisms of myopia in Cohen syndrome mapped to chromosome 8q22. *Invest Ophthalmol Vis Sci, 43* (5), 1686-1693.

Sundin, O.H., Leppert, G.S., Silva, E.D., Yang, J.M., Dharmaraj, S., Maumenee, I.H., Santos, L.C., Parsa, C.F., Traboulsi, E.I., Broman, K.W., Dibernardo, C., Sunness, J.S., Toy, J., & Weinberg, E.M. (2005). Extreme hyperopia is the result of null mutations in MFRP, which encodes a Frizzled-related protein. *Proc Natl Acad Sci U S A, 102* (27), 9553-9558.

Syvanen, A.C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet, 2* (12), 930-942.

Tamm, E.R., Russell, P., Epstein, D.L., Johnson, D.H., & Piatigorsky, J. (1999). Modulation of myocilin/TIGR expression in human trabecular meshwork. *Invest Ophthalmol Vis Sci, 40* (11), 2577-2582.

Tang, K., Fu, D.J., Julien, D., Braun, A., Cantor, C.R., & Koster, H. (1999). Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci U S A, 96* (18), 10016-10020.

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

Tang, W.C., Yip, S.P., Lo, K.K., Ng, P.W., Choi, P.S., Lee, S.Y., & Yap, M.K. (2007). Linkage and association of myocilin (MYOC) polymorphisms with high myopia in a Chinese population. *Mol Vis, 13*, 534-544.

Tang, Z., Wang, K., Tan, W., Ma, C., Li, J., Liu, L., Guo, Q., & Meng, X. (2005). Real-time investigation of nucleic acids phosphorylation process using molecular beacons. *Nucleic Acids Res, 33* (11), e97.

Taylor, S.D., Otlowski, M.F., Barlow-Stewart, K.K., Treloar, S.A., Stranger, M., & Chenoweth, K. (2004). Investigating genetic discrimination in Australia: opportunities and challenges in the early stages. *New Genet Soc, 23* (2), 225-239.

Teikari, J.M., O'Donnell, J., Kaprio, J., & Koskenvuo, M. (1991). Impact of heredity in myopia. *Hum Hered, 41* (3), 151-156.

Thomas, D.C., Casey, G., Conti, D.V., Haile, R.W., Lewinger, J.P., & Stram, D.O. (2009). Methodological Issues in Multistage Genome-wide Association Studies. *Stat Sci, 24* (4), 414-429.

Thorn, F., Cruz, A.A., Machado, A.J., & Carvalho, R.A. (2005). Refractive status of indigenous people in the northwestern Amazon region of Brazil. *Optom Vis Sci, 82* (4), 267-272.

Toh, T., Liew, S.H., MacKinnon, J.R., Hewitt, A.W., Poulsen, J.L., Spector, T.D., Gilbert, C.E., Craig, J.E., Hammond, C.J., & Mackey, D.A. (2005). Central corneal thickness is highly heritable: the twin eye studies. *Invest Ophthalmol Vis Sci, 46* (10), 3718-3722.

Tomarev, S.I., Tamm, E.R., & Chang, B. (1998). Characterization of the mouse Myoc/Tigr gene. *Biochem Biophys Res Commun, 245* (3), 887-893.

Tripathi, R.C., Parapuram, S.K., Tripathi, B.J., Zhong, Y., & Chalam, K.V. (1999). Corticosteroids and glaucoma risk. *Drugs Aging, 15* (6), 439-450.

Tsai, Y.Y., Chiang, C.C., Lin, H.J., Lin, J.M., Wan, L., & Tsai, F.J. (2008). A PAX6 gene polymorphism is associated with genetic predisposition to extreme myopia. *Eye (Lond), 22* (4), 576-581.

van den Oord, E.J., Jiang, Y., Riley, B.P., Kendler, K.S., & Chen, X. (2003). FP-TDI SNP scoring by manual and statistical procedures: a study of error rates and types. *Biotechniques, 34* (3), 610-616, 618-620, 622 passim.

Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L., Wu, G., Birney, E., & Stein, L. (2007). Reactome: a knowledge base of biologic pathways and processes. *Genome Biol, 8* (3), R39.

