

**Discovery of genetic determinants for  
refractive error**

Rupal L. Shah

PhD 2018

# Table of Contents

<b>Table of Tables</b> .....	<b>v</b>
<b>Table of Figures</b> .....	<b>xii</b>
<b>List of Abbreviations</b> .....	<b>xvii</b>
<b>Acknowledgements</b> .....	<b>xxi</b>
<b>Summary</b> .....	<b>xxii</b>
<b>Chapter 1 General Introduction</b> .....	<b>1</b>
1.1 Refractive errors .....	2
1.1.1 <i>Definitions / classification of refractive errors</i> .....	2
1.1.2 <i>Worldwide prevalence</i> .....	3
1.1.3 <i>Genetic and environmental influences on refractive errors</i> .....	6
1.1.4 <i>Current treatments</i> .....	7
1.1.5 <i>Why investigate refractive errors?</i> .....	10
1.2 Features of complex traits .....	10
1.2.1 <i>Genetic variants</i> .....	11
1.2.2 <i>Genotyping and Imputation</i> .....	12
1.2.3 <i>Linkage Disequilibrium</i> .....	14
1.3 Genetic investigations .....	15
1.3.1 <i>Family / twin studies</i> .....	16
1.3.2 <i>Association Studies</i> .....	20
1.3.3 <i>Heritability</i> .....	22
1.4 What has been discovered so far by each of these methods? .....	26
1.4.1 <i>Heritability</i> .....	26
1.4.2 <i>Linkage studies</i> .....	27
1.4.3 <i>Genome-wide Association Studies (GWAS)</i> .....	29
1.5 Overall Aims .....	32
<b>Chapter 2 General Methods</b> .....	<b>34</b>
2.1 ALSPAC cohort of children and their mothers .....	35
2.1.1 <i>Phenotypes</i> .....	35
2.1.2 <i>Genotyping and Imputation</i> .....	36

2.2	UK Biobank Study .....	36
2.2.1	<i>Phenotypes</i> .....	37
2.2.2	<i>Genotyping and Imputation</i> .....	38
2.2.3	<i>Exclusion Criteria</i> .....	40
2.2.4	<i>Covariates</i> .....	42
2.2.5	<i>“High-confidence” Variants</i> .....	42
2.3	Statistical Methods .....	43
2.3.1	<i>Single Marker Tests</i> .....	43
2.3.2	<i>Genomic Control</i> .....	47
2.3.3	<i>Meta-analysis of GWAS Results</i> .....	50
2.3.4	<i>Conditional Analysis</i> .....	54
2.3.5	<i>Gene-based and Gene-set Analyses</i> .....	55
2.3.6	<i>SNP-Heritability</i> .....	60
2.3.7	<i>Multiple Testing Correction</i> .....	68
<b>Chapter 3</b>	<b>X-Chromosome Wide Association Study for Refractive Error .....</b>	<b>72</b>
3.1	Introduction .....	73
3.2	Methods .....	74
3.2.1	<i>Quality Control</i> .....	74
3.2.2	<i>Single Marker Tests</i> .....	75
3.2.3	<i>Sex-Specific Effects</i> .....	76
3.2.4	<i>Gene-based and Gene-set Analyses</i> .....	77
3.2.5	<i>Power Calculation</i> .....	78
3.3	Results .....	79
3.3.1	<i>Single Marker Tests</i> .....	79
3.3.2	<i>Sex-Specific Effects</i> .....	85
3.3.3	<i>Gene-based and Gene-set Analyses</i> .....	89
3.3.4	<i>Power Calculation</i> .....	93
3.4	Discussion .....	96
<b>Chapter 4</b>	<b>CREAM Corneal Astigmatism GWAS Meta-analysis .....</b>	<b>101</b>
4.1	Introduction .....	102
4.2	Methods .....	104
4.2.1	<i>Study Cohorts</i> .....	104

4.2.2	<i>Quality Control</i>	108
4.2.3	<i>Meta-analysis of GWAS Results</i>	109
4.2.4	<i>Gene-based and Gene-set Analyses</i>	110
4.2.5	<i>SNP-Heritability and Genetic Correlation</i>	111
4.3	Results	112
4.3.1	<i>Meta-analysis of GWAS Results</i>	112
4.3.2	<i>Gene-based and Gene-set Analyses</i>	124
4.3.3	<i>SNP-Heritability and Genetic Correlation</i>	129
4.4	Discussion	131
<b>Chapter 5</b>	<b>Estimating SNP-heritability of Refractive Errors</b>	<b>137</b>
5.1	Introduction	138
5.2	Methods	139
5.2.1	<i>SNP-Heritability Estimation</i>	139
5.2.2	<i>Observed to Liability Scale Conversion</i>	140
5.2.3	<i>Consideration of Uncorrected Population Effects</i>	140
5.2.4	<i>Quantification of Dominance Effects</i>	141
5.3	Results	142
5.3.1	<i>Consideration of Uncorrected Population Effects</i>	142
5.3.2	<i>Quantification of Dominance Effects</i>	143
5.3.3	<i>Determining Optimal Trait Definitions</i>	146
5.4	Discussion	150
<b>Chapter 6</b>	<b>UK Biobank GWAS for Corneal and Refractive Astigmatism</b>	<b>156</b>
<b>6.1</b>	<b>Introduction</b>	<b>157</b>
<b>6.2</b>	<b>Methods</b>	<b>158</b>
6.2.1	<i>UK Biobank Sample</i>	158
6.2.2	<i>Consideration of Astigmatism Axis</i>	161
6.2.3	<i>Single Marker Tests</i>	162
6.2.4	<i>Gene-based and Gene-set Analyses</i>	164
6.2.5	<i>SNP-Heritability and Genetic Correlation</i>	165
6.2.6	<i>Phenotypic Correlation</i>	165
6.3	Results	166
6.3.1	<i>Consideration of Astigmatism Axis</i>	166

6.3.2	<i>Single Marker Tests</i> .....	172
6.3.3	<i>Gene-based and Gene-set Analyses</i> .....	197
6.3.4	<i>SNP-Heritability and Genetic Correlation</i> .....	202
6.4	Discussion .....	206
<b>Chapter 7</b>	<b>Investigation of myopic primate retina differentially expressed genes in humans</b> .....	<b>213</b>
7.1	Introduction.....	214
7.2	Methods .....	216
7.2.1	<i>Study Sample</i> .....	216
7.2.2	<i>Gene-based Analyses</i> .....	216
7.2.3	<i>Replication in the UK Biobank Study</i> .....	217
7.3	Results .....	218
7.3.1	<i>Gene-based Analyses</i> .....	218
7.3.2	<i>Replication in the UK Biobank Study</i> .....	222
7.3.3	<i>Exploration of wider flanking regions</i> .....	226
7.4	Discussion .....	229
<b>Chapter 8</b>	<b>Overall Conclusions</b> .....	<b>235</b>
<b>References</b>	.....	<b>241</b>
<b>Appendices</b>	.....	<b>277</b>
Appendix A	Acknowledgements and study Information for cohorts investigated in <i>Chapter 4 CREAM Corneal Astigmatism GWAS Meta-analysis</i> .	278
Appendix B	EasyQC style plots for study-wide quality control in <i>Chapter 4 CREAM Corneal Astigmatism GWAS Meta-analysis</i> .....	287
Appendix C	MAGMA gene-based and gene-set results from <i>Chapter 4 CREAM Corneal Astigmatism GWAS Meta-analysis</i> .....	296
Appendix D	Regional association plots for loci demonstrating genome-wide significant association ( $P < 5 \times 10^{-8}$ ) in <i>Chapter 6 UK Biobank GWAS for Corneal and Refractive Astigmatism</i> .....	300
Appendix E	Summary of MAGMA gene-based results from <i>Chapter 7 Investigation of myopic primate retina differentially expressed genes in humans</i> ...	308

## Table of Tables

Table 1.1: Summary of myopia (MYP) loci listed in the Online Mendelian Inheritance in Man (OMIM) database (URL: <a href="https://omim.org/">https://omim.org/</a> ).....	28
Table 2.1: Summary of the UK Biobank Sample. ....	40
Table 2.2: Summary of exclusions from the UK Biobank sample due to ocular conditions or surgery. ....	41
Table 2.3: Example of false discovery rates in action. Number of genes in the list ( $m$ ) = 10; False discovery rate ( $q$ ) = 0.05; $iq/m$ = critical value; $FDR = Pm/i$ . Genes in bold = significant associations based on FDR. ....	70
Table 3.1: Summary of refractive error in the XWAS analysis sample. ....	74
Table 3.2: Summary of results for variants on chromosome X achieving association test P-values less than $1 \times 10^{-5}$ (FEMALES). EAF =: effect allele frequency; Effect = effect size per copy of <i>Effect Allele</i> ; SE: standard error of beta estimate; N = 2,146. ....	80
Table 3.3: Summary of results for variants on chromosome X from meta-analysis of association tests for MSE at age 15 years ( $P < 1 \times 10^{-5}$ ). EAF = effect allele frequency; Effect = effect size per copy of allele B; SE = standard error of beta estimate; N = 4,070 for all variants. ....	83
Table 3.4: Summary of the 10 variants demonstrating the greatest difference in effect size (beta) between the sexes. EAF = effect allele frequency; SE = standard error of beta estimate; P SexDiff = p-value from the test for male and female beta estimates being identical. ....	88
Table 3.5: The 10 genes demonstrating strongest association from the VEGAS2 gene-based association test. Start and stop positions listed include $\pm 50$ kb flanking regions. nSNPs = number of variants included in gene region; Test Statistic = gene-based $\chi^2$ test statistic with nSNPs degrees of freedom; P-value = obtained from <i>Test Statistic</i> and adjusting for LD between variants; FDR = false discovery rate; Lead Variant = variant within gene locus with strongest association signal from previous SNP-based association test. Total number of genes tested = 1,252.....	90
Table 3.6: The 10 genes demonstrating strongest association from the MAGMA gene-based association test. Start and stop positions listed include $\pm 50$ kb flanking regions. nSNPs = number of variants included in gene region; Z-Statistic	

= gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate. Total number of genes tested = 803..... 91

Table 3.7: Gene-sets demonstrating FDR < 0.05 from MAGMA gene-set association test. nGenes = number of genes included in gene set; Beta = gene-set test statistic; SE = standard error; FDR = false discovery rate; Bonferroni Adjusted P-value = P-value multiplied by the number of gene-sets tested. Total number of gene-sets tested = 10,468..... 92

Table 3.8: Power ( $\beta$ ) to detect an effect size of 0.10 D for variants with MAFs ranging between 0.01 and 0.50. Based on a sample size of N = 4,070 for a normally distributed trait with mean (SD) = -0.382 (1.28) D and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency..... 94

Table 4.1: Subject demographics of participating CREAM study groups..... 105

Table 4.2: Instrument for measuring corneal curvature. .... 106

Table 4.3: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all European ancestry participants aged > 25 years. EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list..... 114

Table 4.4: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all European ancestry participants aged < 25 years. EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list..... 115

Table 4.5: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all Asian ancestry participants aged > 25 years. EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list..... 116

Table 4.6: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all Asian ancestry participants aged < 25 years. EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list..... 117

Table 4.7: Variants demonstrating association signals of  $P < 1 \times 10^{-5}$  in each region in the GWAS meta-analysis of all samples (Europeans and Asians of all ages combined). EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list. .... 121

Table 4.8: Power ( $\beta$ ) to detect an odds ratio of 1.1 for variants with MAFs ranging between 0.01 and 0.50. Based on a sample size of  $N = 31,370$ , a case:control ratio of 1:2.3, a trait population prevalence of 43.5% and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency..... 123

Table 4.9: The 10 genes demonstrating strongest association from the VEGAS2 gene-based association test. Start and stop positions listed include  $\pm 50$  kb flanking regions. nSNPs = number of variants included in gene region. Test Statistic = gene-based  $\chi^2$  test statistic with nSNPs degrees of freedom. P-value = obtained from *Test Statistic* and adjusting for LD between variants. FDR = false discovery rate. Lead Variant = variant within gene locus with strongest association signal from previous single marker based association test. Genes shown in bold were also identified in the 10 strongest associated genes using MAGMA (Table 4.10). Total number of genes tested = 24,232. .... 126

Table 4.10: Genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05). Start and stop positions listed include  $\pm 50$  kb flanking regions. nSNPs = number of variants included in gene region. Z-Statistic = gene-based test statistic. P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model. FDR = false discovery rate. Genes shown in bold were also identified in the 10 strongest associated genes using VEGAS2 (Table 4.9). Total number of genes tested = 18,418. .... 127

Table 4.11: SNP-heritability estimated in samples of European ancestry from the CREAM consortium using LDSC.  $h^2_{\text{SNP}}$  = SNP-heritability; SE = standard error; P-value = test of the null hypothesis ( $h^2_{\text{SNP}} = 0$ )..... 129

Table 4.12: Genetic correlations between pairs of refractive error traits in samples of European ancestry from the CREAM consortium using LDSC. RA = refractive astigmatism; CA = corneal astigmatism; MSE = mean spherical equivalent. P-values refer to the null hypothesis of zero correlation between traits. .... 130

Table 5.1: Estimating the proportion of  $h^2_{\text{SNP}}$  that is inflated due to uncorrected cryptic relatedness and population stratification. Abbreviations:  $h^2_{\text{joint}}$  = Total SNP-heritability estimated from the joint test; P-value = probability of  $h^2 = 0$ ;  $V_{\text{POP}}$  = proportion of phenotypic variance attributed to population structure (cryptic relatedness and population stratification);  $V_{\text{CR}}$  = proportion of phenotypic variance attributed to cryptic relatedness only;  $V_{\text{PS}}$  = proportion of phenotypic variance attributed to population stratification only.  $N = 25,880$ . 145

Table 5.2: Partitioning trait variance into additive, dominance and environmental components using GREMLd..... 145



Table 5.3: Estimates of SNP-heritability ( $h^2_{\text{SNP}}$ ) for corneal astigmatism, refractive astigmatism and spherical equivalent. Analyses were conducted using GCTA-GREML. $h^2_{\text{SNP}}$ = SNP-heritability; SE = standard error; P-value = test of the null hypothesis ( $h^2_{\text{SNP}} = 0$ ).....	148
Table 6.1: Summary of exclusions from the UK Biobank Sample for GWAS of Corneal and Refractive Astigmatism. MSE = mean spherical equivalent. ....	158
Table 6.2: Participant demographics and distribution of corneal and refractive astigmatism in the analysed UK Biobank sample. ....	159
Table 6.3: Proportion of individuals with corneal and refractive astigmatism described as with-the-rule, against-the-rule and oblique; stratified by age in the UK Biobank sample. WTR = with-the-rule astigmatism; ATR = against-the-rule astigmatism. For negative cylindrical powers, WTR = axis 0-30° or 150-180°; ATR = axis 60-120°; oblique = all other angles. The reverse applies for positive cylindrical powers. ....	167
Table 6.4: Summary of corneal and refractive astigmatism and their respective $J_0$ and $J_{45}$ components stratified by age in the UK Biobank sample.....	170
Table 6.5: Summary of GWAS undertaken. $\lambda_{\text{GC}}$ = genomic inflation factor; $\lambda_{\text{LDSC}}$ = intercept from LD Score regression. The number of independent loci was identified using GCTA-COJO. INT = inverse normal transformed (continuous). ....	173
Table 6.6: Variants achieving association test p-values $< 1 \times 10^{-5}$ in GWAS for Corneal Astigmatism analysed as a continuous trait with BOLT-LMM. EAF = effect allele frequency; HWE P-value = P-value from the Hardy-Weinberg disequilibrium test; SE = standard error. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list.....	177
Table 6.7: Conditional analyses performed using summary statistics and conditioning on lead variant – for genome-wide significant associated loci from Corneal Astigmatism GWAS. GCTA-COJO Lead Variant: lead variant within $\pm 1$ Mb of <i>GWAS Lead Variant</i> location. ....	178
Table 6.8: Variants achieving association test p-values $< 1 \times 10^{-5}$ in GWAS for Refractive Astigmatism analysed as a continuous trait with BOLT-LMM. EAF: effect allele frequency, HWE P-value: P-value from the Hardy-Weinberg disequilibrium test, SE: standard error. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ....	182
Table 6.9: Conditional analyses performed using summary statistics and conditioning on lead variant – for genome-wide significant associated loci from	

Refractive Astigmatism GWAS. GCTA-COJO Lead Variant: lead variant within $\pm 1$ Mb of <i>GWAS Lead Variant</i> location.....	183
Table 6.10: The 10 variants demonstrating strongest association in GWAS for Corneal Astigmatism analysed as an inverse normal transformed trait with BOLT-LMM. Effect sizes are in units of standard deviations. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ....	186
Table 6.11: The 10 variants demonstrating strongest association in GWAS for Refractive Astigmatism analysed as an inverse normal transformed trait with BOLT-LMM. Effect sizes are in units of standard deviations. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ....	187
Table 6.12: The 10 variants demonstrating strongest association in GWAS for Corneal Astigmatism analysed as a continuous trait with PLINK 2.0. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ....	190
Table 6.13: The 10 variants demonstrating strongest association in GWAS for Corneal Astigmatism analysed as a dichotomous trait (cases $\geq 1.00$ D) with PLINK 2.0. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ....	191
Table 6.14: The 10 variants demonstrating strongest association in GWAS for Refractive Astigmatism analysed as a continuous trait with PLINK 2.0. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ...	193
Table 6.15: The 10 variants demonstrating strongest association in GWAS for Refractive Astigmatism analysed as a dichotomous trait (cases $\geq 1.00$ D) with PLINK 2.0. NB: variants within $\pm 500$ kb of listed (lead) variant are not included in this list. ....	194
Table 6.16: Sensitivity analyses investigating the effect of outlier phenotype values on association signals – for genome-wide significant associated loci from Corneal Astigmatism GWAS. <i>All Individuals</i> P-value obtained from analyses using PLINK 2.0. Sample sizes: All Individuals = 69,140; Excluding phenotype extremes / anisometropia = 68,411. ....	196
Table 6.17: Sensitivity analyses investigating the effect of outlier phenotype values on association signals – for genome-wide significant associated loci from Refractive Astigmatism GWAS. <i>All Individuals</i> P-value obtained from analyses using PLINK 2.0. Sample sizes: All Individuals = 70,505; Excluding phenotype extremes / anisometropia = 69,304. ....	196
Table 6.18: Genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Corneal Astigmatism as a continuous	

trait. Start and stop positions listed include $\pm 50$ kb flanking regions; nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from <i>Z-Statistic</i> under the assumption of a normally distributed model; FDR = false discovery rate. ....	198
Table 6.19: Genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Astigmatism as a continuous trait. Start and stop positions listed include $\pm 50$ kb flanking regions; nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from <i>Z-Statistic</i> under the assumption of a normally distributed model; FDR = false discovery rate. ....	200
Table 6.20: Estimates of SNP-heritability ( $h^2_{\text{SNP}}$ ) using GCTA and LSDC. $h^2_{\text{SNP}}$ = SNP-heritability; SE: standard error; P-value = test of the null hypothesis ( $h^2_{\text{SNP}} = 0$ ). ....	202
Table 6.21: Genetic and phenotypic correlations between pairs of refractive error traits. Genotypic correlations were obtained using LDSC and summary statistics from BOLT-LMM analyses (N = 86,335 or 88,005). All phenotypic correlations are Pearson correlations for 63,466 unrelated individuals included in GWAS for these respective traits and with data available for all three refractive error measures CA = corneal astigmatism; RA = refractive astigmatism; MSE = mean spherical equivalent. P-values refer to the null hypothesis of zero genetic correlation between traits. ....	204
Table 6.22: Previously observed associations with ocular traits at the newly discovered susceptibility loci for astigmatism. ....	207
Table 7.1: Candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Error. Start and stop positions refer to the respective gene transcription start and stop sites (no flanking region included); nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from <i>Z-Statistic</i> under the assumption of a normally distributed model; FDR = false discovery rate. ....	220
Table 7.2: Candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Error with the inclusion of 50 kb flanking regions. nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from <i>Z-Statistic</i> under the assumption of a normally distributed model; FDR = false discovery rate. ....	220
Table 7.3: Candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Error with the	

inclusion of 200 kb flanking regions. nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate. .... 221

Table 7.4: Testing for replication of the 4 candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) in the UK Biobank study. Start and stop positions refer to the respective gene transcription start and stop sites (no flanking region included); nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate. .... 223

Table 7.5: Testing for replication of the 9 candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) with the inclusion of 50 kb flanking regions in the UK Biobank study. nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate. .... 223

Table 7.6: Testing for replication of the 17 candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) with the inclusion of 200 kb flanking regions in the UK Biobank study. nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate. .... 224

## Table of Figures

Figure 2.1: Flow diagram of METAL inverse variance weighted meta-analysis procedure. Adapted from Willer et al. (2010).....	52
Figure 2.2: Meta-analysis for a single, simulated variant with Forest Plot. SE = standard error. Meta-analysed values were calculated according to the values from the four contributing studies and the inverse variance weighted procedure outlined in Figure 2.1. Forest Plot circles denote effect size; diamond denotes meta-analysed effect size; all error bars denote 95% confidence intervals. ....	53
Figure 2.3: The normal probability density function for transformation from the observed scale to the liability scale. If cases account for 20% of the individuals sampled (shaded area), the height of the function at the threshold value, $t$ (edge of shaded area) is 0.28 ( $z$ , denoted by dotted line). Adapted from Lee et al. (2011). ....	63
Figure 3.1: Plot of minor allele frequencies (MAFs) between males and females. Red line = line of unity (Females MAF = Males MAF). $N = 4,070$ across 267,185 variants. ....	76
Figure 3.2: Chromosome X quantile-quantile plot from sex-specific linear regression association tests with MSE at age 15 years. Blue squares = males; red circles = females. Y-axis shows observed negative $\log_{10}$ p-values and X-axis shows expected negative $\log_{10}$ p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).....	81
Figure 3.3: Chromosome X Manhattan and Quantile-Quantile plots from meta-analysed results. Panel A: Horizontal blue line denotes an arbitrary threshold for declaring “suggestive” evidence of association ( $P = 1 \times 10^{-5}$ ). Panel B: Y-axis shows observed negative $\log_{10}$ p-values and X-axis shows expected negative $\log_{10}$ p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected). ....	84
Figure 3.4: Chromosome X Miami plot from sex-specific linear regression association tests for MSE at age 15 years. Upper portion (red) = females; lower portion (blue) = males. Horizontal yellow lines denote suggestive association threshold of $P = 1 \times 10^{-5}$ . ....	86
Figure 3.5: Differences in effect size (beta) estimates for males and females from individual chromosome X association tests. Plot of beta estimates for males vs.	

females from individual chromosome X association tests. Red line = line of unity (Females Beta = Males Beta). N = 4,070 across 267,185 variants..... 87

Figure 3.6: Power calculations. Panel A: Power to detect variants with an effect size of 0.10 D for the ALSPAC sample with MAFs ranging between 0.01 and 0.50. Based on a sample size of N = 4,070 for a normally distributed trait with mean (SD) = -0.382 (1.28) D and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency. Panel B: Minimum sample sizes required to detect variants with effect sizes of 0.10 D with 60-80% power. Based on a normally distributed trait with mean MSE (SD): -0.382 (1.28) D and Type I error rate of  $5 \times 10^{-8}$ . MAF = minor allele frequency..... 95

Figure 4.1: Quantile-quantile plots for the separate ancestry/age strata fixed effects meta-analyses. Panel A: European ancestry, aged > 25 years; Panel B: European ancestry, aged < 25 years; Panel C: Asian ancestry, aged > 25 years; Panel D: Asian ancestry, aged < 25 years. Y-axes show observed negative  $\log_{10}$  p-values and X-axes show expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected). ..... 112

Figure 4.2: Manhattan plots for the separate ancestry/age strata fixed effects meta-analyses. Panel A: European ancestry, aged > 25 years; Panel B: European ancestry, aged < 25 years; Panel C: Asian ancestry, aged > 25 years; Panel D: Asian ancestry, aged < 25 years. Y-axes show negative  $\log_{10}$  p-values and X-axes show genomic position. Red line corresponds to  $P = 5 \times 10^{-8}$ , blue line corresponds to  $P = 1 \times 10^{-5}$ . ..... 113

Figure 4.3: Manhattan and Quantile-Quantile plots for fixed effects meta-analysis of European and Asian participants of all ages combined (N = 31,370). Panel A: Manhattan plot: Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Panel B: Quantile-Quantile plot: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected). ..... 119

Figure 4.4: Region plot for the locus demonstrating strongest association in the GWAS fixed-effects meta-analysis for European and Asian participants of all ages combined (N = 31,370). Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant rs7673984 (highlighted in purple)..... 120

Figure 4.5: Forest plot and summary table for lead variant rs7673984 across all cohorts. Studies listed above the dotted line are new cohorts not included in the only prior GWAS for corneal astigmatism (Fan et al., 2011). EAF = effect

allele frequency; OR = odds ratio. Horizontal bars denote 95% confidence intervals. rs7673984 was excluded from the Rotterdam-I cohort analysis during quality control filtering. .... 122

Figure 4.6: Power calculation. Power to detect variants with an odds ratio of 1.1 for this meta-analysed sample with MAFs ranging between 0.01 and 0.50. Based on a sample size of  $N = 31,370$ , a case:control ratio of 1:2.3, a trait population prevalence of 43.5% and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency. .... 124

Figure 4.7: Quantile-quantile plots for VEGAS2 and MAGMA gene-based analyses. Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected). .... 128

Figure 5.1: Estimating the proportion of  $h^2_{\text{SNP}}$  that is inflated due to uncorrected cryptic relatedness and population stratification.  $h^2_{\text{sep}}$  = SNP-heritability estimated from the individual chromosome tests;  $h^2_{\text{joint}}$  = Total SNP-heritability estimated from the joint test (all chromosomes pooled together). Blue lines denote linear regression lines. .... 144

Figure 5.2: Estimates of SNP-heritability ( $h^2_{\text{SNP}}$ ) using GCTA. Error bars represent the standard error of the  $h^2_{\text{SNP}}$  estimate. .... 149

Figure 6.1: Distribution of Corneal and Refractive Astigmatism in the analysed UK Biobank sample ( $N = 86,335$  and  $88,005$  respectively). Lower panels are boxplots representing their respective distributions (upper). Diamond symbol = mean astigmatism power. .... 160

Figure 6.2: Prevalence of with-the-rule and against-the-rule astigmatism as a function of age. Astigmatism was defined as with-the-rule when the negative cylindrical power was at an angle of  $0-30^\circ$  or  $150-180^\circ$ , against-the rule for angles  $60-120^\circ$ , and oblique for all other angles. The reverse applied for positive cylindrical powers. ATR = against the rule; WTR = with-the-rule. "NA" values denote individuals with no recorded astigmatism axis. Corneal astigmatism:  $N = 86,335$ ; Refractive Astigmatism:  $N = 88,005$ . .... 168

Figure 6.3: Corneal astigmatism, refractive astigmatism and their respective vector transformed values as a function of age in the analysed UK Biobank sample. Astigmatism magnitude = yellow squares,  $J_0$  (horizontal-vertical astigmatism component) = red circles;  $J_{45}$  (oblique astigmatism component) = orange triangles. Horizontal dashed lines denote Astigmatism Power = 0 D. Points denote mean values for each age group. Error bars denote 95% confidence intervals. .... 171

Figure 6.4: Manhattan and Quantile-Quantile plots for GWAS of Corneal Astigmatism using BOLT-LMM. Panel A: Manhattan plot: Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Panel B: Quantile-Quantile plot: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected). ..... 175

Figure 6.5. Regional association plots for loci demonstrating genome-wide significant association ( $P < 5 \times 10^{-8}$ ) in GWAS for Corneal Astigmatism using BOLT-LMM (N = 86,335). In order of chromosome (top row: rs12032649, rs196052; bottom row: rs1129038, rs62075722). Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple). NB: rs14879552 is a synonym for rs12032649. Larger versions of these individual plots are available in Appendix B. .... 176

Figure 6.6: Manhattan and Quantile-Quantile plots for GWAS of Refractive Astigmatism using BOLT-LMM. Panel A: Manhattan plot: Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Panel B: Quantile-Quantile plot: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected). ..... 180

Figure 6.7: Regional association plots for loci demonstrating genome-wide significant association ( $P < 5 \times 10^{-8}$ ) in GWAS for Refractive Astigmatism using BOLT-LMM (N = 88,005). In order of chromosome (top row: rs12196123, rs1129038; bottom row: rs34635363). Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple). Larger versions of these individual plots are available in Appendix B. .... 181

Figure 6.8: Manhattan and Quantile-Quantile plots for GWAS of Corneal and Refractive Astigmatism using BOLT-LMM as inverse normal transformed traits. Panels A and B: Corneal Astigmatism; Panels C and D: Refractive Astigmatism. Manhattan plots (A and C): Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Quantile-Quantile plots (B and D): Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected). ..... 185

Figure 6.9: Manhattan and Quantile-Quantile plots for GWAS of Corneal Astigmatism using PLINK 2.0 as continuous and dichotomous (cases  $\geq 1.00$  D) traits. Panels A and B: Continuous trait; Panels C and D: Dichotomous trait. Manhattan plots (A and C): Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates



$P = 1 \times 10^{-5}$ . Quantile-Quantile plots (B and D): Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected). ..... 189

Figure 6.10: Manhattan and Quantile-Quantile plots for GWAS of Refractive Astigmatism using PLINK 2.0 as continuous and dichotomous (cases  $\geq 1.00$  D) traits. Panels A and B: Continuous trait; Panels C and D: Dichotomous trait. Manhattan plots (A and C): Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Quantile-Quantile plots (B and D): Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected). ..... 192

Figure 6.11: Phenotypic correlations between pairs of refractive error traits. Panel A: Corneal Astigmatism and Refractive Astigmatism. Diagonal red line = line of unity (Corneal Astigmatism = Refractive Astigmatism); Panel B: Corneal Astigmatism and Spherical Equivalent. Horizontal dashed lines denote Spherical Equivalent = 0 D; Panel C: Refractive Astigmatism and Spherical Equivalent. Horizontal dashed lines denote Spherical Equivalent = 0 D. All correlations reported are Pearson correlations for respective pairs of refractive error measures in the 63,466 unrelated individuals included in GWAS for these respective traits who had data available for all three refractive error measures. .... 205

Figure 7.1: Comparison of  $-\log_{10}$  p-values from the gene-based analyses of the 111 candidate myopia genes using GWAS summary statistics from the CREAM consortium and UK Biobank with flanking regions of 0 kb, 50 kb and 200 kb. X-axes show expected negative  $\log_{10}$  p-values from analyses of CREAM consortium summary statistics. Y-axes show negative  $\log_{10}$  p-values from analyses of UK Biobank summary statistics. Red line: line of unity (CREAM  $-\log_{10}$  p-value = UK Biobank  $-\log_{10}$  p-value). Dashed lines denote separation between genes that demonstrated significant association using the CREAM consortium data (FDR < 0.05; right of lines) and all other candidate genes tested (left of lines). Genes tested for replication in the UK Biobank study are located to the right of the dashed lines. .... 225

Figure 7.2: Quantile-quantile plots of  $-\log_{10}$  p-values from the gene-based analyses of the 111 candidate myopia genes using GWAS summary statistics from the CREAM consortium and UK Biobank. Y-axes show observed negative  $\log_{10}$  p-values and X-axes show expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Black line: line of unity (observed = expected). ..... 228

## List of Abbreviations

95% CI	95% Confidence Interval
AD	Autosomal Dominant
AIM	Ancestry Informative Marker
ALSPAC	Avon Longitudinal Study of Parents and Children
ATR	Against-The-Rule (Astigmatism)
CA	Corneal Astigmatism
Chr	Chromosome
cM	Centimorgan
CNV	Copy Number Variant
COJO	COnditional and JOint
CREAM	Consortium for Refractive Error and Myopia
D	Dioptres
df	Degrees of Freedom
DZ	Dizygotic (twins)
EAF	Effect Allele Frequency
ECM	Extracellular Matrix
eQTL	Expression Quantitative Trait Locus
FDR	False Discovery Rate
GCTA	Genome-wide Complex Trait Analysis (software application)
GREML	Genetic relatedness matrix Restricted Maximum Likelihood
GRM	Genetic Relatedness Matrix
GTEx	Genotype-Tissue Expression

GWAS	Genome-wide Association Study
$H_0$	Null Hypothesis
$H^2$	Broad-sense Heritability
$h^2$	Narrow-sense Heritability
$h^2_{\text{GWAS}}$	Narrow-sense heritability estimates obtained using variants demonstrating genome-wide significant association in GWAS
$h^2_{\text{joint}}$	Total SNP-heritability estimated from the joint test
$h^2_{\text{sep}}$	Total SNP-heritability estimated from the individual chromosome tests
$h^2_{\text{SNP}}$	Narrow-sense heritability estimates obtained using commonly occurring genetic variants (SNP-heritability)
HRC	Haplotype Reference Consortium
HWE	Hardy-Weinberg Equilibrium
INDEL	Insertion/Deletion
INFO	Imputation Quality
INT	Inverse Normal Transformation
kb	Kilobase pairs
LD	Linkage Disequilibrium
LDSC	LD Score Regression (software application)
LMM	Linear Mixed Model
LOCO	Leave One Chromosome Out
LRT	Likelihood Ratio Test
MAF	Minor Allele Frequency
MAGMA	Multi-marker Analysis of GenoMic Annotation

Mb	Megabase pairs
MSE	Mean Spherical Equivalent
MSigDB	Molecular Signatures Database
MYP	Myopia
MZ	Monozygotic (twins)
N	Sample size
OMIM	Online Mendelian Inheritance in Man
OR	Odds Ratio
PCA	Principal Components Analysis
QC	Quality Control
QQ	Quantile-quantile
$r^2$	Square of the statistical correlation between two loci (measure of LD)
RA	Refractive Astigmatism
REML	Restricted Maximum Likelihood
SE	Standard Error
SMR	Summary data–based Mendelian randomisation
SNP	Single Nucleotide Polymorphism
UKBB	UK Biobank
$V_A$	Proportion of phenotypic variance attributed to additive genetic effects only
$V_C$	Proportion of phenotypic variance attributed to common environmental effects only
$V_{CR}$	Proportion of phenotypic variance attributed to cryptic relatedness only

$V_D$	Proportion of phenotypic variance attributed to dominance genetic effects only
$V_E$	Proportion of phenotypic variance attributed to unique environmental effects only
VEGAS	VErsatile Gene-based Association Study (software application)
$V_P$	Phenotypic variance
$V_{POP}$	Proportion of phenotypic variance attributed to population structure (cryptic relatedness and population stratification)
$V_{PS}$	Proportion of phenotypic variance attributed to population stratification only.
WTR	With-The-Rule (Astigmatism)
XWAS	X-chromosome Wide Association Study
$\beta$	Beta coefficient (Effect size) per unit predictor (e.g. per copy of effect allele)
$\lambda_{GC}$	Genomic Inflation Factor
$\lambda_{LDSC}$	Intercept from LD Score regression
$\chi^2$	Chi-squared

## Acknowledgements

I thank my supervisors Jez Guggenheim and Marcela Votruba for their contributions to my research. Thanks also to my advisor, Julie Albon, and my research group: Yu “Yvonne” Huang, Neema Ghorbani Mojarrad, Denis Plotnikov and Alfred Pozarickij.

Thanks also to my fellow postgraduate students and staff at the Cardiff School of Optometry and Vision Sciences for their advice and support, and for allowing me to distract them – especially those resident past and present in 2.11! I also thank Erik T. Viking for his never ending support and “assistance”.

Lastly, many thanks go to the Hong Kong Polytechnic University (PolyU) and Cardiff University for providing me with the opportunity and means to undertake this project, without which this thesis would not exist.

## Summary

Refractive errors such as myopia are the leading cause of reversible visual impairment worldwide with their prevalence rapidly increasing, resulting in greater burden on public health services. The aim of this series of investigations was to leverage the latest statistical methods and large-scale cohorts available in order to develop our understanding of the genetic determinants for the refractive error traits of spherical equivalent, corneal astigmatism and refractive astigmatism.

Investigation of genetic variants on the X-chromosome, a region often neglected in genome-wide association studies (GWAS), identified four genes demonstrating association in a gene-based analysis of spherical equivalent for a cohort of teenagers.

Meta-analysis of GWAS results for corneal astigmatism including European and Asian ancestry cohorts performed on behalf of the CREAM consortium successfully replicated the previously identified association near the *PDGFRA* gene (lead variant: rs7673984, odds ratio = 1.12,  $P = 5.55 \times 10^{-9}$ ).

The availability of data from the UK Biobank facilitated the largest GWAS for corneal and refractive astigmatism performed to date (N = 86,335 and 88,005 respectively). Here, GWAS for these traits identified four and two novel loci associated with corneal and refractive astigmatism respectively. Each of these loci had previously been associated with other ocular traits including myopia. Phenotypic variance explained by common genetic variants was relatively low for corneal and refractive astigmatism at ~6% and ~5% respectively, thus proposing a greater role for rare variants in explaining astigmatism variance due to genetics.

Lastly, in order to link identified variants and genes functionally influenced in myopia development, several candidate myopia genes identified from a primate myopia model demonstrated enrichment with refractive error associated variants in human samples.

Overall, the findings from these investigations are a starting point in guiding further research into the complex biological mechanisms underlying refractive error development.

# **Chapter 1 General Introduction**



## Chapter 1 General Introduction

### 1.1 Refractive errors

#### 1.1.1 *Definitions / classification of refractive errors*

There are three main classes of refractive errors: myopia, hyperopia and astigmatism.

##### 1.1.1.1 *Myopia and Hyperopia*

Myopia occurs when light focuses in front of the retina in the non-accommodated eye, whereas hyperopia occurs when light focuses behind the retina. In both instances, the refractive error results in impaired vision if uncorrected. This altered focusing is due to a mismatch between the power of the refractive components (primarily the cornea and lens) and the axial length of the eye (Young, Metlapally and Shay, 2007). Myopia is primarily caused by axial elongation of the eye (Grosvenor and Scott, 1993; Saw et al., 1996).

##### 1.1.1.2 *Astigmatism*

Astigmatism occurs when the eye fails to bring light from a point source object to a single point focus on the retina, resulting in impaired vision when viewing objects at any distance. If uncorrected in childhood, astigmatism is a risk factor for amblyopia development (Read, Collins and Carney, 2007; Harvey, 2009). In the human eye, the two major sources of astigmatism are the cornea and the crystalline lens. Astigmatism can be described as either “refractive”, which encompasses all contributing sources, or “corneal”, which is restricted to the corneal component,

the major contributing source in most cases of moderate and high astigmatism (Read et al., 2007).

#### *1.1.1.3 Mean Spherical Equivalent (also known as Spherical Equivalent)*

As astigmatism occurs in conjunction with myopia or hyperopia, clinicians often combine these two refractive error components into a single value known as the spherical equivalent refractive error. Spherical equivalent is the combination of the spherical and cylindrical components obtained from refraction, referring to the degree of myopia or hyperopia and the degree of astigmatism respectively. Spherical equivalent refractive error is calculated using the formula: spherical power +  $\frac{1}{2}$  cylindrical power.

#### **1.1.2 Worldwide prevalence**

Globally, uncorrected refractive error is the leading cause of reversible visual impairment and the second leading cause of blindness after cataract (Flaxman et al., 2017). The prevalence of refractive errors such as myopia and astigmatism has been on the increase over the last 50 years in the United Kingdom and especially in East Asia (He, Zheng and Xiang, 2009; Sun et al., 2012; O'Donoghue et al., 2015). Overall, it is projected that a third of the global population will have at least -0.50 Dioptres (D) of myopia by the year 2020, with this projected to increase to half of the global population by 2050 (Holden et al., 2016).

It should be noted that prevalence estimates tend to vary across ethnicities and with age, but also are dependent on trait definitions.

### *1.1.2.1 Myopia*

In a meta-analysis of refractive error prevalence studies in European adults over the age of 25 years, myopia of at least -0.75 D was found to be present in over 30% of the population (Williams et al., 2015b). The prevalence of high myopia (spherical equivalent  $\leq -6.00$  D) was 2.7% (Williams et al., 2015b). It was also noted that at least -0.75 D of myopia was present in almost half of those aged between 25 and 29 years of age (Williams et al., 2015b). On closer inspection of these cohorts, there was a noticeable cohort effect whereby myopia was more prevalent in cohorts of individuals born in more recent years than those of an equivalent age born in earlier decades (Williams et al., 2015a). As with these European ancestry cohorts, myopia was found to be as prevalent in a meta-analysis of Asian ancestry cohorts (~28%), with prevalence approaching 50% in the subgroup of individuals aged 20-29 years (Pan et al., 2015). However, these prevalence estimates were derived using a myopia threshold of -0.50 D, obtained from studies spanning the Asian continent, rather than restricting to Chinese or Japanese individuals in whom myopia tends to occur at a higher prevalence (Pan et al., 2015). No birth year (cohort) effect was identified by Pan et al. (2015). However, year-of-birth cohort effects similar to those identified in European ancestry individuals have been observed in studies from Singapore and Taiwan (Sherwin and Mackey, 2013).

In European ancestry children, the prevalence of myopia ( $< -0.50$  D) is relatively low at 1-5% at age 6-7 years (Ojaimi et al., 2005; O'Donoghue et al., 2010; Logan et al., 2011), increasing towards 20% by age 13 years (O'Donoghue et al., 2010; Logan et al., 2011). However, prevalence rates of myopia are approaching epidemic levels in

parts of East Asia (Morgan et al., 2018). In a study of Chinese school children, the prevalence of myopia of at least -0.50 D was low in those aged 5 years (~5%) with myopia prevalence increasing dramatically from age 8 years up to 15 years, affecting over 70% of 15 year-old children (He et al., 2004). Similar results were found in other cohorts of Chinese and Singaporean schoolchildren (Quek et al., 2004; Lam et al., 2012; Li et al., 2013; Guo et al., 2016). In studies of young adults (16-25 years old), myopia prevalence estimates were greater still, exceeding 90% in some instances (Wu et al., 2001; Jung et al., 2012; Lee et al., 2013a; Lee et al., 2013b).

#### *1.1.2.2 Astigmatism*

In a study of 6-year-old children based in Australia, the prevalence of refractive astigmatism of at least 1.00 D was 4.8% with greater prevalence and magnitude in children of Asian ancestry than those of European origin (11.2% vs. 3.6%) (Huynh et al., 2006). In a subsequent study of 12 year-old children in the same location, the prevalence of refractive astigmatism of at least 1.00 D was largely stable at 6.7%, and, as with the younger cohort, refractive astigmatism was more prevalent in those of East Asian descent (11.2% vs. 5.6%) (Huynh et al., 2007). There were no significant differences between males and females with respect to the prevalence of refractive astigmatism in the Huynh et al. study. In European ancestry children in Northern Ireland, the prevalence of refractive astigmatism ( $\geq 1.00$  D) was relatively stable between those aged 6 years and those aged 12 years, but was higher at 20-25% (O'Donoghue et al., 2011). However, this may be due to the increased prevalence of myopia in this cohort (Northern Ireland: 17.7% vs. Australia: 5.1%)

(Rose et al., 2008; O'Donoghue et al., 2011). In Chinese schoolchildren, refractive astigmatism of at least 0.75 D was found to be present in at least a quarter of children (He et al., 2004; He et al., 2007; Li et al., 2013).

In studies of adults, astigmatism of at least 1.00 D was found to be prevalent in 24% of European ancestry individuals, with prevalence increasing up to 50% in individuals aged 70 years and above (Williams et al., 2015b). Similar prevalence rates for those aged below and over 60 years were identified in a study of individuals resident in the United States (Vitale et al., 2008). In those of Asian ancestry, the prevalence of refractive astigmatism of > 1.00 D was 25-30% in individuals aged 40-59 years, with this level increasing in those over 60 years of age (Wong et al., 2000; Tan et al., 2011; Kim et al., 2013; Pan et al., 2013). As with children, there was little difference in prevalence of refractive astigmatism between males and females (Pan et al., 2013; Williams et al., 2015b).

### ***1.1.3 Genetic and environmental influences on refractive errors***

Whilst improved access to eye care services has resulted in greater reporting of refractive errors, thus accounting for some of the increased prevalence described above, environmental factors have been suggested to be the primary contributors to this rapid increase in prevalence rates (Holden et al., 2016).

Environmental influences known to influence myopia development include: higher educational attainment, reduced time spent outdoors and increased time spent reading; however, there is inconsistent evidence in support of the latter (Ip et al.,

2008; Williams et al., 2008b; Guggenheim et al., 2012; Sherwin et al., 2012; French et al., 2013; Li et al., 2015c; O'Donoghue et al., 2015; Cuellar-Partida et al., 2016; Shah et al., 2017). Understanding of the environmental influences on astigmatism development is less well understood than for myopia. Suggestions for potential causes include eyelid tension, which in turn alters the corneal shape and reduces with age, and visual feedback mechanisms, which may induce astigmatic changes during myopia or hyperopia development (Read et al., 2007).

With respect to genetics, early studies of refractive errors in twins suggested genetic factors make a notable contribution to the development of refractive errors (Sorsby, Sheridan and Leary, 1962; Sorsby and Fraser, 1964). Early studies identifying the role of genetics in astigmatism development came to mixed conclusions. Some studies suggested little or no involvement of genetic factors, while others suggesting a central role of genetic factors (Mash, Hegmann and Spivey, 1975; Teikari et al., 1989; Clementi et al., 1998). It should be noted that the differences in the genetic contribution to astigmatism may be due to phenotypic uncertainty within the respective studies. Attempts to ascertain the genetic contribution to refractive error, and identify genetic loci for refractive error development are described in *Section 1.4*.

#### **1.1.4 Current treatments**

Currently, the primary forms of managing refractive errors are through the optical correction methods of spectacles and contact lenses (Foster and Jiang, 2014). These forms of correction are well tolerated by wearers; however, they are not a

permanent solution and need to be replaced regularly. Alternatively, laser refractive surgery methods can provide a longer-term solution. However, there is a considerable cost implication to the individual and it is not suitable for all cases, especially when the refractive error has not yet stabilised.

It should be noted that the above mentioned methods do not rectify the axial elongation that commonly occurs with myopia (Gwiazda, 2009). As a result, individuals who use these forms of refractive correction are still at risk of potentially sight threatening conditions later in life that are due to this increased axial length. Such conditions include: primary open angle glaucoma, retinal detachment and myopic maculopathy (The Eye Disease Case-Control Study Group, 1993; Marcus et al., 2011; Flitcroft, 2012).

In recent years, clinical trials of myopia retardation strategies have had some successes. For instance, use of atropine eye drops has been trialled with notable success (Huang et al., 2016). In a meta-analysis of 16 myopia control interventions, high (0.5-1%), medium (0.1%) and low (0.01%) dose atropine eye drops demonstrated greatest reduction in myopia progression in terms of refractive error change and axial length elongation compared to single vision spectacle wearing controls during the treatment period (0.45-0.55 D/year and 0.14-0.21 mm/year respectively) with minimal difference between the three dose categories (Huang et al., 2016). Similar results were identified in a wider meta-analysis on the use of atropine eye drops for myopia control conducted by Pineles et al. (2017). Pineles et al. (2017) also examined reported effects after the cessation of treatment and

identified that medium-to-high dose atropine (0.1-1%) resulted in increased axial elongation 1 year after cessation in comparison to low dose atropine (0.01%) where such effects were minimal in comparison to untreated children. This post-treatment “rebound” effect, in combination with the reduced side-effects from using lower doses and its near-equivalent efficacy to higher doses during the treatment period, gives support in favour of using 0.01% atropine as a method of myopia control in children. Alternatively, use of orthokeratology contact lenses reduced axial elongation by approximately 0.36mm compared to controls wearing single vision spectacles during the 2-year treatment period in Chinese schoolchildren aged 6-10 years-old (Cho and Cheung, 2012). Similar results for orthokeratology were found in trials of other Asian and European cohorts as summarised by Li et al. (2016). Effects post-cessation of orthokeratology treatment are currently under investigation however, it has been suggested that axial elongation does resume upon cessation (Cho and Cheung, 2017).

Despite these successes, both orthokeratology and atropine treatment require compliance from children and their parents / guardians in order to maximise efficacy; complications can arise from either strategy such as ocular infections and photosensitivity; and lastly, there is still the potential for rebound effects whereby axial elongation resumes on cessation of the intervention (Chia et al., 2014; Cho and Cheung, 2017; Pineles et al., 2017). Randomised controlled trials whereby children have been allocated additional outdoor time in the school timetable have also shown promise in reducing the incidence of myopia in Chinese schoolchildren (Wu et al., 2013; He et al., 2015). Further trials based on the theme of increasing



outdoor exposure are in progress including one which aims to give schoolchildren levels of light exposure typically experienced outdoors within a classroom setting (Zhou et al., 2017b). However, the role of spending time outdoors in retarding progression in already myopic children is still unclear (Li et al., 2015b; Xiong et al., 2017).

#### **1.1.5 Why investigate refractive errors?**

As refractive errors have such a high prevalence in the population, especially for those of younger age who will require decades of management for these disorders, the causal factors contributing to refractive errors need to be found. This knowledge would allow at-risk individuals to be targeted for preventative treatment interventions, and enable improved preventative strategies to be developed. Ultimately, such approaches will lessen the burden of refractive errors in future generations.

### **1.2 Features of complex traits**

Traits are often described as being either Mendelian or complex (Lander and Schork, 1994). Mendelian traits follow the classical pattern of having dominant or recessive modes of inheritance and are typically caused by mutation at a single genetic locus resulting in impaired or loss of function of that gene. Although rare, high myopia, in the presence or absence of other ocular or systemic pathology, can sometimes be classed as a Mendelian trait (Tang, Yap and Yip, 2008). Complex traits on the other hand can be under the influence of multiple genetic loci, gene-gene and gene-environment interactions and non-genetic effects (Lander and Schork,

1994). Examples of complex traits include common trait variations within the population such as eye colour, height and refractive error. Following, are the primary genetic features that need to be considered when attempting to discover causal genetic features of complex traits.

### **1.2.1 Genetic variants**

The majority of phenotypic differences between individuals arise as a result of differences in our genomes. These genetic differences can occur in the form of single nucleotide substitutions in the genome, or more complex insertions/deletions of one or more nucleotides (Frazer et al., 2009).

The simplest and most frequently occurring type of genetic variant is a single nucleotide polymorphism (SNP) (Wang et al., 1998). This is where the commonly occurring nucleotide at a specific locus in the genetic sequence is substituted for an alternative nucleotide. For example, at a specific position in the human genome, most of the population may possess the 'A' nucleotide but some individuals may instead possess an alternative nucleotide at that position (e.g. 'T'). The different variant nucleotides are referred to as alleles. Often, SNPs do not cause changes to the amino acid sequence (so-called "synonymous" variants), however, there are many occasions where a SNP does alter the amino acid sequence ("non-synonymous" variants) (Hunt et al., 2009). This occurs if the variant alters a codon, a sequence of three nucleotides, to one that codes for an alternative amino acid. Whether synonymous or not, either type of SNP can result in subtle variations to an individual's phenotypes (Hunt et al., 2009). Millions of these polymorphisms are

known to exist throughout the human genome with many more yet to be discovered. SNPs are defined as being common if the rarer allele is present in at least 1% of the population (Wang et al., 1998).

Other more complex variations occurring in the genome include INDELS and copy number variations (CNVs). These sources account up to 20% of all variations in the human genome (Frazer et al., 2009). INDELS refer to insertions and/or deletions of nucleotides to the genetic sequence. These modifications can be as small as a single nucleotide or cover several hundred nucleotides (Mullaney et al., 2010). Unlike SNPs, INDELS have the potential to cause frame-shift mutations. Frame-shift mutations occur when the additional (or deleted) nucleotides are not a multiple of three (the length of a codon), resulting in a change to the codon sequence starting from the site of the INDEL, thus altering the amino acid sequence from this point and ultimately the functionality of the affected gene (Mullaney et al., 2010). Copy number variants (CNVs) are structural variations whereby there is either a deletion or replication of an extensive section of genomic sequence, usually of at least one kilobase (kb) in length (Redon et al., 2006). Such alterations in copy number can result in functional effects due to altered gene expression resulting from loss of one or both copies of a gene, or the presence of additional copies of that gene.

### ***1.2.2 Genotyping and Imputation***

Ascertaining the genotype of a specific individual at a particular genetic locus can be performed either directly (“genotyped”) or inferred (“imputed”). Large-scale direct genotyping is usually achieved by using SNP arrays, which are able to genotype

individuals at hundreds of thousands if not millions of SNP loci simultaneously (Rabbee and Speed, 2006). SNPs that “tag” complex variants such as INDELs are increasingly being included in genotyping arrays. Genotyping is not the same as reading the entire DNA sequence of an individual, as only known locations of genetic variation are assessed during genotyping, and therefore will not detect all SNP variants (LaFramboise, 2009). Sequencing of the whole genome is still, to date, more costly, and therefore array-based genotyping is still widely used for larger-scale studies in order to collect data about common variants spread throughout the human genome.

As current genotyping methods cover only a small fraction of all known variants, imputation methods can be applied in order to increase the number of variants available for phenotypic association testing (Howie, Donnelly and Marchini, 2009). Imputation is the inference of unknown genomic variants based on other known variants (Marchini and Howie, 2010). Reference panels are required for imputation with the most commonly-used panels being those from the International HapMap Consortium and the 1000 Genomes Project (International HapMap et al., 2010; The 1000 Genomes Project Consortium et al., 2012). These reference panels each contain genomic maps for hundreds, now thousands of individuals, with 2,504 genomes included in Phase 3 of the 1000 Genomes Project (Sudmant et al., 2015). These genomic maps spread across multiple population ancestries and were obtained through either genotyping or sequencing methods, respectively.

For each variant, imputation aims to identify which one of two alleles would most likely be present in a particular individual at that locus. Imputation software such as IMPUTE2 and MaCH essentially match haplotypes between the individuals requiring imputation and the reference panels. The alleles chosen at each untyped variant are those present on the best fitting reference haplotype (Howie et al., 2009; Li et al., 2010). One of the major limitations of imputation relates to variants with low minor allele frequency (MAF). These rarer variants occur less frequently in the reference panels and therefore, there is reduced accuracy when imputing these SNPs due to incomplete haplotype matching between the individual and the reference panels. To aid users, imputation software also provides metrics relating to the certainty of imputation for each variant. This enables researchers to exclude variants that are of poor imputation quality, as poorly imputed variants can affect the accuracy of subsequent association analyses.

### **1.2.3 *Linkage Disequilibrium***

As mentioned above, during imputation, untyped variants are inferred through knowledge of haplotypes (Marchini and Howie, 2010). These, in turn, arise through patterns of linkage disequilibrium (LD) throughout the genome (Wall and Pritchard, 2003).

Linkage disequilibrium refers to the non-random statistical correlation between two alleles (Goldstein and Weale, 2001; Slatkin, 2008). This means that if two variants are in LD with each other, and one of these variants has demonstrated significant association for a particular trait, association of the other variant to the trait can be

inferred with near certainty – and it may be that this alternative locus is the actual causal variant (Hirschhorn and Daly, 2005; Wang et al., 2010; Hormozdiari et al., 2015). This is because variants in strong LD are usually inherited together as part of the same “haplotype block” (Goldstein and Weale, 2001; Wall and Pritchard, 2003). Complete linkage equilibrium, on the other hand, occurs when the presence of an allele at one locus is completely independent of that at another locus.

The magnitude of LD between two loci can be determined by examining the relationship between the frequencies of alleles at these loci and the possible haplotypes (Wall and Pritchard, 2003). One commonly used measure of LD is the  $r^2$  value and it represents the square of the statistical correlation between the two loci (Devlin and Risch, 1995). Values range between one (complete linkage disequilibrium) and zero (complete linkage equilibrium). The  $r^2$  value is calculated by applying Equation 1.1.

---

**Equation 1.1: Calculating LD between two biallelic loci.**  $r^2$  = linkage disequilibrium between loci  $p$  and  $q$ ;  $p_A$  = frequency of allele  $A$  at the first locus ( $p$ );  $p_a$  = frequency of the alternative allele ( $a$ ) at the first locus ( $p$ );  $q_B$  = frequency of allele  $B$  at the second locus ( $q$ );  $q_b$  = frequency of the alternative allele ( $b$ ) at the second locus ( $q$ );  $x_{AB}$ ,  $x_{ab}$ ,  $x_{Ab}$  and  $x_{aB}$  refer to the probabilities of the four possible haplotypes ( $x$ ). Adapted from Devlin and Risch (1995)

### 1.3 Genetic investigations

Investigations on the genetic aspects of various traits including refractive errors have traditionally focused on family pedigrees or twins, often with a substantial

proportion of those sampled exhibiting the investigated trait. More recently, investigations have moved on to concentrate on population-based cohorts, representative of the local demographic, usually consisting of samples of unrelated individuals. In these latter investigations, the findings are a more accurate representation of traits in the general population; however, much larger study samples are required in order to have sufficient statistical power to detect variants associated with the trait of interest.

### **1.3.1 Family / twin studies**

#### *1.3.1.1 Heritability*

Family studies consist of groups of multi-generational (both parents and their offspring) or sib-ship (same generation) family units with comparisons made between members within each family (Chen et al., 2007a).

In sib-ship family structures, there is increased sharing of family environment during childhood, which, this could result in a greater proportion of phenotypic variance being attributed to genetic effects. This is likely due to misclassification of shared environmental effects as genetic effects and possible gene-environment interactions that may not have been accounted for (Tenesa and Haley, 2013). In both nuclear and sib-ship family studies, age-related effects within and between pairings need to be taken into consideration as these effects could affect heritability estimates.

Alternatively, twin-based studies use pairs of monozygotic (MZ) and dizygotic (DZ) twins to determine the influence of genetics on the trait of interest, whilst automatically eliminating the confounding effect of age within each pairing (Martin, Boomsma and Machin, 1997). Age-related effects between twin pairs, however, do need to be accounted for in these studies. On average, all genetic variants are shared between MZ twins and only half are shared between DZ twins (Boomsma, Busjahn and Peltonen, 2002). Thus, comparisons can be made between these two groups in order to estimate the proportion of variance attributed between genetic and environmental effects.

An important assumption made in twin studies of heritability is that of equal environments. It is assumed that each member of a twin pairing experiences identical conditions in their shared family environment and this assumption is applied equally for both MZ and DZ twins (Kendler et al., 1993; Derks, Dolan and Boomsma, 2006). Thus, when estimating heritability according to phenotypic correlation within twin pairs, a greater correlation between MZ twins versus DZ twins is assumed to be a result of genetic effects. Conversely, if the phenotypic correlation between MZ twins is less than half of that for DZ twins, common environmental factors are assumed to be the predominantly causal feature (Hill, Goddard and Visscher, 2008). The equal environment assumption is often considered flawed, however, as it is possible for each member of the twin pair to partake in separate activities differentially depending on whether the pair are MZ or DZ twins. As a result, there is the potential for actual environmental effects to be mistaken for genetic effects in heritability estimates obtained through twin studies



(Hettema, Neale and Kendler, 1995). Moreover, estimates of heritability using twin pairs have the potential to misallocate the variance due to epistatic or gene-environment interactions, which is typically assigned as additive genetic effects in these models.

In spite of these limitations, family and twin based studies are very useful for determining the breakdown of the causes of phenotypic variance into their respective genetic and environmental components. This is because the entire genomes are compared according to assumptions of the proportion of variants shared between members of a family according to relatedness. However, these studies can only estimate the proportion of phenotypic variance attributed to genetic and environmental causes and are unable to identify specific causal loci or environmental factors associated with phenotypic variation.

#### *1.3.1.2 Linkage studies*

Linkage analysis attempts to identify a genetic region that harbours a variant (or variants) that influence the investigated trait (Ferreira, 2004). In order to facilitate this, linkage studies typically investigate multi-generational families (pedigrees) where several members are affected by the trait. In these studies, the transmission of a trait and the causal genetic loci can be traced through the generations of the pedigree (Balding, 2006). Variants in strong LD are transmitted together from one generation to the next, with continuous genetic sequences usually transmitted together in an “LD block”. These blocks are separated by recombination events

whereby there is a crossing over of the alleles from a parent during transmission (Weiss and Clark, 2002).

Linkage analysis tests whether there is co-segregation of phenotype and a specified genetic polymorphism (Pulst, 1999). If co-segregation occurs more often than would be expected by chance it can be inferred that a causal variant for this trait likely resides nearby. Unfortunately, the regions identified through linkage studies often span across hundreds of genes, therefore it is typically not possible to narrow down identified regions to the level of a single gene (Boehnke, 1994). Instead, candidate gene studies need to be performed on selected genes within the linked regions.

Linkage studies are particularly strong at identifying rare variants of large effect that contribute to the trait since such variants would be commonly occurring in the investigated pedigree (Manolio et al., 2009). However, as these are rare variants, it is often the case that a susceptibility locus identified in one family may not demonstrate linkage in other families. Therefore, findings from a linkage study are often harder to replicate to a statistically significant level in further pedigrees.

In recent times, there has been a resurgence in the use of linkage studies to identify new susceptibility loci (Bailey-Wilson and Wilson, 2011; Ott, Wang and Leal, 2015). This has in part been due to the reduced costs of genetic sequencing, thus allowing for identification of specific, potentially causal variants within candidate loci, and improved ability to detect novel rare variants (Bailey-Wilson and Wilson, 2011).

### **1.3.2 Association Studies**

Whilst linkage and candidate gene studies have proved to be of value in identifying susceptibility loci, these studies are only able to identify rare causal loci that have a relatively large effect on phenotype (Hirschhorn and Daly, 2005). In addition, the identified loci only account for a small fraction of all of the heritability as estimated from family and twin based studies (Hirschhorn and Daly, 2005).

Over the last decade, advances in genotyping and imputation techniques have led to the use of genome-wide association studies (GWAS) as an alternative method for identifying novel susceptibility loci (Visscher et al., 2017). GWAS are hypothesis-free investigations, individually examining common variants throughout the genome for association with the trait of interest (Stranger, Stahl and Raj, 2011). GWAS are hypothesis free because all available loci are investigated systematically, with no prior assumption as to which loci are more likely to demonstrate significant association. As a greater proportion of the genome is assessed through this method than in candidate gene studies, there is increased opportunity to identify novel susceptibility loci (Stranger et al., 2011).

GWAS can be performed on either continuous or dichotomous traits, irrespective of the presumed mode of inheritance. Traditionally, an additive mode of inheritance is assumed for all variants when undertaking a GWAS, whereby the estimated phenotypic effect is doubled if an individual has two copies of a risk allele over having a single copy of that same risk allele at a particular locus. For example, if having one copy of a risk allele were estimated to change the refractive error of a

person by -0.10 D, having two copies of the same risk allele would be estimated to change the refractive error of that person by -0.20 D. However, if a particular variant has a dominant mode of inheritance resulting in the same phenotypic effect regardless of an individual having one or two copies of the risk allele, estimations of phenotypic effect per copy of that risk allele will be incorrect unless this alternative mode of inheritance is applied prior to analysis.

Whilst GWAS allow for the identification of new susceptibility loci at locations throughout the genome in a single analysis, there are many limitations to using such studies. For instance, association test p-values for individual variants to demonstrate significant genome-wide association are very stringent ( $P < 5 \times 10^{-8}$ ). This choice of statistical threshold takes into account the expected number of independent variants, the number of variants tested, and the patterns of LD between these variants (Risch and Merikangas, 1996; Dudbridge and Gusnanto, 2008; Hoggart et al., 2008). Therefore, large sample sizes are required in order to have sufficient statistical power to detect true associated variants that often have modest effect sizes (e.g. odds ratio < 1.5) (Risch and Merikangas, 1996; Hoggart et al., 2008). It is assumed that quantitative traits are normally distributed (Goh and Yap, 2009). If a trait is not normally distributed, a normalising transformation such as a “Box-Cox”, “logarithmic” or an “inverse normal” transformation can be applied prior to analysis (Goh and Yap, 2009; Buzkova, 2013). Transforming quantitative phenotypes can, however, make interpretation of measures of phenotypic effect relative to allelic frequency much harder, since these measures will no longer be in the dimensions of the original trait. Additional technical considerations also need to

be made when conducting GWAS. GWAS are sensitive to allele frequencies in the sample and therefore population stratification effects can result in spurious associations if not accounted for. There will be uncertainty in the true genotype present at imputed loci, as by definition imputed variants for an individual are inferred, thus resulting in the potential for some variants to be incorrectly imputed. This can be mitigated through quality control processes, whereby variants with low imputation quality scores can be excluded from analysis. Rare variants (MAF < 0.01) are often excluded from analyses since studies would lack sufficient power to identify true association of these rare variants. Yet collectively, these rare variants may provide a significant contribution to the investigated phenotype and therefore, their exclusion reduces the overall number of identifiable associated loci.

Individual GWAS have limited power to detect loci demonstrating significant association with their trait of interest. In order to enhance the ability to detect significant susceptibility loci, the results of multiple GWAS for a single trait from multiple samples are collated and analysed in meta-analyses (Evangelou and Ioannidis, 2013).

### **1.3.3 Heritability**

For any sample, phenotypic variation is due to a combination of genetic and environmental factors (Falconer and Mackay, 1996; Hill et al., 2008). Genetic factors can be divided into additive and non-additive (dominance and epistatic) effects, with environmental factors divided into common and unique environmental effects (Equation 1.2).

**Equation 1.2: Phenotypic variance ( $V_P$ ) as a sum of its parts:** additive genetic ( $V_A$ ) and non-additive ( $V_D$ ) genetic effects, common ( $V_C$ ) and unique ( $V_E$ ) environmental effects.

Additive genetic effects occur when the number of copies of an allele has a linear effect on an individual's phenotype, and each locus acts independently of the others. Non-additive genetic effects include dominance genetic effects, where the resultant effect is dependent on interactions between the pair of alleles at a single locus; and epistasis, where the effect of an allele at one locus is dependent on its interactions with alleles at other loci (Visscher, Hill and Wray, 2008).

Heritability is the proportion of phenotypic variation attributable to genetic factors (Visscher et al., 2008). This can be described in two ways: narrow-sense and broad-sense heritability. Narrow-sense heritability ( $h^2$ ) is the proportion of phenotypic variance attributed solely to additive effects (Equation 1.3), whereas broad-sense heritability ( $H^2$ ) is defined as the proportion of phenotypic variance attributed to all genetic effects. In most instances, the term heritability refers specifically to the narrow-sense variety ( $h^2$ ) (Falconer and Mackay, 1996).

—

**Equation 1.3: Narrow-sense heritability ( $h^2$ ) is the proportion of phenotypic variance ( $V_P$ ) that is attributed to additive genetic ( $V_A$ ) effects.**

By definition, heritability is population specific, as it is sensitive to environmental factors; thus it is not a constant value and can vary with age (Visscher et al., 2008).

Understanding the heritability of a trait is important as it gives an understanding as to the possible causes of variation of that trait in a specific population. For example, for a highly heritable trait such as human height (Macgregor et al., 2006; Yang et al., 2010), it can be inferred that much of the variation in height in a certain population is due to genetic variation within that population. Conversely, if a trait was to exhibit low heritability, it can be inferred that differing environmental (non-genetic) exposures within a population are mostly responsible for their phenotypic variation. This information can be used to direct further research into establishing the causal mechanisms of particular complex traits (Koran et al., 2014) or aid the development of strategies to predict individuals who are more likely to develop these traits (Sanfilippo et al., 2010).

Twin and family-based studies have been used to estimate trait heritability and partition trait variation into its component parts for decades. More recently, statistical methods have been developed to estimate heritability in large samples of unrelated individuals. Here, studies have estimated the proportion of phenotypic variance attributed to variants demonstrating genome-wide significant association in GWAS or from a wider selection of commonly occurring genetic variants (Visscher, 2008; Yang et al., 2010; Yang et al., 2011a). However, there is often a considerable discrepancy in heritability estimates obtained from these newer SNP-based methods and traditional twin / family study based methods. For example,

twin studies of human height have estimated heritability to be approximately 0.80 (Silventoinen et al., 2003). This is in contrast to the estimates of 0.16 and  $\sim 0.50$ , obtained respectively from variants that have demonstrated genome-wide significant association for this trait (Wood et al., 2014) or from a collection of commonly occurring variants irrespective of their strength of association (Yang et al., 2010; Wood et al., 2014). The discrepancy between heritability estimates for a trait from twin / family studies and from variants identified to demonstrate association with that trait is a common problem faced in the field of complex trait genetics and is described as the “missing heritability” (Maher, 2008; Manolio et al., 2009). It is important to note that heritability estimates obtained from twin / family studies include the contribution of all genetic variants, irrespective of their allele frequencies and effect sizes, whereas estimates obtained using variants identified using GWAS ( $h^2_{\text{GWAS}}$ ) only consider the contribution of a few variants to the phenotypic variance. This means that most variants contributing to the phenotype and its variance are not included when estimating  $h^2_{\text{GWAS}}$ , due to their low allele frequency or because they have effect sizes too small to demonstrate sufficiently strong association in GWAS (Pritchard, 2001; Bodmer and Bonilla, 2008; Frazer et al., 2009). In addition, the true causal marker may not be in complete linkage disequilibrium with any variants included in GWAS, thus resulting in the true marker’s contribution not being fully determined (Yang et al., 2010). As described above, an important assumption made in twin studies is the common environmental assumption. This assumption presumes that both individuals in pairs of monozygotic or dizygotic twins are exposed to identical environments (Kim et al., 2015), thus if monozygotic twins are more phenotypically correlated than dizygotic



twins, any differences between them must be due to genetic factors (Scarr and Carter-Saltzman, 1979; Kendler et al., 1993). However, this is often not the case as differences in environmental exposures have been reported to be greater between pairs of dizygotic twins than their monozygotic counterparts, as reviewed by Richardson and Norgate (2005). Further explanations include epigenetic variations, poorly tagged structural variants, gene-gene and gene-environment interactions, all of which are not directly considered in GWAS yet do contribute to the genetic component of phenotypic variance in twin / family studies (Johannes, Colot and Jansen, 2008; Frazer et al., 2009; Manolio et al., 2009; Eichler et al., 2010).

#### **1.4 What has been discovered so far by each of these methods?**

##### **1.4.1 Heritability**

With respect to refractive errors, twin and family based investigations have demonstrated spherical equivalent (and myopia) to be highly heritable (50-91%) and determined primarily by additive effects (Teikari et al., 1991; Hammond et al., 2001; Lyhne et al., 2001; Wojciechowski et al., 2005; Chen et al., 2007a; Klein et al., 2009; Lopes et al., 2009; Baird, Schache and Dirani, 2010; Kim et al., 2013). Genome-wide significant associations from GWAS for spherical equivalent and myopia, only account for 3-10% of the variance in spherical equivalent (Verhoeven et al., 2013; Miraldi Utz, 2017). However, investigation of commonly occurring variants in unrelated individuals estimated the heritability of refractive error to be approximately 0.35 (Guggenheim et al., 2015).

Twin and family studies of refractive and corneal astigmatism have been found each trait to be moderately heritable (50-65%) with this predominantly composed of dominance effects (Hammond et al., 2001; Dirani et al., 2006; Grijibovski et al., 2006); however, one study of corneal astigmatism in Koreans estimated heritability to be slightly greater (70%) and to be composed of solely additive genetic effects (Kim et al., 2013). Estimates of heritability for astigmatism using the newer SNP-based methods have not been published to date.

#### **1.4.2 Linkage studies**

Linkage studies of refractive errors have typically concentrated on high myopia (< -6.00 D). However, there are some studies that have investigated low-moderate myopia (< -1.00 D) (Stambolian et al., 2004; Stambolian et al., 2005; Chen et al., 2007b; Ciner et al., 2008; Schache et al., 2009) or spherical equivalent refractive error as a continuous trait (Hammond et al., 2004; Klein et al., 2007). To date, 25 loci have been designated as myopia (MYP) associated loci, as listed in the Online Mendelian Inheritance in Man (OMIM) database (URL: <https://omim.org/>) and summarised in Table 1.1.

**Table 1.1: Summary of myopia (MYP) loci listed in the Online Mendelian Inheritance in Man (OMIM) data**

<b>MYP Locus</b>	<b>Region</b>	<b>Low/High Myopia</b>	<b>References</b>
MYP1	Xq28	High	Guo et al. (2010); Ratnamala et al. (2011)
MYP2	18p11.31	High	Young et al. (1998b); Young et al. (2001)
MYP3	12q21-q23	High	Young et al. (1998a); Farbrother et al. (2004); Nurnbe
MYP5	17q21-q22	High	Paluru et al. (2003)
MYP6	22q13.33	Low	Stambolian et al. (2004); Klein et al. (2007); Tran-Viet
MYP7	11p13	Low	Hammond et al. (2004)
MYP8	3q26	Low	Hammond et al. (2004); Andrew et al. (2008)
MYP9	4q12	Low	Hammond et al. (2004)
MYP10	8p23	Low	Hammond et al. (2004); Stambolian et al. (2005)
MYP11	4q22-q27	High	Zhang et al. (2005)
MYP12	2q37.1	Low/High	Paluru et al. (2005); Chen et al. (2007b); Schache et a
MYP13	Xq23-q27.2	High	Zhang et al. (2006); Zhang et al. (2007)
MYP14	1p36	Low	Wojciechowski et al. (2006)
MYP15	10q21.1	High	Nallasamy et al. (2007)
MYP16	5p15.33-p15.2	High	Lam et al. (2008)
MYP17 (MYP4)	7p15	Low/High	Ciner et al. (2008); Paget et al. (2008)
MYP18	14q22.1-q24.2	High	Yang et al. (2009)
MYP19	5p15.1-p13.3	High	Ma et al. (2010)
MYP20	13q12.12	High	Shi et al. (2011b)
MYP21	1p22.2	High	Shi et al. (2011a); Tran-Viet et al. (2012); Xiang et al. (
MYP22	4q35.1	High	Zhao et al. (2013)
MYP23	4p16.3	High	Aldahmesh et al. (2013); Jiang et al. (2014)
MYP24	12q13.3	High	Guo et al. (2014); Jiang et al. (2014)
MYP25	5q31.1	High	Guo et al. (2015)
MYP26	Xq13.1	High	Xiao et al. (2016)

Using a combination of linkage analysis and whole-exome sequencing in recent years, high myopia susceptibility loci have been identified at the genes, *SLC39A5* and *P4HA2* in Chinese families with autosomal dominant (AD) non-syndromic high myopia (Guo et al., 2014; Guo et al., 2015). These loci have since been designated as MYP24 and MYP25 respectively (Table 1.1). The increasing number of potential susceptibility loci has added much complexity to understanding the polygenic nature of myopia (Farbrother et al., 2004).

Linkage studies have not been performed for astigmatism, despite the suggestion that such studies may be useful in identifying astigmatism susceptibility loci (Clementi et al., 1998; Dirani et al., 2008)

### **1.4.3 Genome-wide Association Studies (GWAS)**

#### **1.4.3.1 Spherical Equivalent / Myopia**

Early GWAS for refractive errors were conducted in relatively small samples of Asian ancestry. In these studies, potential high myopia susceptibility loci were identified near the *BLID* gene and in the region of the *CTNND2* gene (Nakanishi et al., 2009; Li et al., 2011a). However, these loci did not reach genome-wide significance. It should be noted that this shortfall in statistical significance is likely to be due to the size of the cohorts in each study (N = 4,155 and 2,741 respectively). Li et al. (2011b) identified a genome-wide significant locus for high myopia in the MYP11 myopia linkage locus in their Asian ancestry cohort and two discrete loci (near the *GJD2* and *RASGRF1* genes) demonstrating genome-wide significant association with myopia

were identified on chromosome 15 in separate studies of European ancestry cohorts (Hysi et al., 2010; Solouki et al., 2010).

In larger scale studies, Kiefer et al. (2013) conducted a GWAS for refractive error including over 45,000 customers from the database of the 23andMe direct-to-consumer genetic testing company. For the purposes of their analysis, refractive error was considered as a binary trait with patients reporting if they had been diagnosed as being near-sighted and their age at diagnosis, if applicable. Twenty-two variants were identified to demonstrate genome-wide significant association with the age of myopia onset; all but two of these were not identified in previous studies. In a smaller replication dataset of 8,323 individuals, half of these variants were found to demonstrate significant association. Of the 22 variants identified in the initial cohort, most of these were in or near genes responsible for extracellular matrix (ECM) structure, photoreceptor function, ocular growth, neuronal development and signalling pathways. In turn, altered function of these genes could affect the normal course of ocular development, resulting in changes in cell proliferation, axial length and other structural abnormalities. A similarly sized meta-analysis conducted by the Consortium of Refractive Error and Myopia (CREAM) included 32 population-based cohorts of European and Asian ancestries (total 45,931 participants: 36,636 of European ancestry and 9,295 of Asian ancestry) (Verhoeven et al., 2013). This study identified 24 novel loci demonstrating association with refractive error, with many of these loci identified in both European ancestry and Asian ancestry individuals. The findings from this analysis showed considerable overlap with the GWAS conducted independently and almost

concurrently by the 23andMe group (Kiefer et al., 2013), with 25 loci identified as demonstrating potential association with myopia susceptibility in both studies, despite being conducted through differing methodologies (Kiefer et al., 2013; Verhoeven et al., 2013). A more recent GWAS for near-sightedness conducted by Pickrell et al. (2016) identified 183 associated loci in their sample of almost 200,000 individuals registered on the 23andMe database. However, their study only reported the marker demonstrating strongest association at the top 50 loci of these 183 loci.

#### 1.4.3.2 Astigmatism

Early GWAS for corneal astigmatism showed promise, with the identification of a single locus in the promoter region of *PDGFRA* on chromosome 4 (lead variant: rs7677751,  $P = 7.87 \times 10^{-9}$ ) associated with this trait in 8,513 individuals of Asian ancestry by Fan et al. (2011). Association of the *PDGFRA* region with corneal astigmatism was replicated in a sample of European children (N = 1,968; Guggenheim et al. (2013a)). Variants in the *PDGFRA* region have also demonstrated association in studies of corneal curvature (a trait from which corneal astigmatism is derived) in this sample of European children, in Australians of European ancestry (N = 2,801; Mishra et al. (2012)) and in 10,008 Asian ancestry individuals (Han et al., 2011). Furthermore, variants at the *FRAP1* gene locus on chromosome 1 have demonstrated genome-wide significant association with corneal curvature in this Asian ancestry sample (Han et al., 2011).

The above studies were relatively small scale and therefore had low statistical power to detect variants of small effect size at genome-wide significance. In order to improve the ability to identify associated genetic variants of small effect size, meta-analyses of GWAS for refractive astigmatism, a trait with greater volume of data available from research studies, have been performed in recent years. However, these analyses have not been as fruitful as anticipated. Lopes et al. (2013) identified a susceptibility locus near the *VAX2* gene in their meta-analysis of 22,100 European ancestry individuals. Despite the known function of *VAX2* with respect to early ocular development, giving support to the notion of it being an astigmatism susceptibility locus (Barbieri et al., 1999; Zagozewski et al., 2014), this locus did not demonstrate genome-wide significant association. Furthermore, a larger-scale GWAS meta-analysis for refractive astigmatism conducted by CREAM with 45,931 participants (36,636 of European ancestry and 9,295 of Asian ancestry), identified only a single locus demonstrating genome-wide significant association, located near the *NRXN1* gene on chromosome 2 (lead variant: rs1401327,  $P = 3.92 \times 10^{-8}$ ) (Li et al., 2015a). These results are in contrast to those for the highly heritable refractive traits of spherical equivalent and myopia, which have identified nearly two hundred variants demonstrating genome-wide significant association (Kiefer et al., 2013; Verhoeven et al., 2013; Pickrell et al., 2016).

## **1.5 Overall Aims**

By developing our understanding of the genetic determinants of refractive errors and the biological processes that underlie their development, it is anticipated that more effective therapeutic and preventative strategies can be developed, with the

possibility of tailoring strategies to individual needs. Furthermore, models predicting at-risk individuals can be developed and improved through improved understanding of the genetic architecture underlying these traits.

The aim of the following series of investigations was to leverage the latest statistical methods and large-scale cohorts available in order to further understand the genetic determinants for the refractive error traits of spherical equivalent, corneal astigmatism and refractive astigmatism.

Specifically, studies were conducted with the following aims:

- Examine genetic variants on the X-chromosome for association with spherical equivalent refractive error in a cohort of teenage children.
- Conduct a meta-analysis of GWAS results for corneal astigmatism including European ancestry and Asian ancestry cohorts (on behalf of the CREAM consortium).
- Utilise data from the UK Biobank study to estimate SNP-heritability for spherical equivalent, corneal astigmatism and refractive astigmatism.
- Conduct GWAS for corneal astigmatism and refractive astigmatism using data from the UK Biobank study.
- Investigate whether candidate myopia genes identified from a primate myopia model are enriched with refractive error associated variants in human samples.



## **Chapter 2 General Methods**

## **Chapter 2 General Methods (Study cohorts, tools and models used in these investigations)**

### **2.1 ALSPAC cohort of children and their mothers**

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a population-based birth cohort study (Boyd et al., 2013). Pregnant women with expected delivery dates between 1st April 1991 and 31st December 1992, residing in the former county of Avon, UK, were eligible for enrolment. This cohort consisted of 14,541 pregnancies, from which there were 13,988 children alive at 1 year of age. All participants provided informed consent. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. The ALSPAC website contains details of all the data that is available through a fully searchable data dictionary (URL: <http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>).

Phenotyping, genotyping and imputation were carried out by members of the ALSPAC team.

#### **2.1.1 Phenotypes**

Non-cycloplegic autorefraction (Canon R50 instrument, Canon USA Inc., Lake Success, NY) was performed at clinic visits scheduled at ages 7, 10, 11, 12 and 15 years. Mean spherical equivalent (MSE) was averaged for the two eyes. Outlier autorefraction values were removed as described by Northstone et al. (2013) and the child's precise age was recorded at each clinic visit. As outlined by Guggenheim et al. (2013a), corneal astigmatism was calculated from corneal curvature

measurements (Zeiss IOLmaster instrument, Carl Zeiss Meditec, Welwyn Garden City, UK) obtained at the clinic visit scheduled at age 15 years. Corneal curvature measurements were introduced towards the latter stages of the age 15 year assessment visits; hence only a minority of participants underwent this test.

### **2.1.2 Genotyping and Imputation**

ALSPAC children were genotyped using the Illumina HumanHap550 quad chip. Following quality control (individual call rate > 0.97, SNP call rate > 0.95, minor allele frequency (MAF) > 0.01, Hardy-Weinberg equilibrium (HWE) p-value >  $1 \times 10^{-7}$ , non-European clustering individuals removed), 477,482 SNP genotypes were available for 8,935 children after excluding participants who had withdrawn consent. Haplotypes were phased (using a combined sample of these children and their mothers, when the mother's genotype data was available) using ShapeIT (v2.r644). Imputation was performed using IMPUTE v2.2.2 against all 2,186 reference haplotypes (including non-Europeans) in the Dec 2013 release of the 1000 Genomes reference haplotypes (Phase 1 Version 3). All variants were mapped to NCBI human genome build 37 (hg19/GRCh37) coordinates.