Vatavuk, Z., Skunca Herman, J., Bencic, G., Andrijevic Derk, B., Lacmanovic Loncar, V., Petric Vickovic, I., Bucan, K., Mandic, K., Mandic, A., Skegro, I., Pavicic Astalos, J., Merc, I., Martinovic, M., Kralj, P., Knezevic, T., Barac-Juretic, K., & Zgaga, L. (2009). Common variant in myocilin gene is associated with high myopia in isolated population of Korcula

Island, Croatia. *Croat Med J, 50* (1), 17-22.

Veerappan, S., Pertile, K.K., Islam, A.F., Schache, M., Chen, C.Y., Mitchell, P., Dirani, M., & Baird, P.N. Role of the hepatocyte growth factor gene in refractive error. *Ophthalmology, 117* (2), 239-245 e231-232.

Vella, A., Cooper, J.D., Lowe, C.E., Walker, N., Nutland, S., Widmer, B., Jones, R., Ring, S.M., McArdle, W., Pembrey, M.E., Strachan, D.P., Dunger, D.B., Twells, R.C., Clayton, D.G., & Todd, J.A. (2005). Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *Am J Hum Genet, 76* (5), 773-779.

Vessey, K.A., Rushforth, D.A., & Stell, W.K. (2005). Glucagon- and secretin-related peptides differentially alter ocular growth and the development of form-deprivation myopia in chicks. *Invest Ophthalmol Vis Sci, 46* (11), 3932-3942.

Via, M., Gignoux, C., & Burchard, E.G. (2010) The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med, 2* (1), 3.

Vidic, B., & Lerchner, H. (1990). [Long-term biometric results in glaucoma in children]. *Fortschr Ophthalmol, 87* (1), 25-27.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol, 34* (3), 275-305.

Visscher, P.M., & Le Hellard, S. (2003). Simple method to analyze SNP-based association studies using DNA pools. *Genet Epidemiol, 24* (4), 291-296.

von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B., & Bork, P. (2007). STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res, 35* (Database issue), D358-362.

von Noorden, G.K., & Lewis, R.A. (1987). Ocular axial length in unilateral congenital cataracts and blepharoptosis. *Invest Ophthalmol Vis Sci, 28* (4), 750-752.

Vongphanit, J., Mitchell, P., & Wang, J.J. (2002). Prevalence and progression of myopic retinopathy in an older population. *Ophthalmology, 109* (4), 704-711.

Wacholder, S., Rothman, N., & Caporaso, N. (2002). Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prev, 11* (6), 513-520.

Wallman, J., Gottlieb, M.D., Rajaram, V., & Fugate-Wentzek, L.A. (1987). Local retinal regions control local eye growth and myopia. *Science, 237* (4810), 73-77.

Wallman, J., & McFadden, S. (1995). Monkey eyes grow into focus. *Nat Med, 1* (8), 737-739.

Wallman, J., Turkel, J., & Trachtman, J. (1978). Extreme myopia produced by modest change in early visual experience. *Science, 201* (4362), 1249-1251.

Wallman, J., Wildsoet, C., Xu, A., Gottlieb, M.D., Nickla, D.L., Marran, L., Krebs, W., & Christensen, A.M. (1995). Moving the retina: choroidal modulation of refractive state. *Vision Res, 35* (1), 37-50.

Wang, E.T., Kodama, G., Baldi, P., & Moyzis, R.K. (2006a). Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc Natl Acad Sci U S A, 103* (1), 135-140.

Wang, G.B., Zhang, Q.J., Xiao, X.S., Li, J.Z., Zhang, F.S., Li, S.Q., Li, W., Li, T., Jia, X.Y., Guo, L., & Guo, X.M. (2004). [Variations of the zinc finger protein 161 gene in Chinese with or without high myopia]. *Yi Chuan, 26* (2), 155-159.

Wang, H.Z., Chen, M.T., Chang, C.H., Tsai, M.C., Wu, W.C., & Chung, C.B. (1994a). The changes of ocular axial length and corneal curvatures after scleral buckling for retinal detachment. *Gaoxiong Yi Xue Ke Xue Za Zhi, 10* (2), 77-83.