## **2.2 UK Biobank Study**

UK Biobank is a prospective study following the health and wellbeing of 502,633 participants resident in the UK aged between 40 and 69 years-old at the baseline recruitment visit (during the period 2006 to 2010). UK Biobank received ethical approval from the NHS Research Ethics Committee (Reference: 11/NW/0382). Baseline assessment was undertaken at one of 22 assessment centres distributed

across the UK (Allen et al., 2014; Sudlow et al., 2015). Approximately 20,000 participants also attended the first repeat assessment visit (during the period 2012 to 2013). Demographic information and medical history were ascertained through touch-screen questionnaires. Participants also underwent a wide range of physical and cognitive assessments, including blood sampling (for DNA) and, for participants recruited towards the end of the recruitment period, an ophthalmic examination. Phenotyping, genotyping and imputation were carried out by members of the UK Biobank team.

### **2.2.1 Phenotypes**

119,806 participants had keratometry readings taken for at least one eye using the Tomey RC 5000 autorefractor-keratometer (Tomey Corp., Nagoya, Japan). Up to six measurements were taken for each eye using 6mm diameter keratometry mires, from which corneal astigmatism was derived (see below). 130,521 participants had non-cycloplegic autorefraction performed for at least one eye using the same autorefractor-keratometer, with up to ten measurements taken for each eye. Refractive astigmatism was derived from the autorefraction cylindrical power. Spherical equivalent was recorded as the spherical power plus half of the cylindrical power from autorefraction. All keratometry/autorefractor measurements flagged with an error code 'E' (indicating 'Lower reliability data') were recoded as missing before taking the mean trait values for each eye individually across assessment centre visits, then the mean of both eyes for each individual. The mean corneal astigmatism and mean refractive astigmatism for each individual were also categorised as dichotomous variables using a grid of thresholds to define

case/control status, from 0.50 to 1.50 D, in 0.25 D steps. After the exclusion of unreliable readings, 119,799 participants had measures for corneal astigmatism, and 130,459 participants had measures for refractive astigmatism and spherical equivalent refractive error.

### **2.2.2 Genotyping and Imputation**

Participant DNA samples were genotyped by UK Biobank researchers at approximately 800,000 genetic variants using one of two genotyping arrays, the UK BiLEVE Axiom array or the UK Biobank Axiom array. Genetic data were released in two waves. In the UK Biobank “Interim 150K” release, data were made available for 152,725 samples imputed at 72,355,689 variants using IMPUTE2 (Howie, Marchini and Stephens, 2011) with a merged 1000 Genomes Project Phase 3 and UK10K Project haplotype reference panel (The UK10K Consortium et al., 2015; Wain et al., 2015; Davies et al., 2016). Further details of the imputation protocol can be found at the following URL: <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=157020>.

Of these 152,725 genotyped participants, 141,751 were of European ancestry based on principal components analysis (PCA), and were non-outliers for heterozygosity (defined as autosomal heterozygosity within four standard deviations of the mean for the full set of European ancestry samples). Data for these individuals were taken forward for SNP-heritability estimation.

The second wave of genetic data released from the UK Biobank consisted of information for all participants whose data passed quality control filters, resulting in

genotype information for 488,377 participants. Imputation for the second wave was performed prior to release as described by Bycroft et al. (2017). Briefly, imputation was carried out using IMPUTE4, an updated version of IMPUTE2 (Howie et al., 2011; Bycroft et al., 2017) with a reference panel comprising of the Haplotype Reference Consortium (HRC) reference panel and a merged 1000 Genomes Project Phase 3 and UK10K Project haplotype reference panel (The UK10K Consortium et al., 2015; McCarthy et al., 2016; Bycroft et al., 2017). Due to uncertainty about the reliability of the 1000 Genomes and UK10K imputations, for the present work only the ~40 million variants present in the HRC imputation panel were utilised. All variants were mapped to NCBI human genome build 37 (hg19/GRCh37) coordinates.

Of the 488,377 genotyped participants, 409,728 self-reported White British ancestry and were non-outliers for heterozygosity (heterozygosity within four standard deviations of the mean of the White British ancestry subset defined by Bycroft et al. (2017)). These individuals were taken forward for the GWAS analyses.

Table 2.1 summarises the number of participants in the UK Biobank study with keratometry, autorefraction and genotype data available.

**Table 2.1: Summary of the UK Biobank Sample.**

	<b>Before QC</b>	<b>After QC</b>
Total Sample	502,633	-
Keratometry performed	119,806	119,799
Autorefractometry performed	130,521	130,459
Genotype data available (interim release)	152,725	141,751 <sup>a</sup>
Genotype data available (full release)	488,377	409,728 <sup>b</sup>

<sup>a</sup> After restricting to European ancestry based on PCA and non-outliers for heterozygosity

<sup>b</sup> After restricting to self-reported White British ancestry and non-outlier for heterozygosity

### **2.2.3 Exclusion Criteria**

In order to minimise the effects of ocular pathology or surgery affecting keratometry/autorefractometry readings, individuals were excluded from analyses if at any visit, they self-reported having had any ocular conditions or surgery as listed in Table 2.2. Note that individuals may have met multiple criteria for exclusion. Further exclusions on the basis of extreme phenotype values or anisometropia (differences between eyes within an individual) were not made in the primary analyses.

**Table 2.2: Summary of exclusions from the UK Biobank sample due to ocular conditions or surgery.**

		Corneal Astigmatism	Refractive Astigmatism Spherical Equivalent
<b>Individuals with phenotype data available<sup>a</sup></b>		<b>119,799</b>	<b>130,450</b>
<b>Exclusion Criteria</b>	<b>UK Biobank Field Code</b>	<b>No. of individuals excluded<sup>b</sup></b>	
Any injury or trauma resulting in loss of vision	6148	841	932
Cataract extraction/lens implant	20004	1,658	1,832
Glaucoma surgery/trabeculectomy	20004	149	173
Any eye surgery in the last four weeks	5181	7	8
Cataract surgery <sup>c</sup>	5324	2,863	3,156
Refractive laser eye surgery <sup>c</sup>	5325	2,822	3,010
Surgery for glaucoma or high eye pressure <sup>c</sup>	5326	208	252
Corneal graft surgery <sup>c</sup>	5327	274	308
Cataract	6148	-	7,462
Other serious eye condition	6148	-	2,749
Retinal operation/vitreotomy	20004	-	874
Any eye surgery	20004	-	1,130
<b>Total Excluded<sup>b</sup></b>		<b>6,610</b>	<b>14,413</b>
<b>Individuals remaining</b>		<b>113,189</b>	<b>116,037</b>

<sup>a</sup> After excluding for unreliable readings (see *Section 2.2.1*)

<sup>b</sup> Individuals may have met multiple exclusion criteria

<sup>c</sup> Individuals responding “Don’t know” to these questions were also excluded



#### **2.2.4 Covariates**

For all analyses using the UK Biobank data sets, the mean age of the participant at the assessment centre visit(s) when eye measures were recorded, was included as a quantitative covariate. Mean spherical equivalent was included as an additional quantitative covariate for all analyses of corneal and refractive astigmatism. Genotyping array (0 = BiLEVE, 1 = UK Biobank Axiom) and sex (0 = female, 1 = male) were included as binary covariates. The sex of each individual was matched between self-reported and genetically inferred sex values. If these values did not match, sex was recoded as missing for that individual. Individuals with missing values for either the investigated trait or any of the covariates were excluded from analysis.

#### **2.2.5 “High-confidence” Variants**

For the mixed model analyses carried out using BOLT-LMM and GCTA (*Chapters 5 and 6*), a set of approximately 890,000 “high-confidence” variants in LD was generated using PLINK 2.0 (Chang et al., 2015). All variants with an “rs” prefix that were directly genotyped or imputed in at least 99% of individuals, with minor allele frequency (MAF) > 0.005 and imputation quality (INFO) > 0.90 were LD-pruned to obtain list of variants for creating genetic relatedness matrices (GRMs). This LD-pruning was performed using the PLINK command [–indep-pairwise 50 5 0.1] whereby  $r^2$  values between pairs of markers within 50-marker-wide windows were assessed, with one marker from the pair removed if  $r^2 > 0.1$ . Window start positions were located in 5-marker steps.

## **2.3 Statistical Methods**

Throughout this series of investigations, a variety of statistical methods have been utilised. This section outlines the overall purpose of these tests and the software applications used to perform them.

### **2.3.1 Single Marker Tests**

Several software applications are available to perform single marker tests, commonly known as genome-wide association studies (GWAS) when all autosomes are investigated. Commonly used applications that were implemented in this series of investigations include SNPTEST v2.5, PLINK v1.9/2.0 and BOLT-LMM v2.3 (Purcell et al., 2007; Marchini and Howie, 2010; Chang et al., 2015; Loh et al., 2015; Loh et al., 2018). For all single marker association tests performed here, they used linear or logistic regression approaches or, as seen in *Chapter 6*, newer, mixed linear models.

#### *2.3.1.1 Linear and Logistic Regression*

Linear and logistic regression models are the simplest methods of performing single marker association tests whilst allowing the inclusion of covariates (such as age, sex or ethnicity). These covariates are likely predictors of the trait of interest and may also share independent genetic associations with it; therefore they must be accounted for in order to prevent confounding when testing for associations of genetic variants to the trait of interest (Clarke et al., 2011; Bush and Moore, 2012). Linear regression was employed for continuous traits with effect sizes reported as the average increase (or decrease) in trait value for each copy of the “effect” allele possessed ( $\beta_g$ ).

**Equation 2.1: Linear Regression Equation.**  $Y$  = expected outcome (trait);  $\beta_0$  = intercept of linear regression line;  $\beta_g X_g$  = effect size per copy of the effect allele  $g$ ;  $\beta_c X_c$  = effect size per unit of the covariate  $c$  (e.g. age, sex);  $\varepsilon$  = residual error term.

Conversely, for dichotomous traits, logistic regression was employed as there are only two trait classifications, with variant effects reported as odds ratios. Odds ratios state the average factor by which the likelihood of an individual having case status increases (or decreases) for each copy of the “effect” allele possessed (Equation 2.2). If the likelihood of having case status is equal to that for being unaffected for a given variant, the odds ratio will be 1.

---

**Equation 2.2: Odds Ratio for a given genotype.**  $P_{case}$  = probability of being a case;  $P_{control}$  = probability of being a control.

Rather than estimating the expected trait value for a given genotype as is the case for linear regression, logistic regression estimates the natural logarithm ( $\ln$ ) of the odds ratio, given the predictors (Equation 2.3; Lever, Krzywinski and Altman (2016)).

---

**Equation 2.3: Logistic Regression Equation.**  $\ln(P_{case}/P_{control})$  = log odds ratio given the predictor ( $X$ );  $\beta_0$  = intercept of logistic regression line;  $\beta_g X_g$  = effect size per copy of the effect allele  $g$ ;  $\beta_c X_c$  = effect size per unit of the covariate  $c$  (e.g. age, sex);  $\varepsilon$  = residual error term. Adapted from Lever et al. (2016).

In all instances, an additive model has been applied whereby reported effects are an average per copy of the effect allele possessed (i.e. if  $\beta$  is the magnitude of effect of a particular variant if an individual were heterozygous for this variant, then the magnitude of effect would be  $2\beta$  for an individual who is homozygous for this same allele).

### 2.3.1.2 Mixed Linear Models

As implemented in BOLT-LMM v2.3 (Loh et al., 2015; Loh et al., 2018), mixed linear models are an alternative to standard linear regression models for single marker association tests.

**Equation 2.4: Mixed Linear Model.**  $Y$  = expected outcome (trait);  $\beta_0$  = intercept of linear regression line;  $\beta_g X_g$  = effect size per copy of the effect allele  $g$ ;  $\beta_c X_c$  = effect size per unit of the covariate  $c$  (e.g. age, sex);  $G$  = genetic effects (included as a random effect);  $\varepsilon$  = residual error term.

Mixed linear models share features of standard linear regression models (Equation 2.4 vs. Equation 2.1), with the terms from linear regression forming the fixed effects portion of the linear mixed model. However, a key advantage of mixed linear models over linear regression is that residual population stratification and relatedness within the study sample can be accounted for as random effects (term  $G$  in Equation 2.4), which (if otherwise left unaccounted for) can lead to reduced power to detect associations or an excess of false positive association signals (Yang et al., 2014). This consideration of population effects therefore facilitates the

inclusion of related individuals who would traditionally be excluded from analyses, thus improving association study power.

Mixed linear models traditionally use pre-constructed GRMs in order to account for this residual population structure (Yang et al., 2014); however, BOLT-LMM on the other hand computes the parameters required for the random effects portion of the model during analyses. This improves computational efficiency in part due to the leave-one-chromosome-out (LOCO) procedure implemented by BOLT-LMM, which would otherwise be computationally intensive if multiple GRMs were required to be stored in memory (as is the case for the alternative software application GCTA-LOCO (Yang et al., 2014)). In LOCO analyses, variants situated on the same chromosome as the variant tested are not included in the computation of the random effects portion of the model (Yang et al., 2014). This is because, the variant tested would be considered twice in the model – once as a fixed effect (by default as the tested variant) and secondly as a random effect (either directly, or indirectly through variants in LD acting as proxies). This over-fitting would dampen the true association signal of the tested variant and therefore, use of the LOCO procedure improves the power to identify associations.

BOLT-LMM has the ability to employ two different models when performing single marker association tests; the standard “infinitesimal” and the Bayesian “non-infinitesimal” models (Loh et al., 2015). The “infinitesimal” model assumes that all variants are causal and are of small, normally distributed effect sizes. This model is equivalent to alternative mixed linear models used for single marker association

tests. The “non-infinitesimal” model on the other hand is a more complex model and is only applied if BOLT-LMM computes that there will be an expected improvement in performance compared to the standard “infinitesimal” model (Loh et al., 2015). This Bayesian model assumes effect sizes of variants do not follow a pre-defined single normal distribution, as is the case with frequentist tests (i.e. standard linear/logistic regression or the BOLT-LMM “infinitesimal” model), but rather they can be fitted to two normal distributions that consider the presence of a small number of variants of large effect, with the remaining variants assumed to be of small effect size. These probability distributions are calculated by the software based on the input data and parameters from these distributions are subsequently incorporated into their mixed linear models (Stephens and Balding, 2009). For all single marker association tests performed using BOLT-LMM, the standard “infinitesimal” model was used as the software computed there would be no gain in performance from the “non-infinitesimal” model.

### **2.3.2 Genomic Control**

It is important to take into consideration sources of genomic inflation when performing any genetic association test. Genomic inflation occurs when association test p-values are more significant than would be expected under the null hypothesis of uniformly distributed p-values obtained by chance (Devlin and Roeder, 1999; Zheng, Freidlin and Gastwirth, 2006). Sources of genomic inflation include genotyping errors, unadjusted population stratification and relatedness between sample individuals, all of which can result in false associations; and true polygenicity of the investigated trait (Yang et al., 2011c).

### 2.3.2.1 Genomic Inflation Factor ( $\lambda_{GC}$ )

For each GWAS performed, the genomic inflation factor ( $\lambda_{GC}$ ) was determined by taking the median  $\chi^2$  test statistic observed and dividing it by the expected median  $\chi^2$  test statistic of  $\sim 0.456$  (Equation 2.5) (Devlin, Roeder and Wasserman, 2001).

---

### Equation 2.5: Calculating genomic inflation factor ( $\lambda_{GC}$ ) (Devlin et al., 2001).

In the event that  $\lambda_{GC} > 1$ , studies may decide to adjust their association test statistics in order to account for this inflation by dividing all association test  $\chi^2$  statistics by  $\lambda_{GC}$  (Hinrichs, Larkin and Suarez, 2009). In the quality control protocol outlined by Winkler et al. (2014) and implemented in the software package EasyQC, studies with  $\lambda_{GC} > 1.1$  were deemed to be in need of reviewing with respect to unadjusted population structure. However, adjustment of results using  $\lambda_{GC}$  can be overly conservative in cases of true polygenicity (Bulik-Sullivan et al., 2015b).

### 2.3.2.2 LD Score Regression Intercept ( $\lambda_{LDSC}$ )

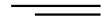
As an alternative method to overcome this problem of assessing genomic inflation in the presence of polygenicity, the software application LD Score regression (LDSC) uses single marker association test statistics to calculate an alternative genomic inflation statistic,  $\lambda_{LDSC}$  (Bulik-Sullivan et al., 2015b). This alternative statistic is obtained from the intercept of its regression analysis and is an estimate of inflation in association test statistics due to residual population structure or genotyping

errors. In order to perform this LD Score regression, reference LD scores were obtained using European ancestry individuals from the 1000 Genomes Project but restricted to variants in the HapMap3 reference panel, since these variants were known to be commonly occurring and well imputed (International HapMap et al., 2010; Bulik-Sullivan et al., 2015b).

The LD score for a variant is defined as the sum of pairwise  $r^2$  values between itself and all variants within 1cM of that variant. As multiple variants may have equal LD scores, variants were grouped into quantiles with the mean LD score of the variants within that quantile regressed against the mean  $\chi^2$  test statistic of those same variants. From this regression, the intercept of the regression line (i.e. the expected mean  $\chi^2$  test statistic in the event of a marker in complete linkage equilibrium with all other markers within 1cM), is defined as the inflation due to residual population structure ( $\lambda_{LDSC}$ ). The slope of this regression line provides an estimate of the SNP-heritability explained by the variants included in the LD score estimation (Bulik-Sullivan et al., 2015b).

In addition, the LDSC software package has the ability to calculate genetic correlations between pairs of traits (Equation 2.6; Bulik-Sullivan et al. (2015a)). The SNP-heritabilities for the respective traits are taken from the LD score regression method described above.





**Equation 2.6: Estimating Genetic Correlation using LDSC.**  $r_g$  = genetic correlation;  
 $\sigma_g^2$  = genetic covariance;  $h^2_i$  = SNP-heritability for trait  $i$  (Bulik-Sullivan et al., 2015a).

### **2.3.3 Meta-analysis of GWAS Results**

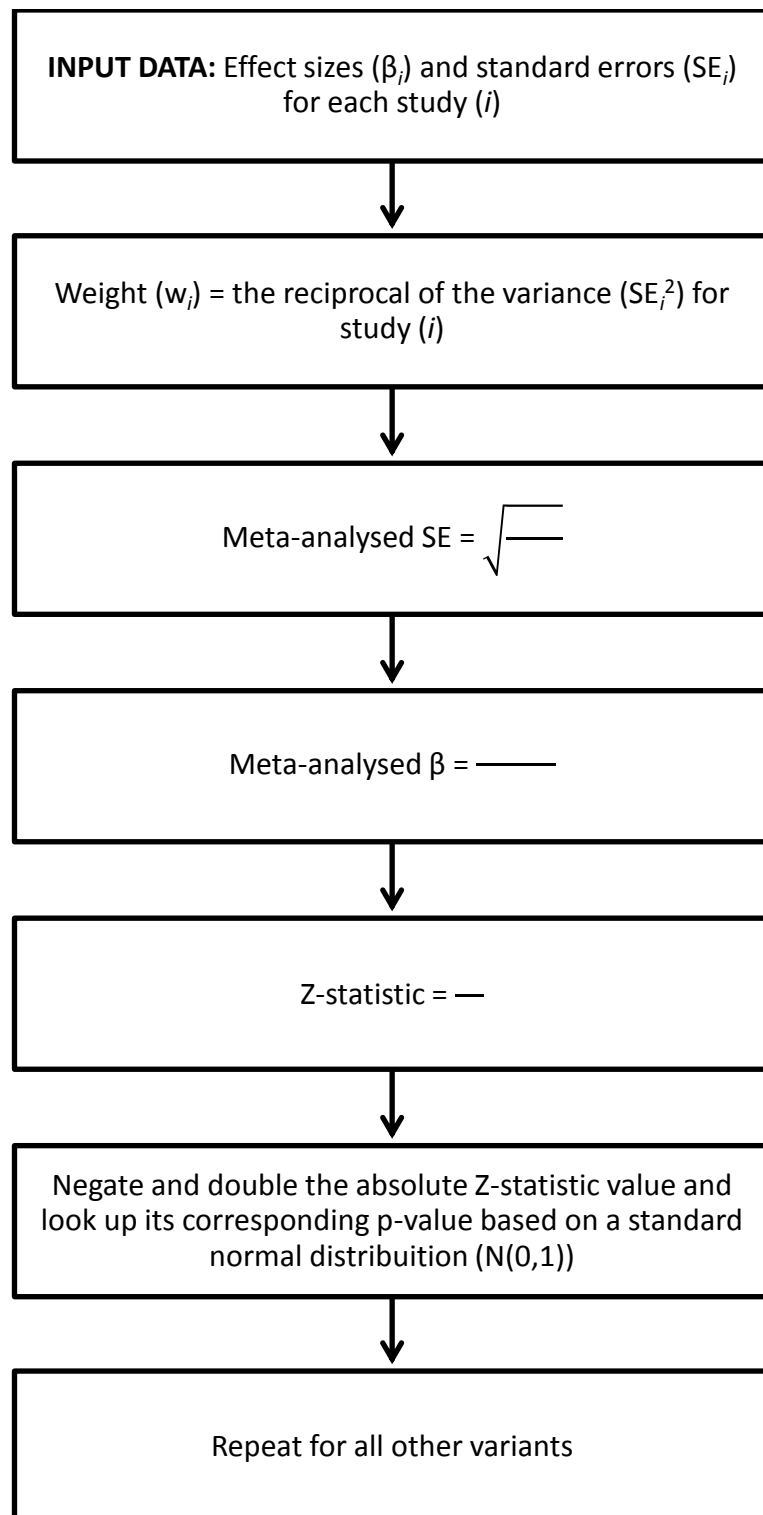
As individual investigations may lack the necessary statistical power to detect associations due to their sample size, meta-analyses have often been performed in order to combine the results of independent individual studies, therefore improving the ability to detect associations (Evangelou and Ioannidis, 2013). For the meta-analyses performed in the following chapters, the METAL software application was used (Willer, Li and Abecasis, 2010). METAL combines association test summary statistics using a fixed-effects, inverse variance weighted approach. This method weights effect sizes by their respective standard errors as outlined in Figure 2.1.

Figure 2.2 demonstrates a meta-analysis of a single variant across four simulated studies. Weights for each study were calculated according to the inverse variance weighted method as outlined in Figure 2.1, in order to then calculate the meta-analysed effect size, its respective standard error and p-value. Also included in Figure 2.2 is a Forest Plot that displays the effect sizes and their respective 95% confidence intervals for this specific variant in each of the studies included in the meta-analysis. Based on a nominal significance threshold of  $P = 0.05$ , the individual studies each appear to have non-significant results (described visually as the error bars crossing the vertical line of no effect), yet when meta-analysed according to

the inverse variance weighted approach, the effect of this variant becomes statistically non-zero (Effect (SE) = 0.115 (0.037),  $P = 1.89 \times 10^{-3}$ ).

This inverse variance weighted method assumes effect sizes and their respective standard errors to be consistent across studies. An alternative fixed-effects meta-analysis method, particularly useful if effect sizes were not reported using consistent units of measurement, weights studies according to their respective sample sizes (Willer et al., 2010).

**Figure 2.1: Flow diagram of METAL inverse variance weighted meta-analysis procedure.** Adapted from Willer et al. (2010).



**Figure 2.2: Meta-analysis for a single, simulated variant with Forest Plot.** SE = standard error. Meta-analysis to the values from the four contributing studies and the inverse variance weighted procedure outlined in F effect size; diamond denotes meta-analysed effect size; all error bars denote 95% confidence intervals.

Study	Effect	SE	Z Statistic	P-value	Weight
1	0.107	0.056	1.897	0.058	314.357
2	0.094	0.096	0.974	0.330	108.009
3	0.128	0.092	1.398	0.162	118.928
4	0.131	0.072	1.813	0.070	191.185
<b>Meta-Analysis</b>	<b>0.115</b>	<b>0.037</b>	<b>3.107</b>	<b>1.89 x 10<sup>-3</sup></b>	

-0.1    0

The meta-analyses performed here utilised a fixed effects model; however, an alternative random effects model can sometimes be more appropriate for the set of study data to be meta-analysed. Random effects models take into consideration the degree of heterogeneity (inconsistency) of effect sizes between studies which may influence results of meta-analyses (DerSimonian and Laird, 1986; Magi and Morris, 2010). This heterogeneity is considered in software applications such as GWAMA (Magi and Morris, 2010) by inclusion of the Cochran's Q statistic for each variant when calculating study weights, however, it should be noted that inclusion of this extra parameter does increase the variance of the estimated meta-analysed effect size (Magi and Morris, 2010), thus reducing the strength of the meta-analysed association signal. Therefore, use of a random effects model is appropriate in situations where there is known variability in the design of the contributing studies and the distribution of effect sizes is more important than identification of a single, precise effect size that can be applied to all studies (Munafo and Flint, 2004).

#### **2.3.4 Conditional Analysis**

To ascertain whether loci achieving genome-wide significant association were driven by a single causal variant or by multiple causal variants in the region, conditional analyses were performed on the single marker test summary statistics using GCTA-COJO (Yang et al., 2012). The variant demonstrating the strongest degree of association at a genome-wide significant locus was included as a covariate in the association test model and the association analysis repeated for all variants within  $\pm 1$  Mb of this marker. The association signals obtained in the conditional analysis would be greatly diminished compared to the original analysis in the event

of there being only a single causal locus. In the presence of multiple causal variants at a locus, variants tagging additional causal variants, independent of the conditioned variant would continue to demonstrate significant association in the GCTA-COJO analysis.

### ***2.3.5 Gene-based and Gene-set Analyses***

Many GWAS investigations have primarily concentrated on single marker tests. However, translating the association signals identified by a GWAS into potential causal mechanisms or biological functions is not straightforward. Combining test results across whole genes or collections of genes, aims to provide greater understanding of the functional consequences of genetic variants with respect to the trait of interest (Neale and Sham, 2004; Cantor, Lange and Sinsheimer, 2010).

Two main software applications available to perform gene-based association tests using association test summary statistics are VEGAS2 and MAGMA (de Leeuw et al., 2015; Mishra and Macgregor, 2015). For analyses undertaken here, both applications were employed.

For both software applications, genes were defined according to NCBI build 37 (hg19/GRCh37) coordinates, which is the corresponding build to which all variants were mapped to during their respective imputation phases. Furthermore, flanking regions of 50 kb (unless otherwise stated) were appended to the gene transcription start and stop sites. These flanking regions were included as variants in these regions can have an effect on gene regulation and expression, not only for the

nearest gene but other nearby genes too (Guo and Jamison, 2005; Schork et al., 2013; Brodie, Azaria and Ofran, 2016; Corradin et al., 2016). For both applications, LD patterns between variants included in the analysis for each gene were estimated using an ancestry matched reference panel (see below).

#### 2.3.5.1 VEGAS2

Developed by Liu et al. (2010b) and updated by Mishra and Macgregor (2015), VEGAS2 (VErsatile Gene-based Association Study 2) computes gene-based test statistics by initially converting single marker association test p-values into upper tail  $\chi^2$  test statistics with 1 degree of freedom before summing for all variants within each gene locus to give a single value. This gene-based  $\chi^2$  test statistic with  $n$  degrees of freedom, with  $n$  defined as the number of variants within the gene locus, is then examined under the null hypothesis of no association. However, variants within a gene locus are rarely in complete linkage equilibrium (i.e. independent of each other) and therefore, this correlation between variants also needs to be considered. These LD patterns ( $\Sigma$ ) are considered using an  $n \times n$  matrix of pairwise LD values which are estimated from an ancestry matched reference panel. In order to obtain p-values for the gene-based association tests, VEGAS2 performs a two-step process. Firstly, values following a multivariate normal distribution with a mean of zero and covariance matrix ( $\Sigma$ ) are simulated by the software. These simulated values are then compared against the summated gene-based  $\chi^2$  test statistic described above. Gene-based p-values are defined as the proportion of simulations where the simulated test statistic is greater than the original (observed) gene-based test statistic (Equation 2.7).

---

**Equation 2.7: VEGAS2 Gene-Based P-value Calculation.**

As the observed gene-based test statistic is based on the upper tail of the  $\chi^2$  distribution, few simulations would be expected to have greater  $\chi^2$  values in the event that the gene is associated with the trait of interest.

It is important to note that VEGAS2 uses simulations rather than permutations to determine the association signal. Before the development of VEGAS, gene-based association tests such as those performed by PLINK [command `--set-test`] were computationally intensive. In the method implemented by PLINK (Purcell et al., 2007), rather than performing simulations to generate alternative  $\chi^2$  test statistics to compare against, individual's phenotype data was permuted over 1000s of iterations to generate several single marker association test statistics which were subsequently used to generate a single gene-based test statistic. As with VEGAS, the proportion of permuted gene-based  $\chi^2$  test statistics greater than the observed  $\chi^2$  test statistic was defined as the gene-based p-value.

In order to maximise efficiency, the number of simulations performed varies depending on the p-value obtained after each round of simulations. Initially, 1000 simulations are performed before generating the gene-based p-value for a particular gene. If this p-value is less than 0.1, 10,000 simulations are performed before reviewing again. If this new gene-based p-value is less than 0.001, 1 million simulations are performed. In the event that no simulated test statistics are greater



than the observed test statistic after 1 million simulations, no further simulations are performed and the gene is assigned a reported p-value of  $P < 1 \times 10^{-6}$ . This is because the Bonferroni adjusted p-value threshold has already been crossed (0.05 / number of genes) and this is known to be an overly conservative threshold due to gene regions not being fully independent of each other since some variants may contribute to more than one gene (Liu et al., 2010b).

Definitions of gene loci and reference files for estimating LD patterns were built in to the VEGAS2 software application. Specifically, gene loci were defined according to a list of all RefSeq genes obtained from the UCSC table browser (Karolchik et al., 2004), and LD patterns were estimated by VEGAS2 using reference files composed of data for the 379 unrelated individuals of European ancestry from Phase 1, Version 3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2012; Mishra and Macgregor, 2015). The European ancestry dataset was used specifically as the cohorts included in these analyses were restricted to those of European ancestry only. The 1000 Genomes Project reference dataset is notably larger than the previously used HapMap2 reference dataset as utilised in the initial release of VEGAS (379 vs. 90 individuals), therefore providing greater accuracy when estimating LD patterns (Mishra and Macgregor, 2015).

#### 2.3.5.2 MAGMA

Unlike VEGAS2, gene definitions and reference files for estimating the LD structure of variants are not built in to the MAGMA (Multi-marker Analysis of GenoMic Annotation; de Leeuw et al. (2015)) software application, however these files are

available for download from the MAGMA website (URL: <https://ctg.cncr.nl/software/magma>). From here, gene definitions (to build 37 coordinates) originally obtained from the NCBI Entrez Gene database (Maglott et al., 2011) were downloaded, alongside reference files for European ancestry individuals from the 1000 Genomes Project.

An initial “Annotation” step must be run in MAGMA in order to assign variants to genes for gene-based analyses. It is at this stage that flanking regions can be appended to the gene transcription start and stop sites so that variants that may affect a particular gene’s regulatory processes are also included.

As GWAS summary statistics (variant ID labels and association test p-values) from single marker tests were used as input data in the absence of raw genotype data, as was the case in *Chapter 4*, the “snp-wise=mean” model was implemented by default. All variants assigned to a specific gene from the annotation step were included in the analysis of that gene with its observed gene-based test statistic determined by converting the respective variant p-values to  $-\log_{10}$  values (i.e. a variant with  $P = 1 \times 10^{-6}$  would be converted to a value of 6) before summing. The respective gene-based p-values are subsequently computed by MAGMA using these gene-based test statistics, the LD structure of the variants analysed and a sampling distribution appropriate for the test statistic as determined by MAGMA.

MAGMA also implements gene-set analysis which facilitates assessment of collections of genes that contribute to a common feature (e.g. a biological

pathway). For the gene-set analyses undertaken in the following chapters, “competitive” gene-set analysis was performed with Z-statistics from the MAGMA gene-based analysis used as input data. From the Molecular Signatures Database (MSigDB) (Subramanian et al., 2005), all gene-sets and their definitions (i.e. which genes contribute to each gene-set) were downloaded and used in these gene-set analyses.

In competitive gene-set analysis implemented by MAGMA, a linear regression is performed with genes coded as “1” if they are in the gene-set analysed or “0” if they are not, and the outcome taken as the gene-based Z-statistic. The mean difference in association between genes within the gene-set and those outside of it ( $\beta_s$ ) is tested against the null hypothesis of no difference in association ( $H_0$ : mean  $\beta_s = 0$ ) (Equation 2.8; de Leeuw et al. (2015)).

**Equation 2.8: Computing test statistics for competitive gene-set analyses in MAGMA.**  $Z$  = gene-set test statistic;  $\beta_0$  = intercept of the linear regression model;  $S$  = number of genes in the gene-set;  $\beta_s$  = difference in association between genes within the gene-set and those outside it;  $\beta_c$  = matrix of covariates (e.g. no. of variants within a gene);  $\epsilon$  = residuals with correlations aligned with the gene-gene correlations to account for LD between genes (de Leeuw et al., 2015).

### **2.3.6 SNP-Heritability**

SNP-heritability estimates were calculated using Genome-wide Complex Trait Analysis (GCTA) (Yang et al., 2011a). In the most commonly-employed analysis approach, GCTA is used to estimate the proportion of trait variance due to additive effects of commonly occurring SNPs ( $h^2_{\text{SNP}}$ ) (GCTA-GREML). This is done by

considering pairwise relatedness between individuals (in the form of a GRM) as a random effect in a mixed linear model fitted using restricted maximum likelihood (REML) (Yang et al., 2011a). The model fitted is equivalent to the mixed linear model described in Equation 2.4 and allows for the inclusion of covariates; however the variance of the trait due to genetic effects ( $\sigma_g^2$ ) from this model is of interest. Here, the trait variance ( $V_p$ ) is defined as the sum of the variance due to genetic variants included in estimation of the GRM whilst considering the variance-covariance structure of the genetic variants between individuals ( $A_g\sigma_g^2$ ) and the residual variance which is attributable to non-genetic effects ( $I\sigma_\epsilon^2$ ) (Equation 2.9; Yang et al. (2011a)). If multiple GRMs are fitted as in the case of a joint analysis,  $A_g\sigma_g^2$  becomes  $\sum A_i\sigma_i^2$ , whereby  $i$  denotes the individual variance components with a single GRM is constructed for each ( $A_i$ ).

**Equation 2.9: Estimating trait variation using GCTA.**  $V_p$  = trait variance;  $A_g$  = genetic relationship matrix (GRM);  $\sigma_g^2$  = variance due to variants contributing to the GRM;  $I$  = identity matrix of residual effects (diagonal elements = 1, off diagonal elements = 0);  $\sigma_\epsilon^2$  = variance due to non-genetic effects (Yang et al., 2011a).

In order to construct the GRM required to estimate variance due to additive genetic effects, a relatedness coefficient between each pair of individuals was calculated using Equation 2.10, which provides a weighted count of how many variant genotypes are common to both individuals. A relatedness coefficient of less than 0.025, which corresponds to the level of genetic similarity expected for second or third cousins, is generally used to define “unrelated” individuals (Yang et al., 2010).

---

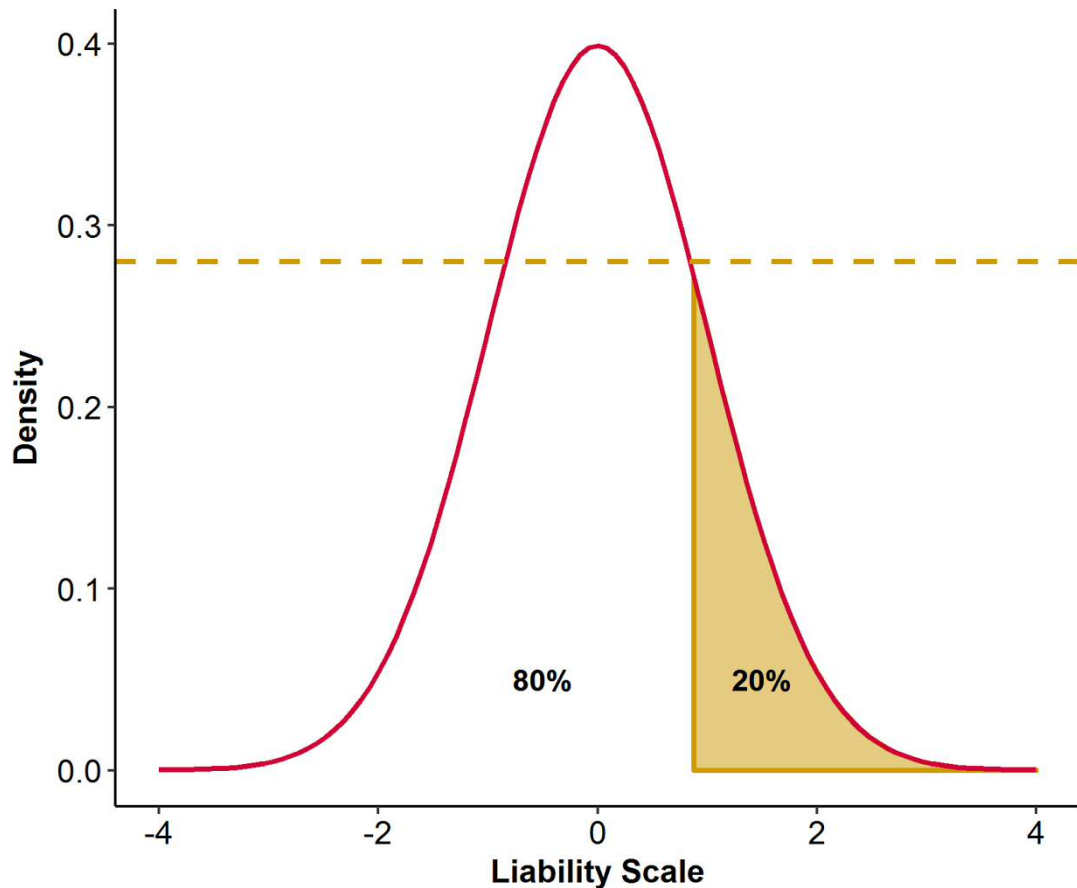
**Equation 2.10: Determining genetic relatedness between two individuals,  $j$  and  $k$ .**  $A$  = genetic relationship coefficient,  $i$  = variant number (1, 2, 3... N),  $p_i$  = frequency of the reference allele of variant  $i$ ,  $x_{ij}$  genotype of individual  $j$  at variant  $i$ ,  $x_{ik}$  genotype of individual  $k$  at variant  $i$  (Yang et al., 2011a).

To determine whether SNP-heritability estimates are greater than zero, since heritability is restricted to a continuous scale of values between 0 and 1; p-values are determined by halving the p-value obtained using the LRT (likelihood ratio test) statistic based on a one-tailed  $\chi^2$  test with 1 degree of freedom. This LRT statistic is computed by GCTA during analysis however it can be closely approximated by:  $(\text{estimated } h^2 / SE)^2$  (Yang et al., 2011a).

#### *2.3.6.1 Observed to Liability Scale Conversion*

As heritability is traditionally estimated on the 'observed scale' (see below for definition), conversion to the 'liability scale' was required for the heritability estimates of the different dichotomous traits to be compared either against each other or across different case threshold definitions (Lee et al., 2011). The liability scale is a continuous scale used for discrete variables, for example cases and controls for a particular phenotype. It is assumed that the liability is normally distributed across the study sample, with the area under the normal distribution curve representing the sample (Figure 2.3).

**Figure 2.3: The normal probability density function for transformation from the observed scale to the liability scale.** If cases account for 20% of the individuals sampled (shaded area), the height of the function at the threshold value,  $t$  (edge of shaded area) is 0.28 ( $z$ , denoted by dotted line). Adapted from Lee et al. (2011).



The threshold value ( $t$ ) on the liability scale is set according to the prevalence of the phenotype, i.e. the number of cases relative to the sample size. For example, if a particular phenotype is presumed to have a prevalence of 20%, the threshold on the liability scale is set such that the 20% of the area under the curve is set to the right of the threshold value with the remaining 80% of the area under the curve to the left of the threshold value (Figure 2.3). This transformation allows a dichotomous trait to be accurately expressed on a continuous scale, which is the same type of scale used when estimating heritability. As a result, ‘heritability on the

liability scale' is independent of the prevalence of the investigated phenotype (Falconer and Mackay, 1996). It has been reported by Lee et al. (2011) that without such a conversion, heritability estimates for dichotomous traits are prone to a downward bias due to uncorrected measurement errors. Transformation to the liability scale is performed automatically by GCTA for dichotomous traits (based on Equation 2.11; (Lee et al., 2011)).

---

**Equation 2.11: Transformation from heritability on the observed scale ( ) to heritability on the liability scale ( ) with consideration of ascertainment bias.**  $K$  = population prevalence,  $P$  = sample prevalence,  $z$  = height of the standard normal probability density function at threshold  $t$  (Lee et al., 2011).

This transformation also takes into consideration ascertainment bias whereby there is a difference between the population and sample prevalence rates.

#### *2.3.6.2 Consideration of Uncorrected Population Effects*

Effects such as population stratification and cryptic relatedness within a sample may affect the results and interpretations of genetic studies if not accounted for. The major principal components have commonly been included as covariates when performing such studies to adjust for population stratification (Price et al., 2006; Price et al., 2010). However, mixed linear models like those utilised by GCTA, can also be used to account for uncorrected population stratification and cryptic relatedness (Yu et al., 2006; Kang et al., 2010; Zhang et al., 2010).

For SNP-heritability estimation using GCTA or equivalent methods, there is no general consensus regarding the inclusion or exclusions of principal components to account for population stratification within the examined sample. However, some authors have suggested that the inclusion of principal components may not fully account for population stratification, and may in fact make SNP-heritability estimates less accurate due to over-fitting (Browning and Browning, 2011; Goddard et al., 2011; Dandine-Roulland et al., 2016; Krishna Kumar et al., 2016).

Strategies have been proposed to estimate the degree of inflation of SNP-heritability estimates due to uncorrected population stratification and cryptic relatedness, based on the principle that genetic regions (e.g. chromosomes) are independent of each other if there are no related individuals present and the sample is homogenous (Yang et al., 2011b). If related individuals are present and/or a non-homogenous sample is analysed, these regions will no longer be independent since there will be some correlation between them through ancestry informative markers (AIMs). AIMs are variants whose allele frequencies are highly correlated within population subgroups. For example, if there are loci where particular alleles are found at high frequency in individuals of European ancestry, yet these same alleles are much rarer in all other population subgroups, these alleles would be deemed to be specific to individuals of European ancestry. It is expected that these AIMs are randomly distributed throughout the genome; hence, regions will be correlated with each other due to these population effects if left unaccounted for (Yang et al., 2011b). Thus, SNP-heritability estimates from an individual region ( $h^2_{sep}$ ) would be inflated due to correlation from variants outside this region in



individual analyses; yet, when all regions are combined in a joint analysis, these correlated effects are taken into consideration as estimates for each region ( $h^2_{\text{joint}}$ ) are obtained by conditioning on all other regions. As a result, estimates for each region would be independent of all other regions included in the model and inflation from these population effects would be eliminated (Yang et al., 2011b).

To ascertain the extent to which SNP-heritability estimates obtained may be biased (inflated) due to uncorrected population stratification or residual cryptic relatedness, the method proposed by Yang et al. (2011b) was utilised. First, the genome was split into individual chromosomes, with SNP-heritability estimated for each chromosome separately and then jointly for all chromosomes. The difference between the heritability estimates for the individual and joint analyses for each chromosome was regressed against chromosome length (obtained from UCSC Genome Browser (Kent et al., 2002) to NCBI human genome build 37 (hg19 / GRCh37) coordinates) (Figure 5.1). Finally, an estimate for the proportion of variance attributable to population structure across the whole genome was obtained by applying Equation 2.12 (Yang et al., 2011b):

**Equation 2.12: Estimating the proportion of variance attributable to population structure across the whole genome.**  $b_0$  = intercept;  $b_1$  = gradient;  $C$  = chromosome;  $L_C$  = chromosome length (Yang et al., 2011b).

### 2.3.6.3 Quantification of Dominance Effects

In addition, GCTA can be used to estimate the proportion of trait variance due to partitioned additive and dominance effect components (GCTA-GREMLd) (Zhu et al., 2015). Here, separate GRMs are constructed by GCTA prior to SNP-heritability estimation; one based on additive effects and another based on dominance effects (Equations 2.10 and 2.13). As with the aforementioned joint analysis for estimating uncorrected population effects, a joint analysis is performed in order to partition the overall genetic component of trait variance into these two constituent parts.

In order to compute the GRM based on dominance effects, Equation 2.10 is adjusted to account for the recoded genotypic values (for the genotypes AA, AB and BB respectively: additive effects ( $x$ ) = 0, 1, 2; dominance effects ( $x'_D$ ) = 0,  $2p$ , or  $(4p - 2)$  whereby  $p$  is the frequency of allele B) thus becoming Equation 2.13:

— —————

**Equation 2.13: Determining genetic relatedness between two individuals,  $j$  and  $k$ , based on dominance effects.**  $D$  = genetic relationship coefficient,  $i$  = variant number (1, 2, 3... N),  $p_i$  = frequency of the reference allele of variant  $i$ ,  $x'_{D(ij)}$  genotype of individual  $j$  at variant  $i$ ,  $x'_{D(ik)}$  genotype of individual  $k$  at variant  $i$  (Zhu et al., 2015).

As a result, Equation 2.9 becomes:

**Equation 2.14: Estimating trait variation due to additive and dominance effects using GCTA-GREMLd.**  $V_p$  = trait variance;  $A$  = additive effects GRM;  $\sigma_A^2$  = variance due to additive effects of variants contributing to the GRM;  $D$  = dominance effects GRM;  $\sigma_D^2$  = variance due to dominance effects of variants contributing to the GRM;  $I$  = identity matrix of residual effects (diagonal elements = 1, off diagonal elements = 0);  $\sigma_\epsilon^2$  = variance due to non-genetic effects (Zhu et al., 2015).

### **2.3.7 Multiple Testing Correction**

When performing a large number of independent statistical tests, many tests would achieve p-values lower than the nominal significance threshold of  $P = 0.05$  by chance alone. As a result of this multiple testing, significance thresholds are typically adjusted (or alternative methods of defining noteworthy results are utilised) in order to account for this. The main methods of multiple testing correction are outlined below.

#### *2.3.7.1 Bonferroni Adjustment*

The Bonferroni adjustment is the simplest method of multiple testing correction. This method takes the nominal significance threshold and divides it by the number of tests performed in that analysis. For example, if 100 genetic variants were tested for association with a given trait and the nominal significance threshold was defined as 0.05, the adjusted significance threshold should be  $0.05/100$ , which is 0.0005. Therefore, if a particular variant achieved an association test p-value of 0.0034 (which would be deemed a “significant” result if using the nominal threshold of 0.05 without considering the other 99 tests performed); this same variant would be deemed as a non-significant association after Bonferroni adjustment, as its p-value is above the adjusted threshold of 0.0005.

#### *2.3.7.2 Genome-wide Significance Threshold*

The genome-wide significance threshold of  $P = 5 \times 10^{-8}$  is commonly applied for genome-wide association studies (Sham and Purcell, 2014). This choice of significance threshold is based on the expected number of independent variants in

the genome, the number of variants tested for association, and the linkage disequilibrium (LD) structure of human populations (Risch and Merikangas, 1996; Dudbridge and Gusnanto, 2008; Hoggart et al., 2008). Generally, p-values less than  $1 \times 10^{-5}$  have been proposed to demonstrate suggestive association (Welter et al., 2014). Such variants may not attain genome-wide significance due to an insufficient sample size, yet they may be candidates worthy of follow-up analyses in an independent cohort.

### *2.3.7.3 False Discovery Rates*

False discovery rates (FDRs), as outlined by Benjamini and Hochberg (1995), are an alternative method to account for multiple tests performed. Rather than determining the likelihood of the null hypothesis of no association being accepted (i.e. a true positive result) as is the case in the previous methods, use of a false discovery rate controls for the proportion of occasions where there is an incorrect rejection of the null hypothesis. In other words, use of false discovery rates allows for control over the proportion of false positive results. This procedure determines likely associations by initially sorting the results for all association tests (e.g. for all genes) in ascending order of association test p-value ( $P$ ). For all genes, the ranking (position) of the gene in this sorted list ( $i$ ) is noted and multiplied by the pre-defined false discovery rate ( $q$ ), generally 0.05 (5%), before dividing by the total number of genes in the list ( $m$ ). This value ( $iq/m$ ) is defined as the critical value for that gene. The lowest gene in the list whereby  $P \leq$  critical value and all others above it are defined as demonstrating likely association with the trait of interest. Of the genes

demonstrating this level of association, no more than  $q$  (i.e. 5%) will be false positive results.

Table 2.3 demonstrates a simulated scenario where false discovery rates have been utilised. In this example, 10 genes were tested for association with a particular trait and their gene-based association test p-values were sorted in ascending order of association test p-value. The column labelled “*Rank*” denotes the position of the genes in this sorted list ( $i$ ). If the pre-defined false discovery rate ( $q$ ) is 0.05, the lowest ranked gene with an association test p-value less than its respective row’s critical value of  $0.005i$  is gene H ( $0.019 < (4 \times 0.05) / 10$ ). Thus, genes J, F, C and H would be deemed to be noteworthy associations.

**Table 2.3: Example of false discovery rates in action.** Number of genes in the list ( $m$ ) = 10; False discovery rate ( $q$ ) = 0.05;  $iq/m$  = critical value;  $FDR = Pm/i$ . Genes in bold = significant associations based on FDR.

Gene	P-value	Rank	$iq/m$	FDR
<b>J</b>	<b>0.002</b>	<b>1</b>	<b>0.005</b>	<b>0.015</b>
<b>F</b>	<b>0.002</b>	<b>2</b>	<b>0.010</b>	<b>0.008</b>
<b>C</b>	<b>0.017</b>	<b>3</b>	<b>0.015</b>	<b>0.024</b>
<b>H</b>	<b>0.019</b>	<b>4</b>	<b>0.020</b>	<b>0.048</b>
E	0.116	5	0.025	0.231
B	0.129	6	0.030	0.215
A	0.500	7	0.035	0.715
I	0.518	8	0.040	0.648
G	0.780	9	0.045	0.867
D	0.954	10	0.050	0.954

For analyses where false discovery rates have been utilised, results have displayed the calculated false discovery rate (e.g. Table 2.3, column “*FDR*”) for each test (e.g. each gene). These were obtained by rearranging the equation such that the

unknown variable is the false discovery rate ( $q$ ). Genes with calculated values of  $q$  that were less than the pre-defined false discovery rate threshold of 0.05 (unless otherwise stated) were deemed as significant associations.

## **Chapter 3 X-Chromosome Wide Association Study for Refractive Error**

## Chapter 3 X-Chromosome Wide Association Study for Refractive

### Error

#### 3.1 Introduction

With the exception of the study by Pickrell et al. (2016), which investigated variants associated with self-reported myopia, GWAS for refractive error have solely concentrated on autosomal variants with no consideration of variants on chromosome X, despite knowledge that three loci associated with non-syndromic myopia have been identified on this chromosome (MYP1, MYP13 and MYP26) (Zhang et al., 2006; Zhang et al., 2007; Guo et al., 2010; Ratnamala et al., 2011; Xiao et al., 2016). As outlined below, this is mostly due to the difficulties of running an association study for chromosome X (Wise, Gyi and Manolio, 2013). Firstly, there are two copies of each autosomal chromosome in an individual; however, males only have one copy of the X-chromosome, and therefore the standard GWAS analysis method needs to be modified in order to account for this. In addition, whilst females have two copies of the X-chromosome, it is often the case that one copy is inactivated, yet this inactivation is variable and may not be complete (Carrel and Willard, 2005; Konig et al., 2014).

The aim of this investigation is to perform a chromosome X-wide association study, using a population-based sample of unrelated individuals, to identify potential susceptibility loci for mean spherical equivalent (MSE) at age 15 years.



## 3.2 Methods

Participants were from the Avon Longitudinal Study of Parents and Children (ALSPAC), a population-based birth cohort study (Boyd et al., 2013). After exclusion of participants who had withdrawn consent, the sample for this current investigation consisted of 8,935 individuals with genotype data, of whom 4,577 were males and 4,358 females. Of these individuals, 4,070 had measures of MSE at age 15 years, obtained through non-cycloplegic autorefraction (Table 3.1).

**Table 3.1: Summary of refractive error in the XWAS analysis sample.**

	<b>Males</b>	<b>Females</b>	<b>Total</b>
<b>N total (%)</b>	4,577 (51.2%)	4,358 (48.8%)	8,935
<b>N with MSE at age 15 years (%)</b>	1,924 (47.3%)	2,146 (52.7%)	4,070
<b>Mean MSE (SD) (D)</b>	-0.370 (1.30)	-0.392 (1.26)	-0.382 (1.28)

Phenotyping, genotyping and imputation were carried out by members of the ALSPAC team and as outlined in *Section 2.1*.

After imputation, genotypes were available for 1,250,218 variants across chromosome X. All variants were mapped to NCBI build 37 genome coordinates.

### 3.2.1 Quality Control

To maintain a high level of data quality, quality control was performed on all variants for all genotyped individuals. Individuals missing information for 10% or more of the variants were excluded from the analysis. Variants were excluded from the analysis if they were not genotyped successfully in at least 95% of individuals, their imputation quality (IMPUTE2 INFO) metric was < 0.5, their MAF was < 0.01 or

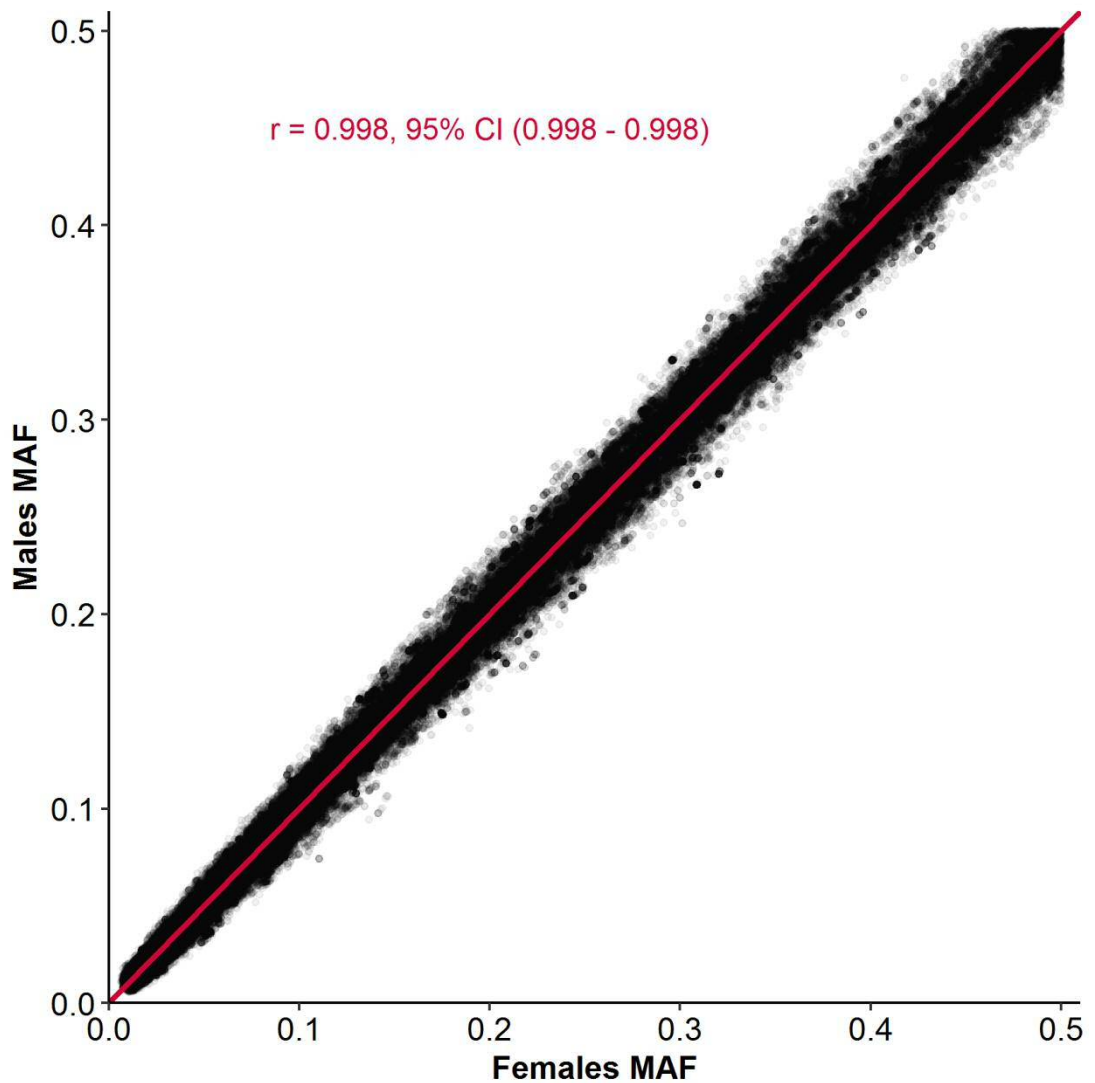
their HWE p-value in females was  $< 1 \times 10^{-6}$ . As regards HWE, it should be noted that the comparatively lenient threshold applied was chosen since it takes into consideration the very high number of variants tested; an analogous value has been utilised in other investigations, e.g. Davidson et al. (2014). Furthermore, if variants are in HWE in females and allele frequencies are equal between males and females then allele frequencies should remain constant through future generations (Zheng et al., 2007). In order to verify this assumption, the correlation in MAFs between males and females was confirmed to be high, as shown in Figure 3.1 ( $r = 0.998$ ; 95% CI: 0.998-0.998).

### **3.2.2 Single Marker Tests**

For all variants passing quality control, tests for association were performed separately for males and females using a frequentist linear regression model and expected counts used in the presence of missing genotypes, implemented in SNPTEST v2.5. Genotypes for males were coded as 0, 2 (AA, BB); whereas females were coded as 0, 1, 2 (AA, AB, BB). As the ALSPAC cohort of children was age matched by design, and non-ancestry matched individuals were excluded by ALSPAC researchers, additional covariates were not included in analyses.

The separate association test results for males and females were then combined in a fixed effects, standard error-weighted meta-analysis using METAL (Willer et al., 2010). Manhattan and quantile-quantile (QQ) plots were generated using the summary statistics generated from each stage of analysis.

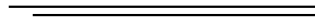
**Figure 3.1: Plot of minor allele frequencies (MAFs) between males and females.** Red line = line of unity (Females MAF = Males MAF). N = 4,070 across 267,185 variants.



### 3.2.3 Sex-Specific Effects

In order to evaluate sex-specific effects, the separate male and female association test results were further analysed using the R package EasyStrata (Winkler et al., 2015). This evaluation took the beta estimates and their standard errors from each SNPTTEST analysis to generate a Z-statistic value (Equation 3.1). From this test statistic, a p-value was obtained for each locus relating to the difference in beta estimates between the two sexes (i.e. the likelihood of beta estimates being

identical between males and females), based on a standard normal distribution. Lower p-values signified loci whose beta estimates were more likely to be different between males and females. A false discovery rate of 5% was applied in order to define whether a sex based effect was present after accounting for multiple testing.



**Equation 3.1:** Calculating the test statistic determining differences in beta estimates between males and females ( $Z$ ).  $\beta$  = beta (effect) estimate; SE = standard error from association test summary statistics.

### **3.2.4 Gene-based and Gene-set Analyses**

Gene-based tests were subsequently performed for this data set using VEGAS2 and MAGMA (outlined in *Section 2.3.5*). In both instances, LD patterns were estimated using European ancestry reference panels, specifically reference files composed of data for the 379 unrelated individuals of European ancestry from Phase 1, Version 3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2012).

Potential functional properties of X-chromosome genes associated with myopia in the MAGMA analysis were further investigated using competitive gene-set analysis in MAGMA (as outlined in *Section 2.3.5.2*). Adjustment for multiple testing was applied using a false discovery rate of 5% for these gene-based and gene-set test results.

### **3.2.5 Power Calculation**

In order to determine how much statistical power this current investigation had, a power calculation was performed using QUANTO (Gauderman, 2002). The following assumptions were applied when performing the calculation: a type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ , an effect size of 0.10 D, a MAF range between 0.01 and 0.50, a sample size of 4,070 individuals, and that MSE had a normal distribution with a mean (SD) of -0.382 (1.28) D. The Type I error rate was selected to match the p-value required to declare significance for the association test performed after adjustment for multiple testing and accounting for LD. The effect size estimate was based on the average effect size from a meta-analysis for refractive error undertaken by the CREAM consortium (Verhoeven et al., 2013). Sample size and MSE estimates were obtained from the ALSPAC sample used for this investigation.

An additional power calculation was performed in order to determine the minimum sample size required in order to have 60-80% power to detect an effect size of 0.10 D. As above, the population MSE estimate was based on the ALSPAC sample mean (SD) of -0.382 (1.28) D.

### **3.3 Results**

After performing quality control on the study data, no individuals were excluded based on the filter of having at least 90% of variants genotyped successfully. At the marker level, 839,176 variants had poor imputation quality (IMPUTE2 INFO < 0.5), 957,399 variants had minor allele frequencies (MAFs) below 0.01, and three variants had HWE p-values below  $1 \times 10^{-6}$  in females. As a result, 983,033 variants were excluded from analysis due to being outside these quality control thresholds, leaving 267,185 variants passing quality control.

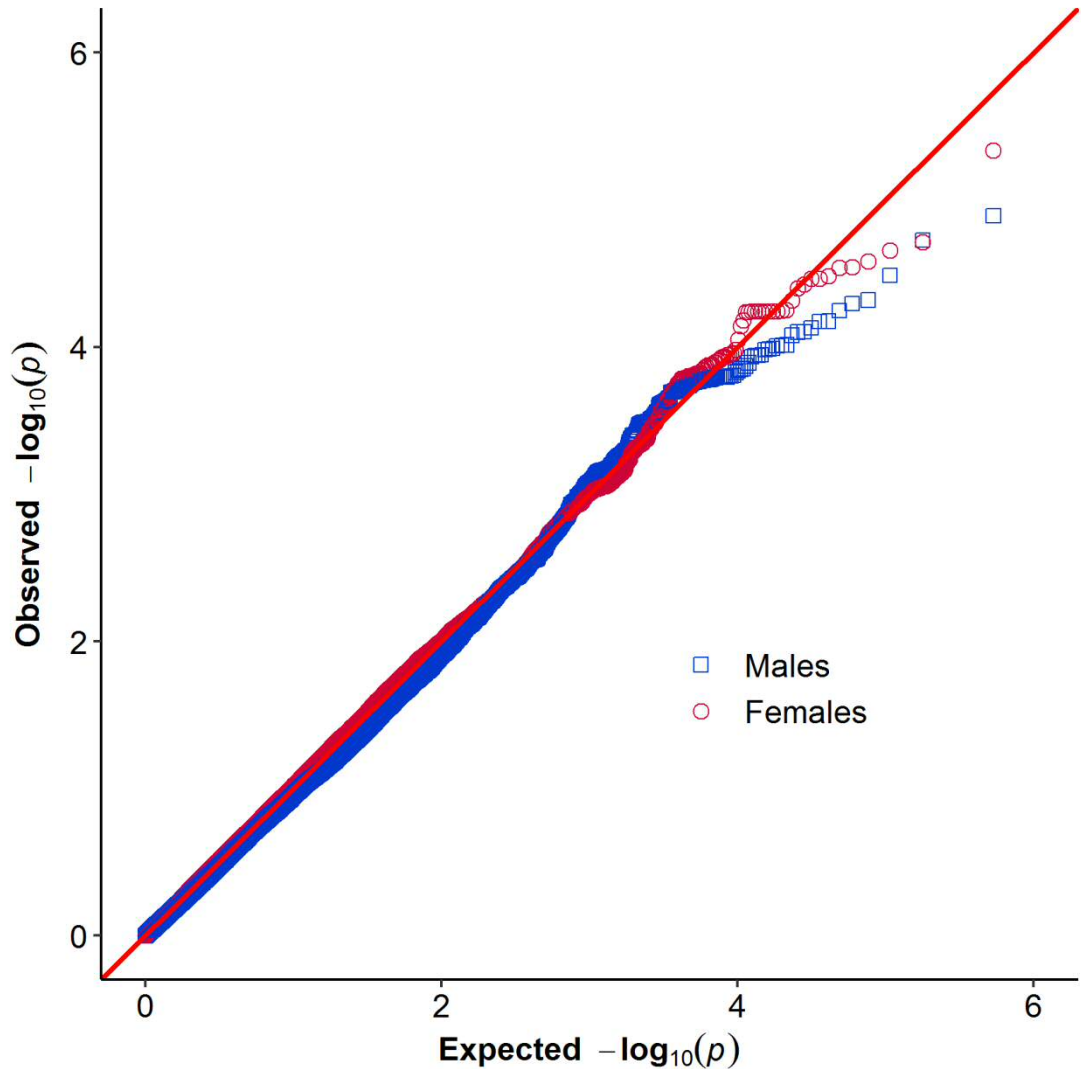
#### **3.3.1 Single Marker Tests**

Linear regression was performed separately for males and females for each of the 267,185 chromosome X variants that passed quality control. In these individual analyses, no variants achieved genome-wide significant association ( $P < 5 \times 10^{-8}$ ). A single variant (rs188930264) demonstrated a p-value of less than  $1 \times 10^{-5}$  for association with MSE at age 15 years for females (Table 3.2). No variants demonstrated  $P < 1 \times 10^{-5}$  in the male-only analysis.

**Table 3.2: Summary of results for variants on chromosome X achieving association test P-values less than allele frequency; Effect = effect size per copy of *Effect Allele*; SE: standard error of beta estimate; N = 2,146.**

<b>Variant</b>	<b>Position</b>	<b>Effect Allele</b>	<b>Other Allele</b>	<b>EAF</b>	<b>Effect</b>	<b>SE</b>	<b>P-value</b>	<b>Nearest Gene</b>
rs188930264	11510031	T	A	0.015	0.921	0.201	$4.63 \times 10^{-6}$	<i>ARHGAP6</i>

**Figure 3.2: Chromosome X quantile-quantile plot from sex-specific linear regression association tests with MSE at age 15 years.** Blue squares = males; red circles = females. Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).



Genomic inflation factors ( $\lambda_{GC}$ ) for males and females were 0.912 and 1.092 respectively. This can be partially observed by the slight deviation of observed p-values from the red line (null hypothesis) in the quantile-quantile plot, which shows a graphical summary of the sex-specific association test results for all variants examined (Figure 3.2). Due to this relatively large discrepancy in  $\lambda_{GC}$  values between the sexes, further investigations were conducted to identify a possible cause. Firstly,



as MSE was originally analysed as a quantitative trait, the analysis was repeated with MSE coded as a binary trait, with cases defined as individuals with MSE of  $\leq -1.00$  D as used in previous investigations of children from this cohort (Guggenheim et al., 2012; Guggenheim et al., 2013b; Guggenheim et al., 2014; Shah et al., 2017). Tests for association were repeated for this binary trait using logistic regression, with  $\lambda_{GC}$  values now 1.044 and 1.003 for males and females, respectively. In addition, an alternative, hypothetical, quantitative trait was simulated for the 4070 individuals included in the original XWAS. This new trait was designed to follow a standard normal distribution, centred on a mean of zero and standard deviation of one. This simulated trait breaks the relationship between genotype and phenotype; thus, the resulting association test p-values would be expected to follow the null hypothesis of being uniformly distributed ( $\lambda_{GC} = 1$ ). The  $\lambda_{GC}$  values for the simulated quantitative trait were 1.016 and 1.021 for males and females respectively. Together, the above results for the binary trait and the simulated quantitative trait support the validity of the methods used to calculate  $\lambda_{GC}$  values for this XWAS investigation. Thus, the observed discrepancy in  $\lambda_{GC}$  between the sexes (0.912 vs. 1.092) is likely to be a chance finding.

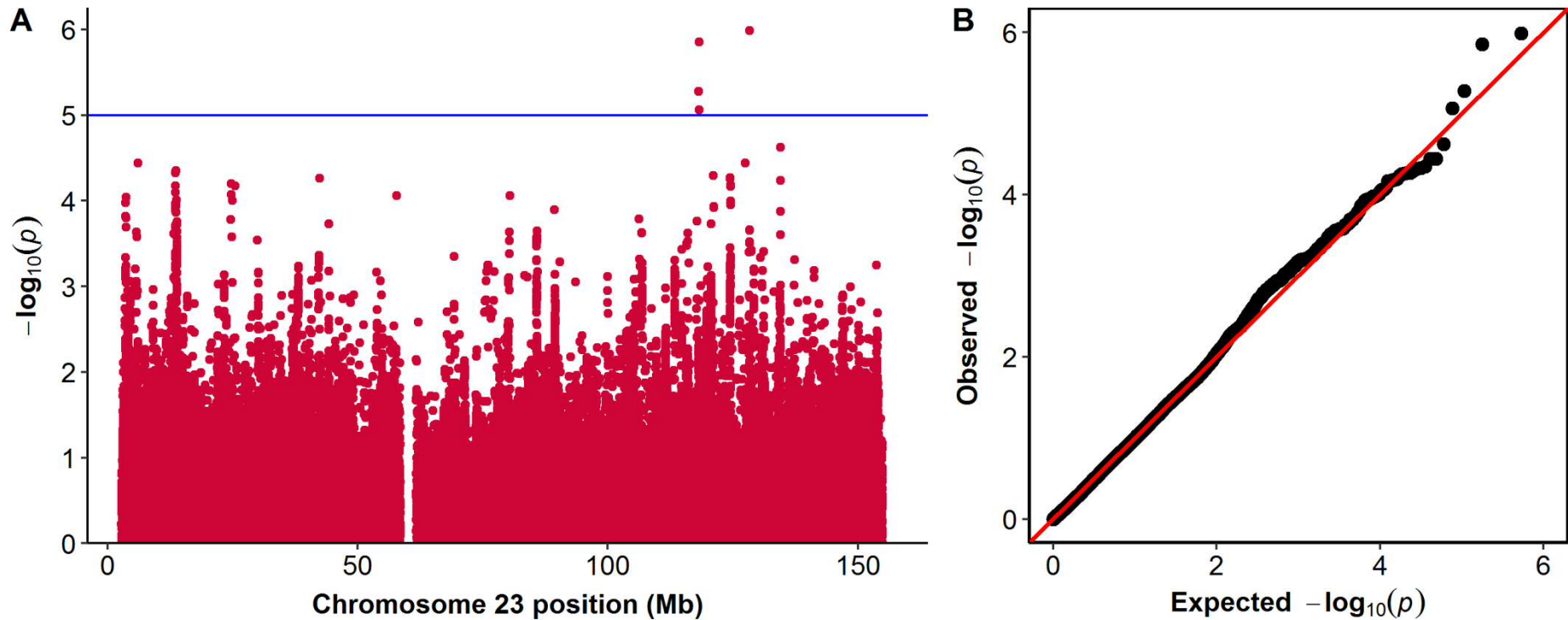
In the meta-analysis of the XWAS summary statistics for males and females, no marker achieved  $P < 5 \times 10^{-8}$  (Table 3.3, Figure 3.3).

**Table 3.3: Summary of results for variants on chromosome X from meta-analysis of association tests for MSE at age 15 years ( $P < 1 \times 10^{-5}$ ).**

EAF = effect allele frequency; Effect = effect size per copy of allele B; SE = standard error of beta estimate; N = 4,070 for all variants.

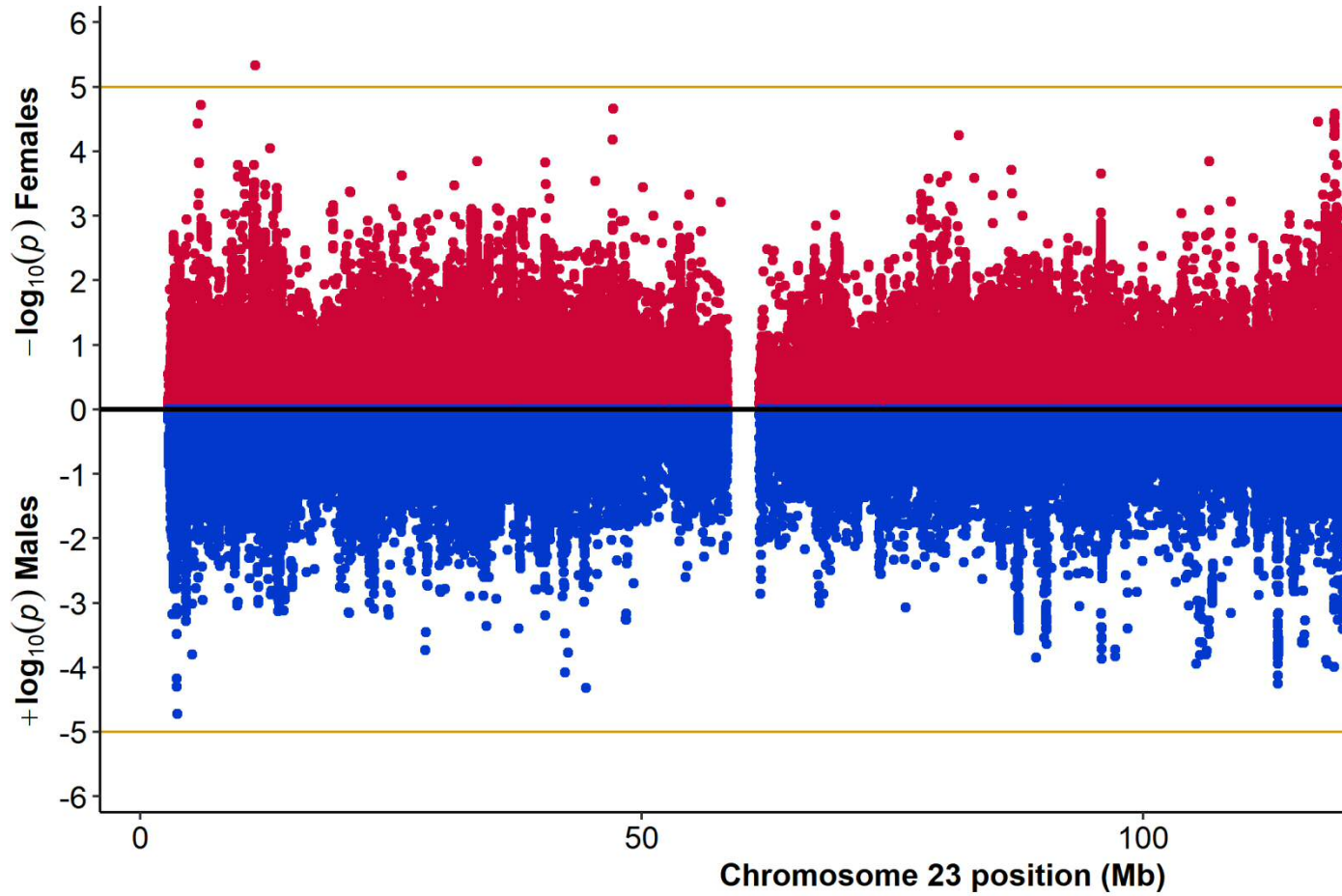
Variant	Position	Effect Allele	Other Allele	EAF	Effect	SE	P-value	Nearest Gene
rs145471572	128366278	G	C	0.012	-0.654	0.134	$1.04 \times 10^{-6}$	<i>RPS26P56</i>
rs189623102	118251135	C	T	0.010	-0.579	0.120	$1.41 \times 10^{-6}$	<i>KIAA1210</i>
rs58142779	118182646	A	C	0.012	-0.541	0.119	$5.27 \times 10^{-6}$	<i>LOC727838</i>
rs147218029	118330315	A	G	0.010	-0.671	0.151	$8.67 \times 10^{-6}$	<i>ARL5AP1</i>

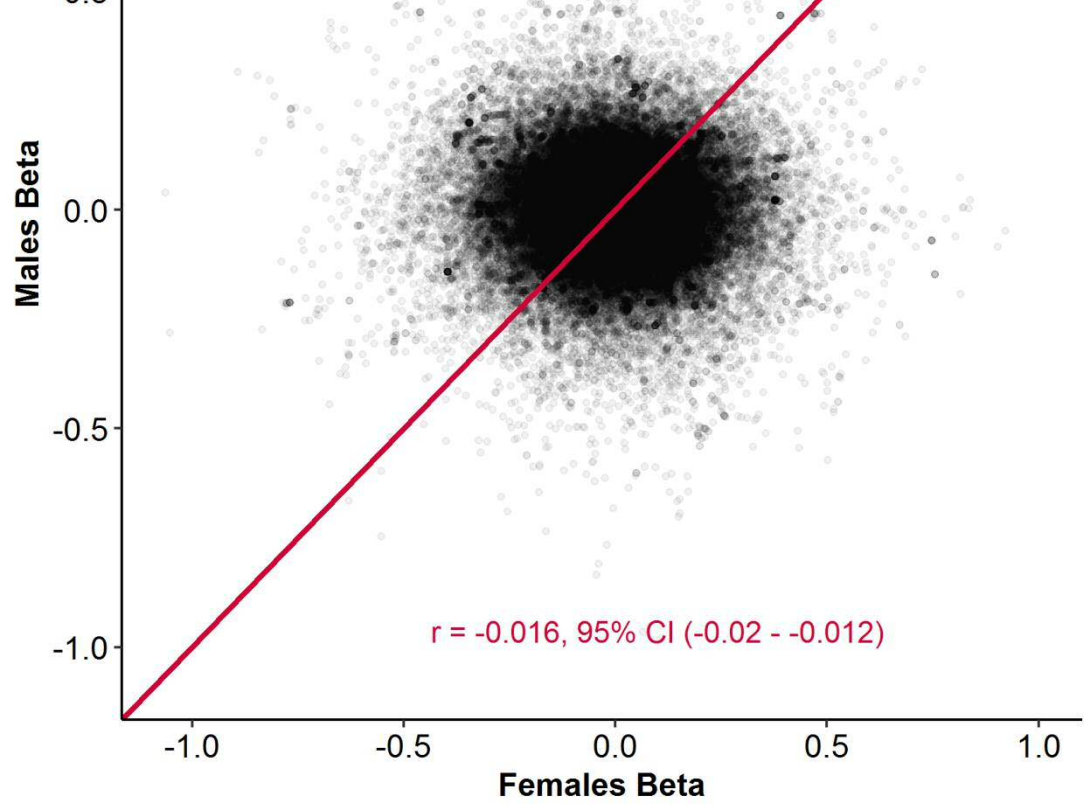
**Figure 3.3: Chromosome X Manhattan and Quantile-Quantile plots from meta-analysed results.** Panel A: Horizontal blue line denotes an arbitrary threshold for declaring “suggestive” evidence of association ( $P = 1 \times 10^{-5}$ ). Panel B: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected).



of effect sizes (beta) estimated from the individual association tests showed virtually no correlation between males and females ( $r = -0.016$ ; Figure 3.5). Moreover, Welch's unpaired samples  $t$ -test identified no significant difference in mean effect size between males and females ( $t = -0.769$ ,  $df = 498,640$ ,  $P = 0.442$ ). Further evaluation of these differences in effect size between the sexes showed that, for a false discovery rate of 5%, there were no significant differences. The ten variants demonstrating the greatest difference in effect size between the sexes are shown in Table 3.4.

**Figure 3.4: Chromosome X Miami plot from sex-specific linear regression association tests for MSE at a** females; lower portion (blue) = males. Horizontal yellow lines denote suggestive association threshold of  $P =$





**Table 3.4: Summary of the 10 variants demonstrating the greatest difference in effect size (beta) between frequency; SE = standard error of beta estimate; P SexDiff = p-value from the test for male and female beta**

Variant	Position	Effect Allele	Other Allele	Males				N	P
				N	EAF	Beta	SE		
rs194287	119052285	T	C	1924	0.145	0.077	0.042	2146	0
rs194289	119053638	A	G	1924	0.145	0.077	0.042	2146	0
rs194284	119050148	C	T	1924	0.181	0.087	0.040	2146	0
rs5910718	119087645	A	G	1924	0.169	-0.067	0.039	2146	0
rs5910714	119083274	C	T	1924	0.169	-0.067	0.039	2146	0
rs138808720	119096689	C	G	1924	0.244	-0.064	0.035	2146	0
rs194320	119078123	T	C	1924	0.170	0.069	0.039	2146	0
rs74945102	119039414	G	A	1924	0.297	0.110	0.033	2146	0
rs5910730	119101709	C	T	1924	0.175	-0.059	0.040	2146	0
rs5910728	119101395	G	A	1924	0.235	-0.066	0.035	2146	0

association signal differed between the two methods. Specifically, no genes from the VEGAS2 analysis demonstrated evidence of association, yet MAGMA analysis suggested the genes *GPM6B*, *PRPS1*, *ZNF449* and *NRK* as likely candidate genes.

A competitive gene-set analysis was also performed using MAGMA. In this analysis, in which 10,468 gene-sets were tested for association with MSE at age 15 years, a total of 13 gene sets achieved a FDR < 0.05. Of these 13 gene sets, five had association test p-values < 0.05 after Bonferroni correction (Table 3.7). The gene *GPM6B* was common to all of the five aforementioned gene-sets.



**Table 3.5: The 10 genes demonstrating strongest association from the VEGAS2 gene-based association test.** Start and stop positions listed include  $\pm 50$  kb flanking regions. nSNPs = number of variants included in gene region; Test Statistic = gene-based  $\chi^2$  test statistic with *nSNPs* degrees of freedom; P-value = obtained from *Test Statistic* and adjusting for LD between variants; FDR = false discovery rate; Lead Variant = variant within gene locus with strongest association signal from previous SNP-based association test. Total number of genes tested = 1,252.

Gene	Start	Stop	nSNPs	Test Statistic	P-value	FDR	Lead Variant
<i>GPM6B</i>	13739061	14006831	468	1514.83	$2.59 \times 10^{-4}$	0.324	rs6633386
<i>ZNF449</i>	134428695	134547338	131	701.04	$9.60 \times 10^{-4}$	0.411	rs5930699
<i>ZNF75D</i>	134332535	134528012	177	794.71	$1.30 \times 10^{-3}$	0.411	rs5975507
<i>PRPS1</i>	106821653	106944256	49	254.12	$1.60 \times 10^{-3}$	0.411	rs9887704
<i>BCORL1</i>	129089163	129242058	94	279.94	$1.64 \times 10^{-3}$	0.411	rs875080
<i>MIR6086</i>	13558410	13658465	159	524.77	$2.43 \times 10^{-3}$	0.507	rs3747418
<i>LOC100129520</i>	124403968	124506950	151	933.25	$3.19 \times 10^{-3}$	0.571	rs3135278
<i>LOC100506790</i>	134480353	134581672	97	411.68	$4.04 \times 10^{-3}$	0.632	rs5930699
<i>KIAA1210</i>	118162597	118334542	178	485.18	$6.18 \times 10^{-3}$	0.640	rs189623102
<i>FRMPD3</i>	106715679	106898474	35	95.19	$6.43 \times 10^{-3}$	0.640	rs66874155

**Table 3.6: The 10 genes demonstrating strongest association from the MAGMA gene-based association test.** Start and stop positions listed include  $\pm 50$  kb flanking regions. nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate. Total number of genes tested = 803.

Gene	Start	Stop	nSNPs	Z-Statistic	P-value	FDR
<i>GPM6B</i>	13739062	14006861	479	3.82	$6.59 \times 10^{-5}$	0.045
<i>PRPS1</i>	106821654	106944256	52	3.58	$1.74 \times 10^{-4}$	0.045
<i>ZNF449</i>	134428696	134547338	157	3.57	$1.79 \times 10^{-4}$	0.045
<i>NRK</i>	105016536	105252602	201	3.51	$2.22 \times 10^{-4}$	0.045
<i>ZNF75D</i>	134369719	134528034	185	3.33	$4.29 \times 10^{-4}$	0.069
<i>LOC100129520</i>	124403698	124509063	160	2.85	$2.22 \times 10^{-3}$	0.296
<i>FRMPD3</i>	106719819	106898474	37	2.76	$2.91 \times 10^{-3}$	0.334
<i>BCORL1</i>	129064277	129242058	132	2.64	$4.11 \times 10^{-3}$	0.413
<i>CTAG1A</i>	153763418	153865075	41	2.47	$6.74 \times 10^{-3}$	0.559
<i>KIAA1210</i>	118162598	118334542	196	2.43	$7.45 \times 10^{-3}$	0.559

**Table 3.7: Gene-sets demonstrating FDR < 0.05 from MAGMA gene-set association test.** nGenes = number of genes included in gene set; Beta = gene-set test statistic; SE = standard error; FDR = false discovery rate; Bonferroni Adjusted P-value = P-value multiplied by the number of gene-sets tested. Total number of gene-sets tested = 10,468.

Gene-set	nGenes	Beta	SE	P-value	FDR	Bonferroni Adjusted P-value
CEBALLOS_TARGETS_OF_TP53_AND_MYC_DN	2	3.80	0.778	6.45 x 10 <sup>-7</sup>	0.002	0.007
WALLACE_PROSTATE_CANCER_DN	1	3.94	0.808	7.08 x 10 <sup>-7</sup>	0.002	0.007
ROSS_LEUKEMIA_WITH_MLL_FUSIONS	1	3.94	0.808	7.08 x 10 <sup>-7</sup>	0.002	0.007
PLASARI_TGFB1_SIGNALING_VIA_NFIC_10HR_UP	1	3.94	0.808	7.08 x 10 <sup>-7</sup>	0.002	0.007
RNTCANNRNNYNATTW_UNKNOWN	3	2.59	0.574	3.73 x 10 <sup>-6</sup>	0.008	0.039
GSE15324_NAIVE_VS_ACTIVATED_ELF4_KO_CD8_TCELL_DN	2	2.45	0.563	7.68 x 10 <sup>-6</sup>	0.013	0.080
GSE43955_TGFB_IL6_VS_TGFB_IL6_IL23_TH17_ACT_CD4_TCELL_60H_UP	4	1.77	0.409	8.89 x 10 <sup>-6</sup>	0.013	0.093
GCM_AQP4	2	2.60	0.605	1.04 x 10 <sup>-5</sup>	0.014	0.109
FINETTI_BREAST_CANCERS_KINOME_BLUE	1	2.33	0.565	2.15 x 10 <sup>-5</sup>	0.023	0.226
GSE41867_DAY6_VS_DAY15_LCMV_CLONE13_EFFECTOR_CD8_TCELL_UP	5	1.57	0.382	2.21 x 10 <sup>-5</sup>	0.023	0.232
HADDAD_B_LYMPHOCYTE_PROGENITOR	3	2.06	0.51	2.89 x 10 <sup>-5</sup>	0.027	0.302
GSE3920_IFNA_VS_IFNB_TREATED_ENDOTHELIAL_CELL_DN	6	1.53	0.394	5.49 x 10 <sup>-5</sup>	0.048	0.575

### **3.3.4 Power Calculation**

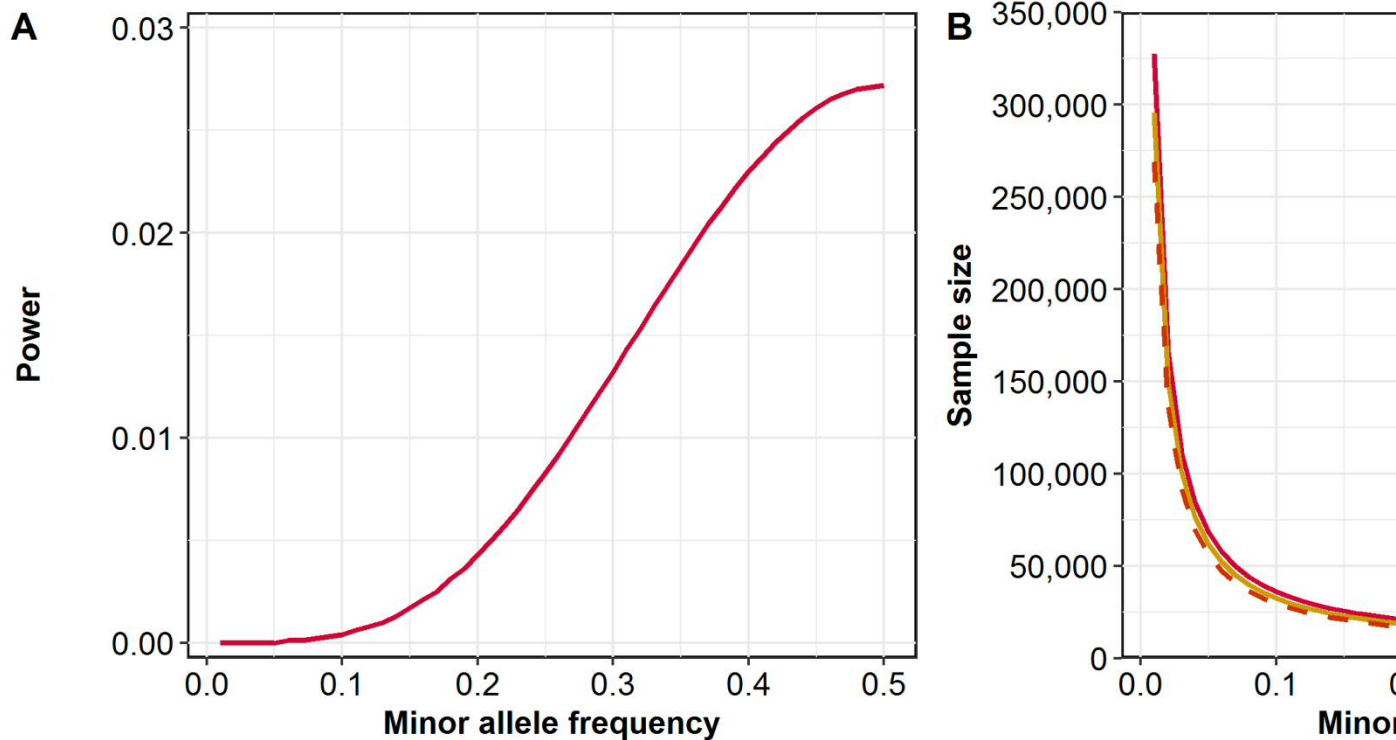
The power to detect associations in this current investigation was calculated in order to identify likely causes for the lack of evidence supporting variation on chromosome X with MSE at age 15 years. Table 3.8 shows that based on the sample size used for the current investigation, there was only 2.72% power to detect variants with an effect size of 0.10 D and MAF of 0.50 at  $p\text{-value} < 5 \times 10^{-8}$ . The power to detect associations decreased steadily for less common variants (Table 3.8, Figure 3.6A).

As the desired power to detect variants with true associations is typically defined as 80% (Cohen, 1992; Hong and Park, 2012), the power calculation was repeated to identify the sample size required to achieve this level of power. As shown in Figure 3.6B, a sample of approximately 35,000 individuals would have the ability to successfully detect 80% of true associations across variants with  $\text{MAF} > 0.1$ . A sample size of approximately 70,000 individuals would be required to detect variants with a MAF of 5%. To achieve 60% power for variants with MAF of 5%, the sample size required would be just over 55,000 individuals.

**Table 3.8: Power ( $\beta$ ) to detect an effect size of 0.10 D for variants with MAFs ranging between 0.01 and 0.50.** Based on a sample size of N = 4,070 for a normally distributed trait with mean (SD) = -0.382 (1.28) D and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency.

<b>MAF</b>	<b>Power (%)</b>	<b>MAF</b>	<b>Power (%)</b>
<b>0.01</b>	< 0.01	<b>0.26</b>	0.92
<b>0.02</b>	< 0.01	<b>0.27</b>	1.02
<b>0.03</b>	< 0.01	<b>0.28</b>	1.12
<b>0.04</b>	< 0.01	<b>0.29</b>	1.22
<b>0.05</b>	< 0.01	<b>0.30</b>	1.32
<b>0.06</b>	0.01	<b>0.31</b>	1.43
<b>0.07</b>	0.01	<b>0.32</b>	1.53
<b>0.08</b>	0.02	<b>0.33</b>	1.64
<b>0.09</b>	0.03	<b>0.34</b>	1.74
<b>0.10</b>	0.04	<b>0.35</b>	1.84
<b>0.11</b>	0.06	<b>0.36</b>	1.94
<b>0.12</b>	0.08	<b>0.37</b>	2.04
<b>0.13</b>	0.10	<b>0.38</b>	2.13
<b>0.14</b>	0.13	<b>0.39</b>	2.22
<b>0.15</b>	0.17	<b>0.40</b>	2.30
<b>0.16</b>	0.21	<b>0.41</b>	2.37
<b>0.17</b>	0.25	<b>0.42</b>	2.44
<b>0.18</b>	0.31	<b>0.43</b>	2.50
<b>0.19</b>	0.36	<b>0.44</b>	2.56
<b>0.20</b>	0.43	<b>0.45</b>	2.61
<b>0.21</b>	0.50	<b>0.46</b>	2.65
<b>0.22</b>	0.57	<b>0.47</b>	2.68
<b>0.23</b>	0.65	<b>0.48</b>	2.70
<b>0.24</b>	0.74	<b>0.49</b>	2.71
<b>0.25</b>	0.83	<b>0.50</b>	2.72

**Figure 3.6: Power calculations.** Panel A: Power to detect variants with an effect size of 0.10 D for the between 0.01 and 0.50. Based on a sample size of  $N = 4,070$  for a normally distributed trait with mean (SD) rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency. Panel B: Minimum sample sizes required to detect variant 80% power. Based on a normally distributed trait with mean MSE (SD):  $-0.382 (1.28)$  D and Type I error frequency.



### 3.4 Discussion

In summary, no variants on chromosome X demonstrated evidence of association with MSE at age 15 years in this sample of 4,070 unrelated individuals. However, the genes *GPM6B*, *PRPS1*, *ZNF449* and *NRK* should be considered as potential candidates warranting further investigation.

*GPM6B* (Neuronal membrane glycoprotein M6-b), located at Xp22.2, has been associated with extracellular matrix assembly, negative regulation of focal adhesion assembly, positive regulation of bone mineralisation and regulation of actin cytoskeleton organisation (Drabek et al., 2011). Located at Xq22.3 are the genes *PRPS1* (Phosphoribosyl Pyrophosphate Synthetase 1) and *NRK* (Nike Related Kinase). *PRPS1* is associated with X-linked recessive Charcot-Marie-Tooth disease-5 (Kim et al., 2007) and non-syndromic deafness (Liu et al., 2010c); while *NRK* is known to be expressed during the latter stages of embryogenesis (Nakano et al., 2000). In addition, zinc finger proteins 449 and 75 (*ZNF449* and *ZNF75D*) appear to play roles as transcription factors with both genes located at Xq26.3 (Luo et al., 2006) and in the list of the ten most highly-ranked genes from gene-based analyses (Tables 3.5 and 3.6). These genes reside close to the known myopia susceptibility locus, *MYP13* (Xq23-25).

Although not highlighted as a suggestive association in the gene-based test, the gene, *CHRD1* (Xq23) resides within the highlighted gene-set RNTCANNRNNYNATTW\_UNKNOWN. *CHRD1* is known to be expressed in the retina during ocular development and has been associated with megalocornea (Webb et

al., 2012; Davidson et al., 2014). A *CHRD1* gene knockdown investigation in *Xenopus* observed increased eye size as a feature in some cases (Pfirrmann et al., 2015).

The gene-sets demonstrating strongest association are related to cell differentiation, transcription factors and growth factors. As myopia is most commonly characterised by an axial elongation of the eye, alterations to these biological processes are plausible causes for myopia development.

A key consideration for this study is the use of non-cycloplegic autorefraction. Whilst the ideal scenario is to use cycloplegia when obtaining refractive error measures in children (Fotedar et al., 2007; Sankaridurg et al., 2017), longitudinal studies such as ALSPAC that investigate a wide range of traits, not just ocular traits, require methods that retain the near visual acuity of participants after their ocular assessment. Such studies also need to minimise dropout rates over time, and this is helped by avoidance of cycloplegic drops. However, the lack of cycloplegia can increase the variability in autorefraction results due to the retained accommodative response, which also results in a myopic shift and reduced repeatability of non-cycloplegic autorefraction readings (Williams et al., 2008a). A previous comparison of non-cycloplegic autorefraction and spectacle prescriptions from optometrists in this sample of ALSPAC children found a mean difference ( $\pm$  SD) of  $-0.22 \pm 0.84$  D (Northstone et al., 2013). This comparison suggests that this myopia shift using non-cycloplegic autorefraction should not greatly influence the mean spherical equivalent measures of the cohort overall; however, the relatively high variability in



these measures does further reduce the power of this study to detect likely associated variants.

Although Bonferroni adjustment is the simplest method of applying multiple testing correction to association test p-values, its primary assumption is that all association tests are independent of each other. Thus, this method of adjustment is overly conservative in genetic studies as variants and genes are not independent entities but are related to each other through patterns of LD. Hence, the genome-wide significance p-value threshold of  $5 \times 10^{-8}$  was chosen to define sufficiently strong association for SNP-based tests and a FDR of 5% was applied for gene-based tests.

Additionally, it is often the case that association studies apply adjustment for genomic inflation arising due to population stratification or other unaccounted confounding (Price et al., 2010). However, genomic control correction was not performed in this investigation, since there was no evidence for inflation of p-values; and this adjustment has been shown to be overly simplistic as it involves applying the same level of correction to all variants, whereas in reality variants are affected to different extents (Price et al., 2006; Schork et al., 2013).

Gene-based tests can be more powerful than single variant tests for complex traits where multiple variants are likely to be causal (Neale and Sham, 2004). On their own, individual variants exert a very small effect on phenotype, yet it is often the case that associated variants are enriched in specific regions. Therefore, pooling variants within a defined locus combines their individual effects and increases the

likelihood of a true association being detected. In addition, as the number of genes is considerably less than the number of individual variants, fewer tests are performed. Thus less stringent multiple testing correction is required for gene-based tests (Liu et al., 2010b).

Gene-based tests performed using VEGAS2 and MAGMA yielded results with different strength association signals after adjustment for multiple testing. It is important to note that despite both methods using 1000 Genomes project reference data mapped to NCBI build 37 coordinates, subtle differences could have arisen from the use of differing releases of the same genome build. Here, this genome build-related phenomenon led to different numbers of genes being tested for association with VEGAS2 (1,252 genes) and MAGMA (803 genes). The two gene-based tests also use different methods to account for LD between SNPs in each gene.

Whilst this XWAS failed to demonstrate association of variants on chromosome X with MSE at age 15 years, this is most likely due to the relatively small sample size used (Figure 3.6). As there were over 260,000 variants included in the SNP-based test after quality control and approximately 1000 genes in the gene-based investigations, stringent multiple testing correction was required to determine whether the association signal detected was likely to represent a true association between variant/gene and phenotype.

GWAS that have successfully identified variants associated with refractive error have analysed sample sizes from 4,270 to 45,771 (Hysi et al., 2010; Solouki et al., 2010; Kiefer et al., 2013; Verhoeven et al., 2013). As demonstrated by the second power calculation (Figure 3.6B), a sample of approximately 150,000 individuals would be required to detect successfully 80% of true associations across variants with  $MAF > 0.1$  and effect size  $> 0.10$  D. With ever-increasing sample sizes, this ability to detect true associations extends to rarer variants. As data from larger cohorts, for example UK Biobank (Sudlow et al., 2015), become available, power to detect true associations of variants on chromosome X with MSE will improve considerably.

Despite this current investigation having failed to identify any genetic variants on chromosome X associated with myopia, there are potential candidate loci warranting further investigation for their role in myopia development in the general population (Craddock, O'Donovan and Owen, 2008). Thus, it is anticipated some of the X-chromosome variants associated with refractive error will be detected when a GWAS is repeated using data from the UK Biobank.

**Chapter 4 CREAM Corneal  
Astigmatism GWAS Meta-analysis**

## Chapter 4 CREAM Consortium GWAS Meta-analysis for Corneal Astigmatism

### 4.1 Introduction

Despite knowledge that genetic factors make a notable contribution to the development of corneal and refractive astigmatism, with the proportion of trait variance attributable to genetic effects (heritability) estimated at 50-65% (Clementi et al., 1998; Hammond et al., 2001; Dirani et al., 2006; Grijbovski et al., 2006; Kim et al., 2013), genome-wide association studies for astigmatism have lagged behind those for spherical equivalent refractive error and myopia.

Early GWAS for corneal astigmatism showed promise with identification of a single locus in the promoter region of *PDGFRA* on chromosome 4 (lead variant: rs7677751,  $P = 7.87 \times 10^{-9}$ ) associated with this trait in 8,513 individuals of Asian ancestry by Fan et al. (2011). One other study has successfully replicated association of this locus with corneal astigmatism in a European ancestry cohort (Guggenheim et al., 2013a); however, no other loci have demonstrated genome-wide significant association ( $P < 5 \times 10^{-8}$ ) with this trait. A large-scale GWAS for the related trait of refractive astigmatism undertaken by the Consortium for Refractive Error and Myopia (CREAM) in 45,931 individuals identified only a single locus demonstrating genome-wide significant association, near the *NRXN1* gene on chromosome 2 (lead variant: rs1401327,  $P = 3.92 \times 10^{-8}$ ) (Li et al., 2015a). These results are in contrast to those for the highly heritable refractive traits of spherical equivalent and myopia, for which over a hundred variants demonstrating genome-wide significant

association have been identified (Kiefer et al., 2013; Verhoeven et al., 2013; Pickrell et al., 2016).

Since refractive astigmatism is a combination of corneal and lenticular influences, it could be argued that this paucity in identifying genome-wide significant variants may be due to phenotypic uncertainty. As a result, the CREAM consortium organised a GWAS meta-analysis study for corneal astigmatism, using the same phenotype definitions as the earlier study by Fan et al. (2011) in order to identify novel susceptibility loci for this trait. I carried out the work described in this chapter on behalf of CREAM.

## **4.2 Methods**

This investigation followed an analysis plan that was agreed upon by members of CREAM before starting work. The plan was designed to standardise methods across participating CREAM groups in order to facilitate meta-analysis. All study groups known to CREAM with relevant genotype and phenotype data were invited to contribute to the study.

### **4.2.1 Study Cohorts**

A total of 31,370 individuals participated in this meta-analysis of corneal astigmatism and their demographics are summarised in Table 4.1. This sample consisted of 22,250 individuals of European ancestry from 14 studies and 9,120 individuals of Asian ancestry from 8 studies. Of these individuals, 5,470 European and 947 Asian participants were under 25 years of age.

Acknowledgements and study information for each of the participating cohorts in this meta-analysis are included in Appendix A. Ethical approval for the study was obtained locally by each CREAM study group, and participants gave informed consent. All phenotyping, genotyping, imputation and individual GWAS were undertaken by the respective study groups prior to submission of their summary statistics to a central site for meta-analysis.

**Table 4.1: Subject demographics of participating CREAM study groups.**

Study	Ancestry	N (Cases/Controls)	%Female	Age (Years)	Corneal
				Mean (SD)	Median (IQ)
ALSPAC	European	2279 (985/1294)	53.1	15.5 (0.3)	0.683 (0.469-0.907)
BMES	European	1238 (720/518)	41.8	73.3 (7.6)	0.863 (0.565-1.161)
EPIC	European	857 (456/401)	58.5	68.7 (7.4)	0.780 (0.527-1.033)
FITSA	European	127 (62/65)	100	67.9 (3.1)	0.733 (0.530-0.936)
GenerationR	European	2071 (981/1090)	49.9	6.09 (0.4)	0.725 (0.480-0.970)
GHS1 <sup>1</sup>	European	2398 (1003/1395)	48.7	55.9 (10.8)	0.650 (0.400-0.900)
GHS2 <sup>1</sup>	European	851 (383/468)	50.9	55.1 (10.8)	0.650 (0.450-0.850)
RAINE	European	1028 (407/621)	50.9	20.0 (0.4)	0.649 (0.445-0.853)
Rotterdam-I	European	5537 (2064/3473)	59.3	69.5 (9.2)	0.601 (0.334-0.868)
Rotterdam-II	European	1982 (633/1349)	53.8	64.8 (8.0)	0.539 (0.294-0.784)
Rotterdam-III	European	2925 (1180/1745)	56.2	57.0 (6.9)	0.618 (0.356-0.880)
OGP-A <sup>2</sup>	European	92 (37/55)	44.6	16.0 (4.5)	0.682 (0.512-0.852)
OGP-B <sup>2</sup>	European	446 (181/265)	43.7	50.6 (15.4)	0.650 (0.430-0.870)
TwinsUK	European	419 (201/218)	92.7	64.0 (10.5)	0.729 (0.476-1.000)
BES-610K <sup>3</sup>	Asian	553 (240/313)	65.6	62.1 (8.4)	0.666 (0.407-0.925)
BES-OmniE <sup>3</sup>	Asian	469 (208/261)	60.1	64.7 (9.5)	0.676 (0.429-0.923)
SCES-610K <sup>3</sup>	Asian	1745 (787/958)	48.7	57.6 (9.0)	0.703 (0.476-0.930)
SCES-OmniE <sup>3</sup>	Asian	545 (257/288)	48.6	59.2 (8.8)	0.723 (0.470-0.976)
SCORM	Asian	947 (768/179)	48.6	10.8 (0.8)	1.205 (0.851-1.559)
SIMES	Asian	1778 (750/1028)	51.7	59.5 (10.8)	0.662 (0.432-0.892)
SINDI	Asian	2261 (814/1447)	48.6	56.5 (9.1)	0.614 (0.411-0.817)
STARS	Asian	822 (525/297)	50.0	38.5 (5.3)	1.000 (0.625-1.375)

<sup>1</sup>Association tests were undertaken separately for samples recruited in different waves.

<sup>2</sup>Association tests were undertaken separately for different age strata (stratum A, age > 3 and < 25 years; stratum B, age > 25 and < 45 years; stratum C, age > 45 years).

<sup>3</sup>Association tests were undertaken separately for samples genotyped on different platforms.



#### 4.2.1.1 Phenotypes

Anterior corneal curvature was measured using keratometry (the keratometer used by each CREAM study group is listed in Table 4.2), with corneal astigmatism defined as the difference in corneal curvature between the steepest and flattest meridians. Corneal astigmatism was averaged between the two eyes, except for individuals with data available for only one eye. To convert keratometry readings from millimetres to dioptres, a conversion factor of 332 divided by the keratometry reading (in mm) was applied (Bennett and Rabbetts, 1989). Individuals known to have keratoconus, corneal scarring, ocular surgery, or any corneal/ocular condition that would impair keratometry were excluded from the analysis.

**Table 4.2: Instrument for measuring corneal curvature.**

<b>Study</b>	<b>Instrument</b>
ALSPAC	IOLmaster
BMES	IOLmaster
EPIC	IOLmaster
FITSA	IOLmaster
GenerationR	IOLmaster
GHS 1	Lenstar LS 900
GHS 2	Lenstar LS 900
RAINE	IOLmaster
Rotterdam-I	Javal Keratometer/Lenstar LS 900
Rotterdam-II	Javal Keratometer/Lenstar LS 900
Rotterdam-III	Javal Keratometer/Lenstar LS 900
OGP-A	IOLmaster
OGP-B	IOLmaster
TwinsUK	Visionix VX-120
BES-610K	Lenstar LS 900
BES-OmniE	Lenstar LS 900
SCES-610K	IOLmaster
SCES-OmniE	IOLmaster
SCORM	Canon RK-5
SIMES	IOLmaster
SINDI	IOLmaster
STARS	IOLmaster

As per Fan et al. (2011), corneal astigmatism was considered as a dichotomous trait with cases defined as individuals with corneal astigmatism  $> 0.75$  D, and controls defined as those with corneal astigmatism  $\leq 0.75$  D.

#### *4.2.1.2 Genotyping and Imputation*

DNA samples were extracted from blood or saliva and genotyped on a high-density SNP platform, as described by Li et al. (2015a). Imputation was subsequently performed by each participating study group using an ancestry matched reference panel from the 1000 Genomes Project and the IMPUTE2 or Minimac software packages (Howie et al., 2009; Li et al., 2010). All variants for all cohorts were mapped to NCBI Build 37 coordinates.

Quality-control filtering of variants was performed following Fan et al. (2016). In general, variants with per-study missingness  $> 0.05$ , minor allele frequency (MAF)  $< 0.05$ , Hardy–Weinberg disequilibrium test p-value  $< 1 \times 10^{-6}$ , poor imputation quality (IMPUTE2 info  $\leq 0.5$  or Minimac Rsq  $\leq 0.5$ ) were excluded, along with samples with per-study missingness  $> 0.05$ , extreme heterozygosity, sex mismatch, and those with unaccounted-for relatedness or outlying ancestry.

#### *4.2.1.3 Genome-wide Association Studies*

Participating study groups performed genome-wide single marker association tests using logistic regression models in their respective cohorts for the phenotype of corneal astigmatism case/control status. If genetic variants were coded as 0, 1 or 2, the analysis was performed using PLINK (Purcell et al., 2007); however, if variants

were coded as imputed dosages on the 0-2 scale, the mach2dat (Li et al., 2010) or ProbABEL (Aulchenko, Struchalin and van Duijn, 2010) software packages were employed. Age was included as a continuous covariate and sex as a binary covariate in analyses. If individual studies had evidence of population stratification from Q-Q plots or the genomic control inflation factor ( $\lambda_{GC}$ ), the first five major principal components were also included as continuous covariates. For cohorts including related individuals, genetic background was accounted for by treating this as a random effect in the analysis model. GWAS were performed separately for individuals in the younger (3 years < age < 25 years) and older (age  $\geq$  25 years) age groups and for individuals of European and Asian ancestries.

#### **4.2.2 Quality Control**

Prior to meta-analysis, all GWAS summary statistics received were reviewed to ensure all marker-level quality control had been performed as requested (i.e. exclusion of variants with MAF < 0.05, imputation quality score < 0.5, HWE test p-value <  $1 \times 10^{-6}$ ).

Following this initial quality control procedure, all summary statistics underwent study-wide quality control procedures as implemented in EasyQC (Winkler et al., 2014). Quality control plots and metrics were generated and evaluated including effect allele frequency (EAF) plots, p-value versus Z-score (P-Z) plots, standard error versus sample size (SE-N) plots, and genomic control inflation factors (Appendix B). Queries regarding the summary statistics were resolved through discussion with

study group analysts, and, where indicated, imputation or association testing was repeated by the respective study groups.

### **4.2.3 Meta-analysis of GWAS Results**

Meta-analyses were initially performed separately according to ancestry and age group as performed by Li et al. (2015a), namely: European ancestry under 25 years old (4 cohorts), European ancestry over 25 years-old (10 cohorts), Asian ancestry under 25 years old (1 cohort), and Asian ancestry over 25 years-old (7 cohorts). In all instances, a fixed-effects, inverse variance weighted approach was used in METAL as outlined in *Section 2.3.3*. To determine whether association signals were significant, a p-value of  $5 \times 10^{-8}$  was adopted for declaring genome-wide significant association in the GWAS meta-analyses (*Section 2.3.7*). Regional association plots for genome-wide significant loci were created using LocusZoom (Pruim et al., 2010). Conditional analysis was performed on the summary statistics from meta-analysis using GCTA-COJO (Yang et al., 2012) as outlined in *Section 2.3.4*.

#### **4.2.3.1 Power Calculation**

The statistical power of the all-cohorts meta-analysis was calculated using the software application QUANTO (Gauderman, 2002). This power calculation was based on an unmatched case-control design, with a case:control ratio of 1:2.3, as determined from the all-cohorts meta-analysed sample data. The following assumptions were applied when performing the calculation: a type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ , an odds ratio of 1.1, a MAF range between 0.01 and 0.50, a sample size of 31,370 individuals, and a population trait prevalence of 43.5%. The sample size

was taken as that used for this meta-analysis, with the population prevalence calculated as the average for all of the population-based studies contributing to this analysis.

#### **4.2.4 Gene-based and Gene-set Analyses**

Gene-based tests were subsequently performed for individuals of European ancestry using VEGAS2 and MAGMA (outlined in *Section 2.3.5*). In both instances, LD patterns were estimated using European ancestry reference panels, specifically reference files composed of data for the 379 unrelated individuals of European ancestry from Phase 1, Version 3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2012). These analyses were restricted to individuals of European ancestry since gene-based tests require consideration of LD patterns between variants within gene regions, which vary for different ancestries. In addition, the sample size contributing to the European ancestry meta-analysis was much greater than for the respective Asian ancestry meta-analysis (22,250 vs. 9,120 individuals). In addition to inclusion of 50 kb flanking regions to the gene transcription start and stop sites, exploratory analyses using flanking regions of 200 kb were also conducted in gene-based tests using MAGMA.

Potential functional properties of genes associated with corneal astigmatism in the MAGMA analysis were further investigated using competitive gene-set analysis in MAGMA. Adjustment for multiple testing was applied using a false discovery rate of 5% for these gene-based and gene-set test results.

#### **4.2.5 SNP-Heritability and Genetic Correlation**

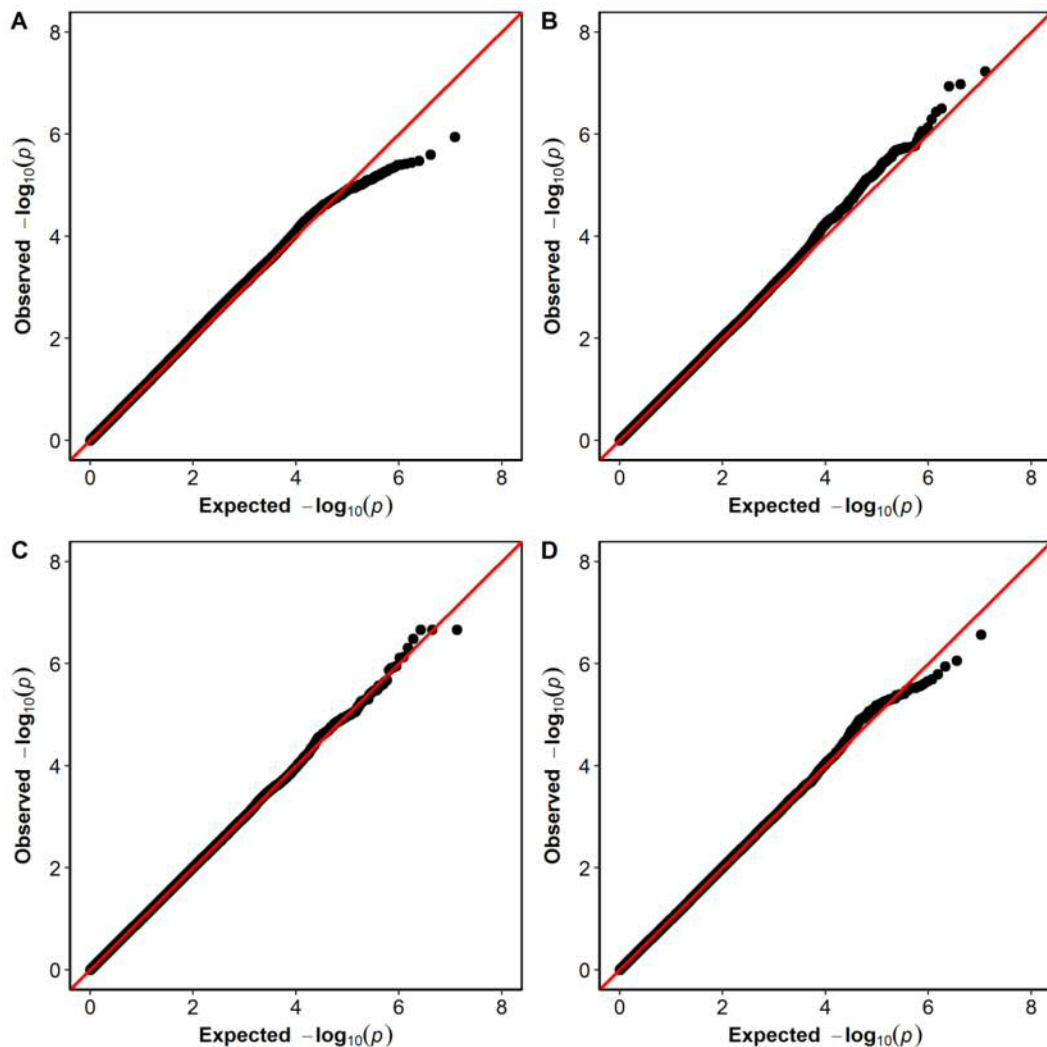
Using LDSC (Bulik-Sullivan et al., 2015a; Bulik-Sullivan et al., 2015b), the shared genetic contributions (i.e. genetic correlation) between corneal astigmatism and the two related traits of refractive astigmatism and spherical equivalent refractive error were quantified (see *Section 2.3.2.2* for details of the LD score regression technique). The heritability explained by commonly-occurring genetic variants (SNP-heritability) for each of the three traits was also estimated, as part of the LDSC analysis. GWAS summary statistics for refractive error and refractive astigmatism were obtained from previous CREAM meta-analysis studies of these traits (Verhoeven et al., 2013; Li et al., 2015a). The LDSC analyses were restricted to individuals of European ancestry since large sample sizes are required and reference LD scores need to be obtained from an ancestry-matched reference panel. As analyses of corneal astigmatism and refractive astigmatism were performed as dichotomous trait studies, their respective case prevalences in the general population were required. These were taken as 42% and 26% respectively, and were calculated as the average for the European ancestry population-based studies contributing to the respective analyses.

## 4.3 Results

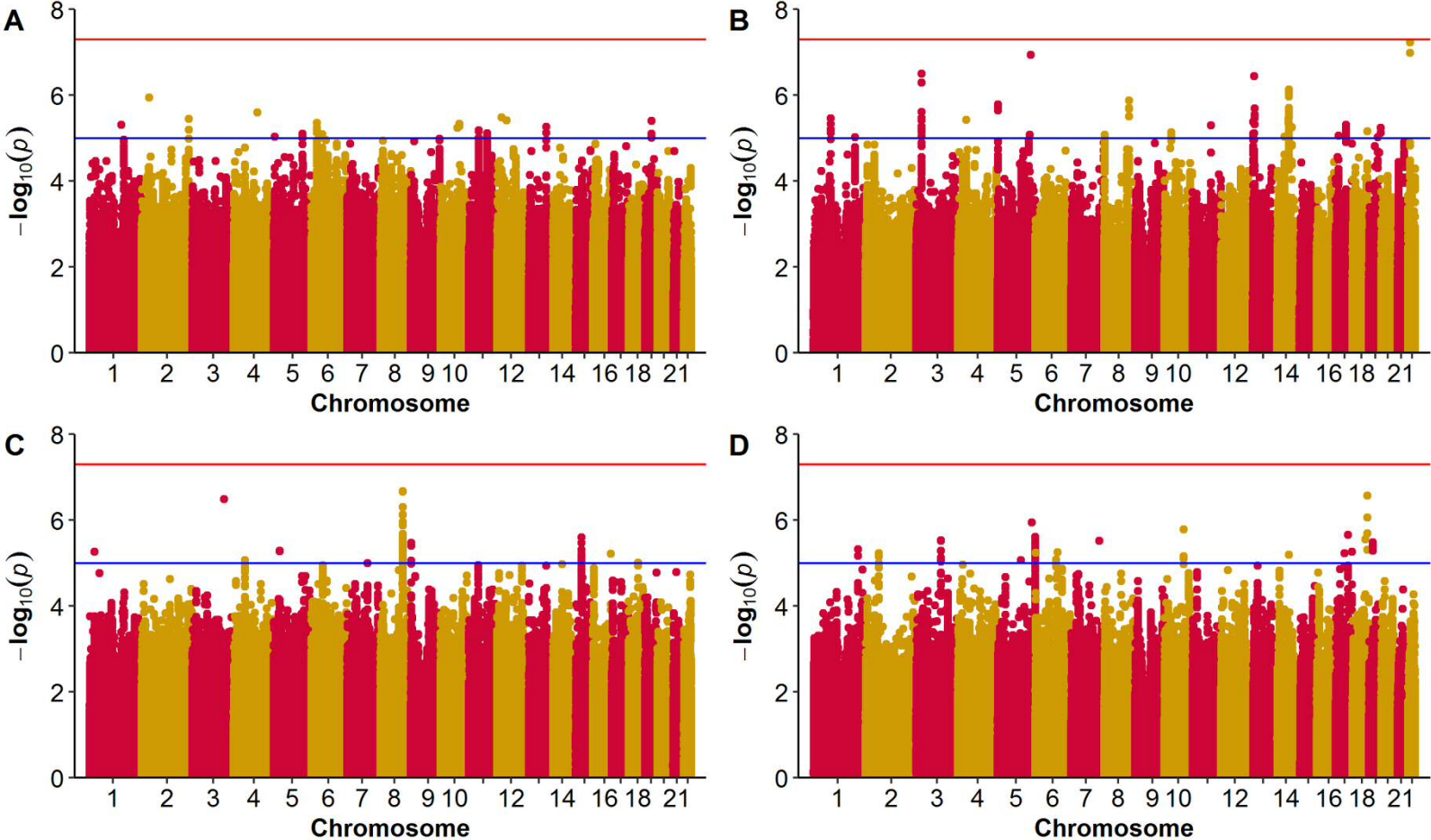
### 4.3.1 Meta-analysis of GWAS Results

Inverse variance weighted meta-analyses were performed for each of the four strata (age under/over 25 years and European/Asian ancestry) across approximately six million genetic variants. In each of the four strata, no locus achieved the pre-defined genome-wide significant level of association ( $P < 5 \times 10^{-8}$ ) (Tables 4.3-4.6, Figure 4.1 and Figure 4.2).

**Figure 4.1: Quantile-quantile plots for the separate ancestry/age strata fixed effects meta-analyses.** Panel A: European ancestry, aged > 25 years; Panel B: European ancestry, aged < 25 years; Panel C: Asian ancestry, aged > 25 years; Panel D: Asian ancestry, aged < 25 years. Y-axes show observed negative  $\log_{10}$  p-values and X-axes show expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).



**Figure 4.2: Manhattan plots for the separate ancestry/age strata fixed effects meta-analyses.** Panel A: European ancestry, aged > 25 years; Panel B: European ancestry, aged < 25 years; Panel C: Asian ancestry, aged > 25 years; Panel D: Asian ancestry, aged < 25 years. Y-axes show negative log<sub>10</sub> p-values and X-axes show genomic position. Red line corresponds to  $P = 5 \times 10^{-8}$ , blue line corresponds to  $P = 1 \times 10^{-5}$ .





**Table 4.3: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all European ancestry participants aged > 25 years.** EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	OR (95% CI)	P-value	Nearest Gene
rs1620100	2	36179888	T	C	0.41	0.91 (0.87-0.94)	$1.14 \times 10^{-6}$	<i>MRPL50P1</i>
rs138222255	4	113801232	A	T	0.89	0.82 (0.76-0.89)	$2.52 \times 10^{-6}$	<i>ANK2</i>
rs73076614	12	22633729	T	C	0.95	1.24 (1.13-1.36)	$3.32 \times 10^{-6}$	<i>C2CD5</i>
rs10187347	2	227348324	T	C	0.47	0.91 (0.88-0.95)	$3.58 \times 10^{-6}$	<i>MIR5702</i>
rs11183146	12	46024937	A	G	0.54	0.84 (0.78-0.90)	$3.82 \times 10^{-6}$	<i>LINC00938</i>
rs4804959	19	31803331	T	C	0.61	1.12 (1.07-1.17)	$3.95 \times 10^{-6}$	<i>TSHZ3</i>
rs9393856	6	27767825	T	G	0.08	0.82 (0.75-0.89)	$4.37 \times 10^{-6}$	<i>TRNAQ10</i>
rs2184695	10	89890390	A	C	0.53	0.91 (0.87-0.95)	$4.66 \times 10^{-6}$	<i>MED6P1</i>
rs35587414	1	153174958	T	C	0.15	1.15 (1.08-1.23)	$4.84 \times 10^{-6}$	<i>LELP1</i>
rs5806282	13	102849565	I	R	0.75	1.12 (1.07-1.17)	$5.41 \times 10^{-6}$	<i>FGF14</i>

**Table 4.4: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all European ancestry participants aged < 25 years.** EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	OR (95% CI)	P-value	Nearest Gene
rs6005414	22	27783185	A	G	0.53	0.80 (0.74-0.87)	$5.83 \times 10^{-8}$	<i>LINC02554</i>
rs17056435	5	158453783	A	G	0.05	2.19 (1.64-2.92)	$1.14 \times 10^{-7}$	<i>EBF1</i>
rs2596618	3	24221506	A	G	0.11	1.38 (1.22-1.57)	$3.16 \times 10^{-7}$	<i>THRB</i>
rs12859952	13	26381179	A	G	0.77	0.77 (0.69-0.85)	$3.61 \times 10^{-7}$	<i>ATP8A2</i>
rs9652372	14	75377345	T	C	0.09	0.70 (0.61-0.81)	$7.43 \times 10^{-7}$	<i>RPS6KL1</i>
rs139812140	8	118821440	D	R	0.16	0.75 (0.67-0.85)	$1.34 \times 10^{-6}$	<i>EXT1</i>
rs7702605	5	2653180	C	G	0.88	0.75 (0.67-0.85)	$1.66 \times 10^{-6}$	<i>IRX2</i>
rs12130807	1	82036836	A	C	0.09	0.71 (0.61-0.82)	$3.43 \times 10^{-6}$	<i>ADGRL2</i>
rs202172201	4	40109940	I	R	0.09	1.41 (1.22-1.62)	$3.79 \times 10^{-6}$	<i>N4BP2</i>
rs9581136	13	19639265	T	C	0.06	0.65 (0.54-0.78)	$4.20 \times 10^{-6}$	<i>GTF2IP3</i>

**Table 4.5: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all Asian ancestry participants aged > 25 years.** EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	OR (95% CI)	P-value	Nearest Gene
rs16875983	8	108285287	T	G	0.25	1.22 (1.13-1.32)	$2.14 \times 10^{-7}$	<i>ANGPT1</i>
rs67687099	3	152747454	D	R	0.20	1.37 (1.22-1.55)	$3.25 \times 10^{-7}$	<i>HMG2P13</i>
rs35026266	15	48769044	T	C	0.31	0.84 (0.79-0.91)	$2.52 \times 10^{-6}$	<i>FBN1</i>
rs10809667	9	1209532	A	T	0.23	1.20 (1.11-1.30)	$3.37 \times 10^{-6}$	<i>RPS27AP14</i>
rs56738713	5	29586419	A	G	0.08	1.37 (1.20-1.57)	$5.17 \times 10^{-6}$	<i>UBL5P1</i>
rs3924436	1	24908959	A	G	0.41	0.86 (0.81-0.92)	$5.42 \times 10^{-6}$	<i>NCMAP</i>
rs1687660	16	86416646	C	G	0.52	0.86 (0.81-0.92)	$6.11 \times 10^{-6}$	<i>LINC00917</i>
rs58435984	4	55127990	T	C	0.78	0.84 (0.77-0.90)	$8.58 \times 10^{-6}$	<i>PDGFRA</i>
rs141310268	18	45633672	D	R	0.14	0.80 (0.72-0.88)	$9.87 \times 10^{-6}$	<i>ZBTB7C</i>
rs817755	7	98202672	A	C	0.91	1.42 (1.21-1.65)	$1.00 \times 10^{-5}$	<i>NPTX2</i>

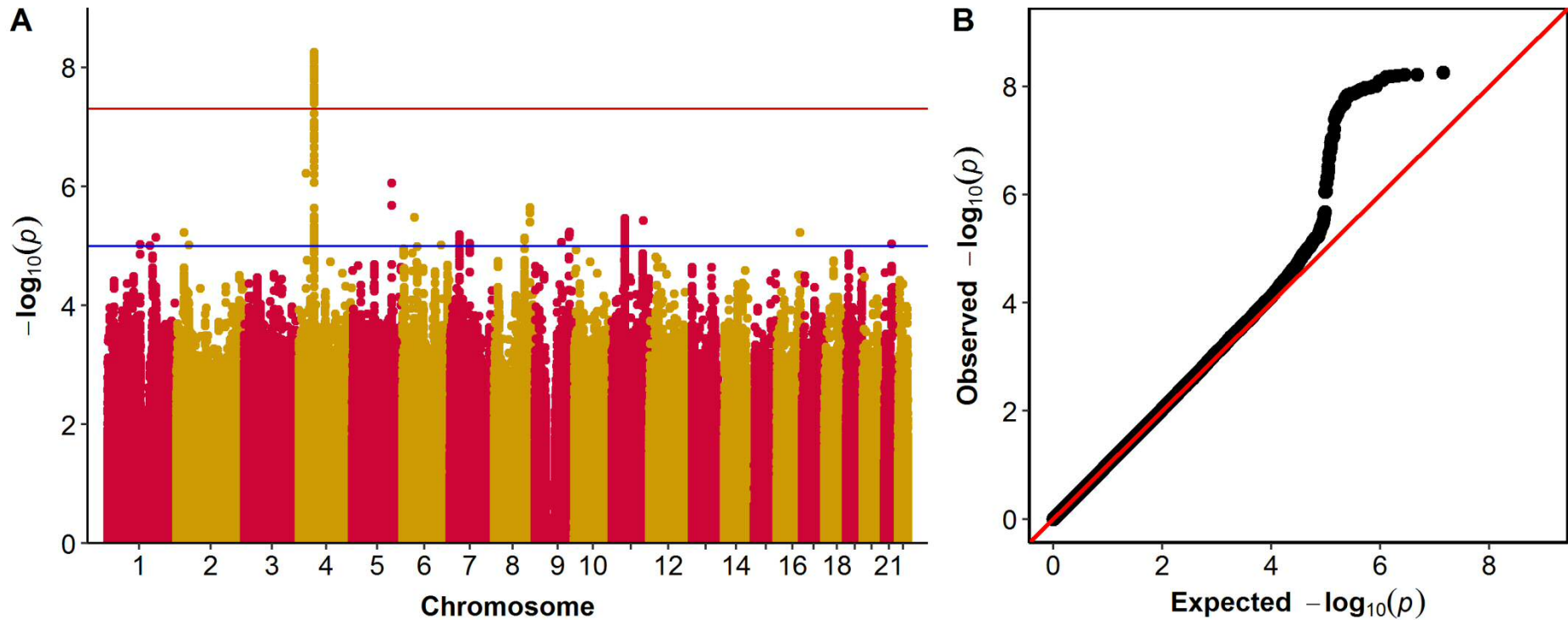
**Table 4.6: The 10 variants demonstrating strongest association in each region in the GWAS meta-analysis of all Asian ancestry participants aged < 25 years.** EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	OR (95% CI)	P-value	Nearest Gene
rs72971923	18	69822333	A	G	0.15	0.32 (0.21-0.50)	$2.68 \times 10^{-7}$	<i>LOC101927537</i>
rs2069368	5	162870726	C	G	0.94	7.30 (3.28-16.26)	$1.14 \times 10^{-6}$	<i>CCNG1</i>
rs7905017	10	92868457	A	G	0.62	1.98 (1.50-2.62)	$1.63 \times 10^{-6}$	<i>LINC00502</i>
rs11079429	17	59472403	A	G	0.29	0.48 (0.35-0.65)	$2.23 \times 10^{-6}$	<i>BCAS3</i>
rs9957	5	179290154	C	G	0.74	1.92 (1.47-2.53)	$2.48 \times 10^{-6}$	<i>TBC1D9B</i>
rs17072824	18	62003510	A	G	0.10	0.35 (0.23-0.54)	$2.80 \times 10^{-6}$	<i>LINC01924</i>
rs779593	3	118029874	T	C	0.69	1.81 (1.41-2.32)	$2.95 \times 10^{-6}$	<i>LOC101926968</i>
rs143004236	7	135420794	A	G	0.05	0.13 (0.06-0.31)	$3.00 \times 10^{-6}$	<i>FAM180A</i>
rs11085245	19	18859757	A	G	0.83	0.45 (0.32-0.63)	$3.31 \times 10^{-6}$	<i>CRTC1</i>
rs12144639	1	213817311	A	G	0.22	0.53 (0.40-0.70)	$4.75 \times 10^{-6}$	<i>PROX1-AS1</i>

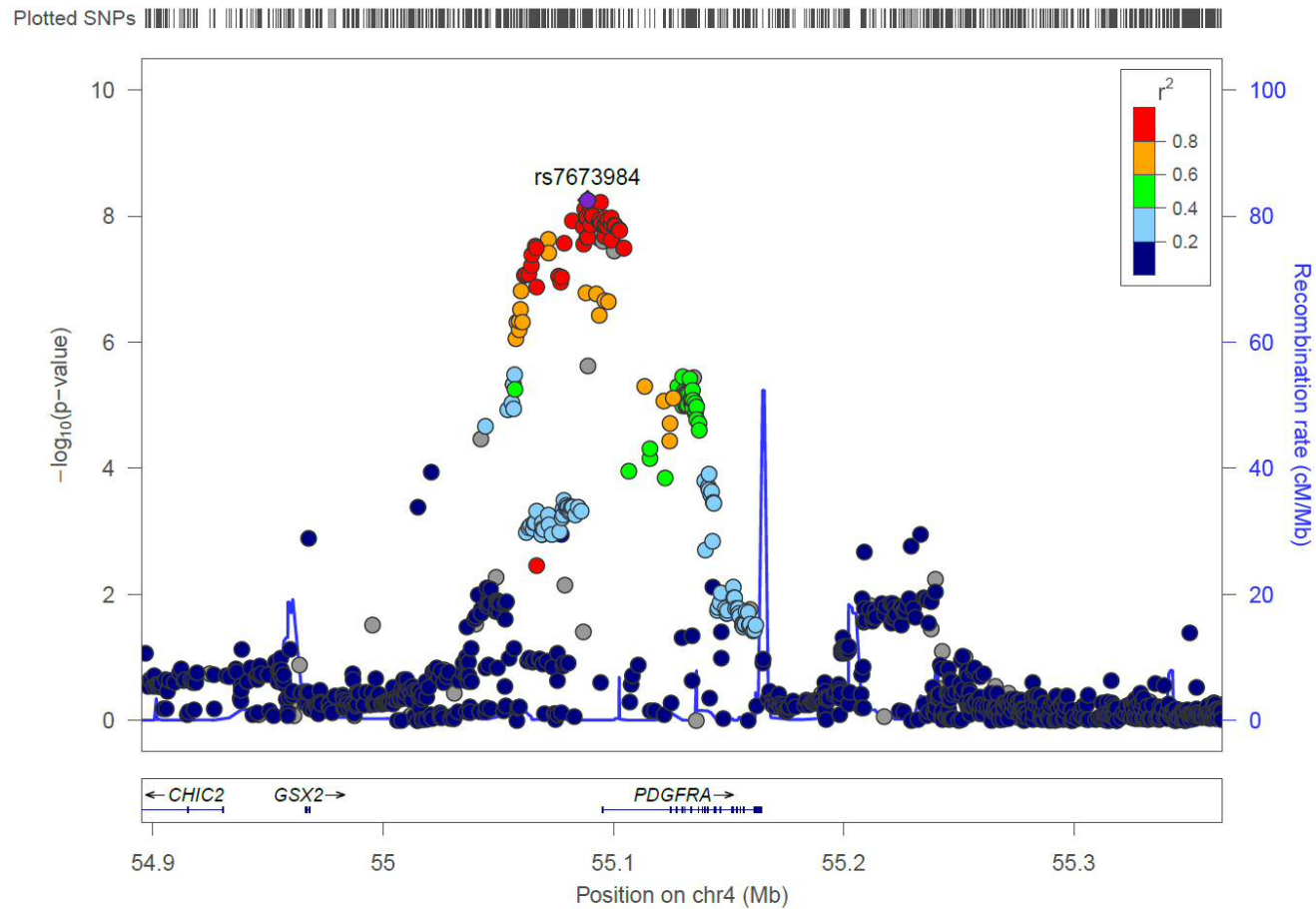
4 (Figure 4.3 and Figure 4.4), replicating the previous findings of Fan et al. (2011). Table 4.7 lists the strongest associated (lead) variants in regions demonstrating suggestive association ( $P < 1 \times 10^{-5}$ ) from this all-cohorts meta-analysis, with regions defined as  $\pm 500$  kb of the lead variant.

Both the European and Asian meta-analyses contributed to the association signal at the *PDGFRA* locus. The variant demonstrating strongest association in the all-cohorts meta-analysis, rs7673984, had an effect size (odds ratio) of OR = 1.15 (95% CI: 1.07-1.24;  $P = 1.76 \times 10^{-4}$ ) for the Asian ancestry cohort meta-analysis, OR = 1.11 (95% CI: 1.06-1.16;  $P = 5.64 \times 10^{-6}$ ) for the European ancestry cohort meta-analysis, and OR = 1.12 (95% CI: 1.08-1.16;  $P = 5.55 \times 10^{-9}$ ) in the all-cohorts (Asians and Europeans combined) meta-analysis. The association of rs7673984 in the individual cohorts examined is summarised in Figure 4.5. Conditional analysis using GCTA-COJO yielded no additional association signals at the *PDGFRA* locus independent of rs7673984.

**Figure 4.3: Manhattan and Quantile-Quantile plots for fixed effects meta-analysis of European and Asian participants of all ages combined (N = 31,370).** Panel A: Manhattan plot: Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Panel B: Quantile-Quantile plot: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).



**Figure 4.4: Region plot for the locus demonstrating strongest association in the GWAS fixed-effects meta-analysis for European and Asian participants of all ages combined (N = 31,370).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant rs7673984 (highlighted in purple).



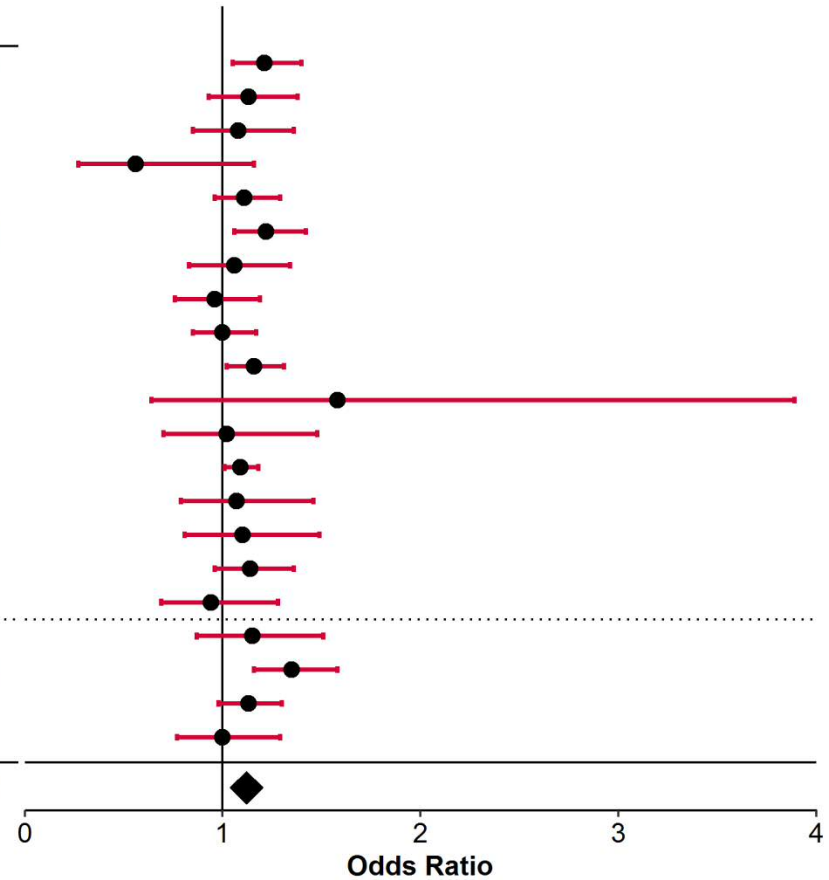
**Table 4.7: Variants demonstrating association signals of  $P < 1 \times 10^{-5}$  in each region in the GWAS meta-analysis of all samples (Europeans and Asians of all ages combined).** EAF = effect allele frequency; OR = odds ratio. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	OR (95% CI)	P-value	Nearest Gene
rs7673984	4	55088761	T	C	0.22	1.12 (1.08-1.16)	$5.55 \times 10^{-9}$	<i>PDGFRA</i>
rs34751092	4	24129037	A	G	0.28	1.09 (1.05-1.13)	$6.07 \times 10^{-7}$	<i>PPARGC1A</i>
rs630203	5	141444269	T	G	0.74	0.92 (0.88-0.95)	$8.83 \times 10^{-7}$	<i>MRPL11P2</i>
rs75607298	8	128611496	A	G	0.72	1.14 (1.08-1.21)	$2.28 \times 10^{-6}$	<i>CASC11</i>
rs62401199	6	43813341	T	C	0.14	1.15 (1.09-1.22)	$3.29 \times 10^{-6}$	<i>LINC01512</i>
rs753992	11	47349846	A	G	0.29	0.91 (0.87-0.95)	$3.48 \times 10^{-6}$	<i>MADD</i>
rs3214101	11	114009408	A	T	0.68	1.08 (1.05-1.12)	$3.75 \times 10^{-6}$	<i>ZBTB16</i>
rs10985068	9	123629724	C	G	0.12	1.13 (1.07-1.19)	$5.87 \times 10^{-6}$	<i>PHF19</i>
rs62128379	2	26960055	T	C	0.85	1.13 (1.07-1.19)	$6.00 \times 10^{-6}$	<i>KCNK3</i>
rs9939114	16	84023972	A	G	0.05	0.54 (0.41-0.70)	$6.01 \times 10^{-6}$	<i>NECAB2</i>
rs60083876	7	34228819	A	T	0.95	1.31 (1.17-1.48)	$6.53 \times 10^{-6}$	<i>BMPER</i>
rs859362	1	175495090	T	C	0.19	1.10 (1.05-1.15)	$7.14 \times 10^{-6}$	<i>TNR</i>
rs11775037	8	108317615	A	G	0.20	1.10 (1.05-1.14)	$7.31 \times 10^{-6}$	<i>ANGPT1</i>
rs7036824	9	96149894	T	C	0.94	0.83 (0.77-0.90)	$8.78 \times 10^{-6}$	<i>C9orf129</i>
rs142168171	7	71253651	I	R	0.09	1.37 (1.19-1.57)	$9.14 \times 10^{-6}$	<i>CALN1</i>
rs7278671	21	41047876	A	G	0.51	0.93 (0.90-0.96)	$9.31 \times 10^{-6}$	<i>B3GALT5</i>
rs191640722	1	119264997	C	G	0.09	0.87 (0.82-0.93)	$9.53 \times 10^{-6}$	<i>LOC100421281</i>
rs36107906	2	44162800	D	R	0.29	1.09 (1.05-1.13)	$9.61 \times 10^{-6}$	<i>LRPPRC</i>
rs4896367	6	138807281	T	C	0.72	1.09 (1.05-1.14)	$9.75 \times 10^{-6}$	<i>NHSL1</i>
rs35587414	1	153174958	T	C	0.15	1.13 (1.07-1.20)	$9.79 \times 10^{-6}$	<i>LELP1</i>



**Figure 4.5: Forest plot and summary table for lead variant rs7673984 across all cohorts.** Studies listed above the dotted line are new cohorts not included in the only prior GWAS for corneal astigmatism (Fan et al., 2011). EAF = effect allele frequency; OR = odds ratio. Horizontal bars denote 95% confidence intervals. rs7673984 was excluded from the Rotterdam-I cohort analysis during quality control filtering.

Study	N	EAF	OR (95%CI)	P-value
ALSPAC	2279	0.22	1.21 (1.05-1.40)	8.97 x 10 <sup>-3</sup>
BMES	1240	0.21	1.13 (0.93-1.38)	0.222
EPIC	857	0.23	1.08 (0.85-1.36)	0.526
FITSA	127	0.16	0.56 (0.27-1.16)	0.117
GenerationR	2071	0.22	1.11 (0.96-1.29)	0.166
GHS1	2398	0.21	1.22 (1.06-1.42)	6.43 x 10 <sup>-3</sup>
GHS2	851	0.21	1.06 (0.83-1.34)	0.658
RAINE	1028	0.21	0.96 (0.76-1.19)	0.689
Rotterdam-II	1982	0.22	1.00 (0.85-1.17)	0.979
Rotterdam-III	2925	0.22	1.16 (1.02-1.31)	0.024
OGP-A	92	0.14	1.58 (0.64-3.89)	0.324
OGP-B	446	0.14	1.02 (0.70-1.48)	0.921
TwinsUK	422	0.21	1.09 (1.01-1.18)	0.025
BES-610K	553	0.20	1.07 (0.79-1.46)	0.643
BES-OmniE	469	0.23	1.10 (0.81-1.49)	0.535
SCES-610K	1745	0.19	1.14 (0.96-1.36)	0.147
SCES-OmniE	545	0.19	0.94 (0.69-1.28)	0.706
SCORM	947	0.20	1.15 (0.87-1.51)	0.332
SIMES	1778	0.26	1.35 (1.16-1.58)	1.18 x 10 <sup>-4</sup>
SINDI	2261	0.26	1.13 (0.98-1.30)	0.092
STARS	822	0.20	1.00 (0.77-1.29)	0.999
<b>Meta Analysis</b>	<b>25838</b>	<b>0.22</b>	<b>1.12 (1.08-1.16)</b>	<b>5.55 x 10<sup>-9</sup></b>



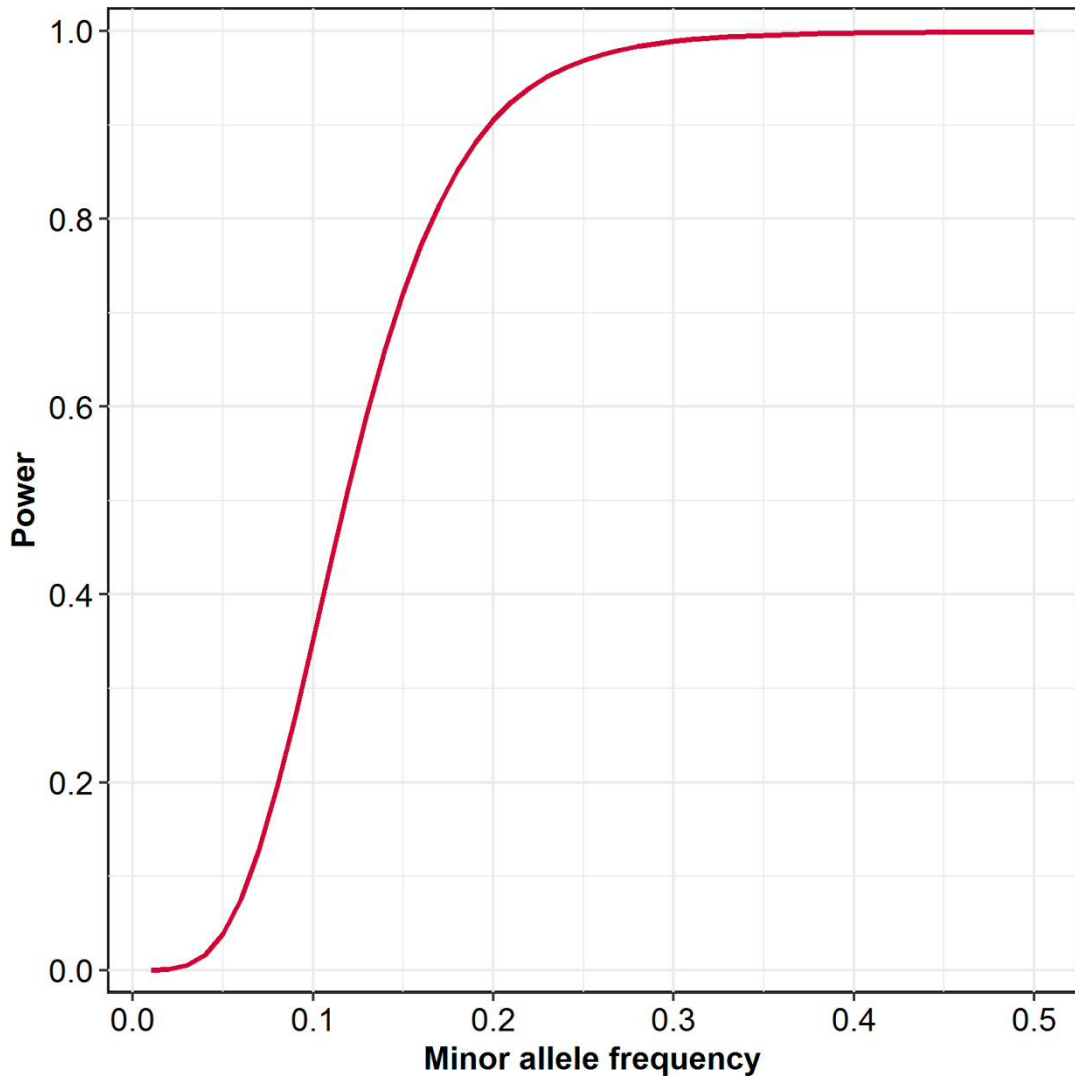
#### 4.3.1.1 Power Calculation

As this meta-analysis of all samples did not identify any novel loci associated with corneal astigmatism, the power to detect associations in this investigation was calculated. Table 4.8 shows that based on the sample size used for this meta-analysis; there was at least 80% power to detect variants with an odds ratio of 1.1 and MAF  $\geq 0.17$  at  $p\text{-value} < 5 \times 10^{-8}$ . The power to detect associations decreased steadily for less common variants with power being only 35% for variants with a MAF of 0.1 (Table 4.8, Figure 4.6).

**Table 4.8: Power ( $\beta$ ) to detect an odds ratio of 1.1 for variants with MAFs ranging between 0.01 and 0.50.** Based on a sample size of  $N = 31,370$ , a case:control ratio of 1:2.3, a trait population prevalence of 43.5% and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency.

<b>MAF</b>	<b>Power (%)</b>	<b>MAF</b>	<b>Power (%)</b>
<b>0.01</b>	0.01	<b>0.26</b>	97.50
<b>0.02</b>	0.10	<b>0.27</b>	97.98
<b>0.03</b>	0.52	<b>0.28</b>	98.36
<b>0.04</b>	1.63	<b>0.29</b>	98.66
<b>0.05</b>	3.89	<b>0.30</b>	98.90
<b>0.06</b>	7.58	<b>0.31</b>	99.09
<b>0.07</b>	12.82	<b>0.32</b>	99.24
<b>0.08</b>	19.43	<b>0.33</b>	99.36
<b>0.09</b>	27.07	<b>0.34</b>	99.46
<b>0.10</b>	35.30	<b>0.35</b>	99.54
<b>0.11</b>	43.66	<b>0.36</b>	99.60
<b>0.12</b>	51.76	<b>0.37</b>	99.65
<b>0.13</b>	59.31	<b>0.38</b>	99.70
<b>0.14</b>	66.13	<b>0.39</b>	99.73
<b>0.15</b>	72.11	<b>0.40</b>	99.76
<b>0.16</b>	77.24	<b>0.41</b>	99.78
<b>0.17</b>	81.57	<b>0.42</b>	99.80
<b>0.18</b>	85.16	<b>0.43</b>	99.82
<b>0.19</b>	88.10	<b>0.44</b>	99.83
<b>0.20</b>	90.49	<b>0.45</b>	99.84
<b>0.21</b>	92.41	<b>0.46</b>	99.85
<b>0.22</b>	93.94	<b>0.47</b>	99.85
<b>0.23</b>	95.16	<b>0.48</b>	99.86
<b>0.24</b>	96.13	<b>0.49</b>	99.86
<b>0.25</b>	96.89	<b>0.50</b>	99.86

**Figure 4.6: Power calculation.** Power to detect variants with an odds ratio of 1.1 for this meta-analysed sample with MAFs ranging between 0.01 and 0.50. Based on a sample size of  $N = 31,370$ , a case:control ratio of 1:2.3, a trait population prevalence of 43.5% and Type I error rate ( $\alpha$ ) of  $5 \times 10^{-8}$ . MAF = minor allele frequency.



#### **4.3.2 Gene-based and Gene-set Analyses**

In order to identify potential candidate genes and biological mechanisms enriched with variants attaining low but not necessarily genome-wide significant p-values from GWAS, gene-based and gene-set tests were performed using VEGAS2 and MAGMA. As these gene-based analyses require ancestry-matched reference panels,

these analyses were conducted using the results from the meta-analysis of all European ancestry cohorts of all ages. No genes demonstrated significant association from the VEGAS2 analyses. The 10 genes demonstrating strongest association from this analysis are shown in Table 4.9. From the MAGMA analysis, 17 genes demonstrated significant association (FDR < 0.05; Table 4.10). Despite the difference in strength of association, there was a high degree of overlap between the results of the two programs, with the genes *ACP2*, *CLDN7*, *ELP5*, and *CTDNEP1* demonstrating the strongest association in both analyses (Table 4.9 and Table 4.10). In the MAGMA gene-based test, these four genes and *TNFAIP8L3* achieved  $P < 0.05$  after stringent Bonferroni correction, whereas this was not the case for VEGAS (Table 4.9 and Table 4.10). Quantile-Quantile plots for the respective VEGAS2 and MAGMA gene-based analyses (Figure 4.7) show results from both methods follow similar distributions at higher  $-\log_{10}$  p-values and no systematic inflation of p-values from the MAGMA gene-based analysis.

A further exploratory gene-based analysis that included markers up to 200 kb upstream or downstream of each gene – an approach that has been successful for certain traits (Brodie et al., 2016) – failed to identify any additional genes associated with corneal astigmatism (at a FDR < 0.05) (Appendix C).

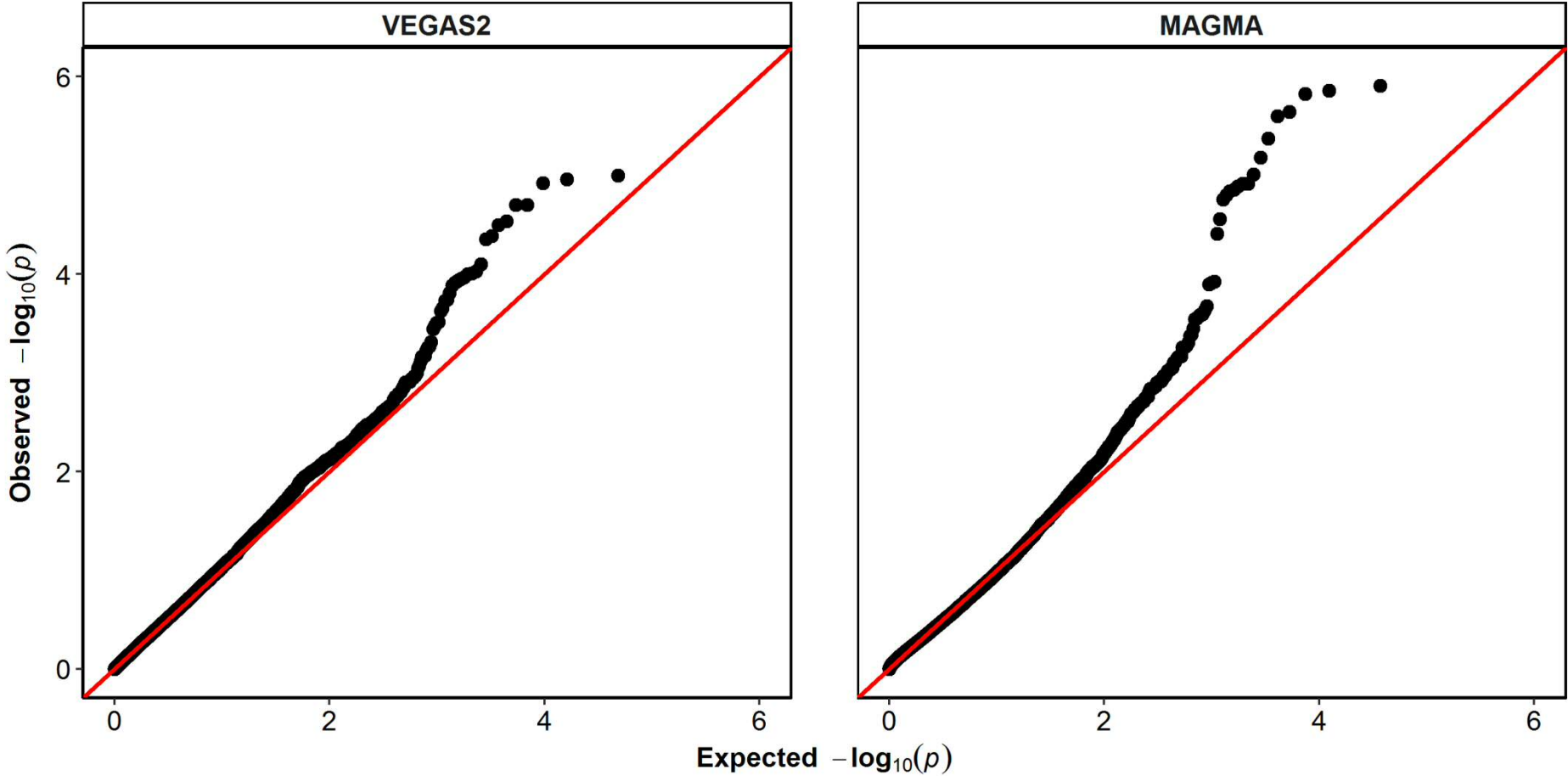
**Table 4.9: The 10 genes demonstrating strongest association from the VEGAS2 gene-based association test.** Start and stop positions listed include  $\pm 50$  kb flanking regions. nSNPs = number of variants included in gene region. Test Statistic = gene-based  $\chi^2$  test statistic with *nSNPs* degrees of freedom. P-value = obtained from *Test Statistic* and adjusting for LD between variants. FDR = false discovery rate. Lead Variant = variant within gene locus with strongest association signal from previous single marker based association test. Genes shown in bold were also identified in the 10 strongest associated genes using MAGMA (Table 4.10). Total number of genes tested = 24,232.

Gene	Chromosome	Start	Stop	nSNPs	Test Statistic	P-value	FDR	Bonferroni Adjusted P-value	Lead Variant
<b>ACP2</b>	11	47210852	47320457	175	1680.63	$1.00 \times 10^{-5}$	0.097	0.242	rs3758670
<i>MADD</i>	11	47240926	47401582	271	2472.89	$1.10 \times 10^{-5}$	0.097	0.267	rs2697920
<i>NR1H3</i>	11	47219850	47340584	191	1929.38	$1.20 \times 10^{-5}$	0.097	0.291	rs3758670
<i>DDB2</i>	11	47186492	47310769	193	1638.98	$2.00 \times 10^{-5}$	0.097	0.485	rs3758670
<b>CLDN7</b>	17	7113221	7216512	123	980.06	$2.00 \times 10^{-5}$	0.097	0.485	rs222836
<b>CTDNEP1</b>	17	7096905	7205259	114	1013.67	$2.90 \times 10^{-5}$	0.111	0.703	rs222836
<b>ELP5</b>	17	7105371	7213259	123	1012.32	$3.20 \times 10^{-5}$	0.111	0.775	rs222836
<b>GABARAP</b>	17	7093737	7195753	96	952.50	$4.10 \times 10^{-5}$	0.118	0.994	rs222836
<i>TYR</i>	11	88861039	89078927	402	3017.77	$4.40 \times 10^{-5}$	0.118	1	rs12808354
<b>PHF23</b>	17	7088346	7192825	104	903.47	$7.90 \times 10^{-5}$	0.174	1	rs222836

**Table 4.10: Genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05).** Start and stop positions listed include  $\pm 50$  kb flanking regions. nSNPs = number of variants included in gene region. Z-Statistic = gene-based test statistic. P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model. FDR = false discovery rate. Genes shown in bold were also identified in the 10 strongest associated genes using VEGAS2 (Table 4.9). Total number of genes tested = 18,418.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<b>ELP5</b>	17	7105372	7213259	123	4.71	$1.23 \times 10^{-6}$	0.009	0.023
<b>CLDN7</b>	17	7113222	7216863	124	4.69	$1.39 \times 10^{-6}$	0.009	0.026
<b>CTDNEP1</b>	17	7096906	7205259	114	4.67	$1.50 \times 10^{-6}$	0.009	0.028
<i>TNFAIP8L3</i>	15	51298798	51447473	212	4.59	$2.26 \times 10^{-6}$	0.009	0.042
<b>ACP2</b>	11	47210853	47320457	175	4.56	$2.52 \times 10^{-6}$	0.009	0.046
<b>GABARAP</b>	17	7093738	7195753	96	4.45	$4.23 \times 10^{-6}$	0.013	0.078
<b>PHF23</b>	17	7088347	7192825	104	4.36	$6.59 \times 10^{-6}$	0.017	0.121
<i>DVL2</i>	17	7078661	7187867	112	4.27	$9.80 \times 10^{-6}$	0.021	0.181
<i>SLC2A4</i>	17	7134986	7241367	155	4.22	$1.21 \times 10^{-5}$	0.021	0.223
<i>ACADVL</i>	17	7070444	7178586	104	4.22	$1.22 \times 10^{-5}$	0.021	0.224
<i>NR1H3</i>	11	47219851	47340584	191	4.21	$1.29 \times 10^{-5}$	0.021	0.238
<i>MADD</i>	11	47240927	47401582	271	4.19	$1.40 \times 10^{-5}$	0.021	0.258
<i>DLG4</i>	17	7043209	7173369	139	4.18	$1.45 \times 10^{-5}$	0.021	0.267
<i>YBX2</i>	17	7141571	7247876	158	4.16	$1.59 \times 10^{-5}$	0.021	0.292
<i>DDB2</i>	11	47186493	47310769	193	4.14	$1.74 \times 10^{-5}$	0.021	0.321
<i>PSMD5</i>	9	123528331	123655299	148	4.03	$2.77 \times 10^{-5}$	0.032	0.510
<i>EIF5A</i>	17	7160318	7265782	179	3.95	$3.89 \times 10^{-5}$	0.042	0.717

**Figure 4.7: Quantile-quantile plots for VEGAS2 and MAGMA gene-based analyses.** Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).



Using these results from the MAGMA gene-based test (with either  $\pm 50$  kb or  $\pm 200$  kb flanking regions), gene-set analysis was performed using MAGMA for the European ancestry cohorts (Appendix C). However, no gene sets were identified as demonstrating a greater level of association with corneal astigmatism than would be expected by chance.

### 4.3.3 SNP-Heritability and Genetic Correlation

LDSC was used to quantify SNP-heritability and the genetic correlation between corneal astigmatism and two related traits, refractive astigmatism and spherical equivalent refractive error (Table 4.11 and Table 4.12). The SNP-heritability ( $h^2_{\text{SNP}}$ ) estimates for corneal and refractive astigmatism ( $h^2_{\text{SNP}} = 0.056$  and  $0.014$ , respectively) were considerably lower than for spherical equivalent ( $h^2_{\text{SNP}} = 0.233$ ). Furthermore, the SNP-heritability estimates for each of these astigmatism traits were not significantly different from zero ( $P > 0.05$ ).

**Table 4.11: SNP-heritability estimated in samples of European ancestry from the CREAM consortium using LDSC.**  $h^2_{\text{SNP}}$  = SNP-heritability; SE = standard error; P-value = test of the null hypothesis ( $h^2_{\text{SNP}} = 0$ ).

Trait	No. of Variants	Sample Size	$h^2_{\text{SNP}}$	SE	P-value
<b>Corneal Astigmatism</b>	1,024,525	22,250	0.056	0.038	0.15
<b>Refractive Astigmatism</b>	1,056,658	31,968	0.014	0.022	0.53
<b>Spherical Equivalent</b>	1,056,658	37,382	0.233	0.018	$2.60 \times 10^{-40}$

The genetic correlation analysis suggested there was a high genetic correlation between spherical equivalent and refractive astigmatism but weak correlation between spherical equivalent and corneal astigmatism (Table 4.12); however due to the high standard errors between all pairs of traits, these estimates were very



imprecise. The lack of precision in genetic correlation estimates from LDSC may be due to the low (and not significantly different from zero) SNP-heritability estimates for the respective astigmatism traits. Therefore, considerably larger sample sizes and/or a more homogenous sample may be required to improve precision when estimating SNP-heritability and genetic correlation.

**Table 4.12: Genetic correlations between pairs of refractive error traits in samples of European ancestry from the CREAM consortium using LDSC.** RA = refractive astigmatism; CA = corneal astigmatism; MSE = mean spherical equivalent. P-values refer to the null hypothesis of zero correlation between traits.

<b>Trait Pairs</b>	<b>No. of Variants</b>	<b>Genetic Correlation (SE)</b>	<b>P-value</b>
<b>RA and CA</b>	934,512	0.233 (0.70)	0.741
<b>MSE and CA</b>	1,024,525	-0.024 (0.16)	0.882
<b>RA and MSE</b>	1,056,658	0.773 (0.65)	0.235

#### 4.4 Discussion

This meta-analysis of GWAS for corneal astigmatism in a combined sample of Europeans and Asians identified a single genome-wide significant locus in the promoter region of the *PDGFRA* gene, thus replicating the previous discovery of this corneal astigmatism associated locus in a predominantly Asian sample by Fan et al. (2011). Despite a fourfold increase in sample size ( $N = 31,370$  versus  $N = 8,513$ ) compared to the only previous GWAS meta-analysis for corneal astigmatism (Fan et al., 2011), the standard, single marker analysis performed here did not identify any novel loci. GWAS for spherical equivalent and other morphological traits in equivalently sized samples have identified dozens of independent risk loci (Kiefer et al., 2013; Verhoeven et al., 2013), yet this paucity of genome-wide significant loci for corneal astigmatism mirrors that observed in a previous large-scale GWAS for refractive astigmatism conducted by Li et al. (2015a).