Wang, I.J., Chiang, T.H., Shih, Y.F., Hsiao, C.K., Lu, S.C., Hou, Y.C., & Lin, L.L. (2006b). The association of single nucleotide polymorphisms in the 5'-regulatory region of the lumican gene with susceptibility to high myopia in Taiwan. *Mol Vis, 12*, 852-857.

Wang, J., Chuang, K., Ahluwalia, M., Patel, S., Umblas, N., Mirel, D., Higuchi, R., & Germer, S. (2005). High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques, 39* (6), 885-893.

Wang, P., Li, S., Xiao, X., Guo, X., & Zhang, Q. (2009a). An evaluation of OPTC and EPYC as candidate genes for high myopia. *Mol Vis, 15*, 2045-2049.

Wang, P., Li, S., Xiao, X., Jia, X., Jiao, X., Guo, X., & Zhang, Q. (2009b). High myopia is not associated with the SNPs in the TGIF, lumican, TGFB1, and HGF genes. *Invest Ophthalmol Vis Sci, 50* (4), 1546-1551.

Wang, Q., Klein, B.E., Klein, R., & Moss, S.E. (1994b). Refractive status in the Beaver Dam Eye Study. *Invest Ophthalmol Vis Sci, 35* (13), 4344-4347.

Wasson, J., Skolnick, G,. Love-Gregory, L., Permutt, M.A. (2002) Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. *Biotechniques, 32*(5):1144-1146

Wensor, M., McCarty, C.A., & Taylor, H.R. (1999). Prevalence and risk factors of myopia in Victoria, Australia. *Arch Ophthalmol, 117* (5), 658-663.

Wiesel, T.N., & Raviola, E. (1977). Myopia and eye enlargement after neonatal lid fusion in monkeys. *Nature, 266* (5597), 66-68.

Wiesen, M.H., Bogdanovich, S., Agarkova, I., Perriard, J.C., & Khurana, T.S. (2007). Identification and characterization of layer-specific differences in extraocular muscle m-bands. *Invest Ophthalmol Vis Sci, 48* (3), 1119-1127.

Wigginton, J.E., Cutler, D.J., & Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet, 76* (5), 887-893.

Wildsoet, C.F. (1997). Active emmetropization--evidence for its existence and ramifications for clinical practice. *Ophthalmic Physiol Opt, 17* (4), 279-290.

Wiltshire, S., Morris, A.P., & Zeggini, E. (2008). Examining the statistical properties of fine-scale mapping in large-scale association studies. *Genet Epidemiol, 32* (3), 204-214.

Wojciechowski, R., Bailey-Wilson, J.E., & Stambolian, D. (2010). Association of matrix metalloproteinase gene polymorphisms with refractive error in Amish and Ashkenazi families. *Invest Ophthalmol Vis Sci, 51* (10), 4989-4995.

Wojciechowski, R., Moy, C., Ciner, E., Ibay, G., Reider, L., Bailey-Wilson, J.E., & Stambolian, D. (2006). Genomewide scan in Ashkenazi Jewish families demonstrates

evidence of linkage of ocular refraction to a QTL on chromosome 1p36. *Hum Genet, 119* (4), 389-399.

Wolffsohn, J.S., Gilmartin, B., Li, R.W., Edwards, M.H., Chat, S.W., Lew, J.K., & Yu, B.S. (2003). Nearwork-induced transient myopia in preadolescent Hong Kong Chinese. *Invest Ophthalmol Vis Sci, 44* (5), 2284-2289.

Wolford, J.K., Blunt, D., Ballecer, C., & Prochazka, M. (2000). High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC). *Hum Genet, 107* (5), 483-487.

Wong, T.Y., Klein, B.E., Klein, R., Knudtson, M., & Lee, K.E. (2003). Refractive errors, intraocular pressure, and glaucoma in a white population. *Ophthalmology, 110* (1), 211-217.

Wu, G., Hua, L., Zhu, J., Mo, Q.H., & Xu, X.M. (2003). Rapid, accurate genotyping of beta-thalassaemia mutations using a novel multiplex primer extension/denaturing high-performance liquid chromatography assay. *Br J Haematol, 122* (2), 311-316.

Wu, H., Yong R., Tan, S., & Yap, E. (2000). Further evidence of association between trabecular meshwork-induced glucocorticoid response (TIGR/Myocilin) gene and severe myopia. *Invest Ophthalmol Vis Sci, 41* (Abstract), 170.