SNP-heritability ( $h^2$ ) estimates for corneal and refractive astigmatism from LD Score regression were 0.056 and 0.014 respectively. These are the first reported estimates for these respective traits and are much lower than those obtained for spherical equivalent ( $h^2 = 0.233$ ), therefore suggesting that additive genetic effects of commonly occurring variants make a relatively small contribution to the variation of astigmatism between individuals. It should be noted that these SNP-heritability estimates are considerably lower than those reported from family and twin studies since SNP-heritability refers to the contribution of the variants tested rather than all sources of genetic variation as is the case in family and twin studies of heritability.

In the study by Fan et al. (2011) that originally identified the association between variants close to the *PDGFRA* gene and corneal astigmatism, the authors speculated that the underlying causal mechanism was common to populations of diverse ancestry and not specifically to those of Asian ancestry. This was based on the knowledge that their investigation included individuals of Indian ancestry, who are more closely genetically related to Europeans than East Asians (Fan et al., 2011). The results from the investigation conducted here support this theory.

Association of genetic variants in the *PDGFRA* gene region with corneal astigmatism have been successfully replicated in a European ancestry cohort by Guggenheim et al. (2013a) (N = 1968); however there was a failure to replicate in a smaller study of Australians of European ancestry (N = 1013; Yazar et al. (2013b)). Guggenheim et al. (2013a) also identified this locus to be associated with axial length and corneal curvature in their European ancestry cohort. Furthermore, the association of variants in the *PDGFRA* region with corneal curvature has been identified in additional European and Asian ancestry cohorts (Han et al., 2011; Mishra et al., 2012); however, a large-scale meta-analysis of GWAS for refractive astigmatism failed to demonstrate genome-wide significant association with this locus (N = 45,287; Li et al. (2015a)). The investigation undertaken here adds to this complex body of evidence in support of genetic variants in the vicinity of the *PDGFRA* gene influencing eye size and corneal astigmatism, not just in Asian ancestry populations but also in European ancestry populations, however, the underlying mechanism of action remains uncertain.

In contrast to the single marker analyses, gene-based analyses have provided new insight into the genetic basis of corneal astigmatism, implicating the genes *ACP2*, *CLDN7*, *CTDNEP1*, *ELP5*, and *TNFAIP8L3*. Three of these five genes: *CLDN7*, *CTDNEP1*, and *ELP5*, are tightly clustered on chromosome 17, with their respective gene-based association signals sharing many variants in common. Therefore, a parsimonious interpretation is that only one of the genes has a causal association with astigmatism with the other two genes likely false-positive associations identified due to the contribution of overlapping variants that influence the causal gene. Of these three genes, *CLDN7*, which encodes the claudin-7 membrane protein (Hewitt, Agarwal and Morin, 2006), appears to be the most biologically plausible candidate. Claudins are responsible for tight junction formation and function (Tsukita and Furuse, 2000), with claudin-7 being the subtype present in human corneal epithelium and endothelium (Inagaki et al., 2013). At present, the mechanism by which claudin-7 may contribute to the development of corneal astigmatism is unclear. The acid phosphatase 2, lysosomal gene (*ACP2*), located on chromosome 11, codes for the beta subunit of the degradative enzyme, lysosomal acid phosphatase (LAP). Activity of the LAP enzyme has been identified to be enhanced in keratoconic corneas (Sawaguchi et al., 1989; Maruyama et al., 2001). The *TNFAIP8L3* gene on chromosome 15 codes for TNF alpha-induced protein 8 like 3, which is preferentially expressed in secretory epithelial cells (Cui et al., 2015). *TNFAIP8L3* is implicated as a negative regulator of inflammation (and carcinogenesis) through its role in TNF $\alpha$  and phospholipid signalling. Based on this evidence, the *CLDN7*, *ACP2*, and *TNFAIP8L3* genes are promising susceptibility genes for corneal astigmatism. It is important to note that whilst statistical support for the

above three genes was much stronger in the MAGMA analysis than in the VEGAS2 analysis, the two software applications similarly ranked the genes demonstrating strongest association. This commonality between the MAGMA and VEGAS2 gene-based test results provides greater confidence that these findings are robust than would be the case for findings identified using either software program alone, as the statistical models and hypothesis tests used by the two programs differ (see *Section 2.3.5* for how these respective tests work). In addition, the increased number of genes tested in the VEGAS2 analysis compared to that in the MAGMA analysis means that there is a greater multiple testing penalty applied to the VEGAS2 gene-based test results.

The strengths of this investigation are that this meta-analysis included data from multiple population samples, with follow-up gene-based and gene-set analyses undertaken in order to identify novel biological insights into the genetics of astigmatism. Weaknesses were, firstly, that despite the inclusion of both European and Asian ancestry cohorts, trans-ethnic meta-analysis (Wang et al., 2013) was not performed due to the small size of the Asian ancestry sample compared to the European ancestry sample. Trans-ethnic meta-analysis takes into consideration heterogeneity in effect sizes between cohorts; however this approach also factors in correlated effect sizes within ancestry groups such that variants demonstrating association in one ancestry group but not in another are still identifiable as true associations when meta-analysed (Wang et al., 2013). A second weakness of this investigation was the broad age spectrum of the participants (cohort mean age range = 6 to 73 years). This latter point is an important consideration since

astigmatism (both corneal and refractive) does not remain constant during life, with changes in orientation occurring with age (Read et al., 2007). For example, in childhood, astigmatism tends to be with-the-rule (WTR), whereas in older adults this orientation typically changes to against-the-rule (ATR) (Sanfilippo et al., 2015; Shao et al., 2017). This change in orientation is more notable for corneal astigmatism than for refractive astigmatism. Furthermore, the magnitude of corneal astigmatism has a tendency to be greater in the youngest (under 10 years-old) and oldest age groups (over 80 years-old) but it generally stable for most of life unlike refractive astigmatism which tends to increase from middle-age (over 50 years-old) (Sanfilippo et al., 2015). To overcome much of this variation, the design of this study was such that only the magnitude of corneal astigmatism was considered (i.e. no consideration of astigmatism axis) and utilisation of a case-control classification, thus reducing the impact of subtle changes in astigmatism that commonly occurs with age.

The effect of spherical refractive error was not considered in this analysis despite it being shown previously that increasing magnitudes of astigmatism are correlated with increasing magnitudes of spherical refractive error (Kronfeld and Devney, 1930; Guggenheim and Farbrother, 2004). However, the potential common role of refractive error associated genetic variants with astigmatism was unclear at the time of planning, but has since been suggested as a possibility from the results of the refractive astigmatism meta-analysis conducted by the CREAM consortium (Li et al., 2015a). Furthermore, inclusion of spherical equivalent (or the spherical refractive component only) as an additional covariate would reduce the statistical

power to detect corneal astigmatism associated variants, and therefore, a parsimonious model was favoured for this study.

In conclusion, this GWAS meta-analysis for corneal astigmatism has replicated the discovery of a genome-wide significant locus near the *PDGFRA* gene (Fan et al., 2011) and provided strong evidence that this locus is relevant in both Asians and Europeans (Table 4.7, Figure 4.5). Three novel candidate genes, *CLDN7*, *ACP2*, and *TNFAIP8L3*, have been identified using gene-based analyses, utilising data from across genomic regions rather than examining one genetic variant at a time. These novel genes warrant further investigation in order to understand their potential role in the pathogenesis of corneal astigmatism. Lastly, utilising the recently developed LD Score regression method to estimate SNP-heritability, it has been revealed that SNP-heritability of corneal (and refractive astigmatism) is considerably lower than that for spherical equivalent refractive error (Table 4.11). This implies that rare variants and/or non-genetic (environmental) factors have a greater influence on astigmatism than spherical equivalent, or that non-additive genetic effects make a substantial contribution to the heritability of astigmatism as suggested in previous studies (Hammond et al., 2001; Dirani et al., 2006; Grjibovski et al., 2006).

## **Chapter 5 Estimating SNP-heritability of Refractive Errors**



## Chapter 5 Estimating SNP-heritability of Refractive Errors

### 5.1 Introduction

In Chapter 4, SNP-heritability estimates for corneal and refractive astigmatism estimated using summary statistics from the respective CREAM GWAS, were found to be low:  $h^2_{\text{SNP}} = 0.056$  and  $0.014$ , respectively. The subsequent availability of genetic data from the UK Biobank enabled an opportunity to further explore and replicate these SNP-heritability analyses in this new sample.

In addition, the investigation performed here aimed to test the hypothesis that SNP-heritability for spherical equivalent refractive error is mostly due to additive effects whereas dominance effects are the main contributors to the SNP-heritability of corneal and refractive astigmatism. The hypotheses were tested using the software application GCTA, in a sample of individuals of White European ancestry with genotype data available from the UK Biobank interim release ( $N = 141,751$ ). In addition, the investigation aimed to provide “optimal classification criteria” for defining cases of corneal and refractive astigmatism based on these heritability estimates, which could then be applied in subsequent GWAS of these respective traits (*Chapter 6*).

## 5.2 Methods

Analyses were restricted to the individuals of White European ancestry with genetic data available from the UK Biobank interim data release as outlined in *Section 2.2.2*. Individuals were excluded from analyses if they had, or reported having, any eye problems or disorders that could bias or invalidate autorefractometry or keratometry (see *Section 2.2.3* for further details). Individuals missing covariate data were also excluded.

### 5.2.1 SNP-Heritability Estimation

SNP-heritability ( $h^2_{\text{SNP}}$ ) estimates were obtained for all three phenotypes (corneal astigmatism, refractive astigmatism and spherical equivalent;  $N = 27,737$  for corneal astigmatism, and 28,403 for refractive astigmatism and spherical equivalent) using the list of approximately 890,000 “high-confidence” variants (defined in *Section 2.2.5*) and the GCTA software application (Yang et al., 2011a). Phenotypes were considered as continuous traits or as dichotomous traits defined using each of the following case thresholds: 0.50 D to 1.50 D, in 0.25 D steps for corneal and refractive astigmatism; and -0.50 D to -1.50 D, in 0.25 D steps for spherical equivalent. Sample prevalences for the respective traits and case thresholds are listed in Table 5.3. Discrete covariates included in SNP-heritability estimation were: genotyping array and sex. The mean age of the participant when eye measures were recorded was included as a quantitative covariate. Mean spherical equivalent was included as an additional quantitative covariate for all analyses of corneal and refractive astigmatism (as outlined in *Section 2.2.4*).

### **5.2.2 Observed to Liability Scale Conversion**

As heritability is estimated on the ‘observed scale’ (see *Section 2.3.6* for definitions), conversion to the ‘liability scale’ was required for the heritability estimates of the different dichotomous traits to be compared either against each other or across different case threshold definitions (Lee et al., 2011). To facilitate this, approximate population prevalence estimates were obtained from the full sample of UK Biobank individuals with valid phenotype measures available, irrespective of ancestry or exclusion criteria. Sample prevalences, defined as the proportion of cases in the investigated sample relative to the total investigated sample size, were calculated by GCTA during analysis. As both sample and population prevalences were included, this transformation automatically took into consideration the presence of any ascertainment bias, although this would be negligible since the population prevalence was obtained from the full UK Biobank dataset.

### **5.2.3 Consideration of Uncorrected Population Effects**

To gauge the extent to which SNP-heritability estimates obtained here may be biased (inflated) due to uncorrected population stratification or residual cryptic relatedness, the method proposed by Yang et al. (2011b) was utilised (see *Section 2.3.6*). To facilitate this, individual GRMs were generated using PLINK 1.9 for each of the 22 autosomal chromosomes using individuals who were included in both GRMs for corneal astigmatism and the refractive phenotypes (N = 25,900). For this analysis, all traits were considered in their continuous form only.

#### **5.2.4 Quantification of Dominance Effects**

Additional SNP-heritability estimates were obtained in an attempt to partition SNP-heritability into separate additive and dominance components. GCTA-GREMLd (Zhu et al., 2015) was used to generate additive and dominance GRMs for corneal astigmatism and the refractive phenotypes using the same unrelated individuals as used for the initial SNP-heritability estimation (N = 27,737 and 28,403 respectively). For this analysis, all traits were considered in their continuous form only.

## 5.3 Results

### 5.3.1 Consideration of Uncorrected Population Effects

In order to test for inflation of SNP-heritability estimates due to uncorrected population stratification or cryptic relatedness, the following linear regression was performed: the difference between heritability estimates from individual analyses of each chromosome versus a joint analysis of all chromosomes ( $h^2_{\text{sep}} - h^2_{\text{joint}}$ ), was regressed on chromosome length (Table 5.1, Figure 5.1). In the case of no cryptic relatedness or population stratification being present, the intercept and gradient would each equal zero. Conversely, in the presence of uncorrected cryptic relatedness or population stratification, the intercept or gradient of the regression line would be greater than zero, respectively. From the joint analyses, SNP-heritability estimates obtained using the “high-confidence” variants (see Section 2.2.5) and individuals present in both initial GRMs for corneal astigmatism and the refractive phenotypes (N = 25,880) were  $h^2_{\text{SNP}} = 0.074$ , 0.049 and 0.391 for corneal astigmatism, refractive astigmatism and spherical equivalent respectively. As listed in Table 5.1, the respective intercepts and gradients were  $2.25 \times 10^{-4}$  ( $P = 0.377$ ) and  $2.92 \times 10^{-6}$  ( $P = 0.109$ ) for corneal astigmatism;  $1.67 \times 10^{-4}$  ( $P = 0.647$ ) and  $4.60 \times 10^{-6}$  ( $P = 0.082$ ) for refractive astigmatism; and  $1.86 \times 10^{-3}$  ( $P = 0.009$ ) and  $8.87 \times 10^{-6}$  ( $P = 0.062$ ) for spherical equivalent. By applying Equation 2.12, the total proportions of phenotypic variance attributable to population structure (cryptic relatedness + population stratification) were  $6.36 \times 10^{-4}$  ( $2.35 \times 10^{-4} + 4.01 \times 10^{-4}$ ),  $8.06 \times 10^{-4}$  ( $1.75 \times 10^{-4} + 6.31 \times 10^{-4}$ ) and  $3.17 \times 10^{-3}$  ( $1.95 \times 10^{-3} + 1.22 \times 10^{-3}$ ) for corneal astigmatism, refractive astigmatism and spherical equivalent, respectively (Table 5.1). Despite the non-zero intercept value for spherical equivalent ( $P < 0.05$ ), its

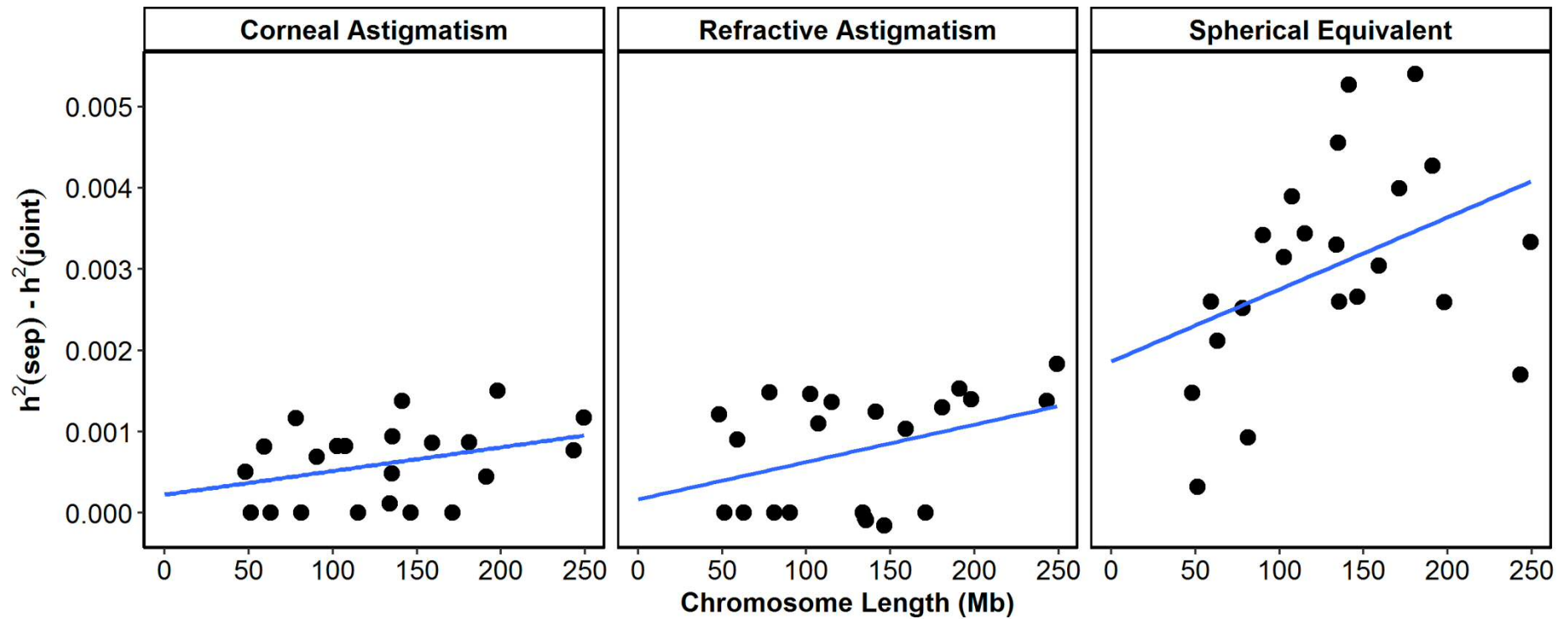
effect on the overall trait variance can be considered as negligible as variation due to population effects is less than 1% of the SNP-heritability estimated for this trait.

In summary, the SNP-heritability estimates for corneal astigmatism, refractive astigmatism and spherical equivalent obtained here do not appear to be notably inflated due to unadjusted population structure. Thus, inclusion of principal components for population stratification adjustment was not deemed necessary for these traits in this sample.

### **5.3.2 Quantification of Dominance Effects**

Since twin and family studies have suggested the presence of dominance effects contributing to the variance of astigmatism (Hammond et al., 2001; Dirani et al., 2006; Grijibovski et al., 2006), GREMLd analyses were performed to partition SNP-heritability estimates into separate additive and dominance components. For all three traits, estimates for additive effects were in line with the  $h^2_{\text{joint}}$  estimates obtained previously; however, the contribution of dominance effects appeared to be negligible (Table 5.2).

**Figure 5.1: Estimating the proportion of  $h^2_{\text{SNP}}$  that is inflated due to uncorrected cryptic relatedness and population stratification.**  $h^2_{\text{sep}}$  = SNP-heritability estimated from the individual chromosome tests;  $h^2_{\text{joint}}$  = Total SNP-heritability estimated from the joint test (all chromosomes pooled together). Blue lines denote linear regression lines.



**Table 5.1: Estimating the proportion of  $h^2_{\text{SNP}}$  that is inflated due to uncorrected cryptic relatedness and population stratification.**

Abbreviations:  $h^2_{\text{joint}}$  = Total SNP-heritability estimated from the joint test; P-value = probability of  $h^2 = 0$ ;  $V_{\text{POP}}$  = proportion of phenotypic variance attributed to population structure (cryptic relatedness and population stratification);  $V_{\text{CR}}$  = proportion of phenotypic variance attributed to cryptic relatedness only;  $V_{\text{PS}}$  = proportion of phenotypic variance attributed to population stratification only. N = 25,880.

Trait	$h^2_{\text{joint}}$	95% CI	Intercept	P-value	Gradient	P-value	$V_{\text{POP}} (V_{\text{CR}} + V_{\text{PS}})$
<b>Corneal Astigmatism</b>	0.074	0.030-0.119	$2.25 \times 10^{-4}$	0.377	$2.92 \times 10^{-6}$	0.109	<b><math>6.36 \times 10^{-4}</math></b> ( $2.35 \times 10^{-4} + 4.01 \times 10^{-4}$ )
<b>Refractive Astigmatism</b>	0.049	0.006-0.093	$1.67 \times 10^{-4}$	0.647	$4.60 \times 10^{-6}$	0.082	<b><math>8.06 \times 10^{-4}</math></b> ( $1.75 \times 10^{-4} + 6.31 \times 10^{-4}$ )
<b>Spherical Equivalent</b>	0.391	0.344-0.437	$1.86 \times 10^{-3}$	0.009	$8.87 \times 10^{-6}$	0.062	<b><math>3.17 \times 10^{-3}</math></b> ( $1.95 \times 10^{-3} + 1.22 \times 10^{-3}$ )

**Table 5.2: Partitioning trait variance into additive, dominance and environmental components using GREMLd.**

Trait	Sample Size	Additive	95% CI	Dominance	95% CI	Environment
<b>Corneal Astigmatism</b>	27,714	0.056	0.015-0.096	0.000	0.000-0.036	0.944
<b>Refractive Astigmatism</b>	28,378	0.046	0.007-0.085	0.000	0.000-0.035	0.954
<b>Spherical Equivalent</b>	28,378	0.368	0.326-0.411	0.000	0.000-0.033	0.632



### **5.3.3 Determining Optimal Trait Definitions**

Refractive errors have often been considered as dichotomous traits in genetic and epidemiological studies; however, the choice of the threshold used to define case/control status has varied from study to study (Hammond et al., 2001; He et al., 2004; Quek et al., 2004; Huynh et al., 2007; Vitale et al., 2008; Dirani et al., 2010; Fan et al., 2011; Li et al., 2015a). In order to determine an optimal trait definition for detecting commonly-occurring genetic variants with additive effects on refractive errors, SNP-heritability estimates were calculated with GCTA for corneal astigmatism, refractive astigmatism and spherical equivalent classified either as continuous or dichotomous traits, and using a grid of case thresholds for the latter (namely, 0.50, 0.75, 1.00, 1.25 and 1.50 D of astigmatism or -0.50, -0.75, -1.00, -1.25 and -1.50 D of spherical equivalent). Following previous precedents (Schulze and McMahon, 2004; Corvin, Craddock and Sullivan, 2010; Koran et al., 2014), this approach was predicated on the assumption that the trait definition capturing the greatest SNP-heritability would be the one most likely to highlight genome-wide significant loci in subsequent GWAS of these traits.

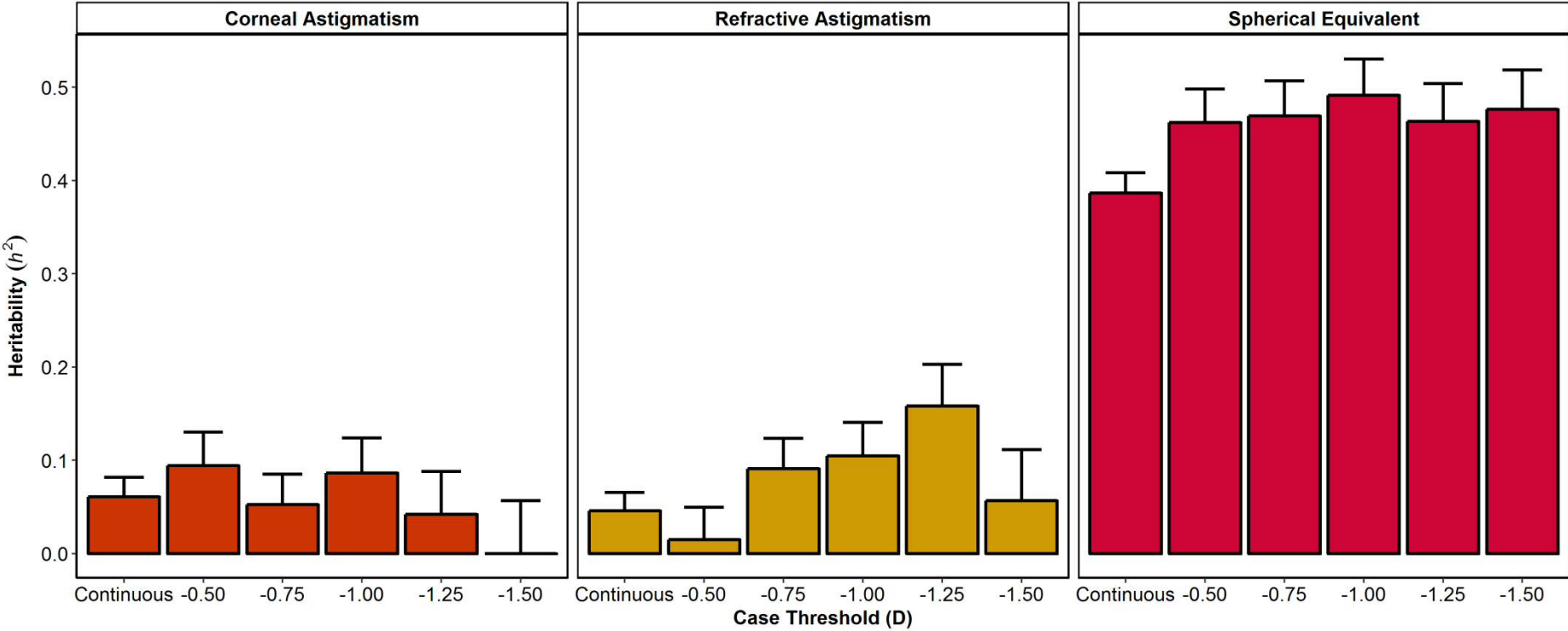
For corneal astigmatism, SNP-heritability was greatest using a case definition threshold of 0.50 D ( $h^2_{\text{SNP}} = 0.094$ ) and negligible for a case threshold of 1.50 D (Table 5.3). However, there appeared to be no meaningful difference in SNP-heritability across the range of trait definitions tested, since all of the confidence intervals overlapped (Table 5.3, Figure 5.2). For refractive astigmatism, SNP-heritability estimates were generally higher than those for corneal astigmatism ( $h^2_{\text{SNP}}$ : range 0.015-0.158; Table 5.3, Figure 5.2). Using case thresholds of increasing

magnitude between 0.50 D and 1.25 D inclusive yielded increasing SNP-heritability estimates, although the large standard errors meant that, again, there was no statistical support for meaningful differences across the range of case thresholds tested. SNP-heritability estimates for spherical equivalent were similar for all case thresholds of examined ( $h^2_{\text{SNP}}$ : range 0.462-0.491; Table 5.3, Figure 5.2) with considerable overlap between estimates. Estimates of SNP-heritability were numerically lower but with much narrower standard errors when astigmatism and spherical equivalent were modelled as continuous traits compared to dichotomous trait analyses (continuous trait  $h^2_{\text{SNP}}$  (SE): corneal astigmatism = 0.061 (0.021); refractive astigmatism = 0.046 (0.020); spherical equivalent = 0.387 (0.022)), although once more, not to a sufficient extent to attain statistical support (Table 5.3, Figure 5.2).

**Table 5.3: Estimates of SNP-heritability ( $h^2_{\text{SNP}}$ ) for corneal astigmatism, refractive astigmatism and spherical equivalent.** Analyses were conducted using GCTA-GREML.  $h^2_{\text{SNP}}$  = SNP-heritability; SE = standard error; P-value = test of the null hypothesis ( $h^2_{\text{SNP}} = 0$ ).

Trait	Case Threshold (D)	Case Prevalence		$h^2_{\text{SNP}}$	SE	P-value
		Population	Sample			
<b>Corneal Astigmatism</b> (N = 27,707)	Continuous	-	-	0.061	0.021	$1.19 \times 10^{-3}$
	0.50	0.72	0.70	0.094	0.036	$3.53 \times 10^{-3}$
	0.75	0.46	0.45	0.053	0.033	0.051
	1.00	0.28	0.27	0.086	0.038	$9.86 \times 10^{-3}$
	1.25	0.17	0.17	0.042	0.046	0.175
	1.50	0.11	0.10	0.000	0.057	0.500
<b>Refractive Astigmatism</b> (N = 28,378)	Continuous	-	-	0.046	0.020	$7.74 \times 10^{-3}$
	0.50	0.73	0.71	0.015	0.035	0.332
	0.75	0.47	0.45	0.091	0.032	$2.00 \times 10^{-3}$
	1.00	0.30	0.28	0.105	0.036	$1.45 \times 10^{-3}$
	1.25	0.19	0.18	0.158	0.045	$1.24 \times 10^{-4}$
	1.50	0.12	0.11	0.057	0.055	0.143
<b>Spherical Equivalent</b> (N = 28,378)	Continuous	-	-	0.387	0.022	$5.05 \times 10^{-85}$
	-0.50	0.35	0.33	0.462	0.036	$2.61 \times 10^{-43}$
	-0.75	0.30	0.29	0.469	0.038	$1.74 \times 10^{-41}$
	-1.00	0.27	0.27	0.491	0.039	$7.04 \times 10^{-42}$
	-1.25	0.25	0.24	0.463	0.040	$1.30 \times 10^{-34}$
	-1.50	0.23	0.22	0.476	0.042	$2.20 \times 10^{-33}$

Figure 5.2: Estimates of SNP-heritability ( $h^2_{\text{SNP}}$ ) using GCTA. Error bars represent the standard error of the  $h^2_{\text{SNP}}$  estimate.



## 5.4 Discussion

This investigation of SNP-heritability in a large European ancestry cohort demonstrated that corneal and refractive astigmatism have low SNP-heritabilities ( $h^2_{\text{SNP}} = 0.06$  and  $0.05$  respectively), while spherical equivalent demonstrated a moderate SNP-heritability ( $h^2_{\text{SNP}} = 0.39$ ). For all three traits, the SNP-heritability models best supported additive effects predominantly contributing to trait variation with a negligible contribution from dominance effects.

The lack of statistical support in favour of a particular threshold for defining astigmatism cases and controls (or myopic vs. non-myopic individuals) (Table 5.3) meant that the selection of an optimal threshold for subsequent GWAS analyses had to be made arbitrarily. Hence, a threshold value of 1.00 D of astigmatism was adopted to define case status, since this value has been widely used in the literature (Huynh et al., 2007). The smaller standard errors obtained for analyses of astigmatism and spherical equivalent when analysed as continuous traits suggest that this coding scheme, i.e. continuous traits rather than dichotomous traits, should be used as the primary outcome for GWAS to be conducted using the full genetic data release from the UK Biobank.

With regard to spherical equivalent, the estimate of SNP-heritability obtained here ( $h^2_{\text{SNP}} = 0.387$ ;  $P < 1 \times 10^{-10}$ ) was slightly greater than a previously published estimate, which suggested a SNP-heritability of 0.35 (Guggenheim et al., 2015). It is important to note that the sample used to generate this previous estimate consisted of children aged 7-15 years-old whereas the current investigation utilised

a much older sample (40-69 years old) and heritability estimates are sensitive to population demographics and changes in environmental exposures, such as age and amount of time spent outdoors (Visscher et al., 2008). More recently, the CREAM consortium estimated SNP-heritability to be between 0.17 and 0.21 in their European ancestry cohorts, but much lower at  $\sim 0.05$  in their Asian ancestry sample (Tedja et al., 2018). The samples included in the investigation by Tedja et al. (2018) were all aged 25 years and above but were obtained from study groups in various locations globally, thus resulting in reduced homogeneity in their sample compared to the UK Biobank White British sample. To date, no additional estimates of SNP-heritability for spherical equivalent / myopia or astigmatism have been published, apart from conference abstracts (Miyake et al., 2013; Hysi et al., 2014).

In general, SNP-heritability estimates for spherical equivalent have been reported to be approximately a third to half of heritability estimates obtained from twin studies (Hammond et al., 2001; Dirani et al., 2006); however, SNP-heritability estimates for corneal and refractive astigmatism from this study are approximately a tenth of heritability estimates reported from twin studies of these respective traits (Hammond et al., 2001; Dirani et al., 2006; Grjibovski et al., 2006). This large difference in heritability estimate between the SNP-based study conducted here and prior twin studies for the astigmatism traits may be due to the differences in the variants investigated by these two methods. Heritability in twin studies captures the effects of all variants, irrespective of their allele frequency and effect size, whereas rare variants are excluded from SNP-heritability estimation (Yang et al., 2017). Therefore, it is likely that genetic contribution to variance of astigmatism is

predominantly due to rare variants (MAF < 0.01), gene-gene or gene-environment interaction effects, or shared environment effects.

Previously, twin studies have suggested the heritability of corneal and refractive astigmatism to be predominantly due to dominance effects (Hammond et al., 2001; Dirani et al., 2006; Grijbovski et al., 2006). However, the investigation conducted here failed to reveal the presence of a non-zero dominance effects component for the SNP-heritability of either trait. Zhu et al. (2015) and Nolte et al. (2017) obtained similar results across a wide range of traits when using GCTA-GREMLd, i.e. the dominance effects component of SNP-heritability was negligible despite prior evidence for the presence of a dominance effects component in twin studies of these traits. Using an alternative method, Zaitlen et al. (2013) also suggested dominance effects made only a small contribution to heritability estimates across several traits, however, the exact contribution could not be fully quantified. Considering these dominance effects results for other complex traits, the causes behind them and those from the investigation here can be rationalised in several ways. One of the simpler arguments is that the sample sizes used in these more recent studies were insufficient to deliver the necessary level of statistical power to detect a non-zero dominance effects component (Zhu et al., 2015); however, the sample sizes used in the analyses here were more than double those of the other studies using these methods, thus boosting the ability to detect a non-zero dominance effects component of SNP-heritability. Another possible explanation for the lack of a dominance effects component relates to the model used to calculate the SNP-heritability estimates. As suggested by Huang and Mackay (2016), the

partitioning of trait variance into its separate components is dependent on the model(s) applied, since effects of one variance component can also contribute to another component, for example some dominance effects can be misallocated as additive effects.

The models applied when using GCTA-GREMLd utilise the classical interpretations of additive and dominance effects (Huang and Mackay, 2016); however, use of alternative, albeit arbitrarily defined definitions, could be utilised as suggested by Huang and Mackay (2016). With these alternative definitions, the presence of a statistically supported non-zero dominance effects component to SNP-heritability may become evident for the traits examined here.

The first part of this investigation interrogated the extent to which uncorrected population effects inflated SNP-heritability estimates. Although the results appear reassuring at first glance, when interpreting these results more generally it is important to note the potential limitations and the assumptions that were made. Firstly, it must be remembered that the SNP-heritability estimates presented here refer to the contribution of a subset of genetic variants to phenotypic variance, and thus would be likely to be lower than twin / family study estimates whereby all genetic contributors are considered. Furthermore, should an alternative selection of variants be used, such as using only directly genotyped variants, there is potential for SNP-heritability estimates to differ. Secondly, the population prevalences used when converting from the observed to the liability scale were obtained from the full UK Biobank sample. This is not the ideal approach, as population prevalences



should be obtained from an independent, ancestry matched source; however no single study has reported population-based prevalence values for each of the thresholds tested here. Vitale et al. (2008) reported 31% prevalence for refractive astigmatism  $\geq 1.00$  D in non-Hispanic white ancestry individuals aged  $\geq 40$  years. However, when that sample was split by age into those aged 40-59 years and  $\geq 60$  years, there was a considerable difference in refractive astigmatism prevalence (28% vs. 50%). In a meta-analysis of European population cohorts, Williams et al. (2015b) reported a lower prevalence of refractive astigmatism  $\geq 1.00$  D (mean age-standardised prevalence 24%); however, there was a notable increase in prevalence with age throughout their age ranges (median age 62 years). Collier Wakefield, Annoh and Nanavaty (2016) investigated corneal astigmatism in pre-operative cataract patients in the UK. Here, higher prevalences of corneal astigmatism were reported for all thresholds, however their sample was of an older age range (mean age 72 years) than other published studies and that of the UK Biobank sample (UK Biobank median age 58 years). These previously published prevalence estimates for astigmatism suggest that using population prevalences based on the full UK Biobank cohort can be justified since these prevalences are not markedly different after considering the increasing magnitude of astigmatism with age.

With respect to heritability, from the evidence presented both here and in previous publications, it appears that the partitioning of variance into separate genetic components lacks sufficient accuracy to be meaningful and any inferences based on heritability should ideally be confined to the total genetic contribution rather than the partitioned components. As SNP-heritability estimation methods have stated

previously, using an additive effects model (as is most commonly done), captures almost all of the total genetic contribution of the considered variants to trait variation and should be deemed sufficient for this approach (Polderman et al., 2015; Huang and Mackay, 2016).

Overall, this investigation of SNP-heritability in a large European ancestry cohort demonstrated that both corneal and refractive astigmatism have a low SNP-heritability, whereas spherical equivalent has a moderate SNP-heritability. Furthermore, for all of the traits examined here, the SNP-heritability appears to be of an additive nature with a negligible contribution from dominance effects.

## **Chapter 6 UK Biobank GWAS for Corneal and Refractive Astigmatism**

## Chapter 6 UK Biobank GWAS for Corneal and Refractive Astigmatism

### 6.1 Introduction

Considering the paucity of loci demonstrating genome-wide significant association with either corneal or refractive astigmatism in previous GWAS for these respective traits, it was earlier speculated whether this may be overcome using a larger and/or more homogenous sample (*Section 4.3.3*).

The availability of the full release of genotype data for the UK Biobank cohort (N = 488,377), approximately 23% of whom had data on corneal and/or refractive astigmatism from non-cycloplegic autorefraction, has provided an opportunity to identify genetic variants associated with corneal or refractive astigmatism using a comprehensive approach and at a larger scale than had been possible previously.

## 6.2 Methods

### 6.2.1 UK Biobank Sample

Analyses were restricted to the individuals of White British ancestry with genetic data available from the full data release as outlined in *Section 2.2.2*. Individuals were excluded from analyses if they had, or reported having, any eye problems or disorders that could impair autorefractometry or keratometry (see *Section 2.2.3* for further details), any missing covariate data, or they had withdrawn consent from participating in the UK Biobank study (Table 6.1).

**Table 6.1: Summary of exclusions from the UK Biobank Sample for GWAS of Corneal and Refractive Astigmatism.** MSE = mean spherical equivalent.

Total No. of Participants	502,633	
	<b>Corneal Astigmatism</b>	<b>Refractive Astigmatism</b>
Genotype data available (full release) <sup>a</sup>	409,728	409,728
Phenotype data available <sup>b</sup>	113,189	116,046
Withdrawn consent	14	14
Individuals with both genotype and phenotype data available	86,422	88,072
Excluded for missing sex covariate data	63	67
Excluded for missing MSE covariate data	24	0
<b>Final Sample Size</b>	<b>86,335</b>	<b>88,005</b>

<sup>a</sup> After restricting to self-reported White British ancestry and non-outlier for heterozygosity (see *Section 2.2.2*)

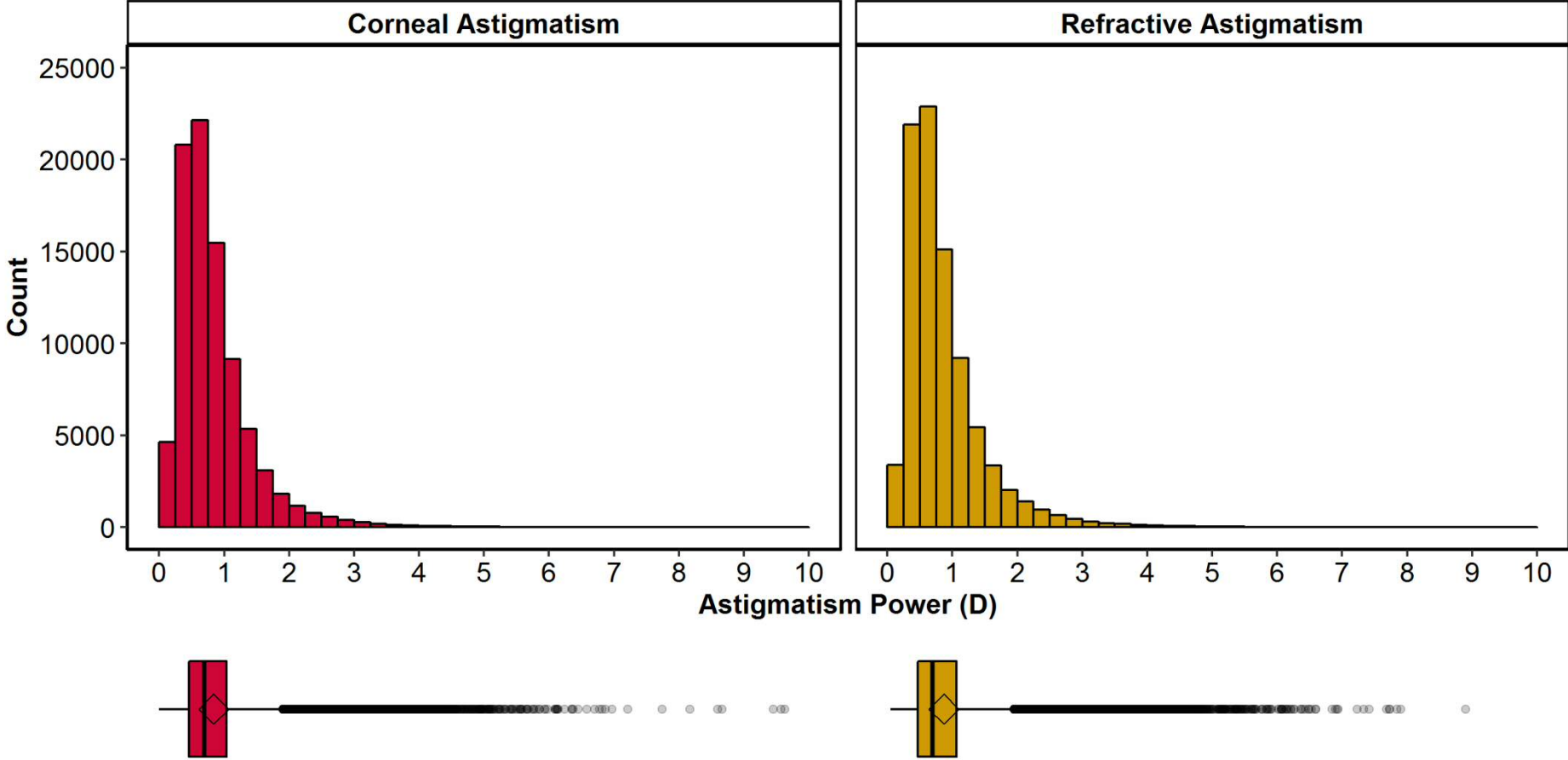
<sup>b</sup> After excluding for unreliable readings, ocular pathology / eye surgery (see *Section 2.2.3*)

The demographics of the analysed samples and the distribution of the corneal astigmatism and refractive astigmatism phenotypes are summarised in Table 6.2 and Figure 6.1.

**Table 6.2: Participant demographics and distribution of corneal and refractive astigmatism in the analysed UK Biobank sample.**

Trait	N	% Female	Age (Years)	Astigmatism Power (D)	
			Mean (SD)	Median (IQR)	Range
<b>Corneal Astigmatism</b>	86,355	54.1	58.1 (7.9)	0.695 (0.460-1.038)	0.000-9.627
<b>Refractive Astigmatism</b>	88,005	53.0	58.0 (7.8)	0.698 (0.470-1.062)	0.052-8.897

**Figure 6.1: Distribution of Corneal and Refractive Astigmatism in the analysed UK Biobank sample (N = 86,335 and 88,005 respectively).** Lower panels are boxplots representing their respective distributions (upper). Diamond symbol = mean astigmatism power.



### **6.2.2 Consideration of Astigmatism Axis**

Astigmatism, whether corneal or refractive, is a vectorial variable: the cylindrical power and its axis. As discussed in *Section 4.4*, the axes of corneal and refractive astigmatism typically change from WTR to ATR with increasing age. This phenomenon was verified in the UK Biobank cohort using data for the individuals available for inclusion in the single marker tests of corneal and refractive astigmatism (N = 86,335 and 88,005 respectively). WTR astigmatism was defined as negative cylindrical power axes between 0-30° and 150-180°. ATR astigmatism was defined as negative cylindrical power axes between 60-120°. All other axes were defined as oblique astigmatism. For cylindrical powers recorded using positive power notation, the reverse definitions applied for WTR and ATR astigmatism. “NA” astigmatism axis values were recorded if the cylindrical power was recorded as 0 or “NA” during keratometry/autorefraction.

The cylindrical power and its axis can be decomposed into a pair of dioptric values using vector notation (Thibos, Wheeler and Horner, 1997). The resultant values,  $J_0$  and  $J_{45}$ , were considered as alternative traits for genetic analysis.  $J_0$  refers to the horizontal-vertical (WTR-ATR) component of astigmatism, whereas  $J_{45}$  refers to oblique astigmatism. Positive  $J_0$  values represent WTR astigmatism, with negative  $J_0$  values representing ATR astigmatism. To obtain these vector transformed values, Equation 6.1 was applied to each individual measurement before taking the mean  $J_0$  and  $J_{45}$  values for each eye individually across assessment centre visits:



— —

**Equation 6.1: Vector transformation of astigmatism power and axis to its respective  $J_0$  and  $J_{45}$  components.**  $C$  = cylindrical (astigmatism) power (including its sign);  $\theta$  = axis of cylindrical power,  $C$ . Adapted from Thibos et al. (1997).

### **6.2.3 Single Marker Tests**

Genome-wide single marker association tests were undertaken for corneal astigmatism ( $N = 86,335$ ) and for refractive astigmatism ( $N = 88,005$ ), each considered as continuous traits using the standard “infinitesimal” mixed linear model approach implemented in BOLT-LMM v2.3 (Loh et al., 2015; Loh et al., 2018). This approach was utilised since mixed linear models provide greater power in association tests than standard linear regression (Yang et al., 2014). Residual population structure and cryptic relatedness were accounted for as random effects in this approach using binary genetic data files (PLINK format .bed, .bim and .fam files) for the “high-confidence” variants genotyped or imputed in these individuals (*Section 2.2.5*). Therefore, related individuals were not excluded from the BOLT-LMM analyses. Regional association plots for genome-wide significant loci were created using LocusZoom (Pruim et al., 2010). Conditional analysis was performed on the summary statistics from these BOLT-LMM analyses using GCTA-COJO as outlined in *Section 2.3.4*.

As corneal and refractive astigmatism are not normally distributed traits, analyses were also performed using BOLT-LMM after applying an inverse normal transformation to the phenotypes of the individuals included in the respective single marker tests. The inverse normal transformation was applied as this rank-

based method tends to provide better approximations of normality in comparison to alternative transformation methods such as Box-Cox and log-transformation (Goh and Yap, 2009). Sensitivity analyses were performed using PLINK 2.0, with corneal and refractive astigmatism considered as continuous traits (as in BOLT-LMM analyses) and then by classifying astigmatism as a dichotomous trait using a threshold value of astigmatism  $\geq 1.00$  D to define case status. As PLINK 2.0 used linear/logistic regression methods to run association analyses, these tests were restricted to unrelated individuals. A pairwise relatedness threshold of 0.025 was applied to remove one of each pair of related individuals from these PLINK 2.0 analyses. In total, 69,140 and 70,505 unrelated individuals were tested for association with corneal astigmatism and refractive astigmatism respectively in these PLINK 2.0 analyses. Case prevalences were 0.27 and 0.28 respectively for these dichotomous trait analyses. Sensitivity analyses were also performed for any identified genome-wide significant loci using the above samples of unrelated individuals to determine whether these associations were driven by individuals with extreme phenotypes or anisometropia. Individuals with more than a 2 D difference in astigmatism power between the two eyes or mean astigmatism power greater than 4 D were excluded, with corneal and refractive astigmatism considered as continuous traits (as in BOLT-LMM analyses).

For all analyses, variant filtering followed the approach of Fan et al. (2016), whereby variants with missingness  $> 0.01$ , MAF  $< 0.01$ , Hardy-Weinberg disequilibrium test p-value  $< 1 \times 10^{-6}$  or imputation quality  $< 0.4$  were excluded, along with samples with missingness  $> 0.05$ . In all instances, covariates included were: genotyping array, sex,

the mean age of the participant when eye measures were recorded, and mean spherical equivalent (as outlined in *Section 2.2.4* and implemented in the SNP-heritability estimation of the interim sample from the UK Biobank (*Chapter 5*)). Inflation of test statistics due to unadjusted population effects were determined by calculating genomic inflation factors and the intercept from LD Score regression as outlined in *Section 2.3.2*.

For loci achieving the genome-wide significance threshold of  $P < 5 \times 10^{-8}$ , previously identified associations with other ocular traits were identified via the NHGRI-EBI Catalog of published genome-wide association studies (MacArthur et al., 2017).

#### **6.2.4 Gene-based and Gene-set Analyses**

Gene-based and gene-set tests were performed using MAGMA (outlined in *Section 2.3.5*). Genes were defined according to NCBI build 37 (hg19/GRCh37) coordinates, with the inclusion of a 50 kb flanking region added to the transcription start/stop positions. LD patterns were estimated by MAGMA using an ancestry matched reference file, specifically the reference files composed of data for the 503 unrelated individuals of European ancestry from Phase 3 of the 1000 Genomes Project. For the gene-based tests, multiple testing was accounted for by applying a false discovery rate threshold of 5%.

Potential functional properties of genes associated with corneal astigmatism in the MAGMA analysis were further investigated using competitive gene-set analysis in

MAGMA. Adjustment for multiple testing was applied using a false discovery rate of 5% for these gene-based and gene-set test results.

#### **6.2.5 SNP-Heritability and Genetic Correlation**

Using LDSC (Bulik-Sullivan et al., 2015a; Bulik-Sullivan et al., 2015b), the SNP-heritabilities of corneal astigmatism and refractive astigmatism were quantified using summary statistics from the single marker association tests conducted using BOLT-LMM. Genetic correlations between pairs of the three refractive error traits: corneal astigmatism, refractive astigmatism and spherical equivalent were also quantified using this method. Summary statistics for spherical equivalent were obtained from single marker association tests conducted using BOLT-LMM for the same variants and individuals as performed for refractive astigmatism. In all instances reference LD scores were obtained for individuals of European ancestry.

#### **6.2.6 Phenotypic Correlation**

Pearson correlation coefficients were calculated for all unrelated individuals included in the single marker association tests who had data available for all refractive error traits investigated (i.e. corneal astigmatism, refractive astigmatism and spherical equivalent; N = 63,466).

## 6.3 Results

### 6.3.1 *Consideration of Astigmatism Axis*

Changes in astigmatism power and axis with increasing age were examined in the UK Biobank sample using data for the individuals available for inclusion in the single marker tests of corneal and refractive astigmatism (N = 86,335 and 88,005 respectively). For both corneal and refractive astigmatism, the proportion of individuals with WTR astigmatism decreased with increasing age, while the proportion of individuals with ATR astigmatism simultaneously increased (Table 6.3, Figure 6.2). The proportion of individuals with oblique axes remained relatively constant with age for both corneal and refractive astigmatism, with any changes in these proportions likely due to individuals transitioning between WTR and ATR classifications. The apparent increase in individuals classified as “NA” is due to increasing numbers of individuals with phenotype data available for only a single eye.

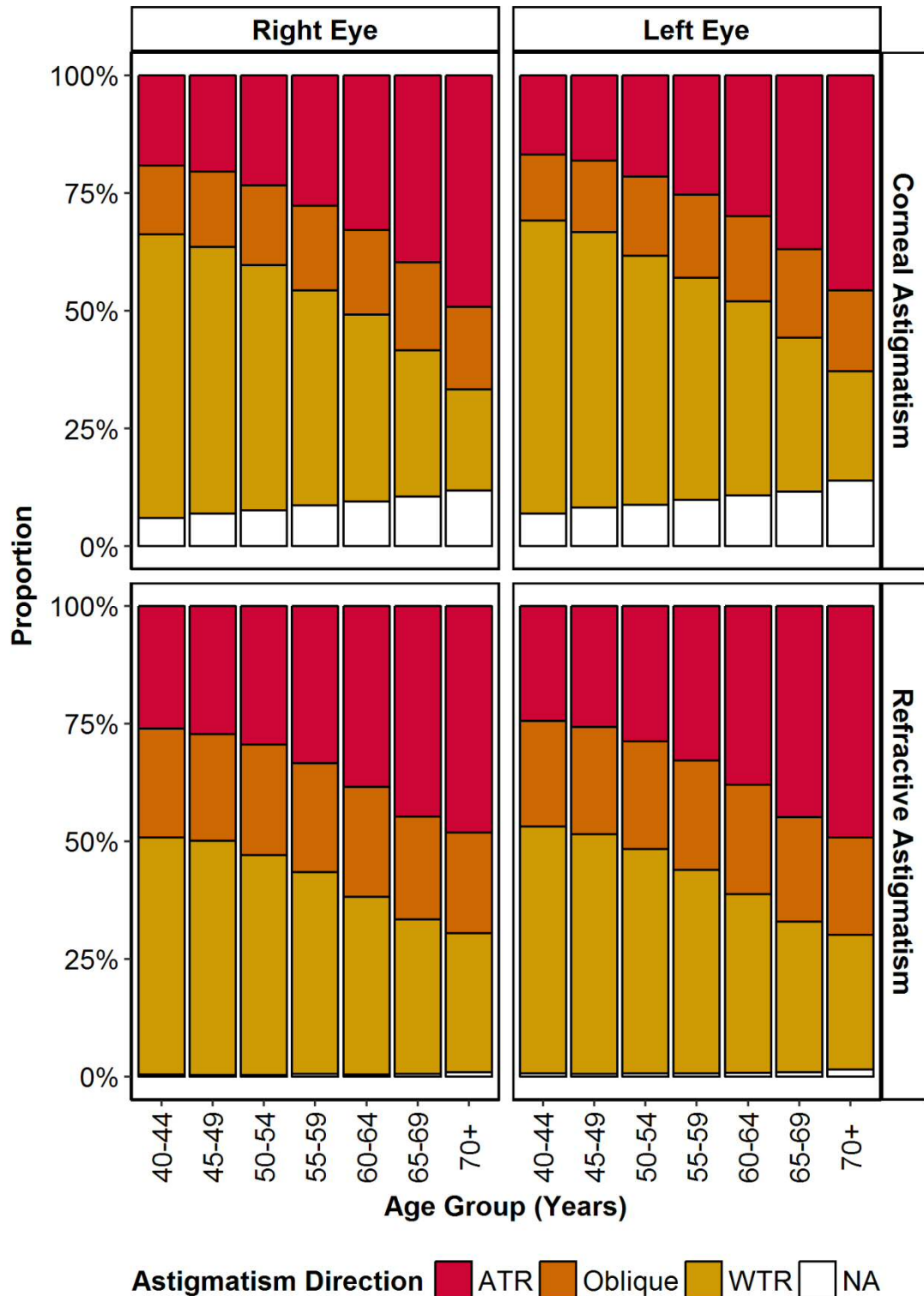
**Table 6.3: Proportion of individuals with corneal and refractive astigmatism described as with-the-rule, against-the-rule, or oblique, stratified by age in the UK Biobank sample.** WTR = with-the-rule astigmatism; ATR = against-the-rule astigmatism; oblique = oblique astigmatism powers, WTR = axis 0-30° or 150-180°; ATR = axis 60-120°; oblique = all other angles. The reverse applies for refractive astigmatism.

Corneal Astigmatism							
Age Group (Years)	Sample Size	Right Eye (%)				Left Eye (%)	
		WTR	Oblique	ATR	NA <sup>a</sup>	WTR	Oblique
<b>Overall</b>	86,335	43.9	17.5	29.8	8.8	45.5	17.3
<b>40-44</b>	6,705	60.2	14.6	19.2	6.0	62.2	14.0
<b>45-49</b>	9,996	56.6	16.1	20.4	6.9	58.5	15.2
<b>50-54</b>	12,167	52.0	17.0	23.4	7.6	52.8	16.9
<b>55-59</b>	15,052	45.6	18.0	27.7	8.7	47.2	17.7
<b>60-64</b>	23,210	39.7	18.0	32.9	9.5	41.1	18.1
<b>65-69</b>	17,884	31.1	18.7	39.8	10.5	32.7	18.8
<b>70+</b>	1,321	21.5	17.5	49.2	11.8	23.2	17.2

Refractive Astigmatism							
Age Group (Years)	Sample Size	Right Eye				Left Eye	
		WTR	Oblique	ATR	NA <sup>a</sup>	WTR	Oblique
<b>Overall</b>	88,005	41.2	22.9	35.4	0.5	41.6	22.8
<b>40-44</b>	6,929	50.4	23.0	26.1	0.4	52.4	22.4
<b>45-49</b>	10,320	49.7	22.7	27.2	0.4	51.0	22.8
<b>50-54</b>	12,590	46.7	23.5	29.5	0.4	47.7	22.8
<b>55-59</b>	15,532	42.9	23.1	33.4	0.6	43.2	23.2
<b>60-64</b>	23,715	37.8	23.3	38.5	0.5	38.0	23.3
<b>65-69</b>	17,687	32.8	21.8	44.8	0.6	31.9	22.2
<b>70+</b>	1,232	29.5	21.4	48.1	1.0	28.7	20.7

<sup>a</sup> "NA" astigmatism axis values were recorded if the cylindrical power was recorded as 0 or "NA" during keratometry.

**Figure 6.2: Prevalence of with-the-rule and against-the-rule astigmatism as a function of age.** Astigmatism was defined as with-the-rule when the negative cylindrical power was at an angle of 0-30° or 150-180°, against-the rule for angles 60-120°, and oblique for all other angles. The reverse applied for positive cylindrical powers. ATR = against the rule; WTR = with-the-rule. “NA” values denote individuals with no recorded astigmatism axis. Corneal astigmatism: N = 86,335; Refractive Astigmatism: N = 88,005.



In addition, vector transformation was applied to corneal and refractive astigmatism for this sample in order to combine their respective magnitude and direction parameters. As shown in Table 6.4 and Figure 6.3, the magnitude of corneal astigmatism tended to remain stable with increasing age (mean: 0.831 0.845 D); however, it did appear to increase in the oldest (70+) age group. After vector transformation, the  $J_0$  component of corneal astigmatism was found to gradually decrease with age and change from WTR (positive  $J_0$  values) to ATR (negative  $J_0$  values). The  $J_{45}$  component of corneal astigmatism appeared to remain relatively stable with increasing age.

Conversely, for refractive astigmatism, there was a gradual increase in magnitude with age from a mean of 0.73 – 0.87 D up to 1.10 – 1.17 D. The  $J_0$  and  $J_{45}$  vector components of refractive astigmatism followed similar patterns as their corneal counterparts with a gradual decrease in the  $J_0$  component, changing from WTR (positive  $J_0$  values) to ATR (negative  $J_0$  values) and relative stability in the  $J_{45}$  vector component with increasing age (Table 6.4, Figure 6.3).

As vector transformed values of astigmatism only provide detail regarding specific meridians (i.e.  $J_0$  = the horizontal-vertical aspect and  $J_{45}$  = the oblique aspect of astigmatism), performing analyses on one component would result in loss of information regarding the other contributing component. As a result, subsequent analyses of corneal and refractive astigmatism in the UK Biobank sample were restricted to their respective magnitudes, irrespective of axis.

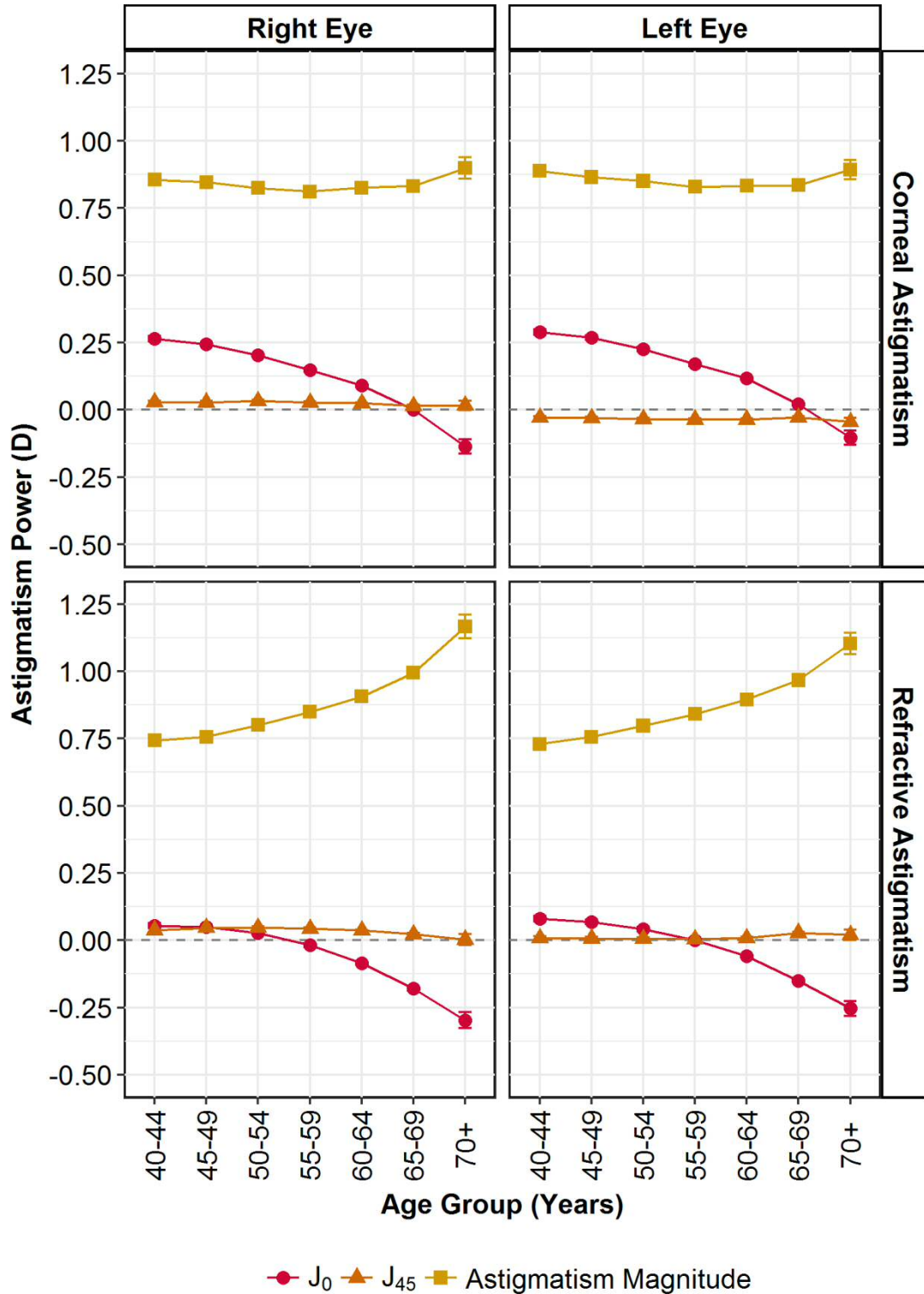


**Table 6.4: Summary of corneal and refractive astigmatism and their respective J<sub>0</sub> and J<sub>45</sub> components stratified by age group in the study sample.**

<b>Corneal Astigmatism (D)</b>							
<b>Age Group (Years)</b>	<b>Sample Size</b>	<b>Right Eye</b>			<b>Left Eye</b>		
		<b>Mean (SD)</b>	<b>Mean J<sub>0</sub> (SD)</b>	<b>Mean J<sub>45</sub> (SD)</b>	<b>Mean (SD)</b>	<b>Mean J<sub>0</sub> (SD)</b>	<b>Mean J<sub>45</sub> (SD)</b>
<b>Overall</b>	86,335	0.831 (0.636)	0.127 (0.439)	0.025 (0.250)	0.845 (0.644)	0.151 (0.438)	-0.025 (0.250)
<b>40-44</b>	6,705	0.856 (0.638)	0.265 (0.399)	0.029 (0.231)	0.888 (0.656)	0.290 (0.403)	-0.029 (0.231)
<b>45-49</b>	9,996	0.846 (0.644)	0.244 (0.401)	0.028 (0.245)	0.866 (0.649)	0.268 (0.398)	-0.028 (0.245)
<b>50-54</b>	12,167	0.825 (0.636)	0.203 (0.411)	0.033 (0.240)	0.851 (0.648)	0.225 (0.412)	-0.033 (0.240)
<b>55-59</b>	15,052	0.812 (0.619)	0.148 (0.419)	0.027 (0.246)	0.829 (0.642)	0.170 (0.423)	-0.027 (0.246)
<b>60-64</b>	23,210	0.826 (0.636)	0.091 (0.444)	0.025 (0.251)	0.833 (0.647)	0.117 (0.441)	-0.025 (0.251)
<b>65-69</b>	17,884	0.832 (0.642)	0.002 (0.452)	0.014 (0.263)	0.835 (0.631)	0.022 (0.450)	-0.014 (0.263)
<b>70+</b>	1,321	0.898 (0.689)	-0.135 (0.457)	0.016 (0.290)	0.894 (0.620)	-0.103 (0.458)	-0.016 (0.290)

<b>Refractive Astigmatism (D)</b>							
<b>Age Group (Years)</b>	<b>Sample Size</b>	<b>Right Eye</b>			<b>Left Eye</b>		
		<b>Mean (SD)</b>	<b>Mean J<sub>0</sub> (SD)</b>	<b>Mean J<sub>45</sub> (SD)</b>	<b>Mean (SD)</b>	<b>Mean J<sub>0</sub> (SD)</b>	<b>Mean J<sub>45</sub> (SD)</b>
<b>Overall</b>	88,005	0.872 (0.695)	-0.052 (0.466)	0.037 (0.289)	0.860 (0.702)	-0.029 (0.463)	0.037 (0.289)
<b>40-44</b>	6,929	0.743 (0.682)	0.053 (0.424)	0.037 (0.259)	0.729 (0.693)	0.080 (0.419)	0.037 (0.259)
<b>45-49</b>	10,320	0.756 (0.677)	0.050 (0.423)	0.047 (0.265)	0.756 (0.692)	0.069 (0.421)	0.047 (0.265)
<b>50-54</b>	12,590	0.800 (0.666)	0.028 (0.434)	0.047 (0.273)	0.798 (0.692)	0.042 (0.437)	0.047 (0.273)
<b>55-59</b>	15,532	0.849 (0.690)	-0.018 (0.459)	0.043 (0.285)	0.841 (0.698)	0.001 (0.458)	0.043 (0.285)
<b>60-64</b>	23,715	0.907 (0.693)	-0.086 (0.473)	0.037 (0.295)	0.895 (0.699)	-0.059 (0.470)	0.037 (0.295)
<b>65-69</b>	17,687	0.994 (0.703)	-0.180 (0.481)	0.023 (0.314)	0.967 (0.701)	-0.150 (0.476)	0.023 (0.314)
<b>70+</b>	1,232	1.166 (0.781)	-0.297 (0.516)	0.003 (0.356)	1.103 (0.720)	-0.253 (0.492)	0.003 (0.356)

**Figure 6.3: Corneal astigmatism, refractive astigmatism and their respective vector transformed values as a function of age in the analysed UK Biobank sample.** Astigmatism magnitude = yellow squares,  $J_0$  (horizontal-vertical astigmatism component) = red circles;  $J_{45}$  (oblique astigmatism component) = orange triangles. Horizontal dashed lines denote Astigmatism Power = 0 D. Points denote mean values for each age group. Error bars denote 95% confidence intervals.



### **6.3.2 Single Marker Tests**

After restricting the analysis sample to individuals of White British ancestry and applying exclusions for missing covariates and eye disorders with the potential to alter the level of astigmatism, there were 86,355 individuals available for inclusion in the GWAS for corneal astigmatism and 88,005 individuals in the GWAS for refractive astigmatism. After applying marker restrictions, there were 5,901,671 and 5,900,115 variants available for inclusion in the corneal and refractive astigmatism analyses, respectively.

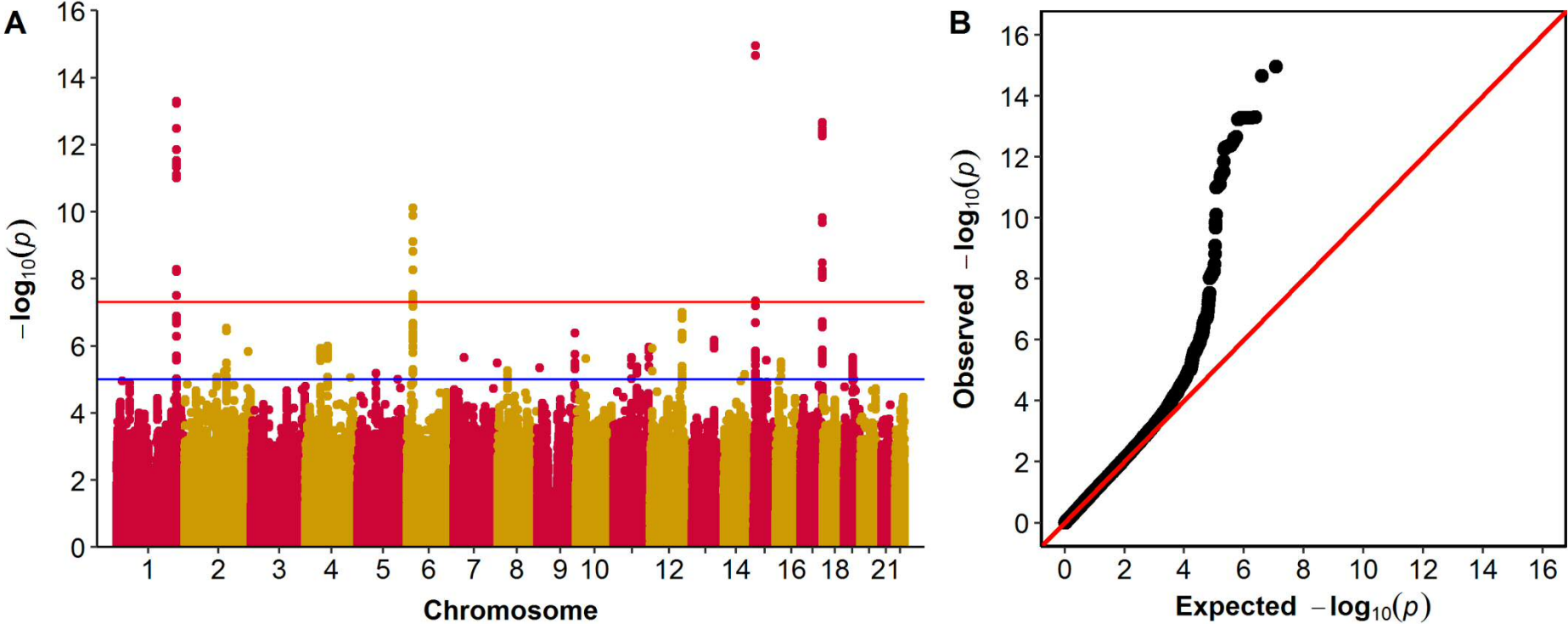
For the primary analyses using the infinitesimal mixed linear model approach implemented in BOLT-LMM, GWAS analyses identified 89 and 45 variants achieving genome-wide significant association ( $P < 5 \times 10^{-8}$ ) for corneal astigmatism and refractive astigmatism respectively (Table 6.5). Specifically, for corneal astigmatism, genome-wide significant variants clustered in four regions (Table 6.6, Figure 6.4, Figure 6.5), while for refractive astigmatism, they clustered in three regions (Table 6.8, Figure 6.6, Figure 6.7). Note that regions were defined as  $\pm 500$  kb of the lead variant.

**Table 6.5: Summary of GWAS undertaken.**  $\lambda_{GC}$  = genomic inflation factor;  $\lambda_{LDSC}$  = intercept from LD Score regression; independent loci was identified using GCTA-COJO. INT = inverse normal transformed (continuous).

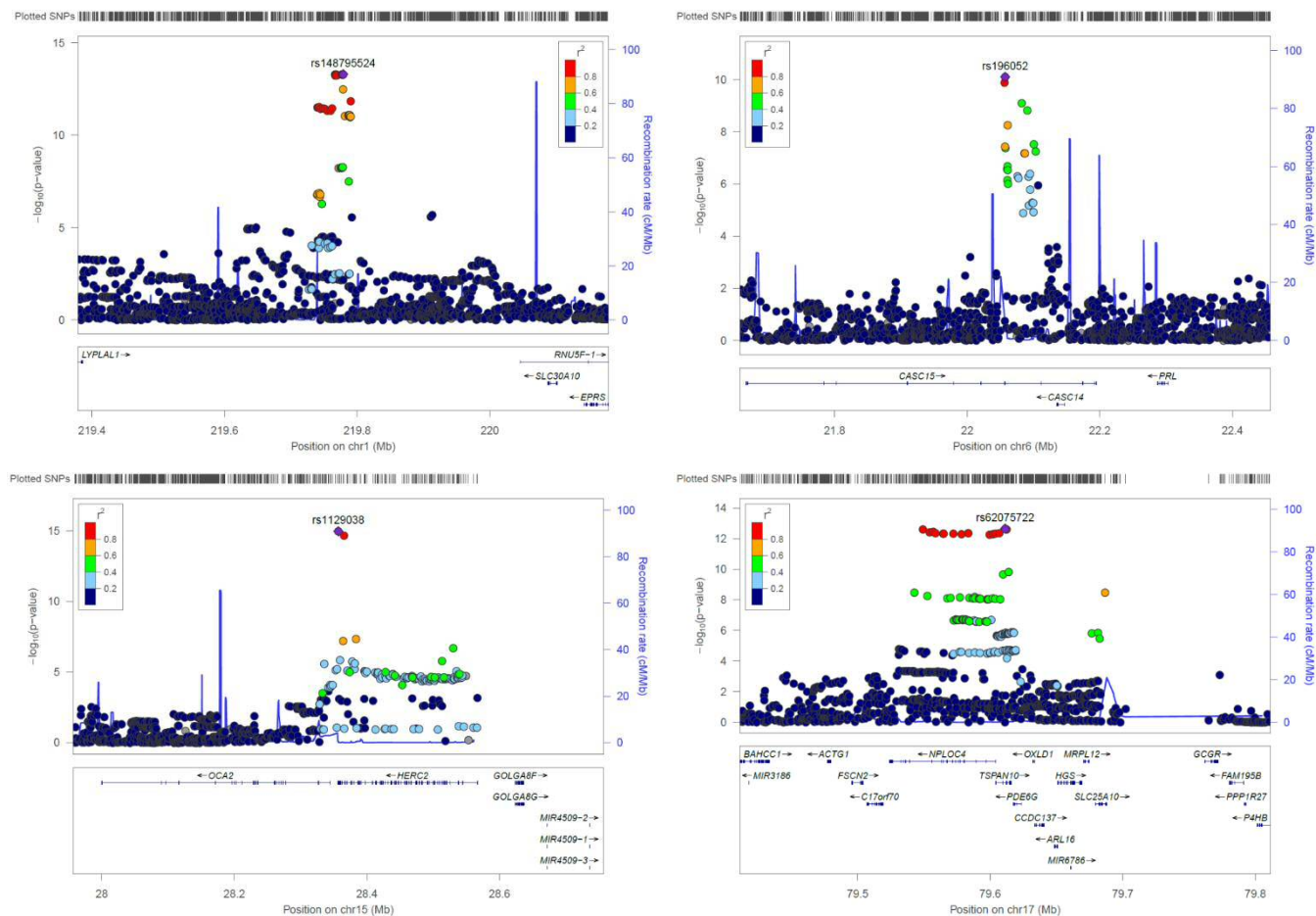
Trait		Method	No. of Variants	Sample Size	$\lambda_{GC}$	$\lambda_{LDSC}$	No. of Genome-wide Significant Variants ( $P < 5 \times 10^{-8}$ )
<b>Corneal Astigmatism</b>	Continuous	BOLT-LMM	5,901,671	86,335	1.094	1.023	89
	INT	BOLT-LMM	5,901,671	86,335	1.045	1.020	64
	Continuous	PLINK 2.0	5,921,785	69,140	1.057	1.020	54
	Cases $\geq 1.00$ D	PLINK 2.0	5,921,785	69,140	1.030	1.001	49
<b>Refractive Astigmatism</b>	Continuous	BOLT-LMM	5,900,115	88,005	1.045	1.005	45
	INT	BOLT-LMM	5,900,115	88,005	1.045	1.001	60
	Continuous	PLINK 2.0	5,919,636	70,505	1.042	1.003	13
	Cases $\geq 1.00$ D	PLINK 2.0	5,919,636	70,505	1.016	0.997	1

The most strongly associated marker located within the only previously identified genome-wide significant locus for corneal astigmatism, the promoter region of the *PDGFRA* gene at 4q12, was rs4864857 ( $P = 1.20 \times 10^{-6}$ ). Conditional analyses conducted by conditioning on the lead variant at each of the four novel genome-wide significant loci suggested that these four association signals were each driven by a single causal variant (Table 6.7). However, the strength of the association signal at the *HERC2* locus when conditioning on lead variant rs1129038 did yield a suggestive association signal within the adjacent *OCA2* gene (lead variant: rs1800407,  $P = 9.88 \times 10^{-6}$ ).

**Figure 6.4: Manhattan and Quantile-Quantile plots for GWAS of Corneal Astigmatism using BOLT-LMM.** Panel A: Manhattan plot: Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Panel B: Quantile-Quantile plot: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).



**Figure 6.5. Regional association plots for loci demonstrating genome-wide significant association ( $P < 5 \times 10^{-8}$ ) in GWAS for Corneal Astigmatism using BOLT-LMM (N = 86,335). In order of chromosome (top row: rs12032649, rs196052; bottom row: rs1129038, rs62075722). Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple). NB: rs14879552 is a synonym for rs12032649. Larger versions of these individual plots are available in Appendix B.**



**Table 6.6: Variants achieving association test p-values < 1 x 10<sup>-5</sup> in GWAS for Corneal Astigmatism analysed as a continuous trait with BOLT-LMM.** EAF = effect allele frequency; HWE P-value = P-value from the Hardy-Weinberg disequilibrium test; SE = standard error. NB: variants within ±500 kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	HWE P-value	Effect (SE)	P-value	Nearest Gene
rs1129038	15	28356859	C	T	0.215	0.621	-0.028 (0.003)	1.10 x 10 <sup>-15</sup>	<i>HERC2</i>
rs12032649	1	219778959	T	G	0.614	0.138	-0.022 (0.003)	5.00 x 10 <sup>-14</sup>	<i>ZC3H11B</i>
rs62075722	17	79611271	A	G	0.358	0.594	0.022 (0.003)	2.20 x 10 <sup>-13</sup>	<i>TSPAN10</i>
rs196052	6	22057200	T	A	0.622	0.477	0.019 (0.003)	7.80 x 10 <sup>-11</sup>	<i>LINC00340</i>
rs61935843	12	116617757	C	A	0.918	0.873	0.028 (0.005)	1.00 x 10 <sup>-7</sup>	<i>MED13L</i>
rs1579050	2	153364527	A	G	0.425	0.017	0.015 (0.003)	3.00 x 10 <sup>-7</sup>	<i>FMNL2</i>
rs10993820	9	136707730	A	G	0.791	0.803	-0.018 (0.003)	4.20 x 10 <sup>-7</sup>	<i>VAV2</i>
rs9517490	13	99584305	T	C	0.300	0.703	0.015 (0.003)	6.80 x 10 <sup>-7</sup>	<i>DOCK9</i>
rs1353386	4	81947080	A	C	0.148	0.808	0.019 (0.004)	1.00 x 10 <sup>-6</sup>	<i>BMP3</i>
rs7931326	11	130276347	C	G	0.935	0.716	-0.028 (0.006)	1.10 x 10 <sup>-6</sup>	<i>ADAMTS8</i>
rs4864857	4	55089814	T	C	0.784	0.166	-0.017 (0.003)	1.20 x 10 <sup>-6</sup>	<i>PGDFRA</i>
rs112947941	12	6997808	A	G	0.931	0.074	0.028 (0.006)	1.20 x 10 <sup>-6</sup>	<i>DSTNP2</i>
rs12473604	2	232401893	G	A	0.775	0.258	-0.016 (0.003)	1.50 x 10 <sup>-6</sup>	<i>NMUR1</i>
rs35313216	11	66224195	G	A	0.928	0.304	0.026 (0.005)	2.20 x 10 <sup>-6</sup>	<i>LOC100505524</i>
rs11084579	19	31802723	G	A	0.666	0.270	0.014 (0.003)	2.20 x 10 <sup>-6</sup>	<i>TSHZ3</i>
rs10279904	7	36806587	C	T	0.987	0.363	-0.058 (0.0012)	2.20 x 10 <sup>-6</sup>	<i>AOAH</i>
rs138016380	10	34449466	C	A	0.976	0.821	-0.044 (0.009)	2.40 x 10 <sup>-6</sup>	<i>PARD3</i>
rs11639295	15	67460757	C	T	0.706	0.902	0.015 (0.003)	2.70 x 10 <sup>-6</sup>	<i>SMAD3</i>
rs16971637	16	19155288	A	C	0.959	0.931	-0.033 (0.007)	3.00 x 10 <sup>-6</sup>	<i>ITPRIPL2</i>
rs117023057	7	158823501	G	A	0.984	0.092	-0.054 (0.012)	3.20 x 10 <sup>-6</sup>	<i>VIPR2</i>
rs2445565	11	86803194	G	C	0.470	0.119	-0.013 (0.003)	4.20 x 10 <sup>-6</sup>	<i>TMEM135</i>
rs12551905	9	7760772	T	C	0.985	0.908	-0.054 (0.012)	4.50 x 10 <sup>-6</sup>	<i>C9orf123</i>
rs149846728	8	36770379	G	A	0.949	0.332	0.030 (0.006)	5.40 x 10 <sup>-6</sup>	<i>KCNU1</i>



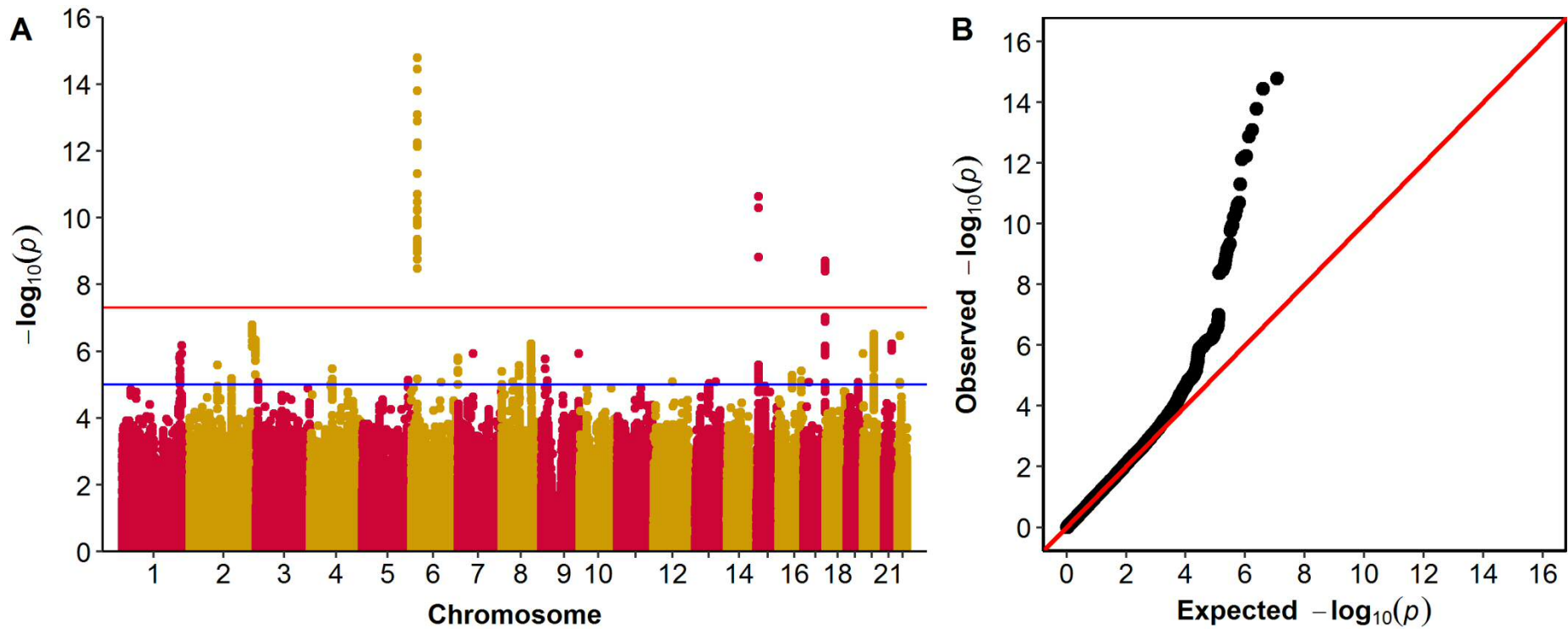
Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	HWE P-value	Effect (SE)	P-value	Nearest Gene
rs62169220	2	145225071	A	G	0.850	0.354	-0.018 (0.004)	5.80 x 10 <sup>-6</sup>	ZEB2
rs830557	5	67608743	C	T	0.529	0.368	0.013 (0.003)	6.50 x 10 <sup>-6</sup>	PIK3R1
rs56274409	14	96690828	A	T	0.944	0.567	-0.028 (0.006)	7.20 x 10 <sup>-6</sup>	BDKRB2
rs6741982	2	117793229	A	G	0.961	0.748	-0.033 (0.007)	8.70 x 10 <sup>-6</sup>	MTND2P21
rs6536686	4	163731498	C	T	0.192	0.921	0.016 (0.004)	8.80 x 10 <sup>-6</sup>	MIR4454
rs57770499	19	36260996	G	A	0.820	0.808	-0.016 (0.004)	1.00 x 10 <sup>-5</sup>	C19orf55
rs13181991	5	146163470	C	T	0.917	0.327	0.023 (0.005)	1.00 x 10 <sup>-5</sup>	PPP2R2B

**Table 6.7: Conditional analyses performed using summary statistics and conditioning on lead variant – for genome-wide significant associated loci from Corneal Astigmatism GWAS.** GCTA-COJO Lead Variant: lead variant within ±1 Mb of GWAS Lead Variant location.