Wu, H., Yu, X.H., & Yap, E.P. (1999). Allelic association between trabecular meshwork-induced glucocorticoid response (TIGR) gene and severe sporadic myopia. *Invest Ophthalmol Vis Sci, 40* (Abstract 3148), S600.

Wu, H.M., Seet, B., Yap, E.P., Saw, S.M., Lim, T.H., & Chia, K.S. (2001). Does education explain ethnic differences in myopia prevalence? A population-based study of young adult males in Singapore. *Optom Vis Sci, 78* (4), 234-239.

Wu, S.Y., Nemesure, B., & Leske, M.C. (2000). Glaucoma and myopia. *Ophthalmology, 107* (6), 1026-1027.

Wu, W.M., Tsai, H.J., Pang, J.H., Wang, H.S., Hong, H.S., & Lee, Y.S. (2005). Touchdown thermocycling program enables a robust single nucleotide polymorphism typing method based on allele-specific real-time polymerase chain reaction. *Anal Biochem, 339* (2), 290-296.

Xiao, W., & Oefner, P.J. (2001). Denaturing high-performance liquid chromatography: A review. *Hum Mutat, 17* (6), 439-474.

Xu, L., Li, J., Cui, T., Hu, A., Fan, G., Zhang, R., Yang, H., Sun, B., & Jonas, J.B. (2005). Refractive error in urban and rural adult Chinese in Beijing. *Ophthalmology, 112* (10), 1676-1683.

Xu, Z., & Taylor, J.A. (2009). SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res, 37* (Web Server issue), W600-605.

Yamane, T., Mok, J., Oka, A., Okada, E., Nishizaki, R., Meguro, A., Yonemoto, J., Kulski, J.K., Ohno, S., Inoko, H., & Mizuki, N. (2007). Lack of association with high myopia and the MYP2 locus in the Japanese population by high resolution microsatellite analysis on chromosome 18. *Clin Ophthalmol, 1* (3), 311-316.

Yanovitch, T., Li, Y.J., Metlapally, R., Abbott, D., Viet, K.N., & Young, T.L. (2009).

Hepatocyte growth factor and myopia: genetic association analyses in a Caucasian population. *Mol Vis, 15*, 1028-1035.

Yap, M., Wu, M., Liu, Z.M., Lee, F.L., & Wang, S.H. (1993). Role of heredity in the genesis of myopia. *Ophthalmic Physiol Opt, 13* (3), 316-319.

Yip, C.W., Leung, K.L., Wong, D., Cheung, D.T., Chu, M.Y., Tang, H.S., & Kam, K.M. (2006). Denaturing HPLC for high-throughput screening of rifampicin-resistant Mycobacterium tuberculosis isolates. *Int J Tuberc Lung Dis, 10* (6), 625-630.

Yip, S.P., Pun, S.F., Leung, K.H., & Lee, S.Y. (2003). Rapid, simultaneous genotyping of five common Southeast Asian beta-thalassemia mutations by multiplex minisequencing and denaturing HPLC. *Clin Chem, 49* (10), 1656-1659.

You, F.M., Huo, N., Gu, Y.Q., Luo, M.C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J., & Anderson, O.D. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics, 9*, 253.

Young, T.L. (2004). Dissecting the genetics of human high myopia: a molecular biologic approach. *Trans Am Ophthalmol Soc, 102*, 423-445.

Young, T.L., Atwood, L.D., Ronan, S.M., Dewan, A.T., Alvear, A.B., Peterson, J., Holleschau, A., & King, R.A. (2001). Further refinement of the MYP2 locus for autosomal dominant high myopia by linkage disequilibrium analysis. *Ophthalmic Genet, 22* (2), 69-75.

Young, T.L., Deeb, S.S., Ronan, S.M., Dewan, A.T., Alvear, A.B., Scavello, G.S., Paluru, P.C., Brott, M.S., Hayashi, T., Holleschau, A.M., Benegas, N., Schwartz, M., Atwood, L.D., Oetting, W.S., Rosenberg, T., Motulsky, A.G., & King, R.A. (2004). X-linked high myopia associated with cone dysfunction. *Arch Ophthalmol, 122* (6), 897-908.

Young, T.L., Guo, X.D., King, R.A., Johnson, J.M., & Rada, J.A. (2003). Identification of genes expressed in a human scleral cDNA library. *Mol Vis, 9*, 508-514.

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

Young, T.L., Ronan, S.M., Alvear, A.B., Wildenberg, S.C., Oetting, W.S., Atwood, L.D., Wilkin, D.J., & King, R.A. (1998a). A second locus for familial high myopia maps to chromosome 12q. *Am J Hum Genet, 63* (5), 1419-1424.

Young, T.L., Ronan, S.M., Drahozal, L.A., Wildenberg, S.C., Alvear, A.B., Oetting, W.S., Atwood, L.D., Wilkin, D.J., & King, R.A. (1998b). Evidence that a locus for familial high myopia maps to chromosome 18p. *Am J Hum Genet, 63* (1), 109-119.

Yuan, H.Y., Chiou, J.J., Tseng, W.H., Liu, C.H., Liu, C.K., Lin, Y.J., Wang, H.H., Yao, A., Chen, Y.T., & Hsu, C.N. (2006). FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res, 34* (Web Server issue), W635-641.

Yue, P., & Moult, J. (2006). Identification and analysis of deleterious human SNPs. *J Mol Biol, 356* (5), 1263-1274.

Yura, T. (1998). The relationship between the types of axial elongation and the prevalence of lattice degeneration of the retina. *Acta Ophthalmol Scand, 76* (1), 90-95.

Zadnik, K. (1997). The Glenn A. Fry Award Lecture (1995). Myopia development in childhood. *Optom Vis Sci, 74* (8), 603-608.

Zadnik, K., Jones, L.A., Irvin, B.C., Kleinstein, R.N., Manny, R.E., Shin, J.A., & Mutti, D.O. (2000). Myopia and ambient night-time lighting. CLEERE Study Group. Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error. *Nature, 404* (6774), 143-144.

Zadnik, K., & Mutti, D.O. (1995). How applicable are animal myopia models to human juvenile onset myopia? *Vision Res, 35* (9), 1283-1288.

Zadnik, K., Satariano, W.A., Mutti, D.O., Sholtz, R.I., & Adams, A.J. (1994). The effect of parental history of myopia on children's eye size. *JAMA, 271* (17), 1323-1327.

Zayats, T., Guggenheim, J.A., Hammond, C.J., & Young, T.L. (2008). Comment on 'A PAX6 gene polymorphism is associated with genetic predisposition to extreme myopia'. *Eye (Lond), 22* (4), 598-599; author reply 599.

Zayats, T., Yanovitch, T., Creer, R.C., McMahon, G., Li, Y.J., Young, T.L., & Guggenheim, J.A. (2009). Myocilin polymorphisms and high myopia in subjects of European origin. *Mol Vis, 15*, 213-222.

Zechi-Ceide, R.M., Jesus Oliveira, N.A., Guion-Almeida, M.L., Antunes, L.F., Richieri-Costa, A., & Passos-Bueno, M.R. (2008). Clinical evaluation and COL2A1 gene analysis in 21 Brazilian families with Stickler syndrome: identification of novel mutations, further genotype/phenotype correlation, and its implications for the diagnosis. *Eur J Med Genet, 51* (3), 183-196.

Zejmo, M., Forminska-Kapuscik, M., Pieczara, E., Filipek, E., Mrukwa-Kominek, E., Samochowiec-Donocik, E., Leszczynski, R., & Smuzynska, M. (2009). Etiopathogenesis and management of high-degree myopia. Part I. *Med Sci Monit, 15* (9), RA199-202.

Zeng, D., & Lin, D.Y. (2005). Estimating haplotype-disease associations with pooled genotype data. *Genet Epidemiol, 28* (1), 70-82.

Zha, Y., Leung, K.H., Lo, K.K., Fung, W.Y., Ng, P.W., Shi, M.G., Yap, M.K., & Yip, S.P. (2009). TGFB1 as a susceptibility gene for high myopia: a replication study with new findings. *Arch Ophthalmol, 127* (4), 541-548.