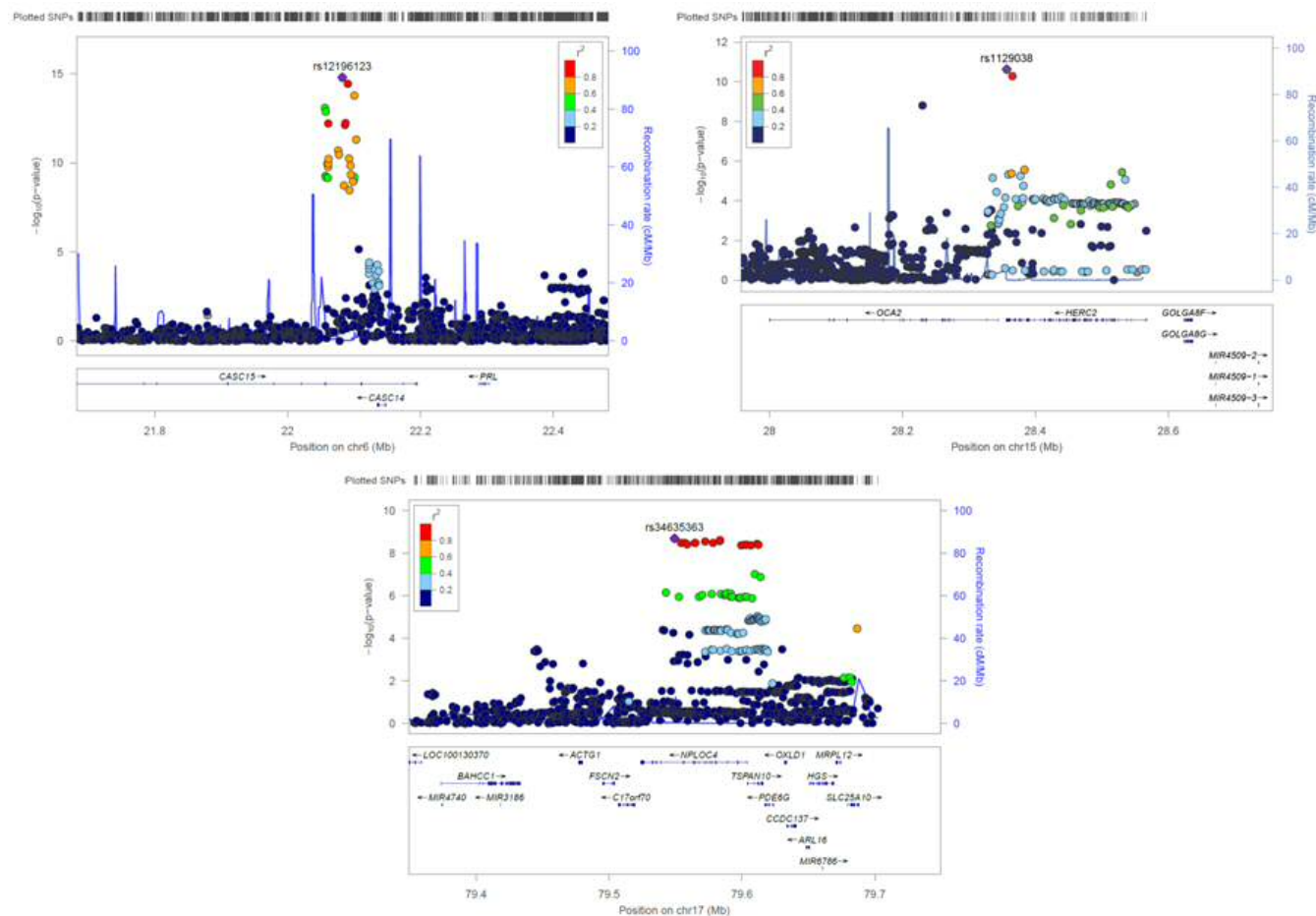
Chromosome	GWAS Lead Variant	P-value	GCTA-COJO Lead Variant	P-value
1	rs12032649	5.00 x 10 <sup>-14</sup>	NA (co-linearity error)	NA
6	rs196052	7.80 x 10 <sup>-11</sup>	rs114074527	1.84 x 10 <sup>-4</sup>
15	rs1129038	1.10 x 10 <sup>-15</sup>	rs1800407	9.88 x 10 <sup>-6</sup>
17	rs62075722	2.20 x 10 <sup>-13</sup>	rs113458760	3.23 x 10 <sup>-4</sup>

also suggested these association signals were each driven by a single causal marker with the exception of the association signal at *HERC2* which appeared to be driven by an additional independent causal marker within the *OCA2* gene at rs1800407 ( $P = 9.03 \times 10^{-15}$ ) (Table 6.9). Conditioning on both rs1129038 and rs1800407 at the *HERC2/OCA2* locus resulted in a suggestive association signal at rs7497044 ( $P = 1.90 \times 10^{-6}$ ), an intronic variant within the nearby *GABRG3* gene. In a previous meta-analysis of GWAS for corneal curvature in European ancestry cohorts from Australia (Mishra et al., 2012), marker rs17137734 within *GABRG3* achieved suggestive association ( $P = 9 \times 10^{-6}$ ). Pairwise LD between variants rs7497044, rs17137734, rs1129038 and rs1800407 varies from low to negligible in Europeans ( $r^2 = 0.0-0.1$ ).

**Figure 6.6: Manhattan and Quantile-Quantile plots for GWAS of Refractive Astigmatism using BOLT-LMM.** Panel A: Manhattan plot: Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Panel B: Quantile-Quantile plot: Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line: line of unity (observed = expected).



**Figure 6.7: Regional association plots for loci demonstrating genome-wide significant association ( $P < 5 \times 10^{-8}$ ) in GWAS for Refractive Astigmatism using BOLT-LMM (N = 88,005).** In order of chromosome (top row: rs12196123, rs1129038; bottom row: rs34635363). Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple). Larger versions of these individual plots are available in Appendix B.



**Table 6.8: Variants achieving association test p-values < 1 x 10<sup>-5</sup> in GWAS for Refractive Astigmatism analysed as a continuous trait with BOLT-LMM.** EAF: effect allele frequency, HWE P-value: P-value from the Hardy-Weinberg disequilibrium test, SE: standard error. NB: variants within ±500 kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	HWE P-value	Effect (SE)	P-value	Nearest Gene
rs12196123	6	22082263	C	T	0.443	0.659	-0.023 (0.003)	1.60 x 10 <sup>-15</sup>	<i>LINC00340</i>
rs1129038	15	28356859	C	T	0.215	0.047	-0.023 (0.004)	2.30 x 10 <sup>-11</sup>	<i>HERC2</i>
rs34635363	17	79549250	G	A	0.641	0.759	-0.018 (0.003)	2.00 x 10 <sup>-9</sup>	<i>NPLOC4</i>
rs10177414	2	228211470	T	C	0.598	0.020	-0.015 (0.003)	1.60 x 10 <sup>-7</sup>	<i>MFF</i>
rs6029691	20	40094364	C	G	0.688	0.621	-0.016 (0.003)	3.10 x 10 <sup>-7</sup>	<i>CHD6</i>
rs139743	22	25299429	A	G	0.582	0.675	-0.015 (0.003)	3.50 x 10 <sup>-7</sup>	<i>SGSM1</i>
rs77008212	2	239307113	A	G	0.913	0.612	-0.026 (0.005)	4.70 x 10 <sup>-7</sup>	<i>TRAF3IP1</i>
rs141045115	21	42387103	G	T	0.971	0.795	-0.043 (0.009)	6.00 x 10 <sup>-7</sup>	<i>LINC00323</i>
rs10435539	8	109167551	G	A	0.795	0.051	0.018 (0.004)	6.10 x 10 <sup>-7</sup>	<i>AURKBPS1</i>
rs116771750	1	219699050	T	C	0.965	0.589	-0.039 (0.008)	6.70 x 10 <sup>-7</sup>	<i>ZC3H11B</i>
rs79999086	20	1058226	T	C	0.980	0.245	-0.051 (0.011)	1.20 x 10 <sup>-6</sup>	<i>LCN1P2</i>
rs17172445	7	55189215	G	T	0.973	0.687	-0.043 (0.009)	1.20 x 10 <sup>-6</sup>	<i>EGFR</i>
rs11244084	9	136191010	C	T	0.926	0.684	-0.027 (0.005)	1.20 x 10 <sup>-6</sup>	<i>PSMF1</i>
rs115732928	1	214154088	A	T	0.952	0.115	-0.033 (0.007)	1.30 x 10 <sup>-6</sup>	<i>PROX1-AS1</i>
rs57717978	6	170267973	C	T	0.880	0.354	-0.021 (0.004)	1.60 x 10 <sup>-6</sup>	<i>LOC101929541</i>
rs10494951	1	212429259	G	A	0.806	0.347	0.018 (0.004)	1.60 x 10 <sup>-6</sup>	<i>LINC00574</i>
rs141720143	9	13237186	A	C	0.987	0.349	-0.060 (0.013)	1.70 x 10 <sup>-6</sup>	<i>MPDZ</i>
rs77909168	2	100555866	C	G	0.985	0.872	-0.056 (0.012)	2.60 x 10 <sup>-6</sup>	<i>AFF3</i>
rs56288719	8	64825393	T	C	0.962	0.207	-0.036 (0.008)	2.70 x 10 <sup>-6</sup>	<i>LOC286184</i>
rs6535231	4	81951911	G	A	0.041	0.386	0.034 (0.007)	3.40 x 10 <sup>-6</sup>	<i>BMP3</i>
rs192290664	16	82799444	A	G	0.978	0.410	-0.046 (0.010)	3.90 x 10 <sup>-6</sup>	<i>CDH13</i>
rs12547340	8	1149028	G	T	0.790	0.445	0.016 (0.004)	4.00 x 10 <sup>-6</sup>	<i>DLGAP2</i>
rs149069109	16	48972368	C	T	0.977	0.920	-0.044 (0.010)	5.20 x 10 <sup>-6</sup>	<i>KLF8P1</i>

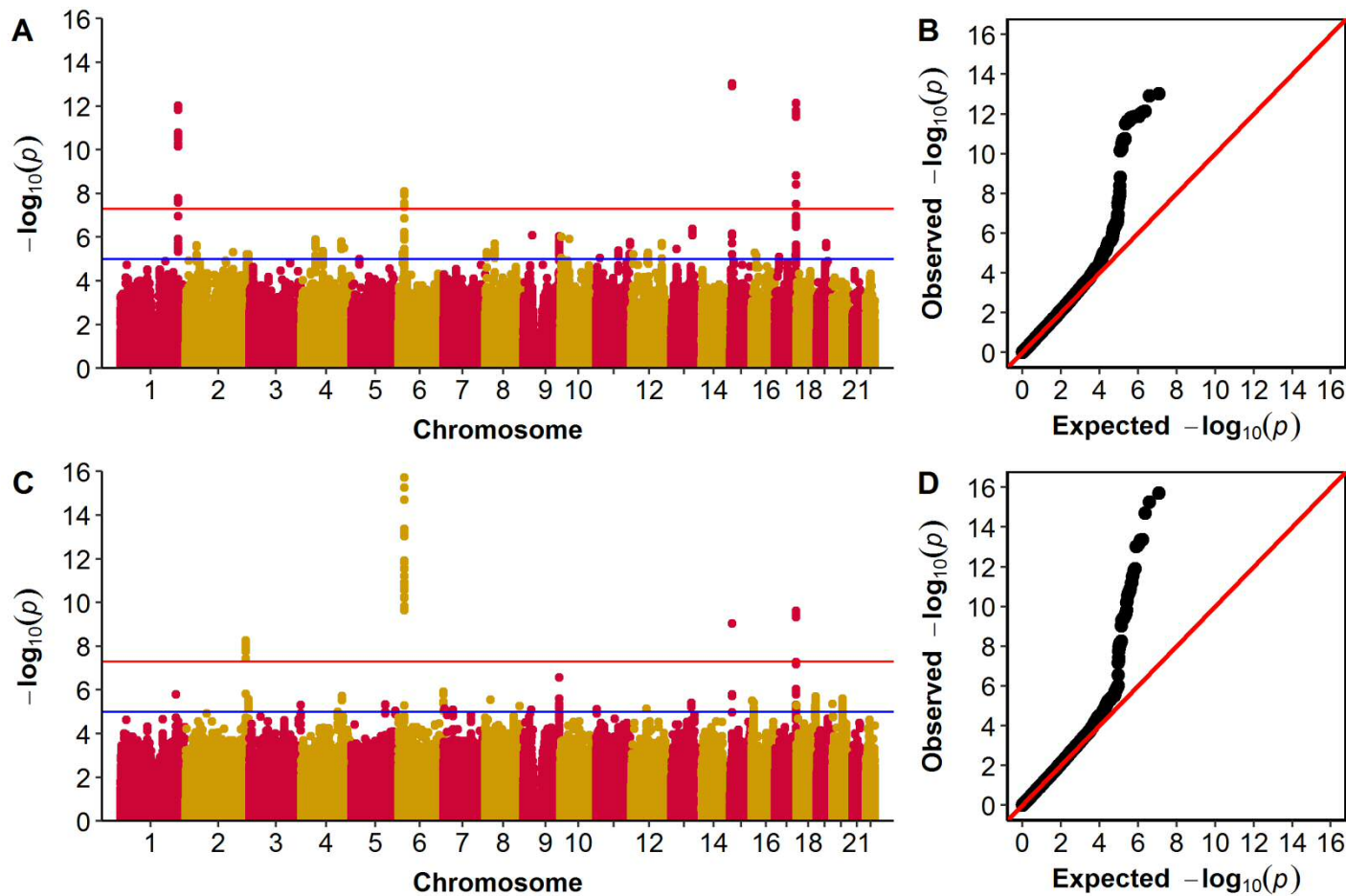
Variant	Chromosome	Position	Effect Allele	Other Allele	EAf	HWE P-value	Effect (SE)	P-value	Nearest Gene
rs6434068	2	153357541	G	C	0.428	0.647	0.013 (0.003)	6.50 x 10 <sup>-6</sup>	<i>FMNL2</i>
rs891933	5	167591402	C	T	0.537	0.329	0.013 (0.003)	7.30 x 10 <sup>-6</sup>	<i>TENM2</i>
rs116945318	9	21638723	A	C	0.985	0.779	-0.055 (0.012)	7.50 x 10 <sup>-6</sup>	<i>KHSRPP1</i>
rs974420	13	93152458	G	A	0.621	0.012	0.013 (0.003)	8.10 x 10 <sup>-6</sup>	<i>GPC5</i>
rs34314196	8	40998854	T	A	0.874	0.724	-0.019 (0.004)	8.20 x 10 <sup>-6</sup>	<i>SFRP1</i>
rs117949737	12	69299466	G	A	0.974	0.594	0.040 (0.009)	8.30 x 10 <sup>-6</sup>	<i>CPM</i>
rs8104928	19	42130284	A	G	0.972	0.185	-0.039 (0.009)	8.40 x 10 <sup>-6</sup>	<i>CEACAM4</i>
rs75819168	17	19651119	C	G	0.984	0.814	-0.051 (0.011)	8.50 x 10 <sup>-6</sup>	<i>ALDH3A1</i>
rs117812342	6	108473446	A	C	0.976	0.766	-0.042 (0.009)	8.60 x 10 <sup>-6</sup>	<i>OSTM1-AS1</i>
rs55939894	3	7963288	A	G	0.813	0.845	-0.016 (0.004)	8.70 x 10 <sup>-6</sup>	<i>LOC101927394</i>
rs513910	13	69095822	T	C	0.726	0.995	0.014 (0.003)	9.10 x 10 <sup>-6</sup>	<i>RPS3AP52</i>
rs62316885	4	75558798	C	T	0.962	0.879	-0.033 (0.008)	1.00 x 10 <sup>-5</sup>	<i>AREGB</i>

**Table 6.9: Conditional analyses performed using summary statistics and conditioning on lead variant – for genome-wide significant associated loci from Refractive Astigmatism GWAS.** GCTA-COJO Lead Variant: lead variant within ±1 Mb of GWAS Lead Variant location.

Chromosome	GWAS Lead Variant	P-value	GCTA-COJO Lead Variant	P-value
6	rs12196123	1.60 x 10 <sup>-15</sup>	rs148281502	2.29 x 10 <sup>-4</sup>
15	rs1129038	2.30 x 10 <sup>-11</sup>	rs1800407	9.03 x 10 <sup>-15</sup>
17	rs34635363	2.00 x 10 <sup>-9</sup>	rs75248478	9.17 x 10 <sup>-4</sup>

Since corneal and refractive astigmatism are not normally distributed traits, inverse normal transformations were applied to the phenotype data for the individuals included in the BOLT-LMM single marker tests. Results for corneal astigmatism revealed no change to the genome-wide significant associations identified compared to the untransformed trait (Table 6.10, Figure 6.8). For refractive astigmatism, three of the four loci identified previously continued to demonstrate genome-wide significant association, the exception being the association signal at *HERC2*, which was independent of the nearby association signal identified at *OCA2* for the untransformed trait. Furthermore, a locus at the *MPP* gene on chromosome 2 that demonstrated near genome-wide significant association for the untransformed refractive astigmatism trait now exceeded this significance threshold when tested using the transformed trait (lead variant: rs10177414; untransformed trait p-value =  $1.60 \times 10^{-7}$ , transformed trait p-value =  $5.40 \times 10^{-9}$ ) (Table 6.11, Figure 6.8).

**Figure 6.8: Manhattan and Quantile-Quantile plots for GWAS of Corneal and Refractive Astigmatism using BOLT-LMM as inverse normal transformed traits.** Panels A and B: Corneal Astigmatism; Panels C and D: Refractive Astigmatism. Manhattan plots (A and C): Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Quantile-Quantile plots (B and D): Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected).





**Table 6.10: The 10 variants demonstrating strongest association in GWAS for Corneal Astigmatism analysed as an inverse normal transformed trait with BOLT-LMM.** Effect sizes are in units of standard deviations. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

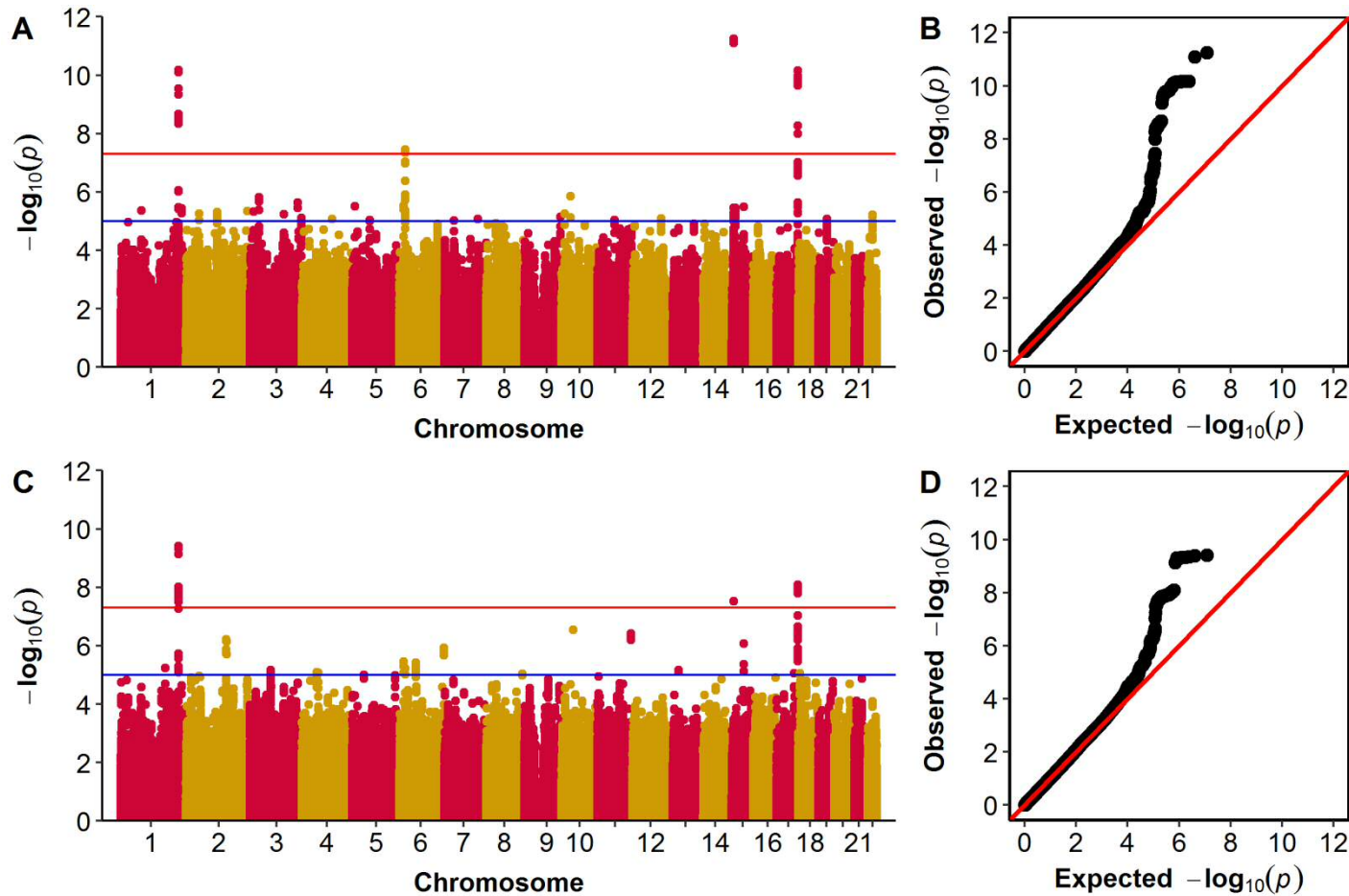
Variant	Chromosome	Position	Effect Allele	Other Allele	Effect (SE)	P-value	Nearest Gene
rs1129038	15	28356859	C	T	-0.043 (0.006)	$9.10 \times 10^{-14}$	<i>HERC2</i>
rs62075722	17	79611271	A	G	0.036 (0.005)	$7.30 \times 10^{-13}$	<i>TSPAN10</i>
rs12028838	1	219778675	G	T	-0.034 (0.005)	$9.70 \times 10^{-13}$	<i>ZC3H11B</i>
rs196052	6	22057200	T	A	0.029 (0.005)	$7.90 \times 10^{-9}$	<i>LINC00340</i>
rs9517490	13	99584305	T	C	0.026 (0.005)	$4.20 \times 10^{-7}$	<i>DOCK9</i>
rs75261924	9	34210395	T	C	0.089 (0.018)	$8.30 \times 10^{-7}$	<i>UBAP1</i>
rs10993820	9	136707730	A	G	-0.029 (0.006)	$9.00 \times 10^{-7}$	<i>VAV2</i>
rs11251716	10	3141970	C	G	0.024 (0.005)	$9.90 \times 10^{-7}$	<i>PFKP</i>
rs61859066	10	33285862	T	G	-0.044 (0.009)	$1.20 \times 10^{-6}$	<i>LOC101929475</i>
rs4864857	4	55089814	T	C	-0.028 (0.006)	$1.30 \times 10^{-7}$	<i>PDGFRA</i>

**Table 6.11: The 10 variants demonstrating strongest association in GWAS for Refractive Astigmatism analysed as an inverse normal transformed trait with BOLT-LMM.** Effect sizes are in units of standard deviations. NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	Effect (SE)	P-value	Nearest Gene
rs12196123	6	22082263	C	T	-0.038 (0.005)	$1.90 \times 10^{-16}$	<i>LINC00340</i>
rs34635363	17	79549250	G	A	-0.030 (0.005)	$2.40 \times 10^{-10}$	<i>NPLOC4</i>
rs1800407	15	28230318	C	T	-0.050 (0.008)	$9.10 \times 10^{-10}$	<i>OCA2</i>
rs10177414	2	228211470	T	C	-0.027 (0.005)	$5.40 \times 10^{-9}$	<i>MFF</i>
rs11244084	9	136191010	C	T	-0.045 (0.009)	$2.70 \times 10^{-7}$	<i>LCN1P2</i>
rs73792446	6	170265697	C	T	-0.034 (0.007)	$1.20 \times 10^{-6}$	<i>SGSM1</i>
rs10494951	1	212429259	G	A	0.028 (0.006)	$1.60 \times 10^{-6}$	<i>LINC00574</i>
rs1993423	4	154824112	T	C	0.022 (0.005)	$1.90 \times 10^{-6}$	<i>SFRP2</i>
rs966122	18	72538519	G	T	0.036 (0.007)	$2.00 \times 10^{-6}$	<i>ZNF407</i>
rs6029691	20	40094364	C	G	-0.023 (0.005)	$2.50 \times 10^{-6}$	<i>CHD6</i>

the three loci originally associated with refractive astigmatism (*LINC00340* and *HERC2*) continued to demonstrate genome-wide significant association (Tables 6.12 and 6.13, Figure 6.9). Additional sensitivity analyses were performed using logistic regression for the same groups of individuals and covariates as analysed by linear regression, with cases defined as individuals with corneal or refractive astigmatism  $\geq 1.00$  D. Here, three of the four previously identified loci, near the genes *ZC3H11B*, *HERC2* and *TSPAN10/NPLOC4*, demonstrated genome-wide significant association for corneal astigmatism, while only the *LINC00340* locus continued to demonstrate genome-wide significant association for refractive astigmatism (Tables 6.14 and 6.15, Figure 6.10). In all instances, the association signals were reduced using linear and logistic regression compared to the mixed linear model analyses. This was likely due to the substantial drop in sample size necessitated by standard regression based methods, which cannot account for relatedness between individuals.

**Figure 6.9: Manhattan and Quantile-Quantile plots for GWAS of Corneal Astigmatism using PLINK 2.0 as continuous and dichotomous (cases  $\geq 1.00$  D) traits.** Panels A and B: Continuous trait; Panels C and D: Dichotomous trait. Manhattan plots (A and C): Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Quantile-Quantile plots (B and D): Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected).



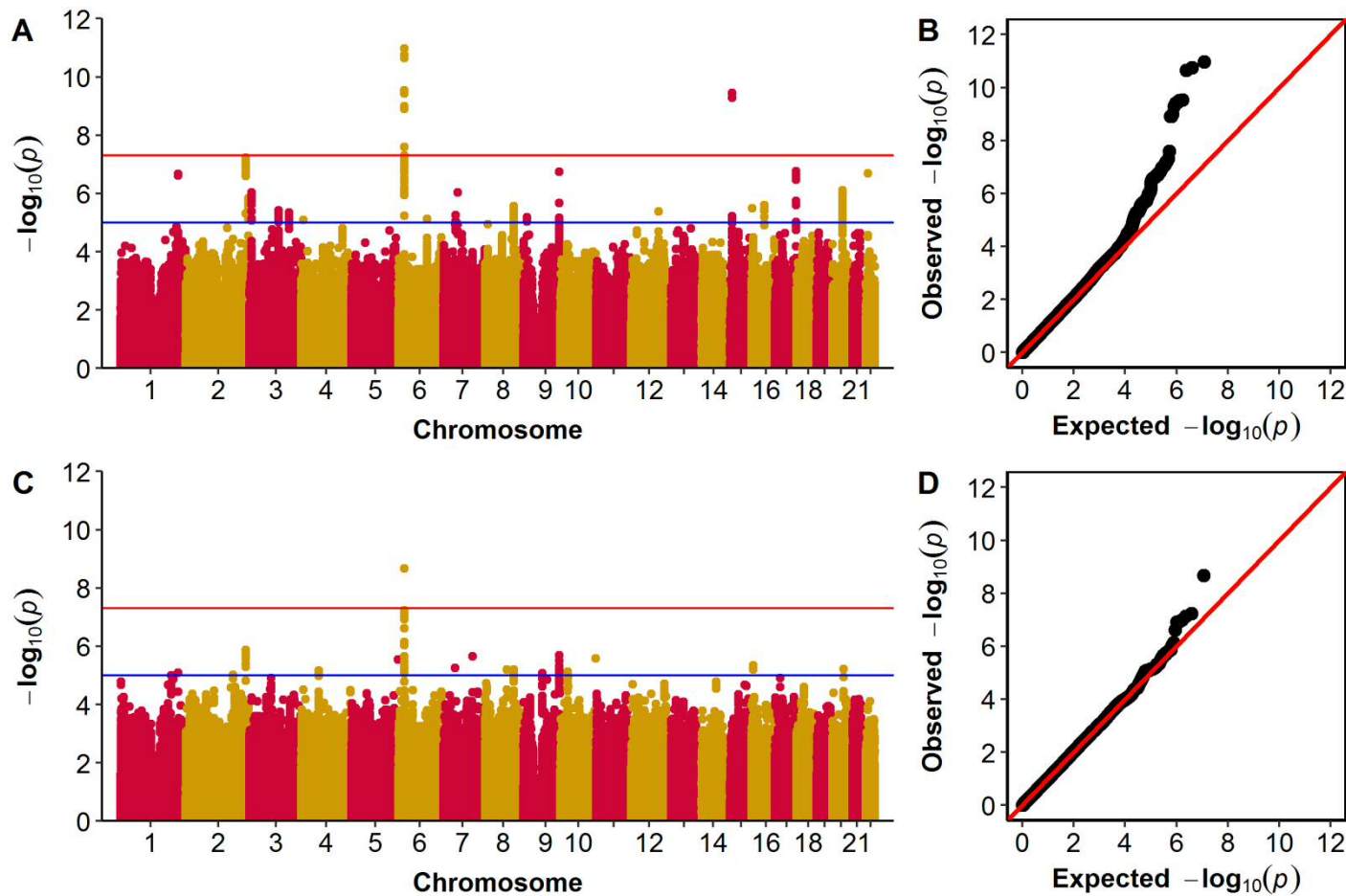
**Table 6.12: The 10 variants demonstrating strongest association in GWAS for Corneal Astigmatism analysed as a continuous trait with PLINK 2.0.** NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	Effect (SE)	P-value	Nearest Gene
rs1129038	15	28356859	T	C	0.026 (0.004)	$5.73 \times 10^{-12}$	<i>HERC2</i>
rs12032649	1	219778959	G	T	0.021 (0.003)	$6.60 \times 10^{-11}$	<i>ZC3H11B</i>
rs34635363	17	79549250	A	G	0.021 (0.003)	$7.09 \times 10^{-11}$	<i>NPLOC4</i>
rs196052	6	22057200	A	T	-0.018 (0.003)	$3.53 \times 10^{-8}$	<i>LINC00340</i>
rs138016380	10	34449466	A	C	0.050 (0.010)	$1.40 \times 10^{-6}$	<i>PAR3</i>
rs142566675	3	35248521	A	C	0.077 (0.016)	$1.51 \times 10^{-6}$	<i>LOC101928135</i>
rs75380963	3	181554299	T	C	0.079 (0.017)	$2.29 \times 10^{-6}$	<i>LOC101929598</i>
rs115822495	6	15681708	T	C	-0.072 (0.015)	$2.94 \times 10^{-6}$	<i>DTNBP1</i>
rs11955682	5	11416298	T	G	0.047 (0.010)	$3.15 \times 10^{-6}$	<i>CTNND2</i>
rs11639295	15	67460757	T	C	-0.016 (0.003)	$3.24 \times 10^{-6}$	<i>SMAD3</i>

**Table 6.13: The 10 variants demonstrating strongest association in GWAS for Corneal Astigmatism analysed as a dichotomous trait (cases  $\geq 1.00$  D) with PLINK 2.0.** NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	Odds Ratio (95% CI)	P-value	Nearest Gene
rs72629670	1	219768131	G	A	0.925 (0.901-0.950)	$3.95 \times 10^{-10}$	<i>ZC3H11B</i>
rs62075722	17	79611271	G	A	1.075 (1.051-1.100)	$8.14 \times 10^{-9}$	<i>TSPAN10</i>
rs12913832	15	28365618	G	A	0.920 (0.891-0.950)	$2.96 \times 10^{-8}$	<i>HERC2</i>
rs11238956	10	44749854	C	T	0.937 (0.912-0.962)	$2.89 \times 10^{-7}$	<i>C10orf142</i>
rs1623169	11	128584320	T	C	1.088 (1.056-1.121)	$3.80 \times 10^{-7}$	<i>FLI1</i>
rs6434068	2	153357541	C	G	1.063 (1.039-1.088)	$6.18 \times 10^{-7}$	<i>FMNL2</i>
rs11639295	15	67460757	T	C	1.069 (1.042-1.095)	$8.57 \times 10^{-7}$	<i>SMAD3</i>
rs2763272	6	168832091	T	C	0.932 (0.904-0.960)	$1.19 \times 10^{-6}$	<i>SMOC2</i>
rs115822495	6	15681708	T	C	1.343 (1.219-1.468)	$3.57 \times 10^{-6}$	<i>DTNBP1</i>
rs10223780	6	62865025	G	C	1.059 (1.035-1.083)	$3.90 \times 10^{-6}$	<i>KHDRBS2</i>

**Figure 6.10: Manhattan and Quantile-Quantile plots for GWAS of Refractive Astigmatism using PLINK 2.0 as continuous and dichotomous (cases  $\geq 1.00$  D) traits.** Panels A and B: Continuous trait; Panels C and D: Dichotomous trait. Manhattan plots (A and C): Red line indicates  $P = 5 \times 10^{-8}$ ; blue line indicates  $P = 1 \times 10^{-5}$ . Quantile-Quantile plots (B and D): Y-axis shows observed negative  $\log_{10}$  p-values and X-axis shows expected negative  $\log_{10}$  p-values according to the null hypothesis of no genetic association. Red line = line of unity (observed = expected).



**Table 6.14: The 10 variants demonstrating strongest association in GWAS for Refractive Astigmatism analysed as a continuous trait with PLINK 2.0.** NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	Effect (SE)	P-value	Nearest Gene
rs12196123	6	22082263	T	C	-0.022 (0.003)	$1.07 \times 10^{-11}$	<i>LINC00340</i>
rs1129038	15	28356859	T	C	-0.024 (0.004)	$3.63 \times 10^{-10}$	<i>HERC2</i>
rs10177414	2	228211470	C	T	-0.018 (0.003)	$5.95 \times 10^{-8}$	<i>MFF</i>
rs34635363	17	79549250	A	G	-0.018 (0.003)	$1.75 \times 10^{-7}$	<i>NPLOC4</i>
rs11244084	9	136191010	T	C	0.032 (0.006)	$1.78 \times 10^{-7}$	<i>LCN1P2</i>
rs139743	22	25299429	G	A	-0.017 (0.003)	$2.03 \times 10^{-7}$	<i>SGSM1</i>
rs116771750	1	219699050	C	T	-0.046 (0.009)	$2.11 \times 10^{-7}$	<i>ZC3H11B</i>
rs2072970	20	40112488	A	G	-0.017 (0.003)	$7.93 \times 10^{-7}$	<i>CHD6</i>
rs55939894	3	7963288	G	A	-0.020 (0.004)	$9.09 \times 10^{-7}$	<i>LOC101927394</i>
rs17172445	7	55189215	T	G	-0.049 (0.010)	$9.43 \times 10^{-7}$	<i>EGFR</i>



**Table 6.15: The 10 variants demonstrating strongest association in GWAS for Refractive Astigmatism analysed as a dichotomous trait (cases  $\geq 1.00$  D) with PLINK 2.0.** NB: variants within  $\pm 500$  kb of listed (lead) variant are not included in this list.

Variant	Chromosome	Position	Effect Allele	Other Allele	Odds Ratio (95% CI)	P-value	Nearest Gene
rs10946507	6	22100367	A	G	0.930 (0.906-0.954)	$2.15 \times 10^{-9}$	<i>LINC00340</i>
rs10177414	2	228211470	C	T	0.942 (0.918-0.966)	$1.34 \times 10^{-6}$	<i>MFF</i>
rs11244084	9	136191010	T	C	0.898 (0.853-0.942)	$2.07 \times 10^{-6}$	<i>LCN1P2</i>
rs73430835	7	111628421	C	T	0.777 (0.673-0.882)	$2.19 \times 10^{-6}$	<i>DOCK4</i>
rs2026781	10	133841039	C	G	1.067 (1.040-1.094)	$2.62 \times 10^{-6}$	<i>JAKMIP3</i>
rs114330092	5	178661923	T	G	0.866 (0.807-0.926)	$2.81 \times 10^{-6}$	<i>ADAMTS2</i>
rs12051285	16	7460115	G	T	1.057 (1.034-1.081)	$4.61 \times 10^{-6}$	<i>RBFOX1</i>
rs192526513	7	44989410	A	G	1.245 (1.150-1.340)	$5.54 \times 10^{-6}$	<i>MYO1G</i>
rs151198248	20	42389025	T	C	1.335 (1.210-1.460)	$6.19 \times 10^{-6}$	<i>LOC101927200</i>
rs76638496	8	82917849	A	G	0.829 (0.748-0.911)	$6.39 \times 10^{-6}$	<i>SNX16</i>

there was a greater decrease in effect size for all variants using these stricter inclusion criteria, and only the association signal on chromosome 6 continued to demonstrate genome-wide significant association (Table 6.17); however, the association signal on chromosome 17 did not demonstrate genome-wide significance when all unrelated individuals were considered. Based on these sensitivity analyses, it appears that the associations identified for corneal astigmatism are robust against the presence of extreme phenotype values or differences within individuals than refractive astigmatism. This can also be seen in the dichotomous trait analyses, whereby cases with astigmatism of 1.25 D or 5.00 D were considered equally despite the latter being an extreme trait value. In these dichotomous trait analyses, the number of genome-wide significant associations for refractive astigmatism reduced to a single genome-wide significant association on chromosome 6 (Table 6.15), whereas three of the four loci identified for corneal astigmatism continued to demonstrate such strong association (Table 6.13).

**Table 6.16: Sensitivity analyses investigating the effect of outlier phenotype values on association signals associated loci from Corneal Astigmatism GWAS.** *All Individuals* P-value obtained from analyses using PLINK = 69,140; Excluding phenotype extremes / anisometropia = 68,411.

Chromosome	GWAS Lead Variant	Effect Allele	All Individuals		Excluding Phenotype Extremes / Anisometropia	
			Effect (SE)	P-value	Effect (SE)	P-value
1	rs12032649	G	0.021 (0.003)	$6.60 \times 10^{-11}$	0.019 (0.003)	4.26 x 10 <sup>-11</sup>
6	rs196052	A	-0.018 (0.003)	$3.53 \times 10^{-8}$	-0.015 (0.003)	4.16 x 10 <sup>-8</sup>
15	rs1129038	C	-0.026 (0.004)	$5.73 \times 10^{-12}$	-0.021 (0.003)	4.18 x 10 <sup>-12</sup>
17	rs62075722	A	0.021 (0.003)	$1.61 \times 10^{-10}$	0.019 (0.003)	6.69 x 10 <sup>-11</sup>

**Table 6.17: Sensitivity analyses investigating the effect of outlier phenotype values on association signals associated loci from Refractive Astigmatism GWAS.** *All Individuals* P-value obtained from analyses using PLINK = 70,505; Excluding phenotype extremes / anisometropia = 69,304.

Chromosome	GWAS Lead Variant	Effect Allele	All Individuals		Excluding Phenotype Extremes / Anisometropia	
			Effect (SE)	P-value	Effect (SE)	P-value
6	rs12196123	C	-0.022 (0.003)	$1.07 \times 10^{-11}$	-0.016 (0.003)	6.60 x 10 <sup>-12</sup>
15	rs1129038	C	-0.024 (0.004)	$3.63 \times 10^{-10}$	-0.015 (0.003)	5.06 x 10 <sup>-10</sup>
17	rs34635363	A	0.018 (0.003)	$1.75 \times 10^{-7}$	0.013 (0.003)	6.63 x 10 <sup>-8</sup>

nine genes at the *TSPAN10/NPLOC4* locus (17q25.3; FDR =  $2.10 \times 10^{-6}$ ) along with the genes *HERC2* (15q13.1; FDR =  $2.30 \times 10^{-4}$ ), *PDGFRA* (4q12; FDR =  $5.21 \times 10^{-4}$ ), and *B3GNT7* (2q37.1; FDR =  $3.66 \times 10^{-3}$ ) (Table 6.18). For refractive astigmatism, gene-based analysis identified 35 genes with FDR < 0.05. Of these genes, seven were clustered at the gene dense *TSPAN10/NPLOC4* locus (17q25.3; FDR =  $5.00 \times 10^{-3}$ ). Additional genes identified included *TMEM211* (22q11.23; FDR =  $4.96 \times 10^{-3}$ ), *PROX1* (1q32.3; FDR =  $4.96 \times 10^{-3}$ ), *HERC2* (15q13.1; FDR =  $5.00 \times 10^{-3}$ ), and *PLAUR* (19q13.31; FDR =  $5.69 \times 10^{-3}$ ) (Table 6.19). The *VAX2* gene located at a suggested susceptibility locus for refractive astigmatism identified by Lopes et al. (2013) at (2p13.3) also achieved FDR < 0.05 (FDR = 0.014).

Gene-set analyses in MAGMA yielded non-significant findings after correction for multiple testing for both astigmatism traits.

**Table 6.18: Genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Corneal Astigmatism as a continuous trait.** Start and stop positions listed include  $\pm 50$  kb flanking regions; nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>NPLOC4</i>	17	79473913	79654138	716	6.34	$1.15 \times 10^{-10}$	$2.10 \times 10^{-6}$	$2.10 \times 10^{-6}$
<i>TSPAN10</i>	17	79559349	79665779	485	5.71	$5.75 \times 10^{-9}$	$5.27 \times 10^{-5}$	$1.05 \times 10^{-4}$
<i>PDE6G</i>	17	79567489	79673607	479	5.46	$2.36 \times 10^{-8}$	$1.44 \times 10^{-6}$	$4.33 \times 10^{-4}$
<i>HERC2</i>	15	28306183	28617313	185	5.33	$5.01 \times 10^{-8}$	$2.30 \times 10^{-4}$	$9.19 \times 10^{-4}$
<i>CCDC137</i>	17	79583761	79690937	470	5.24	$7.88 \times 10^{-8}$	$2.89 \times 10^{-4}$	$1.45 \times 10^{-3}$
<i>OXLD1</i>	17	79582066	79683618	476	5.16	$1.24 \times 10^{-7}$	$3.78 \times 10^{-4}$	$2.27 \times 10^{-3}$
<i>PDGFRA</i>	4	55045264	55214412	324	5.07	$1.99 \times 10^{-7}$	$5.21 \times 10^{-4}$	$3.65 \times 10^{-3}$
<i>C17orf70</i>	17	79456911	79569429	402	4.93	$4.08 \times 10^{-7}$	$9.34 \times 10^{-4}$	$7.47 \times 10^{-3}$
<i>HGS</i>	17	79600962	79719151	423	4.82	$7.24 \times 10^{-7}$	$1.48 \times 10^{-3}$	0.013
<i>ARL16</i>	17	79598224	79700954	437	4.68	$1.47 \times 10^{-6}$	$2.69 \times 10^{-3}$	0.027
<i>B3GNT7</i>	2	232210335	232315875	168	4.59	$2.19 \times 10^{-6}$	$3.66 \times 10^{-3}$	0.040
<i>FSCN2</i>	17	79431022	79554156	393	4.57	$2.43 \times 10^{-6}$	$3.72 \times 10^{-3}$	0.045
<i>BMP3</i>	4	81902119	82028685	194	4.40	$5.30 \times 10^{-6}$	$7.47 \times 10^{-3}$	0.097
<i>TUT1</i>	11	62291431	62409109	102	4.10	$2.07 \times 10^{-5}$	0.027	0.380
<i>ISCU</i>	12	108905239	109013160	288	4.08	$2.25 \times 10^{-5}$	0.027	0.413
<i>EML3</i>	11	62319690	62430492	101	4.07	$2.33 \times 10^{-5}$	0.027	0.427
<i>VAV2</i>	9	136577016	136907496	941	4.03	$2.80 \times 10^{-5}$	0.028	0.513
<i>ROM1</i>	11	62330213	62432592	88	4.03	$2.84 \times 10^{-5}$	0.028	0.520
<i>GANAB</i>	11	62342298	62464198	110	4.02	$2.87 \times 10^{-5}$	0.028	0.526
<i>B3GAT3</i>	11	62332768	62439647	97	4.00	$3.11 \times 10^{-5}$	0.028	0.570
<i>MTA2</i>	11	62310675	62419312	90	3.98	$3.40 \times 10^{-5}$	0.030	0.623

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>EEF1G</i>	11	62277073	62391460	103	3.96	$3.79 \times 10^{-5}$	0.030	0.695
<i>ASB1</i>	2	239285382	239410891	300	3.95	$3.89 \times 10^{-5}$	0.030	0.714
<i>ARMC9</i>	2	232013294	232288606	542	3.95	$3.93 \times 10^{-5}$	0.030	0.720
<i>SNED1</i>	2	241888213	242088515	437	3.88	$5.19 \times 10^{-5}$	0.038	0.952
<i>INTS5</i>	11	62364320	62470774	106	3.85	$5.93 \times 10^{-5}$	0.041	1
<i>C2orf83</i>	2	228424806	228548036	508	3.84	$6.09 \times 10^{-5}$	0.041	1
<i>SMAD3</i>	15	67308036	67537533	448	3.82	$6.70 \times 10^{-5}$	0.044	1
<i>C1orf54</i>	1	150194687	150303335	119	3.79	$7.42 \times 10^{-5}$	0.044	1
<i>GLRA3</i>	4	175495367	175800465	769	3.78	$7.81 \times 10^{-5}$	0.044	1
<i>TMEM119</i>	12	108933622	109041894	260	3.78	$7.91 \times 10^{-5}$	0.044	1
<i>NCL</i>	2	232269459	232379208	233	3.77	$8.16 \times 10^{-5}$	0.044	1
<i>APH1A</i>	1	150187799	150291609	99	3.77	$8.17 \times 10^{-5}$	0.044	1
<i>COL6A5</i>	3	130014359	130253690	707	3.76	$8.48 \times 10^{-5}$	0.044	1
<i>MAP2K6</i>	17	67360838	67588470	296	3.76	$8.49 \times 10^{-5}$	0.044	1
<i>CIART</i>	1	150204930	150309505	135	3.75	$8.88 \times 10^{-5}$	0.044	1
<i>LGALS14</i>	19	40144946	40250088	295	3.75	$8.94 \times 10^{-5}$	0.044	1

**Table 6.19: Genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Astigmatism as a continuous trait.** Start and stop positions listed include  $\pm 50$  kb flanking regions; nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>TMEM211</i>	22	25281208	25385314	144	4.90	$4.72 \times 10^{-7}$	$4.96 \times 10^{-3}$	$8.65 \times 10^{-3}$
<i>PROX1</i>	1	214111278	214264853	278	4.88	$5.41 \times 10^{-7}$	$4.96 \times 10^{-3}$	$9.92 \times 10^{-3}$
<i>HERC2</i>	15	28306183	28617313	186	4.76	$9.86 \times 10^{-7}$	$5.00 \times 10^{-3}$	0.018
<i>TSPAN10</i>	17	79559349	79665779	484	4.74	$1.09 \times 10^{-6}$	$5.00 \times 10^{-3}$	0.020
<i>NPLOC4</i>	17	79473913	79654138	717	4.60	$2.15 \times 10^{-6}$	$5.69 \times 10^{-3}$	0.039
<i>PDE6G</i>	17	79567489	79673607	478	4.60	$2.15 \times 10^{-6}$	$5.69 \times 10^{-3}$	0.039
<i>PLAUR</i>	19	44100247	44224498	254	4.59	$2.17 \times 10^{-6}$	$5.69 \times 10^{-3}$	0.040
<i>OXLD1</i>	17	79582066	79683618	475	4.38	$5.87 \times 10^{-6}$	0.013	0.108
<i>CCDC137</i>	17	79583761	79690937	469	4.35	$6.71 \times 10^{-6}$	0.014	0.123
<i>VAX2</i>	2	71077720	71210576	281	4.33	$7.59 \times 10^{-6}$	0.014	0.139
<i>MFF</i>	2	228139867	228272552	144	4.09	$2.13 \times 10^{-5}$	0.031	0.391
<i>HGS</i>	17	79600962	79719151	278	4.07	$2.36 \times 10^{-5}$	0.031	0.432
<i>MDH1</i>	2	63765743	63884331	186	4.06	$2.44 \times 10^{-5}$	0.031	0.447
<i>ATP6V1B1</i>	2	71112998	71242561	484	4.04	$2.65 \times 10^{-5}$	0.031	0.485
<i>EMC9</i>	14	24558081	24660837	717	4.04	$2.67 \times 10^{-5}$	0.031	0.490
<i>PSME2</i>	14	24562574	24665855	144	4.04	$2.68 \times 10^{-5}$	0.031	0.492
<i>CALR</i>	19	12999414	13105304	278	4.02	$2.95 \times 10^{-5}$	0.032	0.541
<i>FARSA</i>	19	12983284	13094558	186	3.97	$3.57 \times 10^{-5}$	0.035	0.654
<i>ARL16</i>	17	79598224	79700954	484	3.97	$3.62 \times 10^{-5}$	0.035	0.664
<i>ZNF277</i>	7	111796643	112033989	717	3.95	$3.92 \times 10^{-5}$	0.036	0.718
<i>PSME1</i>	14	24555295	24658176	478	3.93	$4.28 \times 10^{-5}$	0.036	0.784

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>ASB1</i>	2	239285382	239410891	254	3.93	4.32 x 10 <sup>-5</sup>	0.036	0.793
<i>RSPO2</i>	8	108861544	109145913	475	3.89	5.06 x 10 <sup>-5</sup>	0.040	0.927
<i>PPP2R5A</i>	1	212408879	212585205	469	3.86	5.60 x 10 <sup>-5</sup>	0.041	1
<i>CYP7B1</i>	8	65458529	65761348	281	3.86	5.63 x 10 <sup>-5</sup>	0.041	1
<i>FAM169A</i>	5	74023399	74213832	144	3.84	6.03 x 10 <sup>-5</sup>	0.043	1
<i>OCA2</i>	15	27950021	28394483	278	3.82	6.57 x 10 <sup>-5</sup>	0.045	1
<i>FITM1</i>	14	24550675	24652058	186	3.81	6.87 x 10 <sup>-5</sup>	0.045	1
<i>ZNF407</i>	18	72215106	72827628	484	3.78	7.84 x 10 <sup>-5</sup>	0.048	1
<i>POGZ</i>	1	151325200	151495753	717	3.78	7.86 x 10 <sup>-5</sup>	0.048	1
<i>SF3A1</i>	22	30677977	30802931	478	3.77	8.17 x 10 <sup>-5</sup>	0.048	1
<i>GFM2</i>	5	73967029	74113196	254	3.76	8.34 x 10 <sup>-5</sup>	0.048	1
<i>RAD23A</i>	19	13006628	13114457	475	3.76	8.58 x 10 <sup>-5</sup>	0.048	1
<i>RNF31</i>	14	24566101	24679870	469	3.75	8.99 x 10 <sup>-5</sup>	0.048	1
<i>PPP1R2</i>	3	195191221	195320224	281	3.73	9.54 x 10 <sup>-5</sup>	0.050	1



### 6.3.4 SNP-Heritability and Genetic Correlation

LD Score regression-based SNP-heritability estimates for corneal and refractive astigmatism, calculated from the GWAS summary statistics for the continuous trait analyses described above, were 0.036 (SE = 0.006,  $P = 4.34 \times 10^{-10}$ ) and 0.034 (SE = 0.006,  $P = 2.71 \times 10^{-9}$ ), respectively (Table 6.20). These estimates were lower – albeit with overlapping 95% confidence intervals – compared to the equivalent estimates obtained directly using GCTA (Table 6.20).

**Table 6.20: Estimates of SNP-heritability ( $h^2_{\text{SNP}}$ ) using GCTA and LDSC.**  $h^2_{\text{SNP}}$  = SNP-heritability; SE: standard error; P-value = test of the null hypothesis ( $h^2_{\text{SNP}} = 0$ ).

Trait	Method	No. of Variants	Sample Size	$h^2_{\text{SNP}}$	SE	P-value
Corneal	GCTA	732,404	27,707	0.061	0.021	$1.19 \times 10^{-3}$
Astigmatism	LDSC	864,048	86,355	0.036	0.006	$4.34 \times 10^{-10}$
Refractive	GCTA	732,404	28,378	0.046	0.020	$7.74 \times 10^{-3}$
Astigmatism	LDSC	863,851	88,005	0.034	0.006	$2.71 \times 10^{-9}$

The genetic correlation between corneal and refractive astigmatism, calculated using LD Score regression, was 0.851 (SE = 0.068,  $P = 1.37 \times 10^{-35}$ ). In contrast, genetic correlations between the astigmatism traits and spherical equivalent were much weaker at -0.108 between corneal astigmatism and spherical equivalent; and -0.104 between refractive astigmatism and spherical equivalent (Table 6.21). In both of these instances, the genetic correlations were not significantly different from zero ( $P = 0.067$  and  $0.071$  respectively; Table 6.21).

Comparing genotypic correlations for the UK Biobank data against those obtained for the CREAM meta-analyses revealed the genetic correlation between corneal and

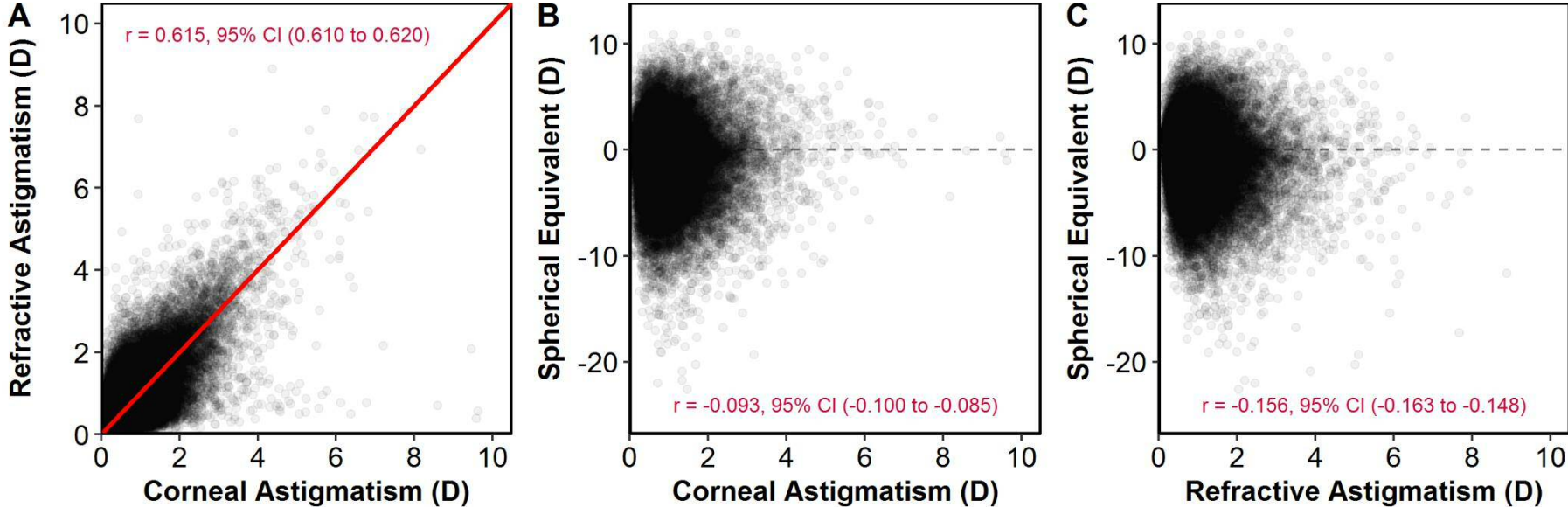
refractive astigmatism was much greater in the UK Biobank dataset (0.851 vs. 0.233). In addition, the genetic correlation between corneal astigmatism and spherical equivalent was in line with that obtained from the CREAM consortium data (Section 4.3.3) at -0.108; however the genetic correlation between refractive astigmatism and spherical equivalent was much reduced for the UK Biobank data at -0.104. A likely explanation for the differences in genetic correlations between the two datasets is the greater uncertainty in the SNP-heritability estimation for refractive astigmatism in the CREAM consortium data, which is subsequently used to calculate genetic correlation.

All phenotypic correlations were significantly different from the null hypothesis of no correlation (Table 6.21). Figure 6.11 shows the distributions of the respective refractive errors for all unrelated individuals (N = 63,466).

**Table 6.21: Genetic and phenotypic correlations between pairs of refractive error traits.** Genotypic correlations were obtained using LDSC and summary statistics from BOLT-LMM analyses (N = 86,335 or 88,005). All phenotypic correlations are Pearson correlations for 63,466 unrelated individuals included in GWAS for these respective traits and with data available for all three refractive error measures CA = corneal astigmatism; RA = refractive astigmatism; MSE = mean spherical equivalent. P-values refer to the null hypothesis of zero genetic correlation between traits.

Trait Pairs	No. of Variants	Genetic Correlation (SE)	P-value	Phenotypic Correlation	95% CI
CA and RA	862,521	0.851 (0.068)	$1.37 \times 10^{-35}$	0.615	0.610 to 0.620
CA and MSE	862,524	-0.108 (0.059)	0.067	-0.093	-0.100 to -0.085
RA and MSE	863,831	-0.104 (0.057)	0.071	-0.156	-0.163 to -0.148

**Figure 6.11: Phenotypic correlations between pairs of refractive error traits.** Panel A: Corneal Astigmatism and Refractive Astigmatism. Diagonal red line = line of unity (Corneal Astigmatism = Refractive Astigmatism); Panel B: Corneal Astigmatism and Spherical Equivalent. Horizontal dashed lines denote Spherical Equivalent = 0 D; Panel C: Refractive Astigmatism and Spherical Equivalent. Horizontal dashed lines denote Spherical Equivalent = 0 D. All correlations reported are Pearson correlations for respective pairs of refractive error measures in the 63,466 unrelated individuals included in GWAS for these respective traits who had data available for all three refractive error measures.



## 6.4 Discussion

The GWAS analyses undertaken here for corneal and refractive astigmatism are the largest performed to date, and have led to the discovery of four novel genome-wide significant loci associated with corneal astigmatism, and two novel genome-wide significant loci associated with refractive astigmatism.

It is notable that the novel loci for refractive or corneal astigmatism identified in the GWAS analyses undertaken here have previously shown association with other ocular traits. Variants near the protein coding gene *ZC3H11B* (zinc finger CCCH-type containing 11B) on chromosome 1 (1q41) have previously demonstrated association with pathological (high) myopia in Asian ancestry cohorts and with axial length in both European and Asian ancestry individuals (Fan et al., 2012; Cheng et al., 2013). Ocular expression of *ZC3H11B* has been identified in human retinal and scleral tissues (Fan et al., 2012). *LINC00340*, also known as *CASC15* (cancer susceptibility 15), is a long, non-coding RNA transcript located on chromosome 6 (6p22.3). In previous meta-analysis of GWAS from European and Asian ancestry cohorts, this locus demonstrated genome-wide significant association with spherical equivalent refractive error (Fan et al., 2016) and suggestive association ( $P < 1 \times 10^{-5}$ ) with refractive astigmatism (Li et al., 2015a). For both studies, associations at the locus appear to be largely driven by signals from European-ancestry cohorts, with little association demonstrated by their Asian ancestry counterparts. The protein coding gene *HERC2* (HECT and RLD domain containing E3 ubiquitin protein ligase 2) and its neighbouring gene *OCA2* (Oculocutaneous albinism type 2) on chromosome 15 (15q13.1) have both previously demonstrated association with eye, skin and hair

pigmentation (Kayser et al., 2008; Sturm and Larsson, 2009; Liu et al., 2010a). *TSPAN10* (Tetraspanin 10), also known as Oculospanin, is a protein coding gene located within a gene-dense region on chromosome 17 (17q25.3). This gene regulates the transmembrane metalloprotease *ADAM10* as part of the Notch signalling pathway (Charrin et al., 2014). Ocular expression of *TSPAN10* has been identified in the iris, ciliary body and retinal pigment epithelium (Wistow et al., 2002) and this locus has previously demonstrated genome-wide significant association with eye colour, myopia and age-related macular degeneration (Liu et al., 2010a; Fritsche et al., 2016; Pickrell et al., 2016).

Table 6.22 contains a summary of the loci achieving genome-wide significant association in this investigation and previously identified associations of these loci with other ocular traits, as reported in the NHGRI-EBI Catalog of published genome-wide association studies (MacArthur et al., 2017). With exception of the association signal at *HERC2/OCA2*, the majority of the astigmatism susceptibility loci have demonstrated association with refractive error traits in previous GWAS analyses.

**Table 6.22: Previously observed associations with ocular traits at the newly discovered susceptibility loci for astigmatism.**

Gene	Region	Previous Associations	References
<i>ZC3H11B</i>	1q41	Axial length Pathological myopia	Cheng et al. (2013) Fan et al. (2012)
<i>LINC00340</i>	6p22.3	Refractive astigmatism Spherical equivalent	Li et al. (2015a) Fan et al. (2016)
<i>HERC2</i>	15q13.1	Eye colour	Kayser et al. (2008)
<i>TSPAN10/NPLOC4</i>	17q25.3	Myopia	Pickrell et al. (2016)
		Advanced age-related macular degeneration Eye colour	Fritsche et al. (2016) Liu et al. (2010a)

In the earlier GWAS for corneal astigmatism (*Chapter 4*), conducted using data from CREAM, a genome-wide significant locus was identified at the *PDGFRA* gene. When analyses were restricted to individuals of European ancestry, this locus did not attain genome-wide significant association but was highly suggestive of association with corneal astigmatism (lead variant: rs7673984; OR (95% CI) = 1.11 (1.06-1.16);  $P = 5.64 \times 10^{-6}$ ). In analyses of the UK Biobank cohort with corneal astigmatism considered as a continuous trait, this locus demonstrated an equivalent degree of association (lead variant: rs4864857; effect (SE): 0.017 (0.003);  $P = 1.20 \times 10^{-6}$ ), and achieved a false discovery rate of 0.014 (1.4%) in the MAGMA gene-based test. This, together with evidence from association studies in other European ancestry cohorts (Mishra et al., 2012; Guggenheim et al., 2013a) strongly suggests that the *PDGFRA* gene locus should be considered as a susceptibility locus for corneal astigmatism in individuals of European ancestry.

The primary GWAS analyses were conducted using mixed linear models as implemented in BOLT-LMM (Loh et al., 2015). Mixed linear models have the advantage over standard linear regression that they can correct for residual population stratification and relatedness within the study sample, which can otherwise lead to reduced power or an excess of false positive association signals (Yang et al., 2014). Due to the increased sample size the mixed linear model approach allows, the genome-wide significant association signals obtained here were stronger than those obtained from standard linear regression. An important limitation of using mixed linear models for association studies is that they can produce results with high Type I error rates for dichotomous traits (Yang et al.,

2014; Chen et al., 2016); hence, corneal and refractive astigmatism were only considered as continuous traits for the mixed model analyses. Analysis methods work under the assumption that the trait is normally distributed; however, transforming corneal and refractive astigmatism values using an inverse normal transformation produced similar results to those obtained using the untransformed traits (Table 6.10, Table 6.11, Figure 6.8). Furthermore, the results were similar to those obtained using PLINK 2.0 with corneal and refractive astigmatism considered as continuous traits and as dichotomous traits defined using a threshold of 1.00 D for assigning case status (Tables 6.12-6.15, Figure 6.9, Figure 6.10).

Recent attempts have been made by software developers to overcome this Type I error rate for dichotomous traits through development of alternative models founded on the principles of mixed linear models (Chen et al., 2016; Zhou et al., 2017a). The alternative models (e.g. GMMAT and SAIGE) require initial fitting of a null logistic mixed model to determine model parameters, akin to those for the “non-infinitesimal” model from BOLT-LMM, before association test statistics are computed. As these models are still in their infancy, they were not utilised here. However, should the differences in Type I error rates of these models compared to standard logistic regression become comparable to linear regression vs. mixed linear models, all traits will benefit from the advantages of mixed model association tests.

As increasing magnitudes of astigmatism are correlated with increasing magnitudes of spherical refractive error (Kronfeld and Devney, 1930; Guggenheim and



Farbrother, 2004), spherical equivalent was included as a covariate in order to negate the effects of this correlation as a potential driver of association signals. Using spherical equivalent is more conservative an adjustment than using the spherical refractive component only, as this also adjusts for the contribution of refractive astigmatism to the overall refractive error of the individual (Guggenheim and Farbrother, 2004).

It should be noted that these genetic analyses of corneal and refractive astigmatism have solely considered the magnitude of astigmatism and not its axis. Astigmatism axis could be considered in a number of ways, such as analysing the vector transformed traits of  $J_0$  and  $J_{45}$  or by segregating the sample by astigmatism axis category (WTR, ATR or Oblique). However, these methods are not perfect as each option results in loss of data that may be considered important in understanding the development of astigmatism. In the case of using vector notation,  $J_0$  and  $J_{45}$  values tend to be considered together rather than isolation as it may be the case that an individual with  $J_0$  (or  $J_{45}$ ) = 0 D may not have 0 D of astigmatism overall. Whilst performing analyses using these alternative vector transformed traits may give insight into potential genetic variants associated with increased oblique astigmatism ( $J_{45}$ ) or astigmatism favouring the horizontal (or vertical) meridian ( $J_0$ ), combining the results of these analyses would provide results in the context of magnitude alone, as has been performed in the investigations in this section.

Alternatively, if the sample were to be segregated by axis category, with analyses performed for individuals with WTR astigmatism separately to those with ATR

astigmatism, variants demonstrating association with one specific category of astigmatism may be identified. However, there would be loss of power to detect true associated variants due to the reduced sample sizes in each group. A compromise could be to include astigmatism axis (either quantitative or categorical) as an additional covariate in the analysis model. However, care must be taken if analyses are performed using data from both eyes since astigmatism axes are rarely identical between eyes of an individual, and tend to follow a mirror symmetry pattern (e.g. right eye = 70°, left eye = 110°) (Guggenheim et al., 2008). In this case, categorising astigmatism axis would be favoured. However, added complexity arises if the axis for one eye is categorised as “WTR” and the other eye is categorised as “Oblique”, despite there being only a 1-2° difference in axes between the eyes after accounting for mirror symmetry (e.g. right eye = 29°, class = WTR; left eye = 149°, class = Oblique).

Furthermore, there may be individuals in the sample with no astigmatism identified (i.e. spherical corneal surface or spherical refractive error only). As these individuals would not have data available for astigmatism axis, they would not be included in the vector transformed or axis category analyses described above, thus resulting in loss of data that may help understanding of why people don't develop astigmatism. As a result of these considerations, maximising sample size and preference for a parsimonious analysis model – thus maximising statistical power, analysis was restricted to consideration of astigmatism magnitude alone.

In summary, these are the largest genome-wide association studies for corneal and refractive astigmatism to date and identified four novel loci for corneal astigmatism, two of which are also novel loci for refractive astigmatism. It was notable that all of these novel loci have previously been associated with different ocular traits (Table 6.22), most prominently spherical equivalent refractive error. However, the astigmatism association signals were genome-wide significant even after adjusting for the effects of spherical equivalent, confirming that they represent independent associations, thus lending further support to the concept of shared genetic susceptibility for myopia and astigmatism.

**Chapter 7 Investigation of myopic  
primate retina differentially expressed  
genes in humans**

## **Chapter 7 Investigation of myopic primate retina differentially expressed genes in humans**

### **7.1 Introduction**

The preceding chapters described experiments aiming to identify genetic variants associated with the refractive error traits spherical equivalent, corneal astigmatism and refractive astigmatism. Despite some successes in identifying candidate loci associated with these respective traits, and successful attempts to identify candidate genes enriched with variants through gene-based analyses, these analyses did not definitively link specific candidate genes to biological processes relevant to disease mechanisms.

One way to examine whether a candidate gene is implicated in the pathogenesis of a disorder is to assess changes in the expression level of the gene in individuals developing the disease (Westra and Franke, 2014). It follows that identifying expression quantitative trait loci (eQTLs), which are genetic variants associated with altered expression of specific genes (Nica and Dermitzakis, 2013), could aid determining which variants highlighted from GWAS are the true causal variants. Knowledge of such causal variants can then aid future development of potential genetic therapeutic and preventative interventions.

A key limitation when conducting analyses of gene expression is that the responses are tissue and time dependent (Chowers et al., 2003; Nica and Dermitzakis, 2008; Dimas et al., 2009). Thus, samples from which gene expression is examined are best

obtained from the appropriate tissue (e.g. retina or cornea, in the case of refractive errors) at a time when refractive error development is most active (i.e. adolescence). Collecting such samples is unfeasible in humans due to the invasive nature of the process; however such samples can be obtained from animal models.

For example, Tkatchenko et al. (2006) carried out an experiment in which juvenile rhesus macaques and green monkeys were form deprived via lid fusion, resulting in the induction of myopia in the form-deprived eyes. Significant differential expression of 119 genes was identified between the retinas of the lid fused and fellow control eyes of these primates.

The aim of the following investigation was to ascertain whether the human orthologues of the candidate genes identified by Tkatchenko et al. (2006) were enriched with naturally occurring genetic variants associated with spherical equivalent refractive error. The gene-based statistical methods employed in the preceding chapters were utilised for this purpose.

## **7.2 Methods**

### **7.2.1 Study Sample**

Results from a meta-analysis of GWAS for refractive error performed by the Consortium for Refractive Error and Myopia (CREAM) were used for this investigation (Tedja et al., 2018). The meta-analysis results used were restricted to cohorts of European ancestry only. This sample was obtained by Tedja et al. (2018) through combining cohorts from study groups known to CREAM and customers from the 23andMe genetic testing company database. Variants were included in the analysis carried out by Tedja et al. (2018) if they had a MAF  $\geq 0.01$  in the CREAM study cohorts or  $\geq 0.005$  in the 23andMe cohort, and an imputation quality score from IMPUTE2 info  $\geq 0.3$  or Minimac Rsq  $\geq 0.3$ .

This resulted in a total of 11,691,823 variants across the genome tested for association with refractive error in 148,218 individuals.

### **7.2.2 Gene-based Analyses**

To test whether the differentially expressed genes identified by Tkatchenko et al. (2006) were human myopia susceptibility genes, gene-based association tests were performed using MAGMA (outlined in *Section 2.3.5*) utilising the data from the CREAM refractive error GWAS. The candidate genes investigated were restricted to those with human autosomal orthologues, resulting in a total of 111 candidate genes. Genes were defined according to NCBI build 37 (hg19/GRCh37) coordinates. Due to the wide range of distances between variants and the genes they influence (Brodie et al., 2016), analyses were performed using the following flanking regions

appended to the gene's translation start and stop sites: 0 kb, 50 kb and 200 kb. LD patterns were estimated by MAGMA using an ancestry matched reference file, specifically the reference files composed of data for the 379 unrelated individuals of European ancestry from Phase 1, version 3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2012). Multiple testing was accounted for by applying a false discovery rate threshold of 5%. Further exploratory analyses were also carried out with even larger flanking regions (500 kb 1 Mb and 2 Mb).

### ***7.2.3 Replication in the UK Biobank Study***

Genes demonstrating  $FDR < 0.05$  were examined for replication in the independent, UK Biobank study cohort. Here, gene-based analyses were performed using results from a GWAS for spherical equivalent for 88,005 White British individuals as described in *Section 6.2.5*.



## 7.3 Results

### 7.3.1 Gene-based Analyses

In order to statistically ascertain whether the candidate genes from Tkatchenko et al. (2006) were enriched with variants attaining low but not necessarily genome-wide significant p-values from the GWAS of refractive error in humans, gene-based tests were performed in MAGMA. In total 111 autosomal genes were tested for association with refractive error.

Restricting gene regions to within translation start and stop sites (0 kb flanking regions) resulted in four of the 111 candidate genes demonstrating significant association (FDR < 0.05). These four genes were *LBH*, *CLU*, *DDIT4* and *SEPT4* (Table 7.1).

Initially, a flanking region of 50 kb was appended to the gene transcription start and stop sites. This resulted in nine of the candidate genes demonstrating significant association with refractive error (FDR < 0.05; Table 7.2). Included in this list of genes were the four genes demonstrating association when restricting to variants within the gene transcription locus, as well as the genes *TSR3*, *ERLEC1*, *SNW1*, *CNTF* and *SPON1* (Table 7.2).

Increasing flanking regions to include variants 200 kb upstream and downstream of gene regions resulted in 17 of the 111 candidate genes demonstrating significant association (FDR < 0.05; Table 7.3). Included in these 17 genes were all nine genes

demonstrating association when tested with the inclusion of the smaller, 50 kb, flanking region.

**Table 7.1: Candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Error.** Start and stop positions refer to the respective gene transcription start and stop sites (no flanking region included); nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>CLU</i>	8	27454434	27472328	40	5.21	$9.23 \times 10^{-8}$	$1.02 \times 10^{-5}$	$1.02 \times 10^{-5}$
<i>DDIT4</i>	10	74033677	74035797	4	4.60	$2.10 \times 10^{-6}$	$8.32 \times 10^{-5}$	$2.31 \times 10^{-4}$
<i>LBH</i>	2	30454397	30482899	102	4.59	$2.27 \times 10^{-6}$	$8.32 \times 10^{-5}$	$2.50 \times 10^{-4}$
<i>SEPT4</i>	17	56597611	56618179	39	4.39	$5.72 \times 10^{-6}$	$1.57 \times 10^{-4}$	$6.29 \times 10^{-4}$

**Table 7.2: Candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Error with the inclusion of 50 kb flanking regions.** nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>TSR3</i>	16	1349241	1451873	387	5.33	$4.83 \times 10^{-8}$	$5.31 \times 10^{-6}$	$5.31 \times 10^{-6}$
<i>DDIT4</i>	10	73983677	74085797	221	5.14	$1.38 \times 10^{-7}$	$7.60 \times 10^{-6}$	$1.52 \times 10^{-5}$
<i>LBH</i>	2	30404397	30532899	386	4.38	$5.93 \times 10^{-6}$	$2.18 \times 10^{-4}$	$6.53 \times 10^{-4}$
<i>CLU</i>	8	27404434	27522328	392	3.90	$4.78 \times 10^{-5}$	$1.27 \times 10^{-3}$	$5.26 \times 10^{-3}$
<i>SEPT4</i>	17	56547611	56668179	206	3.86	$5.76 \times 10^{-5}$	$1.27 \times 10^{-3}$	$6.33 \times 10^{-3}$
<i>ERLEC1</i>	2	53964068	54095956	398	3.07	$1.06 \times 10^{-3}$	0.019	0.117
<i>SNW1</i>	14	78133942	78277542	625	2.93	$1.71 \times 10^{-3}$	0.026	0.188
<i>CNTF</i>	11	58340146	58443206	280	2.90	$1.89 \times 10^{-3}$	0.026	0.207
<i>SPON1</i>	11	13934184	14339679	1192	2.77	$2.77 \times 10^{-3}$	0.034	0.304

**Table 7.3: Candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) for Refractive Error with the inclusion of 200 kb flanking regions.** nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>DDIT4</i>	10	73833677	74235797	821	5.66	$7.40 \times 10^{-9}$	$3.97 \times 10^{-7}$	$8.21 \times 10^{-7}$
<i>ANKRD28</i>	3	15508743	16101053	1663	5.65	$7.88 \times 10^{-9}$	$3.97 \times 10^{-7}$	$8.74 \times 10^{-7}$
<i>CNIH1</i>	14	54690279	55108322	964	5.60	$1.07 \times 10^{-8}$	$3.97 \times 10^{-7}$	$1.19 \times 10^{-6}$
<i>TSR3</i>	16	1199241	1601873	1765	5.10	$1.66 \times 10^{-7}$	$4.44 \times 10^{-6}$	$1.85 \times 10^{-5}$
<i>HNRNPA3</i>	2	177876063	178288687	982	5.07	$2.00 \times 10^{-7}$	$4.44 \times 10^{-6}$	$2.22 \times 10^{-5}$
<i>E2F4</i>	16	67026068	67432821	552	3.69	$1.12 \times 10^{-4}$	$1.86 \times 10^{-3}$	0.012
<i>SNW1</i>	14	77983942	78427542	1978	3.68	$1.17 \times 10^{-4}$	$1.86 \times 10^{-3}$	0.013
<i>SEPT4</i>	17	56397611	56818179	814	3.36	$3.85 \times 10^{-4}$	$5.34 \times 10^{-3}$	0.043
<i>LBH</i>	2	30254397	30682899	1308	3.19	$7.10 \times 10^{-4}$	$8.46 \times 10^{-3}$	0.079
<i>RAB18</i>	10	27593103	28031166	1578	3.17	$7.62 \times 10^{-4}$	$8.46 \times 10^{-3}$	0.085
<i>SPON1</i>	11	13784184	14489679	2031	3.04	$1.16 \times 10^{-3}$	0.012	0.129
<i>DGKE</i>	17	54711460	55146036	1159	2.96	$1.55 \times 10^{-3}$	0.014	0.172
<i>CLU</i>	8	27254434	27672328	1426	2.91	$1.78 \times 10^{-3}$	0.015	0.198
<i>CNTF</i>	11	58190146	58593206	1136	2.83	$2.30 \times 10^{-3}$	0.018	0.256
<i>KIFAP3</i>	1	169690461	170243882	2061	2.78	$2.74 \times 10^{-3}$	0.020	0.304
<i>ERLEC1</i>	2	53814068	54245956	1385	2.76	$2.86 \times 10^{-3}$	0.020	0.318
<i>LRPPRC</i>	2	43913363	44423144	2040	2.51	$6.05 \times 10^{-3}$	0.040	0.672

Examining the flanking region of 50 kb revealed eight of the nine genes replicating significant association in the UK Biobank study. The gene failing to replicate in this sample was *CNTF* (Table 7.5). This failure of the *CNTF* gene to replicate significant association in the UK Biobank study was repeated when 200 kb flanking regions were applied. With the exception of *KIFAP3*, all other 15 genes identified in the CREAM sample for the 200 kb flanking region successfully replicated in the UK Biobank study (Table 7.6).

Spearman's rank correlations between the CREAM and UK Biobank  $-\log_{10}$  p-values obtained from their respective gene-based analyses were moderate but non-zero at 0.337 ( $P = 3.62 \times 10^{-4}$ ), 0.456 ( $P = 7.58 \times 10^{-7}$ ) and 0.511 ( $P = 1.87 \times 10^{-8}$ ) for the results using 0 kb, 50 kb and 200kb flanking regions respectively. Figure 7.1 displays a comparison of  $-\log_{10}$  p-values from the MAGMA gene-based analyses for the CREAM and UK Biobank summary statistics at these three flanking regions.

**Table 7.4: Testing for replication of the 4 candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) in the UK Biobank study.** Start and stop positions refer to the respective gene transcription start and stop sites (no flanking region included); nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>SEPT4</i>	17	56597611	56618179	36	6.37	$9.22 \times 10^{-11}$	$3.69 \times 10^{-10}$	$3.69 \times 10^{-10}$
<i>LBH</i>	2	30454397	30482899	76	5.28	$6.35 \times 10^{-8}$	$1.27 \times 10^{-7}$	$2.54 \times 10^{-7}$
<i>DDIT4</i>	10	74033677	74035797	4	5.08	$1.88 \times 10^{-7}$	$2.50 \times 10^{-7}$	$7.50 \times 10^{-7}$
<i>CLU</i>	8	27454434	27472328	30	3.47	$2.56 \times 10^{-4}$	$2.56 \times 10^{-4}$	$1.03 \times 10^{-3}$

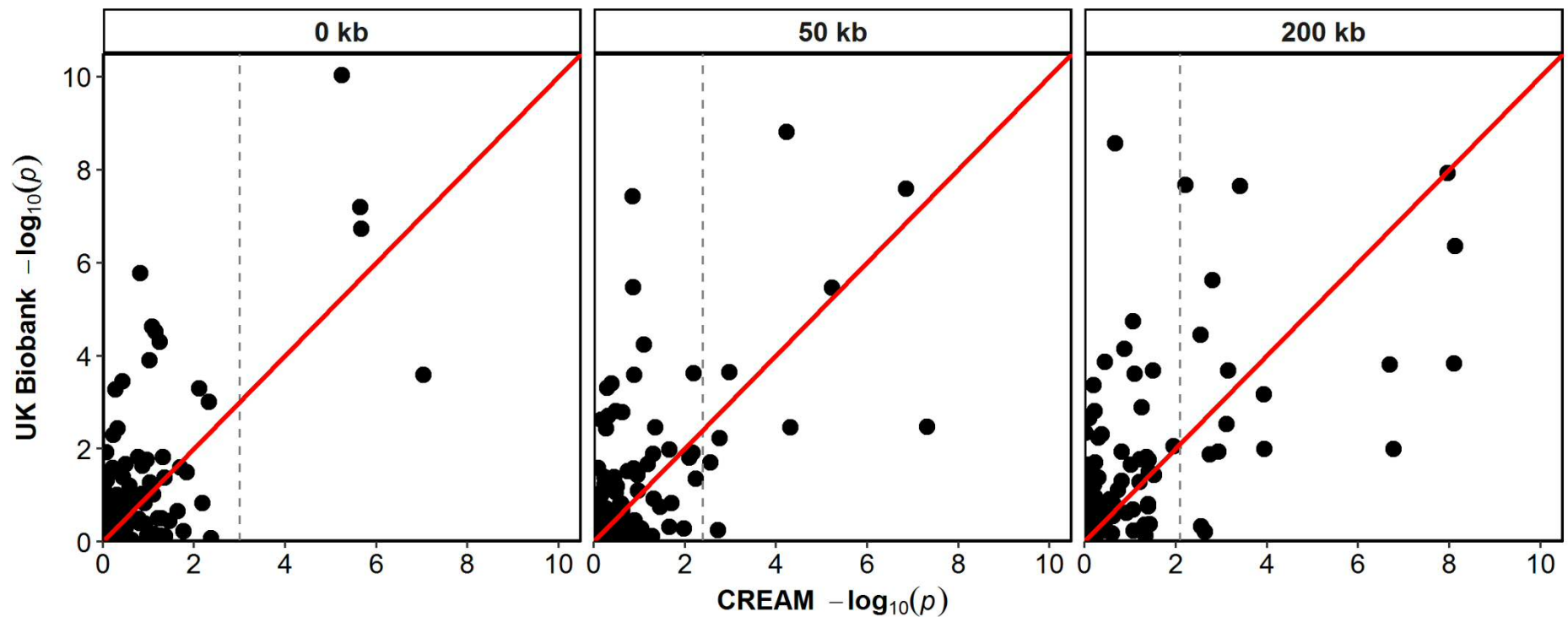
**Table 7.5: Testing for replication of the 9 candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) with the inclusion of 50 kb flanking regions in the UK Biobank study.** nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>SEPT4</i>	17	56547611	56668179	181	5.93	$1.56 \times 10^{-9}$	$1.40 \times 10^{-8}$	$1.40 \times 10^{-8}$
<i>DDIT4</i>	10	73983677	74085797	157	5.44	$2.60 \times 10^{-8}$	$1.17 \times 10^{-7}$	$2.34 \times 10^{-7}$
<i>LBH</i>	2	30404397	30532899	227	4.49	$3.49 \times 10^{-6}$	$1.05 \times 10^{-5}$	$3.14 \times 10^{-5}$
<i>ERLEC1</i>	2	53964068	54095956	340	3.51	$2.23 \times 10^{-4}$	$5.02 \times 10^{-4}$	$2.01 \times 10^{-3}$
<i>TSR3</i>	16	1349241	1451873	275	2.71	$3.34 \times 10^{-3}$	$5.19 \times 10^{-3}$	0.030
<i>CLU</i>	8	27404434	27522328	246	2.70	$3.46 \times 10^{-3}$	$5.19 \times 10^{-3}$	0.031
<i>SNW1</i>	14	78133942	78277542	531	2.52	$5.93 \times 10^{-3}$	$7.62 \times 10^{-3}$	0.053
<i>SPON1</i>	11	13934184	14339679	878	2.05	0.020	0.023	0.180
<i>CNTF</i>	11	58340146	58443206	256	-0.17	0.568	0.568	1

**Table 7.6: Testing for replication of the 17 candidate myopia genes demonstrating strongest association from the MAGMA gene-based association test (FDR < 0.05) with the inclusion of 200 kb flanking regions in the UK Biobank study.** nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from *Z-Statistic* under the assumption of a normally distributed model; FDR = false discovery rate.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>CNIH1</i>	14	54690279	55108322	618	5.58	$1.18 \times 10^{-8}$	$1.27 \times 10^{-7}$	$2.01 \times 10^{-7}$
<i>LRPPRC</i>	2	43913363	44423144	1408	5.48	$2.14 \times 10^{-8}$	$1.27 \times 10^{-7}$	$3.64 \times 10^{-7}$
<i>SEPT4</i>	17	56397611	56818179	658	5.47	$2.24 \times 10^{-8}$	$1.27 \times 10^{-7}$	$3.81 \times 10^{-7}$
<i>DDIT4</i>	10	73833677	74235797	642	4.92	$4.36 \times 10^{-7}$	$1.85 \times 10^{-6}$	$7.41 \times 10^{-6}$
<i>DGKE</i>	17	54711460	55146036	771	4.57	$2.40 \times 10^{-6}$	$8.18 \times 10^{-6}$	$4.09 \times 10^{-5}$
<i>ERLEC1</i>	2	53814068	54245956	1077	3.97	$3.57 \times 10^{-5}$	$1.01 \times 10^{-4}$	$6.07 \times 10^{-4}$
<i>ANKRD28</i>	3	15508743	16101053	1278	3.62	$1.45 \times 10^{-4}$	$3.26 \times 10^{-4}$	$2.47 \times 10^{-3}$
<i>HNRNPA3</i>	2	1.78E+08	1.78E+08	762	3.61	$1.54 \times 10^{-4}$	$3.26 \times 10^{-4}$	$2.61 \times 10^{-3}$
<i>LBH</i>	2	30254397	30682899	858	3.53	$2.06 \times 10^{-4}$	$3.89 \times 10^{-4}$	$3.50 \times 10^{-3}$
<i>SNW1</i>	14	77983942	78427542	1588	3.21	$6.71 \times 10^{-4}$	$1.14 \times 10^{-3}$	0.011
<i>RAB18</i>	10	27593103	28031166	1134	2.75	$2.96 \times 10^{-3}$	$4.58 \times 10^{-3}$	0.050
<i>TSR3</i>	16	1199241	1601873	1110	2.32	0.010	0.013	0.172
<i>E2F4</i>	16	67026068	67432821	465	2.32	0.010	0.013	0.174
<i>SPON1</i>	11	13784184	14489679	1453	2.27	0.012	0.014	0.196
<i>CLU</i>	8	27254434	27672328	1098	2.22	0.013	0.015	0.224
<i>KIFAP3</i>	1	1.7E+08	1.7E+08	1616	0.09	0.466	0.495	1
<i>CNTF</i>	11	58190146	58593206	984	-0.30	0.617	0.617	1

**Figure 7.1: Comparison of  $-\log_{10}$  p-values from the gene-based analyses of the 111 candidate myopia genes using GWAS summary statistics from the CREAM consortium and UK Biobank with flanking regions of 0 kb, 50 kb and 200 kb.** X-axes show expected negative  $\log_{10}$  p-values from analyses of CREAM consortium summary statistics. Y-axes show negative  $\log_{10}$  p-values from analyses of UK Biobank summary statistics. Red line: line of unity (CREAM  $-\log_{10}$  p-value = UK Biobank  $-\log_{10}$  p-value). Dashed lines denote separation between genes that demonstrated significant association using the CREAM consortium data (FDR < 0.05; right of lines) and all other candidate genes tested (left of lines). Genes tested for replication in the UK Biobank study are located to the right of the dashed lines.





### **7.3.3 Exploration of wider flanking regions**

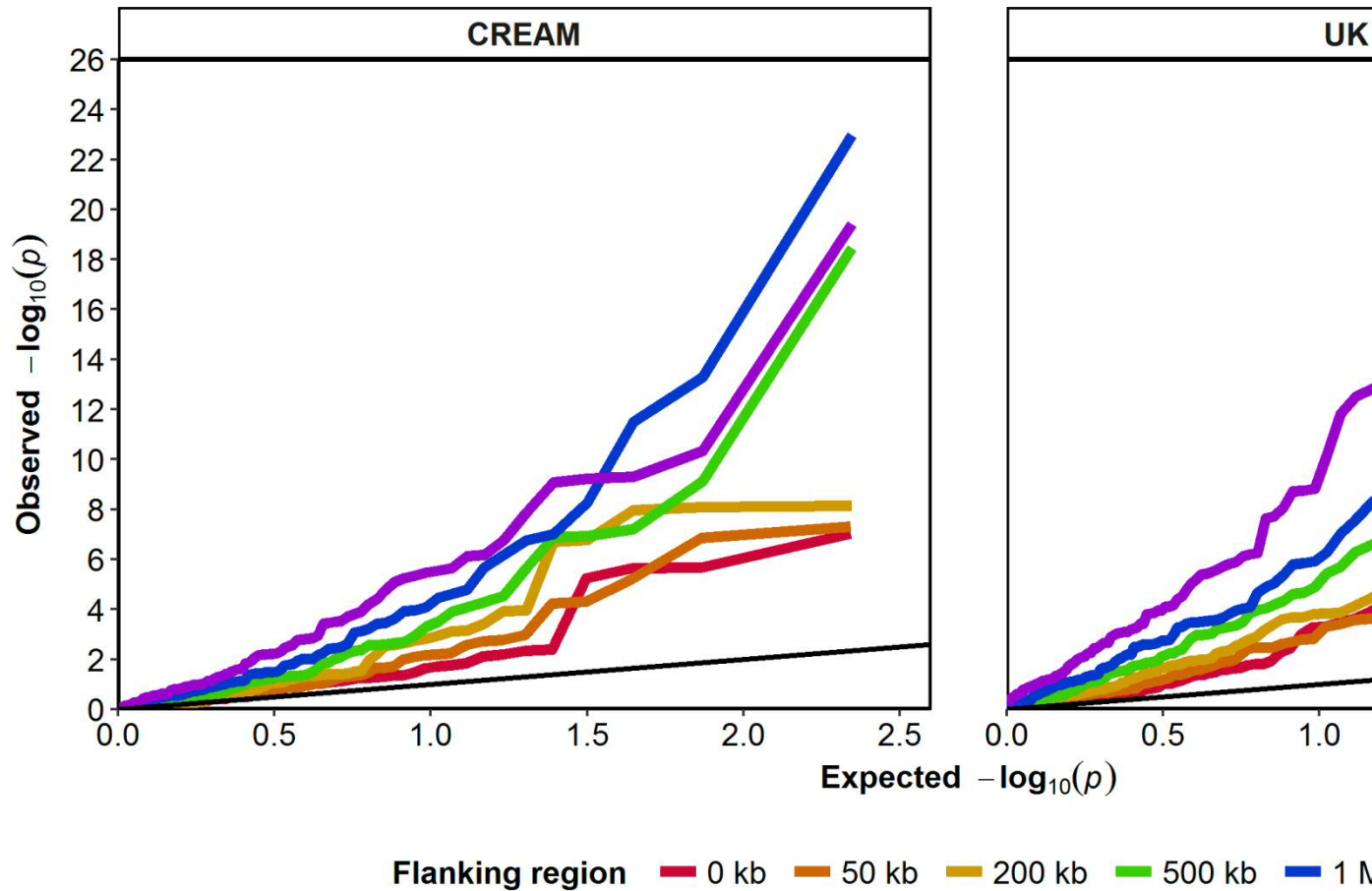
As an exploratory analysis of wider flanking regions that could harbour variants influencing gene enhancers and repressors, wider flanking regions of 500 kb, 1 Mb and 2 Mb were applied, initially in the gene-based analyses of the CREAM refractive error data. These increased flanking regions outside of the gene transcription start and stop sites resulted in 23, 30 and 44 of the 111 candidate genes demonstrating association with a FDR < 0.05 respectively (Appendix E). Overall, 52 of the 111 candidate genes demonstrated significant association (FDR < 0.05) with refractive error when tested with any one of the applied flanking regions. Of these genes, *DDIT4* and *SEPT4* demonstrated significant association with refractive error using all six of the flanking regions. A table summarising gene-based test results for the CREAM refractive data at all 111 candidate genes with all iterations of flanking regions can be found in Appendix E.

Replication analysis performed using data from the UK Biobank study revealed 17 out of 23 genes, 17 out of 30 genes and 21 out of 44 genes continued to demonstrate significant association when the flanking regions of 500 kb, 1 Mb and 2 Mb respectively were applied.

Investigation of all 111 candidate myopia genes in UK Biobank identified 15, 20, 30, 44, 55 and 73 genes demonstrating significant association (FDR < 0.05) with spherical equivalent for the respective flanking regions tested (Appendix E). Overall, 83 out of the 111 genes tested demonstrated significant association when using any one of the applied flanking regions. Thirteen genes demonstrated significant

association with all six of the applied flanking regions when using data from the UK Biobank. Figure 7.2 shows quantile-quantile plots of the results from the respective gene-based analyses with each line denoting a specific flanking region.

**Figure 7.2: Quantile-quantile plots of  $-\log_{10} p$ -values from the gene-based analyses of the 111 candidate r summary statistics from the CREAM consortium and UK Biobank. Y-axes show observed negative  $\log_{10} p$ -v negative  $\log_{10} p$ -values according to the null hypothesis of no genetic association. Black line: line of unity (ob**



## 7.4 Discussion

Gene-based tests performed using human refractive error GWAS summary statistics and restricted to 111 candidate myopia genes suggested positive association of several of these genes, as shown by replication in an independent sample.

In this investigation, candidate genes were examined with six different flanking regions appended to their respective transcription start and stop sites (0 kb, 50 kb, 200 kb, 500 kb, 1 Mb and 2 Mb). Four of the candidate myopia genes demonstrated significant association when tested with no flanking region appended (0 kb). It could be suggested that the genetic variants influencing these four genes are likely to reside within the gene itself, with the causal variant(s) possibly non-synonymous variants or splice variants which alter the resultant codon sequence and ultimately the end gene product (Hunt et al., 2009). By including flanking regions to the gene transcription start and stop sites, variants that may influence gene regulatory elements can be captured, alongside variants that may be in LD with an untyped causal variant. Often, these variants are located within a few kilobases of the transcription start / stop site; however, gene enhancers and repressors have been known to be identified hundreds of kilobases away from their target gene and even further beyond (Elkon and Agami, 2017). A complication of using wider flanking regions however is that the increased number of variants included in the gene-based test can result in increased statistical “noise” from the uninformative variants that are included, thus making it harder for genes influenced by local variants to continue to demonstrate significant association. Since refractive error is a polygenic trait, with 167 independent loci demonstrating significant association in

the recent GWAS for this trait conducted by the CREAM consortium (Tedja et al., 2018), it is highly likely that the candidate myopia genes will be under the influence of a combination of effects, with some genes influenced by variants in their immediate vicinity and other genes influenced by more distal variants.

Not all of the candidate myopia genes demonstrated significant enrichment of refractive error-associated variants in this investigation. A potential explanation for this may be that these genes could not demonstrate sufficient statistical support due to other, non-informative variants included in the gene-based test masking the effects of the true influencing variants. Alternatively, the biological process resulting in the altered gene expression in the myopic eyes may be due to a more complex mechanism. For example, rather than a variant directly acting on the functional properties of a specific gene, the variant may exert its effects via an intermediary interacting gene. Alternatively, there may be no naturally-occurring variants in some genes with important roles in myopia. In addition, some candidate genes from the Tkatchenko et al. (2006) study could have been false-positives with no true relationship to myopia. Finally, some of the candidate genes may play an important role in emmetropisation in the primate myopia model, but their human orthologues may not harbour refractive error associated genetic variants.

Additionally, some of the candidate myopia genes demonstrating association with refractive error in the MAGMA gene-based analysis may have been false positives. This is because the flanking regions applied to the gene regions may have captured variants that do indeed influence gene expression, but expression of an alternative,

nearby gene rather than the specified candidate gene. This reasoning can also be applied to the moderate correlation and different numbers of significant associations identified between the discovery and replication samples since summary statistics were used for two slightly different traits. The summary statistics used from the analysis performed by Tedja et al. (2018) refer to a standardised refractive error trait, whereas the summary statistics from the UK Biobank data refer to the raw spherical equivalent trait. Additional differences in the gene-based associations between the discovery and replication samples include the number of variants tested for association in their respective GWAS, the origin of the samples and the sample size itself, and the power of the respective studies. All of these differences influence the respective GWAS and in turn their respective gene-based analyses.

A follow-up investigation by Tkatchenko et al. (2015) of one of the 111 candidate genes explored here, *APLP2*, identified that this gene (and the flanking 100 kb region) was indeed enriched with variants associated with refractive error in cohorts of children and adults. In the investigation described in this chapter, however, *APLP2* failed to demonstrate enrichment with refractive error associated variants using MAGMA, except when applying a 1-2 Mb flanking region to the gene (Appendix E). It should be noted that whilst Tkatchenko et al. (2015) included a 100 kb flanking region to the gene, a flanking region of 50 kb would be sufficient to capture the effects of the cluster of variants at the 5' promoter region they hypothesised to influence the expression of this gene in the retina.

It is important to note that the candidate genes investigated here were prioritised from an experiment using a form deprivation model of myopia. As the candidate genes appearing on such a list are dependent on the biological mechanisms influenced by the intervention, the prioritised genes may differ when alternative methods of inducing myopic changes in animal models are employed. Such methods include using translucent diffusers to induce form deprivation whilst maintaining some – albeit reduced – retinal illumination, using defocusing lenses which maintain normal levels of retinal illumination and are more realistic of the human experience, or using changes to the temporal frequency or chromatic content of the illumination (Stone and Khurana, 2010).

Furthermore, the primate myopia model is an example of environmentally induced myopia. Conversely, human refractive error GWAS summary statistics are from examination of the association of genetic polymorphisms with refractive error. As a result, these two methods are not examining like-for-like biological mechanisms; however there is a considerable degree of overlap since both methods capture some effects of gene-gene and gene-environment interactions.

Gene expression is a dynamic process, responding to changes in environment and can be variable depending on the tissue, time of day / year, and over the course one's lifespan (Chowers et al., 2003; Lopez-Maury, Marguerat and Bahler, 2008). Considering this complex nature of gene expression, the ideal experimental scenario would involve gaining access to the tissue of interest (i.e. human retina) at the time when trait changes are likely to be greatest (i.e. in adolescence) and

investigate the association between expression of genes in these samples and the trait of interest (i.e. refractive error). However, in the real world, obtaining such biological samples would involve highly invasive procedures with the potential to cause long-term harm to individuals taking part, thus rendering such studies unfeasible. Animal models therefore provide a valuable opportunity to address questions such as those posed here regarding gene expression despite the non-natural method of trait simulation.

With the constant growth of open access gene expression databases such as the Genotype-Tissue Expression (GTEx) portal (GTEx Consortium, 2017), there are increasing numbers of human samples and tissue types with gene expression data available and summary statistics from expression quantitative trait loci (eQTL) studies. An ideal follow-up investigation would be to use summary statistics from a retina eQTL study. Here, variants associated with expression of each of the candidate myopia genes identified by Tkatchenko et al. (2006) would be compared against variants associated with refractive error. Variants demonstrating significant association with both retinal gene expression and refractive error would thus be deemed likely candidates underlying the causal mechanism. Such causal gene(s) would serve as a potential targets for future genetic intervention methods. This approach, also known as summary data-based Mendelian randomisation (SMR) analysis, has been proposed by Zhu et al. (2016). They identified 104 genes as being likely candidate genes influencing the traits of height, body mass index (BMI), waist-to-hip ratio adjusted by BMI, rheumatoid arthritis and schizophrenia. They also noted that the majority of these candidate genes were not the nearest gene to the



GWAS associated variant. However, the lack of an eQTL database for whole retina, let alone specific retinal cell types, currently limits the adaptation of this approach for refractive error traits.

## **Chapter 8 Overall Conclusions**

## Chapter 8 Overall Conclusions

The investigations conducted in the preceding chapters have identified novel candidate susceptibility loci for the refractive error traits of spherical equivalent, corneal astigmatism and refractive astigmatism. In addition, utilising recently developed techniques to estimate the proportion of variance attributed to genetic effects has suggested corneal and refractive astigmatism may be considerably less heritable than previously thought, or under greater influence of multiple rare genetic variants.

An important point to remember stemming from the results of the single marker tests for astigmatism in the UK Biobank cohort is the potential for shared genetic susceptibility to different refractive traits. Much of the focus on myopia research is concentrated on *retinal* changes in gene expression, as highlighted by animal models of myopia, and the assumption that the majority of myopic changes occur towards the posterior pole (Wallman and Winawer, 2004; Li and Zhang, 2017; Wojciechowski and Cheng, 2018). One possible argument for the shared genetic susceptibility of astigmatism and myopia is through the concept of visual feedback driving changes at the posterior retina and sclera (Wallman and Winawer, 2004). However, if genetic variants influenced biological mechanisms that also regulate *corneal (or lenticular)* shape or structural properties, they may also cause altered image focusing at the retina. In turn, this blurred image would feedback into the visually guided ocular growth mechanism, resulting in physical changes at the level of the retina and sclera.

Despite identification of novel candidate susceptibility loci, the functional mechanisms by which these loci influence the development of their respective traits remain uncertain. As highlighted in the previous chapter, attempts at bridging the gap between genetic variants, genes and biological functions could be undertaken through methods such as summary data–based Mendelian randomisation (SMR) analysis. At present, it is not yet feasible to run a retinal eQTL study or to use summary statistics from a pre-existing study; however tissue specific gene expression banks are constantly growing. Donations are generally from older persons where typical refractive error development had long ceased, yet overlapping results of a retinal eQTL study from this group with results of eQTL studies from alternative tissues such as blood or skin in adolescents may still be helpful at determining the causal genes influencing refractive errors.

Complementary to this, a multi-omics approach could be performed whereby changes in the epigenome, transcriptome, proteome and metabolome of ocular cells are investigated simultaneously in order to identify the biological mechanisms that are altered in refractive error development (Ritchie et al., 2015). However, the dynamic nature of such processes (Feinberg, 2007) – and the difficulty in obtaining appropriate samples from adolescent humans – currently means that such approaches are not feasible. Furthermore, the knowledge gained from equivalent studies using alternative cell types or samples from older aged groups may be limited.

The series of investigations conducted throughout this thesis have not considered the role of environmental factors. As stated in *Section 1.1.3*, environmental factors such as higher educational attainment and the time spent outdoors are known to influence myopia development. Gene-environment interactions, with education level as the environment feature investigated, have been identified with respect to spherical equivalent refractive error (Fan et al., 2016), yet the biological mechanisms underlying how such environmental factors influence myopia development are still under investigation. Such interactions are not easily identifiable as shown by the limited replication from gene-environment interaction GWAS of other complex traits (McAllister et al., 2017); however, interactions and their effects may be more apparent when investigating the downstream biological processes such as from a multi-omics investigation.

In addition, these investigations have utilised data from participant samples with a wide age range. Refractive errors tend to vary throughout life. These changes start from the emmetropisation process during childhood, through to the development of myopic changes during adolescence into early adulthood, before stabilising in later life. For instance, the X-chromosome wide association study for refractive error (*Chapter 3*) utilised participants from a birth-cohort study. Whilst these participating children were all of similar ages (~15 years) at the time when refraction data was obtained, for many of them, their refractive development had not fully stabilised. As a result, some children in the cohort would still be described as “progressing myopes”, and their spherical equivalent refractive error would continue to increase in the direction of myopia over the next few years. This may

have resulted in fewer genetic variants demonstrating association with refractive error due to the trait still developing. In the case of the analyses for corneal and refractive astigmatism (*Chapters 4-6*), these analyses utilised data from studies of participants with widely-differing ages. As demonstrated in the UK Biobank sample (*Section 6.3.1*) and other aging cohorts (Asano et al., 2005; Sanfilippo et al., 2015), there is a tendency for refractive astigmatism to increase in magnitude with increasing age from the age of 40 years, whereas, corneal astigmatism demonstrates relative stability in magnitude in adults until 70 years of age. Along with the more complex phenotype that is refractive astigmatism, this change in magnitude with age may in part explain the higher number of genetic variants demonstrating significant association for corneal astigmatism than for refractive astigmatism in the UK Biobank sample (*Chapter 6*). For all refractive error traits, analyses would ideally be performed using cohorts of ages where the investigated trait is stable. For example, analyses of spherical equivalent refractive error or refractive astigmatism may be best performed using cohorts of young adults in the age range of 25-40 years. On the other hand, analyses of corneal astigmatism magnitude may be performed using a wider age range, including children and adults up to 70 years of age. However, if the axis of corneal astigmatism is also to be considered in analyses, this age range should be restricted to a maximum age of 40 years, similar to that suggested for spherical equivalent and refractive astigmatism. However, employing such age restrictions on study cohorts can notably reduce the available sample size and therefore investigators have to balance the reduced statistical power of a smaller sample against the reduction in statistical noise for age-dependent traits. Some of these age-related influences can be mediated by

inclusion of age as an additional covariate, as has been performed in the analyses conducted in this series of investigations. However, there may still be some residual effects that have not been fully accounted for.

Ultimately, identifying the altered biological processes in refractive error development, whether they are altered gene transcription or altered expression of certain proteins, could serve as potential targets for therapeutic or preventative intervention strategies. The findings from the investigations undertaken in this thesis are only a starting point in guiding further research into the complex biological mechanisms underlying refractive error development.

## References



## References

- Aldahmesh M A, Khan A O, Alkuraya H, Adly N, Anazi S, Al-Saleh A A, Mohamed J Y et al. (2013) Mutations in LRPAP1 are associated with severe myopia in humans. *American Journal of Human Genetics* 93: 313-320.
- Allen N E, Sudlow C, Peakman T, Collins R, and Biobank U K (2014) UK biobank data: come and get it. *Science Translational Medicine* 6: 224ed224.
- Andrew T, Maniatis N, Carbonaro F, Liew S H, Lau W, Spector T D, and 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 Genetics* 4: e1000220.
- Asano K, Nomura H, Iwano M, Ando F, Niino N, Shimokata H, and Miyake Y (2005) Relationship between astigmatism and aging in middle-aged and elderly Japanese. *Japanese Journal of Ophthalmology* 49: 127-133.
- Aulchenko Y S, Struchalin M V, and van Duijn C M (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 11: 134.
- Bailey-Wilson J E, and Wilson A F (2011) Linkage analysis in the next-generation sequencing era. *Human Heredity* 72: 228-236.
- Baird P N, Schache M, and Dirani M (2010) The GENes in Myopia (GEM) study in understanding the aetiology of refractive errors. *Progress in Retinal and Eye Research* 29: 520-542.
- Balding D J (2006) A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* 7: 781-791.
- Barbieri A M, Lupo G, Bulfone A, Andreazzoli M, Mariani M, Fougousse F, Consalez G G et al. (1999) A homeobox gene, *vax2*, controls the patterning of the eye dorsoventral axis. *Proceedings of the National Academy of Sciences of the United States of America* 96: 10729-10734.
- Benjamini Y, and Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 57: 289-300.

Bennett A G, and Rabbetts R B (1989) *Clinical Visual Optics*. Second Edition ed. London: Butterworths.

Biino G, Balduini C L, Casula L, Cavallo P, Vaccargiu S, Parracciani D, Serra D et al. (2010) Analysis of 12,517 inhabitants of a Sardinian geographic isolate reveals that predispositions to thrombocytopenia and thrombocytosis are inherited traits. *Haematologica* 96: 96-101.

Bodmer W, and Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics* 40: 695-701.

Boehnke M (1994) Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *American Journal of Human Genetics* 55: 379-390.

Boomsma D, Busjahn A, and Peltonen L (2002) Classical twin studies and beyond. *Nature Reviews Genetics* 3: 872-882.

Boyd A, Golding J, Macleod J, Lawlor D A, Fraser A, Henderson J, Molloy L et al. (2013) Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* 42: 111-127.

Brodie A, Azaria J R, and Ofran Y (2016) How far from the SNP may the causative genes be? *Nucleic Acids Research* 44: 6046-6054.

Browning S R, and Browning B L (2011) Population structure can inflate SNP-based heritability estimates. *American Journal of Human Genetics* 89: 191-193; author reply 193-195.

Bulik-Sullivan B, Finucane H K, Anttila V, Gusev A, Day F R, Loh P R, ReproGen C et al. (2015a) An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 47: 1236-1241.

Bulik-Sullivan B K, Loh P R, Finucane H K, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N et al. (2015b) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* 47: 291-295.

Bush W S, and Moore J H (2012) Chapter 11: Genome-wide association studies. *PLoS Computational Biology* 8: e1002822.

Buzkova P (2013) Linear regression in genetic association studies. *PLoS ONE* 8: e56976.

Bycroft C, Freeman C, Petkova D, Band G, Elliott L T, Sharp K, Motyer A et al. (2017) Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*.

Cantor R M, Lange K, and Sinsheimer J S (2010) Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *American Journal of Human Genetics* 86: 6-22.

Carrel L, and Willard H F (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434: 400-404.

Chang C C, Chow C C, Tellier L C, Vattikuti S, Purcell S M, and Lee J J (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.

Charrin S, Jouannet S, Boucheix C, and Rubinstein E (2014) Tetraspanins at a glance. *Journal of Cell Science* 127: 3641-3648.

Chen C Y, Scurrah K J, Stankovich J, Garoufalos P, Dirani M, Pertile K K, Richardson A J et al. (2007a) Heritability and shared environment estimates for myopia and associated ocular biometric traits: the Genes in Myopia (GEM) family study. *Human Genetics* 121: 511-520.

Chen C Y, Stankovich J, Scurrah K J, Garoufalos P, Dirani M, Pertile K K, Richardson A J et al. (2007b) Linkage replication of the MYP12 locus in common myopia. *Investigative Ophthalmology and Visual Science* 48: 4433-4439.

Chen H, Wang C, Conomos M P, Stilp A M, Li Z, Sofer T, Szpiro A A et al. (2016) Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *American Journal of Human Genetics* 98: 653-666.

Cheng C Y, Schache M, Ikram M K, Young T L, Guggenheim J A, Vitart V, MacGregor S et al. (2013) Nine loci for ocular axial length identified through genome-wide

association studies, including shared loci with refractive error. *American Journal of Human Genetics* 93: 264-277.

Chia A, Chua W H, Wen L, Fong A, Goon Y Y, and Tan D (2014) Atropine for the treatment of childhood myopia: changes after stopping atropine 0.01%, 0.1% and 0.5%. *American Journal of Ophthalmology* 157: 451-457 e451.

Cho P, and Cheung S W (2012) Retardation of myopia in Orthokeratology (ROMIO) study: a 2-year randomized clinical trial. *Investigative Ophthalmology and Visual Science* 53: 7077-7085.

Cho P, and Cheung S W (2017) Discontinuation of orthokeratology on eyeball elongation (DOEE). *Cont Lens Anterior Eye* 40: 82-87.

Chowers I, Liu D, Farkas R H, Gunatilaka T L, Hackam A S, Bernstein S L, Campochiaro P A et al. (2003) Gene expression variation in the adult human retina. *Human Molecular Genetics* 12: 2881-2893.

Ciner E, Wojciechowski R, Ibay G, Bailey-Wilson J E, and Stambolian D (2008) Genomewide scan of ocular refraction in African-American families shows significant linkage to chromosome 7p15. *Genetic Epidemiology* 32: 454-463.

Clarke G M, Anderson C A, Pettersson F H, Cardon L R, Morris A P, and Zondervan K T (2011) Basic statistical analysis in genetic case-control studies. *Nature Protocols* 6: 121-133.

Clementi M, Angi M, Forabosco P, Di Gianantonio E, and Tenconi R (1998) Inheritance of astigmatism: evidence for a major autosomal dominant locus. *American Journal of Human Genetics* 63: 825-830.

Cohen J (1992) Statistical power analysis. *Current Directions in Psychological Science* 1: 98-101.

Collier Wakefield O, Annoh R, and Nanavaty M A (2016) Relationship between age, corneal astigmatism, and ocular dimensions with reference to astigmatism in eyes undergoing routine cataract surgery. *Eye (London, England)* 30: 562-569.

Corradin O, Cohen A J, Luppino J M, Bayles I M, Schumacher F R, and Scacheri P C (2016) Modeling disease risk through analysis of physical interactions between

genetic variants within chromatin regulatory circuitry. *Nature Genetics* 48: 1313-1320.

Corvin A, Craddock N, and Sullivan P F (2010) Genome-wide association studies: a primer. *Psychological Medicine* 40: 1063-1077.

Craddock N, O'Donovan M C, and Owen M J (2008) Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. *Molecular Psychiatry* 13: 649-653.

Cuellar-Partida G, Lu Y, Kho P F, Hewitt A W, Wichmann H E, Yazar S, Stambolian D et al. (2016) Assessing the Genetic Predisposition of Education on Myopia: A Mendelian Randomization Study. *Genetic Epidemiology* 40: 66-72.

Cui J, Hao C, Zhang W, Shao J, Zhang N, Zhang G, and Liu S (2015) Identical expression profiling of human and murine TIPE3 protein reveals links to its functions. *Journal of Histochemistry and Cytochemistry* 63: 206-216.

Dandine-Roulland C, Bellenguez C, Debette S, Amouyel P, Genin E, and Perdry H (2016) Accuracy of heritability estimations in presence of hidden population stratification. *Scientific Reports* 6: 26471.

Davidson A E, Cheong S S, Hysi P G, Venturini C, Plagnol V, Ruddle J B, Ali H et al. (2014) Association of *CHRD1* mutations and variants with X-linked megalocornea, Neuhauser syndrome and central corneal thickness. *PLoS ONE* 9: e104163.

Davies G, Marioni R E, Liewald D C, Hill W D, Hagenaars S P, Harris S E, Ritchie S J et al. (2016) Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Molecular Psychiatry* 21: 758-767.

de Leeuw C A, Mooij J M, Heskes T, and Posthuma D (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology* 11: e1004219.

Derks E M, Dolan C V, and Boomsma D I (2006) A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Res Hum Genet* 9: 403-411.

DerSimonian R, and Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177-188.

Devlin B, and Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29: 311-322.

Devlin B, and Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997-1004.

Devlin B, Roeder K, and Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. *Theoretical Population Biology* 60: 155-166.

Dimas A S, Deutsch S, Stranger B E, Montgomery S B, Borel C, Attar-Cohen H, Ingle C et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325: 1246-1250.

Dirani M, Chamberlain M, Shekar S N, Islam A F, Garoufalos P, Chen C Y, Guymer R H et al. (2006) Heritability of refractive error and ocular biometrics: the Genes in Myopia (GEM) twin study. *Investigative Ophthalmology and Visual Science* 47: 4756-4761.

Dirani M, Chan Y H, Gazzard G, Hornbeak D M, Leo S W, Selvaraj P, Zhou B et al. (2010) Prevalence of refractive error in Singaporean Chinese children: the strabismus, amblyopia, and refractive error in young Singaporean Children (STARS) study. *Investigative Ophthalmology and Visual Science* 51: 1348-1355.

Dirani M, Islam A, Shekar S N, and Baird P N (2008) Dominant genetic effects on corneal astigmatism: the genes in myopia (GEM) twin study. *Investigative Ophthalmology and Visual Science* 49: 1339-1344.

Drabek K, van de Peppel J, Eijken M, and van Leeuwen J P (2011) GPM6B regulates osteoblast function and induction of mineralization by controlling cytoskeleton and matrix vesicle release. *Journal of Bone and Mineral Research* 26: 2045-2051.

Dudbridge F, and Gusnanto A (2008) Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology* 32: 227-234.

Eichler E E, Flint J, Gibson G, Kong A, Leal S M, Moore J H, and Nadeau J H (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics* 11: 446-450.

Elkon R, and Agami R (2017) Characterization of noncoding regulatory DNA in the human genome. *Nature Biotechnology* 35: 732-746.

Evangelou E, and Ioannidis J P (2013) Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics* 14: 379-389.

Falconer D S, and Mackay T F C (1996) *Introduction to Quantitative Genetics*. 4th ed. Harlow: Longman.

Fan Q, Barathi V A, Cheng C Y, Zhou X, Meguro A, Nakata I, Khor C C et al. (2012) Genetic variants on chromosome 1q41 influence ocular axial length and high myopia. *PLoS Genetics* 8: e1002753.

Fan Q, Verhoeven V J, Wojciechowski R, Barathi V A, Hysi P G, Guggenheim J A, Hohn R et al. (2016) Meta-analysis of gene-environment-wide association scans accounting for education level identifies additional loci for refractive error. *Nat Commun* 7: 11008.

Fan Q, Zhou X, Khor C C, Cheng C Y, Goh L K, Sim X, Tay W T et al. (2011) Genome-wide meta-analysis of five Asian cohorts identifies PDGFRA as a susceptibility locus for corneal astigmatism. *PLoS Genetics* 7: e1002402.

Farbrother J E, Kirov G, Owen M J, Pong-Wong R, Haley C S, and Guggenheim J A (2004) Linkage analysis of the genetic loci for high myopia on 18p, 12q, and 17q in 51 U.K. families. *Investigative Ophthalmology and Visual Science* 45: 2879-2885.

Feinberg A P (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447: 433-440.

Ferreira M A (2004) Linkage analysis: principles and methods for the analysis of human quantitative traits. *Twin Research* 7: 513-530.

Flaxman S R, Bourne R R A, Resnikoff S, Ackland P, Braithwaite T, Cicinelli M V, Das A et al. (2017) Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. *Lancet Glob Health* 5: e1221-e1234.

Flitcroft D I (2012) The complex interactions of retinal, optical and environmental factors in myopia aetiology. *Progress in Retinal and Eye Research* 31: 622-660.

Foong A W, Saw S M, Loo J L, Shen S, Loon S C, Rosman M, Aung T et al. (2007) Rationale and methodology for a population-based study of eye diseases in Malay

people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiology* 14: 25-35.

Foster P J, and Jiang Y (2014) Epidemiology of myopia. *Eye (London, England)* 28: 202-208.

Fotedar R, Rochtchina E, Morgan I, Wang J J, Mitchell P, and Rose K A (2007) Necessity of cycloplegia for assessing refractive error in 12-year-old children: a population-based study. *American Journal of Ophthalmology* 144: 307-309.

Frazer K A, Murray S S, Schork N J, and Topol E J (2009) Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* 10: 241-251.

French A N, Morgan I G, Mitchell P, and Rose K A (2013) Risk factors for incident myopia in Australian schoolchildren: the Sydney adolescent vascular and eye study. *Ophthalmology* 120: 2100-2108.

Fritsche L G, Igl W, Bailey J N, Grassmann F, Sengupta S, Bragg-Gresham J L, Burdon K P et al. (2016) A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nature Genetics* 48: 134-143.

Gauderman W J (2002) Sample size requirements for association studies of gene-gene interaction. *American Journal of Epidemiology* 155: 478-484.

Goddard M E, Lee S H, Yang J, Wray N R, and Visscher P M (2011) Response to Browning and Browning. *American Journal Of Human Genetics* 89: 193-195.

Goh L, and Yap V B (2009) Effects of normalization on quantitative traits in association test. *BMC Bioinformatics* 10: 415.

Goldstein D B, and Weale M E (2001) Population genomics: linkage disequilibrium holds the key. *Current Biology* 11: R576-579.

Grijbovski A M, Magnus P, Midelfart A, and Harris J R (2006) Epidemiology and heritability of astigmatism in Norwegian twins: an analysis of self-reported data. *Ophthalmic Epidemiology* 13: 245-252.



Grosvenor T, and Scott R (1993) Three-year changes in refraction and its components in youth-onset and early adult-onset myopia. *Optometry and Vision Science* 70: 677-683.

GTEx Consortium (2017) Genetic effects on gene expression across human tissues. *Nature* 550: 204.

Guggenheim J A, and Farbrother J E (2004) The association between spherical and cylindrical component powers. *Optometry and Vision Science* 81: 62-63.

Guggenheim J A, McMahon G, Kemp J P, Akhtar S, St Pourcain B, Northstone K, Ring S M et al. (2013a) A genome-wide association study for corneal curvature identifies the platelet-derived growth factor receptor alpha gene as a quantitative trait locus for eye size in white Europeans. *Molecular Vision* 19: 243-253.

Guggenheim J A, McMahon G, Northstone K, Mandel Y, Kaiserman I, Stone R A, Lin X et al. (2013b) Birth order and myopia. *Ophthalmic Epidemiology* 20: 375-384.

Guggenheim J A, Northstone K, McMahon G, Ness A R, Deere K, Mattocks C, Pourcain B S et al. (2012) Time outdoors and physical activity as predictors of incident myopia in childhood: a prospective cohort study. *Investigative Ophthalmology and Visual Science* 53: 2856-2865.

Guggenheim J A, St Pourcain B, McMahon G, Timpson N J, Evans D M, and Williams C (2015) Assumption-free estimation of the genetic contribution to refractive error across childhood. *Molecular Vision* 21: 621-632.

Guggenheim J A, Williams C, Northstone K, Howe L D, Tilling K, St Pourcain B, McMahon G et al. (2014) Does vitamin D mediate the protective effects of time outdoors on myopia? Findings from a prospective birth cohort. *Investigative Ophthalmology and Visual Science* 55: 8550-8558.

Guggenheim J A, Zayats T, Prashar A, and To C H (2008) Axes of astigmatism in fellow eyes show mirror rather than direct symmetry. *Ophthalmic and Physiological Optics* 28: 327-333.

Guo H, Jin X, Zhu T, Wang T, Tong P, Tian L, Peng Y et al. (2014) SLC39A5 mutations interfering with the BMP/TGF-beta pathway in non-syndromic high myopia. *Journal of Medical Genetics* 51: 518-525.

Guo H, Tong P, Liu Y, Xia L, Wang T, Tian Q, Li Y et al. (2015) Mutations of P4HA2 encoding prolyl 4-hydroxylase 2 are associated with nonsyndromic high myopia. *Genetics in Medicine* 17: 300-306.

Guo L, Yang J, Mai J, Du X, Guo Y, Li P, Yue Y et al. (2016) Prevalence and associated factors of myopia among primary and middle school-aged students: a school-based study in Guangzhou. *Eye (London, England)* 30: 796-804.

Guo X, Xiao X, Li S, Wang P, Jia X, and Zhang Q (2010) Nonsyndromic high myopia in a Chinese family mapped to MYP1: linkage confirmation and phenotypic characterization. *Archives of Ophthalmology* 128: 1473-1479.

Guo Y, and Jamison D C (2005) The distribution of SNPs in human gene regulatory regions. *BMC Genomics* 6: 140.

Gwiazda J (2009) Treatment options for myopia. *Optometry and Vision Science* 86: 624-628.

Hammond C J, Andrew T, Mak Y T, and 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. *American Journal of Human Genetics* 75: 294-304.

Hammond C J, Snieder H, Gilbert C E, and Spector T D (2001) Genes and environment in refractive error: the twin eye study. *Investigative Ophthalmology and Visual Science* 42: 1232-1236.

Han S, Chen P, Fan Q, Khor C C, Sim X, Tay W T, Ong R T et al. (2011) Association of variants in FRAP1 and PDGFRA with corneal curvature in Asian populations from Singapore. *Human Molecular Genetics* 20: 3693-3698.

Harvey E M (2009) Development and treatment of astigmatism-related amblyopia. *Optometry and Vision Science* 86: 634-639.

He M, Huang W, Zheng Y, Huang L, and Ellwein L B (2007) Refractive error and visual impairment in school children in rural southern China. *Ophthalmology* 114: 374-382.

He M, Xiang F, Zeng Y, Mai J, Chen Q, Zhang J, Smith W et al. (2015) Effect of Time Spent Outdoors at School on the Development of Myopia Among Children in China: A Randomized Clinical Trial. *JAMA* 314: 1142-1148.

He M, Zeng J, Liu Y, Xu J, Pokharel G P, and Ellwein L B (2004) Refractive error and visual impairment in urban children in southern china. *Investigative Ophthalmology and Visual Science* 45: 793-799.

He M, Zheng Y, and Xiang F (2009) Prevalence of myopia in urban and rural children in mainland China. *Optometry and Vision Science* 86: 40-44.

Hettema J M, Neale M C, and Kendler K S (1995) Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. *Behavior Genetics* 25: 327-335.

Hewitt K J, Agarwal R, and Morin P J (2006) The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer* 6: 186.

Hill W G, Goddard M E, and Visscher P M (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* 4: e1000008.

Hinrichs A L, Larkin E K, and Suarez B K (2009) Population stratification and patterns of linkage disequilibrium. *Genetic Epidemiology* 33 Suppl 1: S88-92.

Hirschhorn J N, and Daly M J (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95-108.

Hofman A, Brusselle G G, Darwish Murad S, van Duijn C M, Franco O H, Goedegebure A, Ikram M A et al. (2015) The Rotterdam Study: 2016 objectives and design update. *European Journal of Epidemiology* 30: 661-708.

Hoggart C J, Clark T G, De Iorio M, Whittaker J C, and Balding D J (2008) Genome-wide significance for dense SNP and resequencing data. *Genetic Epidemiology* 32: 179-185.

Hohn R, Kottler U, Peto T, Blettner M, Munzel T, Blankenberg S, Lackner K J et al. (2015) The ophthalmic branch of the Gutenberg Health Study: study design, cohort profile and self-reported diseases. *PLoS ONE* 10: e0120476.

Holden B A, Fricke T R, Wilson D A, Jong M, Naidoo K S, Sankaridurg P, Wong T Y et al. (2016) Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. *Ophthalmology* 123: 1036-1042.

Hong E P, and Park J W (2012) Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics* 10: 117-122.

Hormozdiari F, Kichaev G, Yang W Y, Pasaniuc B, and Eskin E (2015) Identification of causal genes for complex traits. *Bioinformatics* 31: i206-213.

Howie B, Marchini J, and Stephens M (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1: 457-470.

Howie B N, Donnelly P, and Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics* 5: e1000529.

Huang J, Wen D, Wang Q, McAlinden C, Flitcroft I, Chen H, Saw S M et al. (2016) Efficacy Comparison of 16 Interventions for Myopia Control in Children: A Network Meta-analysis. *Ophthalmology* 123: 697-708.

Huang W, and Mackay T F (2016) The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLoS Genetics* 12: e1006421.

Hunt R, Sauna Z E, Ambudkar S V, Gottesman M M, and Kimchi-Sarfaty C (2009) Silent (synonymous) SNPs: should we care about them? *Methods in Molecular Biology* 578: 23-39.

Huynh S C, Kifley A, Rose K A, Morgan I, Heller G Z, and Mitchell P (2006) Astigmatism and its components in 6-year-old children. *Investigative Ophthalmology and Visual Science* 47: 55-64.

Huynh S C, Kifley A, Rose K A, Morgan I G, and Mitchell P (2007) Astigmatism in 12-year-old Australian children: comparisons with a 6-year-old population. *Investigative Ophthalmology and Visual Science* 48: 73-82.

Hysi P G, Mangino M, Nag A, Williams K M, and Hammond C J (2014) High Genomic Coverage through NGS Increases Refractive Error Phenotypic Variance Explained by Genes. *Investigative Ophthalmology & Visual Science* 55: 4512-4512.

Hysi P G, Young T L, Mackey D A, Andrew T, Fernandez-Medarde A, Solouki A M, Hewitt A W et al. (2010) A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. *Nature Genetics* 42: 902-905.

Inagaki E, Hatou S, Yoshida S, Miyashita H, Tsubota K, and Shimmura S (2013) Expression and distribution of claudin subtypes in human corneal endothelium. *Investigative Ophthalmology and Visual Science* 54: 7258-7265.

International HapMap C, Altshuler D M, Gibbs R A, Peltonen L, Altshuler D M, Gibbs R A, Peltonen L et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52-58.

Ip J M, Saw S M, Rose K A, Morgan I G, Kifley A, Wang J J, and Mitchell P (2008) Role of near work in myopia: findings in a sample of Australian school children. *Investigative Ophthalmology and Visual Science* 49: 2903-2910.

Jaddoe V W, Mackenbach J P, Moll H A, Steegers E A, Tiemeier H, Verhulst F C, Witteman J C et al. (2006) The Generation R Study: Design and cohort profile. *European Journal of Epidemiology* 21: 475-484.

Jiang D, Li J, Xiao X, Li S, Jia X, Sun W, Guo X et al. (2014) Detection of mutations in LRPAP1, CTSH, LEPREL1, ZNF644, SLC39A5, and SCO2 in 298 families with early-onset high myopia by exome sequencing. *Investigative Ophthalmology and Visual Science* 56: 339-345.

Johannes F, Colot V, and Jansen R C (2008) Epigenome dynamics: a quantitative genetics perspective. *Nature Reviews Genetics* 9: 883-890.

Jung S K, Lee J H, Kakizaki H, and Jee D (2012) Prevalence of myopia and its association with body stature and educational level in 19-year-old male conscripts in seoul, South Korea. *Investigative Ophthalmology and Visual Science* 53: 5579-5583.

Kang H M, Sul J H, Service S K, Zaitlen N A, Kong S Y, Freimer N B, Sabatti C et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42: 348-354.

Kaprio J, and Koskenvuo M (2002) Genetic and environmental factors in complex diseases: the older Finnish Twin Cohort. *Twin Research* 5: 358-365.

Karolchik D, Hinrichs A S, Furey T S, Roskin K M, Sugnet C W, Haussler D, and Kent W J (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Research* 32: D493-496.

Kayser M, Liu F, Janssens A C, Rivadeneira F, Lao O, van Duijn K, Vermeulen M et al. (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *American Journal of Human Genetics* 82: 411-423.

Kendler K S, Neale M C, Kessler R C, Heath A C, and Eaves L J (1993) A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* 23: 21-27.

Kent W J, Sugnet C W, Furey T S, Roskin K M, Pringle T H, Zahler A M, and Haussler D (2002) The human genome browser at UCSC. *Genome Research* 12: 996-1006.

Khawaja A P, Chan M P, Hayat S, Broadway D C, Luben R, Garway-Heath D F, Sherwin J C et al. (2013) The EPIC-Norfolk Eye Study: rationale, methods and a cross-sectional analysis of visual impairment in a population-based cohort. *BMJ Open* 3: e002684.

Kiefer A K, Tung J Y, Do C B, Hinds D A, Mountain J L, Francke U, and Eriksson N (2013) Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS Genetics* 9: e1003299.

Kim H J, Sohn K M, Shy M E, Krajewski K M, Hwang M, Park J H, Jang S Y et al. (2007) Mutations in PRPS1, which encodes the phosphoribosyl pyrophosphate synthetase enzyme critical for nucleotide biosynthesis, cause hereditary peripheral neuropathy with hearing loss and optic neuropathy (cmtx5). *American Journal of Human Genetics* 81: 552-558.

Kim M H, Zhao D, Kim W, Lim D H, Song Y M, Guallar E, Cho J et al. (2013) Heritability of myopia and ocular biometrics in Koreans: the healthy twin study. *Investigative Ophthalmology and Visual Science* 54: 3644-3649.

Kim Y, Lee Y, Lee S, Kim N H, Lim J, Kim Y J, Oh J H et al. (2015) On the Estimation of Heritability with Family-Based and Population-Based Samples. *Biomed Res Int* 2015: 671349.

Klein A P, Duggal P, Lee K E, Klein R, Bailey-Wilson J E, and Klein B E (2007) Confirmation of linkage to ocular refraction on chromosome 22q and identification of a novel linkage region on 1q. *Archives of Ophthalmology* 125: 80-85.

Klein A P, Suktitipat B, Duggal P, Lee K E, Klein R, Bailey-Wilson J E, and Klein B E (2009) Heritability analysis of spherical equivalent, axial length, corneal curvature, and anterior chamber depth in the Beaver Dam Eye Study. *Archives of Ophthalmology* 127: 649-655.

Konig I R, Loley C, Erdmann J, and Ziegler A (2014) How to include chromosome X in your genome-wide association study. *Genetic Epidemiology* 38: 97-103.

Koran M E, Thornton-Wells T A, Jahanshad N, Glahn D C, Thompson P M, Blangero J, Nichols T E et al. (2014) Impact of family structure and common environment on heritability estimation for neuroimaging genetics studies using Sequential Oligogenic Linkage Analysis Routines. *J Med Imaging (Bellingham)* 1: 014005.

Krishna Kumar S, Feldman M W, Rehkopf D H, and Tuljapurkar S (2016) Limitations of GCTA as a solution to the missing heritability problem. *Proceedings of the National Academy of Sciences of the United States of America* 113: E61-70.

Kronfeld P C, and Devney C (1930) The frequency of astigmatism. *Archives of Ophthalmology* 4: 873-884.

LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research* 37: 4181-4193.

Lam C S, Lam C H, Cheng S C, and Chan L Y (2012) Prevalence of myopia among Hong Kong Chinese schoolchildren: changes over two decades. *Ophthalmic and Physiological Optics* 32: 17-24.

Lam C Y, Tam P O, Fan D S, Fan B J, Wang D Y, Lee C W, Pang C P et al. (2008) A genome-wide scan maps a novel high myopia locus to 5p15. *Investigative Ophthalmology and Visual Science* 49: 3768-3778.

Lander E S, and Schork N J (1994) Genetic dissection of complex traits. *Science* 265: 2037-2048.

Lavanya R, Jeganathan V S, Zheng Y, Raju P, Cheung N, Tai E S, Wang J J et al. (2009) Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiology* 16: 325-336.

Lee J H, Jee D, Kwon J W, and Lee W K (2013a) Prevalence and risk factors for myopia in a rural Korean population. *Investigative Ophthalmology and Visual Science* 54: 5466-5471.

Lee S H, Wray N R, Goddard M E, and Visscher P M (2011) Estimating missing heritability for disease from genome-wide association studies. *American Journal of Human Genetics* 88: 294-305.

Lee Y Y, Lo C T, Sheu S J, and Lin J L (2013b) What factors are associated with myopia in young adults? A survey study in Taiwan Military Conscripts. *Investigative Ophthalmology and Visual Science* 54: 1026-1033.

Lever J, Krzywinski M, and Altman N (2016) Logistic regression. *Nature Methods* 13: 541-542.

Li J, and Zhang Q (2017) Insight into the molecular genetics of myopia. *Molecular Vision* 23: 1048-1080.

Li Q, Wojciechowski R, Simpson C L, Hysi P G, Verhoeven V J, Ikram M K, Hohn R et al. (2015a) Genome-wide association study for refractive astigmatism reveals genetic co-determination with spherical equivalent refractive error: the CREAM consortium. *Human Genetics* 134: 131-146.

Li S M, Kang M T, Wu S S, Liu L R, Li H, Chen Z, and Wang N (2016) Efficacy, Safety and Acceptability of Orthokeratology on Slowing Axial Elongation in Myopic Children by Meta-Analysis. *Current Eye Research* 41: 600-608.

Li S M, Li H, Li S Y, Liu L R, Kang M T, Wang Y P, Zhang F et al. (2015b) Time Outdoors and Myopia Progression Over 2 Years in Chinese Children: The Anyang Childhood Eye Study. *Investigative Ophthalmology and Visual Science* 56: 4734-4740.

Li S M, Li S Y, Kang M T, Zhou Y, Liu L R, Li H, Wang Y P et al. (2015c) Near Work Related Parameters and Myopia in Chinese Children: the Anyang Childhood Eye Study. *PLoS ONE* 10: e0134514.



Li S M, Liu L R, Li S Y, Ji Y Z, Fu J, Wang Y, Li H et al. (2013) Design, methodology and baseline data of a school-based cohort study in Central China: the Anyang Childhood Eye Study. *Ophthalmic Epidemiology* 20: 348-359.

Li Y, Willer C J, Ding J, Scheet P, and Abecasis G R (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology* 34: 816-834.

Li Y J, Goh L, Khor C C, Fan Q, Yu M, Han S, Sim X et al. (2011a) Genome-wide association studies reveal genetic variants in CTNND2 for high myopia in Singapore Chinese. *Ophthalmology* 118: 368-375.

Li Z, Qu J, Xu X, Zhou X, Zou H, Wang N, Li T et al. (2011b) A genome-wide association study reveals association between common variants in an intergenic region of 4q25 and high-grade myopia in the Chinese Han population. *Human Molecular Genetics* 20: 2861-2868.

Liu F, Wollstein A, Hysi P G, Ankra-Badu G A, Spector T D, Park D, Zhu G et al. (2010a) Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genetics* 6: e1000934.

Liu J Z, McRae A F, Nyholt D R, Medland S E, Wray N R, Brown K M, Investigators A et al. (2010b) A versatile gene-based test for genome-wide association studies. *American Journal of Human Genetics* 87: 139-145.

Liu X, Han D, Li J, Han B, Ouyang X, Cheng J, Li X et al. (2010c) Loss-of-function mutations in the PRPS1 gene cause a type of nonsyndromic X-linked sensorineural deafness, DFN2. *American Journal of Human Genetics* 86: 65-71.

Logan N S, Shah P, Rudnicka A R, Gilmartin B, and Owen C G (2011) Childhood ethnic differences in ametropia and ocular biometry: the Aston Eye Study. *Ophthalmic and Physiological Optics* 31: 550-558.

Loh P-R, Kichaev G, Gazal S, Schoech A P, and Price A L (2018) Mixed model association for biobank-scale data sets. *bioRxiv*.

Loh P R, Tucker G, Bulik-Sullivan B K, Vilhjalmsjon B J, Finucane H K, Salem R M, Chasman D I et al. (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* 47: 284-290.

Lopes M C, Andrew T, Carbonaro F, Spector T D, and Hammond C J (2009) Estimating heritability and shared environmental effects for refractive error in twin and family studies. *Investigative Ophthalmology and Visual Science* 50: 126-131.

Lopes M C, Hysi P G, Verhoeven V J, Macgregor S, Hewitt A W, Montgomery G W, Cumberland P et al. (2013) Identification of a candidate gene for astigmatism. *Investigative Ophthalmology and Visual Science* 54: 1260-1267.

Lopez-Maury L, Marguerat S, and Bahler J (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* 9: 583-593.

Luo K, Li J, Cui Y, Xu M, Yuan J, Tang W, Wan B et al. (2006) Identification and characterization of the human SCAN domain zinc-finger gene ZNF449. *Molecular Biology Reports* 33: 51-57.

Lyhne N, Sjolie A K, Kyvik K O, and 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. *British Journal of Ophthalmology* 85: 1470-1476.

Ma J H, Shen S H, Zhang G W, Zhao D S, Xu C, Pan C M, Jiang H et al. (2010) Identification of a locus for autosomal dominant high myopia on chromosome 5p13.3-p15.1 in a Chinese family. *Molecular Vision* 16: 2043-2054.

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H et al. (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research* 45: D896-D901.

Macgregor S, Cornes B K, Martin N G, and Visscher P M (2006) Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Human Genetics* 120: 571-580.

Magi R, and Morris A P (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11: 288.

Maglott D, Ostell J, Pruitt K D, and Tatusova T (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research* 39: D52-57.

Maher B (2008) Personal genomes: The case of the missing heritability. *Nature* 456: 18-21.

Manolio T A, Collins F S, Cox N J, Goldstein D B, Hindorff L A, Hunter D J, McCarthy M I et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.

Marchini J, and Howie B (2010) Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* 11: 499-511.

Marcus M W, de Vries M M, Junoy Montolio F G, and Jansonius N M (2011) Myopia as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. *Ophthalmology* 118: 1989-1994 e1982.

Martin N, Boomsma D, and Machin G (1997) A twin-pronged attack on complex traits. *Nature Genetics* 17: 387-392.

Maruyama Y, Wang X, Li Y, Sugar J, and Yue B Y (2001) Involvement of Sp1 elements in the promoter activity of genes affected in keratoconus. *Investigative Ophthalmology and Visual Science* 42: 1980-1985.

Mash A J, Hegmann J P, and Spivey B E (1975) Genetic analysis of indices of corneal power and corneal astigmatism in human populations with varying incidences of strabismus. *Investigative Ophthalmology* 14: 826-832.

McAllister K, Mechanic L E, Amos C, Aschard H, Blair I A, Chatterjee N, Conti D et al. (2017) Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases. *American Journal of Epidemiology* 186: 753-761.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood A R, Teumer A, Kang H M et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 48: 1279-1283.

Miraldi Utz V (2017) Nature versus nurture: A systematic approach to elucidate gene-environment interactions in the development of myopic refractive errors. *Ophthalmic Genetics* 38: 117-121.

Mishra A, and Macgregor S (2015) VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* 18: 86-91.

Mishra A, Yazar S, Hewitt A W, Mountain J A, Ang W, Pennell C E, Martin N G et al. (2012) Genetic variants near PDGFRA are associated with corneal curvature in Australians. *Investigative Ophthalmology and Visual Science* 53: 7131-7136.

Miyake M, Yamashiro K, Nakanishi H, Nakata I, Akagi-Kurashige Y, Kumagai K, Tsujikawa A et al. (2013) Heritability Estimation of Axial Length and Refractive Error Explained by Genome-Wide Single Nucleotide Polymorphisms. *Investigative Ophthalmology & Visual Science* 54: 1734-1734.

Moayyeri A, Hammond C J, Valdes A M, and Spector T D (2013) Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *International Journal of Epidemiology* 42: 76-85.

Morgan I G, French A N, Ashby R S, Guo X, Ding X, He M, and Rose K A (2018) The epidemics of myopia: Aetiology and prevention. *Progress in Retinal and Eye Research* 62: 134-149.

Mullaney J M, Mills R E, Pittard W S, and Devine S E (2010) Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics* 19: R131-R136.

Munafo M R, and Flint J (2004) Meta-analysis of genetic association studies. *Trends in Genetics* 20: 439-444.

Nakanishi H, Yamada R, Gotoh N, Hayashi H, Yamashiro K, Shimada N, Ohno-Matsui K et al. (2009) A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1. *PLoS Genetics* 5: e1000660.

Nakano K, Yamauchi J, Nakagawa K, Itoh H, and Kitamura N (2000) NESK, a member of the germinal center kinase family that activates the c-Jun N-terminal kinase pathway and is expressed during the late stages of embryogenesis. *Journal of Biological Chemistry* 275: 20533-20539.

Nallasamy S, Paluru P C, Devoto M, Wasserman N F, Zhou J, and Young T L (2007) Genetic linkage study of high-grade myopia in a Hutterite population from South Dakota. *Molecular Vision* 13: 229-236.

Neale B M, and Sham P C (2004) The future of association studies: gene-based analysis and replication. *American Journal of Human Genetics* 75: 353-362.

Nica A C, and Dermitzakis E T (2008) Using gene expression to investigate the genetic basis of complex disorders. *Human Molecular Genetics* 17: R129-134.

Nica A C, and Dermitzakis E T (2013) Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 368: 20120362.

Nolte I M, van der Most P J, Alizadeh B Z, de Bakker P I, Boezen H M, Bruinenberg M, Franke L et al. (2017) Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *European Journal of Human Genetics* 25: 877-885.

Northstone K, Guggenheim J A, Howe L D, Tilling K, Paternoster L, Kemp J P, McMahon G et al. (2013) Body stature growth trajectories during childhood and the development of myopia. *Ophthalmology* 120: 1064-1073 e1061.

Nurnberg G, Jacobi F K, Broghammer M, Becker C, Blin N, Nurnberg P, Stephani U et al. (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. *International Journal of Molecular Medicine* 21: 429-438.

O'Donoghue L, Kapetanankis V V, McClelland J F, Logan N S, Owen C G, Saunders K J, and Rudnicka A R (2015) Risk Factors for Childhood Myopia: Findings From the NICER Study. *Investigative Ophthalmology and Visual Science* 56: 1524-1530.

O'Donoghue L, McClelland J F, Logan N S, Rudnicka A R, Owen C G, and Saunders K J (2010) Refractive error and visual impairment in school children in Northern Ireland. *British Journal of Ophthalmology* 94: 1155-1159.

O'Donoghue L, Rudnicka A R, McClelland J F, Logan N S, Owen C G, and Saunders K J (2011) Refractive and corneal astigmatism in white school children in northern ireland. *Investigative Ophthalmology and Visual Science* 52: 4048-4053.

Ojaimi E, Rose K A, Morgan I G, Smith W, Martin F J, Kifley A, Robaei D et al. (2005) Distribution of ocular biometric parameters and refraction in a population-based study of Australian children. *Investigative Ophthalmology and Visual Science* 46: 2748-2754.

Ott J, Wang J, and Leal S M (2015) Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics* 16: 275-284.

Paget S, Julia S, Vitezica Z G, Soler V, Malecaze F, and Calvas P (2008) Linkage analysis of high myopia susceptibility locus in 26 families. *Molecular Vision* 14: 2566-2574.

Paluru P, Ronan S M, Heon E, Devoto M, Wildenberg S C, Scavello G, Holleschau A et al. (2003) New locus for autosomal dominant high myopia maps to the long arm of chromosome 17. *Investigative Ophthalmology and Visual Science* 44: 1830-1836.

Paluru P C, Nallasamy S, Devoto M, Rappaport E F, and Young T L (2005) Identification of a novel locus on 2q for autosomal dominant high-grade myopia. *Investigative Ophthalmology and Visual Science* 46: 2300-2307.

Pan C W, Dirani M, Cheng C Y, Wong T Y, and Saw S M (2015) The age-specific prevalence of myopia in Asia: a meta-analysis. *Optometry and Vision Science* 92: 258-266.

Pan C W, Wong T Y, Lavanya R, Wu R Y, Zheng Y F, Lin X Y, Mitchell P et al. (2011) Prevalence and risk factors for refractive errors in Indians: the Singapore Indian Eye Study (SINDI). *Investigative Ophthalmology and Visual Science* 52: 3166-3173.

Pan C W, Zheng Y F, Anuar A R, Chew M, Gazzard G, Aung T, Cheng C Y et al. (2013) Prevalence of refractive errors in a multiethnic Asian population: the Singapore epidemiology of eye disease study. *Investigative Ophthalmology and Visual Science* 54: 2590-2598.

Pfiffmann T, Emmerich D, Ruukonen P, Quandt D, Buchen R, Fischer-Zirnsak B, Hecht J et al. (2015) Molecular mechanism of CHRDL1-mediated X-linked megalocornea in humans and in *Xenopus* model. *Human Molecular Genetics* 24: 3119-3132.

Pickrell J K, Berisa T, Liu J Z, Segurel L, Tung J Y, and Hinds D A (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics* 48: 709-717.

Pineles S L, Kraker R T, VanderVeen D K, Hutchinson A K, Galvin J A, Wilson L B, and Lambert S R (2017) Atropine for the Prevention of Myopia Progression in Children: A Report by the American Academy of Ophthalmology. *Ophthalmology* 124: 1857-1866.

Polderman T J, Benyamin B, de Leeuw C A, Sullivan P F, van Bochoven A, Visscher P M, and Posthuma D (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47: 702-709.

Price A L, Patterson N J, Plenge R M, Weinblatt M E, Shadick N A, and Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38: 904-909.

Price A L, Zaitlen N A, Reich D, and Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* 11: 459-463.

Pritchard J K (2001) Are rare variants responsible for susceptibility to complex diseases? *American Journal of Human Genetics* 69: 124-137.

Pruim R J, Welch R P, Sanna S, Teslovich T M, Chines P S, Gliedt T P, Boehnke M et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26: 2336-2337.

Pulst S M (1999) Genetic linkage analysis. *Archives of Neurology* 56: 667-672.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M A, Bender D, Maller J et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559-575.

Quek T P, Chua C G, Chong C S, Chong J H, Hey H W, Lee J, Lim Y F et al. (2004) Prevalence of refractive errors in teenage high school students in Singapore. *Ophthalmic and Physiological Optics* 24: 47-55.

Rabbee N, and Speed T P (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22: 7-12.

Ratnamala U, Lyle R, Rawal R, Singh R, Vishnupriya S, Himabindu P, Rao V et al. (2011) Refinement of the X-linked nonsyndromic high-grade myopia locus MYP1 on Xq28 and exclusion of 13 known positional candidate genes by direct sequencing. *Investigative Ophthalmology and Visual Science* 52: 6814-6819.

Read S A, Collins M J, and Carney L G (2007) A review of astigmatism and its possible genesis. *Clinical & Experimental Optometry* 90: 5-19.

Redon R, Ishikawa S, Fitch K R, Feuk L, Perry G H, Andrews T D, Fiegler H et al. (2006) Global variation in copy number in the human genome. *Nature* 444: 444-454.

Richardson K, and Norgate S (2005) The equal environments assumption of classical twin studies may not hold. *British Journal of Educational Psychology* 75: 339-350.

Risch N, and Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273: 1516-1517.

Ritchie M D, Holzinger E R, Li R, Pendergrass S A, and Kim D (2015) Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics* 16: 85-97.

Rose K A, Morgan I G, Ip J, Kifley A, Huynh S, Smith W, and Mitchell P (2008) Outdoor activity reduces the prevalence of myopia in children. *Ophthalmology* 115: 1279-1285.

Sanfilippo P G, Hewitt A W, Hammond C J, and Mackey D A (2010) The heritability of ocular traits. *Survey of Ophthalmology* 55: 561-583.

Sanfilippo P G, Yazar S, Kearns L, Sherwin J C, Hewitt A W, and Mackey D A (2015) Distribution of astigmatism as a function of age in an Australian population. *Acta Ophthalmol* 93: e377-385.

Sankaridurg P, He X, Naduvilath T, Lv M, Ho A, Smith E, 3rd, Erickson P et al. (2017) Comparison of noncycloplegic and cycloplegic autorefraction in categorizing refractive error data in children. *Acta Ophthalmol* 95: e633-e640.

Saw S M, Katz J, Schein O D, Chew S J, and Chan T K (1996) Epidemiology of myopia. *Epidemiologic Reviews* 18: 175-187.

Saw S M, Shankar A, Tan S B, Taylor H, Tan D T, Stone R A, and Wong T Y (2006) A cohort study of incident myopia in Singaporean children. *Investigative Ophthalmology and Visual Science* 47: 1839-1844.

Sawaguchi S, Yue B Y, Sugar J, and Gilboy J E (1989) Lysosomal enzyme abnormalities in keratoconus. *Archives of Ophthalmology* 107: 1507-1510.

Scarr S, and Carter-Saltzman L (1979) Twin method: defense of a critical assumption. *Behavior Genetics* 9: 527-542.



Schache M, Chen C Y, Pertile K K, Richardson A J, Dirani M, Mitchell P, and Baird P N (2009) Fine mapping linkage analysis identifies a novel susceptibility locus for myopia on chromosome 2q37 adjacent to but not overlapping MYP12. *Molecular Vision* 15: 722-730.

Schache M, Richardson A J, Mitchell P, Wang J J, Rochtchina E, Viswanathan A C, Wong T Y et al. (2013) Genetic association of refractive error and axial length with 15q14 but not 15q25 in the Blue Mountains Eye Study cohort. *Ophthalmology* 120: 292-297.

Schork A J, Thompson W K, Pham P, Torkamani A, Roddey J C, Sullivan P F, Kelsoe J R et al. (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genetics* 9: e1003449.

Schulze T G, and McMahon F J (2004) Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Human Heredity* 58: 131-138.

Shah R L, Huang Y, Guggenheim J A, and Williams C (2017) Time Outdoors at Specific Ages During Early Childhood and the Risk of Incident Myopia. *Investigative Ophthalmology and Visual Science* 58: 1158-1166.

Sham P C, and Purcell S M (2014) Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics* 15: 335-346.

Shao X, Zhou K J, Pan A P, Cheng X Y, Cai H X, Huang J H, and Yu A Y (2017) Age-Related Changes in Corneal Astigmatism. *Journal of Refractive Surgery* 33: 696-703.

Sherwin J C, and Mackey D A (2013) Update on the epidemiology and genetics of myopic refractive error. *Expert Review of Ophthalmology* 8: 63-87.

Sherwin J C, Reacher M H, Keogh R H, Khawaja A P, Mackey D A, and Foster P J (2012) The association between time spent outdoors and myopia in children and adolescents: a systematic review and meta-analysis. *Ophthalmology* 119: 2141-2151.

Shi Y, Li Y, Zhang D, Zhang H, Li Y, Lu F, Liu X et al. (2011a) Exome sequencing identifies ZNF644 mutations in high myopia. *PLoS Genetics* 7: e1002084.

Shi Y, Qu J, Zhang D, Zhao P, Zhang Q, Tam P O S, Sun L et al. (2011b) Genetic variants at 13q12.12 are associated with high myopia in the Han Chinese population. *American Journal of Human Genetics* 88: 805-813.

Silventoinen K, Sammalisto S, Perola M, Boomsma D I, Cornes B K, Davis C, Dunkel L et al. (2003) Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Research* 6: 399-408.

Slatkin M (2008) Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* 9: 477-485.

Solouki A M, Verhoeven V J, van Duijn C M, Verkerk A J, Ikram M K, Hysi P G, Despret D D et al. (2010) A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14. *Nature Genetics* 42: 897-901.

Sorsby A, and Fraser G R (1964) Statistical Note on the Components of Ocular Refraction in Twins. *Journal of Medical Genetics* 1: 47-49.

Sorsby A, Sheridan M, and Leary G A (1962) *Refraction and its components in twins*. London: Her Majesty's Stationery Office.

Stambolian D, Ciner E B, Reider L C, Moy C, Dana D, Owens R, Schlifka M et al. (2005) Genome-wide scan for myopia in the Old Order Amish. *American Journal of Ophthalmology* 140: 469-476.

Stambolian D, Ibay G, Reider L, Dana D, Moy C, Schlifka M, Holmes T et al. (2004) Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. *American Journal of Human Genetics* 75: 448-459.

Stephens M, and Balding D J (2009) Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics* 10: 681-690.

Stone R A, and Khurana T S (2010) Gene profiling in experimental models of eye growth: clues to myopia pathogenesis. *Vision Research* 50: 2322-2333.

Stranger B E, Stahl E A, and Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187: 367-383.

Sturm R A, and Larsson M (2009) Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res* 22: 544-562.

Subramanian A, Tamayo P, Mootha V K, Mukherjee S, Ebert B L, Gillette M A, Paulovich A et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15545-15550.

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* 12: e1001779.

Sudmant P H, Rausch T, Gardner E J, Handsaker R E, Abyzov A, Huddleston J, Zhang Y et al. (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* 526: 75-81.

Sun J, Zhou J, Zhao P, Lian J, Zhu H, Zhou Y, Sun Y et al. (2012) High prevalence of myopia and high myopia in 5060 Chinese university students in Shanghai. *Investigative Ophthalmology and Visual Science* 53: 7504-7509.

Tan C S, Chan Y H, Wong T Y, Gazzard G, Niti M, Ng T P, and Saw S M (2011) Prevalence and risk factors for refractive errors and ocular biometry parameters in an elderly Asian population: the Singapore Longitudinal Aging Study (SLAS). *Eye (London, England)* 25: 1294-1301.

Tang W C, Yap M K, and Yip S P (2008) A review of current approaches to identifying human genes involved in myopia. *Clinical & Experimental Optometry* 91: 4-22.

Tedja M S, Wojciechowski R, Hysi P G, Eriksson N, Furlotte N A, Verhoeven V J M, Iglesias A I et al. (2018) Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nature Genetics* 50: 834-848.

Teikari J, O'Donnell J J, Kaprio J, and Koskenvuo M (1989) Genetic and environmental effects on oculometric traits. *Optometry and Vision Science* 66: 594-599.

Teikari J M, O'Donnell J, Kaprio J, and Koskenvuo M (1991) Impact of heredity in myopia. *Human Heredity* 41: 151-156.

Tenesa A, and Haley C S (2013) The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics* 14: 139-149.

The 1000 Genomes Project Consortium, Abecasis G R, Auton A, Brooks L D, DePristo M A, Durbin R M, Handsaker R E et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.

The Eye Disease Case-Control Study Group (1993) Risk factors for idiopathic rhegmatogenous retinal detachment. The Eye Disease Case-Control Study Group. *American Journal of Epidemiology* 137: 749-757.

The UK10K Consortium, Walter K, Min J L, Huang J, Crooks L, Memari Y, McCarthy S et al. (2015) The UK10K project identifies rare variants in health and disease. *Nature* 526: 82-90.

Thibos L N, Wheeler W, and Horner D (1997) Power vectors: an application of Fourier analysis to the description and statistical analysis of refractive error. *Optometry and Vision Science* 74: 367-375.

Tkatchenko A V, Tkatchenko T V, Guggenheim J A, Verhoeven V J, Hysi P G, Wojciechowski R, Singh P K et al. (2015) APLP2 Regulates Refractive Error and Myopia Development in Mice and Humans. *PLoS Genetics* 11: e1005432.

Tkatchenko A V, Walsh P A, Tkatchenko T V, Gustinich S, and Raviola E (2006) Form deprivation modulates retinal neurogenesis in primate experimental myopia. *Proceedings of the National Academy of Sciences of the United States of America* 103: 4681-4686.

Tran-Viet K N, Powell C, Barathi V A, Klemm T, Maurer-Stroh S, Limviphuvadh V, Soler V et al. (2013) Mutations in SCO2 are associated with autosomal-dominant high-grade myopia. *American Journal of Human Genetics* 92: 820-826.

Tran-Viet K N, St Germain E, Soler V, Powell C, Lim S H, Klemm T, Saw S M et al. (2012) Study of a US cohort supports the role of ZNF644 and high-grade myopia susceptibility. *Molecular Vision* 18: 937-944.

Tsukita S, and Furuse M (2000) Pores in the wall: claudins constitute tight junction strands containing aqueous pores. *Journal of Cell Biology* 149: 13-16.

Verhoeven V J, Hysi P G, Wojciechowski R, Fan Q, Guggenheim J A, Hohn R, MacGregor S et al. (2013) Genome-wide meta-analyses of multi-ancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nature Genetics* 45: 314-318.

Visscher P M (2008) Sizing up human height variation. *Nature Genetics* 40: 489-490.

Visscher P M, Hill W G, and Wray N R (2008) Heritability in the genomics era-- concepts and misconceptions. *Nature Reviews Genetics* 9: 255-266.

Visscher P M, Wray N R, Zhang Q, Sklar P, McCarthy M I, Brown M A, and Yang J (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics* 101: 5-22.

Vitale S, Ellwein L, Cotch M F, Ferris F L, 3rd, and Sperduto R (2008) Prevalence of refractive error in the United States, 1999-2004. *Archives of Ophthalmology* 126: 1111-1119.

Wain L V, Shrine N, Miller S, Jackson V E, Ntalla I, Artigas M S, Billington C K et al. (2015) Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respiratory Medicine* 3: 769-781.

Wall J D, and Pritchard J K (2003) Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* 4: 587-597.

Wallman J, and Winawer J (2004) Homeostasis of eye growth and the question of myopia. *Neuron* 43: 447-468.

Wang D G, Fan J B, Siao C J, Berno A, Young P, Sapolsky R, Ghandour G et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077-1082.

Wang K, Dickson S P, Stolle C A, Krantz I D, Goldstein D B, and Hakonarson H (2010) Interpretation of association signals and identification of causal variants from genome-wide association studies. *American Journal of Human Genetics* 86: 730-742.

Wang X, Chua H X, Chen P, Ong R T, Sim X, Zhang W, Takeuchi F et al. (2013) Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Human Molecular Genetics* 22: 2303-2311.

Wang Y X, Zhang J S, You Q S, Xu L, and Jonas J B (2014) Ocular diseases and 10-year mortality: the Beijing Eye Study 2001/2011. *Acta Ophthalmol* 92: e424-428.

Webb T R, Matarin M, Gardner J C, Kelberman D, Hassan H, Ang W, Michaelides M et al. (2012) X-linked megalocornea caused by mutations in *CHRD1* identifies an essential role for ventroptin in anterior segment development. *American Journal of Human Genetics* 90: 247-259.

Weiss K M, and Clark A G (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics* 18: 19-24.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* 42: D1001-1006.

Westra H J, and Franke L (2014) From genome to function by studying eQTLs. *Biochimica et Biophysica Acta* 1842: 1896-1902.

Willer C J, Li Y, and Abecasis G R (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190-2191.

Williams C, Miller L, Northstone K, and Sparrow J M (2008a) The use of non-cycloplegic autorefractometry data in general studies of children's development. *British Journal of Ophthalmology* 92: 723-724.

Williams C, Miller L L, Gazzard G, and Saw S M (2008b) A comparison of measures of reading and intelligence as risk factors for the development of myopia in a UK cohort of children. *British Journal of Ophthalmology* 92: 1117-1121.

Williams K M, Bertelsen G, Cumberland P, Wolfram C, Verhoeven V J, Anastasopoulos E, Buitendijk G H et al. (2015a) Increasing Prevalence of Myopia in Europe and the Impact of Education. *Ophthalmology* 122: 1489-1497.

Williams K M, Verhoeven V J, Cumberland P, Bertelsen G, Wolfram C, Buitendijk G H, Hofman A et al. (2015b) Prevalence of refractive error in Europe: the European Eye Epidemiology (E(3)) Consortium. *European Journal of Epidemiology* 30: 305-315.

Winkler T W, Day F R, Croteau-Chonka D C, Wood A R, Locke A E, Magi R, Ferreira T et al. (2014) Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols* 9: 1192-1212.

Winkler T W, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, and Heid I M (2015) EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* 31: 259-261.

Wise A L, Gyi L, and Manolio T A (2013) eXclusion: toward integrating the X chromosome in genome-wide association analyses. *American Journal of Human Genetics* 92: 643-647.

Wistow G, Bernstein S L, Wyatt M K, Fariss R N, Behal A, Touchman J W, Bouffard G et al. (2002) Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: over 6000 non-redundant transcripts, novel genes and splice variants. *Molecular Vision* 8: 205-220.

Wojciechowski R, and Cheng C Y (2018) Involvement of Multiple Molecular Pathways in the Genetics of Ocular Refraction and Myopia. *Retina* 38: 91-101.

Wojciechowski R, Congdon N, Bowie H, Munoz B, Gilbert D, and West S K (2005) Heritability of refractive error and familial aggregation of myopia in an elderly American population. *Investigative Ophthalmology and Visual Science* 46: 1588-1592.

Wojciechowski R, Moy C, Ciner E, Ibay G, Reider L, Bailey-Wilson J E, and Stambolian D (2006) Genomewide scan in Ashkenazi Jewish families demonstrates evidence of linkage of ocular refraction to a QTL on chromosome 1p36. *Human Genetics* 119: 389-399.

Wong T Y, Foster P J, Hee J, Ng T P, Tielsch J M, Chew S J, Johnson G J et al. (2000) Prevalence and risk factors for refractive errors in adult Chinese in Singapore. *Investigative Ophthalmology and Visual Science* 41: 2486-2494.

Wood A R, Esko T, Yang J, Vedantam S, Pers T H, Gustafsson S, Chu A Y et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 46: 1173-1186.

Wu H-M, Seet B, Yap E P-H, Saw S-M, Lim T-H, Chia, and Kee-Seng (2001) Does Education Explain Ethnic Differences in Myopia Prevalence? A Population-Based

Study of Young Adult Males in Singapore. *Optometry and Vision Science* 78: 234-239.

Wu P C, Tsai C L, Wu H L, Yang Y H, and Kuo H K (2013) Outdoor activity during class recess reduces myopia onset and progression in school children. *Ophthalmology* 120: 1080-1085.