Zhang, Q., Beltran, W.A., Mao, Z., Li, K., Johnson, J.L., Acland, G.M., & Aguirre, G.D. (2003). Comparative analysis and expression of CLUL1, a cone photoreceptor-specific gene. *Invest Ophthalmol Vis Sci, 44* (10), 4542-4549.

Zhang, Q., Guo, X., Xiao, X., Jia, X., Li, S., & Hejtmancik, J.F. (2005). A new locus for autosomal dominant high myopia maps to 4q22-q27 between D4S1578 and D4S1612. *Mol Vis, 11*, 554-560.

Zhang, Q., Guo, X., Xiao, X., Jia, X., Li, S., & Hejtmancik JF. (2006) Novel locus for X linked recessive high myopia maps to Xq23-q25 but outside MYP1. *J Med Genet. 43*(5), e20.

Zhang, Q., Li, S., Xiao, X., Jia, X., & Guo, X. (2007). Confirmation of a genetic locus for X-linked recessive high myopia outside MYP1. *J Hum Genet, 52* (5), 469-472.

Zhou, G., & Williams, R.W. (1999). Eye1 and Eye2: gene loci that modulate eye size, lens weight, and retinal area in the mouse. *Invest Ophthalmol Vis Sci, 40* (5), 817-825.

Zhou, G., Kamahori, M., Okano, K., Chuan, G., Harada, K., Kambara, H. (2001). Quantitative detection of single nucleotide polymorphisms for a pooled sample by a bioluminometric assay coupled with modified primer extension reactions (BAMPER). *Nucleic Acids Res, 29*(19), E93.

Zhou, J., & Young, T.L. (2005). Evaluation of Lipin 2 as a candidate gene for autosomal dominant 1 high-grade myopia. *Gene, 352*, 10-19.

Zhou, X., Shen, M., Xie, J., Wang, J., Jiang, L., Pan, M., Qu, J., & Lu, F. (2008). The development of the refractive status and ocular growth in C57BL/6 mice. *Invest Ophthalmol Vis Sci, 49* (12), 5208-5214.

Zhu, G., Hewitt, A.W., Ruddle, J.B., Kearns, L.S., Brown, S.A., Mackinnon, J.R., Chen, C.Y., Hammond, C.J., Craig, J.E., Montgomery, G.W., Martin, N.G., & Mackey, D.A. (2008). Genetic dissection of myopia: evidence for linkage of ocular axial length to chromosome 5q. *Ophthalmology, 115* (6), 1053-1057 e1052.

Ziegler, A., Konig, I.R., & Thompson, J.R. (2008). Biostatistical aspects of genome-wide association studies. *Biom J, 50* (1), 8-28.

Zou, G., & Zhao, H. (2003). Haplotype frequency estimation in the presence of genotyping errors. *Hum Hered, 56* (1-3), 131-138.

Appendices

The information sheets (in both English and Chinese versions) and the consent forms (only in Chinese version).

The Hong Kong Polytechnic University
Faculty of Health and Social Studies

Research programme on myopia genetics

Information Sheet

Aim:

In Hong Kong, about 70% of young adults have myopia (or short-sightedness). Previous studies have shown the contribution of both genetic and environmental factors to myopia. We are interested in identifying genes and other factors predisposing humans to myopia. This study is the first of its kind in Hong Kong.

The target subjects are high myopes and emmmetropic control. The refractive errors of high myopic subjects must be at least –8 dioptres whereas the control non-myopic. Please think seriously before deciding to participate.

Method:

Each participant will be asked to give the necessary personal information (including sex, medical history, etc.), and offered ***FREE*** eye examination (about 1.5 hours) and health check (about 0.5 hour), all performed by ***qualified personnel***.

Eye examination. The following measurements will be made using standard optometric procedures: refractive status of the eye, the ocular aberration, the corneal curvature, the dimensions of the eyeball, the intraocular pressure and the interpupillary distance.

For accurate refraction, one or two drops of an eyedrop (0.5% cyclopentolate or 1% tropicamide) will be instilled in your eyes. This eyedrop will relax the focusing power of your eyes and things will look brighter because your pupil will be dilated. This could last 6-12 hours. For the measurement of your eyeball size, another eyedrop (Benoxinate, 0.4%) will be instilled on your eyes to numb the cornea. The cornea will regain sensation within 30 minutes. If you are allergic to any eyedrop or have a history of glaucoma, you must let us know.