Xiang X, Wang T, Tong P, Li Y, Guo H, Wan A, Xia L et al. (2014) New ZNF644 mutations identified in patients with high myopia. *Molecular Vision* 20: 939-946.

Xiao X, Li S, Jia X, Guo X, and Zhang Q (2016) X-linked heterozygous mutations in ARR3 cause female-limited early onset high myopia. *Molecular Vision* 22: 1257-1266.

Xiong S, Sankaridurg P, Naduvilath T, Zang J, Zou H, Zhu J, Lv M et al. (2017) Time spent in outdoor activities in relation to myopia prevention and control: a meta-analysis and systematic review. *Acta Ophthalmol* 95: 551-566.

Yang J, Benjamin B, McEvoy B P, Gordon S, Henders A K, Nyholt D R, Madden P A et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42: 565-569.

Yang J, Ferreira T, Morris A P, Medland S E, Genetic Investigation of A T C, Replication D I G, Meta-analysis C et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* 44: 369-375, S361-363.

Yang J, Lee S H, Goddard M E, and Visscher P M (2011a) GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88: 76-82.

Yang J, Manolio T A, Pasquale L R, Boerwinkle E, Caporaso N, Cunningham J M, de Andrade M et al. (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* 43: 519-525.

Yang J, Weedon M N, Purcell S, Lettre G, Estrada K, Willer C J, Smith A V et al. (2011c) Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* 19: 807-812.



Yang J, Zaitlen N A, Goddard M E, Visscher P M, and Price A L (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* 46: 100-106.

Yang J, Zeng J, Goddard M E, Wray N R, and Visscher P M (2017) Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics* 49: 1304-1310.

Yang Z, Xiao X, Li S, and Zhang Q (2009) Clinical and linkage study on a consanguineous Chinese family with autosomal recessive high myopia. *Molecular Vision* 15: 312-318.

Yazar S, Forward H, McKnight C M, Tan A, Soloshenko A, Oates S K, Ang W et al. (2013a) Raine eye health study: design, methodology and baseline prevalence of ophthalmic disease in a birth-cohort study of young adults. *Ophthalmic Genetics* 34: 199-208.

Yazar S, Mishra A, Ang W, Kearns L S, Mountain J A, Pennell C, Montgomery G W et al. (2013b) Interrogation of the platelet-derived growth factor receptor alpha locus and corneal astigmatism in Australians of Northern European ancestry: results of a genome-wide association study. *Molecular Vision* 19: 1238-1246.

Young T L, Atwood L D, Ronan S M, Dewan A T, Alvear A B, Peterson J, Holleschau A et al. (2001) Further refinement of the MYP2 locus for autosomal dominant high myopia by linkage disequilibrium analysis. *Ophthalmic Genetics* 22: 69-75.

Young T L, Metlapally R, and Shay A E (2007) Complex trait genetics of refractive error. *Archives of Ophthalmology* 125: 38-48.

Young T L, Ronan S M, Alvear A B, Wildenberg S C, Oetting W S, Atwood L D, Wilkin D J et al. (1998a) A second locus for familial high myopia maps to chromosome 12q. *American Journal of Human Genetics* 63: 1419-1424.

Young T L, Ronan S M, Drahozal L A, Wildenberg S C, Alvear A B, Oetting W S, Atwood L D et al. (1998b) Evidence that a locus for familial high myopia maps to chromosome 18p. *American Journal of Human Genetics* 63: 109-119.

Yu J, Pressoir G, Briggs W H, Vroh Bi I, Yamasaki M, Doebley J F, McMullen M D et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203-208.

Zagozewski J L, Zhang Q, Pinto V I, Wigle J T, and Eisenstat D D (2014) The role of homeobox genes in retinal development and disease. *Developmental Biology* 393: 195-208.

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, and Price A L (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genetics* 9: e1003520.

Zhang Q, Guo X, Xiao X, Jia X, Li S, and Hejtmancik J F (2005) A new locus for autosomal dominant high myopia maps to 4q22-q27 between D4S1578 and D4S1612. *Molecular Vision* 11: 554-560.

Zhang Q, Guo X, Xiao X, Jia X, Li S, and Hejtmancik J F (2006) Novel locus for X linked recessive high myopia maps to Xq23-q25 but outside MYP1. *Journal of Medical Genetics* 43: e20.

Zhang Q, Li S, Xiao X, Jia X, and Guo X (2007) Confirmation of a genetic locus for X-linked recessive high myopia outside MYP1. *Journal of Human Genetics* 52: 469-472.

Zhang Z, Ersoz E, Lai C Q, Todhunter R J, Tiwari H K, Gore M A, Bradbury P J et al. (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42: 355-360.

Zhao F, Wu J, Xue A, Su Y, Wang X, Lu X, Zhou Z et al. (2013) Exome sequencing reveals CCDC111 mutation associated with high myopia. *Human Genetics* 132: 913-921.

Zheng G, Freidlin B, and Gastwirth J L (2006) Robust genomic control for association studies. *American Journal of Human Genetics* 78: 350-356.

Zheng G, Joo J, Zhang C, and Geller N L (2007) Testing association for markers on the X chromosome. *Genetic Epidemiology* 31: 834-843.

Zhou W, Nielsen J B, Fritsche L G, Dey R, Elvestad M B, Wolford B N, LeFaive J et al. (2017a) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *bioRxiv*.

Zhou Z, Chen T, Wang M, Jin L, Zhao Y, Chen S, Wang C et al. (2017b) Pilot study of a novel classroom designed to prevent myopia by increasing children's exposure to outdoor light. *PLoS ONE* 12: e0181772.

Zhu Z, Bakshi A, Vinkhuyzen A A, Hemani G, Lee S H, Nolte I M, van Vliet-Ostaptchouk J V et al. (2015) Dominance genetic variation contributes little to the missing heritability for human complex traits. *American Journal of Human Genetics* 96: 377-385.

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson M R, Powell J E, Montgomery G W et al. (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* 48: 481-487.

# Appendices

**Appendix A Acknowledgements and study Information for cohorts  
investigated in *Chapter 4 CREAM Corneal  
Astigmatism GWAS Meta-analysis***

**ALSPAC.** The Avon Longitudinal Study of Parents and Children (ALSPAC) team and authors are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This publication is the work of the authors and JAG and CW will serve as guarantors for the contents of this paper. This research was specifically funded by NIHR Senior Research Fellowship SRF-2015-08-005 (CW) and a Wellcome Trust ISSF Populations Pilot Award (grant 508353/509506). Ethical approval for the ALSPAC study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Please note that the ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary: <http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>. ALSPAC children with available genotype data and corneal curvature phenotype information formed the GWAS sample (Table 4.1). A description of the ALSPAC study cohort is available (Boyd et al., 2013).

**BES.** The Beijing Eye Study (BES) was supported by National Natural Science Foundation of China (grant # 81770890). This publication is the work of the authors and YXW and JBJ will serve as guarantors for the contents of this paper. The study was approved by the Medical Ethics Committee of the Beijing Tongren Hospital. A description of the BES study cohort is available (Wang et al., 2014).

**BMES.** The Blue Mountains Eye Study (BMES) acknowledge funding from the National Health and Medical Research Council of Australia (NHMRC) Senior Research Fellowship 1138585 (PNB). The Centre for Eye Research Australia (CERA) receives Operational Infrastructure Support from the Victorian Government. Details of the BMES cohort have been published previously (Schache et al., 2013).

**EPIC.** The European Prospective Investigation of Cancer (EPIC)-Norfolk infrastructure and core functions are supported by grants from the Medical Research Council (G1000143) and Cancer Research UK (C864/A14136). The clinic for the third health examination was funded by Research into Aging (262). Mr Khawaja was a Wellcome Trust Clinical Research Fellow at the time of analysis. The EPIC-Norfolk Eye Study was performed following the principles of the Declaration of Helsinki and the Research Governance Framework for Health and Social Care. The study was approved by the Norfolk Local Research Ethics Committee (05/Q0101/191) and East Norfolk & Waveney NHS Research Governance Committee (2005EC07L). All participants gave written, informed consent. A description of the EPIC study cohort is available (Khawaja et al., 2013).

**FITSA.** Finnish Twin Study on Aging (FITSA) is a study of genetic and environmental effects on the disablement process in older female twins. The study cohort of 13,888 adult twin pairs started in 1975. Altogether 103 MZ and 114 DZ twin pairs (424 individuals, all women of European ancestry) aged 63–76 years living in Finland took part in multiple laboratory examinations in 2000 and 2003, and responded in questionnaires in 2011. Before the examinations, the subjects provided a written informed consent according to the Declaration of Helsinki. The study protocol was approved by the ethics committee of the Central Hospital District of Central Finland. FITSA was supported by ENGAGE (FP7-HEALTH-F4–2007, 201,413); European Union through the GENOMEUTWIN project (QLG2-CT-2002–01254); the Academy of Finland Center of Excellence in Complex Disease Genetics (213506, 129680); the Academy of Finland Aging Programme; and the Finnish Ministry of Culture and Education and University of Jyväskylä, Silmäsaatiö Foundation and Evald & Hilda Nissi Foundation. For FITSA the contributions of Emmi Tikkanen, Samuli Ripatti, Markku Kauppinen, Taina Rantanen and Jaakko Kaprio are acknowledged. A description of the FITSA cohort has been published (Kaprio and Koskenvuo, 2002).

**Generation R.** The Generation R study is conducted by the Erasmus Medical Centre in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam, the Rotterdam Homecare Foundation, and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (Star-MDC), Rotterdam. We gratefully acknowledge the contribution of the children and parents, as well as the participating general practitioners, hospitals, midwives, and pharmacies in Rotterdam. The Generation R

study is made possible by financial support from the Erasmus Medical Centre, Rotterdam; the Netherlands Organisation for Scientific Research (NWO); the Netherlands Organisation for Health Research and Development (ZonMw); the Dutch Ministry of Education, Culture and Science; the Dutch Ministry of Health, Welfare, and Sports; the European Commission (DG XII); and Uitzicht (Grant 2013–24). The study protocol was approved by the Medical Ethical Committee of the Erasmus Medical Centre, Rotterdam (MEC 217.595/2002/20), and written informed consent was obtained from all participants. Research was conducted according to the declaration of Helsinki. A description of the Generation R study has been published (Jaddoe et al., 2006).

**Gutenberg Health Study (GHS 1 and GHS 2).** The Gutenberg Health Study is a population-based, prospective, observational cohort study in mid-western Germany that includes consecutive follow-ups every five years. The primary study aim is to evaluate and improve cardiovascular risk stratification and the general health status of the population. The baseline examination included a total of 15,010 participants aged 35 to 74 years and took place from 2007 to 2012. The participants were randomly drawn and equally stratified for sex, residence (urban or rural) and for each decade of age. Exclusion criteria were the following: insufficient knowledge of German and physical or mental inability to participate in the examinations in the study center. The study protocol and study documents were approved by the local ethics committee of the Medical Chamber of Rhineland-Palatinate, Germany (reference no. 837.020.07; original vote: 22.3.2007, latest update: 20.10.2015). According to the tenets of the Declaration of Helsinki, written informed consent



was obtained from all participants before their entry into the study. The Gutenberg Health Study is funded through the government of Rhineland-Palatinate (“Stiftung Rheinland-Pfalz für Innovation,” contract AZ 961–386261/733), the research programs “Wissen schafft Zukunft” and “Center for Translational Vascular Biology (CTVB)” of the Johannes Gutenberg-University of Mainz, the National Genome Network “NGFNplus” by the Federal Ministry of Education and Research, Germany (A301GS0833) and its contracts with Boehringer Ingelheim and PHILIPS Medical Systems. We thank all study participants for their willingness to provide data for this research project and we are indebted to all coworkers for their enthusiastic commitment. A description of the ophthalmic arm of the GHS has been published (Hohn et al., 2015).

**OGP.** The Ogliastra Genetic Park (OGP) study authors would like to express their gratitude to all the study participants for their contributions, to the municipal administrations for their economic and logistic support and, to the whole OGP team, which includes interviewers, computer and laboratory technicians, research scientists, physicians and nurses. This research was supported by grant from the Italian Ministry of Education, University and Research (MIUR) no: 5571/DSPAR/2002. The research protocol of the study was approved by the institutional review board of the Italian Ministry of Education, University and Research. It adheres to the tenets of the declaration of Helsinki, furthermore written informed consent was obtained from all participants. A description of the OGP study cohort has been published (Biino et al., 2010).

**RAINE** (Western Australian Pregnancy Cohort). We are grateful to all the study participants. We also thank the Raine Study and Lions Eye Institute (LEI) research staff for cohort coordination and data collection. The core management of the Raine Study is funded by The University of Western Australia (UWA), The Telethon Institute for Child Health Research, Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, Women's and Infant's Research Foundation and Curtin University. Genotyping was funded by Australian National Health and Medical Research Council (NHMRC) project grant 1021105. Support for the REHS was provided by LEI, the Australian Foundation for the Prevention of Blindness and ORIA. SY is supported by NHMRC CJ Martin Early Career Fellowship (#1111437). A description of the RAINE Eye Health Study cohort is available (Yazar et al., 2013a).

**Rotterdam Study (RS1, RS2, RS3).** The Rotterdam Study is a prospective population-based cohort study in the elderly living in Ommoord, a suburb of Rotterdam, the Netherlands. In brief, the Rotterdam Study consists of 3 independent cohorts: RS1, RS2, and RS3. For the current analysis, 5,328 residents aged 55 years and older were included from RS1, 2,009 participants aged 55 and older from RS2, and 1,970 aged 45 and older from RS3. 99% of subjects were of European ancestry. Participants underwent multiple physical examinations with regular intervals from 1991 to present, including a non-dilated automated measurement of refractive error using a Topcon RM-A2000 autorefractor. All measurements in RS-1–3 were conducted after the Medical Ethics Committee of the Erasmus University had approved the study protocols and all participants had given a written informed consent in accordance

with the Declaration of Helsinki. The Rotterdam Study was supported by the Dutch governmental Innovational Research Incentives Scheme Grant (VICI 91815655); Horizon2020 ERC Consolidator Grant (648268); Erasmus Medical Center and Erasmus University, Rotterdam, The Netherlands; Netherlands Organization for Health Research and Development (ZonMw); Uitzicht; the Research Institute for Diseases in the Elderly; the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); the Municipality of Rotterdam; the Netherlands Genomics Initiative/NWO; Center for Medical Systems Biology of NCI; Lijf en Leven; Henkes Stichting; Landelijke Stichting voor Blinden en Slechtzienden; Oogfonds; MaculaFonds. We acknowledge Ada Hooghart, Corina Brussee, Riet Bernaerts-Biskop, Patricia van Hilten, Pascal Arp, Jeanette Vergeer, Marijn Verkerk; Sander Bervoets for their valuable contributions. A description of the Rotterdam study has been published (Hofman et al., 2015).

**SCES, SIMES and SINDI.** The Singapore Chinese Eye Study (SCES), Singapore Malay Eye Study (SiMES) and Singapore Indian Eye Study (SINDI) were supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council, Singapore (08/1/35/19/550 and 09/1/35/19/616). Ching-Yu Cheng is supported by an award from NMRC (CSA/033/2012). Descriptions of the SCES, SIMES and SINDI cohorts have been published (Foong et al., 2007; Lavanya et al., 2009; Pan et al., 2011).

**SCORM.** The Singapore Cohort Study of the Risk Factors for Myopia (SCORM) was supported by the Biomedical Research Council (BMRC) 06/1/21/19/466. A description of the SCORM cohort has been published (Saw et al., 2006).

**STARS.** The Singaporean Chinese in the Strabismus, Amblyopia, and Refractive Error Study (STARS) was supported by National Medical Research Council (NMRC), Singapore (grants 1176/2008). A description of the STARS cohort has been published (Dirani et al., 2010).

**TwinsUK.** The TwinsUK adult twin registry based at St. Thomas' Hospital in London is a volunteer cohort of over 10,000 twins from the general population. Twins largely volunteered unaware of the eye studies, gave fully informed consent under a protocol reviewed by the St. Thomas' Hospital Local Research Ethics Committee. TwinsUK is funded by the Wellcome Trust and the European Community's Seventh Framework Programme (FP7/2007–2013). The study also receives support from the National Institute for Health Research Clinical Research Facility at Guy's and St. Thomas' National Health Service Foundation Trust and National Institute for Health Research Biomedical Research Centre at Guy's and St. Thomas' National Health Service Foundation Trust and King's College London. Keratometry was obtained using the VX-120 ocular diagnostic device (Visionix®, Luneau Technology Group). A description of the TwinsUK study cohort is available (Moayyeri et al., 2013).

**CREAM Meta-analyses.** This work was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health.

**Appendix B EasyQC style plots for study-wide quality control in**

***Chapter 4 CREAM Corneal Astigmatism GWAS***

***Meta-analysis***

**B1-2: Effect Allele Frequency (EAF) plots for each cohort**

B1: EAF plots for the 14 European ancestry cohorts.

B2: EAF plots for the 8 Asian ancestry cohorts

**B3-4: P-value vs. Z-score (P-Z) plots for each cohort**

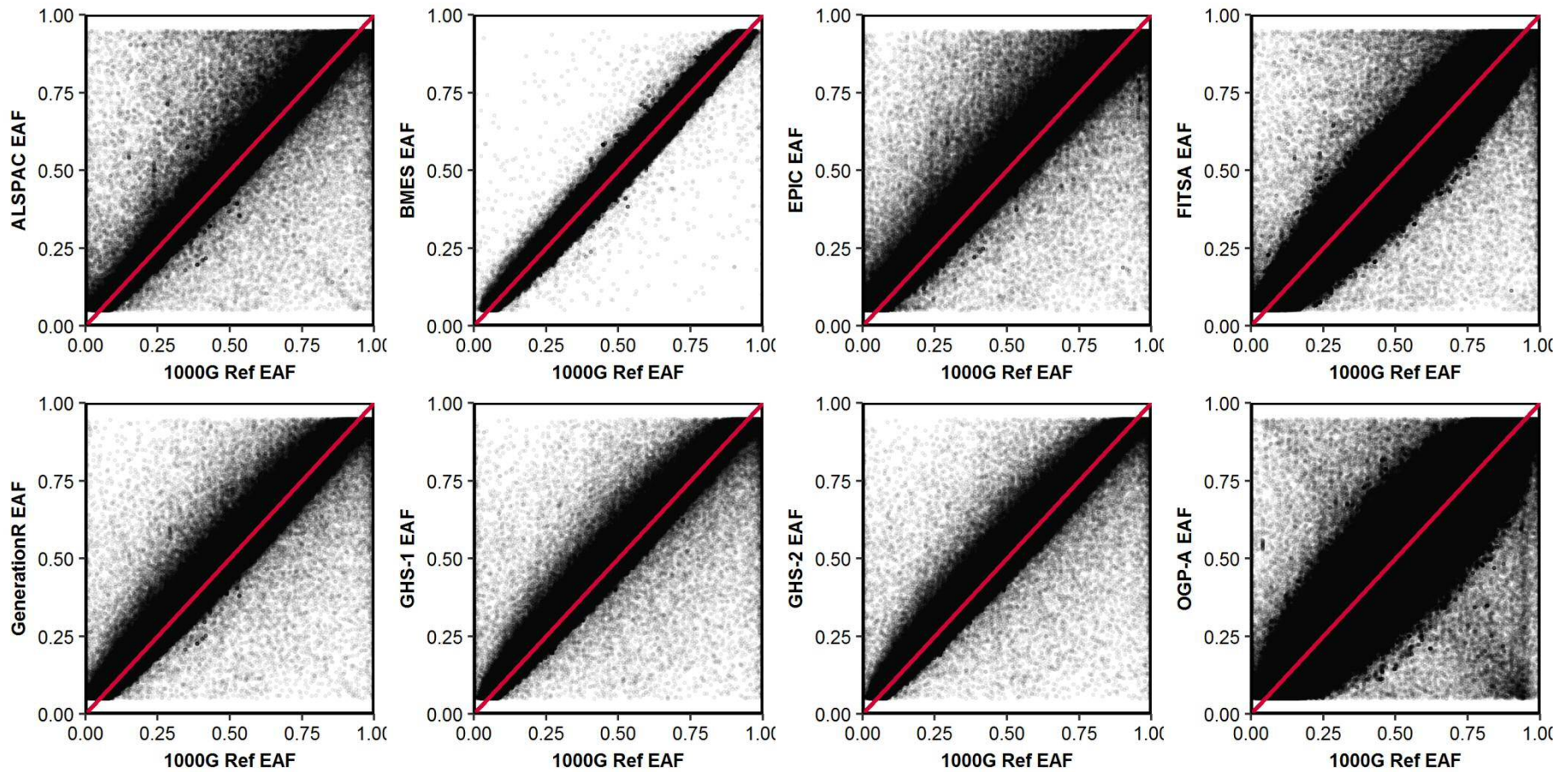
B3: P-Z plots for the 14 European ancestry cohorts.

B4: P-Z plots for the 8 Asian ancestry cohorts

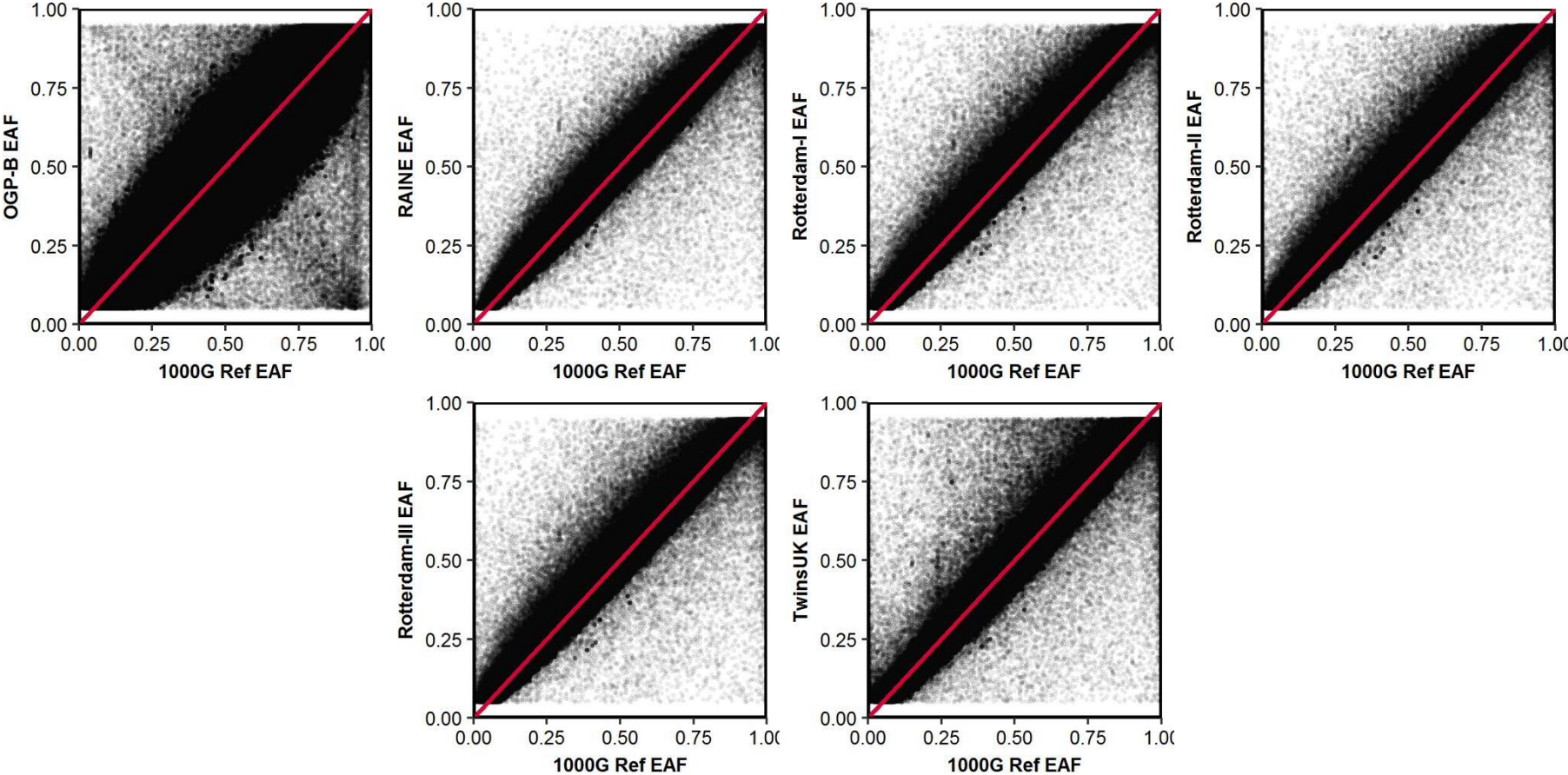
**B5: Standard error vs. sample size (SE-N) plot for all 22 cohorts**

**B6: Lambda-N plot for all 22 cohorts**

**B1: Effect Allele Frequency (EAF) plots for each of the 14 European ancestry cohorts.** Y-axis: cohort allele frequency of the *Effect Allele*. 1000G Ref EAF: allele frequency of the same *Effect Allele* in Phase 3 of the 1000 Genomes Project European samples (N = 503). Red line = line of unity (Study EAF = 1000 Genomes Reference EAF).

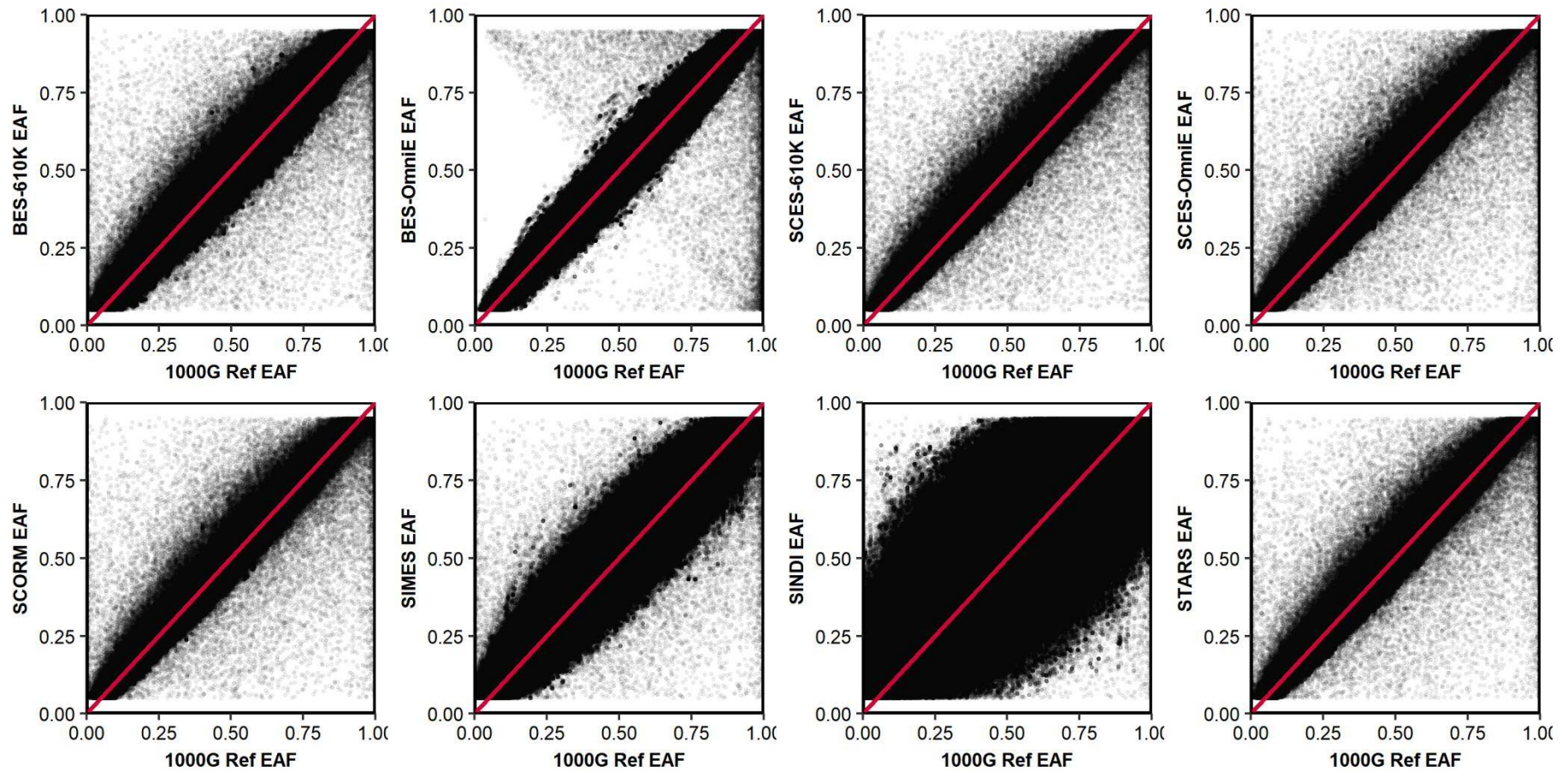


(B1 continued)

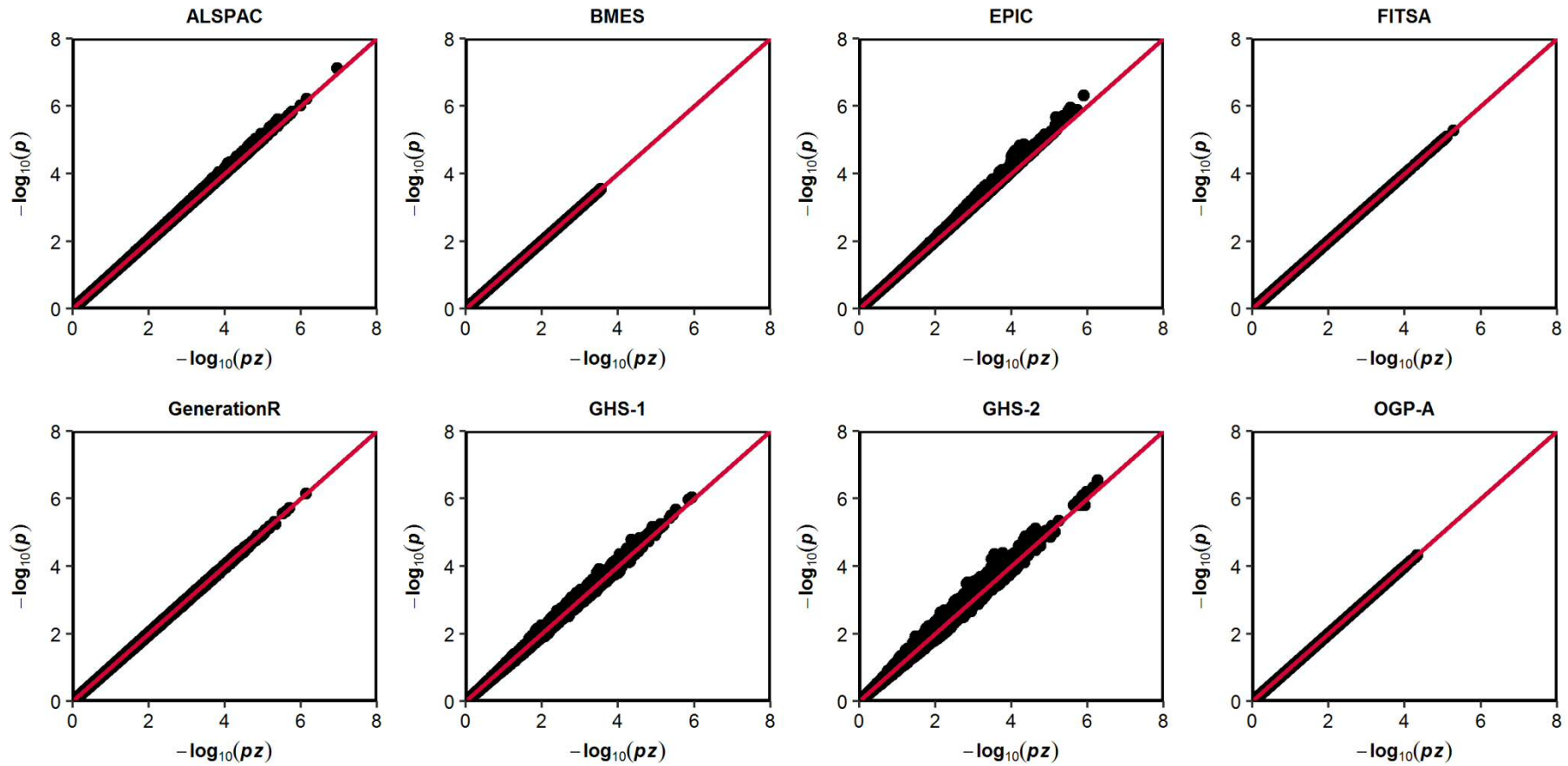




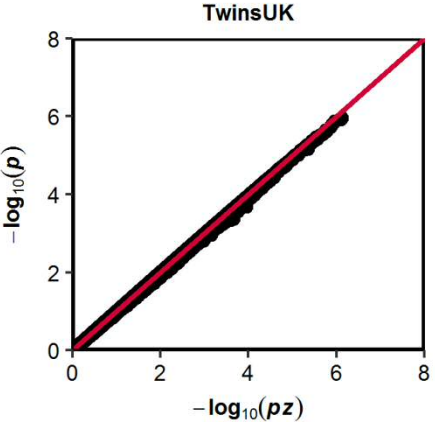
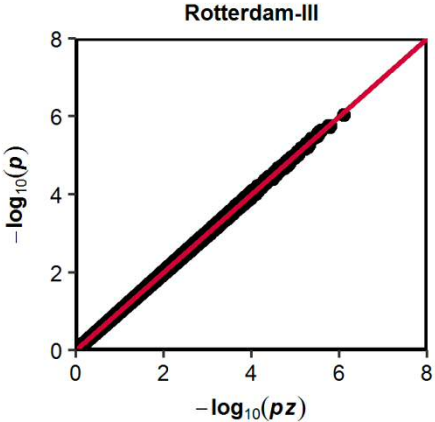
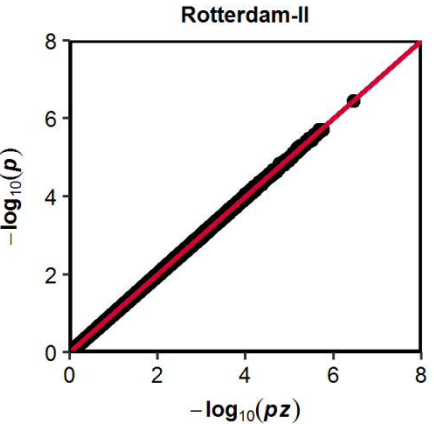
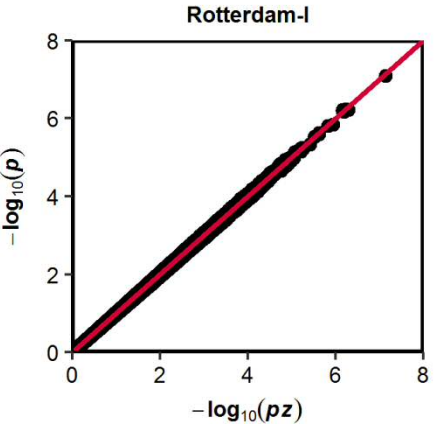
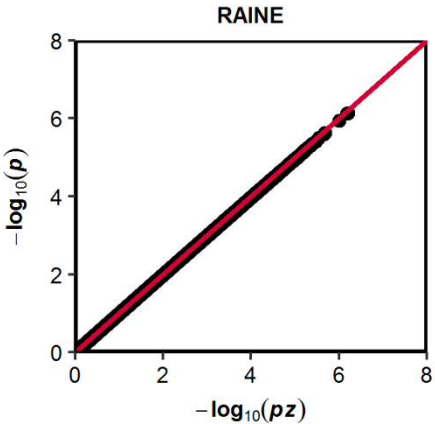
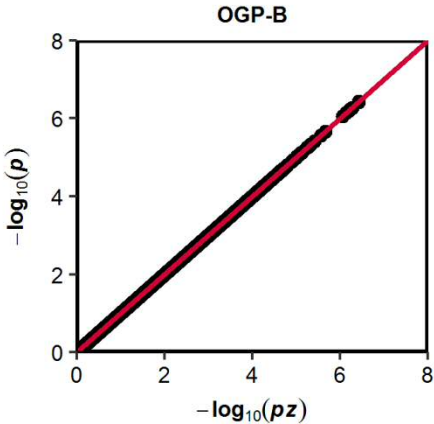
**B2: Effect Allele Frequency (EAF) plots for each of the 8 Asian ancestry cohorts.** Y-axis: cohort allele frequency of the *Effect Allele*; 1000G Ref EAF: allele frequency of the same *Effect Allele* in Phase 1 of the 1000 Genomes Project East Asian samples (N = 504). Red line = line of unity (Study EAF = 1000 Genomes Reference EAF).



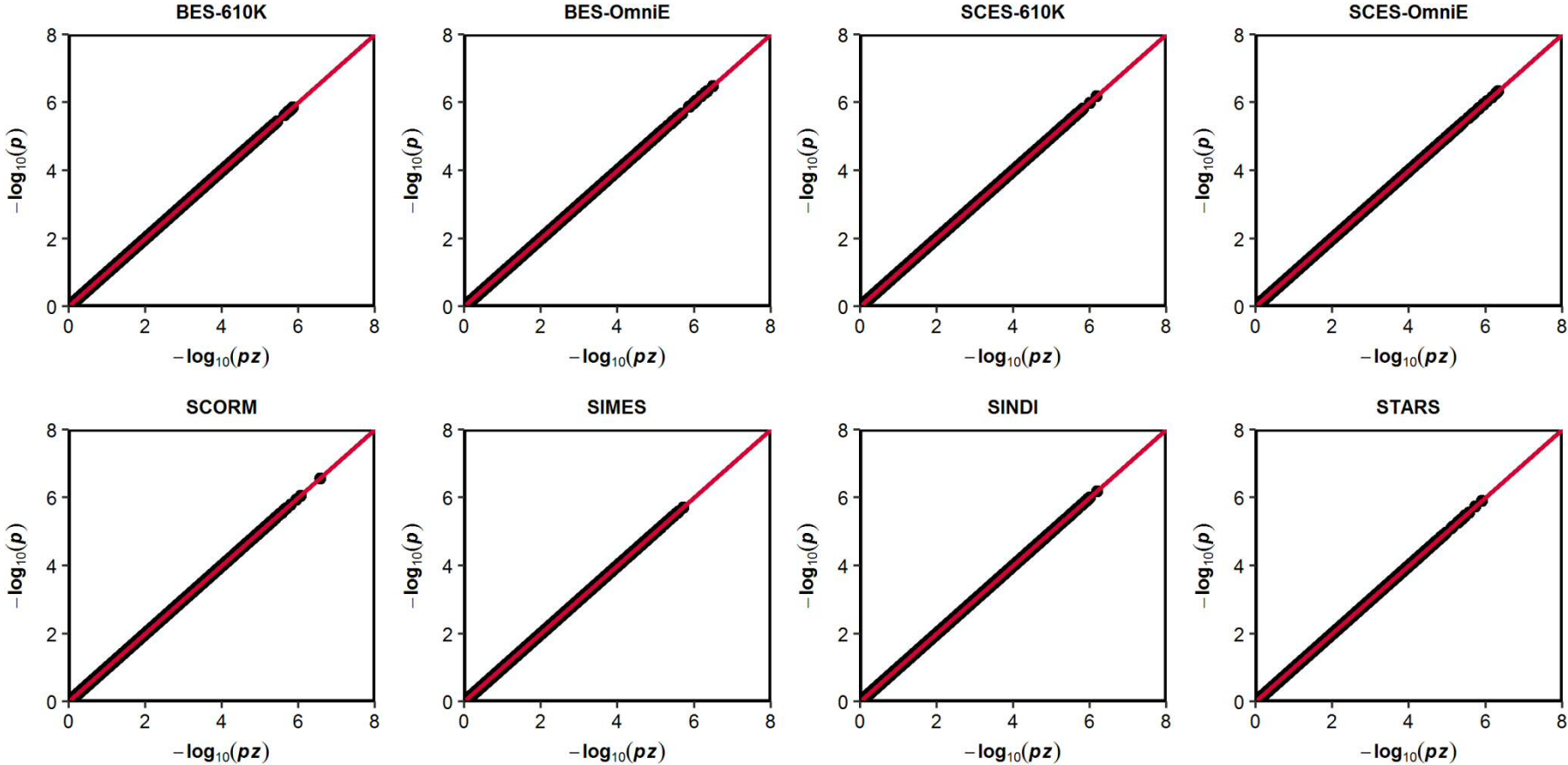
**B3: P-value versus Z-score (P-Z) plots for each of the 14 European ancestry cohorts.** X-axis:  $-\log_{10} p$ -value computed from Z-statistics for each variant tested; Y-axis:  $-\log_{10} p$ -value reported for each variant tested. NB: Z-statistic for each variant =  $\beta/SE$ . Red line = line of unity (observed p-value = p-value determined from Z-statistics).



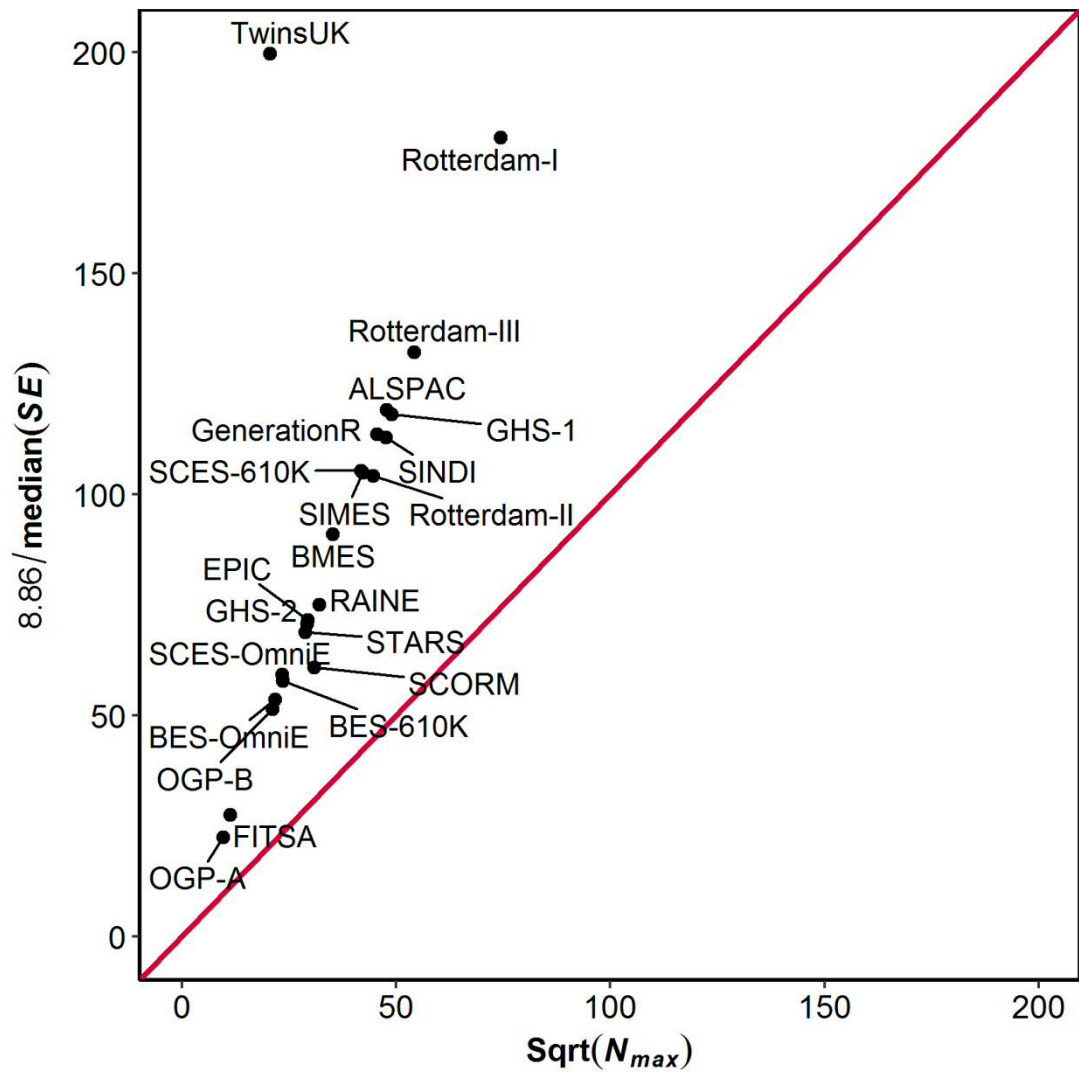
(B3 continued)



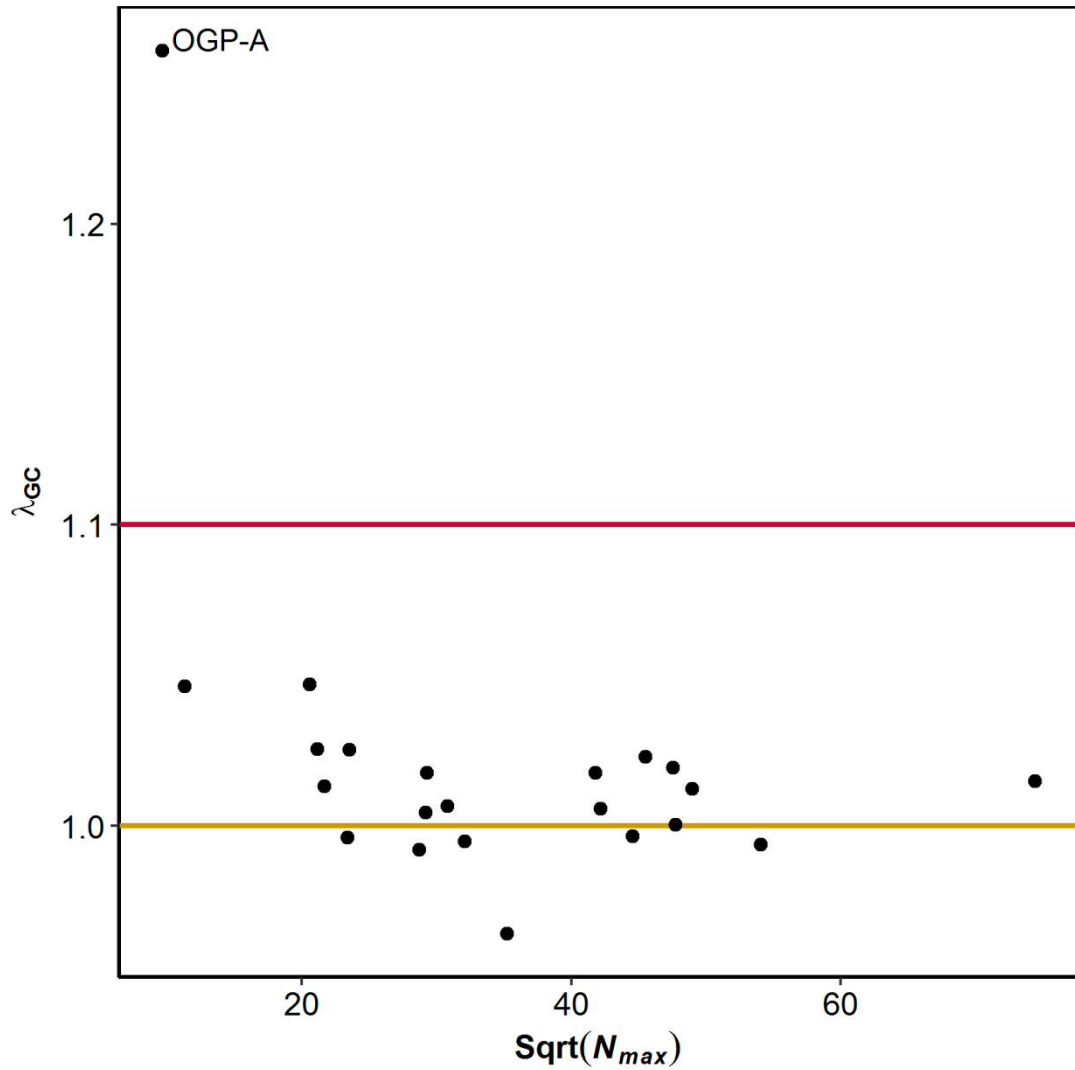
**B3: P-value versus z-score (P-Z) plots for each of the 8 Asian ancestry cohorts.** X-axis:  $-\log_{10} p$ -value computed from Z-statistics for each variant tested; Y-axis:  $-\log_{10} p$ -value reported for each variant tested. NB: Z-statistic for each variant =  $\beta/SE$ . Red line = line of unity (observed p-value = p-value determined from Z-statistics).



**B5: Standard error vs. sample size (SE-N) plot for all 22 cohorts.** Median(SE) = median standard error for variants in the respective study;  $N_{max}$  = Maximum sample size in the respective study. NB: 8.86 = calibration factor based on GWAS chip and 1000 Genomes reference panel for imputation (obtained from Winkler et al. (2014)). With the exception of the TwinsUK study, results from all cohorts follow the same linear relationship between standard error and sample size.



**B6: Lambda-N plot for all 22 cohorts.** Genomic inflation factor ( $\lambda_{GC}$ ) plotted against the square root of the maximum sample size for the cohort ( $N_{max}$ ). Horizontal yellow line ( $\lambda_{GC} = 1.0$ ) denotes optimal (no) genomic inflation; Horizontal red line ( $\lambda_{GC} = 1.1$ ) denotes upper limit of acceptable genomic inflation. NB: the OGP-A study is based on a population isolate and is the smallest cohort in the meta-analysis ( $N = 92$ ).



**Appendix C** MAGMA gene-based and gene-set results from  
*Chapter 4 CREAM Corneal Astigmatism GWAS*  
*Meta-analysis*

C1: The 10 genes demonstrating strongest association from MAGMA gene-based association test. Genes including  $\pm 200$  kb flanking region to gene transcription start and stop sites

**C2-3 The 10 gene-sets demonstrating strongest association from MAGMA gene-set association test**

C2: Genes included  $\pm 50$  kb flanking regions to gene start and stop sites

C3: Genes included  $\pm 200$  kb flanking regions to gene start and stop sites

**C1: The 10 genes demonstrating strongest association from the MAGMA gene-based association test.** Start and stop positions listed include  $\pm 200$  kb flanking regions. nSNPs = number of variants included in gene region; Z-Statistic = gene-based test statistic; P-value = obtained from Z-Statistic under the assumption of a normally distributed model; FDR = false discovery rate. Total number of genes tested = 18,499.

Gene	Chromosome	Start	Stop	nSNPs	Z-Statistic	P-value	FDR	Bonferroni Adjusted P-value
<i>PACSN3</i>	11	46999073	47408010	529	4.48	$3.80 \times 10^{-6}$	0.070	0.070
<i>TNFAIP8L3</i>	15	51148798	51597473	876	4.33	$7.58 \times 10^{-6}$	0.070	0.140
<i>AP4E1</i>	15	51000780	51498097	984	4.07	$2.34 \times 10^{-5}$	0.144	0.433
<i>ARFGAP2</i>	11	46985849	47399054	525	4.00	$3.20 \times 10^{-5}$	0.148	0.593
<i>DDB2</i>	11	47036493	47460769	631	3.86	$5.68 \times 10^{-5}$	0.210	1
<i>C11orf49</i>	11	46758240	47385932	734	3.76	$8.45 \times 10^{-5}$	0.246	1
<i>NEURL4</i>	17	7018947	7432644	749	3.74	$9.31 \times 10^{-5}$	0.246	1
<i>DSE</i>	6	116401231	116962422	1008	3.69	$1.10 \times 10^{-4}$	0.254	1
<i>FBXW2</i>	9	123319254	123755740	567	3.60	$1.58 \times 10^{-4}$	0.325	1
<i>TSPYL1</i>	6	116396022	116801280	706	3.57	$1.81 \times 10^{-4}$	0.335	1



**C2: The 10 gene-sets demonstrating strongest association from MAGMA gene-set association test.** Genes included  $\pm 50$  kb flanking regions added to transcription start and stop sites. nGenes = number of genes included in gene set; Beta = gene-set test statistic; SE = standard error; FDR = false discovery rate; Bonferroni Corrected P-value = P-value multiplied by the number of gene-sets tested. Total number of gene-sets tested = 13,290.

Gene-set	nGenes	Beta	SE	P-value	FDR	Bonferroni Adjusted P-value
NIKOLSKY_BREAST_CANCER_8Q12_Q22_AMPLICON	130	0.36	0.091	5.08E-05	0.675	0.675
JIANG_CORE_DUPLICATION_GENES	7	1.03	0.309	4.32E-04	0.979	1
MYLLYKANGAS_AMPLIFICATION_HOT_SPOT_16	9	0.87	0.264	5.00E-04	0.979	1
V\$AR_01	148	0.17	0.054	7.54E-04	0.979	1
GSE24142_EARLY_THYMIC_PROGENITOR_VS_DN3_THYMOCYTE_UP	195	0.14	0.044	7.75E-04	0.979	1
chr8q21	54	0.40	0.127	8.05E-04	0.979	1
GSE5099_DAY3_VS_DAY7_MCSF_TREATED_MACROPHAGE_DN	160	0.15	0.048	8.15E-04	0.979	1
GSE3982_NEUTROPHIL_VS_TH1_UP	189	0.14	0.045	9.53E-04	0.979	1
AACWWCAANK_UNKNOWN	135	0.18	0.058	1.26E-03	0.979	1
TAGGTCA MIR-192 MIR-215	39	0.32	0.109	1.85E-03	0.979	1

**C3: The 10 gene-sets demonstrating strongest association from MAGMA gene-set association test.** Genes included  $\pm 200$  kb flanking regions added to transcription start and stop sites. nGenes = number of genes included in gene set; Beta = gene-set test statistic; SE = standard error; FDR = false discovery rate; Bonferroni Corrected P = P-value multiplied by the number of gene-sets tested. Total number of gene-sets tested = 13,290.

Gene-set	nGenes	Beta	SE	P-value	FDR	Bonferroni Adjusted P
BIOCARTA_RARRXR_PATHWAY	15	0.36	0.108	4.01E-04	0.953	1
V\$TATA_C	269	0.08	0.025	4.11E-04	0.953	1
NIKOLSKY_BREAST_CANCER_8Q12_Q22_AMPLICON	130	0.25	0.075	4.15E-04	0.953	1
KIM_PTEN_TARGETS_DN	5	0.49	0.145	4.26E-04	0.953	1
MYLLYKANGAS_AMPLIFICATION_HOT_SPOT_16	9	0.51	0.153	4.29E-04	0.953	1
REACTOME_PRE_NOTCH_TRANSCRIPTION_AND_TRANSLATION	26	0.24	0.080	1.33E-03	0.953	1
GSE40068_BCL6_POS_VS_NEG_CXCR5_POS_TFH_DN	190	0.08	0.028	3.13E-03	0.953	1
GSE41978_ID2_KO_VS_ID2_KO_AND_BIM_KO_KLRG1_LOW_EFFECTOR_CD8_TCELL_UP	184	0.08	0.029	3.25E-03	0.953	1
AGGTGCA MIR-500	87	0.12	0.044	3.39E-03	0.953	1
PID_ERB_GENOMIC_PATHWAY	14	0.27	0.098	3.44E-03	0.953	1

D2: Centred on lead variant rs196052 (chromosome 6)

D3: Centred on lead variant rs1129038 (chromosome 15)

D4: Centred on lead variant rs62075722 (chromosome 17)

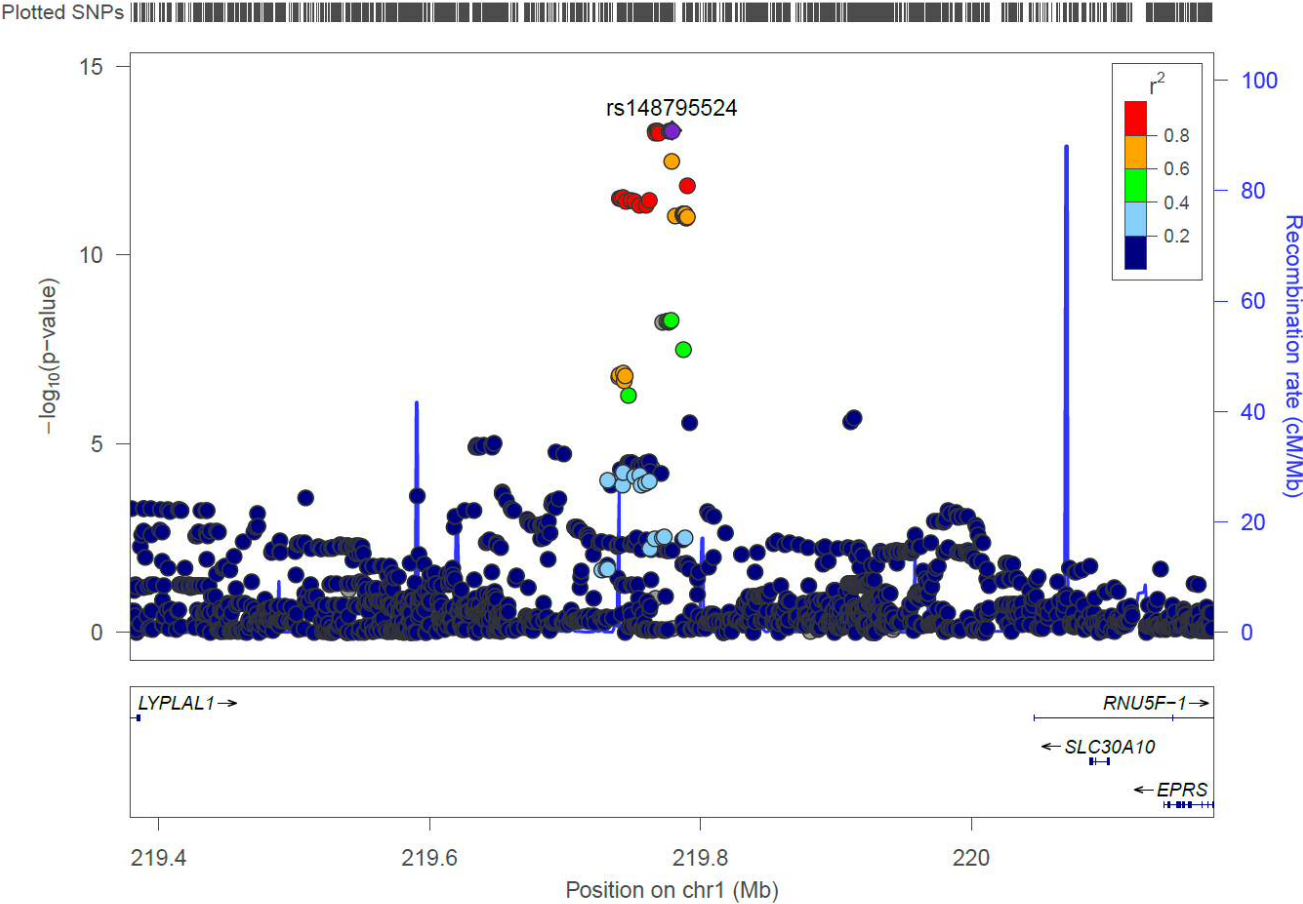
**D5-7: Regional association plots from GWAS for Refractive Astigmatism using  
BOLT-LMM (N = 88,005).**

D5: Centred on lead variant rs12196123 (chromosome 6)

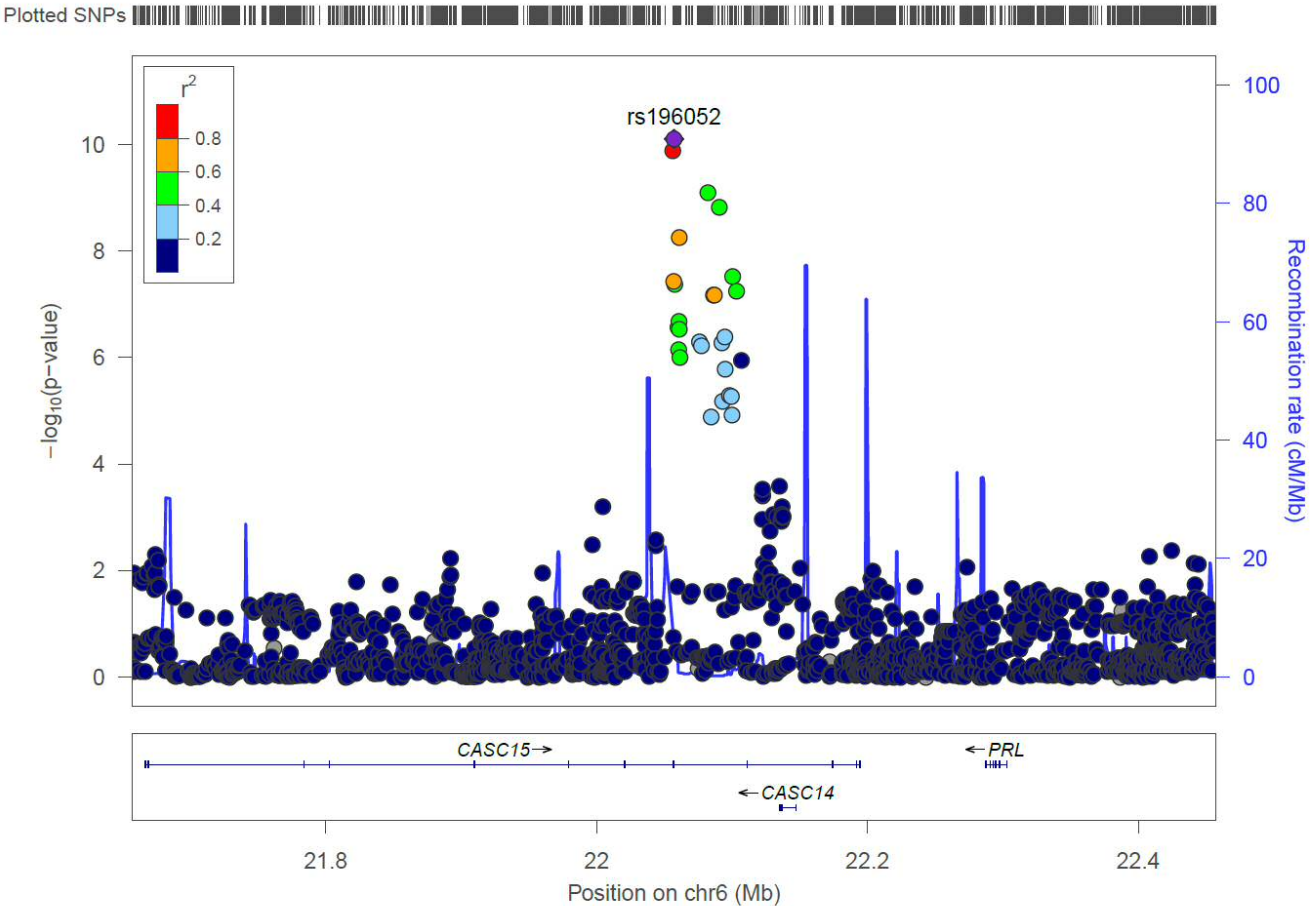
D6: Centred on lead variant rs1129038 (chromosome 15)

D7: Centred on lead variant rs34635363 (chromosome 17)

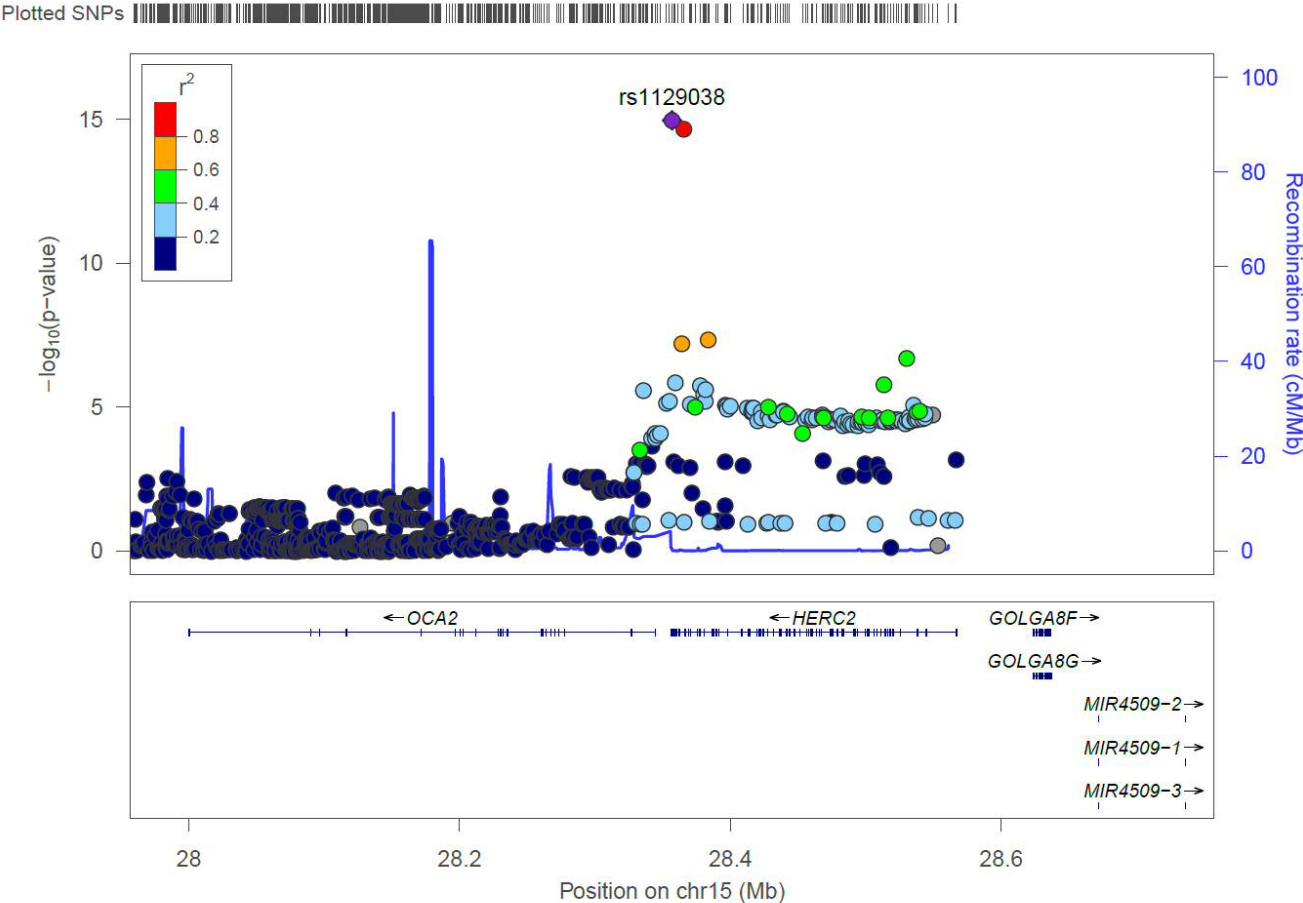
**D1: Regional association plot centred on lead variant rs12032649 (chromosome 1) from GWAS for Corneal Astigmatism using BOLT-LMM (N = 86,335).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple). NB: rs14879552 is a synonym for rs12032649.



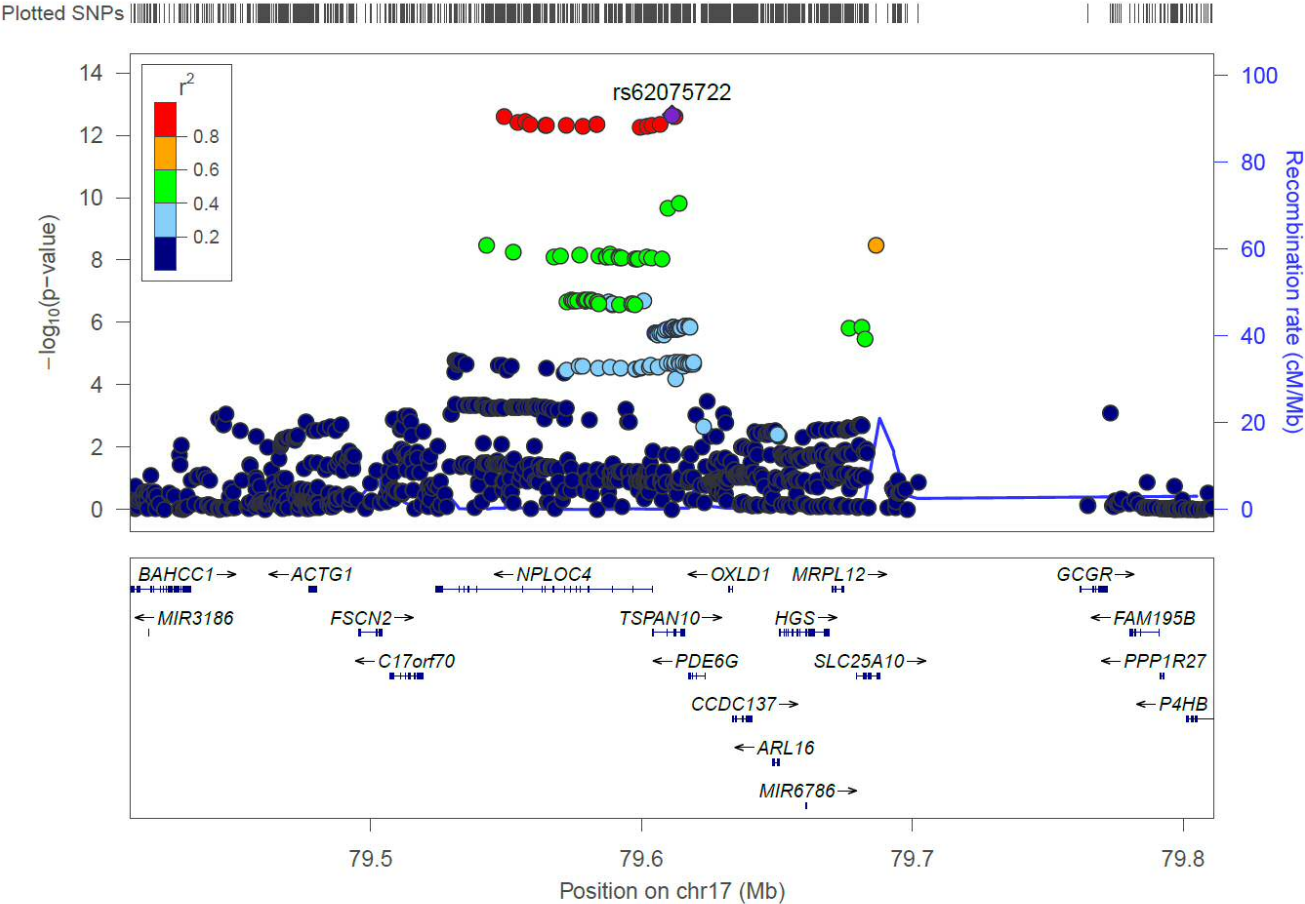
**D2: Regional association plot centred on lead variant rs196052 (chromosome 6) from GWAS for Corneal Astigmatism using BOLT-LMM (N = 86,335).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple).



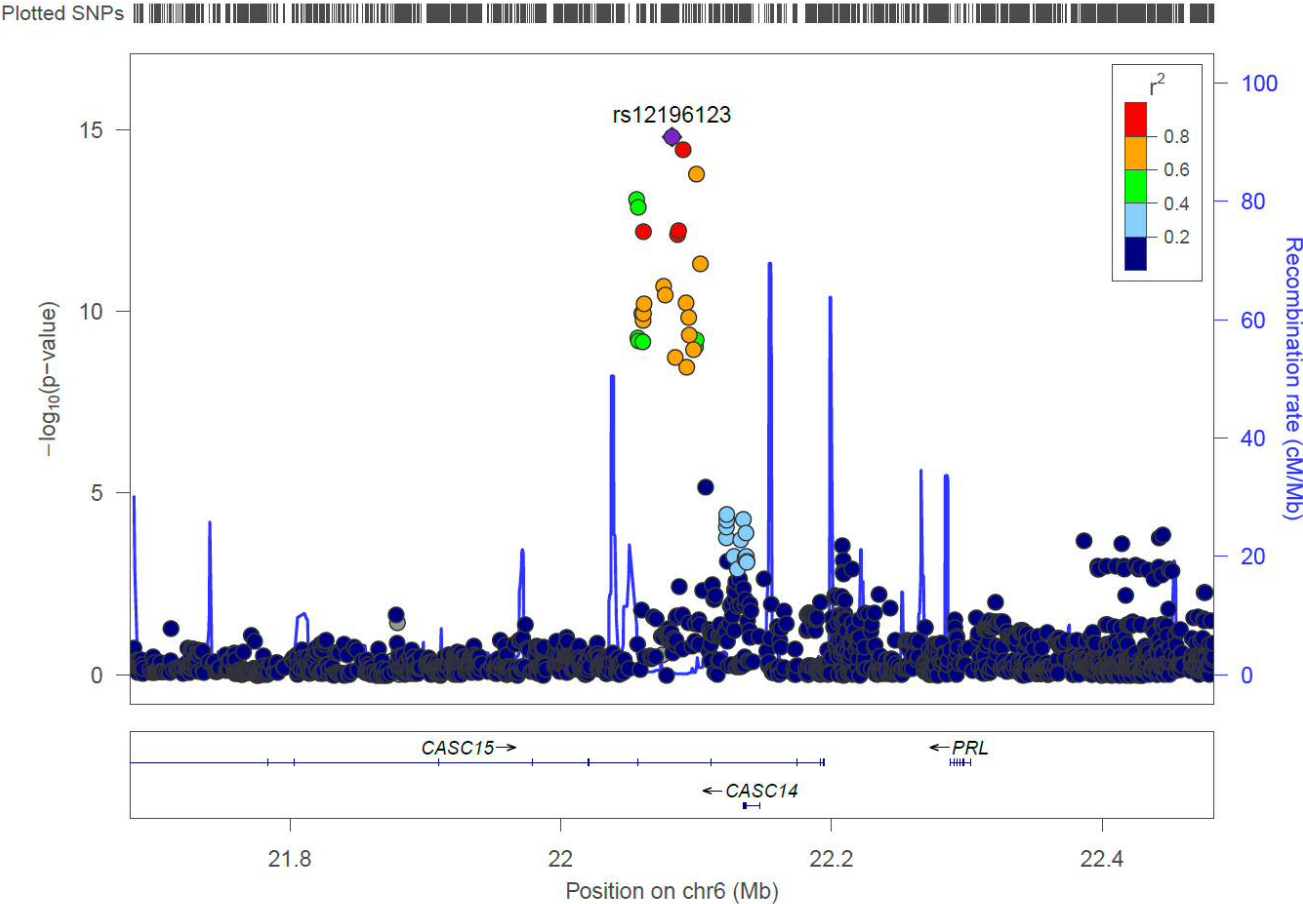
**D3: Regional association plot centred on lead variant rs1129038 (chromosome 15) from GWAS for Corneal Astigmatism using BOLT-LMM (N = 86,335).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple).



**D4: Regional association plot centred on lead variant rs62075722 (chromosome 17) from GWAS for Corneal Astigmatism using BOLT-LMM (N = 86,335).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple).

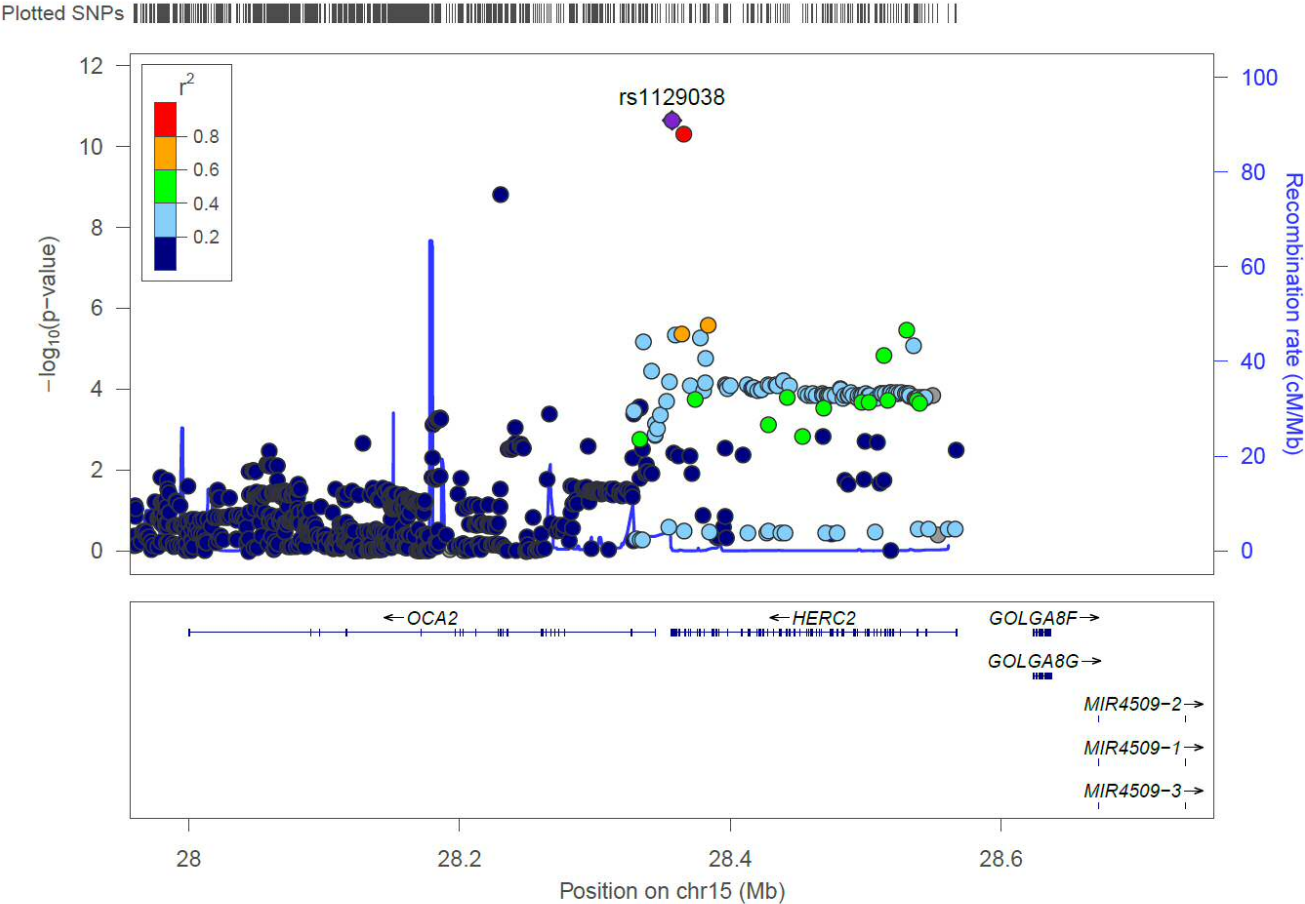


**D5: Regional association plot centred on lead variant rs12196123 (chromosome 6) from GWAS for Refractive Astigmatism using BOLT-LMM (N = 88,005).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple).

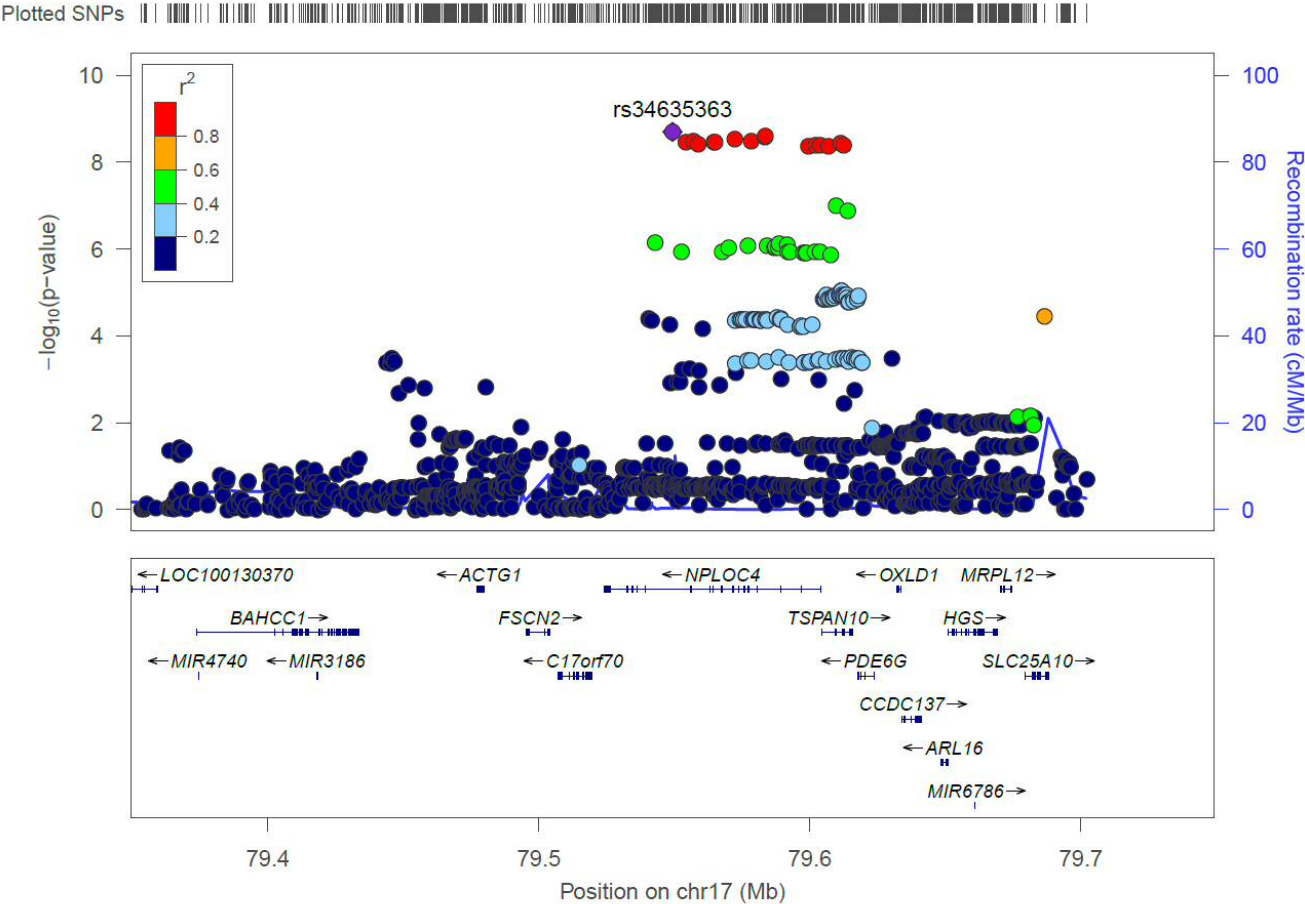




**D6: Regional association plot centred on lead variant rs1129038 (chromosome 15) from GWAS for Refractive Astigmatism using BOLT-LMM (N = 88,005).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple).