The other ocular measurements will be performed in a non-invasive manner without using any eyedrop.

Health Check. The health check includes measurement of weight, height, blood pressure, pulse, head circumference and other head-related measurements. Your blood and urine samples will also be collected for laboratory tests (as part of the health check) and for research use. The laboratory tests include routine urine analysis, blood cholesterol level, complete blood counts and blood groups.

You will be informed of your own results of the eye examination, health check and blood tests. You will be offered a ***free*** consultation session with our physicians in the University Health Service if your laboratory test results are abnormal. All the information and samples collected will only be available to the investigators involved in studies on the genetic factors involved in myopia. Otherwise, all personal information collected will be kept confidential. Data from this study will be published, but individuals will not be identified or identifiable. You may decline to take part or withdraw should you change your mind.

For inquiry or booking, please contact our optometrist Mr.LO KA KIN (2766 4147 or 9846 """). If you have any complaint, please contact the Principal Investigator Dr YIP Shea Ping (2766 7906) or the Human Subjects Ethics Subcommittee.

香港理工大學
醫療及社會科學院
近視遺傳之研究簡介

研究之目的：

香港大概有百分之七十的年青人患有近視，現有之研究顯示遺傳及環境因素可促成近視的形成。我們希望可以找出促成近視的人類基因及其他因素，此乃香港首個同類形的研究。

本研究的對像涉及深近視(近視必須達八百度或以上)及無屈光度數人仕，我們十分希望每一位參加者能夠完成此項研究，請仔細考慮後才決定是否參與。

研究之方式：

每位參與者將會被邀請提供所須之個人資料(包括年齡、病歷等)，然後，我們會爲參與者進行免費的眼睛檢查(約一個小時)及健康檢查(約半小時)。

眼睛檢查：在檢驗眼睛過程中，參與者的屈光度數，眼睛的像差，角膜孤度，眼球結構、眼壓及瞳距將會被量度。

我們會用眼藥水(0.5% cyclopentolate 或 1% tropicamide) 暫時減低參與者的眼睛調節力來進行屈光度數檢驗。因調節能力減低，參與者的對焦能力亦會減低；因瞳孔放大，會對強光較爲敏感；藥力維持約 3-6 小時。另外，眼角膜亦會被眼藥水(Benoxinate 0.4%)局部麻醉以進行眼球結構量度，藥力維持約三十分鐘。如對眼藥水敏感或曾患有青光眼，參與者必須告訴研究人員。

而其他眼睛檢查程序，則無需使用任何眼藥水，也是十分安全。

健康檢查：健康檢查包括體重、高度、血壓、脈博，頭圍及頭部有關的量度。參與者的血液及尿液樣本亦會被抽取以作測試、和供應近視遺傳研究之使用。測試包括全血計數、血壓及尿液常規檢查。

參與者會收到有關之檢查報告；假如參與者的測試結果不正常，我們會給予參與者一次的免費醫療諮詢。所有個人資料均會保密處理。參與者可以隨時退出此項計劃而無須負任何責任。

如有任何查詢或預約，請致電伍保華視光師(2766 4712 或 9846 　　)。
如有任何投訴，可直接聯絡研究主管葉社平博士(2766 7906)或香港理工大學之科研道德委員會。

香港理工大學
醫療及社會科學院

近視遺傳之研究

參與研究同意書

* *

✻ 本人明白所提供之研究簡介。

✻ 本人有被給與發問之機會，而所發問之問題亦已獲得滿意的答覆。

✻ 本人擁有隨時收回此協議而不受任何懲罰之權利。

✻ 本人授權與理工大學醫療保健部之醫生將本人之健康檢查報告，發給此項研究之科研人員作研究之用。本人明白此授權書之影印副本亦同樣有效。

✻ 本人明白此結果將來用作公佈時，本人個人資料及身份絕不會被公開。

* *

本人__________________________同意參與是項研究。

或

簽署__________________________ 簽署__________________________

(參與人仕) (見証人)

日期__________________________

此項研究已獲香港理工大學之科研道德委員會核准。若有任何投訴， 閣下可致函此委員會。