**D7: Regional association plot centred on lead variant rs34635363 (chromosome 17) from GWAS for Refractive Astigmatism using BOLT-LMM (N = 88,005).** Symbol colours denote linkage disequilibrium ( $r^2$ ) values of variants with respect to the lead variant (highlighted in purple).



**for refractive error**

E1: Genes included  $\pm 0$  kb,  $\pm 50$  kb or  $\pm 200$  kb flanking regions to gene start and stop sites

E2: Genes included  $\pm 500$  kb,  $\pm 1$  Mb or  $\pm 2$  Mb flanking regions to gene start and stop sites

**E3-4 MAGMA gene-based association test results for the 111 candidate myopia genes using summary statistics from the UK Biobank GWAS for spherical equivalent**

E3: Genes included  $\pm 0$  kb,  $\pm 50$  kb or  $\pm 200$  kb flanking regions to gene start and stop sites

E4: Genes included  $\pm 500$  kb,  $\pm 1$  Mb or  $\pm 2$  Mb flanking regions to gene start and stop sites

**E1: MAGMA gene-based association test results for the 111 candidate myopia genes using data from the CREAM refractive error meta-analysis and the inclusion of 0 kb, 50 kb or 200 kb flanking regions.** MAGMA P = obtained from *Z-Statistic* under the assumption of a normally distributed model; Rank = ranking of the gene out of 111 according to gene-based test P-values; FDR = false discovery rate. Highlighted results = FDR < 0.05. NB: No variants were mapped within 50 kb of *SMN1*.

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>MINOS1</i>	1	19923471	19956315	0.153	34	0.495	0.022	16	0.143	0.046	27	0.188
<i>CEP85</i>	1	26560644	26605529	0.075	22	0.377	0.090	26	0.382	0.276	47	0.651
<i>STK40</i>	1	36805225	36851525	0.379	52	0.753	0.270	45	0.641	0.237	45	0.583
<i>PTPRF</i>	1	43991708	44089343	0.368	50	0.753	0.410	58	0.777	0.355	52	0.751
<i>JAK1</i>	1	65298906	65432593	0.514	68	0.818	0.450	61	0.797	0.712	84	0.940
<i>SERBP1</i>	1	67873493	67896123	0.459	60	0.812	0.690	86	0.873	0.957	107	0.987
<i>ABCD3</i>	1	94883933	94984219	0.337	48	0.753	0.455	62	0.797	0.774	88	0.968
<i>WDR77</i>	1	111982512	111991915	0.841	98	0.940	0.928	105	0.971	0.844	96	0.971
<i>RPRD2</i>	1	150336624	150449042	0.095	26	0.402	0.130	30	0.454	0.134	39	0.383
<i>DENND4B</i>	1	153901977	153919162	0.186	37	0.553	0.261	43	0.641	0.086	35	0.268
<i>HDGF</i>	1	156711899	156722240	0.070	21	0.367	0.044	19	0.255	0.044	26	0.186
<i>SDHC</i>	1	161284166	161334541	0.382	55	0.753	0.428	59	0.797	0.526	64	0.912
<i>CREG1</i>	1	167510250	167523056	0.383	56	0.753	0.801	96	0.918	0.690	80	0.939
<i>KIFAP3</i>	1	169890461	170043882	0.017	10	0.189	0.011	14	0.083	2.74 x 10 <sup>-3</sup>	15	<b>0.020</b>
<i>NUCKS1</i>	1	205681947	205719372	0.946	105	0.978	0.966	107	0.986	0.155	41	0.419
<i>RNF187</i>	1	228674277	228697598	0.556	73	0.820	0.570	73	0.832	0.773	87	0.968
<i>LGALS8</i>	1	236681514	236716281	0.604	79	0.820	0.538	70	0.832	0.780	89	0.968
<i>RAB10</i>	2	26256729	26360323	0.693	88	0.865	0.819	98	0.919	0.683	79	0.939
<i>LBH</i>	2	30454397	30482899	2.27 x 10 <sup>-6</sup>	3	<b>8.32 x 10<sup>-5</sup></b>	5.93 x 10 <sup>-6</sup>	3	<b>2.18 x 10<sup>-4</sup></b>	7.10 x 10 <sup>-4</sup>	9	<b>8.46 x 10<sup>-3</sup></b>
<i>LRPPRC</i>	2	44113363	44223144	0.083	24	0.378	0.137	33	0.455	6.05 x 10 <sup>-3</sup>	17	<b>0.040</b>
<i>ERLEC1</i>	2	54014068	54045956	7.74 x 10 <sup>-3</sup>	8	0.106	1.06 x 10 <sup>-3</sup>	6	<b>0.019</b>	2.86 x 10 <sup>-3</sup>	16	<b>0.020</b>

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CCDC104</i>	2	55746740	55772216	0.057	19	0.330	0.079	24	0.364	0.080	33	0.268
<i>HNRNPA3</i>	2	178076063	178088687	0.315	45	0.753	0.234	39	0.639	2.00 x 10 <sup>-7</sup>	5	4.44 x 10 <sup>-6</sup>
<i>STRADB</i>	2	202316392	202359501	0.593	76	0.820	0.581	75	0.832	0.312	48	0.702
<i>PKI55</i>	2	217081612	217084915	0.786	94	0.919	0.282	48	0.646	0.187	42	0.494
<i>MRPL44</i>	2	224822121	224832431	0.118	29	0.446	0.200	37	0.594	0.316	50	0.702
<i>BTD</i>	3	15642864	15689147	0.020	11	0.200	8.00 x 10 <sup>-3</sup>	13	0.068	0.011	18	0.068
<i>ANKRD28</i>	3	15708743	15901053	0.049	16	0.330	6.74 x 10 <sup>-3</sup>	12	0.062	7.88 x 10 <sup>-9</sup>	2	3.97 x 10 <sup>-7</sup>
<i>TOP2B</i>	3	25639396	25706396	0.494	66	0.818	0.306	50	0.673	0.061	31	0.216
<i>SEC61A1</i>	3	127770402	127790526	0.285	42	0.745	0.274	47	0.641	0.359	53	0.751
<i>ATP11B</i>	3	182511291	182639423	0.308	43	0.753	0.233	38	0.639	0.498	61	0.895
<i>CCNI</i>	4	77968308	77998719	0.480	64	0.812	0.544	71	0.832	0.050	29	0.191
<i>CCNA2</i>	4	122737599	122745153	0.597	77	0.820	0.603	79	0.833	0.882	99	0.977
<i>CCNB1</i>	5	68462837	68474072	0.529	70	0.818	0.439	60	0.797	0.612	72	0.939
<i>SMN1</i>	5	70220768	70248839	-	-	-	-	-	-	0.574	68	0.926
<i>SCAMP1</i>	5	77656339	77776562	0.079	23	0.378	0.106	27	0.428	0.229	44	0.577
<i>ANKS1A</i>	6	34857038	35085802	0.536	72	0.818	0.506	65	0.832	0.597	71	0.933
<i>DST</i>	6	56322785	56819426	0.259	40	0.712	0.322	51	0.689	0.584	70	0.926
<i>EPB41L2</i>	6	131160487	131384462	0.362	49	0.753	0.350	53	0.716	0.500	62	0.895
<i>AKAP12</i>	6	151561134	151679694	0.163	35	0.513	0.182	36	0.557	0.431	58	0.825
<i>VIP</i>	6	153071932	153080900	0.381	54	0.753	0.248	42	0.641	0.037	21	0.180
<i>CBX3</i>	7	26240831	26253227	0.828	97	0.939	0.882	103	0.935	0.693	82	0.939
<i>SEPT7</i>	7	35840596	35946715	0.055	18	0.330	0.053	22	0.264	0.083	34	0.268
<i>CCDC136</i>	7	128431464	128462187	0.801	95	0.927	0.747	92	0.887	0.962	108	0.987
<i>ATP6V1F</i>	7	128502857	128505903	0.381	53	0.753	0.860	100	0.935	0.978	110	0.987
<i>PODXL</i>	7	131185021	131241376	0.863	101	0.940	0.884	104	0.935	0.839	95	0.971

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CNTNAP2</i>	7	145813453	148118090	0.678	87	0.857	0.677	85	0.873	0.641	75	0.939
<i>LINC00599</i>	8	9757574	9760839	0.230	38	0.667	0.268	44	0.641	0.041	25	0.180
<i>BNIP3L</i>	8	26239894	26270644	0.657	85	0.851	0.663	83	0.873	0.785	90	0.968
<i>CLU</i>	8	27454434	27472328	9.23 x 10 <sup>-8</sup>	1	<b>1.02 x 10<sup>-5</sup></b>	4.78 x 10 <sup>-5</sup>	4	<b>1.27 x 10<sup>-3</sup></b>	1.78 x 10 <sup>-3</sup>	13	<b>0.015</b>
<i>CALB1</i>	8	91070836	91095107	0.611	82	0.820	0.517	66	0.832	0.404	55	0.807
<i>NTRK2</i>	9	87283417	87641985	0.044	15	0.324	0.109	28	0.428	0.407	56	0.807
<i>IARS</i>	9	94972489	95056038	0.450	59	0.812	0.563	72	0.832	0.671	77	0.939
<i>WDR31</i>	9	116075502	116102620	0.370	51	0.753	0.536	69	0.832	0.056	30	0.209
<i>CAMSAP1</i>	9	138700333	138799060	0.728	90	0.890	0.357	54	0.716	0.494	60	0.895
<i>GDI2</i>	10	5807186	5855512	0.107	28	0.422	0.358	55	0.716	0.424	57	0.825
<i>CELF2</i>	10	10838851	11378674	0.948	106	0.978	0.936	106	0.971	0.947	104	0.987
<i>COMMD3</i>	10	22605312	22609246	0.536	71	0.818	0.367	56	0.720	0.397	54	0.807
<i>RAB18</i>	10	27793103	27831166	0.052	17	0.330	5.75 x 10 <sup>-3</sup>	10	0.062	7.62 x 10 <sup>-4</sup>	10	<b>8.46 x 10<sup>-3</sup></b>
<i>DDIT4</i>	10	74033677	74035797	2.10 x 10 <sup>-6</sup>	2	<b>8.32 x 10<sup>-5</sup></b>	1.38 x 10 <sup>-7</sup>	2	<b>7.60 x 10<sup>-6</sup></b>	7.40 x 10 <sup>-9</sup>	1	<b>3.97 x 10<sup>-7</sup></b>
<i>ADD3</i>	10	111765627	111895323	0.150	33	0.495	0.141	34	0.455	0.212	43	0.548
<i>SPON1</i>	11	13984184	14289679	6.58 x 10 <sup>-3</sup>	7	0.103	2.77 x 10 <sup>-3</sup>	9	<b>0.034</b>	1.16 x 10 <sup>-3</sup>	11	<b>0.012</b>
<i>CNTF</i>	11	58390146	58393206	4.17 x 10 <sup>-3</sup>	5	0.086	1.89 x 10 <sup>-3</sup>	8	<b>0.026</b>	2.30 x 10 <sup>-3</sup>	14	<b>0.018</b>
<i>MPEG1</i>	11	58975983	58980494	0.479	63	0.812	0.577	74	0.832	0.519	63	0.912
<i>NCAM1</i>	11	112831969	113149158	0.443	57	0.812	0.406	57	0.777	0.568	67	0.926
<i>ARHGEF12</i>	11	120207264	120360645	0.480	65	0.812	0.696	87	0.873	0.889	101	0.977
<i>APLP2</i>	11	129939716	130014706	0.969	109	0.978	0.995	110	0.995	0.956	106	0.987
<i>ACAD8</i>	11	134123428	134135749	0.597	78	0.820	0.529	68	0.832	0.442	59	0.831
<i>WNK1</i>	12	861759	1020618	0.885	103	0.945	0.740	91	0.887	0.691	81	0.939
<i>GABARAPL1</i>	12	10365435	10375727	0.610	81	0.820	0.635	81	0.856	0.644	76	0.939
<i>LYRM5</i>	12	25348150	25357949	0.933	104	0.978	0.977	109	0.986	0.975	109	0.987

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>TMBIM6</i>	12	50135293	50158717	0.878	102	0.945	0.124	29	0.454	0.048	28	0.189
<i>SPRYD3</i>	12	53458100	53473204	0.472	62	0.812	0.473	64	0.814	0.151	40	0.419
<i>CNOT2</i>	12	70636774	70748773	0.811	96	0.929	0.584	76	0.832	0.578	69	0.926
<i>TAOK3</i>	12	118587606	118810750	0.651	84	0.851	0.758	94	0.888	0.849	97	0.971
<i>ZC3H13</i>	13	46528600	46626896	0.781	93	0.919	0.864	101	0.935	0.923	103	0.987
<i>RB1</i>	13	48877883	49056026	0.316	46	0.753	0.272	46	0.641	0.564	66	0.926
<i>PCDH9</i>	13	66876966	67804468	4.67 x 10 <sup>-3</sup>	6	0.086	6.38 x 10 <sup>-3</sup>	11	0.062	0.031	20	0.175
<i>RBM26</i>	13	79885962	79980393	0.023	12	0.210	0.019	15	0.142	0.062	32	0.216
<i>PSMB5</i>	14	23485752	23504429	0.257	39	0.712	0.590	78	0.832	0.885	100	0.977
<i>CNIH1</i>	14	54890279	54908322	0.122	30	0.446	0.132	32	0.454	1.07 x 10 <sup>-8</sup>	3	3.97 x 10 <sup>-7</sup>
<i>SNW1</i>	14	78183942	78227542	0.015	9	0.178	1.71 x 10 <sup>-3</sup>	7	0.026	1.17 x 10 <sup>-4</sup>	7	1.86 x 10 <sup>-3</sup>
<i>PAPOLA</i>	14	96968713	97033453	0.962	107	0.978	0.877	102	0.935	0.996	111	0.996
<i>STRC</i>	15	43891685	44002286	0.746	92	0.892	0.814	97	0.919	0.823	93	0.971
<i>CCNB2</i>	15	59397284	59417244	0.740	91	0.892	0.750	93	0.887	0.675	78	0.939
<i>ANP32A</i>	15	69070874	69113261	0.967	108	0.978	0.589	77	0.832	0.704	83	0.940
<i>IDH3A</i>	15	78441698	78462884	0.449	58	0.812	0.735	90	0.887	0.351	51	0.751
<i>TSR3</i>	16	1399241	1401873	0.609	80	0.820	4.83 x 10 <sup>-8</sup>	1	5.31 x 10 <sup>-6</sup>	1.66 x 10 <sup>-7</sup>	4	4.44 x 10 <sup>-6</sup>
<i>ZNF267</i>	16	31885079	31928629	0.528	69	0.818	0.519	67	0.832	0.314	49	0.702
<i>E2F4</i>	16	67226068	67232821	0.061	20	0.334	0.035	18	0.212	1.12 x 10 <sup>-4</sup>	6	1.86 x 10 <sup>-3</sup>
<i>NFAT5</i>	16	69599869	69738569	0.093	25	0.402	0.065	23	0.309	0.039	22	0.180
<i>COX4I1</i>	16	85833173	85840608	0.669	86	0.856	0.087	25	0.382	0.121	38	0.353
<i>MPRIP</i>	17	16945790	17095962	0.169	36	0.516	0.177	35	0.556	0.618	73	0.939
<i>TVP23B</i>	17	18684458	18710027	0.034	13	0.291	0.048	20	0.260	0.040	23	0.180
<i>NLK</i>	17	26369009	26523407	0.043	14	0.324	0.130	31	0.454	0.249	46	0.602
<i>FAM134C</i>	17	40731526	40761445	0.268	41	0.719	0.236	40	0.639	0.040	24	0.180

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>ATP5G1</i>	17	46970148	46973233	0.700	89	0.865	0.698	88	0.873	0.749	86	0.966
<i>COX11</i>	17	53029259	53046064	0.312	44	0.753	0.456	63	0.797	0.720	85	0.940
<i>DGKE</i>	17	54911460	54946036	0.462	61	0.812	0.238	41	0.639	1.55 x 10 <sup>-3</sup>	12	<b>0.014</b>
<i>SEPT4</i>	17	56597611	56618179	5.72 x 10 <sup>-6</sup>	4	<b>1.57 x 10<sup>-4</sup></b>	5.76 x 10 <sup>-5</sup>	5	<b>1.27 x 10<sup>-3</sup></b>	3.85 x 10 <sup>-4</sup>	8	<b>5.34 x 10<sup>-3</sup></b>
<i>DHX40</i>	17	57642886	57685713	0.135	32	0.465	0.050	21	0.260	0.095	37	0.286
<i>DDX5</i>	17	62494374	62503042	0.105	27	0.422	0.296	49	0.664	0.539	65	0.921
<i>PRKAR1A</i>	17	66507921	66547457	0.134	31	0.465	0.022	17	0.143	0.030	19	0.175
<i>RBBP8</i>	18	20513295	20606451	0.979	110	0.979	0.969	108	0.986	0.803	92	0.969
<i>OSBPL1A</i>	18	21742009	21977846	0.641	83	0.849	0.792	95	0.917	0.798	91	0.969
<i>RAX2</i>	19	3769087	3772219	0.560	74	0.820	0.835	99	0.928	0.920	102	0.987
<i>CLIP3</i>	19	36505562	36523797	0.512	67	0.818	0.638	82	0.856	0.834	94	0.971
<i>SNAP25</i>	20	10199477	10288068	0.858	100	0.940	0.720	89	0.887	0.953	105	0.987
<i>SYS1</i>	20	43988660	44005442	0.323	47	0.753	0.606	80	0.833	0.869	98	0.977
<i>KCNB1</i>	20	47988505	48099181	0.592	75	0.820	0.676	84	0.873	0.641	74	0.939
<i>OSBPL2</i>	20	60813541	60871269	0.849	99	0.940	0.326	52	0.689	0.087	36	0.268



**E2: MAGMA gene-based association test results for the 111 candidate myopia genes using data from the CREAM refractive error meta-analysis and the inclusion of 500 kb, 1 Mb or 2 Mb flanking regions.** MAGMA P = obtained from *Z-Statistic* under the assumption of a normally distributed model; Rank = ranking of the gene out of 111 according to gene-based test P-values; FDR = false discovery rate. Highlighted results = FDR < 0.05

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>MINOS1</i>	1	19923471	19956315	0.159	49	0.356	0.154	59	0.289	0.033	49	0.075
<i>CEP85</i>	1	26560644	26605529	0.475	74	0.712	0.609	94	0.712	0.633	101	0.696
<i>STK40</i>	1	36805225	36851525	2.70 x 10 <sup>-3</sup>	18	<b>0.017</b>	3.75 x 10 <sup>-3</sup>	24	<b>0.017</b>	0.268	82	0.362
<i>PTPRF</i>	1	43991708	44089343	0.242	56	0.479	0.313	79	0.439	6.27 x 10 <sup>-3</sup>	37	<b>0.018</b>
<i>JAK1</i>	1	65298906	65432593	0.048	33	0.156	0.217	66	0.361	4.95 x 10 <sup>-3</sup>	34	<b>0.016</b>
<i>SERBP1</i>	1	67873493	67896123	0.302	61	0.549	0.018	33	0.059	3.49 x 10 <sup>-3</sup>	33	<b>0.012</b>
<i>ABCD3</i>	1	94883933	94984219	0.014	24	0.065	0.117	54	0.241	0.100	61	0.182
<i>WDR77</i>	1	111982512	111991915	0.813	100	0.897	0.090	50	0.200	6.36 x 10 <sup>-3</sup>	38	<b>0.018</b>
<i>RPRD2</i>	1	150336624	150449042	0.210	52	0.449	0.275	72	0.414	0.278	84	0.365
<i>DENND4B</i>	1	153901977	153919162	0.067	39	0.191	0.039	42	0.102	6.06 x 10 <sup>-5</sup>	18	<b>3.74 x 10<sup>-4</sup></b>
<i>HDGF</i>	1	156711899	156722240	0.061	37	0.183	9.74 x 10 <sup>-3</sup>	27	<b>0.038</b>	4.42 x 10 <sup>-6</sup>	13	<b>3.77 x 10<sup>-5</sup></b>
<i>SDHC</i>	1	161284166	161334541	0.788	97	0.897	0.726	99	0.814	0.707	105	0.748
<i>CREG1</i>	1	167510250	167523056	0.825	102	0.897	0.305	77	0.435	0.518	96	0.599
<i>KIFAP3</i>	1	169890461	170043882	4.93 x 10 <sup>-4</sup>	12	<b>4.56 x 10<sup>-3</sup></b>	1.16 x 10 <sup>-4</sup>	14	<b>9.23 x 10<sup>-4</sup></b>	1.12 x 10 <sup>-3</sup>	26	<b>4.80 x 10<sup>-3</sup></b>
<i>NUCKS1</i>	1	205681947	205719372	0.141	48	0.327	0.083	48	0.189	3.63 x 10 <sup>-4</sup>	24	<b>1.68 x 10<sup>-3</sup></b>
<i>RNF187</i>	1	228674277	228697598	0.946	109	0.963	0.381	85	0.497	0.011	42	<b>0.030</b>
<i>LGALS8</i>	1	236681514	236716281	0.510	77	0.735	0.518	91	0.632	1.55 x 10 <sup>-3</sup>	28	<b>6.14 x 10<sup>-3</sup></b>
<i>RAB10</i>	2	26256729	26360323	0.679	85	0.886	0.037	40	0.102	6.38 x 10 <sup>-3</sup>	39	<b>0.018</b>
<i>LBH</i>	2	30454397	30482899	0.106	44	0.267	0.106	52	0.227	0.106	63	0.186
<i>LRPPRC</i>	2	44113363	44223144	1.26 x 10 <sup>-4</sup>	10	<b>1.40 x 10<sup>-3</sup></b>	1.72 x 10 <sup>-7</sup>	6	<b>3.18 x 10<sup>-6</sup></b>	1.70 x 10 <sup>-5</sup>	16	<b>1.18 x 10<sup>-4</sup></b>
<i>ERLEC1</i>	2	54014068	54045956	4.25 x 10 <sup>-3</sup>	19	<b>0.025</b>	0.012	30	<b>0.043</b>	1.66 x 10 <sup>-3</sup>	30	<b>6.14 x 10<sup>-3</sup></b>

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CCDC104</i>	2	55746740	55772216	$5.20 \times 10^{-5}$	8	<b><math>7.21 \times 10^{-4}</math></b>	$8.68 \times 10^{-4}$	20	<b><math>4.82 \times 10^{-3}</math></b>	$2.95 \times 10^{-4}$	22	<b><math>1.49 \times 10^{-3}</math></b>
<i>HNRNPA3</i>	2	178076063	178088687	$7.32 \times 10^{-10}$	2	<b><math>4.06 \times 10^{-8}</math></b>	$1.07 \times 10^{-23}$	1	<b><math>1.19 \times 10^{-21}</math></b>	$3.75 \times 10^{-20}$	1	<b><math>4.16 \times 10^{-18}</math></b>
<i>STRADB</i>	2	202316392	202359501	0.047	30	0.156	$5.84 \times 10^{-4}$	18	<b><math>3.60 \times 10^{-3}</math></b>	$1.53 \times 10^{-4}$	20	<b><math>8.50 \times 10^{-4}</math></b>
<i>PKI55</i>	2	217081612	217084915	0.125	46	0.302	0.281	74	0.414	0.270	83	0.362
<i>MRPL44</i>	2	224822121	224832431	0.048	32	0.156	0.083	49	0.189	0.383	92	0.458
<i>BTD</i>	3	15642864	15689147	$6.22 \times 10^{-8}$	3	<b><math>2.30 \times 10^{-6}</math></b>	$1.65 \times 10^{-5}$	9	<b><math>2.03 \times 10^{-4}</math></b>	$5.84 \times 10^{-3}$	35	<b>0.018</b>
<i>ANKRD28</i>	3	15708743	15901053	$1.14 \times 10^{-7}$	4	<b><math>2.95 \times 10^{-6}</math></b>	$8.45 \times 10^{-5}$	12	<b><math>7.82 \times 10^{-4}</math></b>	$3.04 \times 10^{-3}$	32	<b>0.011</b>
<i>TOP2B</i>	3	25639396	25706396	$5.90 \times 10^{-3}$	21	<b>0.031</b>	$3.61 \times 10^{-3}$	23	<b>0.017</b>	$1.46 \times 10^{-3}$	27	<b><math>6.01 \times 10^{-3}</math></b>
<i>SEC61A1</i>	3	127770402	127790526	0.047	31	0.156	$9.94 \times 10^{-3}$	29	<b>0.038</b>	$2.57 \times 10^{-3}$	31	<b><math>9.19 \times 10^{-3}</math></b>
<i>ATP11B</i>	3	182511291	182639423	0.594	82	0.804	0.745	100	0.827	0.536	98	0.606
<i>CCNI</i>	4	77968308	77998719	0.326	62	0.584	0.326	80	0.448	0.286	86	0.369
<i>CCNA2</i>	4	122737599	122745153	0.779	95	0.897	0.327	81	0.448	0.226	77	0.326
<i>CCNB1</i>	5	68462837	68474072	0.816	101	0.897	0.818	106	0.857	0.461	94	0.543
<i>SMN1</i>	5	70220768	70248839	0.792	98	0.897	0.588	93	0.701	$3.89 \times 10^{-5}$	17	<b><math>2.54 \times 10^{-4}</math></b>
<i>SCAMP1</i>	5	77656339	77776562	0.044	29	0.156	0.016	32	0.056	0.058	55	0.117
<i>ANKS1A</i>	6	34857038	35085802	0.520	78	0.736	0.764	101	0.836	0.152	71	0.236
<i>DST</i>	6	56322785	56819426	0.437	72	0.672	0.415	88	0.520	0.027	47	0.063
<i>EPB41L2</i>	6	131160487	131384462	0.724	88	0.886	0.042	43	0.108	$8.17 \times 10^{-10}$	5	<b><math>1.81 \times 10^{-8}</math></b>
<i>AKAP12</i>	6	151561134	151679694	$1.98 \times 10^{-3}$	14	<b>0.016</b>	0.073	45	0.177	0.136	69	0.219
<i>VIP</i>	6	153071932	153080900	0.110	45	0.270	0.190	64	0.329	0.038	50	0.084
<i>CBX3</i>	7	26240831	26253227	0.572	81	0.783	0.856	109	0.872	0.934	111	0.934
<i>SEPT7</i>	7	35840596	35946715	0.065	38	0.189	0.100	51	0.218	0.107	64	0.186
<i>CCDC136</i>	7	128431464	128462187	0.724	89	0.886	0.221	68	0.361	0.048	52	0.102
<i>ATP6V1F</i>	7	128502857	128505903	0.695	86	0.886	0.280	73	0.414	0.050	53	0.104
<i>PODXL</i>	7	131185021	131241376	0.375	66	0.631	0.609	95	0.712	0.592	100	0.657

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CNTNAP2</i>	7	145813453	148118090	0.734	92	0.886	0.811	105	0.857	0.726	106	0.761
<i>LINC00599</i>	8	9757574	9760839	$2.48 \times 10^{-3}$	16	<b>0.017</b>	$2.15 \times 10^{-6}$	8	<b><math>2.98 \times 10^{-5}</math></b>	$2.82 \times 10^{-6}$	11	<b><math>2.85 \times 10^{-5}</math></b>
<i>BNIP3L</i>	8	26239894	26270644	0.873	103	0.932	0.768	102	0.836	0.384	93	0.458
<i>CLU</i>	8	27454434	27472328	0.049	35	0.156	0.305	78	0.435	0.305	87	0.389
<i>CALB1</i>	8	91070836	91095107	0.277	59	0.521	0.133	57	0.258	0.238	80	0.327
<i>NTRK2</i>	9	87283417	87641985	0.707	87	0.886	0.787	104	0.840	0.835	109	0.850
<i>IARS</i>	9	94972489	95056038	0.499	76	0.729	0.357	83	0.477	0.338	90	0.416
<i>WDR31</i>	9	116075502	116102620	0.161	50	0.356	0.283	76	0.414	0.345	91	0.421
<i>CAMSAP1</i>	9	138700333	138799060	0.730	90	0.886	0.830	107	0.861	0.333	89	0.416
<i>GDI2</i>	10	5807186	5855512	0.224	55	0.451	0.846	108	0.870	0.799	108	0.822
<i>CELF2</i>	10	10838851	11378674	0.742	93	0.886	0.777	103	0.837	0.746	107	0.774
<i>COMMD3</i>	10	22605312	22609246	0.132	47	0.312	0.396	87	0.505	0.689	103	0.743
<i>RAB18</i>	10	27793103	27831166	$9.82 \times 10^{-3}$	23	<b>0.047</b>	0.129	56	0.256	0.657	102	0.714
<i>DDIT4</i>	10	74033677	74035797	$1.33 \times 10^{-7}$	5	<b><math>2.95 \times 10^{-6}</math></b>	$6.82 \times 10^{-7}$	7	<b><math>1.08 \times 10^{-5}</math></b>	$8.21 \times 10^{-6}$	15	<b><math>6.07 \times 10^{-5}</math></b>
<i>ADD3</i>	10	111765627	111895323	0.437	71	0.672	0.219	67	0.361	0.154	72	0.236
<i>SPON1</i>	11	13984184	14289679	$4.54 \times 10^{-3}$	20	<b>0.025</b>	$7.02 \times 10^{-3}$	26	<b>0.030</b>	0.117	66	0.197
<i>CNTF</i>	11	58390146	58393206	$8.51 \times 10^{-5}$	9	<b><math>1.05 \times 10^{-3}</math></b>	$3.80 \times 10^{-4}$	17	<b><math>2.48 \times 10^{-3}</math></b>	$3.14 \times 10^{-4}$	23	<b><math>1.51 \times 10^{-3}</math></b>
<i>MPEG1</i>	11	58975983	58980494	0.395	68	0.639	$2.34 \times 10^{-4}$	15	<b><math>1.73 \times 10^{-3}</math></b>	$1.23 \times 10^{-4}$	19	<b><math>7.19 \times 10^{-4}</math></b>
<i>NCAM1</i>	11	112831969	113149158	0.368	64	0.631	0.066	44	0.167	0.024	45	0.059
<i>ARHGEF12</i>	11	120207264	120360645	0.938	108	0.963	0.586	92	0.701	0.124	67	0.205
<i>APLP2</i>	11	129939716	130014706	0.886	105	0.932	$5.30 \times 10^{-3}$	25	<b>0.024</b>	$4.73 \times 10^{-11}$	2	<b><math>2.62 \times 10^{-9}</math></b>
<i>ACAD8</i>	11	134123428	134135749	0.391	67	0.639	0.176	62	0.316	0.128	68	0.208
<i>WNK1</i>	12	861759	1020618	0.078	41	0.207	0.196	65	0.334	0.111	65	0.190
<i>GABARAPL1</i>	12	10365435	10375727	0.432	70	0.672	$7.50 \times 10^{-4}$	19	<b><math>4.38 \times 10^{-3}</math></b>	$3.43 \times 10^{-6}$	12	<b><math>3.17 \times 10^{-5}</math></b>
<i>LYRM5</i>	12	25348150	25357949	0.890	106	0.932	0.375	84	0.496	0.103	62	0.184

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>TMBIM6</i>	12	50135293	50158717	0.040	28	0.156	0.031	35	0.096	0.076	57	0.146
<i>SPRYD3</i>	12	53458100	53473204	0.264	57	0.515	0.474	90	0.584	0.531	97	0.606
<i>CNOT2</i>	12	70636774	70748773	0.301	60	0.549	0.038	41	0.102	0.160	74	0.240
<i>TAOK3</i>	12	118587606	118810750	0.081	43	0.210	0.163	60	0.300	0.098	60	0.180
<i>ZC3H13</i>	13	46528600	46626896	0.362	63	0.631	3.62 x 10 <sup>-4</sup>	16	<b>2.48 x 10<sup>-3</sup></b>	6.97 x 10 <sup>-3</sup>	40	<b>0.019</b>
<i>RB1</i>	13	48877883	49056026	0.882	104	0.932	0.956	111	0.956	0.015	44	<b>0.037</b>
<i>PCDH9</i>	13	66876966	67804468	0.070	40	0.193	0.079	47	0.186	0.086	59	0.161
<i>RBM26</i>	13	79885962	79980393	0.031	27	0.129	0.033	38	0.096	0.236	79	0.327
<i>PSMB5</i>	14	23485752	23504429	0.983	111	0.983	0.649	96	0.747	0.705	104	0.748
<i>CNIH1</i>	14	54890279	54908322	3.42 x 10 <sup>-19</sup>	1	<b>3.80 x 10<sup>-17</sup></b>	5.05 x 10 <sup>-14</sup>	2	<b>2.80 x 10<sup>-12</sup></b>	6.09 x 10 <sup>-10</sup>	4	<b>1.69 x 10<sup>-8</sup></b>
<i>SNW1</i>	14	78183942	78227542	2.91 x 10 <sup>-5</sup>	7	<b>4.61 x 10<sup>-4</sup></b>	1.13 x 10 <sup>-4</sup>	13	<b>9.23 x 10<sup>-4</sup></b>	1.64 x 10 <sup>-3</sup>	29	<b>6.14 x 10<sup>-3</sup></b>
<i>PAPOLA</i>	14	96968713	97033453	0.968	110	0.976	0.129	55	0.256	0.318	88	0.401
<i>STRC</i>	15	43891685	44002286	0.927	107	0.962	0.913	110	0.922	0.881	110	0.889
<i>CCNB2</i>	15	59397284	59417244	0.781	96	0.897	0.653	97	0.747	0.236	78	0.327
<i>ANP32A</i>	15	69070874	69113261	0.672	84	0.886	0.109	53	0.228	6.23 x 10 <sup>-3</sup>	36	<b>0.018</b>
<i>IDH3A</i>	15	78441698	78462884	0.221	54	0.451	3.19 x 10 <sup>-12</sup>	3	<b>1.18 x 10<sup>-10</sup></b>	5.11 x 10 <sup>-10</sup>	3	<b>1.69 x 10<sup>-8</sup></b>
<i>TSR3</i>	16	1399241	1401873	3.09 x 10 <sup>-4</sup>	11	<b>3.12 x 10<sup>-3</sup></b>	3.42 x 10 <sup>-5</sup>	11	<b>3.45 x 10<sup>-4</sup></b>	5.67 x 10 <sup>-6</sup>	14	<b>4.50 x 10<sup>-5</sup></b>
<i>ZNF267</i>	16	31885079	31928629	0.628	83	0.840	0.187	63	0.329	0.178	76	0.260
<i>E2F4</i>	16	67226068	67232821	2.25 x 10 <sup>-6</sup>	6	<b>4.17 x 10<sup>-5</sup></b>	5.47 x 10 <sup>-9</sup>	4	<b>1.52 x 10<sup>-7</sup></b>	7.60 x 10 <sup>-7</sup>	9	<b>9.37 x 10<sup>-6</sup></b>
<i>NFAT5</i>	16	69599869	69738569	0.015	25	0.066	0.033	36	0.096	2.23 x 10 <sup>-6</sup>	10	<b>2.48 x 10<sup>-5</sup></b>
<i>COX4I1</i>	16	85833173	85840608	0.273	58	0.521	0.268	70	0.414	0.014	43	<b>0.036</b>
<i>MPRIP</i>	17	16945790	17095962	0.397	69	0.639	0.036	39	0.102	0.029	48	0.068
<i>TVP23B</i>	17	18684458	18710027	0.078	42	0.207	3.31 x 10 <sup>-3</sup>	22	<b>0.017</b>	0.051	54	0.105
<i>NLK</i>	17	26369009	26523407	0.369	65	0.631	0.073	46	0.177	0.025	46	0.060
<i>FAM134C</i>	17	40731526	40761445	2.66 x 10 <sup>-3</sup>	17	<b>0.017</b>	0.026	34	0.086	0.075	56	0.146

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>ATP5G1</i>	17	46970148	46973233	0.021	26	0.091	0.033	37	0.096	0.076	58	0.146
<i>COX11</i>	17	53029259	53046064	0.731	91	0.886	0.714	98	0.809	0.147	70	0.234
<i>DGKE</i>	17	54911460	54946036	$8.22 \times 10^{-3}$	22	<b>0.041</b>	$9.68 \times 10^{-8}$	5	<b><math>2.15 \times 10^{-6}</math></b>	$1.71 \times 10^{-7}$	7	<b><math>2.71 \times 10^{-6}</math></b>
<i>SEPT4</i>	17	56597611	56618179	$2.26 \times 10^{-3}$	15	<b>0.017</b>	$2.49 \times 10^{-5}$	10	<b><math>2.77 \times 10^{-4}</math></b>	$1.38 \times 10^{-8}$	6	<b><math>2.55 \times 10^{-7}</math></b>
<i>DHX40</i>	17	57642886	57685713	0.060	36	0.183	0.015	31	0.054	$6.56 \times 10^{-7}$	8	<b><math>9.10 \times 10^{-6}</math></b>
<i>DDX5</i>	17	62494374	62503042	0.492	75	0.728	0.390	86	0.504	$3.81 \times 10^{-4}$	25	<b><math>1.69 \times 10^{-3}</math></b>
<i>PRKAR1A</i>	17	66507921	66547457	$1.16 \times 10^{-3}$	13	<b><math>9.87 \times 10^{-3}</math></b>	$2.36 \times 10^{-3}$	21	<b>0.012</b>	$8.69 \times 10^{-3}$	41	<b>0.024</b>
<i>RBBP8</i>	18	20513295	20606451	0.801	99	0.897	0.282	75	0.414	0.239	81	0.327
<i>OSBPL1A</i>	18	21742009	21977846	0.442	73	0.672	0.258	69	0.414	0.540	99	0.606
<i>RAX2</i>	19	3769087	3772219	0.048	34	0.156	0.341	82	0.462	0.039	51	0.084
<i>CLIP3</i>	19	36505562	36523797	0.221	53	0.451	0.417	89	0.520	0.279	85	0.365
<i>SNAP25</i>	20	10199477	10288068	0.754	94	0.890	0.135	58	0.258	0.155	73	0.236
<i>SYS1</i>	20	43988660	44005442	0.531	80	0.736	0.165	61	0.300	0.465	95	0.543
<i>KCNB1</i>	20	47988505	48099181	0.526	79	0.736	0.273	71	0.414	0.172	75	0.254
<i>OSBPL2</i>	20	60813541	60871269	0.201	51	0.437	$9.85 \times 10^{-3}$	28	<b>0.038</b>	$1.95 \times 10^{-4}$	21	<b><math>1.03 \times 10^{-3}</math></b>

**E3: MAGMA gene-based association test results for the 111 candidate myopia genes using data from the UK Biobank spherical equivalent GWAS and the inclusion of 0 kb, 50 kb or 200 kb flanking regions.** MAGMA P = obtained from *Z-Statistic* under the assumption of a normally distributed model; Rank = ranking of the gene out of 111 according to gene-based test P-values; FDR = false discovery rate. Highlighted results = FDR < 0.05. NB: No variants were mapped within the *COMMD3* gene or within 200 kb of *SMN1*.

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>MINOS1</i>	1	19923471	19956315	0.400	62	0.657	0.476	74	0.700	0.736	102	0.793
<i>CEP85</i>	1	26560644	26605529	0.688	90	0.834	0.517	79	0.713	0.122	51	0.263
<i>STK40</i>	1	36805225	36851525	0.352	57	0.657	0.173	48	0.396	0.279	69	0.445
<i>PTPRF</i>	1	43991708	44089343	3.57 x 10 <sup>-4</sup>	10	<b>3.89 x 10<sup>-3</sup></b>	3.95 x 10 <sup>-4</sup>	10	<b>4.34 x 10<sup>-3</sup></b>	1.36 x 10 <sup>-4</sup>	10	<b>1.41 x 10<sup>-3</sup></b>
<i>JAK1</i>	1	65298906	65432593	0.737	91	0.871	0.789	101	0.860	0.946	110	0.946
<i>SERBP1</i>	1	67873493	67896123	0.653	88	0.808	0.725	96	0.818	4.55 x 10 <sup>-3</sup>	22	<b>0.023</b>
<i>ABCD3</i>	1	94883933	94984219	0.608	85	0.780	0.699	94	0.818	0.695	99	0.772
<i>WDR77</i>	1	111982512	111991915	0.786	97	0.883	0.803	102	0.866	0.294	72	0.449
<i>RPRD2</i>	1	150336624	150449042	1.25 x 10 <sup>-4</sup>	8	<b>1.71 x 10<sup>-3</sup></b>	2.60 x 10 <sup>-4</sup>	9	<b>3.18 x 10<sup>-3</sup></b>	7.18 x 10 <sup>-5</sup>	9	<b>8.78 x 10<sup>-4</sup></b>
<i>DENND4B</i>	1	153901977	153919162	0.125	36	0.378	0.152	45	0.371	0.206	62	0.366
<i>HDGF</i>	1	156711899	156722240	3.02 x 10 <sup>-5</sup>	6	<b>5.48 x 10<sup>-4</sup></b>	3.48 x 10 <sup>-3</sup>	18	<b>0.021</b>	0.015	31	0.054
<i>SDHC</i>	1	161284166	161334541	0.901	105	0.932	0.922	109	0.930	0.662	96	0.751
<i>CREG1</i>	1	167510250	167523056	0.372	60	0.657	0.832	104	0.880	0.852	105	0.881
<i>KIFAP3</i>	1	169890461	170043882	0.596	83	0.780	0.525	80	0.713	0.466	87	0.583
<i>NUCKS1</i>	1	205681947	205719372	0.607	84	0.780	0.494	77	0.700	0.050	42	0.131
<i>RNF187</i>	1	228674277	228697598	0.454	70	0.695	0.412	72	0.630	0.214	63	0.373
<i>LGALS8</i>	1	236681514	236716281	0.026	23	0.123	3.68 x 10 <sup>-3</sup>	19	<b>0.021</b>	2.23 x 10 <sup>-3</sup>	20	<b>0.012</b>
<i>RAB10</i>	2	26256729	26360323	0.261	48	0.592	0.254	60	0.465	0.060	44	0.146
<i>LBH</i>	2	30454397	30482899	6.35 x 10 <sup>-8</sup>	2	<b>3.46 x 10<sup>-6</sup></b>	3.49 x 10 <sup>-6</sup>	5	<b>7.67 x 10<sup>-5</sup></b>	2.06 x 10 <sup>-4</sup>	14	<b>1.62 x 10<sup>-3</sup></b>
<i>LRPPRC</i>	2	44113363	44223144	2.39 x 10 <sup>-5</sup>	5	<b>5.21 x 10<sup>-4</sup></b>	3.38 x 10 <sup>-6</sup>	4	<b>7.67 x 10<sup>-5</sup></b>	2.14 x 10 <sup>-8</sup>	3	<b>6.16 x 10<sup>-7</sup></b>
<i>ERLEC1</i>	2	54014068	54045956	5.05 x 10 <sup>-4</sup>	11	<b>4.87 x 10<sup>-3</sup></b>	2.23 x 10 <sup>-4</sup>	7	<b>3.18 x 10<sup>-3</sup></b>	3.57 x 10 <sup>-5</sup>	8	<b>4.91 x 10<sup>-4</sup></b>

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CCDC104</i>	2	55746740	55772216	5.05 x 10 <sup>-5</sup>	7	<b>7.86 x 10<sup>-4</sup></b>	5.81 x 10 <sup>-5</sup>	6	<b>1.06 x 10<sup>-3</sup></b>	2.46 x 10 <sup>-4</sup>	15	<b>1.80 x 10<sup>-3</sup></b>
<i>HNRNPA3</i>	2	178076063	178088687	0.022	20	0.118	1.63 x 10 <sup>-3</sup>	13	<b>0.014</b>	1.54 x 10 <sup>-4</sup>	12	<b>1.41 x 10<sup>-3</sup></b>
<i>STRADB</i>	2	202316392	202359501	0.186	41	0.493	0.060	34	0.193	0.262	66	0.434
<i>PKI55</i>	2	217081612	217084915	0.489	75	0.711	0.342	67	0.561	0.078	46	0.186
<i>MRPL44</i>	2	224822121	224832431	0.410	67	0.657	0.585	87	0.739	0.588	93	0.696
<i>BTB</i>	3	15642864	15689147	0.025	22	0.123	0.016	24	0.072	8.95 x 10 <sup>-3</sup>	25	<b>0.039</b>
<i>ANKRD28</i>	3	15708743	15901053	0.015	17	0.092	0.012	22	0.061	1.45 x 10 <sup>-4</sup>	11	<b>1.41 x 10<sup>-3</sup></b>
<i>TOP2B</i>	3	25639396	25706396	0.098	35	0.306	0.064	35	0.201	0.017	32	0.058
<i>SEC61A1</i>	3	127770402	127790526	0.171	40	0.467	0.191	50	0.417	0.389	81	0.521
<i>ATP11B</i>	3	182511291	182639423	0.768	96	0.872	0.545	84	0.713	5.75 x 10 <sup>-3</sup>	24	<b>0.026</b>
<i>CCNI</i>	4	77968308	77998719	0.505	77	0.714	0.656	91	0.793	0.466	88	0.583
<i>CCNA2</i>	4	122737599	122745153	0.932	109	0.932	0.729	97	0.818	0.857	107	0.881
<i>CCNB1</i>	5	68462837	68474072	0.855	102	0.913	0.691	93	0.817	0.726	101	0.791
<i>SCAMP1</i>	5	77656339	77776562	0.095	34	0.305	0.079	38	0.229	0.158	56	0.307
<i>SMN1</i>	5	70220768	70248839	-	-	-	-	-	-	-	-	-
<i>ANKS1A</i>	6	34857038	35085802	5.36 x 10 <sup>-4</sup>	12	<b>4.87 x 10<sup>-3</sup></b>	4.90 x 10 <sup>-4</sup>	11	<b>4.90 x 10<sup>-3</sup></b>	1.57 x 10 <sup>-3</sup>	19	<b>9.07 x 10<sup>-3</sup></b>
<i>DST</i>	6	56322785	56819426	0.062	30	0.227	0.087	39	0.244	0.110	49	0.248
<i>EPB41L2</i>	6	131160487	131384462	0.041	26	0.168	0.068	36	0.209	0.136	52	0.283
<i>AKAP12</i>	6	151561134	151679694	0.318	53	0.640	0.496	78	0.700	0.288	70	0.449
<i>VIP</i>	6	153071932	153080900	0.575	81	0.774	0.151	44	0.371	0.416	84	0.545
<i>CBX3</i>	7	26240831	26253227	0.403	63	0.657	0.618	89	0.764	0.884	109	0.892
<i>SEPT7</i>	7	35840596	35946715	0.747	92	0.871	0.754	100	0.830	0.577	91	0.696
<i>CCDC136</i>	7	128431464	128462187	0.930	108	0.932	0.904	108	0.921	0.854	106	0.881
<i>ATP6V1F</i>	7	128502857	128505903	0.669	89	0.819	0.887	106	0.920	0.879	108	0.892
<i>PODXL</i>	7	131185021	131241376	0.482	74	0.710	0.731	98	0.818	0.323	74	0.481

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CNTNAP2</i>	7	145813453	148118090	0.139	37	0.409	0.099	41	0.265	0.120	50	0.263
<i>LINC00599</i>	8	9757574	9760839	0.919	106	0.932	0.539	82	0.713	0.157	55	0.307
<i>BNIP3L</i>	8	26239894	26270644	0.384	61	0.657	0.215	55	0.424	0.292	71	0.449
<i>CLU</i>	8	27454434	27472328	$2.56 \times 10^{-4}$	9	<b><math>3.10 \times 10^{-3}</math></b>	$3.46 \times 10^{-3}$	17	<b>0.021</b>	0.013	30	<b>0.048</b>
<i>CALB1</i>	8	91070836	91095107	0.329	56	0.640	0.374	69	0.596	0.356	77	0.509
<i>NTRK2</i>	9	87283417	87641985	0.042	27	0.168	0.036	30	0.133	0.201	61	0.362
<i>IARS</i>	9	94972489	95056038	0.645	87	0.808	0.317	63	0.554	0.582	92	0.696
<i>WDR31</i>	9	116075502	116102620	0.405	64	0.657	0.193	51	0.417	1.30E-03	18	<b>7.94E-03</b>
<i>CAMSAP1</i>	9	138700333	138799060	0.546	79	0.753	0.074	37	0.221	0.042	41	0.114
<i>GDI2</i>	10	5807186	5855512	0.017	19	0.099	0.041	32	0.142	$4.93 \times 10^{-3}$	23	<b>0.024</b>
<i>CELF2</i>	10	10838851	11378674	0.074	31	0.261	0.096	40	0.263	0.265	67	0.434
<i>RAB18</i>	10	27793103	27831166	0.310	51	0.640	0.044	33	0.148	$2.96 \times 10^{-3}$	21	<b>0.016</b>
<i>DDIT4</i>	10	74033677	74035797	$1.88 \times 10^{-7}$	3	<b><math>6.81 \times 10^{-6}</math></b>	$2.60 \times 10^{-8}$	2	<b><math>1.37 \times 10^{-6}</math></b>	$4.36 \times 10^{-7}$	5	<b><math>9.59 \times 10^{-6}</math></b>
<i>ADD3</i>	10	111765627	111895323	$1.68 \times 10^{-6}$	4	<b><math>4.59 \times 10^{-5}</math></b>	$3.74 \times 10^{-8}$	3	<b><math>1.37 \times 10^{-6}</math></b>	$2.69 \times 10^{-9}$	1	<b><math>2.96 \times 10^{-7}</math></b>
<i>COMMD3</i>	10	22605312	22609246	-	-	-	0.647	90	0.790	0.238	64	0.407
<i>SPON1</i>	11	13984184	14289679	0.149	39	0.417	0.020	25	0.088	0.012	28	<b>0.044</b>
<i>CNTF</i>	11	58390146	58393206	0.853	100	0.913	0.568	86	0.727	0.617	94	0.722
<i>MPEG1</i>	11	58975983	58980494	0.230	44	0.562	0.400	71	0.619	0.391	82	0.521
<i>NCAM1</i>	11	112831969	113149158	0.751	94	0.871	0.329	65	0.555	0.486	90	0.594
<i>ARHGEF12</i>	11	120207264	120360645	0.413	68	0.657	0.219	57	0.424	0.159	57	0.307
<i>APLP2</i>	11	129939716	130014706	0.232	45	0.562	0.466	73	0.700	0.786	103	0.835
<i>ACAD8</i>	11	134123428	134135749	0.595	82	0.780	0.528	81	0.713	0.195	60	0.358
<i>WNK1</i>	12	861759	1020618	0.240	46	0.568	0.166	47	0.388	0.393	83	0.521
<i>GABARAPL1</i>	12	10365435	10375727	0.622	86	0.788	0.214	52	0.424	0.361	78	0.509
<i>LYRM5</i>	12	25348150	25357949	0.408	65	0.657	0.609	88	0.762	0.485	89	0.594



Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>TMBIM6</i>	12	50135293	50158717	0.328	54	0.640	0.347	68	0.561	0.435	86	0.557
<i>SPRYD3</i>	12	53458100	53473204	3.64 x 10 <sup>-3</sup>	14	<b>0.028</b>	2.00 x 10 <sup>-3</sup>	14	<b>0.016</b>	0.012	29	<b>0.044</b>
<i>CNOT2</i>	12	70636774	70748773	0.048	28	0.187	0.041	31	0.142	0.020	34	0.065
<i>TAOK3</i>	12	118587606	118810750	0.855	101	0.913	0.824	103	0.880	0.373	79	0.519
<i>ZC3H13</i>	13	46528600	46626896	0.462	71	0.695	0.544	83	0.713	0.724	100	0.791
<i>RB1</i>	13	48877883	49056026	0.511	78	0.714	0.393	70	0.617	0.639	95	0.740
<i>PCDH9</i>	13	66876966	67804468	9.89 x 10 <sup>-4</sup>	13	<b>8.29 x 10<sup>-3</sup></b>	2.39 x 10 <sup>-4</sup>	8	<b>3.18 x 10<sup>-3</sup></b>	2.06 x 10 <sup>-4</sup>	13	<b>1.62 x 10<sup>-3</sup></b>
<i>RBM26</i>	13	79885962	79980393	0.222	43	0.562	0.147	43	0.371	0.052	43	0.134
<i>PSMB5</i>	14	23485752	23504429	0.083	32	0.281	0.215	54	0.424	0.032	38	0.092
<i>CNIH1</i>	14	54890279	54908322	0.146	38	0.417	0.027	28	0.105	1.18 x 10 <sup>-8</sup>	2	<b>6.16 x 10<sup>-7</sup></b>
<i>SNW1</i>	14	78183942	78227542	0.032	24	0.146	5.93 x 10 <sup>-3</sup>	20	<b>0.033</b>	6.71 x 10 <sup>-4</sup>	17	<b>4.34 x 10<sup>-3</sup></b>
<i>PAPOLA</i>	14	96968713	97033453	0.465	73	0.695	0.163	46	0.388	0.089	47	0.207
<i>STRC</i>	15	43891685	44002286	0.035	25	0.152	0.026	27	0.105	0.022	35	0.067
<i>CCNB2</i>	15	59397284	59417244	0.826	99	0.909	0.722	95	0.818	0.667	97	0.751
<i>ANP32A</i>	15	69070874	69113261	0.416	69	0.657	0.252	59	0.465	0.385	80	0.521
<i>IDH3A</i>	15	78441698	78462884	0.561	80	0.764	0.333	66	0.555	0.173	58	0.324
<i>TSR3</i>	16	1399241	1401873	0.408	66	0.657	3.34 x 10 <sup>-3</sup>	16	<b>0.021</b>	0.010	26	<b>0.042</b>
<i>ZNF267</i>	16	31885079	31928629	0.503	76	0.714	0.490	76	0.700	0.136	53	0.283
<i>E2F4</i>	16	67226068	67232821	0.316	52	0.640	0.178	49	0.399	0.010	27	<b>0.042</b>
<i>NFAT5</i>	16	69599869	69738569	0.053	29	0.200	0.021	26	0.091	0.017	33	0.058
<i>COX4I1</i>	16	85833173	85840608	0.804	98	0.895	0.901	107	0.921	0.241	65	0.407
<i>MPRIP</i>	17	16945790	17095962	0.015	18	0.092	0.030	29	0.114	0.060	45	0.146
<i>TVP23B</i>	17	18684458	18710027	0.359	58	0.657	0.121	42	0.316	0.031	37	0.092
<i>NLK</i>	17	26369009	26523407	0.767	95	0.872	0.737	99	0.818	0.669	98	0.751
<i>FAM134C</i>	17	40731526	40761445	0.260	47	0.592	0.217	56	0.424	0.174	59	0.324

Gene	Chr	Start	Stop	0 kb			50 kb			200 kb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>ATP5G1</i>	17	46970148	46973233	0.276	49	0.614	0.283	61	0.510	0.090	48	0.207
<i>COX11</i>	17	53029259	53046064	0.329	55	0.640	0.317	62	0.554	0.342	75	0.498
<i>DGKE</i>	17	54911460	54946036	0.871	103	0.913	0.967	110	0.967	$2.40 \times 10^{-6}$	6	$4.41 \times 10^{-5}$
<i>SEPT4</i>	17	56597611	56618179	$9.22 \times 10^{-11}$	1	$1.00 \times 10^{-8}$	$1.56 \times 10^{-9}$	1	$1.72 \times 10^{-7}$	$2.24 \times 10^{-8}$	4	$6.16 \times 10^{-7}$
<i>DHX40</i>	17	57642886	57685713	0.023	21	0.120	0.013	23	0.061	0.022	36	0.067
<i>DDX5</i>	17	62494374	62503042	0.748	93	0.871	0.673	92	0.805	0.421	85	0.545
<i>PRKAR1A</i>	17	66507921	66547457	0.093	33	0.305	0.011	21	0.055	0.036	39	0.103
<i>RBBP8</i>	18	20513295	20606451	0.196	42	0.507	0.236	58	0.448	0.344	76	0.498
<i>OSBPL1A</i>	18	21742009	21977846	0.294	50	0.640	0.560	85	0.724	0.789	104	0.835
<i>RAX2</i>	19	3769087	3772219	0.923	107	0.932	0.481	75	0.700	0.311	73	0.469
<i>CLIP3</i>	19	36505562	36523797	0.871	104	0.913	0.871	105	0.912	0.156	54	0.307
<i>SNAP25</i>	20	10199477	10288068	0.367	59	0.657	0.215	53	0.424	0.038	40	0.104
<i>SYS1</i>	20	43988660	44005442	0.465	72	0.695	0.324	64	0.555	0.270	68	0.436
<i>KCNB1</i>	20	47988505	48099181	$5.00 \times 10^{-3}$	15	<b>0.036</b>	$2.38 \times 10^{-3}$	15	<b>0.017</b>	$4.29 \times 10^{-4}$	16	$2.95 \times 10^{-3}$
<i>OSBPL2</i>	20	60813541	60871269	0.012	16	0.081	$1.54 \times 10^{-3}$	12	<b>0.014</b>	$1.81 \times 10^{-5}$	7	$2.84 \times 10^{-4}$

**E4: MAGMA gene-based association test results for the 111 candidate myopia genes using data from the UK Biobank spherical equivalent GWAS and the inclusion of 500 kb, 1 Mb or 2 Mb flanking regions.** MAGMA P = obtained from Z-Statistic under the assumption of a normally distributed model; Rank = ranking of the gene out of 111 according to gene-based test P-values; FDR = false discovery rate. Highlighted results = FDR < 0.05

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>MINOS1</i>	1	19923471	19956315	0.634	100	0.704	0.268	89	0.334	$8.16 \times 10^{-3}$	64	<b>0.014</b>
<i>CEP85</i>	1	26560644	26605529	0.016	42	<b>0.042</b>	0.050	62	0.089	0.059	80	0.082
<i>STK40</i>	1	36805225	36851525	0.080	59	0.149	0.144	81	0.198	$5.50 \times 10^{-3}$	60	<b>0.010</b>
<i>PTPRF</i>	1	43991708	44089343	$3.62 \times 10^{-6}$	11	<b><math>3.65 \times 10^{-5}</math></b>	$1.26 \times 10^{-5}$	17	<b><math>8.24 \times 10^{-5}</math></b>	$1.88 \times 10^{-9}$	14	<b><math>1.49 \times 10^{-8}</math></b>
<i>JAK1</i>	1	65298906	65432593	0.988	111	0.988	0.990	111	0.990	0.145	97	0.166
<i>SERBP1</i>	1	67873493	67896123	0.036	51	0.077	0.066	65	0.112	0.071	84	0.094
<i>ABCD3</i>	1	94883933	94984219	0.558	95	0.651	0.808	108	0.829	0.368	108	0.379
<i>WDR77</i>	1	111982512	111991915	0.021	45	0.051	$2.73 \times 10^{-4}$	25	<b><math>1.21 \times 10^{-3}</math></b>	$4.85 \times 10^{-14}$	6	<b><math>8.27 \times 10^{-13}</math></b>
<i>RPRD2</i>	1	150336624	150449042	$3.22 \times 10^{-4}$	20	<b><math>1.79 \times 10^{-3}</math></b>	$4.90 \times 10^{-4}$	31	<b><math>1.76 \times 10^{-3}</math></b>	$1.53 \times 10^{-4}$	40	<b><math>4.26 \times 10^{-4}</math></b>
<i>DENND4B</i>	1	153901977	153919162	0.080	58	0.149	$6.37 \times 10^{-3}$	46	<b>0.015</b>	0.014	68	<b>0.023</b>
<i>HDGF</i>	1	156711899	156722240	0.012	38	<b>0.036</b>	0.039	57	0.075	$1.37 \times 10^{-4}$	38	<b><math>4.01 \times 10^{-4}</math></b>
<i>SDHC</i>	1	161284166	161334541	0.458	90	0.564	0.022	55	<b>0.044</b>	$3.29 \times 10^{-3}$	57	<b><math>6.41 \times 10^{-3}</math></b>
<i>CREG1</i>	1	167510250	167523056	0.831	108	0.854	0.767	107	0.796	0.541	110	0.546
<i>KIFAP3</i>	1	169890461	170043882	0.210	71	0.324	0.072	67	0.119	0.135	95	0.158
<i>NUCKS1</i>	1	205681947	205719372	$4.58 \times 10^{-4}$	21	<b><math>2.42 \times 10^{-3}</math></b>	$1.71 \times 10^{-3}$	34	<b><math>5.59 \times 10^{-3}</math></b>	$2.48 \times 10^{-16}$	3	<b><math>9.17 \times 10^{-15}</math></b>
<i>RNF187</i>	1	228674277	228697598	0.311	77	0.448	0.453	98	0.513	$3.55 \times 10^{-3}$	58	<b><math>6.79 \times 10^{-3}</math></b>
<i>LGALS8</i>	1	236681514	236716281	0.011	37	<b>0.032</b>	$9.54 \times 10^{-3}$	49	<b>0.022</b>	$8.06 \times 10^{-5}$	35	<b><math>2.56 \times 10^{-4}</math></b>
<i>RAB10</i>	2	26256729	26360323	0.083	62	0.149	$6.30E \times 10^{-3}$	45	<b>0.015</b>	$7.82 \times 10^{-5}$	34	<b><math>2.55 \times 10^{-4}</math></b>
<i>LBH</i>	2	30454397	30482899	$7.64 \times 10^{-3}$	36	<b>0.024</b>	0.025	56	0.050	0.240	102	0.261
<i>LRPPRC</i>	2	44113363	44223144	$1.66 \times 10^{-7}$	7	<b><math>2.63 \times 10^{-6}</math></b>	$4.87E \times 10^{-9}$	8	<b><math>6.75 \times 10^{-8}</math></b>	$8.36 \times 10^{-9}$	15	<b><math>6.18 \times 10^{-8}</math></b>
<i>ERLEC1</i>	2	54014068	54045956	$5.00 \times 10^{-7}$	9	<b><math>6.17 \times 10^{-6}</math></b>	$5.37 \times 10^{-7}$	11	<b><math>5.42 \times 10^{-6}</math></b>	$2.27 \times 10^{-8}$	17	<b><math>1.48 \times 10^{-7}</math></b>

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CCDC104</i>	2	55746740	55772216	2.05 x 10 <sup>-8</sup>	5	<b>4.55 x 10<sup>-7</sup></b>	5.86 x 10 <sup>-10</sup>	6	<b>1.08 x 10<sup>-8</sup></b>	1.43 x 10 <sup>-15</sup>	4	<b>3.97 x 10<sup>-14</sup></b>
<i>HNRNPA3</i>	2	178076063	178088687	2.07 x 10 <sup>-6</sup>	10	<b>2.30 x 10<sup>-5</sup></b>	3.56 x 10 <sup>-26</sup>	1	<b>3.95 x 10<sup>-24</sup></b>	1.36 x 10 <sup>-25</sup>	1	<b>1.51 x 10<sup>-23</sup></b>
<i>STRADB</i>	2	202316392	202359501	0.448	89	0.559	0.519	99	0.582	0.253	104	0.270
<i>PKI55</i>	2	217081612	217084915	0.082	60	0.149	0.080	70	0.127	8.60 x 10 <sup>-3</sup>	65	<b>0.015</b>
<i>MRPL44</i>	2	224822121	224832431	0.655	102	0.713	0.717	106	0.751	0.219	100	0.244
<i>BTD</i>	3	15642864	15689147	6.45 x 10 <sup>-5</sup>	16	<b>4.47 x 10<sup>-4</sup></b>	3.09 x 10 <sup>-4</sup>	27	<b>1.27 x 10<sup>-3</sup></b>	0.014	69	<b>0.023</b>
<i>ANKRD28</i>	3	15708743	15901053	1.03 x 10 <sup>-4</sup>	18	<b>6.36 x 10<sup>-4</sup></b>	9.95 x 10 <sup>-4</sup>	33	<b>3.35 x 10<sup>-3</sup></b>	0.019	71	<b>0.029</b>
<i>TOP2B</i>	3	25639396	25706396	4.04 x 10 <sup>-3</sup>	31	<b>0.014</b>	5.43 x 10 <sup>-4</sup>	32	<b>1.88 x 10<sup>-3</sup></b>	1.59 x 10 <sup>-5</sup>	30	<b>5.87 x 10<sup>-5</sup></b>
<i>SEC61A1</i>	3	127770402	127790526	0.014	40	<b>0.038</b>	0.014	51	<b>0.031</b>	6.69 x 10 <sup>-5</sup>	33	<b>2.25 x 10<sup>-4</sup></b>
<i>ATP11B</i>	3	182511291	182639423	0.006	33	<b>0.019</b>	0.021	54	<b>0.044</b>	0.042	76	0.062
<i>CCNI</i>	4	77968308	77998719	0.421	85	0.539	0.041	58	0.078	1.50 x 10 <sup>-6</sup>	22	<b>7.56 x 10<sup>-6</sup></b>
<i>CCNA2</i>	4	122737599	122745153	0.839	109	0.855	0.649	102	0.706	0.057	79	0.080
<i>CCNB1</i>	5	68462837	68474072	0.699	105	0.739	0.429	96	0.492	0.555	111	0.555
<i>SCAMP1</i>	5	77656339	77776562	0.072	57	0.141	0.120	78	0.171	0.036	75	0.054
<i>SMN1</i>	5	70220768	70248839	0.409	84	0.539	0.814	109	0.829	1.10 x 10 <sup>-4</sup>	36	<b>3.32 x 10<sup>-4</sup></b>
<i>ANKS1A</i>	6	34857038	35085802	5.98 x 10 <sup>-4</sup>	24	<b>2.76 x 10<sup>-3</sup></b>	8.37 x 10 <sup>-5</sup>	19	<b>4.89 x 10<sup>-4</sup></b>	1.26 x 10 <sup>-6</sup>	21	<b>6.67 x 10<sup>-6</sup></b>
<i>DST</i>	6	56322785	56819426	0.097	63	0.171	0.099	74	0.149	0.073	85	0.096
<i>EPB41L2</i>	6	131160487	131384462	0.423	87	0.539	8.40 x 10 <sup>-3</sup>	48	<b>0.019</b>	1.51 x 10 <sup>-9</sup>	12	<b>1.40 x 10<sup>-8</sup></b>
<i>AKAP12</i>	6	151561134	151679694	5.89 x 10 <sup>-3</sup>	34	<b>0.019</b>	7.35 x 10 <sup>-3</sup>	47	<b>0.017</b>	0.026	72	<b>0.040</b>
<i>VIP</i>	6	153071932	153080900	0.625	99	0.701	0.674	105	0.712	0.166	98	0.187
<i>CBX3</i>	7	26240831	26253227	0.888	110	0.896	0.671	104	0.712	0.306	105	0.323
<i>SEPT7</i>	7	35840596	35946715	0.151	66	0.251	0.105	75	0.155	0.226	101	0.248
<i>CCDC136</i>	7	128431464	128462187	0.267	74	0.400	2.92 x 10 <sup>-3</sup>	42	<b>7.71 x 10<sup>-3</sup></b>	3.72 x 10 <sup>-4</sup>	41	<b>1.01 x 10<sup>-3</sup></b>
<i>ATP6V1F</i>	7	128502857	128505903	0.210	72	0.324	0.018	52	<b>0.038</b>	8.90 x 10 <sup>-4</sup>	49	<b>2.02 x 10<sup>-3</sup></b>
<i>PODXL</i>	7	131185021	131241376	0.071	56	0.140	0.051	63	0.090	3.70 x 10 <sup>-3</sup>	59	<b>6.96 x 10<sup>-3</sup></b>

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>CNTNAP2</i>	7	145813453	148118090	0.120	65	0.206	0.083	71	0.128	0.068	83	0.090
<i>LINC00599</i>	8	9757574	9760839	0.036	52	0.077	$2.52 \times 10^{-3}$	41	<b><math>6.82 \times 10^{-3}</math></b>	$9.58 \times 10^{-4}$	50	<b><math>2.13 \times 10^{-3}</math></b>
<i>BNIP3L</i>	8	26239894	26270644	0.205	70	0.324	0.172	82	0.233	$2.78 \times 10^{-3}$	56	<b><math>5.51 \times 10^{-3}</math></b>
<i>CLU</i>	8	27454434	27472328	0.051	54	0.105	0.019	53	<b>0.040</b>	0.092	89	0.114
<i>CALB1</i>	8	91070836	91095107	0.151	67	0.251	0.095	73	0.144	0.017	70	<b>0.027</b>
<i>NTRK2</i>	9	87283417	87641985	0.322	78	0.451	0.131	79	0.185	0.029	73	<b>0.045</b>
<i>IARS</i>	9	94972489	95056038	0.599	97	0.679	0.389	93	0.464	0.426	109	0.433
<i>WDR31</i>	9	116075502	116102620	$9.41 \times 10^{-4}$	25	<b><math>4.18 \times 10^{-3}</math></b>	$2.49 \times 10^{-3}$	39	<b><math>6.82 \times 10^{-3}</math></b>	0.047	78	0.067
<i>CAMSAP1</i>	9	138700333	138799060	$1.31 \times 10^{-5}$	12	<b><math>1.21 \times 10^{-4}</math></b>	$2.51 \times 10^{-3}$	40	<b><math>6.82 \times 10^{-3}</math></b>	$8.02 \times 10^{-4}$	46	<b><math>1.94 \times 10^{-3}</math></b>
<i>GDI2</i>	10	5807186	5855512	0.030	48	0.068	0.280	91	0.342	0.142	96	0.164
<i>CELF2</i>	10	10838851	11378674	0.393	82	0.528	0.073	68	0.119	0.250	103	0.269
<i>RAB18</i>	10	27793103	27831166	0.015	41	<b>0.042</b>	$3.73 \times 10^{-4}$	30	<b><math>1.38 \times 10^{-3}</math></b>	$6.19 \times 10^{-3}$	62	<b>0.011</b>
<i>DDIT4</i>	10	74033677	74035797	$1.29 \times 10^{-9}$	2	<b><math>5.52 \times 10^{-8}</math></b>	$2.89 \times 10^{-9}$	7	<b><math>4.59 \times 10^{-8}</math></b>	$1.84 \times 10^{-9}$	13	<b><math>1.49 \times 10^{-8}</math></b>
<i>ADD3</i>	10	111765627	111895323	$1.65 \times 10^{-8}$	4	<b><math>4.55 \times 10^{-7}</math></b>	$1.60 \times 10^{-6}$	14	<b><math>1.27 \times 10^{-5}</math></b>	$4.00 \times 10^{-6}$	27	<b><math>1.65 \times 10^{-5}</math></b>
<i>COMMD3</i>	10	22605312	22609246	$1.43 \times 10^{-3}$	29	<b><math>5.46 \times 10^{-3}</math></b>	$3.30 \times 10^{-4}$	28	<b><math>1.28 \times 10^{-3}</math></b>	$1.26 \times 10^{-3}$	52	<b><math>2.69 \times 10^{-3}</math></b>
<i>SPON1</i>	11	13984184	14289679	$2.50 \times 10^{-5}$	14	<b><math>1.98 \times 10^{-4}</math></b>	$1.22 \times 10^{-6}$	12	<b><math>1.13 \times 10^{-5}</math></b>	$3.16 \times 10^{-6}$	25	<b><math>1.40 \times 10^{-5}</math></b>
<i>CNTF</i>	11	58390146	58393206	0.528	93	0.631	0.586	101	0.644	0.046	77	0.066
<i>MPEG1</i>	11	58975983	58980494	0.566	96	0.654	0.430	97	0.492	0.012	67	<b>0.019</b>
<i>NCAM1</i>	11	112831969	113149158	0.422	86	0.539	0.203	85	0.265	0.088	88	0.111
<i>ARHGEF12</i>	11	120207264	120360645	0.176	68	0.287	$2.48 \times 10^{-3}$	38	<b><math>6.82 \times 10^{-3}</math></b>	$1.90 \times 10^{-6}$	23	<b><math>9.16 \times 10^{-6}</math></b>
<i>APLP2</i>	11	129939716	130014706	0.692	104	0.739	$9.28 \times 10^{-8}$	10	<b><math>1.03 \times 10^{-6}</math></b>	$7.09 \times 10^{-15}$	5	<b><math>1.57 \times 10^{-13}</math></b>
<i>ACAD8</i>	11	134123428	134135749	0.273	75	0.404	0.107	76	0.157	0.076	86	0.098
<i>WNK1</i>	12	861759	1020618	0.827	107	0.854	0.878	110	0.886	0.311	106	0.326
<i>GABARAPL1</i>	12	10365435	10375727	0.335	81	0.459	0.209	87	0.266	$8.20 \times 10^{-4}$	47	<b><math>1.94 \times 10^{-3}</math></b>
<i>LYRM5</i>	12	25348150	25357949	0.682	103	0.735	0.077	69	0.124	$2.51 \times 10^{-6}$	24	<b><math>1.16 \times 10^{-5}</math></b>

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>TMBIM6</i>	12	50135293	50158717	0.433	88	0.546	0.312	92	0.376	$9.61 \times 10^{-3}$	66	<b>0.016</b>
<i>SPRYD3</i>	12	53458100	53473204	0.027	46	0.066	0.207	86	0.266	0.065	82	0.088
<i>CNOT2</i>	12	70636774	70748773	0.083	61	0.149	0.047	61	0.085	0.167	99	0.187
<i>TAOK3</i>	12	118587606	118810750	0.038	53	0.080	0.113	77	0.163	$5.54 \times 10^{-3}$	61	<b>0.010</b>
<i>ZC3H13</i>	13	46528600	46626896	0.542	94	0.640	0.280	90	0.342	0.088	87	0.111
<i>RB1</i>	13	48877883	49056026	0.647	101	0.711	0.416	94	0.492	$7.69 \times 10^{-7}$	20	<b><math>4.27 \times 10^{-6}</math></b>
<i>PCDH9</i>	13	66876966	67804468	$5.14 \times 10^{-4}$	22	<b><math>2.59 \times 10^{-3}</math></b>	$2.05 \times 10^{-4}$	23	<b><math>9.91 \times 10^{-4}</math></b>	$5.47 \times 10^{-4}$	43	<b><math>1.41 \times 10^{-3}</math></b>
<i>RBM26</i>	13	79885962	79980393	$5.65 \times 10^{-4}$	23	<b><math>2.72 \times 10^{-3}</math></b>	$9.24 \times 10^{-6}$	16	<b><math>6.41 \times 10^{-5}</math></b>	$1.48 \times 10^{-4}$	39	<b><math>4.21 \times 10^{-4}</math></b>
<i>PSMB5</i>	14	23485752	23504429	0.118	64	0.205	0.061	64	0.105	$7.79 \times 10^{-3}$	63	<b>0.014</b>
<i>CNIH1</i>	14	54890279	54908322	$4.89 \times 10^{-18}$	1	<b><math>5.43 \times 10^{-16}</math></b>	$8.30 \times 10^{-16}$	2	<b><math>4.61 \times 10^{-14}</math></b>	$1.54 \times 10^{-12}$	10	<b><math>1.70 \times 10^{-11}</math></b>
<i>SNW1</i>	14	78183942	78227542	$1.42 \times 10^{-4}$	19	<b><math>8.30 \times 10^{-4}</math></b>	$9.11 \times 10^{-5}$	20	<b><math>5.06 \times 10^{-4}</math></b>	$8.50 \times 10^{-4}$	48	<b><math>1.97 \times 10^{-3}</math></b>
<i>PAPOLA</i>	14	96968713	97033453	0.013	39	<b>0.037</b>	$1.44 \times 10^{-4}$	22	<b><math>7.29 \times 10^{-4}</math></b>	$1.06 \times 10^{-3}$	51	<b><math>2.31 \times 10^{-3}</math></b>
<i>STRC</i>	15	43891685	44002286	$4.59 \times 10^{-3}$	32	<b>0.016</b>	$2.58 \times 10^{-4}$	24	<b><math>1.19 \times 10^{-3}</math></b>	$3.15 \times 10^{-5}$	31	<b><math>1.13 \times 10^{-4}</math></b>
<i>CCNB2</i>	15	59397284	59417244	0.028	47	0.067	$2.14 \times 10^{-3}$	37	<b><math>6.43 \times 10^{-3}</math></b>	$3.85 \times 10^{-5}$	32	<b><math>1.34 \times 10^{-4}</math></b>
<i>ANP32A</i>	15	69070874	69113261	0.599	98	0.679	0.539	100	0.598	$2.29 \times 10^{-3}$	55	<b><math>4.62 \times 10^{-3}</math></b>
<i>IDH3A</i>	15	78441698	78462884	$2.43 \times 10^{-3}$	30	<b><math>9.00 \times 10^{-3}</math></b>	$3.35 \times 10^{-10}$	4	<b><math>8.93 \times 10^{-9}</math></b>	$1.49 \times 10^{-13}$	8	<b><math>2.07 \times 10^{-12}</math></b>
<i>TSR3</i>	16	1399241	1401873	0.322	79	0.451	0.083	72	0.128	0.117	92	0.141
<i>ZNF267</i>	16	31885079	31928629	0.261	73	0.397	0.066	66	0.112	0.121	93	0.145
<i>E2F4</i>	16	67226068	67232821	$2.14 \times 10^{-5}$	13	<b><math>1.83 \times 10^{-4}</math></b>	$4.02 \times 10^{-10}$	5	<b><math>8.93 \times 10^{-9}</math></b>	$6.75 \times 10^{-11}$	11	<b><math>6.81 \times 10^{-10}</math></b>
<i>NFAT5</i>	16	69599869	69738569	$6.85 \times 10^{-3}$	35	<b>0.022</b>	$1.85 \times 10^{-3}$	36	<b><math>5.70 \times 10^{-3}</math></b>	$1.90 \times 10^{-8}$	16	<b><math>1.32 \times 10^{-7}</math></b>
<i>COX4I1</i>	16	85833173	85840608	0.394	83	0.528	0.422	95	0.492	$3.81 \times 10^{-4}$	42	<b><math>1.01 \times 10^{-3}</math></b>
<i>MPRIP</i>	17	16945790	17095962	0.033	50	0.073	0.142	80	0.197	0.097	90	0.120
<i>TVP23B</i>	17	18684458	18710027	0.053	55	0.108	$3.22 \times 10^{-3}$	43	<b><math>8.31 \times 10^{-3}</math></b>	$3.93 \times 10^{-6}$	26	<b><math>1.65 \times 10^{-5}</math></b>
<i>NLK</i>	17	26369009	26523407	0.296	76	0.433	0.179	84	0.237	0.064	81	0.087
<i>FAM134C</i>	17	40731526	40761445	0.017	43	<b>0.044</b>	0.011	50	<b>0.025</b>	$2.16 \times 10^{-3}$	54	<b><math>4.44 \times 10^{-3}</math></b>

Gene	Chr	Start	Stop	500 kb			1 Mb			2 Mb		
				MAGMA P	Rank	FDR	MAGMA P	Rank	FDR	MAGMA P	Rank	FDR
<i>ATP5G1</i>	17	46970148	46973233	1.49 x 10 <sup>-9</sup>	3	<b>5.52 x 10<sup>-8</sup></b>	1.40 x 10 <sup>-6</sup>	13	<b>1.19 x 10<sup>-5</sup></b>	5.63 x 10 <sup>-7</sup>	18	<b>3.47 x 10<sup>-6</sup></b>
<i>COX11</i>	17	53029259	53046064	0.181	69	0.291	0.178	83	0.237	5.92 x 10 <sup>-4</sup>	44	<b>1.49 x 10<sup>-3</sup></b>
<i>DGKE</i>	17	54911460	54946036	4.96 x 10 <sup>-8</sup>	6	<b>9.18 x 10<sup>-7</sup></b>	7.78 x 10 <sup>-11</sup>	3	<b>2.88 x 10<sup>-9</sup></b>	5.22 x 10 <sup>-14</sup>	7	<b>8.2 x 10<sup>-13</sup></b>
<i>SEPT4</i>	17	56597611	56618179	2.44 x 10 <sup>-7</sup>	8	<b>3.39 x 10<sup>-6</sup></b>	2.61 x 10 <sup>-8</sup>	9	<b>3.22 x 10<sup>-7</sup></b>	3.17 x 10 <sup>-13</sup>	9	<b>3.91 x 10<sup>-12</sup></b>
<i>DHX40</i>	17	57642886	57685713	0.030	49	0.069	3.35 x 10 <sup>-4</sup>	29	<b>1.28 x 10<sup>-3</sup></b>	6.26 x 10 <sup>-17</sup>	2	<b>3.47 x 10<sup>-15</sup></b>
<i>DDX5</i>	17	62494374	62503042	0.468	91	0.571	1.14 x 10 <sup>-4</sup>	21	<b>6.03 x 10<sup>-4</sup></b>	6.15 x 10 <sup>-7</sup>	19	<b>3.59 x 10<sup>-6</sup></b>
<i>PRKAR1A</i>	17	66507921	66547457	4.69 x 10 <sup>-5</sup>	15	<b>3.47 x 10<sup>-4</sup></b>	4.40 x 10 <sup>-6</sup>	15	<b>3.25 x 10<sup>-5</sup></b>	8.92 x 10 <sup>-6</sup>	29	<b>3.42 x 10<sup>-5</sup></b>
<i>RBBP8</i>	18	20513295	20606451	0.480	92	0.580	0.659	103	0.710	0.363	107	0.377
<i>OSBPL1A</i>	18	21742009	21977846	0.776	106	0.812	0.044	60	0.081	0.107	91	0.130
<i>RAX2</i>	19	3769087	3772219	1.06 x 10 <sup>-3</sup>	28	<b>4.21 x 10<sup>-3</sup></b>	3.30 x 10 <sup>-3</sup>	44	<b>8.33 x 10<sup>-3</sup></b>	1.11 x 10 <sup>-4</sup>	37	<b>3.32 x 10<sup>-4</sup></b>
<i>CLIP3</i>	19	36505562	36523797	9.87 x 10 <sup>-4</sup>	26	<b>4.21 x 10<sup>-3</sup></b>	3.06 x 10 <sup>-4</sup>	26	<b>1.27 x 10<sup>-3</sup></b>	6.44 x 10 <sup>-4</sup>	45	<b>1.59 x 10<sup>-3</sup></b>
<i>SNAP25</i>	20	10199477	10288068	0.018	44	<b>0.046</b>	0.043	59	0.081	0.036	74	0.053
<i>SYS1</i>	20	43988660	44005442	0.325	80	0.451	0.245	88	0.309	0.130	94	0.153
<i>KCNB1</i>	20	47988505	48099181	1.05 x 10 <sup>-3</sup>	27	<b>4.21 x 10<sup>-3</sup></b>	1.84 x 10 <sup>-3</sup>	35	<b>5.70 x 10<sup>-3</sup></b>	2.09 x 10 <sup>-3</sup>	53	<b>4.37 x 10<sup>-3</sup></b>
<i>OSBPL2</i>	20	60813541	60871269	9.10 x 10 <sup>-5</sup>	17	<b>5.94 x 10<sup>-4</sup></b>	2.22 x 10 <sup>-5</sup>	18	<b>1.37 x 10<sup>-4</sup></b>	6.51 x 10 <sup>-6</sup>	28	<b>2.58 x 10<sup>-5</sup></b>